Review of the study by Nuimura et al.

1. General comments

The study by Nuimura et al. presents a new glacier inventory for glaciers in High Mountain Asia derived from manual delineation of outlines on more than 220 Landsat ETM+ scenes from 1999-2003. The work behind GAMDAM is thus a tremendous collaborative effort filling an important data gap for this region. Unfortunately, the authors have decided to introduce a glacier definition that seemingly excludes all (?) steep rock walls (which is not what is recommended in the GLIMS analysis tutorial). As a first (1) general comment, this makes the new outlines incomparable to other existing datasets and is strongly limiting their usability for other applications when made available.

Another major issue (2) is the rather detailed comparison with the poor-quality glacier outlines of the RGI in this region. The RGI was only designed for global and large regional analysis rather than for analysis on glacier or catchment scale (see RGI technical note). High-quality outlines are available in several subregions for such an analysis from specific studies (and the RGI) and these should be used for a comparison (and cited by themselves).

For clean glaciers more accurate (and consistent) outlines can be derived from automated mapping (e.g. using simple band ratios). I recommend (3) using them for accuracy assessment of the manual digitizing and explain why automated processing was not applied here.

This brings me to two further critical points. All methods need to be properly described (4) for traceability. As an example: just writing (P2803, L13): 'Contour lines were used to identify glacier outlines' is not sufficient. I would like to know how this is working.

And finally (5), the quality control procedure seems to rely on one (?) person ('revised by a second operator') who - in the end - always knows where a glacier outline needs to be and can reshape them to the one and only true position. It is unclear to me how the judgement of this second operator can always be superior and 100% correct? In view of the examples shown on page 7 I have some serious doubts about the quality of the GAMDAM inventory.

I have some further points to criticise (see specific comments), with the above being the most general ones. Of course, it is now difficult to recommend 'Please do the inventory again and apply the correct glacier definition this time'. This might be senseless as the dataset (GGI) might be useful for its intended purpose, but it is important to pay attention to the other points, in particular explaining why certain decisions have been taken and how the applied methods work. As this study is describing a basic dataset (creation, accuracy, key parameters) I suggest moving Figures S1 and S2 (the revised ones) to the main text as they answer key questions about the approach taken. The comparisons with the entire RGI (which is frequently updated so the comparisons presented here will quickly be outdated) should be replaced with more detailed comparisons of independent studies (having provided high-quality datasets) and results from automated mapping. And please do not compare glacier numbers from different inventories. These are basically arbitrary numbers depending on several external factors (e.g. minimum size, drainage divide rules, separation of tributaries) with a very limited scientific meaning. I hope that my suggestions are helpful in revising the ms.

2. Specific comments

P2800

L6: This is not the reason; I think the explanation for the good match is that the same DEM is used here. In the void-filled regions of the SRTM DEM the geolocation of the Landsat scenes could be off by about 5 pixels (150 m) resulting in a relatively poor match with other geocoded datasets (e.g. the GDEM).

L16: For this reason more precise (regional) inventories should be used for a comparison (however, the here applied glacier definition requires to only compare complete glaciers).

P2801

L17: With this purpose in mind I would also include the upper parts of all glaciers as they might belong to their accumulation area (see page 8 example). In particular when later operating with elevation related variables (like in Sakai et al. 2014) the missing accumulation areas would cause a bias. How have glaciers with an interrupted profile been considered?

P2802

L1: This section should introduce the RGI and AGI datasets used for comparison. However, as mentioned above I recommend not using the entire RGI for this due to its obvious regional deficits and ongoing improvement. Please select high-quality outlines from individual (citable) studies (that can also be found in the RGI dataset).

L6: The DEM used for orthorectification is not everywhere identical to the void-filled SRTM DEM but is a merged product (called GLSDEM).

L11: This is fine but it requires that snow and cloud conditions are substantially different (what is difficult for orographic clouds and perennial snow fields). It needs to be described what the differences among the multiple scenes are and how they were combined. To me it seems that in many regions it was not possible to find scenes with appropriate snow conditions (despite the combination of scenes), as the 3-year period is simply too short for this. Just as an example: For the M. Everest scene 140-041 the supplemental xls table indicates that two scenes from 17.10. 2001 and 5.1. 2002 were selected. The first one shows severe seasonal snow cover hiding the real glacier perimeter. Why has this scene be used? Because of the smaller regions in shadow? Please explain the selection/combination process.

L14/15: I have no idea how this can work. Please explain it in the methods section.

P2803

L6/7: The discrimination of snow from clouds is working because of the strong absorption of ice/snow in the SWIR compared to clouds. The moderate absorption in the NIR has nothing to do with it (and actually snow reflectance is still high in the NIR).

L10: This is fine, but why has automated mapping not been used, at least to have an accurate base for the clean ice? This should be explained here. I assume it has something to do with the poor snow conditions in several images and the difficulties in interpreting them?

L13: As indicated above, this is not a description of a method. Please add all relevant details to understand how this is working (maybe better: do not use it).

L20-24: This is not the definition as given in the GLIMS Analysis Tutorial. First, all connected feeders (above the Bergschrund) need to be included, and second, also the unconnected glaciers in the steep headwalls are glaciers (maybe hanging glaciers). Finally, several glaciers might have interrupted profiles (e.g. due to a steep slope) and receive ice through avalanches. Of course, the upper parts of these glaciers have to be mapped as well. Only ice-free rock walls (and those covered by seasonal snow) need to be excluded.

L25: According to my experience, it is much more easy to delineate clean glaciers from the false colour composites (where glaciers appear light blue and thus have good contrast) and interpret debris from the true colour composites (also to have a better comparison with the high-resolution data available in Google Earth). In the example of Fig. S1 b) and c) the contrast issue is well visible: The glaciers in c) are clean but the ice is dirty (polluted) and has thus a much lower reflectance. By the way, also these dirty (but debris-free) glaciers can be accurately mapped automatically (e.g. with a simple band ratio).

P2804

L2: 'delineated by reference to topographic data'? How does this work? Does the 100 m SRTM DEM show the glacier boundary clearer then the Landsat imagery at 30 / 15 m resolution? This is hard to believe. Has a hillshade been used or just the contour lines mentioned before? In the latter case: How have they been used to decide where the boundary is? This needs to be described.

L4: What happens when Google Earth images are snow covered (which is often the case)? Or have all high-resolution images from this source been perfect for all decisions?

L6: What is seasonal ice cover?

L7: 'we referred to topographic data': this is not a description of a method. Apart from the fact that it is an unfortunate decision to define glaciers different from the standard by excluding their steep parts, it needs to be described HOW this method works (I can imagine a threshold value applied to a slope grid, but contour lines?).

L13ff: This section sounds like there is one person (second operator) knowing everything precisely and thus being able to always give advice for correct interpretation to all others. Given the sometimes wide range of interpretation that is possible (e.g. in cold-dry regions where debris-covered glaciers often have no clear boundaries to rock glaciers), I doubt that such a person (or several?) exists. The examples on pages 7 & 8 also illustrate severe difficulties in correct interpretation. In any case, a comparison with the outlines derived from automated mapping is missing (for clean glaciers) and should be provided. This can even be used as a reference dataset (for accuracy determination) as it is free of generalization effects. Furthermore, other regional studies should be considered (see P2802, L1).

L19/20: When specific surface features are obscured by shading, there is no need to assume that there is no ice underneath and exclude these regions. Very likely (as I can judge from Fig. S2a) the orange line is much closer to the correct outline than all others, i.e. that was not a misidentification of the glacier but the correct one.

L22/23: Where these tests performed independently? Please provide details of the method.

P2805

L1: This reads like the second operator has special knowledge that cannot be shared beforehand with those doing the work and that this expertise is always the correct one. Please show examples of what this second operator is correcting to learn from it (or even better: use the same rules for all operators).

L14: There could be a mismatch with the outlines derived from Landsat as the SRTM DEM with the wrongly interpolated data voids has been used for orthorectification.

L20-25: I would place this into the methods section. It is a description of how calculations have been performed.

L23: Hayakawa et al. have not investigated the performance over glaciers. This is in general a different type of terrain due to lack of contrast (snow), more gentle slopes and self similarity of surface features (debris). A subtraction of both DEMs should reveal which DEM is more appropriate for the specific purpose and provide better evidence for the selection.

P2806

L1-12: This discussion of DEM uncertainties is not really a result. I would either describe this when introducing the datasets or mention it in the discussion section. I am also not sure if this evaluation (ICESat comparison) really matters when considering the applications shown in Figs. 4, 7 and 12.

L13: I would not place this section before 4.3. The main results are in 4.3 and they should be described first. Section 4.2 itself starts with a description of methods (that might be better

placed in the methods section). I would move the reminder of it to the discussion section as it belongs to an overall evaluation of the results. But this is maybe a matter of personal taste. In any case, sections 5.1 and 5.2 are results of this study and have to be in the results section (with the suggested changes) rather than in the discussion.

L17: I suggest using the mean value of all digitizations as a reference for calculating the standard deviation. Otherwise it would imply that the digitiztions are not independent and one is always better than all the others (which seems to be confirmed in L21). When the quality in test 5 is very different from all others, I would assume that the rules for digitization have changed (?) and results are not comparable. Please clarify why this final test was superior in quality to all others and only compare what can be compared.

L25/26: If these would be 'real' uncertainties, the outlines would not be worth considering in an inventory (as they should be better than 5% in the mean). But as mentioned before, there is likely a bias in the calculation of the accuracy and this should be corrected first (applying the same rules for all digitizations and then use the mean value as a reference).

P2807

L4/5: How can median elevations be area-weighted? I would understand that only glaciers larger than a certain size are used to calculate a mean value (to reduce the influence of local topographic factors which have a stronger influence on small glaciers) but area-weighting?

L13: As mentioned above, I see section 5.1 and 5.2 of this study as results rather than a discussion and would suggest moving larger parts of it to sections 4.2 and 4.3 (the current 4.3 should be 4.1). As this would result in a missing discussion, the key findings (e.g. the differences between the compared inventories) should be critically assessed in a revised discussion section. I suggest including a discussion of uncertainties and how they impact on the results (e.g. how does median elevation change when 'correct' glacier outlines are used?), how glacier area changes due to a different interpretation of what a glacier is, and where the largest real differences in interpretation are (when comparing inventories of similar quality), among others (e.g. the derived topographic parameters).

L13ff: As mentioned before, please use regions for comparison that are worth a comparison rather than those who are wrong for obvious reasons (and do not compare numbers).

P2808

L11: The RGI has only assimilated the existing datasets rather than interpreting them (i.e. the obvious errors were in the source material).

P2809

L4/5: What has the spectral mapping of glaciers to do with the partitioning? The latter is performed with a DEM.

P2810

L21/22: For clean ice manual delineation is not better than automated methods but more inconsistent and not reproducible. The peer-review process is fine but intransparent (i.e. I do not understand how this works, see comments above). It needs to be explained where the first delineation failed and why the 'second operator' is always right with the interpretation.

P2810

L2/3: 'potentially accounts' sounds like if it is not clear that removing large parts of the accumulation area results in smaller glaciers. I assume that real changes of glacier size since the 1970s are comparably small?

L6: Fig. S1c is not about seasonal snow.

L6: Misinterpretation by whom? The RGI or this inventory? What about comparing glacier outlines in a region with good quality and snow conditions (see suggestions above)?

L12: Please be aware that the excluded headwalls also include glacier (parts) under the seasonal snow (see page 8 example).

L13: I do not understand this sentence. What is meant by 'projections of mass balance by in situ observations'? Does this refer to differences in calculated mass changes due to the different techniques of spatial interpolation and averaging applied? How can exclusion of headwalls improve this? As far as I know, these regions are filtered when altimetry is used, are seldom measured in the field, and have small changes anyway.

Tables

Table 1: Please replace 'Number of excluded small glaciers' with 'Excluded glaciers'

Table 2: Please use the abbreviation AGI also here (instead of the citation). The last column is not the difference in percent (that would be the difference in km^2 divided by the total area), but a normalized value (with negative values being changed to positive). If this way of presenting differences should be retained, I suggest using +4 instead of 104 and -4 instead of 96, etc. This would allow for a more easy comparison. It also needs to be shown on a map (in a new Fig. 1) where these subregions refer to (catchment boundaries). The further study sites for area comparison should be marked in this new Figure as well. And please use a more descriptive caption (see Table 3 comment).

Table 3: Please use a more descriptive caption, e.g. 'Comparison of regionally aggregated total glacier areas from Bolch et al. 2012 and the GGI'.

Figures

Please insert an overview Figure showing important subregions/test sites and outlines of catchments (listed in Tables 2 and 3).

Figs. 2a/b, 3, 5, 6, 8, 9 (with their respective revised content): Please add minor tick marks and show them on both sides. I suggest placing the a), b) etc. annotation outside the plot.

Figs. 10 and S1: Green on light blue is difficult to see, please use another colour (yellow?).

Fig. 2: As a justification for selecting a specific DEM, I suggest to simply subtract the GDEM from the SRTM DEM, add a colour coding in classes of standard deviation, and a layover of glacier outlines (of course, this could only be shown for a subregion). The comparison with ICESat is interesting in general, but in the framework of this study I would suggest showing something more relevant (glacier specific).

Fig. 3: This plot should be recalculated, after applying the same rules to all tests (also number 5) and using the mean area as a reference.

Fig. 4: This plot might change when 'correct' outlines are used. The impact of such a change should be determined for a test region and discussed in the main text.

Fig. 5: I suggest removing this figure. For the GGI vs AGI comparison I would like to see an example (close-up with a few glaciers) showing an overlay of outlines.

Fig. 6: When retaining this figure, I suggest removing the comparison with the entire RGI and focus on more regional comparisons with other high-quality datasets. When hypsometry is shown, it would be nice to indicate how the sampling is done (e.g. in 100 m bins?).

Fig. 7: I suggest removing the comparison with the RGI dataset and instead of c) showing the difference between b) and c)

Fig. 8: General remark: For scatter plots with a high correlation, it is most often more insightful to plot the difference between the values vs the values. As mentioned above, I see little value in comparing glacier numbers (what is the message here?) and I would not compare areas or elevations of the entire RGI. Please use a few subregions where the RGI data are of sufficient quality and compare these. The effect of not considering parts of the accumulation area on median glacier elevation should be analysed in detail, in particular when it is foreseen to use this parameter for climatic interpretation (see Sakai et al. 2014).

Fig. 9: I suggest removing this Figure.

Fig. 10: The comparison with high-quality outlines as shown in a) worries me. I have compiled comparisons with high-resolution screen shots from Google Maps on page 7, showing that the GGI quality is partly rather poor compared to the RGI. It seems that GGI is highly generailzed (like the DCW) and that the second operator has also problems in interpreting glaciers correctly. The poor-quality regions shown in b) and c) have already been documented by Pfeffer et al. (2014) and have been revised in the mean time (for RGI 4.0). There is no need to show them here. The red line on d) is certainly not correct, but why has the glacier in the upper centre and at the Lhotse westface been removed? Please also check the Mt. Everest map from swisstopo or the National Geographic Society for a correct interpretation of glacier extents in this region and see comparison on page 8. Caption: I think the normal way of providing geographic coordinates is latitude, longitude.

Figs. 11/12: I suggest removing these figures.

Figs. S1 and S2: Some coordinates must be given for all images to have a chance to find these example glaciers (as in Fig. 10). As mentioned above, please integrate these figures in the main text. This is the key work of this study and should be properly documented.

Fig. S1: a) Where is the not included steep headwall and the glacier (outline)? b) How is snow discriminated from clouds then? c) glaciers are not debris covered but dirty (i.e. it is clean ice for automated mapping), d) where is the glacier outline here? (this is the interesting stuff!), e) as the thermal signature is not unique and the spatial resolution is much coarser, how is this compensated?, f) where is the glacier in the right hand image?, g) the left image has lots of snow, please adjust the caption for f) and g) to be clear what are the issues here.

Fig. S2: The orange line might be more correct than all others (see example on page 8). Is there an image available showing that there is bare rock under the snow?

Typo issues

P2801, L4: A glacier inventory ...
P2801, L15: Pfeffer et al (with f)
P2801, L19/21: over high mountain Asia (remove 'the', here and elsewhere)
P2802, L7: We selected Landsat scenes
P2802, L9: Reference System 2 (WRS1 is for MSS)
P2804, L17: were delineated differently (not all are inaccurate)
P2806, L4/8: biases: I would call these errors
P2809, 16/17: DN is a unfortunate abbreviation as it is already occupied by 'digital number'.
P2818, caption: 'Footprints of Landsat scenes ...'
P2827, caption Fig. 10 (last line): These are false-colour composites

Cited reference

Sakai, A., T. Nuimura, K. Fujita, S. Takenaka, H. Nagai and D. Lamsal (2014): Climate regime of Asian glaciers revealed by GAMDAM Glacier Inventory. The Cryosphere Discuss., 8, 3629-3663.

With reference to Fig. 10a (centre) I would like to illustrate here (Google Maps screenshots) that the GGI missed many glaciers that are correct in RGI and mapped glaciers where there are no glaciers (yellow example). Please compare yourself and explain your interpretation.

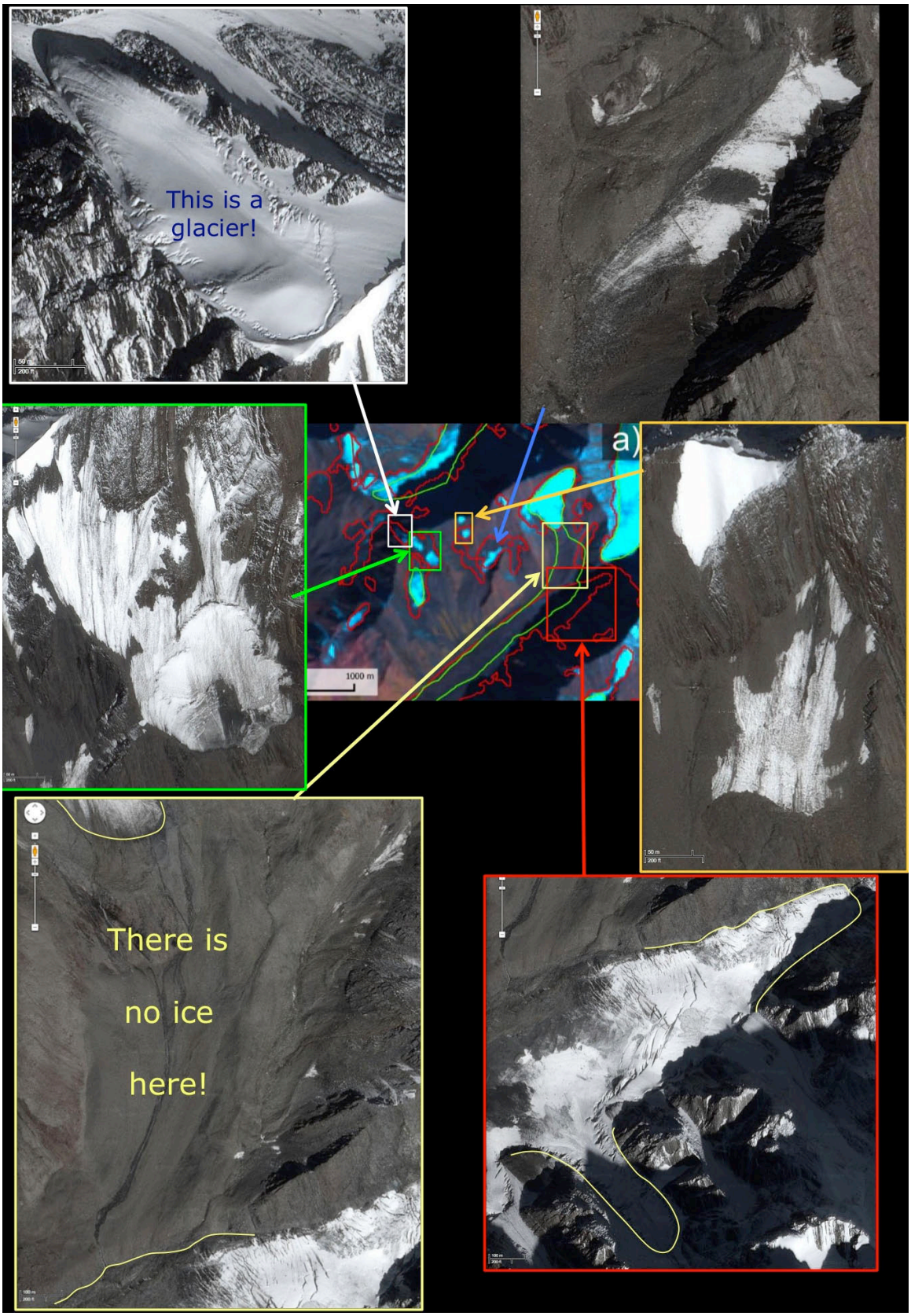

Fig. 10d: It seems that the red line is more correct than the green one (e.g. in the circles).