Review of the study by Smith et al.

General comments

The study by Smith et al. aims at presenting glacier area change rates from the Pamir to the Tien Shan as well as revealing a potential relation of the area changes to regional gradients in precipitation. In my opinion the latter goal is a misconception. A decreasing precipitation trend will results in an upward shift of the ELA, resulting in smaller glaciers if mountains are not getting higher at the same time. As relative area change rates generally increase towards smaller glaciers (shown by numerous studies), higher area reductions from W to E will automatically occur in case glacier size decreases from W to E, i.e. gradients in precipitation are not required to explain a possible trend in area changes. Which basically means that the authors can safely remove all the climate sections and focus on a more thorough analysis and improved presentation of the area changes. I suggest removing the climate analysis also due to the unclear physical relation between climate change and glacier area change. I mean glaciers have a response time and a highly variable thickness distribution, both governing (together with mass balance and flow dynamics) how the area of a glacier changes with time. As nothing of this is investigated here or considered in the discussion, it reinforces to just remove the climatic interpretation of area changes. Finally, I wonder why glacier sensitivity to climatic changes should increase towards drier climates? I so far thought it is vice versa.

Before I list my specific comments below, I would like to elaborate on some more general aspects of the ms that I hope make clear what the larger problems are. As a starting point, my general impression is that the study wanted too much at the same time:

1) Presenting in rich detail a new method of digital image processing for glacier mapping
2) Analyse the results of the method in detail for conclusions about its performance
3) Process a large number of satellite images with this method and derive (area) changes
4) Present aggregated results of the area changes (at least they appear in the discussion)
5) Provide a broad overview on available climate data and trends for the entire region
6) link 4) to 5) using a statistical analysis of trends (but lacking a physical reasoning)

In this endeavour the authors simply lost track of what is important and what not, and this concerns both content and the form of presentation. As an example: glacier classification with the band ratio is basically a one-liner. What is the improvement over previous works when I have to read and understand five pages of complex image processing details (5442-5446), but still get results that have to be controlled for each glacier and adjusted manually? I have no qualms in learning about a new image processing method that outperforms previous methods. This is how it should be. However, in my opinion it should either provide better results or be less complex when the quality is similar and the steps and improvements need to be adequately shown. Figures 2 to 4 provide little help in this regard: The depicted region is too large to clearly see (and discuss) the effects, the steps are only given as text rather than supported by figures (I really want to see a velocity map and the effects of thresholding, as this is new), high-slope areas of glaciers are removed without any comments, the clean ice classification looks poor (Fig. 3), and the generated outlines are noisy and have to be corrected anyway (Fig. 4). Why should the method be used when it requires such a high amount of additional calculations and input datasets (velocity) when the outlines need to be corrected anyway? Why should others use them when the uncorrected ones have poor quality, a large number of glaciers are missing completely and others lost their steep parts or tongues (see my summary figure below)? There could be reasons for all this but they need to be explained.

The contents of the ms has some further flaws in this direction: On P5443/44 is a lengthy description of two methods 'although neither provided significant improvement ...', so why they are presented? Why confirming 'that the RGI was not designed for detailed glacial comparisons' (P5440), but then doing it anyway (Figs. 5 / 6)? Why presenting an analysis of area mismatches vs topographic indices (Fig. 8) when there is no correlation (should there be any?)? Why calculating a regression equation for the area changes of an individual glacier (Figs. 10/11)? I mean the outlines seem to be wrong (missed debris-cover), the changes are 0.5 and 0.1%, and confidence intervals are given that have no relation to anything (same for the median elevation: why is it provided?). Which brings me to the general point that the study often describes unnecessary (and partly strange) details but misses to present what is really of interest (see examples below). I recommend removing the unnecessary parts (climatic analysis), focus on one or two subjects and present these in detail with supporting figures.

Another higher-level critique is the organization of the ms. I am aware that different approaches are possible to present background/motivation, study region/datasets, methods, results, discussion and conclusions, but they should align with the aims of the study. From the title and introduction I would conclude that deriving area changes with a new multi-method approach is a main objective of the study. But the results section only provides a lengthy discussion about algorithm errors (unfortunately only by statistical measures rather than by overlay of outlines from other sources and a related discussion about the differences), while the discussion presents the results of the area changes but starts with mentioning changes of two individual glaciers without presenting the numbers, continues with a wordy description of what has not been done (please do it, this is the important part for glacier specific change assessment), before section 5.2 finally lists some numbers. Please present results in the results section (including the key numbers) and discuss these in the discussion section. Just writing (in 5.2.1) 'These values are presented in Table 2' and then leave it to the reader to analyse why these numbers are there and what they mean might distract readers. And please only compare (i) annual relative area change rates across regions and (ii) data with a temporal overlap. It makes little sense comparing the data that are currently compared in Table 2. The following more specific comments partly come back to the above points (with line numbers).

Specific comments

P5434

L2/3: 'decipher' is really ambitious, I suggest starting with an investigation of mass balance and climate. Please be sure that this will likely never be possible with area changes

L6/7: Here and everywhere else: Please replace all glacial with glacier when speaking about contemporary glaciers.

L11: Here and everywhere else: Please replace all 'the authors' with 'we'. It reads like it refers to somebody else.

L16: Please divide all glaciers into entities and determine their changes collectively. Just using the disconnected one does not work.

L20/1: 'thermally insulated' is generally used for debris covered glaciers. What does 'some of the impacts of changing temperatures' mean? What are the impacts? Should 'changing' be increasing?

P5435

L2: 'understanding of changes': what changes, area changes? What have these to do with 'sustainable water management' and what are the impacts requiring mitigation? I mean runoff in many glacierized regions is increasing despite area shrinkage. What is the connection here?

L5/6: length changes? Please note that this study is about area changes. Why 'though'?

L9: Can it be possible that the authors mix up length and area changes? Please be precise in the terminology.

L11: Why 'attempts'? I mean the data are there and can be used.

L20-23: Glacier outlines are created in time intervals of a few decades to upscale the datasets with a better temporal resolution but limited spatial coverage to the entire mountain range. In general, they are not used for tracking annual changes (in length or area). Please also note that many glaciers change (here: advance/retreat, i.e. length) much less or much more than the given 15-30 m / yr. The required inventory update rate is thus variable while the 1-2 pixel accuracy is relatively fixed (and applies also to other sensors).

L24: Glacier mapping with a band ratio was originally proposed by Hall et al. (1987).

P5436

L16: It seems that basically only velocity is new here, please illustrate how it is working!

L23: Why have the Landsat scenes to be geo-referenced and co-registered? I mean the L1T product from USGS is orthorectified and over glaciers mostly accurate within 1 pixel.

L25: What are 'trends in glacial character'? Please be precise. And what is a 'link' when R^2=0.02?

P5438/39/40

It is interesting to see that after the detailed description of satellite and climate model derived datasets the authors note in L3 (P5440) that climate station data (at high elevations) do not exist in the study region. And at the same time it is assumed that reanalysis data are able to resolve regional variability and trends in steep high mountain topography? They might be physically consistent, yes. But can they providing anything reliable for such an analysis? I suggest to just skip the climate analysis, also for the other reasons listed above.

P5440

L22/3: 'primarily': and if seasonal snow is present? Please describe and illustrate it, this is the interesting part. Is the 10% cloud cover over glaciers? What has been done in this case? When it refers to the entire scene, it is useless information as a scene can have 90% cloud cover and all glaciers are cloud free (e.g. above a fog layer).

P5441

L8: As above, I do not understand why the already orthorectified scenes have to be 'georeferenced and registered'.

L23: 'can be used': Why can? Have they or have they not? I mean this is the methods section where the applied methods should be described. So please describe how this works.

P5442

L3: In general, even neighbouring lakes can have completely different spectral properties, e.g. depending on the sediment load. How can this range be properly mapped with one scene-specific threshold?

L11/12: The threshold value for the band ratio has to be selected scene-specific and might vary from about 1.5 to 2.5. What does TM1 > 250 mean? TM1 is used to improve classification in shadow by separating ice/snow in shadow from bare rock in shadow. Typical threshold values are around 60 (+/-20); 250 seems to be snow in sunlight (which is well mapped by the band ratio)?

L16: Why are they removed? Lakes on a glacier (used above as seed points) are certainly part of the glacier and ice in shadow has to be included as well. Please adjust the glacier mapping accordingly.

P5443

L3: Please describe how this 'additional thresholding' works.

L8: Please describe (in section 2?) which image pairs were used to create the velocity maps (not all have OLI pairs), how the maps look like (for a difficult example) and how the binary mask looks like that is used to improve the mapping.

L26: When neither provided improvement, why describe these methods? Please remove.

P5444

L20: What is meant with 'velocity profiles', velocities? What is 'very different' (5, 50, 250 m/yr)?

L26: Please show this network and illustrate the method! I have no idea how it looks like or why the method works.

P5445

L6: Please show that it is effective.

L11: Which comes back to the point that the analyst has to carefully go through all glaciers and decide if they need additional seed points or not. Which I think is in contrast to an effective (and improved) algorithm (i.e. thousands of glaciers might have to be checked).

L17: section 3.3.4: Please show how these different masks work.

L18: 'are generally accurate': What does this mean? Please provide details.

P5446

L4: Please illustrate the effects of the different filters with an example. A 5x5 filter has already a quite heavy impact on the size of small glaciers.

L9: Which metadata is added? How is it added to each glacier when glacier complexes are not separated into individual glaciers?

L18: What is the difference between the two datasets, what is the base for the digitization, how large are the glaciers in this sample, are debris covered glaciers included, where are these outlines shown (along with an overlay of the automatically generated outlines), why have errors not been calculated with reference to this dataset (i.e. where does the 2% come from)? Please note that the description given here is not sufficient.

L25: Ice caps? Where are ice caps in the study region? Why does snow cover connect glaciers (scenes should be free of snow outside of glaciers)?

L24/25: 'contiguous glacial areas': when they are connected by seasonal snow, they cannot be glacial, the term used for the RGI is 'glacier complexes'. What are 'component parts'? Individual glaciers?

L26: Were watershed boundaries (i.e. drainage divides) only derived for the 2 x 750 glaciers in the control datasets? Why not for all others?

P5447

L4: The entire error analysis section has to be revised as all calculations are based on a glacier definition that is inconsistent with the common understanding of a glacier.

L4: The results section should present the results of the study in regard to the research question (i.e. the area changes presented in 5.2, but with more details).

L9: No, the RGI is not accurate when it comes to the comparison of only 138 glaciers. This is in contrast to its intended purpose. Please show the used outlines with an overlay (the hypsographic plots do not carry any relevant information in this regard) and apply the same glacier definition as in the RGI. It makes no sense to compare hypsographic curves when the dataset created here has removed large parts of the accumulation region.

L15: Where does this conclusion come from? Why does the RGI universally overclassify glacier areas? I mean the here-applied classification arbitrarily removes large parts of the accumulation area of glaciers as well as entire glaciers (see examples in the annotated figure below). Could it be possible that the method applied here universally underclassifies glacier area because the authors apply a wrong glacier definition?

P5448

L1ff: Have the authors really assumed that their dataset is correct and the other one is wrong? This is hard to believe.

P5449

L9: Please use 'we' rather than 'the researchers'.

P5450

L6-8: This is the more general statement, it should be in the beginning. What comes afterwards belongs to the methods and/or datasets description. Please note when writing 'and not in assessing trends within individual glaciers' and then presenting next the results for two individual glaciers (L2), there is a mismatch between what is written and what has been done.

L11: 'require significant manual work': Indeed, this is the point. When there is an easy way of doing it, somebody would likely already have done it. To advance science somewhat, I recommend doing it, at least if the authors decide to refine their glacier definition.

P5451

L1: But it should be possible to provide the time period covered for each scene and a mean annual rate that is comparable across regions.

L16: I fear they are not comparable as individual glaciers have been selected and sometimes only parts of a glacier are considered. For a realistic comparison it is at least required to apply a consistent definition of a glacier.

L23: OLI performs better in images with clouds and snow cover? How is this possible? Does OLI see through clouds and detect the glacier boundary under snow cover?

P5452

L2: 'best suited images': Sure? The example in Fig. 3 shows way too much seasonal snow for deriving accurate glacier outlines.

L3: The statistical analysis presented would even if the correct glacier definition had been used not provide any meaningful assessment of the accuracy of the glacier outlines. For this it is required to show overlays with comparable, accurate and independent datasets and determine glacier area differences in a quantitative way (mean difference, standard

deviation etc.). For outlines that have been corrected manually, it can be recommended to perform three independent digitizations of >10 glaciers and compare these.

L4: I would argue here that everything is accurate in large quantities, in particular when errors have a normal distribution. But I think it is more important that the accuracy of individual glacier outlines is better than 5%. This needs to be shown. Please note that the 3.9% area change reported here is not significant compared to the accuracy (also the potential one) of the dataset. Relative area changes have to be higher for sub-regions than this to be suitable for a trend analysis.

L22: It is interesting to see that over the past few years several studies have claimed that area changes and mass balances can be directly compared and any change in precipitation or temperature will have an immediate impact on area changes. So far I thought that glaciers have a response time (typically of a few decades) and that observed changes in length relate to a climate forcing long ago? For area changes I think one can say that glaciers will shrink when temperatures are rising, but a physical relation is yet missing.

P5433-5456

L1ff: I will not comment on these sections in detail as they should be removed from the study for the reasons discussed above. The statements are partly missing a scientific base (e.g. P5433, L13-19), in particular when considering how the dataset was created and how small the observed changes are.

P5457

L6-8: Fig. 12 does not show 'glacial statistics' at the watershed scale (such as Figs. 5 & 6), it shows (very uncertain) area changes of individual glaciers plotted against their longitude. Figures 10 and 11 (coordinates in F11 are wrong, where is this glacier terminus?) nicely illustrate what the problem with the area changes is: the different outlines look random or maybe like wishful thinking. On what base have they been placed there? As debris cover has not been mapped, the real glacier extent is rather different. I can imagine that these glaciers have not changed at all when debris is correctly considered.

L11: 'the algorithm performs well on debris-covered areas': I disagree from what I see on Fig. 4. Moreover, also the manual editing fails to map the tongues correctly (see examples in the figure below and Figs. 10 and 11 in the ms).

L14: 'step forward': I disagree here as well and think the algorithm is a huge step back. Not the idea itself, but the result. I think adjusting the definition of a glacier to the output of an algorithm to confirm that the methods works well is a bad idea. Any chance to modify the algorithm in a way that the result is compliant with a standard glacier definition?

L17: And here finally it comes, the algorithm was not designed to map individual glaciers precisely. This gives some food for thought. I mean what else should the purpose be? We already have low quality datasets with a rather limited use, why add another one? Should it not at least be sufficiently accurate to perform reliable change assessment (i.e. a key purpose of this study)?

L19-23: It would have been nice to describe snow and cloud conditions of the scenes used in this study. This generic statement is rather useless in this regard.

L24: 'generally too low': MSS has no SWIR band. How should the algorithm work at all?

L26: Once drainage divides are generated, they can be used again and again. If appropriate images are used, there is no need to manually correct outlines. And where are the ice caps (apart from Gregoriev) and why should they be divided?

L27ff: Is the examination of changes of individual glaciers not the purpose of this study (see Fig. 12)? I think this is not a caveat of the algorithm, but of the missing post-processing.

L5448

I think I stop here. It should be clear from my above comments that the conclusions make little sense. When going big is not feasible within the project (high quality outlines for all glaciers in the study region), I recommend reducing the size of the study region and add a detailed analysis of the results.

Tables

Table 1: Scene IDs are nice for a database but hard to decipher for humans. Please provide sensor and date instead (when path row is given on top). A comment on snow and cloud conditions for each scene would be helpful for identification of problematic regions.

Table 2: Please check if numbers can be compared at all in regard to the temporal differences. Provide annual change rates and the number of glaciers analysed for each region.

Figures

Fig. 1: A part of the top can be cropped, maybe add path and row to the Landsat footprints.

Fig. 2: This is more text than figure. Maybe move the text into numbered bullet points in the main text and add for each major step a figure illustrating what happens. I also suggest using a smaller region for illustration to better see the changes. A few coloured outlines on the same image might also enhance visibility of the changes in each step.

Fig. 3: What is the difference between the upper Fig. 2 and this one? Please remove when they are identical. Overlays with the reference dataset (used for accuracy assessment) should be shown somewhere.

Fig. 4: see comments in the example figure below. The caption should read "Landsat ETM+ panchromatic band acquired on xx is shown in the background".

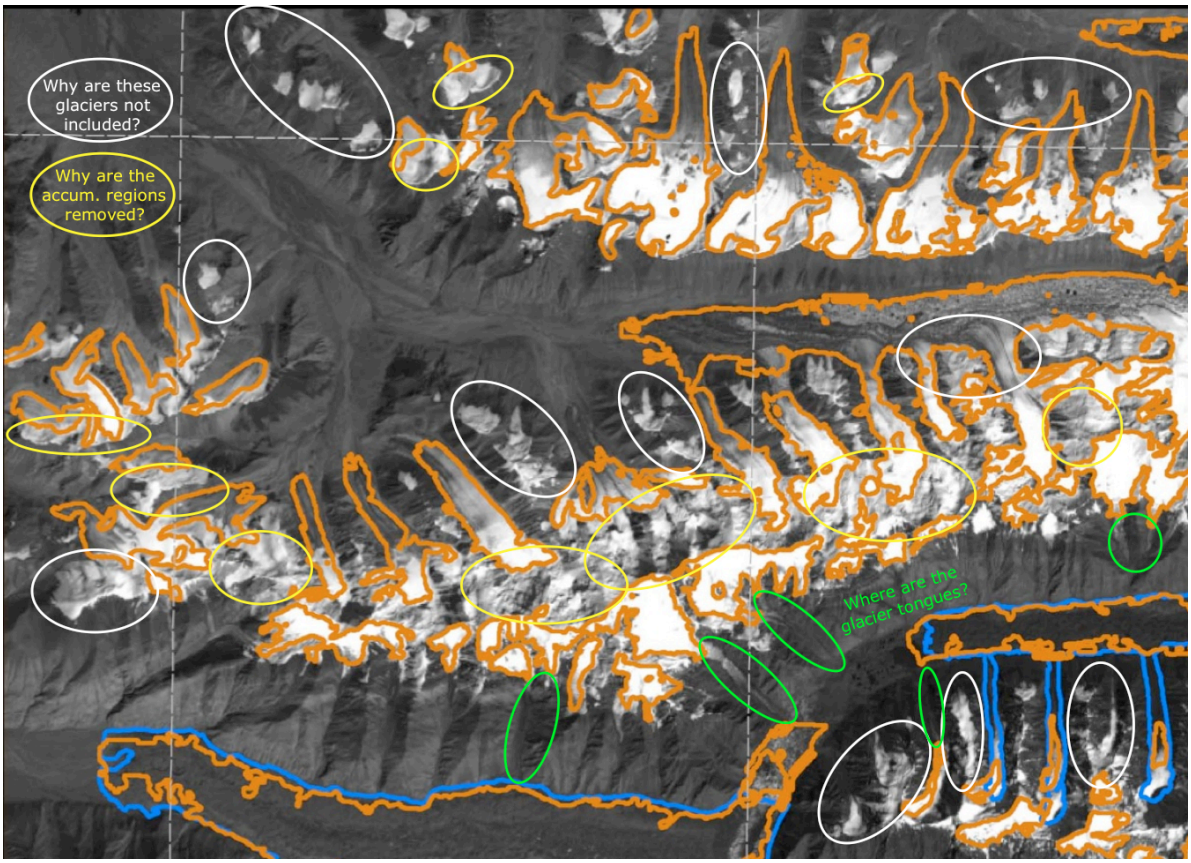
Fig. 5/6/7/9: I think these lumped hypsographic plots say nothing about the accuracy of the created dataset (apart from the gross misinterpretation of glacier extents by the authors). Once glaciers are mapped correctly, outline overlays should be used to illustrate the accuracy of the method and mean differences / standard deviations should be calculated.

Fig. 8: I suggest removing plots that have no meaning. It would have been fine to just say that a correlation was not found (at least when it has been made clear why there should be a correlation). And please use relative area differences to have comparable results.

Fig. 10/11: I have explained above why I think these figures are free of useful information. Please remove the scatter plots with the confidence bounds and use examples where a change can be well followed over a somewhat larger region. These examples just demonstrate that the authors have problems in correctly delineating debris-covered glaciers.

Fig. 12: The graph has +/- values, it is thus a change rate rather than a retreat rate, it should also be named area change rate and values should be given in percent rather than km^2. But apart from this I think the plot does not carry any useful information and can be removed.

Figs. 13/14: Please remove the climatic analysis and focus on the area changes.



Example figure: Selected examples (marked by white, yellow, and green circles) of wrong classifications (from Fig. 4 of the study).