Reply to Editor Comments – Smith et al. – Glacier Mapping in the Tien Shan

Focus on the glacier mapping algorithm only

The sections on the application of the algorithm for change detection, as well as our interpretation of possible links to changing regional climate, have been cut. These sections have been replaced by enhanced treatment of the glacier mapping algorithm, with more illustrative images (Figs 2-8, 13).

Clearly show and discuss the importance of each different processing step for improving the results

We have expanded our in-text explanation of each step of the algorithm in Sections 3.1-3.4, as well as added several new images (Figs 2-8) to illustrate what the delineated areas look like after each step of the algorithm. The final result is compared to the initial spectral classification in Figure 8, and compared with other datasets (manual and RGI) in Figure 13.

Show and discuss the influence of different thresholds (e.g. band ratio, TM 1 ratio, velocity) and clearly justify your chosen ones

We have updated our threshold discussion in Sections 3.3, 3.4.1, 3.4.2, and 3.4.4, and have included images that show the extent of classified and unclassified areas after each step (Figs 2-8). Included here as Figure 1 of this reply is an example velocity map, classified to show values between 0-10 m/yr.

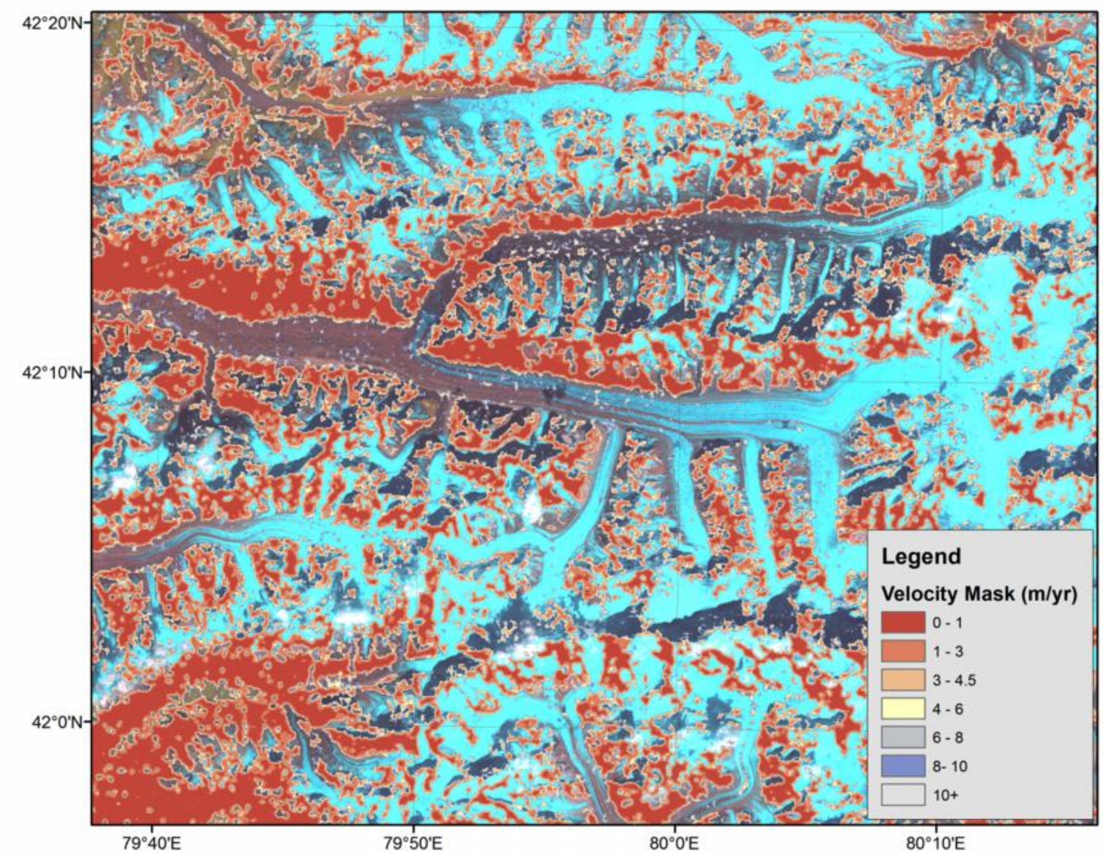

Figure 1 – Velocity map (m/yr) over Landsat 8 B7/5/3, Sept 25, 2013.Showing increased removal of glacier pixels as the velocity threshold is increased.

As can be seen in Figure 1 of this reply, the majority of stable ground is moving below 1 m/yr. However, the amount of glacier pixels which fall within the range 1-4.5 (the general cutoff speed proposed in our algorithm), is minimal, and is generally confined to river sands which only shift slightly over the velocity measurement time period. Removing pixels above this threshold begins to remove a higher percentage of glacier pixels as compared to river sand pixels. It is important to note, however, that minimum velocity thresholds are set on a scene-by-scene basis, and 4.5 m/yr is a general rule of thumb but not always the exact classification cutoff used in the algorithm.

Make sure that all smaller debris-free glaciers are also included in the final results. Put more emphasis on the upper glacier boundary so that the outlines are here also consistent and reliable.

We have broadened our spectral range in our initial classification (described in Section 3.3), which has improved our mapping of small glaciers, as well as the upper boundaries of larger glaciers. This can be seen in Figures 2, 13, and 14.

Debris-covered Inylchek glaciers are special (and maybe therefore also difficult to map) in regard of glacier flow, the periodically draining Merzkacher Lake, calving and surging activities which may need special consideration. Show that your algorithm works also for other debris-covered glaciers in Tien Shan, e.g. the Koxkar and Tomur Glacier in the region south or Tomur/Pobeda/Jengish Chokosu Peak or the glaciers of Kokshal Too.

We have changed our selection of images to show not only the Inylchek glacier (Figures 2, 7), but also the Tomur glacier (Figures 8, 13). We note that the algorithm performs well over all of the large debris-covered glaciers that we examined in the study area.

Best Regards,

Taylor Smith, Bodo Bookhagen, and Forest Cannon

Reply to Reviewer Comments – Smith et al. – Glacier Mapping in the Tien Shan

We would like to thank both reviewers and our editor for their detailed and comprehensive comments, which are very helpful in reframing and reorganizing our manuscript. In this response we will reply first to several general comments which were raised by both reviewers, and then respond to the more specific questions raised by one or more reviewers. We have not included responses to purely formatting or word choice suggestions, but we have taken these edits into account with this revision.

(1) Focus of the Manuscript

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. – F Paul

The second remark is about the focuses of this manuscript. From my understanding, I think this manuscript mainly want to describe a new method to delineate debris covered glacier. But it also focuses on the glacier change, and even on the atmospheric setting. From my view, too many focuses sometimes equals to no focus. It is better if the authors can give up other focuses, return to the main one, and make everything is clear about it, includes give more details. – W Guo

Both reviewers have suggested that the manuscript tries to analyze too many factors, and in trying this it loses focus. In light of this, and several comments asking for clarification and expanded treatment of the algorithm steps, we have dropped the regional climate analysis and instead focus wholly on the algorithm and a more thorough analysis of each step and its benefits and shortcomings in the classification process.

(2) Further Explanation of the Algorithm's Utility

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… 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. There could be reasons for all this but they need to be explained.

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)? - F Paul

Mapping both clean glacial ice and debris-covered glacier areas is a complicated problem. These two glacier types, though contiguous, have completely different spectral signatures, and generally occur in distinct topographic settings. This is not a problem we have completely solved with this algorithm, as there are simply too many variables to map glaciers with complete accuracy. However, our algorithm does improve on traditional digitization methods in two main ways: classification speed and temporal resolution. Using our algorithm, we can delineate glaciers within any given Landsat 5/7/8 scene in a matter of minutes. While these outlines should not be considered perfect, they certainly outperform a band ratio based approach, which is often used as a first pass glacier calculation. Further discussion of the steps of the algorithm is included in the Methods section, with explanatory images included as Figures 2-8 and 13.

Given this speed of classification, it is not unreasonable to calculate glacier outlines on a wide range of Landsat scenes, which can be used to generate a longitudinal dataset. We envision this algorithm not as a final perfect glacial classification step, but rather as an improvement over traditional spectral ratios and as a tool to help densify glacier outline databases in time. We include a preliminary discussion of this in the Discussion section, with an explanatory image in Figure 14.

(3) Changes to the Algorithm in Response to Reviews

3.1. The clean ice area will be significantly underestimated by these methods due to following reasons: A. The unique band ratio threshold (2.0) during the delineation of clean ice glacier area via band ratio method needs to be argued. It cannot always fulfill the demands of precise clean ice delineation from any Landsat image. From my experience, it is too larger in many cases, especially from images acquired during ablation season.

B. The additional 250 of TM1 DN value is too higher because the clean ice area of glacier below equilibrium line during ablation season is fairly dark due to the exists of melted water. This will eliminate many ablated clean ice area. … L11: From my experience, the optimal band ratio threshold to differentiate glacier and non-glacier is not unique, and many glacier surfaces has TM3/TM5 ratio of less than 1.5. Can you make certain that the value of ≥2 is the best to delineate clean ice glacier? Will it underestimate the area of clean ice? – W Guo

In regards to these comments, we have updated our classification algorithm. We have changed our classification threshold for TM band 1 to 60. Using this method, we have reduced the number of small and isolated glaciers which were previously missed by the algorithm. An enhanced discussion of the algorithm steps is found in the Methods section, with an updated Spectral classification image in Figure 2, and a comparison of this spectral classification with the final algorithm outlines in Figure 8. The updated scripts are available via Github, at https://github.com/ttsmith89/GlacierExtraction

C. Actually glacier tends to be survived hiding behinds the hill shadow in North Hemisphere. Therefore the exclusion of area in hill shadow is also problematic, and will exclude many glacier areas, which can be clearly seen from Figure 3 and 4. … L15-19: If you simply exclude all areas in cast shadow, how to appropriately detect the area of glaciers hided by shadow (improper exclusion can be seen from Figure 3 and 4)? In the other hands, is the 90m SRTM DEM has the sufficient resolution to accurately simulate the topography by this method? So I think there are enough reasons to doubt about the precision of the final clean ice glacier outline.

This was a conscious design decision in our algorithm, as we cannot with any accuracy identify glacier areas which are under shadow. We chose to remove shadowed areas which would often be misclassified instead of leaving them in. In essence, we chose to classify the minimum area which could be well-defined by our algorithm instead of overclassifying extraneous glacial areas. There is an option to remove this thresholding step in the updated scripts, if that is the preferred method in a specific use case.

3.2. Even among the new methods to delineate debris covered glacier area, there are also many uncertainties introduced.

A. The topographical filtering step seems too subjective to exclude debris covered area with relatively flat topography. The coarse resolution SRTM also cannot support for this even if the glacier surfaces in this region are fairly rugged… L22: The glacier surface topography has very larger variances for different glaciers and under changing every day, and nobody can make sure their surface is flatter or rougher that other surface. As the authors have stated, the glacier surface only "tends to be rougher than the surrounding areas", but the reality is this assumption is not valid for all glacier surfaces. Therefore, I think this first step to delineate debris covered glacier is ambiguous, and may lead to some exclusion of real debris covered glacier area in certain cases.

Yes, this is something we noticed as well in our classification. We use a very broad filter when looking at roughness so as not to remove glacial areas. We find that this filter mainly removes wide, flat areas, which are far removed from glaciated areas.

B. Although the idea is very innovative, the glacier velocity filtering step will also bring very larger uncertainties, because the unknown accuracy of the glacier velocity extracted. – W Guo

P5436
L16: It seems that basically only velocity is new here, please illustrate how it is working!

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. – F Paul

Again, we use a wide range of glacier velocities, and aim to primarily remove 'stable ground'. These are areas which have not changed significantly over the timespan of our velocity measurement, such as flat plains, hillslopes which do not have glacier coverage, and river beds. River beds in particular are difficult to differentiate from debris-covered glacier tongues, as they share many topographic and spectral characteristics. A map of removed 'stable' areas is found in Figure 4 of the updated manuscript. An example velocity surface can also be seen in this reply, as Figure 1. The data sources table has been updated to indicate which images were used for velocity calculations.

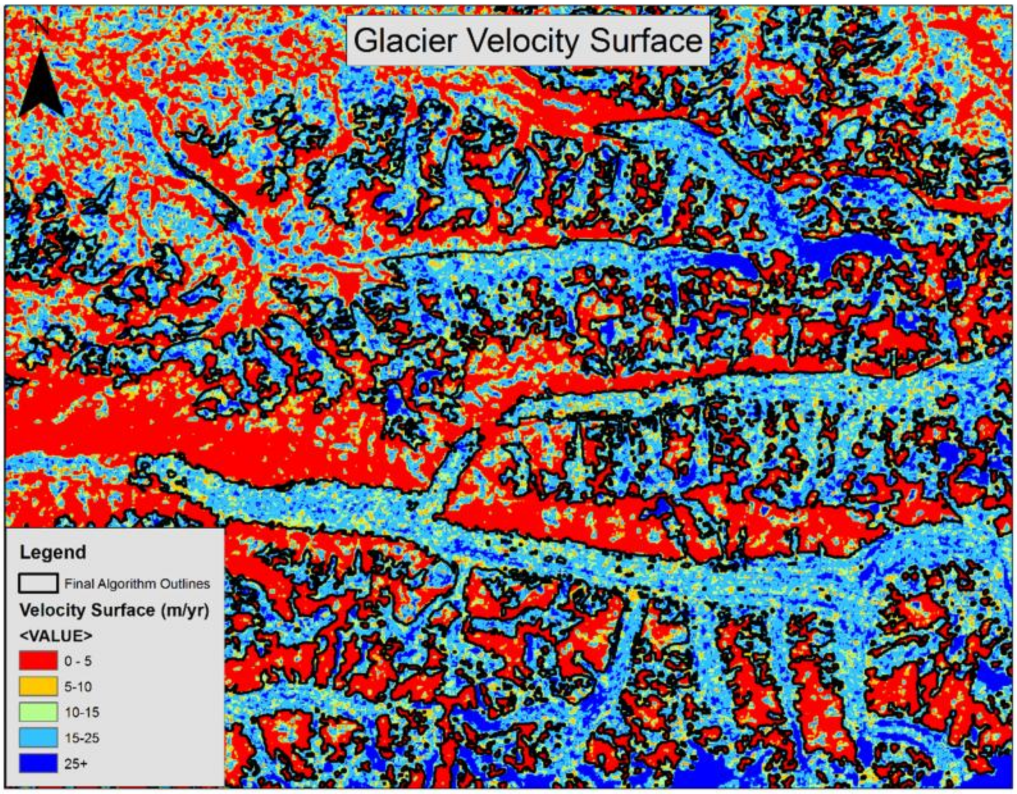

Figure 1- Sample Velocity Map, with classification results overlaid in Black. Averaged velocities over a ten-year timespan, presented in m/yr. Image Pair: LE71470312002230SGS00, LC81470312013268LGN00

Although thresholds are variable for each scene, based on the average speed of glaciers in any given scene, areas moving slower than 4.5m/yr are generally considered 'stable' for this classification and masked out.

C. The spatial filtering step uses the HydroSHEDS 15s river network to remove "potential debris areas". However, from my check, the river network was so sparsely distributed, and the author didn't give the weight of two different distances, I suspect that it will also eliminate many real debris covered glacier area.

Figure 5 in the updated manuscript shows an example of a binary mask which is used for the distance weighting technique. The 'seed points' in this case are our spectrally defined glaciers (black), as well as rivers (blue), which we see go up to, and sometimes well into, the spectrally defined glacier areas. By removing areas far removed from clean ice and these river valleys, we can effectively remove certain non-glacier areas, for example non-glacierized hillslopes.

D. In the final mask step, from my experience, even a single 3×3 median filtering will bring large changes on the shape of final glacier outline. However, according to the author's description, totally 4 kinds and 6 steps of filter, include 3×3 median filter (applied twice), 5×5 median filter, image opening (twice in total), *and image bridging, were applied. After so many filter steps, I wonder whether the resulted glacier outline can represent the true situation of glaciers on the Landsat image. – W Guo*

We have tried to balance accuracy in glacier outlines with accuracy in glacier internal integrity. As can be seen in Figure 7 of the updated manuscript, before any filtering steps are done, many glaciers contain small 'holes', which limit our ability to accurately assess their area and other statistics. To resolve this, we apply filters which remove many of the smaller holes, while attempting to limit changes to glacier outlines. In our updated algorithm, we have implemented fewer filtering steps in an effort to better balance outline and internal consistency in our glacier outlines. These updates are described in the Methods section.

(4) Removal of Extraneous Glacier Discussion

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). – F Paul

We have decided to remove these sections to provide more focus on the algorithm methodology, and will return later to these individual glacier studies in a future publication.

(5) More Conservative Error Discussion

The most important problem of this manuscript is the lack of a convincible referential dataset to evaluate the accuracy of the results from all these methods it illustrated. The authors use manual digitized outlines for ~750 glaciers as the control datasets. But as the authors stated in the paper, the results of manual work has very larger uncertainties, especially from coarser resolution like Landsat images and for the debris covered glacier area. – W Guo

L18-23: The manually created glacier outlines must have very limited accuracy, especially for the debris covered glacier area, and should not be regarded as control datasets to validate the results of the methods used in the study. As has been suggested by Paul et al (2013), the higher resolution images in Google Map can provide excellent referential source. So I suggest the authors choose this better way to validate the results of this study. – W Guo

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. – F Paul

We recognize that manually generated glacier outlines are inherently error prone, and are not a perfect control dataset to use. We attempted to verify these outlines in both ArcGIS against Landsat images, as well as in Google Earth, against higher resolution imagery, to create the highest possibly accuracy. For multiple time steps as we have analyzed, there simply does not exist a better set of glacier outlines. We have removed much of the discussion around comparison statistics between our control dataset and the algorithm outlines and have instead focused on overlaying the different sets of outlines for visual comparison. This can be seen best in Figure 13.

Specific Commentary – W Guo

P5435
L20: 15-30m only corresponds to half to one pixel of most bands of ETM+. If this is the real meaning, it should be clarified by state the panchromatic band of ETM+.
This has been clarified on Line 15.

P5441
L6: The residual error of Landsat images compare to their master image in this step should better be given in proper form, as it is important to determine the accuracy of glacier velocity and glacier change in the following contents.

We refer here to georeferencing, to ensure that our outlines are consistent in space. An updated discussion of this rectification is included in Lines 106-112. This step simply attempts to remove these small sources of error which will then be propagated through the rest of the image processing. It should also be noted that many Landsat 8 images were not subjected to this process, as they are generally very well georeferenced already.

L25: From my experience, the clean ice of glacier always has similar NDWI value with water bodies, how the authors solve this contradiction should be clarified, along with the method to differentiate the frozen lake with glacier surface, because many Landsat images used in the study were acquired during winter (can be read in Table 1).

We have removed the NDWI thresholding step, to bias our algorithm towards overclassification instead of removing glacier pixels.

P5542
L12: I noticed that acquisition date of Landsat 8 OLI images used in the study are fairly different, and the glacier velocity is known has seasonal differences. So how the measured glacier surface velocities were standardized into m yr-1 should be clarified here. Furthermore, as Kääb (2002) didn't gave the accuracy of the measured velocity, and the original thesis that describe the CIAS software are not in English, the authors should better give some illustration on the accuracy achieved by CIAS, because of the importance of glacier velocity in determining debris covered glacier area.

Yes, there will certainly be seasonal differences in glacial velocity. However, there are limited options for clean Landsat scenes with which to create a velocity surface for each Path/Row combination. We chose instead to standardize each velocity surface by the number of years – including fractional years – between the master and slave images used for the velocity calculations. As these acquisitions were several years apart, we were not assessing seasonal velocities, but instead we developed a multi-year average velocity surface. This velocity surface does not perfectly conform to glacier areas, and different parts of each glacier are moving at different speeds. However, we were able to successfully differentiate between 'stable' areas and 'moving' areas, which is an important metric for classifying debris tongues.

In this manuscript we have used the CIAS software, as we did not find significant improvement with other correlation methods, such as those used in COSI-Corr, because of our conservative velocity thresholding. A more detailed comparison of image matching techniques can be found here:

Heid, Torborg & Kääb, Andreas (2012). Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery. Remote Sensing of Environment. ISSN 0034-4257. 118, s 339- 355 . doi: 10.1016/j.rse.2011.11.024

Section 3.3.2: This section is not also well organized, and the section header is totally not coincident with its contents. The first paragraph can be called as real 'topographic filtering'. However, the second and third paragraphs are talking about the glacier surface velocity rather than glacier topography, and should better to be called as 'kinetic filtering' or 'dynamic filtering'; the fourth paragraph is again talking about 'frequential filtering' rather than topography. Besides this, since the experiments on frequential filtering were not successful, I suggest to put it into Discussion section?

We have changed the organization of this section based on this suggestion, and moved the discussion of frequential correlation methods to the Discussion section.

L22: The glacier surface topography has very larger variances for different glaciers and under changing every day, and nobody can make sure their surface is flatter or rougher that other surface. As the authors have stated, the glacier surface only 'tends to be rougher than the surrounding areas', but the reality is that assumption is not valid for all glacier surfaces. Therefore I think this first step to delineate debris covered glacier is ambiguous, and may lead to some exclusion of real debris covered glacier area in certain cases.

While it is true that the standard deviation filter does remove some glacier areas from the 'potential debris areas', it removes far more flat terrain that is not associated with glaciers. We use a very inclusive threshold in this step to leave as much debris tongue area intact as possible. We have tested removing the standard deviation filter, and find generally poorer glacier outlines.

P5543
L26: "Two additional topographic indices …": This paragraph is actually talking about some experiments in frequential domain rather than topography, so "frequential indices" is more appropriate here. Furthermore, as mentioned above, because the experiments illustrated in this and the following paragraphs were not successful, it is suggested to put these two paragraphs into Discussion section.

We have moved this section to the Discussion section of the manuscript. Although we did not use these experimental analyses in our algorithm, we felt it was important to include what did not work as well as what did, in the case that future studies wished to explore similar avenues of research.

P5444

L27: I checked the HydroSHEDS 15s of Asia, and found that this data is very coarse in depict river networks, and can only illustrate larger trunk rivers… So I suspect that if using such river networks to determine the rationality of debris covered glacier area, it will eliminate many debris covered areas of small glaciers in this step, especially those cirque glaciers or small valley glaciers. Even by using 200m buffer, this situation cannot be well overcome. If it is solved by add manual seeds in second step, the manual workloads should be very heavy. – W Guo

P5444

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

Figure 5 in the updated manuscript shows that using clean glacier ice for distance weighting seed points is often inadequate, especially in the case of very long debris tongues. We supplement our spectral outlines with the HydroSHEDS river network so that we do not remove debris zones which flow far down valleys from the clean glacier ice.

P5445

L24-27: As the authors have stated, the determination of proper thresholds of all parameters must be difficult to accurately delineate glaciers because so many methods have been introduced, and most parameters must be site dependent. So I believe that the portability or the universality of these methods must be very limited.

We have found that the algorithm classifies glaciers with high accuracy across the eight Landsat Path/Row combinations investigated. We use a constant set of thresholds for the majority of our filters, with Path/Row specific thresholds for velocity and Elevation. These thresholds are quickly determined by inspection of the topography and velocity surface of any Path/Row combination. We have also simplified the algorithm in this update, so that less input datasets are necessary for glacier classification. An overlay with both the control dataset and the RGI v4.0 can be found in Figure 13.

P5447

Section Results: In my opinion, the results section actually gives no results. Meanwhile, the bulk elevation distribution of different results (Figure 5, 6, and 7) can not give a good description of the accuracy achieved, as well as the situation of Figure 8 and 9. Many actual differences will be covered by this way. In this sense, the area differences for different glacier area classes between the three datasets can provide useful information on their comparison, especially between the results of authors' methods and the manually digitized dataset.

We have decided to save the discussion of regional climate for a forthcoming publication, and instead focus here on our algorithm and the results.

We maintain that the bulk area distribution plots provide useful data on the topographic distribution of glaciers in the study region. As can be seen in Figure 10 of the manuscript, there are certainly a range of misclassified areas. However, these offsets overlap at low elevations, indicating generally good treatment of debris-tongues, and skew towards overclassification at high elevations, indicating that we do not miss many glacier areas and instead provide a 'maximum glacier area'. This is not to say that the glacier extents we present are perfectly constrained *spatially* – there are certainly areas which are misclassified. However, when assessing the elevation, aspect, slope, or curvature distribution of glaciers across a given area, our statistics remain robust. It is for such applications that we propose to use the raw data from our algorithm. For more detailed glacier-by-glacier analysis, manual corrections of glacier outlines should be applied, just as manual corrections are necessary when using a spectral ratio classification of glaciers.

Specific Commentary – F Paul

P5434

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?

Yes, this language was not specific enough. In this, we referred to the general trends of increasing temperature in the region, which have been noted by several authors (e.g., Lioubimtseva and Henebry, 2009; Sorg et al., 2012). Those glaciers at higher elevations have experienced less of this regional warming, as temperatures have not risen as quickly at high elevations. However, this analysis has been cut from the new manuscript.

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?

We refer here to changes in water resource availability. While it is true that in many catchments runoff is increasing despite shrinking glaciers, this trend will not continue forever. As there remains a definite lack of high temporal and spatial resolution analysis of glaciers in the Tien Shan, we posit that increased understanding of how glaciers are responding to climate change in the region will help inform improved water management in the region. However, we have removed this section along with our climate analysis, which will be presented in a future manuscript.

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).

We propose to densify the decadal repeat times of current glacier inventories with our algorithm, which can provide a first pass at higher temporal resolution glacier datasets.

P5436
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.

While this is true of Landsat 8 images, Landsat 5 and 7 images are not as well orthorectified and georeferenced. We use this step to make certain that our base datasets, the Landsat bands, are well rectified in space.

L25: What are 'trends in glacial character'? Please be precise. And what is a 'link' when R2=0.02?

We refer here to our observed relationship between longitude and glacier area change rate. This portion of the manuscript has been cut to focus on a more thorough analysis of the glacier mapping algorithm. These results will be published in a forthcoming manuscript.

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).

All scenes with seasonal snow were manually inspected after processing, and those glaciers which were strongly impacted by snow cover were not included in our statistical analysis. The 10% cloud cover statistic refers to the whole scene, as clouds which do not overlap with glacier areas still impact the efficacy of the classification algorithm. We propose that atmospheric moisture content of a scene changes classification sensitivity. So, even if glaciers are not covered by clouds, the additional atmospheric moisture in the area changes their spectral signature.

We have noted that this is less of an issue with the Landsat 8 satellite, potentially due to the higher spectral sensitivity of the sensors. This discussion has been updated on Lines 367-380.

P5441
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.

The lake delineation section has been updated with more specific language on Lines 123-135.

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?

In our selection of 'index lakes', we ensure that at least three lakes are chosen in each scene, with differing topographic settings (i.e., far removed from a glacier, on top of a glacier, and at the base of a glacier). In addition to this, we add a buffer to the average spectral signature of the lakes, so that we better capture the lakes in any given scene. This explanation has been improved on Lines 123-135.

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.

We have adjusted our glacier mapping algorithm with this in mind.

P5443
L3: Please describe how this 'additional thresholding' works.

The additional thresholding mentioned here is described in the subsequent paragraphs of the manuscript. In essence, we first identify clean ice areas using a spectral ratio, and put these areas aside. Then, we consider the rest of the image, looking at slope, velocity, distance weighting, and statistical filtering. These steps act only on the debris areas, to identify the debris tongues of glaciers. After the thresholding, we add the spectrally-derived areas back to the image to derive our final glacier outlines which have both clean and debris-covered glacier areas. The Methods section of the manuscript has been updated with new text and figures to further explain this.

L8: Please describe 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.

The image pairs have been added to Table 1, and Figure 4 now shows a binary velocity mask. A velocity field image is included with this reply as Figure 1.

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

We have moved this to the Discussion section, as we maintain that steps that do not work should still be included as guides for future researchers.

P5444
L20: What is meant with 'velocity profiles', velocities? What is 'very different' (5, 50, 250 m/yr)?

Velocity profiles here refers to average velocity of this type of pixel, whether it is stable ground, river sand, or dead ice attached to a glacier. These pixels tend to be moving less than 4.5 m/yr, which is used as our threshold metric for velocity on most scenes. This is also often the minimum detection threshold for many velocity measurements. This has been updated on Lines 183-186.

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

Areas masked out by the velocity step are included in Figure 4 of the updated manuscript.

P5445
L6: Please show that it is effective.

Figure 6 in the updated manuscript shows the glacier outlines after the velocity filtering overlaid with the outlines just after the next step: distance weighting. As can be seen in Figure 6, very little area is removed from the main glacierized areas, but instead areas far removed from river valleys and clean glacier ice are discarded (areas seen in black in Figure 6). This is easiest to see in the top left corner of the image.

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).

We have found that only those glaciers which have particularly long debris tongues, such as the Inylchek glacier, really benefit from seed points. These seed points can be clicked before testing the algorithm based solely on the length and thickness of a glacier's debris tongue. An example of which seed points were clicked can be seen in Figure 2 of this reply, alongside the river network and spectral outline seed points.

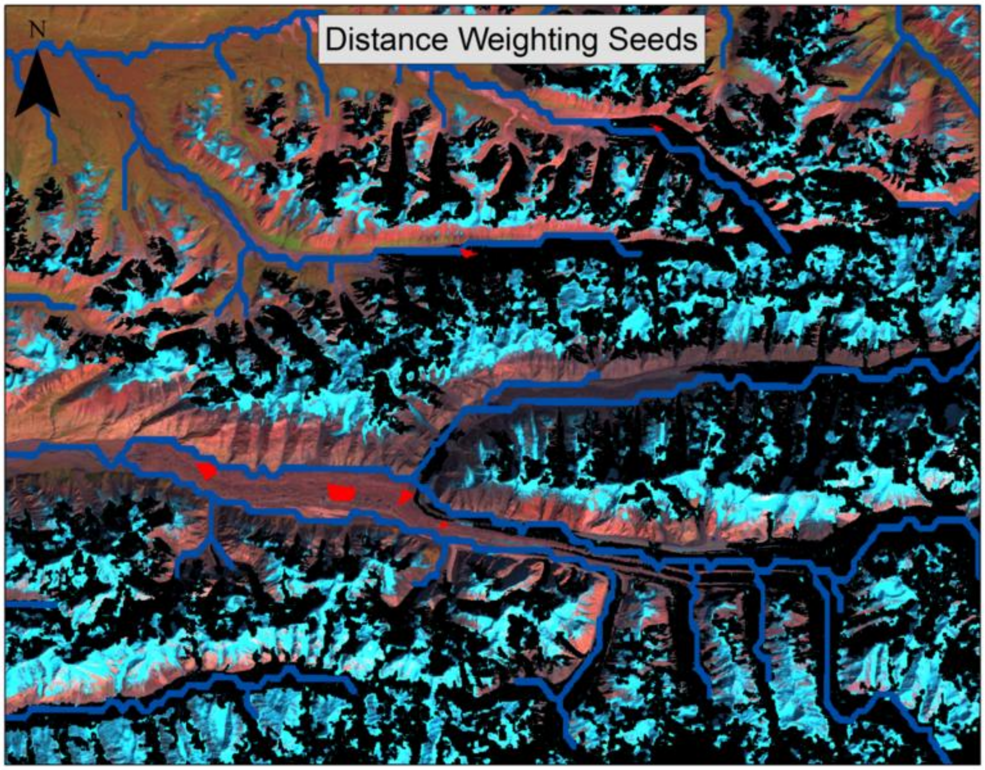

Figure 2 – Distance Weighting Seeds including spectral outlines (Black), HydroSHEDS rivers (Blue), and manual seed points (Red).

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

The updated algorithm has removed these steps, as they unnecessarily complicate the classification process without adding significant classification improvement.

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

Figure 7 in the new manuscript shows the unfiltered algorithm outlines as compared to the filtered outlines. The outlines derived by this point are generally coherent, although they often contain holes.

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.

Yes, this is true. We have considered our filtering necessary to remove holes in the glacier surfaces due to errors in the velocity map, errors in the SRTM, and other inconsistencies. When using a filter such as this there is always a tradeoff where you lose some 'good' data to correct 'bad' data. In light of the comments provided by both reviewers we have minimized the statistical filtering used in the algorithm, the effects of which can be seen in the updated manuscript in Figure 7. An updated description of the statistical filtering can be found in Section 3.4.4.

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

This has been removed from the algorithm, as well as the manuscript. We had originally included metadata such as Area and Perimeter which could then be used to filter out very small (~5 pixels or less) areas to 'clean up' the glacier outlines created. However, we have found that this step is far too processor-intensive for the minimal benefit it provides.

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.

The two datasets cover the same spatial extent (all Path/Row combinations), but cover different times. One was created for ~2000, another for ~2011, so that there would be two temporally independent control datasets. These encompass ~3000 sq km of glacier area, and include numerous debris-covered glaciers. The outlines are shown alongside the RGI and final classification results in Figure 13 of the updated manuscript. An updated discussion of the control datasets can be found in Section 3.5.

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?

Ice caps was a poor word choice, as you are correct that there are not ice caps in the study region. We refer here to connected glacier areas around the peaks of mountains, from which glaciers flow into separate watersheds and river valleys. The wording has been updated on Lines 244-247.

Although we have chosen a series of mostly snow-free images, it is very hard to find an absolutely 'clean' image in the study area without snow, and thus there are some glacier areas which are connected by seasonal snow cover. This snow cover is very hard to remove with our algorithm, and is one of the first things we look for when performing manual corrections.

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?

Watershed boundaries were only derived for a subset of the glacierized areas due to processing constraints. Deriving statistics on such a large number of distinct polygons, against a wide range of background datasets (elevation, slope, aspect), is computationally expensive. This analysis could of course be expanded to use every watershed in the region, but we felt it was more efficient to use a representative sample of watersheds as opposed to using every one.

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).

We have updated our statistical analysis and reorganized the results section to conform to the new focus of the manuscript on the algorithm and not on the climate setting.

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.

When we compare our outlines with the RGI (Figure 13 in the new manuscript), there are clear differences, with our algorithm overclassifying some areas and underclassifying others. However, the time it takes to generate our glacier outlines is quite small compared to the work it takes to create a dataset such as the RGI. We have removed the discussion of the RGI from the bulk-area statistics, and instead only compare it with our results visually.

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?

Our updated manuscript and algorithm includes figures showing the qualitative differences between the RGI and our algorithm results (Figure 13).

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.

We have decided to remove this section to better focus on an analysis of the algorithm as opposed to its utility in assessing individual glaciers. A limited discussion of potential future uses with glacier change detection can be seen in Section 5.2 and Figure 14.

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.

We have removed this section of the manuscript, and will explain these results in a future manuscript.

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.

Statistics are generated from consistent glacier outlines, which are clipped by the watershed boundaries mentioned above. In this way, individual glaciers, or comparable parts of the same glaciers, are analyzed. The glaciers chosen are diverse in size, debris cover, elevation, and topographic setting. We attempted to choose a representative sample of glaciers across the study region.

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?

The increased sensor range onboard Landsat 8 reduces saturation in the spectral bands used for glacier ice detection. This is due to the difference in the sensors (8-bit on Landsat 5/7 and 12-bit on Landsat 8), which allows for better radiometric quantization and a higher signal to noise ratio. The change of Landsat 8 to a pushbroom sensor also results in higher radiometric sensitivity.

Reduced saturation in the spectral domain reduces the impact of atmospheric moisture which is present when there are clouds within the scene. As the glaciers are classified by ratios of spectral bands, having a wider range of possible TM 1, 3, and 5 values reduces errors due to band saturation. This is what we are referring to when we say that OLI performs better on images where snow and clouds are present.

P5452
L2: 'best suited images': Sure? The example in Fig. 3 shows way too much seasonal snow for deriving accurate glacier outlines.

We were limited in our choices of Landsat scenes by the number of cloud covered and snow covered scenes. We attempted to choose those which were best suited to our analysis, but yes, some of them do have seasonal snow cover. We have removed this section in any case, and will return to this analysis in a future manuscript.

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.

We have updated our analysis and included quantitative assessments of glacier outline accuracy in Section 5.2, and Figures 13 and 14.

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.

This is a very good point. We will return to this question in a future manuscript, and here have instead focused on the algorithm process.

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.*

Several authors have noted regional temperature increases over the past few decades (e.g., Lioubimtseva and Henebry, 2009; Sorg et al., 2012). However, we will return to this section in a future manuscript, as we have removed the climate section from this publication.

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.

Debris cover has been mapped by the algorithm, as can be seen from several images presented above. However, we will include this analysis in a future publication, as we focus here on the algorithm.

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).

Examples of debris tongues being well mapped by the algorithm can be seen in Figures 7, 8, 13, and 14 in the manuscript. While we agree that the outlines are not *exactly* accurate, we maintain that they outperform spectral-only delineations by a wide margin.

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

Yes, this is true. This sentence has been removed. We were referring here to the lack of the SWIR band when we mentioned 'lower spectral resolution', but this was not clear enough.

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?

Yes, this is true. It was outside the scope of this study to generate a 'maximum glacier area' polygon for each glacier in the study which could be used as a subsequent clipping mask. This would require a set of high-quality watershed boundaries which would then be modified to reflect breaks between 'main trunk' and 'side valley' glaciers. This is certainly possible, but again, only a limited version of this technique was used in this study due to the time-intensity of generating all of these polygons.

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.

These changes have been included in the updated manuscript.

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.

This table has been removed, as we have focused entirely on the algorithm results.

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

We have updated this figure as Figure 1 in the new manuscript.

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.

We have removed this figure and instead include a list of algorithm steps. Further illustrative images are included as Figures 2-8.

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.

Examples of such outlines have been included as Figures 7,8,13, and 14. We maintain that the bulk hypsometry of glaciers in a given region is a useful statistic, but it has been de-emphasized in the revised manuscript.

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.

This chart has been cut.

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 km2. But apart from this I think the plot does not carry any useful information and can be removed.

This chart has been cut to focus on the algorithm.

Specific Commentary – T. Bolch

Focus on the glacier mapping algorithm only

The sections on the application of the algorithm for change detection, as well as our interpretation of possible links to changing regional climate, have been cut. These sections have been replaced by enhanced treatment of the glacier mapping algorithm, with more illustrative images (Figs 2-8, 13).

Clearly show and discuss the importance of each different processing step for improving the results

We have expanded our in-text explanation of each step of the algorithm in Sections 3.1-3.4, as well as added several new images (Figs 2-8) to illustrate what the delineated areas look like after each step of the algorithm. The final result is compared to the initial spectral classification in Figure 8, and compared with other datasets (manual and RGI) in Figure 13.

Show and discuss the influence of different thresholds (e.g. band ratio, TM 1 ratio, velocity) and clearly justify your chosen ones

We have updated our threshold discussion in Sections 3.3, 3.4.1, 3.4.2, and 3.4.4, and have included images that show the extent of classified and unclassified areas after each step (Figs 2-8). Included here as Figure 3 of this reply is an example velocity map, classified to show values between 0-10 m/yr.

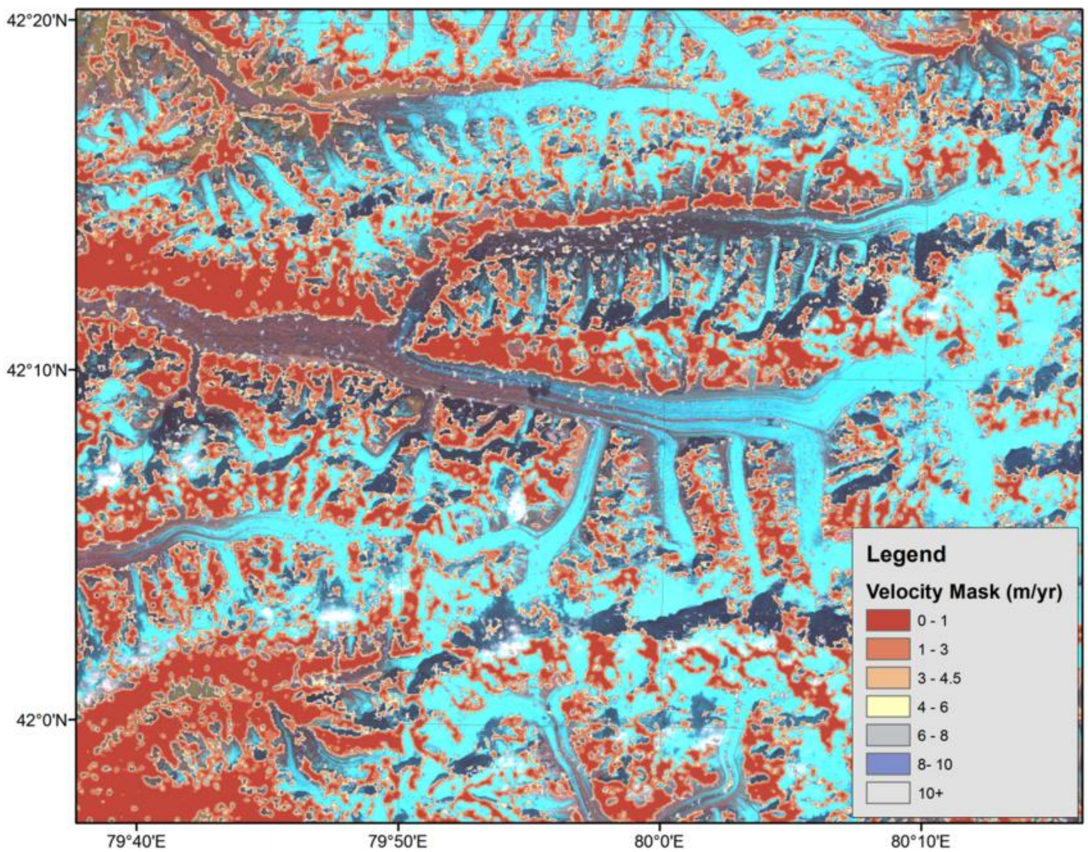

Figure 3 – Velocity map (m/yr) over Landsat 8 B7/5/3, Sept 25, 2013.Showing increased removal of glacier pixels as the velocity threshold is increased.

As can be seen in Figure 3 of this reply, the majority of stable ground is moving below 1 m/yr. However, the amount of glacier pixels which fall within the range 1-4.5 (the general cutoff speed proposed in our algorithm), is minimal, and is generally confined to river sands which only shift slightly over the velocity measurement time period. Removing pixels above this threshold begins to remove a higher percentage of glacier pixels as compared to river sand pixels. It is important to note, however, that minimum velocity thresholds are set on a scene-by-scene basis, and 4.5 m/yr is a general rule of thumb but not always the exact classification cutoff used in the algorithm.

Make sure that all smaller debris-free glaciers are also included in the final results. Put more emphasis on the upper glacier boundary so that the outlines are here also consistent and reliable.

We have broadened our spectral range in our initial classification (described in Section 3.3), which has improved our mapping of small glaciers, as well as the upper boundaries of larger glaciers. This can be seen in Figures 2, 13, and 14.

Debris-covered Inylchek glaciers are special (and maybe therefore also difficult to map) in regard of glacier flow, the periodically draining Merzkacher Lake, calving and surging activities which may need special consideration. Show that your algorithm works also for other debris-covered glaciers in Tien Shan, e.g. the Koxkar and Tomur Glacier in the region south or Tomur/Pobeda/Jengish Chokosu Peak or the glaciers of Kokshal Too.

We have changed our selection of images to show not only the Inylchek glacier (Figures 2, 7), but also the Tomur glacier (Figures 8, 13). We note that the algorithm performs well over all of the large debris-covered glaciers that we examined in the study area.

Best Regards,

Taylor Smith, Bodo Bookhagen, and Forest Cannon