Reply to Reviewer Comments – T Smith et al.

Wanqin Guo – General Comments

1. I still argue that the lack of meaningful control dataset is the critical shortcoming of this study, which significantly lowers down the scientific quality of this paper. Although current control dataset has been improved in Google Earth, its validity to act as a control dataset still need to be proved. The comparison between the algorithm outputs and current control dataset (also act as the verification of the algorithm introduced in the paper) is yet inconvincible to me. A control dataset from existing near simultaneous high resolution satellite images, which was mentioned in my previous referee comments, is again recommended here.

We have endeavored to create the highest possible quality control dataset, keeping in mind that our access to simultaneous high-resolution satellite images is limited. We have reviewed and updated our manually digitized outlines. We also include in this revision a comparison between our results and the Chinese Glacier Inventory v2, where there is overlap between our study area and that dataset.

2. In the results, currently it focuses more on statistical aspects of the comparison between the algorithm outputs and manual control dataset, which cannot clearly illustrate the performance of the algorithm. I recommend the author to provide more glaciologically meaningful comparisons besides current pure geometrical comparisons, such as the differences between glacier with different size (e.g., in different area ranks) and different type (e.g., hanging glacier, cirque glacier, valley glacier, etc.), as well as between clean-ice and debris-covered glaciers. They will provide more detailed information on the suitable application case of this algorithm. Furthermore, because the algorithm presented in this paper is mostly focusing on debris-covered glacier delineation, it is necessary to provide some comparisons with results of previous algorithms, such as Taschner and Ranzi (2002), Paul et al. (2004), Bolch et al. (2007), and Shukla et al. (2010), to better illustration the advancements of this algorithm achieved.

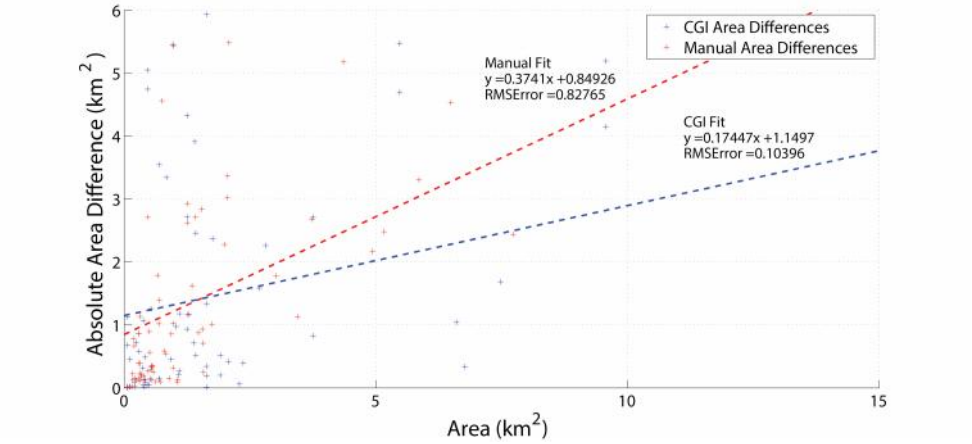

Figure 1: Glacier Area vs Area misclassification, as compared to both a subset of the manual control dataset and the CGI

We have included in this reply a plot illustrating the differences in classification across different size classes, which can be seen in Figure 1. We emphasize, however, that the algorithm was not designed around mapping individual glacier areas, and such a comparison was removed from the original version of the manuscript. As a slight change in which areas are 'connected' by snow, misclassified pixels, or other classification issues can drastically change the reported glacier area, we do not present this data in the updated manuscript. If, for example, a glacier with an area of 10 sq km was connected by a small strip of misclassified area to a glacier of 50 sq km, the reported area would be 60 sq km, which matches poorly if it is compared to either the 10 sq km or 50 sq km glacier area. As this creates a large number of outliers for individual glacier comparisons, we have elected not to present individual-level glacier statistics in the revised manuscript.

Furthermore, in our analysis, we have not split our manual control dataset into debris-covered and non-debris covered sections, but instead have mapped contiguous glacier areas. This matches up with the format of the algorithm and clean-ice outline output, which are not split into individual glacier polygons. Thus, we do not present information on debris cover percentages. We have also added a comparison between our algorithm and several previous algorithms in Section 5.1, and note that despite errors at an individual glacier level, we see strong agreement between the aggregated area of the control datasets and the algorithm datasets.

3. I agree with the author's opinion (Line 343-345) on the limitations of this algorithm that it can only be used at the scale of watersheds, satellite image footprints, or mountain ranges, and to provide "baseline set of glacier outlines which can be corrected manually" (Line 356-357), rather than for compilation of regional glacier inventory. Besides, from my view, the complexity of this algorithm and the involvement of many manual work during its performance (at least seven steps may need human interventions, includes rectification of Landsat images, extraction of velocity fields, lake detections, lowest elevation definitions, velocity and distance threshold determinations, seed lake points definitions, etc.) may largely limit its wide application even within such large scale studies. Many improvements are needed to further promote its automating ability, includes the automatic optimization of the involved thresholds, and selection of the glacier delineation and post-processing methods. These aspects should also be considered by the author. Currently it is hard to say that how much improvements this algorithm can achieve on the accuracy and efficiency of glacier delineation than the widely used glacier delineation methods, i.e., automatic delineation of clean-ice area with manual improvements, and manual digitization of debris-covered area.

We have updated the algorithm once again with this review, with the goal of limiting manual input and simplifying the algorithm as much as possible. An updated workflow has been included in the Methods section, and updated scripts have been provided. We have also added

the option of using NDSI directly in the glacier mapping script, for those cases where this is the more desirable clean-ice delineation method.

4. The text in the last part of Discussion (Line 360-370) truly illustrated the other real limitations of the algorithm introduced in the paper, which put more limitations on its applicability. To make the scientific contribution of this paper clear, I suggest the author to rewrite the ***Abstract*** *and* ***Conclusion*** *section, to better illustrate the advancements and limitations of the algorithm presented in the paper.*

We have re-written those sections to be clearer.

Wanqin Guo – Specific Comments

Abstract: See 4 in ***General Comments****.*

Updated abstract included.

Line 3-5: I don't see many necessities to mention glacier volume change studies here. Furthermore, several literatures should better not be cited here.

These have been removed.

Line 26: "where the debris is sourced from" is inappropriate. Most glacier surface debris comes from accumulation area and subglacial moraine, rather than the hill slopes around glacier tongue.

This wording has been updated.

Line 44: "… and ideal test area" should be "… an ideal test area"?

This has been corrected.

Line 65-68: Which version of the SRTM data was used should be clarified because large difference exists among different versions.

This has been clarified.

Line 80: should the "and any …" be "and many …"?

No, this is correct. Manual debris points are optional, and the algorithm will work without them. They are helpful for very long debris tongues, but are not necessary for every glacier.

Line 76-98: A carefully designed illustration should be better to present the contents of this section. Current organization somehow conflicts with following sections and is hard to follow. The sequential numbers also conflict with the main sections.

A flow diagram was proposed in the original manuscript, but one reviewer asked for something in the current format. We have elected to modify the current chart, in the interest of saving space and presenting a clear listing of the algorithm steps.

Line 112: "as high-quality georeferencing is already included in the L1T product", is this conclusion comes from your validations? If not, please provide some references to prove it!

We have re-run our original analyses without the use of the AROP, and find that the georeferencing and orthorectification provided by the USGS L1T product is sufficient for the algorithm.

Line 114: What is the meaning of "hydrologically corrected DEM"? Please give some illustration.

We refer here to the CGIAR void-filled DEM. We have updated our language accordingly.

Line 141-142: Again same issue as my early comments (see general comments 3.1A in tcd-8-C2680-2014-supplement.pdf) arise here. Your response in tcd-8-C2882-2015-supplement.pdf mentioned you have changed the threshold of TM3/TM5 to 1.5. Is it typing error? The results shown under the correspondent response in tcd-8-C2882-2015-supplement.pdf give overestimation of clean ice in accumulation area, but it seems coming from the extensive seasonal snow cover in the image you used, rather than smaller TM3/TM5 threshold.

This line was a miscitation of previous literature and has been fixed. We use a threshold of TM1>60, and TM3/5>2. We have found this is the most effective pairing for classifying glaciers in our study region.

Line 167-170: The image correlation methods required that the glacier surface has similar structures or spatial patterns. I noticed from Table 1 that 7/8 of the Landsat scenes use image pairs with time interval longer than 10 years. Glacier surface may change dramatically from year to year, no matter the clean-ice or debris-covered area. Ten years must be too long to perform image correlation. So I suspect about the validities of the extracted velocity field to filter the glacier surface pixels. The ideal time interval between the selected image pair no should longer than several years. I suggest the author to reselect the image pairs with short time interval, e.g., 2-3 years.

The velocity filtering as proposed in this manuscript is not meant to be a complex or perfect quality approach. For this, there are more complex algorithms, such as those discussed here:

Dehecq, Amaury, Noel Gourmelen, and Emmanuel Trouve. "Deriving large-scale glacier velocities from a complete satellite archive: Application to the Pamir–Karakoram–Himalaya." *Remote Sensing of Environment* 162 (2015): 55-66.

For our approach, we focus instead on a simple method that provides reasonable results. As the velocity filter is only meant to remove 'stable' pixels, and not add to the classification of debris or glacier area, an imperfect estimate will still provide a useful filter. Furthermore, as we use only a single velocity measurement for each Path/Row image set, it is most important that very stable terrain is identified over the entire time range, not that we have perfect glacier velocities. Providing high quality glacier velocity measurements for each glacier area is outside of the scope of this study.

Line 176-177: 'cloud-free and snow-free images' should better to be 'cloud- and snow-free images'

This has been updated.

Line 256-263: The manual control dataset seems containing main bias itself, especially in the regions under hill shadow, which occupy very large area and can be identified from Figure 13… The bias of the algorithm results towards high elevation areas thus should party attributes to such misclassifications. Although manual digitization of glacier outlines in accumulation area is more tedious than other region, it is worth to be carefully performed to provide better comparison.

We have endeavored to create the highest quality control dataset possible, keeping in mind questions of both scale and accuracy. After careful re-consideration of our control dataset, we have made some few improvements, especially to accumulation zones. We have also included comparisons with the newly released Chinese Glacier Inventory v2 (Figures 10, 14).

Line 264-277: Most reasons of mismatch between the algorithm and manual control datasets described here are reasonable, but I suppose that if the quality of the control dataset can be further improved, the distribution of over- and under-classified glacier area along elevation should be slightly different with current one.

We provide updated figures here, after carefully re-analyzing our control dataset.

Line 264: Maybe "intrinsic" or "inherent" is better than "persistent" here to better represent the meaning of natural shortcomings of the algorithm.

This has been updated.

Line 279-283: Maybe the distances between the vertices of algorithm results and their nearest vertices or lines in manual control dataset are more meaningful here, because generally the

manual digitized glacier outlines contains far less vertices and fewer shape variations than the automatically delineated results.

Yes, we have considered this. The main problem with this approach is that when the spectral vertices are used, the distribution can quite easily be skewed by one or two small (but complex) polygons which are not contiguous with the main glacier area. As we know that the manual dataset's vertices represent actual (or as close to possible) glacier vertices, we can use them to compare with the algorithm outlines. If we were to do the reverse, we wouldn't be able to be certain that the 'master' vertex was well placed along the edge of a glacier.

Line 283: What's the meaning of "Normalized Distance" in Figure 11? Please give some illustrations around here.

We normalize the distance distribution to scale from 0-1 using the maximum distance.

Line 290-298: This section seems duplicates with 4.1. It should better to be merged into 4.1, or be totally removed.

We have moved this section to section 4.1.

Line 300-318: From my view, this section is unnecessary in current paper, because they only described the failed method experiments that give no improvements on the glacier classification. It should better to be totally removed.

We have opted to keep this section in the manuscript, as we believe that having a record of both what works and what does not is useful for the future exploration of this topic.

Line 325-328: This sentence is hard to follow. What are you actually wanted to say here? To substitute the thresholding method of clean ice? Please rewrite it.

This has been updated.

Line 335-338: From my view, the comparison between the algorithm and manual control datasets and RGI in this region is unnecessary. The lower quality of current RGI in this region is well known, therefore should not to be regarded as a reference.

We provided a visual comparison with the RGI for this region to emphasize the utility of a quick, wide-area, classification scheme. In our updated manuscript, we instead show a comparison with the newly released Chinese Glacier Inventory v2, for visual comparison.

Line 350-352: I didn't think the velocity data extracted from different image pairs can be "static in time", because large differences may exist among different images, although the topographic data can be considered as "static". "same areas are captured …" may mostly due to the "static" topographies.

Here static refers to the fact that we only leverage a single velocity measurement. If the algorithm were to use multiple velocity timesteps, this factor would no longer be 'static' in the classification. However, we have tested stepped velocity measurements and do not find enough improvement in classification results to justify the additional processing time.

Line 357-359: How the "processing time" can be "decreased when a large set of Landsat scenes are considered"? "as generating the input velocity surfaces can take longer than processing glacier outlines from dozens of Landsat scenes"? Is it a typing error?

This is correct. Generating a velocity field for an entire Landsat scene can take a few hours on a standard desktop computer (in processing time, not human interaction). However, once this velocity field has been generated, it can be used in the classification of an arbitrary number of Landsat scenes sharing the same Path/Row combination. Each scene classification takes 10-15 minutes, depending on which output datasets must be polygonised. The actual time to generate a binary glacier/not-glacier .TIF file is generally 3-5 minutes or less on a standard desktop computer. This is what we refer to when we say that processing time is decreased, as the *average* processing time will decrease as the number of images used is increased. We have updated our wording to make this clearer.

Line 370-379: These findings are very meaningful. However, they seem not tightly connected to the contents of this paper. They were suggested to be shorten.

We have shortened this section, as it is mostly theoretical.

Line 381-397: See 4 of ***General Comments***

We have updated our conclusions section.

Table 1: "Bold dates indicate use for Velocity profiles" should better to be "Bold dates indicate images used for Velocity profiles"; see Line 167-170 for further comment.

This has been updated.

Figure 4: The higher velocities in the regions along the river and on the hill slopes aside glacier tongues make me further doubt about the validity of velocity fields extracted in this study; also see comment on Line 167-170.

Figure 4 shows low velocity areas, or areas which have not changed significantly between the two Landsat scenes used in image correlation. We have updated this figure to be clearer.

Figure 9: See comment on Line 256-263.
Figure 10: See comment on Line 264-277.
Figure 11: See comments on Line 279-283 and Line 283.

Figure 12: See comment on Line 290-298.
Figure 13: See comment on Line 256-263.

Figures have been updated in the provided manuscript, specific commentary on each update is found in the previous comment replies.

Anonymous – General Comments

My main issue here is that the study basically compares results to either zero (only spectral mapping) or one (full manual delineation) with the obvious results shown in Figs. 9 to 12 on an aggregated level. However, a large number of more or less sophisticated methods have been developed in the meantime for mapping debris-covered glaciers (thermal bands, decision tree classifiers, object based image analysis, use of coherence images, etc.). Apart from one very simple method that is at least mentioned as a base for this study (but also not used for comparison), none of these other methods are mentioned or compared to.

I think it is beyond the purpose of this study to also apply all these methods and compare the results of the say 'velocity - river network approach' presented here to the outcome of these methods. But I expect at least a tabular summary of these other methods with columns listing what they need as an input, where they have been tested, and what their pros and cons are (e.g. regarding processing time, data availability, quality, required post-processing), as well as a direct comparison to at least one of the other methods (e.g. the method described in the study by Paul et al. 2004) that do not require extensive additional processing (like the velocity fields for this study) and can thus be easily implemented. This would also help to much better see what the advantages of the more complex method presented here is over what is already available. In short, I like the idea with the velocity fields and the river seed points but I need to see how results improve by the extra-effort required to get this information, and what this extra effort is (there is only a random note on this in L 358/9).

A discussion of additional algorithms has been added to Section 5.1. We do not provide a comparison with the Paul et al. (2004) methods, as they leverage proprietary software and FORTRAN codes to which we do not have access. This is discussed further below in regards to specific comments on the feasibility of comparisons with other algorithms.

There are two further important points: one is the missing presentation of results (e.g. the total glacier area mapped) and one is the critical discussion of the method in view of recently published datasets (GAMDAM and the new Chinese glacier inventory). Why is it worth applying this method despite these new datasets and considering the intention that the method might not provide accurate outlines for individual glaciers? As a smaller point, I still do not understand why the TM and ETM+ data have to be co-registered to a master image. The level 1T product from USGS is in general highly accurate over glaciers (positional variability < 1 pixel), in particular the ETM+ scenes used to generate the GLS2000 global reference dataset. This extra effort seems distracting for others to apply the method and I think this is a bad idea. Maybe the

authors can check what the effect of NOT additionally geocoding (with AROP) the already orthorectified level 1T product on the results is and remove this part from the pre-processing description if the effects are small.

We have added the total glacier area mapped (Section 4.1). We have also included visual comparison with the Chinese Glacier Inventory v2 (Figure 14). Unfortunately, the GAMDAM database has not yet been published, so we cannot comment on its accuracy. Section 5.1 also now contains a short overview of the utility of our algorithm as compared to static datasets, such as the CGI, RGI, or GAMDAM. We see the most important difference in use case is that our algorithm can be used for multi-temporal studies. While a single, well constrained, and well verified glacier outline is perfect for some use cases, a dataset which is not tied to a single time slice can also be appropriate. For example, change detection studies will need more than a single outline for each glacier.

We have tested our algorithm without the use of the AROP package, and find that it is superfluous. We instead rely on the geocoding in the L1T products. Our original reliance on AROP was due to many Landsat 5 images not being available at Level 1T until recently.

Anonymous – Specific Comments

Abstract: I suggest rewriting the abstract with some care to better motivate the study (what is the key shortcoming that was responsible for this study? Certainly not that there are two (not several) inventories that provide only a one point in time snapshot). Show that the method presented here is not only a significant improvement over purely spectral-based classification (this is not a big deal and applies to several other algorithms as well) but that it is better than (in terms of accuracy, time required or whatever) than the current state of the art. It not only limits longitudinal studies, it limits all studies (so just remove longitudinal). Please use debris-covered glacier tongues instead of 'glacier debris tongues' and remove the sentence breaks. 'The relationship between Landsat band …' should be better specified (e.g. 'such as the band ratio with Landsat using bands …')

The abstract has been updated based on both reviewer comments.

L2: No, its changes in glacier length. We do not yet have any useful relation between area changes and climate change.

We do not discuss glacier length changes in this manuscript, but we do briefly discuss area changes. We argue that both represent a proxy for climate forcing, albeit with different relationships. In this case, we view length and area changes both as proxies for climate forcing.

L3: I would not say more recent studies and although. There is neither a temporal preference nor is it related to remote sensing techniques (which would include DEMs derived from aerial photography. You might say that studies on volume changes have increased with free the availability of DEMs (e.g. SRTM) and altimetry (e.g. ICESat) data.

This has been removed.

L5: Instead of Stocker et al. (2013), I recommend citing Vaughan et al. (2013), which is the Cryosphere Chapter.

This line has been removed based on both reviewer comments.

L8: notably the one from GLIMS (Global Land Ice …) and … the Randolph Glacier Inventory (RGI)

We feel this is a stylistic choice, and have chosen to keep full names outside of the brackets and acronyms inside.

L10: I think it would be important to have a short comment also on the now available GAMDAM and new Chinese Glacier inventories. They both have delineated all the debris-covered parts manually and might be more accurate than the outlines provided from the algorithm presented here. To this end, it should be better described what the further benefits of the method presented here are.

This has been added in Section 5.1, as well as Figures 10 and 14.

L14: The cited study has a focus on technical challenges for glacier mapping. Maybe add another citation for the societal impacts.

This has been added.

L30: When the method presented here builds upon the methods developed by Paul et al. (2004), I think it would make much sense to directly compare the results to the outcome of this earlier method (outline overlay, quality, workload, issues for post processing, etc.). An improvement over the pure spectral mapping is rather easy to achieve.

We have added this comparison in Section 5.1.

L32: I am uncertain if these goals really apply to this study. The method is rather complex (requires different software packages and intense pre-processing) and not really tested globally (e.g. to the stagnant debris-covered glacier in the Mt. Everest region).

We have simplified the algorithm as much as possible, and minimized the number of software packages required. We argue that the diverse types of glaciers present in the Tien Shan constitute a widely-applicable test bed for the algorithm. A truly global test of the algorithm was outside the scope of this study.

L43/44: This is all fine, but what about the variability in surface velocities? I think a critical part of the presented method is its applicability to glaciers flowing very slowly. Are there examples in the sample?

We have not seen any totally stagnant glaciers in our study area. In the case of totally stagnant glaciers (dead ice), using only the horizontal component of the velocity vector, or an alternate measure of 'difference' between two cross-correlated images in the place of a velocity measurement as used here, may be more appropriate. As the velocity cutoff is left to the user to decide, a region-specific velocity cutoff could also be tested.

L59: It could also be possible to have a high diversity of glaciers in only one climatic zone. Or does this statement relate to differing mass balance gradients / temperature regimes?

This statement refers to different temperature and precipitation regimes which are a contributing factor to the differing glacier types in the study region.

L66: Which version of the SRTM DEM has been used, the one with voids or the interpolated version from CGIAR?

We have used the void-filled SRTM V4.1 from CGIAR. This has been clarified in the manuscript.

L68: What method was used for down-sampling?

From the current manuscript (L57): 'The SRTM data and its derivatives were downsampled to 30m to match the resolution of the Landsat images using bilinear resampling.' We have left this as-is.

L76/99: Naming two different sections "Data preparation" seems unfortunate. I suggest using a different name for section 3.1 (pre-processing?).

We have updated this section.

L103/4: Ok, but what has been done in these cases? Where snow/cloud-covered images just excluded or somehow corrected?

Each image in Table 1 was processed, but areas with snow or cloud cover were not used in the statistical analyses presented in this manuscript.

L106-113: I do not understand why this is required? Is the Level 1T orthorectification by USGS so poor? For a region to the south (Karakoram), we found some shaking mountain crests in a 15 year time series of Landsat TM, ETM+ and OLI images, but everything else was precise within a pixel. I would never touch this. As this step with AROP is related to additional workload, making the method presented here less attractive, I strongly recommend checking if it can be removed. Less is more!

We have re-run our analyses without the use of AROP, and have chosen to remove this orthorectification step from the algorithm. See above.

L114: Maybe add what is calculated from the DEM

From L97 of the manuscript: 'The algorithm generates a slope image from the DEM'

L123: Debris-covered glacier tongues tend to …

This section has been removed.

L123: I think this statement needs to be more substantiated. What does 'tend to' mean? Is it the majority (say 90%) or only a few? As these lakes play an important (?) role for the algorithm, the question is what happens when they are not present and to how many glacier tongues this applies. Are the results without lakes as good as with lakes? I think this is an important assessment as global application has been mentioned as a possibility in the introduction and lakes on debris-covered glaciers might not be that common.

We have performed some sensitivity analyses on the algorithm, and find that in most cases the river networks are more useful as seed points than the glacier lakes. When used in concert with a few manual seed points, the use of supra-glacial lakes becomes somewhat superfluous. In the interest of simplifying the algorithm, we have removed this section. An updated algorithm is again provided with the supplement.

L129: All these algorithm-tuning steps performed manually need to be accounted for in the workload budget to allow an honest comparison of the increase in accuracy vs the increase in workload (compared to less complex methods).

Please see the discussion of this in Sections 5.1 and 5.3.

L131: far away from any glacier

This has been updated.

L142: Where does this value of >250 come from? The TM1 threshold is designed to improve classification in regions of shadow. Even fresh snow does not have DNs > 250 in shadow? I might be wrong, but from this statement I have to assume that even the simple band ratio method is not properly implemented in the processing workflow (?), thus providing results for the spectral mapping that are not as good as they could be.

This is a mistyped literature citation, from Hanshaw and Bookhagen (2014). In our algorithm, we use a TM1>60 for our clean ice delineation. This has been updated.

L148: please write 'debris-covered glacier tongues'

This has been updated.

L149: Why 'at high elevations' and not in shadow? What has the elevation to do with it?

Here we refer to steep faces which have a different spectral profile that tend to occur in high elevation areas in the accumulation zone. This is what we refer to here. We have updated the manuscript for clarity.

L153: As mentioned in the general comments, when this method is building on this former study, it would be good to compare the results achieved with the methods developed here against it. The comparison against the pure spectral classification is much less interesting as there are meanwhile so many algorithms doing better.

We have added a discussion of this and other algorithms to Section 5.1.

L161: There are too many decisions merged in Fig. 3, the striped pattern selected for illustration is too imprecise, and the area shown is too large. So at first, please zoom in (to 1/5 of the image), second illustrate the effects of the respective binary masks in three subsets (slope threshold, slope variability, elevation range), and finally use a grey-scale background image and semi-transparent colour-coded areas to visualize the effects of each sub-step. The striped regions now also include the very steep headwalls of glaciers, but this cannot be true when regions steeper than 24 degrees are filtered. So please check and revise the figure.

This figure has been updated.

L166ff: It would be good to show a classified velocity map also in the main paper. This is in my opinion the core of the here presented method and results might sensitively depend on the selection of the correct velocity threshold. Without knowing how the resulting velocity ranges over glaciers look like, it is difficult to imagine how Fig 4 was produced.

We have modified Figure 4 in text.

L177: I do not understand how it was possible to select snow free images for the correlation. The glaciers shown in Fig. 4 are heavily snow-covered and show deep shadows. How was it possible to derive meaningful correlations (from optical images) and henceforth velocities larger than a given threshold in these regions? I do even not see any noise in these regions, which is hard to believe. Please clarify.

We hope that the changes to Figure 4 help explain this. We endeavored to choose the 'cleanest' image available, but no image is perfect in this sense. As we are using the velocity measurement to *remove* stable ground, we look more closely at clean stable ground as opposed to very clean glaciers.

L190: Please add how long the velocity processing normally takes. Without this information it is impossible to see if the extra-effort is worth the improved result.

This has been updated in-text. We found that the method used in the paper took a few hours per velocity field on a standard desktop (no human interaction during this time).

L195ff: The distance filtering seems to be similar to the 8-point neighbourhood filter applied in the 2004 study, basically removing everything that is not connected to glaciers. Again, can an estimate of the required workload for these steps been added to get an impression on the required extra effort.

The method proposed by Paul et al. (2004) uses a proprietary, computationally expensive, 'image polygon growing' algorithm in the PCI software suite. Their method checked for connectivity between two binary images, and then fed the unique ID numbers of each region into a FORTRAN code. Using this, they removed all 'debris' pixels which did not have a connected set of 'glacier' pixels. Our method is simpler, and relies upon distance as opposed to connectivity. We use a set of seed points and remove all pixels outside of a distance threshold. This operates on a binary image, and is computationally inexpensive.

L213: As mentioned before, I only see here the river seed points (in Fig. 5) and the effects of everything in Fig. 6. But I do not see the map with the distances (for the various datasets) that has been used as a base for the removal. What are the distances that have been used as a threshold? Please add this information (parts of it was already shown in the rebuttal).

We have opted to not include a map with the distances between each pixel in the main manuscript, as we already show several images of the sequential processing. We have included such an image here in the reply:

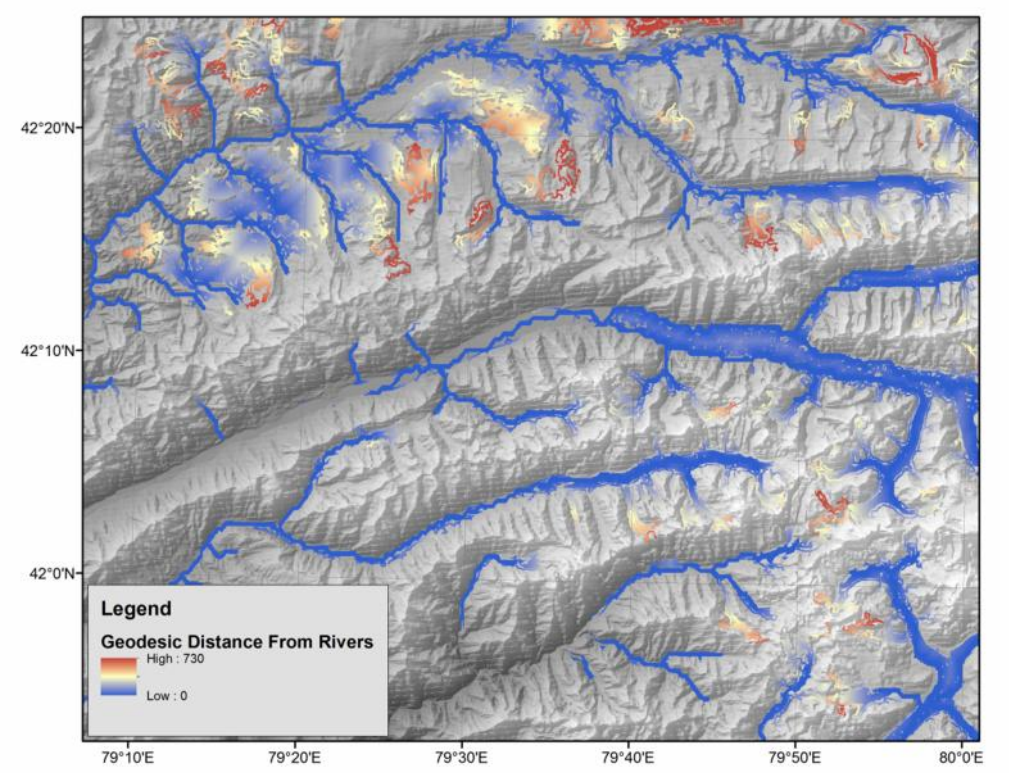

Figure 2: Geodesic distance from rivers.

We use a value of 90 in our distance weighting function, using the Matlab function 'graydist'. This is a geodesic distance transform, and does not conform to a direct 'number of pixels' measurement.

L216-226: As above, it would be nice to see the effects of the filtering steps on the binary masks (for a set of close-ups, not at the scale of Figs. 6 or 7). The filtering seems to be rather massive and I assume there is also some impact on smaller clean ice glacier extents. The effect can be quantified in the results or discussion, but it should be shown that it is a minor one (at least I hope so).

We have included a more close-up view of the filtering impacts in Figure 7. Clean ice is generally not impacted by the filtering, while isolated non-glacier areas are removed and holes in the debris tongues are filled.

L227: Please compare results with what can be achieved (I assume much faster) with the Paul et al. (2004) method. The comparison against the pure spectral classification is has a very limited meaning. Also: Place the spectral outlines on top of the final outlines to better see the real difference.

As Paul et al. (2004) do not provide timing or processor intensity measurements for their method, any comparison between workload is conjecture. However, based on the intensity of implementing a large-scale polygon growing algorithm, their algorithm is likely unsuitable for a

wide-area application as presented in this manuscript. We have included a discussion of this and other algorithms in Section 5.1. The figure has also been updated.

L230: Can a little bit of statistics be provided for this manually digitized dataset, for example the size class distribution and debris cover percentages for each class?

A size-class distribution has been included as Figure 9.

L248: The results section now contains a discussion of algorithm errors, a presentation of a further method (vertex distance matching), and a statistical comparison to a random sampling of glacier areas. There are no numbers about the derived glacier areas and how they compare to other datasets (e.g. the RGI, GAMDAM or new Chinese datasets). Is there a chance to add some results that look more like results? I suggest describing the vertex distance matching method in the methods section.

The vertex matching is not an essential part of the algorithm we propose, but rather a way to analyze the results of the algorithm. As such, we propose to keep the limited description of the matching process alongside the results of the vertex matching. We have added a new section to the results giving some statistics of the mapped area.

L250: I do not understand why the 750 glaciers are now reduced to 215 for comparison. What was wrong with the others? Has this sample still the same size class distribution?

We have reduced the glaciers processed here due to time and computation limitations. It is very computationally expensive to perform the statistical comparisons presented here, so we have chosen a representative sample of glaciers instead of using the entire manual dataset. The random sample comes from the entire 750 glacier control dataset, however. The size class distributions of the 750 and 215 glaciers are the same.

L254: I recognize that the method has not been developed to provide accurate outlines for individual glaciers. However, I find the comparison to elevation distributions too aggregated. Is there a possibility to add a scatter plot (size vs. relative size difference) showing how individual glaciers compare (maybe marking the heavily debris-covered glaciers with a different symbol)? Such a comparison would also justify the performed separation into individual glaciers (which is not required when only hypsometry is compared).

Please see Figure 1 of this reply and the discussion of this above.

L276/7: I recommend a comparison with a dataset derived from a more sophisticated method.

We have added a comparison statistic against the new CGI v2, for glaciers in our study region which also appear in the CGI (Figure 9).

L291: we used 465 …

This has been updated.

300: It would be nice if this part (5.1) of the discussion could expand a little bit on other methods that have been tested previously. Currently it is largely centred on what has been tested in this study. However, there are also object-based classification approaches, neural networks have been tested, and pattern recognition (also: thermal bands, decision tree classifiers, coherence images, and hybrid approaches). Maybe put these into context.

We have expanded this section to compare our method with other published methods.

303: "algorithm, but neither provided" (although sounds as it was clear from the beginning that they will not work)

Based on previous texture-analysis studies (ie, Racoviteanu and Williams, 2012), we hoped to find additional means of discriminating glaciers in the texture or frequency domain. Unfortunately, neither provided significant improvement.

L314: Please note that it is recommended to use glacierized rather than glaciated when referring to contemporary glaciers.

This has been updated.

L329: Please show that it also moves a step forward compared to more complex approaches, (such as Paul et al., 2004), that can be easily implemented here. That inclusion of additional measures (slope, vegetation, neighbourhood analysis) improves classification of debris-covered glaciers over pure spectral approaches is known since 2004.

A discussion of this has been included with our expanded treatment of previous algorithms.

L341: Please add here a short discussion on the results achieved when directly comparing individual glaciers (see comment at L254).

Please see the discussion of this below Figure 1 of this reply.

L345: This might be well the case, but is this something anybody would really do? As far as I know, scientists tend to always want to have the best possible dataset, independent of the scale of their application. And if a freely available source does not satisfy there quality needs, they digitize everything by themselves. Please also note that we now have the GAMDAM inventory and the new Chinese inventory for that region (both not yet in RGI 4.0). Maybe it would be sensible to demonstrate that the dataset created here is not obsolete in this regard.

While it is true that our algorithm output will not match the precision required for small-scale studies, such as mass balance studies, we maintain that it can be very useful for certain types of

studies. While some studies can rely on a single glacier measurement in time, others need a dynamic measure of glacier area. Our algorithm works on any given time period of Landsat coverage, and thus can provide a multi-temporal look at glacier areas. This is missing from the large public datasets.

L346: 'powerful tool': this statement requires some information on the required workload for data processing, considering all steps that are really important (a high workload would imply that the tool is not that powerful).

We have attempted to streamline the algorithm as much as possible, and remove as many manual steps as possible. There are also updated measures of processing time in section 5.1.

L351: 'mostly static': At least in this part of the world, try it in the Karakoram …

We refer here to the use of a single velocity, slope, and elevation measurement through different scene classifications. Thus, some of the input measures are 'static', even if the timeframe has changed.

L354: As mentioned above, demonstrate that the method presented here can be a substitute for algorithms such as Paul et al. (2004). Being better than pure spectral mapping is easy.

We have updated our discussion of this in Section 5.1.

L358/9: Here it is, a statement on processing time. Please expand on this to be transparent.

We have expanded this section.

L364: Indeed, this is now the problem. Please demonstrate why the method presented here is worth to apply anyway.

We propose the raw algorithm output for wide-area glacier studies, such as range-wide hypsometries. For smaller-scale studies, we suggest that the glaciers are manually edited to the researcher's needs.

L371-379: This sounds a little bit theoretical. I hope it can be replaced with a discussion of more practically relevant topics after revision.
L372: The glacier ice detection is mostly based on the SWIR band (where reflectance is very low) and TM1 in shadow, where saturation is also not a problem. Maybe think of a better explanation.

We have shortened this discussion.

L383/4: Please see above, it needs to outperform simple hybrid classifiers to be worth testing.

This has been updated.

L393/4: Please test this in the Mt. Everest range before stating it. Flow velocities are close to zero in this region for many of the heavily debris-covered glaciers.

We refer here to our study area as Central Asia. We have updated this to make it clearer.

L396: What is now the recommendation when I need accurate outlines for a larger region? Use first this algorithm and then manually correct the remaining errors or a native digitization of these parts?

The recommendation is to use our algorithm in studies which require a large number of glacier outlines through *time*. The algorithm can quickly generate a robust baseline across all timesteps, which can then be manually corrected to the needs of the study.

In my review above, I have suggested to add several further figures. These should have the form of a regular square to show at least too images side by side. Compared to Fig. 2, I also suggest showing a close-up (maybe 2/3 of the current Fig. 2). The regions currently shown in the various figures differ. I think this is not a good idea to trace the effects of the different processing steps. I suggest showing the processing steps in only one or maybe two different regions.

Based on reviews of the previous manuscript, we endeavored to show that our algorithm works in a wide range of areas, instead of focusing in on only one glacier or set of glaciers. We have updated the figures following this reviewer's suggestions, however.

Fig. 1: Caption: "and location of eight Landsat image footprints … along with their path/row combinations." By the way, why has scene 146/30 & 31 not been used to close the gap?

The original study was planned around a long swath of Central Asia, in which having a wider east-west range was more desirable than covering every single glacier in the Tien Shan/Pamir region. Thus, we did not include 146/30 and 146/31 in our study.

Fig. 2: Is there also an image available with less seasonal snow? It also seems as if the glaciers in shadow are not properly mapped. I might be wrong, but I think the band ratio can do better (by changing the TM 1 threshold).

Yes, the spectral ratio has trouble perfectly mapping areas in shadow. We have endeavored to use a ratio that is loose enough to catch most shadowed areas, but does not overclassify non-glacier areas.

Fig. 3: See comments to L161. Please show the same region as in Fig. 4 or 5.

This has been updated.

Fig. 4: The red on red is difficult to see. Please use a different colour and add the velocity map the mask is based on. Caption: This is not really a binary mask. It is an overlay of the velocity mask with an RGB composite image showing included regions transparent and excluded regions in red. Please be precise with the caption. Is this really a 7/5/3 composite in the background? It more looks like a 7/4/3 or 5/4/3 composite.

We have updated this figure.

Fig. 5: I suggest to also show the other seed points (e.g. lakes etc.) in this image. Can the image also illustrate what is distant, i.e. which regions will not be considered? Caption: 'The blue lines illustrate the presence of …'

We have updated this figure.

Fig. 6: Legend: The 'unfiltered outlines' look more like 'unfiltered areas' (i.e. the polygons are filled). Legend: 'Velocity threshold' is unclear. Are the black regions those that are still included after a threshold is applied? In this case I am not quite sure what the benefit of the velocity calculation is. I see velocity noise all over the image.

We have updated this figure. While there remains significant noise in the classification after the low-velocity areas have been removed, the velocity step is key for removing areas *near* to glaciers. The results of velocity filtering in areas distant from glaciers are less important, as those areas are removed by the distance weighting step.

Fig. 7: The blue outlines surrounding the blue glaciers are difficult to see, maybe use yellow? Maybe label some main glaciers in Fig. 2 to indicate where they are (Inylcheck, Tomur). Legend: Instead of 'unfiltered outlines' I would label the black polygons 'removed after filtering'.

We have updated this figure.

Fig. 8: As above: maybe use yellow lines instead of blue and place the red ones on top to better see what has been added by the algorithm.

We have updated this figure.

Fig. 9 to 12: Please remove the title from each of the plots and introduce some minor tick marks on the x and y axis on all plots. Please also consider showing two of these plots side-by-side (naming them Fig. 9a and 9 and 10a and b). All captions: I think the interpretation of what the figures show (this indicates) is not required in the caption but should be properly explained in the main text.

We have updated these figures after re-checking our manual control dataset. We have also updated the figures without titles.

Fig. 13: The orange lines are difficult to see, I suggest using yellow instead.

We have updated this figure and changed the color to purple.

Fig. 14: I think this is a bad example for illustrating glacier changes. The Landsat scene from Oct. 5, 2002 suffers from intense seasonal snow. In consequence, most of the 'area' changes visible in the image are due to the reduction in seasonal snow. The only real changes can be found at the terminus of the five clean-ice glaciers along the middle part of the image. At the 'debris-tongues' there is no change at all as far as I can see it. Maybe use a different example. The 'commonly misclassified river sand' needs to be marked (why has this type of misclassification not been removed with a vegetation filter?). The red outlines on the reddish background image are difficult to see, I suggest using yellow instead.

We have updated this figure and use a different image (August 2002). We have tried several filters, but the spectral range of cold, wet, river sand is very similar to that of debris on top of glacier ice. Thus, we have opted to overclassify some areas in river valleys as opposed to removing debris cover zones.

Improving Semi-Automated Glacial Mapping with a Multi-Method Approach: Applications in Central Asia

Taylor Smith[1], Bodo Bookhagen[1], and Forest Cannon[2]
[1]Institute for Earth and Environmental Sciences, Universität Potsdam, Germany
[2]Geography Department, University of California, Santa Barbara, USA

Corresponding authors:
Taylor Smith and Bodo Bookhagen
Institute for Earth and Environmental Sciences
Universität Potsdam
Potsdam-Golm 14476, Germany
Email: tsmith@uni-potsdam.de, bodo@geo.uni-potsdam.de

Abstract.

~~There exist several global glacier databases, but they are often limited to a single glacier-area measurement in time. This limits longitudinal studies of glacier change, particularly on hard to map debris-covered glaciers, which require extensive manual digitization~~Studies of glaciers often require extensive manual digitization in a Geographic Information System (GIS), as current algorithms struggle to delineate glacier areas with debris cover or other irregular spectral profiles. Although several approaches have improved upon spectral band ratio delineation of glacier areas, none have entered wide use due to complexity or computational intensity.

~~To minimize time spent on manual digitization, we designed and implemented~~ In this study, we present and apply a glacier mapping algorithm in ~~the Tien Shan~~ Central Asia which delineates both clean glacier ice ~~– methods which are well documented – and glacier debris tongues, which have distinct spectral signatures from clean glacier ice, and often require extensive manual intervention. This research improves upon methods developed to automatically delineate glacier areas using spectral, topographic, velocity ,~~ and ~~spatial relationships~~debris-covered glacier tongues. The algorithm is built around the unique velocity and topographic characteristics of glaciers, and further leverages spectral and spatial relationship data. We found that the algorithm misclassifies between 2 and 10% of glacier areas, as compared to a ∼750 glacier control dataset, and can reliably classify a given Landsat scene in 3-5 minutes.

The algorithm does not completely solve the difficulties inherent in classifying glacier areas from remotely sensed imagery, but does represent a significant improvement over purely spectral-based classification schemes, such as the ~~relationship between Landsat bands one,~~ band ratio of Landsat 7 bands three and five or the Normalized Difference Snow Index. The main caveats of the algorithm are (1) classification errors at an individual glacier level, (2) reliance on manual intervention to separate connected glacier areas, and (3) dependence on fidelity of the input Landsat data.

1 Introduction

~~Changes in glacier area have long been considered one of the best indications of climate change (i.e. Oerlemans, 2005) , although more recent studies have also assessed glacier volume changes with the advent of new remote sensing techniques (e.g., Berthier et al., 2007; Aizen et al., 2007; Gardelle et al., 2012, 2013; Bolch et~~

This study focuses on mapping glaciers over a large spatial scale using publicly available remotely sensed data. Several high-resolution glacier outline databases have been produced, most notably the Global Land Ice Measurements from Space (GLIMS) project (Armstrong et al., 2005; Raup et al., 2007, 2014), and the recently produced supplemental GLIMS dataset known as the Randolph Glacial Inventory (RGI) v4.0 (Arendt et al., 2012; Pfeffer et al., 2014). Smaller-scale glacier databases are also available, such as the Chinese Glacier Inventory (CGI) v2 (Guo et al., 2014) . A coherent, complete, and accurate global glacier database is important for several reasons, including monitoring global glacier changes driven by climate change, natural hazard detection and assessment, and analysis of the role of glaciers in natural and built environments, including glacier contributions to regional water budgets and hydrologic cycles ~~(Racoviteanu et al., 2009)~~ (Racoviteanu et al., 2009; Stocker, 2013) . Precision in glacier outlines is of utmost importance for monitoring changes in glaciers, which may change less than 15-30 m/yr (∼1-2 pixels of Landsat Enhanced Thematic Mapper (ETM+)/yr). Thus, spatially accurate glacier outlines are imperative for precise glacier change detection (Paul et al., 2004, 2013).

Several methods have been developed to delineate clean glacier ice (i.e. Hall et al., 1987; Paul, 2002; Paul et al., 2002; Racoviteanu et al., 2008a,b; Hanshaw and Bookhagen, 2014), relying primarily on spectral data available on satellites such as Landsat and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). Although significant progress has been made towards automated glacier outline retrieval using satellite imagery, these methods struggle to accurately map debris-covered glaciers, or other glaciers with irregular spectral profiles (Paul et al., 2004; Bolch et al., 2007; Racoviteanu et al., 2008b; Scherler et al., 2011a). Much of this ~~difficultly~~ difficulty stems from the similarities in spectral profiles of debris located on top of a glacier tongue and the surrounding ~~hillslopes where the debris is soureed from~~landscape. The majority of studies examining debris-covered glaciers employ extensive manual digitization in a Geographic Information System (GIS), which is very time consuming, and can introduce significant user-generated errors (Paul et al., 2013; Pfeffer et al., 2014; Raup et al., 2014). Building on the multi-spectral, topographic, and spatially-weighted methods developed by Paul et al. (2004), we present a refined rules-based clas-

sification algorithm based on spectral, topographic, ~~land-cover,~~ velocity, and spatial relationships between glacier areas and the surrounding environment. The algorithm has been designed to be user-friendly, globally applicable, and built upon open-source tools.

2 Study Area and Data Sources

2.1 Study Area

In this study we use a suite of 62 Landsat Thematic Mapper (TM), Enhanced Thematic Mapper+ (ETM+) and Optical Land Imager (OLI) images (1998-2013) across a spatially and topographically diverse set of study sites comprising eight Landsat footprints (Path/Row combinations: 144/30, 145/30, 147/31, 148/31, 149/31, 151/33, 152/32, 153/33) along a ∼1,~~500km~~ 500 km profile from the Central Pamir to the Central and Central-Eastern Tien Shan (Figure 1, Table 1) to analyze the results of our classification algorithm.

The study area contains a wide range of glacier types and elevations, with both small and clean-ice dominated glaciers, as well as large, low-slope, and debris-covered glaciers. The diversity in glacier types in the region provides ~~and~~ an ideal test area, particularly in ~~regards to~~ mapping glaciers with long and irregular debris tongues, such as the Inylchek and Tomur glaciers in the Central Tien Shan (Shangguan et al., 2015) .

The wintertime climate of the study area is controlled by both the Winter Westerly Disturbances (WWDs) and the Siberian High, which dominate regional circulation and create strong precipitation gradients throughout the range, which extends from Uzbekistan in the west through China in the east (Figure 1) (Lioubimtseva and Henebry, 2009; Narama et al., 2010; Bolch et al., 2011; Sorg et al., 2012; Cannon et al., 2014). The western edges of the region tend to receive more winter precipitation in the form of snow, with precipitation concentrated in the spring and summer in the central and eastern reaches of the range (Narama et al., 2010).

~~Sorg et al. (2012) note that due to the interaction of the Siberian High – a semi-permanent thermal high-pressure system which extends from Siberia towards the Tien Shan – and more western continental atmospheric patterns such as WWDs, the study area has a distinct precipitation gradient with decreasing precipitation from the west to east. There is also a strong gradient between outer and inner reaches of the Tien Shan, as moisture is precipitated on the windward (northwest) side of topography, moving eastwards from the Pamirs (Sorg et al., 2012) . These climate zones contribute to the diversity of~~

~~glaciers in the study region.~~

2.2 Data Sources

Our glacier mapping algorithm is based on several datasets. The Landsat 5 (TM), 7 (ETM+), and 8 (OLI) platforms were chosen as the primary spectral data sources, as they provide spatially and temporally extensive coverage of the study area (Table 1). ASTER can also be used as a source of spectral information, but here we chose to focus on the larger footprint and longer timeseries available through the Landsat archive. In addition to spectral data, the 2000 Shuttle Radar Topography Mission V4.1 (SRTM) Digital Elevation Model (DEM) (∼90m, void-filled) was leveraged to provide elevation ~~, slope, and hillshade~~ and slope information (Jarvis et al., 2008). The SRTM data and its derivatives were downsampled to 30m to match the resolution of the Landsat images using bilinear resampling. The USGS Hydrosheds river network (15 second resolution, ∼500m) was also used as an input dataset (Lehner et al., 2008).

3 Methods

Our glacier classification algorithm uses several sequential thresholding steps to delineate glacier outlines. The scripts used in this study are available in the Data Repository, with updates posted to http://github.com/ttsmith89/GlacierExtraction/. It is noted if the step requires manual processing or is part of a script.

1. Data ~~Preparation~~ Pre-processing
    (a) ~~Landsat TM/ETM+ data is aligned to a single base image (Manual)~~ Velocity fields are calculated with Normalized Image Cross Correlation (Manual, can be automatized)
    (b) The Hydrosheds river network is ~~buffered to a polygon feature (Manual~~rasterized (Manual, can be automatized)
    (c) ~~Training lakes and any Manual Debris~~ Optional manual debris points are created (Manual, optional)
    (d) SRTM data is used to create ~~Slope and Hillshade images~~ a hillslope image (Python Script)
    (e) All input datasets are matched to a single extent and spatial resolution (30m) (Python Script)

2. Glacier Classification Steps

   (a) ~~Lakes are classified using the Normalized Difference Water Index (NDWI) (Python Script)~~ Clean-ice glacier outlines are created using Landsat Bands 1,3, and 5 (Matlab Script)

   (b) 'Potential debris areas' are generated from low-slope ~~and low slope-variability~~ areas (Matlab Script)

   (c) ~~Low elevation~~ Low-elevation areas are removed (Matlab Script)

   (d) ~~Low- and high-velocity~~ Low-velocity areas are removed (Matlab Script)

   (e) ~~Distance weighting~~ Distance-weighting metrics are used to remove areas distant from river networks or clean glacier ice (Matlab Script)

   (f) ~~Distance weighting~~ Distance-weighting metrics are used to remove areas ~~distant from glacier lakes~~ very distant from clean glacier ice and manual seed points (Matlab Script)

   (g) The resulting glacier outlines are cleaned with statistical filtering (Matlab Script)

3. Post-processing

   (a) Glacier outlines are exported to ESRI shapefile format for use in a GIS (Python Script)

3.1 Data Preparation

For accurate glacier delineation, we primarily used Landsat images which were free of new snow, and had less than 10% cloud cover. However, we have also included scenes with limited ~~snowcover and cloud cover~~ snow- and cloud-cover in our analysis to understand ~~the impacts of snow and clouds~~ their impacts on our classification algorithm. We find that the presence of fresh snow in images tends to overclassify glacier areas and classify non-permanent snow as glaciers. Additionally, cloud covered glaciers cannot be correctly mapped by the algorithm (Paul et al., 2004; Hanshaw and Bookhagen, 2014).

~~After selecting a series of Landsat images, we co-registered each image to a 'master' image specific to each Landsat Path/Row combination, using the Automated Registration and Orthorectification Package (AROP) (Gao et al., 2009) (Step 1(a)). Master images are denoted in Table 1 with an asterisk. This ensures that glacier outlines are properly matched, and that glacier outlines derived from subsequent Landsat scenes~~ We use the USGS Level 1T orthorectified Landsat scenes to ensure that the derived glacier outlines are consistent in space ~~. This step is considered optional for Landsat TM and ETM+, as the referencing only moves the images on the order of 1-2 pixels, and is not~~

needed for Landsat OLI, as high-quality georeferencing is already included in the L1T product. Once the data were georeferenced and registered, a pair of scripts perform the glacier classification. (Hansen and Loveland, 2012; Nuimura et al., 2014) .

The algorithm uses Landsat imagery, a hydrologically corrected void-filled DEM, a velocity surface derived from image cross-correlation, and the Hydrosheds 15s river network (buffered by 200m and converted to a raster) as the primary inputs (Steps 1(ba) and 1(eb)). The algorithm generates slope and hillshade images a slope image from the DEM according to image date, time, and location, and rectifies additional input datasets described below for processing by resampling and reprojecting each dataset to the same spatial extent and resolution (30m to match the Landsat data) (Steps 1(ed) and 1(fe)). Although the current algorithm leverages a few proprietary Matlab commands, we will continue to update the code with the goal of using only open-source tools and libraries in the future.

3.2 Lake Delineation

Glacial debris tongues tend to host supra-glacier lakes, particularly during the summer months when snow cover is lowest (Quincey et al., 2007; Gardelle et al., 2011) . These lakes are used as seed points for distance-weighting mechanisms to more accurately delineate glacier debris tongues. Lakes are delineated using the NDWI (Gao, 1996) , after which misclassified areas are removed by masking out high slopes and shadowed areas (Step 2(a)) (Huggel et al., 2002; Worni et al., 2013; Nie et al., 2013; Hanshaw and Bookhagen, 2 To increase the modularity of the algorithm, we did not rely on a fixed NDWI value across all datasets, but instead used a manually generated set of index lakes that exhibit the spectral properties desired. Several lakes in each Landsat scene are selected, with a range of geographic settings (e.g., far removed from any glacier, on top of a debris tongue, at the base of a glacier). A small spectral buffer (0.025) is then added to the by-scene NDWI value to account for lakes with differing spectral signatures to those of the index lakes (generally below -0.05). This small manual step greatly increases the number of correctly classified lakes by providing scene-specific NDWI thresholds for each processed image.

3.2 Clean Ice Delineation

Calculations are performed on rasterized versions of each input dataset, which have been standardized to the same matrix size. The first step in the classification process leverages Landsat 7 Bands 1, 3, and 5 (Step 2(ba)). For Landsat 8 OLI images, a slightly different set of bands is used to

conform to OLI's modified spectral range. For simplicity, bands referenced in this publication refer to Landsat 7 ETM+ spectral ranges. The ratio of TM3/TM5 (value $\geq$2), with additional spectral information from TM1 (value >~~250~~25) has been used in previous research as an effective means of delineating glacier areas (e.g., Hall et al., 1987; Hanshaw and Bookhagen, 2014), but is not effective in delineating debris-covered glacier areas (Figure 2). In our algorithm, we use a threshold of TM3/TM5 $\geq$ 2 and TM1 >60 to map clean glacier ice. The end result of this step is the spectrally-derived glacier outlines, which are later integrated back into the workflow before statistical filtering (Figure 2). Here we choose ~~the~~ fairly conservative threshold values to ensure that we do not remove clean glacier ice. We find that increasing the TM1 threshold results in tighter classification of ~~glacier debris~~ debris-covered glacier tongues, but also removes some areas properly classified as glacier, particularly ~~at high elevations~~in steep areas of the accumulation zone. Thus, we err on the side of overclassification with our delineation of clean glacier ice.

3.3 Debris-covered Ice Delineation

3.3.1 Topographic Filtering

Building on the work of Paul et al. (2004), low slope areas (between 1 and 24°) are isolated as areas where debris-covered glaciers are likely to exist (Step 2(~~e~~b)). ~~As glacier surfaces tend to be rougher than the surrounding areas, a standard deviation filter (3x3 kernel size) is also applied to the slope and used to mask out areas of low slope variability.~~ Low elevation areas (automatically defined on a scene-by-scene basis based on the average elevation of clean-ice areas, generally below 2500-3000m in ~~our~~ the study area) are then masked out to decrease processing time (Step 2(~~d~~c)). These thresholding steps are performed independent of the previous, spectrally delineated, glacier outlines. In essence, this step identifies areas where there is the potential for a ~~debris covered~~ debris-covered glacier to exist. Additional thresholding is then performed on this 'potential debris area' subset to identify debris-covered glacier areas (Figure 3).

As can be seen in Figure 3, extensive areas which are not glacier or glacier debris tongue are identified in this step. However, this step generally removes all pixels outside of the main glacierized areas of any scene, and allows the algorithm to work on a subset of the image, thus reducing processing time. The next step uses a generalized velocity surface to subset the 'potential debris area'.

3.3.2 Velocity Filtering

The Correlation Image Analysis Software (CIAS) (Kääb, 2002) tool, which uses a method of statistical image cross-correlation, is used to derive glacier velocities from Landsat Band 8 panchromatic images. This method functions by tracking individual pixels across space and time, and provides a velocity surface at the same resolution as the input datasets (15m) (Step 1(~~b~~a)). The velocity surface is then upsampled using bilinear resampling to provide a consistent velocity estimate across the entire Landsat scene. We then standardized the velocity measurements to m/yr using the capture dates of the two Landsat images. As glacier velocity can change significantly throughout the year, and clean images were not available ~~for~~ at exactly the same intervals for each Path/Row combination, there is some error in our velocity fields. ~~We mitigate this range in velocities by using Path/Row-specific velocity thresholds during our classification process~~However, as the velocity surface is used to remove stable ground, which is generally well-defined despite changes in glacier velocities, errors in the velocity surface do not contribute significantly to glacier classification errors, excepting on slower-moving parts of debris-covered glacier tongues. It is important to note that ~~cloud-free~~ cloud- and snow-free images are essential for this step, as the presence of snow or cloud cover can disrupt the correlation process, resulting in anomalous velocity measurements. An example ~~binary velocity mask~~ velocity surface is shown in Figure 4 (Step 2(~~e~~d)). ~~This mask shows areas which~~ Red areas are removed from the 'potential debris areas' ~~in red~~, as they fall outside of the expected range of debris-tongue velocities.

We only used one multi-year velocity measurement for each path/row combination to derive general areas of movement/stability for glacier classification, as using stepped velocity measurements over smaller time increments did not show a noticeable improvement in glacier classification. This also improved our classification of slow-moving glaciers, which may not change significantly over only a single year. These velocities ranged generally from ~~4-30~~ 4.5-30 m/yr across the different scenes~~, with very few glacier pixels falling outside of this range. Scene-specific thresholds were chosen based on both manual inspection of the velocity surfaces, and their impact on glacier classification on a scene-specific test dataset. Furthermore, a~~. A single velocity threshold of 5 m/yr was used across all scenes to remove stable ground. A method of frequential cross-correlation using the co-registration of optically sensed images and correlation (COSI-Corr) tool (Leprince et al., 2007; Scherler et al., 2011b) was tested and did not show any appreciable improvement in velocity measurements (Heid and Kääb, 2012).

The velocity step is most important for removing hard-to-classify pixels along the edges of glaciers, and wet sands in riverbeds. These regions are often spectrally indistinguishable from debris tongues, but have very different velocity profiles. It is important to note, however, that this step also removes some glacier area, as not all parts of a glacier are moving at the same speed. This can result in small holes in the delineated glaciers, which the algorithm attempts to rectify using statistical filtering. Generating a velocity field is the most computationally expensive step of the algorithm.

3.3.3 Spatial Weighting

After topographic and velocity filtering, a set of ~~spatially-weighted~~ spatially-weighted filters was constructed. The first filtering step uses the Hydrosheds river network to remove 'potential debris areas' which are distant from the center of a given glacier valley (Figure 5, Step 2(~~fe~~)). As glaciers occur along the flowlines of rivers, and the Hydrosheds river network generally delineates flowlines ~~all the way~~ nearly to the peaks of mountains, the river network provides an ideal set of seed points with which to remove misclassified pixels outside of river valleys. A second distance weighting is then performed using ~~both the supra-glacier lakes detected during the lake delineation step, and~~ the clean-ice outlines generated in Step 2(a), as well as any manual seed points provided (Step 2(~~gf~~)). As debris tongues must occur in proximity to ~~one or more of: (1) glacier areas , (2)~~ either glacier areas or the centerlines of valleys, ~~or (3) supra-glacier lakes,~~ these two steps are effective in removing overclassified areas (Figure 6). At this step, it is possible to add manual seed points, which may be necessary for some longer debris tongues. We note that these are optional, and the majority of glaciers do not need the addition of manual seed points. However, for certain ~~glaciers which do not have many lakes, or do not have lakes large enough to be delineated by Landsat~~irregular or cirque glaciers, the addition of manual ~~control~~ seed points has been observed to increase the efficacy of the algorithm. In processing the Landsat imagery presented here, we have not used additional manual seed points.

The spatial weighting step is essential for removing pixels spatially distant from any ~~glacier or lake~~ clean-ice area. In many cases, large numbers of river pixels, and in some cases dry sand pixels, have similar spectral and topographic profiles to debris covered glaciers. This step effectively removes the majority of pixels outside the general glaciated area(s) of a Landsat scene, as can be seen in Figure 6~~, which shows the results of the velocity filtering (black), underneath the results of the spatial filtering (blue)~~.

3.3.4 Statistical Filtering

Once the spatial weighting steps are completed, ~~the glacier outlines are generally accurate. A~~ a set of three filters are then applied, in order to remove isolated pixels, bridge gaps between isolated glacier areas, and fill holes in large contiguous areas (Step 2(~~h~~g)). First, a 3x3 median filter is applied, followed by an 'area opening' filter, which fills holes in contiguous glacier areas. Finally an 'image bridging' filter is applied to connect disjointed areas, and fill holes missed by the area opening filter.

This step is essential for filling holes and reconnecting separated glacier areas. As our initial filtering methods are based on a fixed set of threshold values, there are often glacier pixels which are removed. For example, some pixels in the middle of a debris tongue may be moving more slowly than the provided velocity threshold, and are thus removed. This problem is somewhat, but not completely, mitigated by the statistical filtering (Figure 7).

~~When the final filtered outlines are compared to the original spectral outlines, there is a clear improvement in~~ The improved classification of debris ~~tongues (Figure 8)~~areas between the clean-ice and final algorithm outputs can clearly be seen in Figure 8.

3.4 Creation of Manual Control Datasets

Manual control datasets encompassing ~750 glaciers (~~3000~~ 11,000 km^2) were created to test the efficacy of the glacier mapping algorithm. These datasets were digitized off of Landsat imagery in a GIS, and then corrected with higher resolution imagery in Google Earth. The datasets are coherent in space, but cover two different times (~2000 and ~2011, depending on the dates of the available Landsat scenes). The bulk of the manually digitized glaciers fall within the boundary of Landsat Path/Row combination 147/031, as this is the most heavily glacierized sub-region of our study area. However, we have digitized glaciers throughout the eight Path/Row combinations to avoid biasing our statistics and algorithm to one specific scene extent. We have also considered a wide range of size classes in our manual dataset (<0.5 km^2 to 500+ km^2), as well as both clean and debris-covered glaciers. We note that although the manual datasets here are considered 'perfect', there is inherent error in any manual digitization in a GIS (e.g., Paul et al., 2013). Due to the lack of ground truth information, we have estimated the overall uncertainty of the manual dataset to be 2% based on previous experiments (Paul et al., 2002, 2013). Figure 9 shows the size class distribution of the manual control dataset, with logarithmic area scaling.

Before any comparisons between glaciers can be performed, glacier complexes must be split into

component parts. A set of manually edited watershed boundaries, derived from the SRTM DEM, were used to split both the manual and algorithm datasets into individual glacier areas for analysis. In this way, the diverse datasets and classified glacier areas can be split into the same subset areas for statistical comparison.

4 Results

Over the eight Landsat footprints used in this study, we map ~44,000 km^2 of glaciers over a two distinct time slices. Several additional time periods were mapped, but not included in the statistical analysis presented in this manuscript.

4.1 Statistical Analysis of Algorithm Errors

A subset of 215 glaciers from the manual control datasets of varying size and topographic setting was chosen for more detailed analysis. The unedited, algorithm-generated, glacier outlines were compared against ~~both~~ spectral outlines, which only classify the glacier areas via commonly used spectral subsetting (using TM1, TM3, and TM5, produced in Step 2(b))~~and ,~~ the manual control datasets, and the CGI v2. Figure 10 shows the bulk elevation distributions across 215 glaciers for each dataset in 10m elevation bins.

There is some apparent bias in our algorithm towards low elevation areas, which represent the debris-covered portions of glaciers and are the most difficult areas to classify. There is also a bias in our control dataset towards underclassifying the high elevation areas, which we attribute to user bias in removing isolated rock outcrops within glaciers, as opposed to simply defining accumulation areas as a single polygon. In general, the algorithm and the control dataset are well matched below 4000 meters; above this the spectral dataset and the algorithm dataset begin to align closely and generally follow the manually digitized data. This threshold represents the general transition from debris-covered glaciers to clean glacier ice in the study area. Our algorithm output is also well-matched with the CGI v2, except at very high elevations where it overclassifies some areas as compared to the CGI v2.

In order to examine ~~persistent~~ inherent bias throughout the algorithm classification, under- and over-classified areas for a subset of the control dataset were examined. To determine areas of over-classification (underclassification), the manually (algorithm) generated dataset was subtracted from the algorithm (manual) dataset, leaving only pixels which are overclassified (underclassified). Figure

11 shows the elevation distributions of under and over classified areas. The algorithm tends to consistently overclassify areas across the range of glacier elevations, which we attribute here to differences in manual and algorithm treatment of steep and de-glaciated areas within glacier accumulation zones. Importantly, the algorithm underclassifies a much smaller number of pixels, generally corresponding to areas below 4000m, where debris tongues are dominant. The majority of these pixels are along the edges of glacier debris tongues, which are removed by the algorithm due to their low relative velocity. It is also possible that some of these pixels are 'dead ice', which is difficult to differentiate from debris tongues by visual inspection. The total misclassification of algorithm-derived outlines against two independent manual control datasets are 2% and 10% respectively, which represents a significant improvement from a pure spectral delineation approach.

To investigate sampling bias in our analysis, we used 465 GLIMS glacier identification numbers (centroids, point features) which overlapped with the manual control datasets. A random subset of 100 of these points was chosen for this analysis. As can be seen in Figure 12, similar patterns emerge between the randomly sampled glaciers and the sampling used in other sections of this manuscript. There is evidence of more noise in the random sample, as some glaciers which we avoided due to closeness to wet sand/or other hard-to-classify areas were chosen during the random sampling. However, the relationship between the algorithm and the manual datasets remains significant (Kolmogorov–Smirnov test passes at 99% confidence interval).

4.2 Vertex Distance Matching

To capture changes in the shape of the glacier outlines between the initial spectral classification and the final algorithm output, we computed the distance between pairs of glacier vertices. We first reduced our manual control dataset to component vertices, which were then matched to the closest vertex in the spectral and final algorithm results polygons, respectively. The results of this distance matching can be seen in Figure 13.

The distance distribution for the algorithm dataset shows generally close agreement between the algorithm and manual control datasets. The spectral dataset also contains a large percentage of vertices close to a 1:1 agreement with the manual control dataset, which are primarily those vertices at the upper edges of glaciers, or vertices from small, debris-free glaciers. The difference in these two distributions is attributed to the increased precision with which the algorithm maps debris-covered glacier outlines.

4.3 ~~Comparison to a Random Sampling of Glacial Areas~~

5 Discussion

~~In order to examine sampling bias in our analysis, the researchers used 465 GLIMS glacier identification numbers (centroids, point features) which overlapped with the manual control datasets. A random subset of 100 of these points was chosen for this analysis. As can be seen in Figure 12, similar patterns emerge between the randomly sampled glaciers and the sampling used in other sections of this manuscript. There is evidence of more noise in the random sample, as some glaciers which we avoided due to closeness to wet sand or other hard-to-classify areas were chosen during the random sampling. However, the relationship between the algorithm and the manual datasets remains significant (ks-test passes at 99confidence interval).~~

5.1 Comparison with Previous Glacier Mapping Algorithms

~~Averaged elevation differences for a random sample of glaciers overlapping a manual control dataset (n=100). Shows generally tight agreement between the manual glacier dataset and the algorithm dataset below 4000m, with tighter agreement between the spectral and algorithm datasets above 4000m. Indicates improved mapping of debris-tongues by the algorithm, and similar treatment of clean ice by both the algorithm and the spectrally-delineated glaciers.~~ Several authors have presented alternative debris-covered glacier classification methods and schemes (e.g., Taschner and Ranzi, 2002; Paul et al., 2004; Bolch et al., While all of these methods present improvements over basic clean-ice delineation as proposed by Hall et al. (1987) , they each have shortcomings which limit their range of use. Table 2 shows a comparison of these different methods alongside the algorithm presented in this study.

6 ~~Discussion~~

Our study improves on previous work in three main ways: (1) computational intensity, (2) diversity of study area, and (3) temporal range of our dataset. The methods proposed in this study, excepting the generation of a velocity field, require very little processing power. Once initial input datasets (velocity surface, rasterized river network) have been created, a Landsat scene can be processed in 3-5 minutes. When this is compared with the training dataset creation, computationally expensive classification schemes, and neighborhood analyses employed by other studies, there is a clear improvement in efficiency. Secondly, we analyze a significantly larger glacier area than any of the previous studies,

which has helped us generalize our algorithm and methods to a wide range of topographic and landcover settings. Finally, we process a multi-year dataset, encompassing 62 Landsat scenes with varying landcover and weather settings. This has allowed us to further generalize our algorithm to be effective beyond a single scene or small set of scenes, and to remain effective across a wide spatial and temporal range. The time-dynamic aspect of our algorithm can also provide a complement to time-static wide-area datasets, such as the RGI v4.0, the CGI v2, and the forthcoming GAMDAM datasets (Arendt et al., 2012; Guo et al., 2014; Nuimura et al., 2014) . While these datasets may provide higher-quality manually digitized outlines for specific glaciers, they only provide a single snapshot in time, and are limited to a specific area of coverage.

5.1 Unused Filtering Steps

Two additional topographic indices – spatial Fast Fourier Transforms (FFTs), also known as 2D FFTs, and ASTER surface roughness measurements – were tested during the development of the algorithm, although neither provided significant improvement. We attempted to derive frequential information from several Landsat and ASTER bands, with limited success. Some glaciers exhibit a unique frequency signature when analyzed using spatial FFTs, although these were not consistent across multiple debris-covered glaciers with differing surface characteristics. Additionally, the FFT approach was tested against a principal component analysis (PCA) image derived from all Landsat bands, without significant improvement to the algorithm.

We also attempted to integrate surface roughness measurements using the ASTER satellite, which contains both forward looking (3N - nadir) and backwards looking (3B - backwards) images, primarily intended for the generation of stereoscopic DEMs. The difference in imaging angle provides the opportunity to examine surface roughness by examining changes in shadowed areas (Mushkin et al., 2006; Mushkin and Gillespie, 2011). We found that there are slight surface roughness differences between glaciated and non-glaciated areas on some debris tongues, but that these differences are not significant enough to use as a thresholding metric. Furthermore, the nature of the steep topography limits the efficacy of this method, as valleys which lie parallel the satellite flight path and those which lie perpendicular to the flight path show different results. Thus, the algorithm relies on the velocity and slope thresholds to characterize the topography of the glacier areas.

5.2 Algorithm Use Cases and Caveats

The glacier outlines provided by the algorithm are a useful first pass analysis of glacier area. It is often more efficient to digitize only misclassified areas, as opposed to digitizing entire glacier areas by hand (Paul et al., 2013). Paul et al. (2013) also note that for clean ice, automatically derived glacier outlines tend to be more accurate, and it is only in the more difficult debris-covered and shadowed areas that manual digitization becomes preferable. In the algorithm presented here, clean ice thresholding was implemented using TM1, TM3, and TM5~~, but as the algorithm primarily operates~~. However, because the algorithm operates primarily on 'potential debris areas', ~~as opposed to whole glaciers, different spectral classification schemes could be easily substituted for regionswhere other spectral measurements~~any clean ice classification scheme could be used. For example, in other study regions, or for different satellite sensors, other schemes, such as the Normalized Difference Snow Index, ~~perform best~~may outperform clean ice classification as implemented in this study.

The algorithm moves a step further than spectral-only classification and attempts to classify glacier areas as accurately as possible, including debris-covered areas. As can be seen in Figure 14, the algorithm compares well with both the control dataset and the ~~RGI v4.0~~ CGI v2 – a ~~~198, 000 glacier global dataset derived from multiple data sources~~ high-fidelity, manually edited, dataset – across a range of glacier types (Step 2(a)) (Guo et al., 2014) . However, ~~it does~~ the algorithm outlines do not perfectly align with either dataset. In Figure 14, a tendency to remove pixels along the edge of glacier debris tongues can be observed, which we attribute to the fact that the center of debris tongues often move faster than the edges. Furthermore, both the algorithm results and the manual control dataset underestimate glacier area as compared to the ~~RGI~~CGI, due to the removal of non-clean ice pixels at high altitudes or high slopes, which are generally within the accumulation area of a glacier but ~~rarely are~~ are not always covered by permanent ice. These two types of classification bias are easily rectified with minimal manual intervention. ~~Furthermore~~Some bias between the manual or algorithm datasets and the CGI v2 can also be attributed to the difference in time; while the manual and algorithm datasets share an image date, the CGI v2 was digitized on top of multiple images that may not match up perfectly in time with our datasets. Despite these misclassified areas, the raw algorithm output identifies the furthest reaches of the glacier tongues effectively in most cases, as can be seen in three long debris tongues shown in Figure 14.

Without post-processing, these raw glacier outlines can be used to analyze regional glacier characteristics, such as slope, aspect, and hypsometry. Even if glacier outlines are not perfectly rectified

in space, at the scale of watersheds, satellite image footprints, or mountain ranges, errors of under- and over-classification even out, yielding valuable regional statistics (Figure 10). As the method can be easily modified to fit the topographic and glacier setting of any region, it is a powerful tool for analyzing glacier changes over large scales over the period of Landsat TM, ETM+ and OLI coverage. Small glacier changes are also captured by the algorithm, as can be seen in Figure 15.

Figure 15 also illustrates some potential errors with the algorithm where river sand is sometimes delineated as glacier area. In many cases, the same areas are captured across different timestamps, as the topographic and velocity data used to define 'potential debris areas' is mostly static in time, excepting the distance weighting steps. However, these areas are ~~generally small and~~ easily removed during manual inspection of results.

The second use case for the algorithm is as a substitute for simple spectral ratios. Particularly in regions with numerous debris-covered glaciers, manual digitization of glacier tongues is time consuming. Our algorithm provides a robust baseline set of glacier outlines which can be corrected manually, with minimal extra processing time. ~~Processing time is also decreased when a large set of Landsat scenes are considered, as~~ As generating the input velocity surfaces can take longer than processing glacier outlines from dozens of Landsat scenes, efficiencies are gained when a large number of Landsat scenes are processed. The algorithm as published takes ~3-5 minutes of actual processing time once the base datasets have been created. For a single Path/Row combination, the time to set up the input datasets (velocity surface, manual debris points) is ~4 hours. Once the initial setup has been completed for a given Path/Row combination, an arbitrary number of Landsat scenes can be processed very quickly.

Although the algorithm represents a step ~~towards improved~~ forward in semi-automated glacier classification, there are several important caveats to keep in mind: (1) Lack of data density and temporal range limits the efficacy of individual glacier analysis; the algorithm presented in this paper was not designed with individual glacier studies in mind, and in many cases, such as in mass balance studies, more accurate manual glacier outlines are necessary. Furthermore, (2) the algorithm relies on manual intervention to separate individual glaciers which are connected through overlapping classified areas, or which are part of glacier complexes.

Finally, (3) the algorithm relies heavily on the fidelity of the Landsat images provided, in that glacier outlines on images with ~~cloud cover or snow cover~~ cloud- or snow-cover are less likely to be well defined. This creates a data limitation, as many ~~glaciated~~ glacierized areas are subject to

frequent ~~snow and cloud cover~~cloud- and snow-cover, and thus have a limited number of potentially useful Landsat images for the purpose of this algorithm. ~~We have found that OLI images are generally better classified than TM/ETM+ images, potentially due to the increased sensor range onboard OLI reduces saturation in the spectral bands used for glacier ice detection. The difference in the sensors (8-bit on TM/ETM+ and 12-bit on OLI) allows for better radiometric quantization and a higher signal to noise ratio. The change of OLI 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. However, this conclusion will need to be verified with further data acquisitions, and remains conjecture.~~

6 Conclusions

This study presents an enhanced glacier classification methodology based on the spectral, topographic, and spatial characteristics of glaciers. ~~Our algorithm represents a step forward toward~~ We present a new method of (semi-) automated glacier classification, ~~in that it outperforms spectral-only algorithms in wide use in the glaciology and remote sensing communities~~which is built upon, but unique from, the work of previous authors. Although it does not completely solve the difficulties associated with debris-covered glaciers, it can effectively and rapidly characterize glaciers over a wide area. ~~The first steps of this algorithm, which seek to characterize maximum possible glacier area, are lake delineation, spectral delineation, and topographic filtering. Following these steps~~Following an initial delineation of clean glacier ice, a set of velocity, spatial, and statistical filters are applied to accurately delineate glacier outlines, including their debris-covered areas.

When compared visually and statistically against a ~~control dataset, or a wide-area glacier database such as the Randolph Glacier Inventory v4.0~~manually digitized control dataset and the high-fidelity CGI v2, our algorithm remains robust across the wide range of glacier sizes and types found in Northern and Central Asia. The algorithm developed here will be applicable to a wide range of ~~glaciated~~ glacierized regions, particularly in those regions where debris-covered glaciers are dominant, and extensive manual digitization of glacier areas ~~is~~ has previously been required. The raw algorithm output is usable for rough statistical queries on glacier area, hypsometry, slope, and aspect; however, manual ~~intervention should be applied~~ inspection of algorithm output is necessary before

using algorithm glacier outlines for more in-depth area change or mass balance studies.

Acknowledgements. This work was supported through the Earth Research Institute (UCSB) through a Natural Hazards Research Fellowship, as well as the NSF grant AGS-1116105. We would like to thank Frank Paul~~and Wanqin Guo~~, Wanqin Guo, and one anonymous reviewer for their detailed and helpful reviews, as well as Tobias Bolch for his contribution to the development of the manuscript.

References

Aizen, V., Aizen, E., and Kuzmichonok, V.: Glaciers and hydrological changes in the Tien Shan: simulation and prediction, Environmental Research Letters, 2, 045 019, 2007.

Arendt, A., Bolch, T., Cogley, J., Gardner, A., Hagen, J., Hock, R., Kaser, G., Pfeffer, W., Moholdt, G., Paul, F., et al.: Randolph Glacier Inventory [v2. 0]: A Dataset of Global Glacier Outlines. Global Land Ice Measurements from Space, Boulder Colorado, USA, Digital Media, 2012.

Armstrong, R., Raup, B., Khalsa, S., Barry, R., Kargel, J., Helm, C., and Kieffer, H.: GLIMS glacier database, National Snow and Ice Data Center, Boulder, Colorado, USA, 2005.

Berthier, E., Arnaud, Y., Kumar, R., Ahmad, S., Wagnon, P., and Chevallier, P.: Remote sensing estimates of glacier mass balances in the Himachal Pradesh (Western Himalaya, India), Remote Sensing of Environment, 108, 327–338, 2007.

Bolch, T., Buchroithner, M. F., Kunert, A., and Kamp, U.: Automated delineation of debris-covered glaciers based on ASTER data, in: Geoinformation in Europe (Proc. of 27th EARSel Symposium, 04-07 June 2007), Bozen, Italy, pp. 403–410, 2007.

Bolch, T., Peters, J., Yegorov, A., Pradhan, B., Buchroithner, M., and Blagoveshchensky, V.: Identification of potentially dangerous glacial lakes in the northern Tien Shan, Natural Hazards, 59, 1691–1714, 2011.

Bolch, T., Kulkarni, A., Kääb, A., Huggel, C., Paul, F., Cogley, J., Frey, H., Kargel, J., Fujita, K., Scheel, M., et al.: The state and fate of Himalayan glaciers, Science, 336, 310–314, 2012.

Cannon, F., Carvalho, L., Jones, C., and Bookhagen, B.: Multi-annual variations in winter westerly disturbance activity affecting the Himalaya, Climate Dynamics, pp. 1–15, 2014.

Gao, B.-C.: NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space, Remote sensing of environment, 58, 257–266, 1996.

Gao, F., Masek, J., and Wolfe, R. E.: Automated registration and orthorectification package for Landsat and Landsat-like data processing, Journal of Applied Remote Sensing, 3, 033 515–033 515, 2009.

Gardelle, J., Arnaud, Y., and Berthier, E.: Contrasted evolution of glacial lakes along the Hindu Kush Himalaya mountain range between 1990 and 2009, Global and Planetary Change, 75, 47–55, 2011.

Gardelle, J., Berthier, E., and Arnaud, Y.: Slight mass gain of Karakoram glaciers in the early twenty-first century, Nature geoscience, 5, 322–325, 2012.

Gardelle, J., Berthier, E., Arnaud, Y., and Kääb, A.: Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999-2011, Cryosphere, 7, 2013.

Guo, W., Liu, J. X. S., Shangguan, D., Wu, L., Yao, X., Zhao, J., Liu, Q., Jiang, Z., Li, P., Wei, J., Bao, W., Yu, P., Ding, L., Li, G., Ge, C., and Wang, Y.: The Second Glacier Inventory Dataset of China (Version 1.0), Cold and Arid Regions Science Data Center at Lanzhou, 2014.

Hall, D., Ormsby, J., Bindschadler, R., and Siddalingaiah, H.: Characterization of snow and ice reflectance zones on glaciers using Landsat Thematic Mapper data, Ann. Glaciol, 9, 1–5, 1987.

Hansen, M. C. and Loveland, T. R.: A review of large area monitoring of land cover change using Landsat data, Remote sensing of Environment, 122, 66–74, 2012.

Hanshaw, M. N. and Bookhagen, B.: Glacial areas, lake areas, and snowlines from 1975 to 2012: status of the Cordillera Vilcanota, including the Quelccaya Ice Cap, northern central Andes, Peru, Cryosphere Discussions, 7, 2014.

Heid, T. and Kääb, A.: Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery, Remote Sensing of Environment, 118, 339–355, 2012.

Huggel, C., Kääb, A., Haeberli, W., Teysseire, P., and Paul, F.: Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps, Canadian Geotechnical Journal, 39, 316–330, 2002.

Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E., et al.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org), 2008.

Kääb, A.: Monitoring high-mountain terrain deformation from repeated air-and spaceborne optical data: examples using digital aerial imagery and ASTER data, ISPRS Journal of Photogrammetry and remote sensing, 57, 39–52, 2002.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., and Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, Nature, 488, 495–498, 2012.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, EOS, Transactions American Geophysical Union, 89, 93–94, 2008.

Leprince, S., Ayoub, F., Klingert, Y., and Avouac, J.-P.: Co-registration of optically sensed images and correlation (COSI-Corr): An operational methodology for ground deformation measurements, in: Geoscience and

Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, pp. 1943–1946, IEEE, 2007.
Lioubimtseva, E. and Henebry, G. M.: Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations, Journal of Arid Environments, 73, 963–977, 2009.
Mushkin, A. and Gillespie, A.: Using ASTER Stereo Images to Quantify Surface Roughness, in: Land Remote Sensing and Global Environmental Change, pp. 463–481, Springer, 2011.
Mushkin, A., Gillespie, A., Danilina, I., O'Neal, M., Pietro, L., Abbott, E., and Balick, L.: Using sub-pixel roughness estimates from ASTER stereo images to compensate for roughness effects in the thermal infrared, in: RAQRS II: 2nd International Symposium on Recent Advances in Quantitative Remote Sensing, 2006.
Narama, C., Kääb, A., Duishonakunov, M., and Abdrakhmatov, K.: Spatial variability of recent glacier area changes in the Tien Shan Mountains, Central Asia, using Corona (˜ 1970), Landsat (˜ 2000), and ALOS (˜ 2007) satellite data, Global and Planetary Change, 71, 42–54, 2010.
Nie, Y., Liu, Q., and Liu, S.: Glacial Lake Expansion in the Central Himalayas by Landsat Images, 1990–2010, PloS one, 8, e83 973, 2013.
Nuimura, T., Sakai, A., Taniguchi, K., Nagai, H., Lamsal, D., Tsutaki, S., Kozawa, A., Hoshina, Y., Takenaka, S., Omiya, S., et al.: The GAMDAM Glacier Inventory: a quality controlled inventory of Asian glaciers, The Cryosphere Discussions, 8, 2799–2829, 2014.
Oerlemans, J.: Extracting a climate signal from 169 glacier records, Science, 308, 675–677, 2005.
Paul, F.: Changes in glacier area in Tyrol, Austria, between 1969 and 1992 derived from Landsat 5 Thematic Mapper and Austrian Glacier Inventory data, International Journal of Remote Sensing, 23, 787–799, 2002.
Paul, F., Kaab, A., Maisch, M., Kellenberger, T., and Haeberli, W.: The new remote-sensing-derived Swiss glacier inventory: I. Methods, Annals of Glaciology, 34, 355–361, 2002.
Paul, F., Huggel, C., and Kääb, A.: Combining satellite multispectral image data and a digital elevation model for mapping debris-covered glaciers, Remote Sensing of Environment, 89, 510–518, 2004.
Paul, F., Barrand, N., Baumann, S., Berthier, E., Bolch, T., Casey, K., Frey, H., Joshi, S., Konovalov, V., Bris, R. L., et al.: On the accuracy of glacier outlines derived from remote-sensing data, Annals of Glaciology, 54, 171–182, 2013.
Pfeffer, W. T., Arendt, A., Bliss, A., Bolch, T., Cogley, J., Gardner, A., Hagen, J., Hock, R., Kaser, G., Kienholz, C., Miles, E., Moholdt, G., Mölg, Paul, F., Radić, V., Rastner, P., Raup, B., Rich, J., Sharp, M., and Consortium, T. R.: The Randolph Glacier Inventory: a globally complete inventory of glaciers, Journal of Glaciology, 60, 537–552, 2014.
Quincey, D., Richardson, S., Luckman, A., Lucas, R., Reynolds, J., Hambrey, M., and Glasser, N.: Early recognition of glacial lake hazards in the Himalaya using remote sensing datasets, Global and Planetary Change, 56, 137–152, 2007.
Racoviteanu, A. and Williams, M. W.: Decision tree and texture analysis for mapping debris-covered glaciers in the Kangchenjunga area, Eastern Himalaya, Remote Sensing, 4, 3078–3109, 2012.
Racoviteanu, A. E., Arnaud, Y., Williams, M. W., and Ordonez, J.: Decadal changes in glacier parameters in the Cordillera Blanca, Peru, derived from remote sensing, Journal of Glaciology, 54, 499–510, 2008a.
Racoviteanu, A. E., Williams, M. W., and Barry, R. G.: Optical remote sensing of glacier characteristics: a review with focus on the Himalaya, Sensors, 8, 3355–3383, 2008b.
Racoviteanu, A. E., Paul, F., Raup, B., Khalsa, S. J. S., and Armstrong, R.: Challenges and recommendations in mapping of glacier parameters from space: results of the 2008 Global Land Ice Measurements from Space (GLIMS) workshop, Boulder, Colorado, USA, Annals of Glaciology, 50, 53–69, 2009.
Rastner, P., Bolch, T., Notarnicola, C., and Paul, F.: A comparison of pixel-and object-based glacier classification with optical satellite images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 853–862, 2014.
Raup, B., Kääb, A., Kargel, J. S., Bishop, M. P., Hamilton, G., Lee, E., Paul, F., Rau, F., Soltesz, D., Khalsa, S. J. S., et al.: Remote sensing and GIS technology in the Global Land Ice Measurements from Space (GLIMS) project, Computers & Geosciences, 33, 104–125, 2007.
Raup, B. H., Khalsa, S. J. S., Armstrong, R. L., Sneed, W. A., Hamilton, G. S., Paul, F., Cawkwell, F., Beedle, M. J., Menounos, B. P., Wheate, R. D., et al.: Quality in the GLIMS Glacier Database, in: Global Land Ice Measurements from Space, pp. 163–182, Springer, 2014.
Scherler, D., Bookhagen, B., and Strecker, M. R.: Spatially variable response of Himalayan glaciers to climate change affected by debris cover, Nature Geoscience, 4, 156–159, 2011a.
Scherler, D., Bookhagen, B., and Strecker, M. R.: Hillslope-glacier coupling: The interplay of topography and glacial dynamics in High Asia, Journal of Geophysical Research: Earth Surface, 116, 2011b.
Shangguan, D., Bolch, T., Ding, Y., Kröhnert, M., Pieczonka, T., Wetzel, H., and Liu, S.: Mass changes of

Southern and Northern Inylchek Glacier, Central Tian Shan, Kyrgyzstan, during 1975 and 2007 derived from remote sensing data, The Cryosphere, 9, 703–717, 2015.

Shukla, A., Arora, M., and Gupta, R.: Synergistic approach for mapping debris-covered glaciers using optical–thermal remote sensing data with inputs from geomorphometric parameters, Remote Sensing of Environment, 114, 1378–1387, 2010.

Sorg, A., Bolch, T., Stoffel, M., Solomina, O., and Beniston, M.: Climate change impacts on glaciers and runoff in Tien Shan (Central Asia), Nature Climate Change, 2, 725–731, 2012.

Stocker, D. Q.: Climate change 2013: The physical science basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Summary for Policymakers, IPCC, 2013.

Taschner, S. and Ranzi, R.: Comparing the opportunities of Landsat-TM and Aster data for monitoring a debris covered glacier in the Italian Alps within the GLIMS project, in: Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International, vol. 2, pp. 1044–1046, IEEE, 2002.

Worni, R., Huggel, C., and Stoffel, M.: Glacial lakes in the Indian Himalayas - From an area-wide glacial lake inventory to on-site and modeling based risk assessment of critical glacial lakes, Science of the Total Environment, 468, S71–S84, 2013.

Table 1. Data table listing Landsat capture dates used in this study. Organized by WRS2 Path/Row combinations. Starred dates denote 'Master' images to which others were rectified. Bold dates indicate ~~use~~ images used for ~~Velocity~~ velocity profiles.

	144/030	145/030	147/031
Number of Images	11	10	12
Date Range of Images	2002-2013	1998-2013	2000-2013
LT5 Capture Dates	Jul 31, 2006	Sep 2, 1998*	Aug 19, 2011
	Aug 8, 2009	Oct 4, 1998	Oct 2, 1998
	Sep 27, 1998	Jul 22, 2006	Sep 6, 2006
	Jul 13, 2011	**Aug 10, 2007**	Aug 24, 2007
		Sep 11, 2007	Oct 3, 2010
		Oct 2, 2009	Aug 3, 2011
		Aug 2, 2010	
		Jul 4, 2011	
		Sep 6, 2011	
LE7 Capture Dates	**Sep 14, 2002***		Sep 14, 2000*
	Jul 7, 2000		**Oct 5, 2002**
	Aug 8, 2000		Aug 18, 2002
	Jun 7, 2001		
LC8 Capture Dates	**Oct 22, 2013**	**Sep 27, 2013**	**Sep 25, 2013**
	Aug 19, 2013		Sep 9, 2013
	Sep 4, 2013		May 7, 2014
Projection	WGS 1984 45N		WGS 1984 44N
Comments	Eastern Edge of Study Area		Vicinity of Inylchek Glacier

	148/031	149/031	150/032
Number of Images	13	3	5
Date Range of Images	2002-2013	1999-2013	1998-2013
LT5 Capture Dates	Sep 16, 2007	Sep 7, 2007	Oct 23, 1998
	Sep 11, 2011		Jul 1, 2009
	Aug 22, 1998		
	Aug 12, 2006		
	Sep 13, 2006		
	Jul 30, 2007		
LE7 Capture Dates	**Jul 24, 2002***	**Sep 9, 1999***	**Aug 20, 2001***
	Jul 16, 1999		Sep 24, 2002
	Sep 18, 1999		
	Aug 25, 2002		
LC8 Capture Dates	**Jul 30, 2013**	**Oct 9, 2013**	**Jun 10, 2013**
	Oct 2, 2013		
	May 14, 2014		
Projection	WGS 1984 44N	WGS 1984 43N	WGS 1984 43N
Comments			

	151/033	153/033
Number of Images	5	3
Date Range of Images	1998-2013	1998-2013
LT5 Capture Dates	Sep 28, 1998	Sep 26, 1998
	Sep 10, 2009	
LE7 Capture Dates	Aug 24, 2000*	**Sep 29, 2002***
	Sep 28, 2001	
LC8 Capture Dates	**Oct 7, 2013**	**Oct 5, 2013**
Projection	WGS 1984 43N	WGS 1984 42N
Comments		Towards Pamir Knot

Table 2. Comparison of methods between previous debris-covered glacier mapping studies.

Method	Short Description	Data Inputs	Processing Intensive Steps	Area Covered in Study	Reported Accuracy
Taschner and Ranzi (2002)	Clean-ice detection using Landsat, coupled with ASTER thermal data	Landsat, ASTER	Data resampling, pixel clustering	5.58 km^2, Italian Alps	Not Reported
Paul et al. (2004)	Clean-ice detection using Landsat, coupled topographic analysis and neighborhood analysis	Landsat, ASTER-DEM	Image Polygon Growing neighborhood analysis	23 km^2, Swiss Alps	21% of debris misclassified
Bolch et al. (2007)	A set of training areas based on spectral and topographic information is used to determine classification thresholds	ASTER, ASTER-DEM	Creation and tuning of training dataset	Not reported, Mt. Everest Region	5% total area misclassified
Shukla et al. (2010)	Multiple landcover types mapped using spectral and thermal imagery combined with a DEM	ASTER, AWiFS, DEM	Data conversion and registration, solar illumination analysis, training dataset creation, Maximum Likelihood Classifier	200 km^2, Samudra Tapu glacier, Himachal Pradesh, India	8-14% debris misclassified
Racoviteanu and Williams (2012)	Decision tree classification with ASTER and topographic data, and (2) texture analysis exploiting surface roughness	ASTER, DEM, Quickbird, Worldview2	Training dataset creation, decision tree set-up, principal component analysis	576.4 km^2, Sikkim Himalaya, NE India	(1) 25%, (2) 31% debris misclassified
Rastner et al. (2014)	Comparison of object- and pixel-based methods of glacier mapping. Both methods use spectral and topographic information as inputs	ASTER, Landsat, DEM	Manual threshold definitions, segmentation processing, iterative thresholding	Not reported, three distinct test regions	11.5% (object-based) and 23.4% (pixel-based) misclassified areas for Himalaya region
This Study	Clean-ice detection coupled with topographic, velocity, and distance weighting thresholds	Landsat, SRTM DEM, River Network	Velocity field calculation, optional debris seed point selection	~44,000 km^2, Pamir-Tien Shan	2-10% total area misclassified

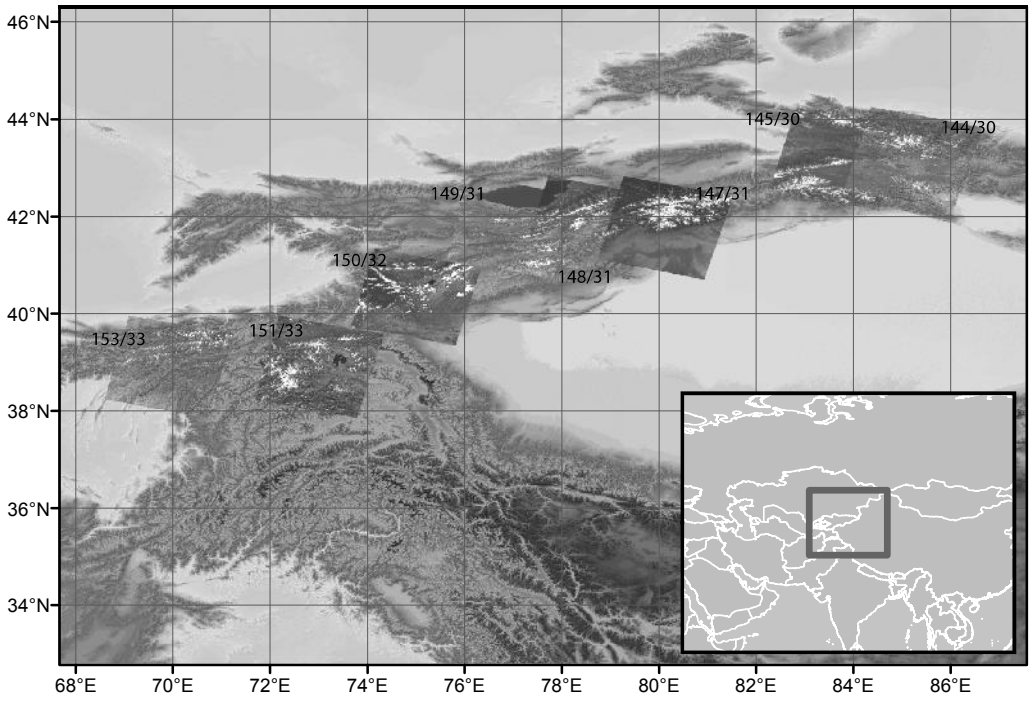

Fig. 1. Greater study area of the Tien Shan, showing SRTM v4.1 topography (Jarvis et al., 2008) and ~~locations~~ location of eight Landsat ~~scenes outlines~~ image footprints (grayscale) used in the study, ~~as well as~~ along with their Path/Row combinations.

Fig. 2. Characteristic example of a debris-covered glacier tongue (Inylchek Glacier). Spectrally-delineated glacier outlines (black), over Landsat bands B7/B5/B3 ~~loaded as Red~~(R/~~Green~~G/~~Blue~~B), from image LC81470312013268LGN00. ~~Showing~~ This shows generally well-mapped clean ice, but poor treatment of debris-covered tongues.

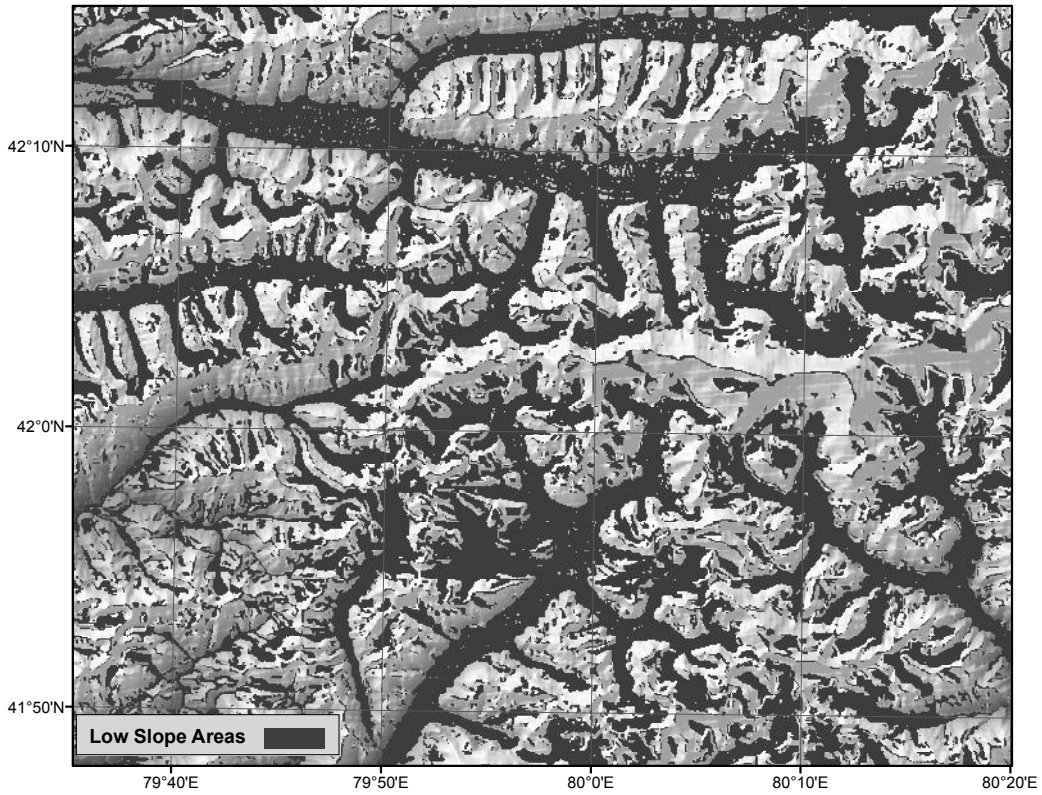

Fig. 3. ~~Black lines~~ Blue areas show 'Potential Debris Areas', as delineated by slopes between 1-24 degrees, ~~and slope variability (via 3x3 standard deviation filter) above 2,~~ with elevations below 2500m removed. Only these areas are examined in the subsequent thresholding steps, to reduce processing time and misclassification errors. SRTM hillshade underneath.

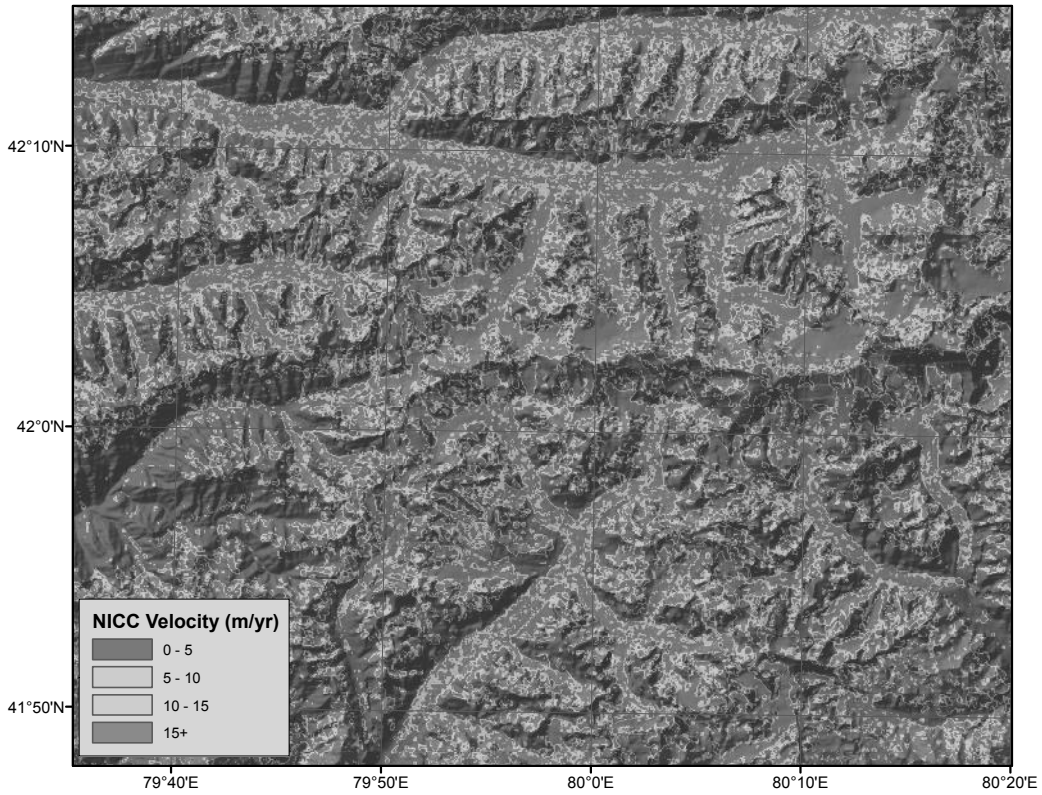

Fig. 4. Example ~~binary~~ of an annual glacial velocity ~~mask~~surface, generated using Normalized Image Cross Correlation (NICC). Areas in red are slow-moving areas and represent stable ground. They are removed from the 'potential debris areas'~~, because they do not show strong surface movements. Debris-tongues and clean-ice areas visible as B7/5/3 image LC81470312013268LGN00~~.'

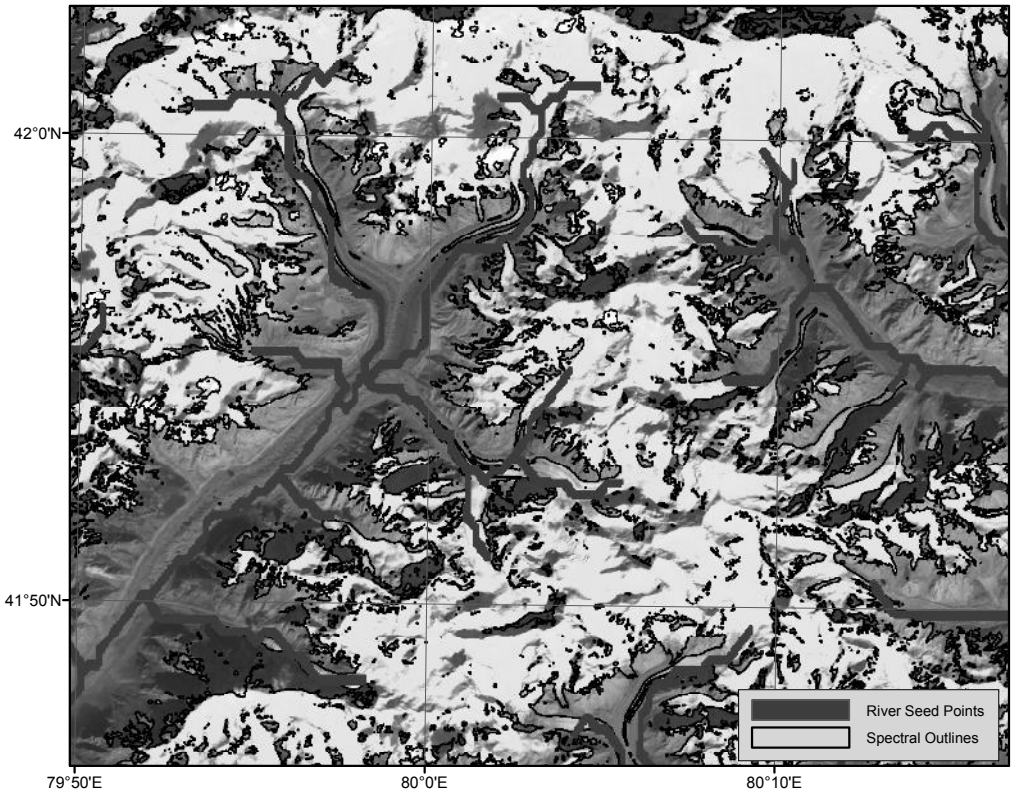

Fig. 5. Example of ~~distance weighting~~ distance-weighting seed ~~points~~ areas used to remove pixels from the 'potential debris areas' which are distant from either a river valley or classified glacier ice. Rivers in blue, spectrally-delineated glaciers are outlines in black. ~~Illustrates~~ The blue lines illustrate the presence of river networks along debris-covered tongues of glaciers where there is little clean glacier ice.

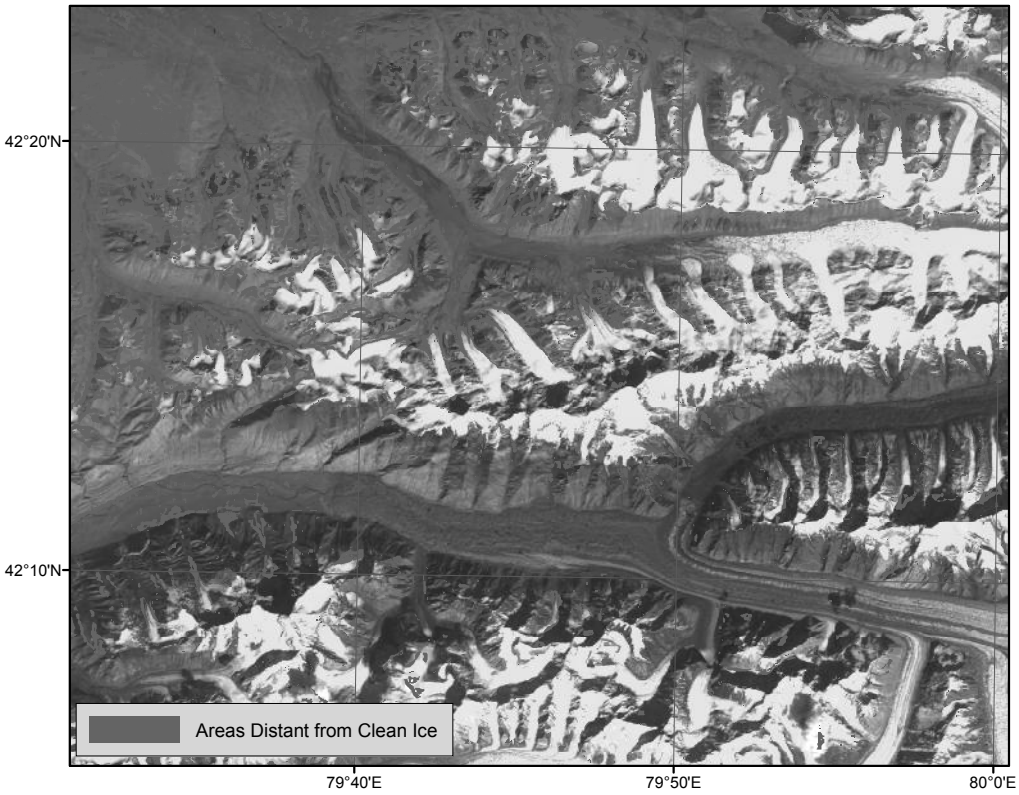

Fig. 6. ~~Example~~ Areas removed by second distance weighting ~~classification to remove areas distant from valleys and glacier areas. Results of the velocity filtering in black underneath the results of the distance filtering in blue, to illustrate the difference in classification after the distance threshold~~ step (red). The geodesic distance ~~weighting primarily~~ algorithm removes ~~pixels distant from glacier~~ isolated areas ~~, such~~ near glaciers as well as ~~the large~~ areas ~~of black in the top left~~distant from any glacier.

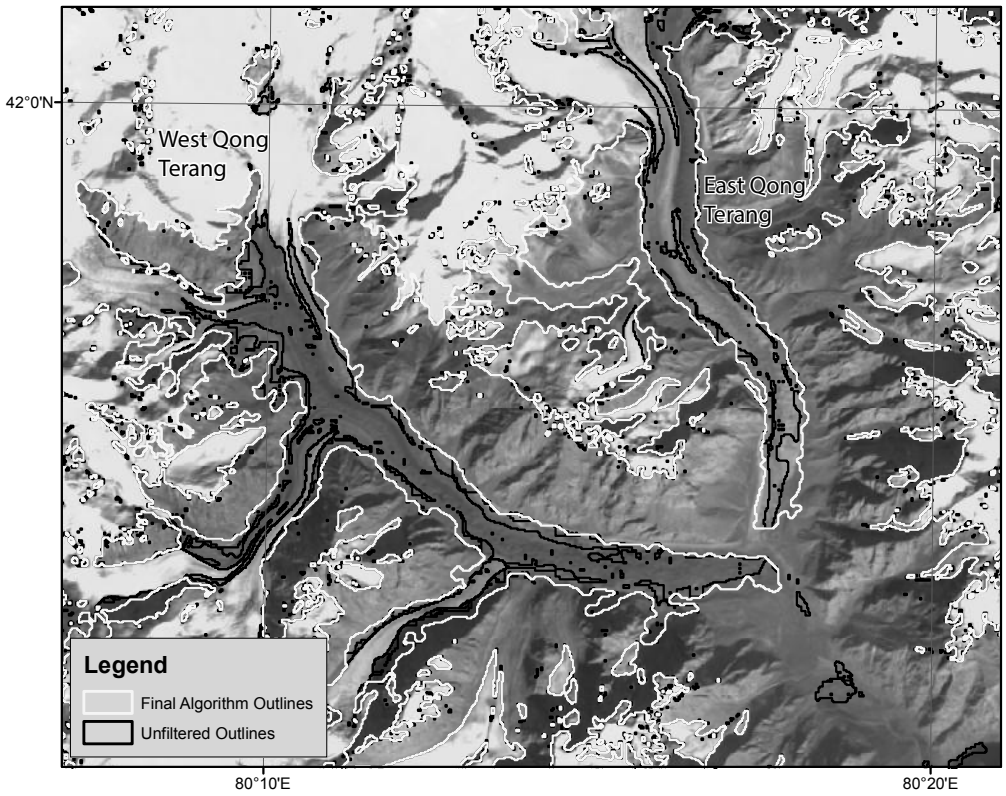

Fig. 7. Impacts of statistical filtering on glacier outlines, with areas in black removed during the filtering process. Primarily small holes in large debris-tongues are removed, while the glacier outlines remain intact during this step. ~~This can be seen on both the Inylchek~~ East and ~~Tomur glaciers in this figure~~West Qong Terang Glaciers.

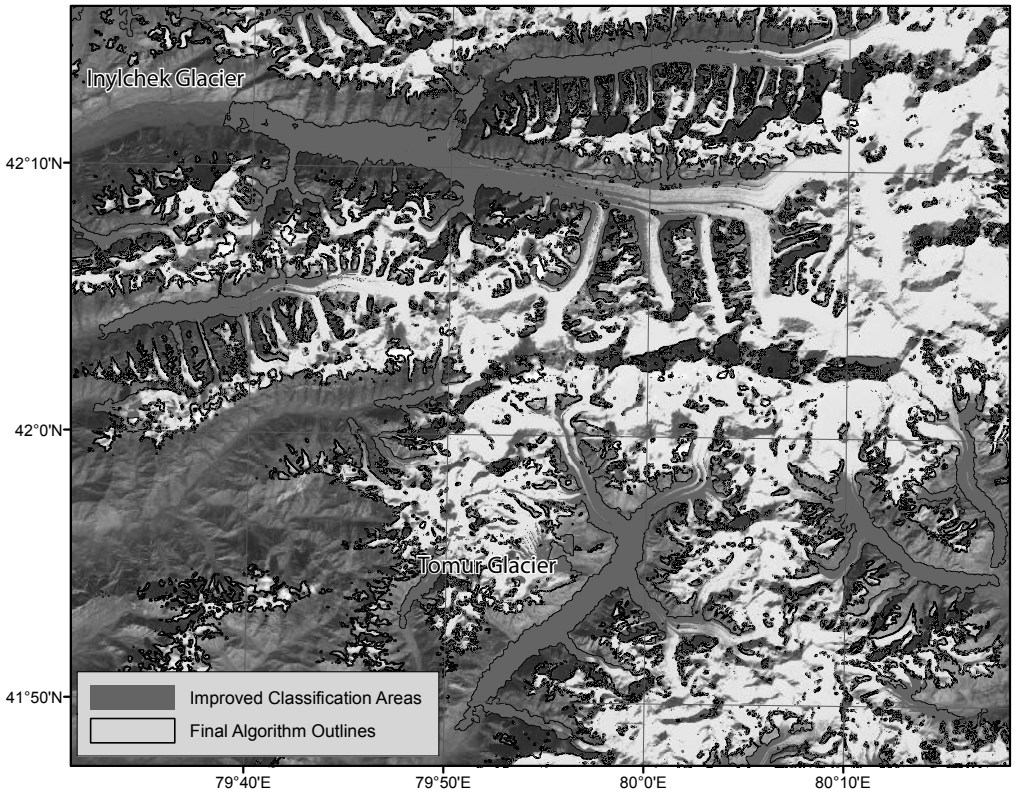

Fig. 8. Final algorithm ~~results~~ outlines (~~blue~~black) ~~,~~ with ~~spectral outlines (~~areas classified after the clean-ice delineation in red~~)~~. Illustrates the improved classification by the algorithm across several large debris tongues. ~~Vicinity of the Tomur glacier,~~ Landsat OLI image captured Sept 25, 2013 as background.

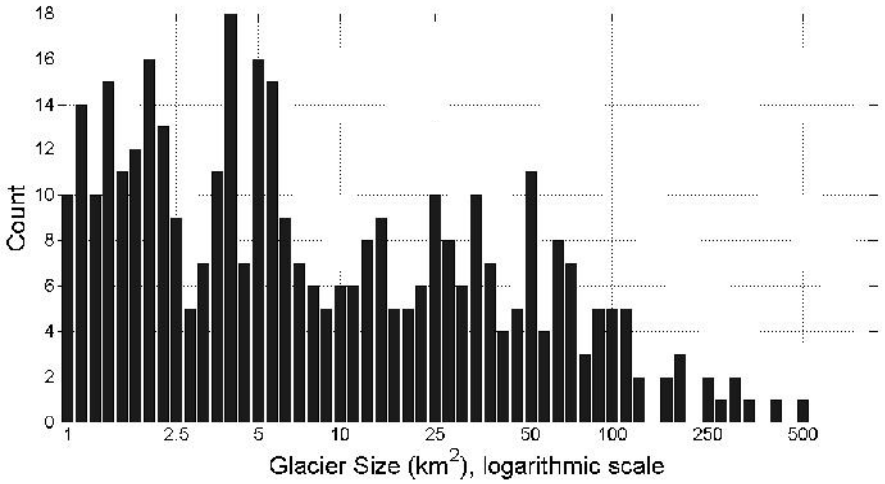

Fig. 9. Glacier size class distribution (n=750) for the manual control dataset. Note the logarithmic x-axis to account for a wide range of glacial sizes.

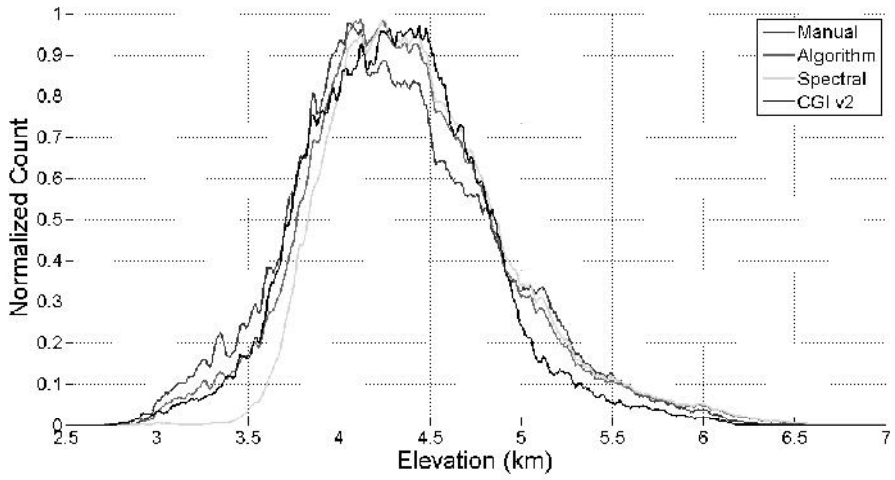

Fig. 10. Bulk elevation distributions of sampled glaciers, with manual delineation (reference dataset, n=215, 4,500 km^2) in blue, algorithm-derived delineation in red, ~~and~~ spectral delineation in green~~. Illustrates high fidelity algorithm outlines in low-elevation areas~~, and ~~generally tight agreement between spectral-delineated and algorithm outlines at elevations above 4000m~~CGI v2 in black. Values have been normalized to maximum probability.

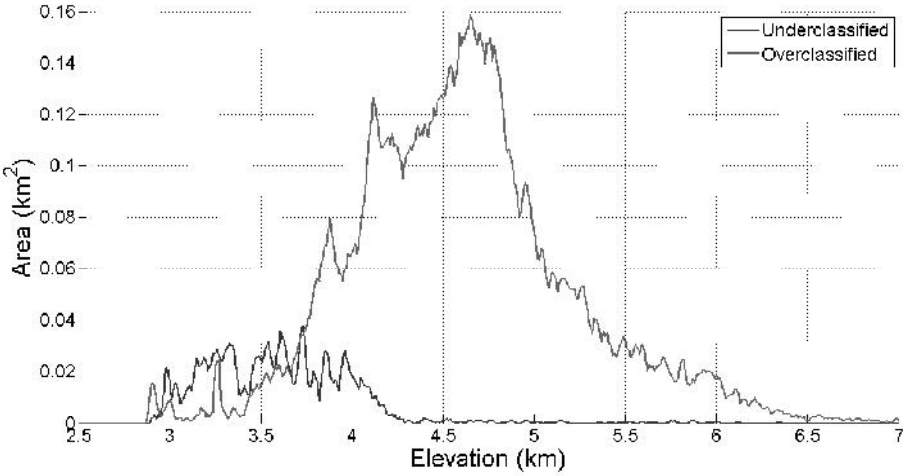

Fig. 11. Elevation distributions of over- and under-classified glacier areas, as compared to a manual control dataset (n=~~21~~575, 330 km^2). Overclassified areas show that the algorithm ~~is weighted towards overclassifying high elevation areas, and~~ does not remove large portions of the accumulation area, but instead adds additional area as compared to the control dataset. Underclassified areas indicate that the algorithm ~~does not perfectly classify~~ identifies less area than the manually-digitized dataset in low-elevation ~~areas~~regions. 5.5% is overclassified, ~~but that it misclassifies a relatively small area percentage~~and 0.8% is underclassified.

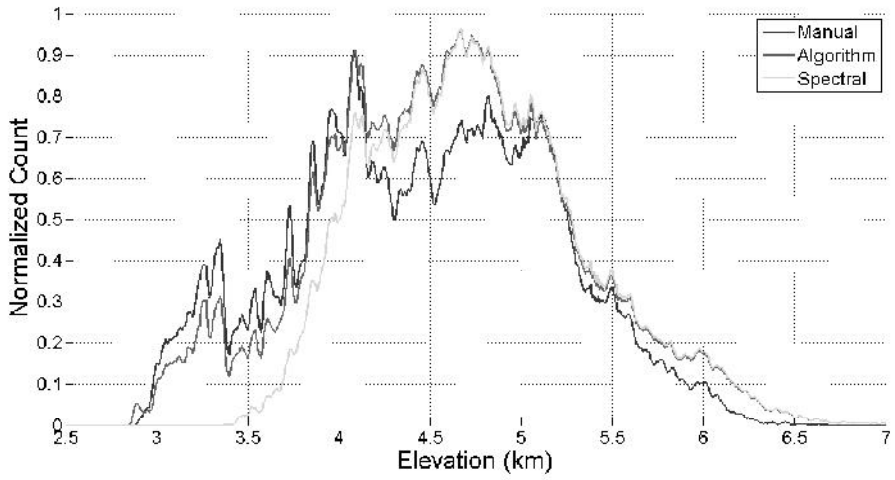

Fig. 12. Averaged elevation differences for a random sample of glaciers overlapping a manual control dataset (n=100, 100 km^2). Shows generally close agreement between the manual glacier dataset and the algorithm dataset below 4000m, with closer agreement between the spectral and algorithm datasets above 4000m. This indicates improved mapping of debris-tongues by the algorithm, and similar treatment of clean ice by both the algorithm and the spectrally-delineated glaciers.

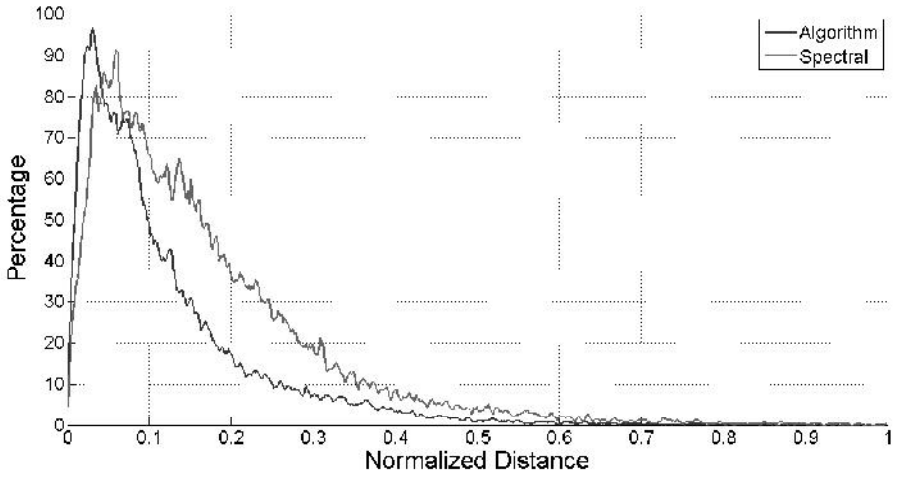

Fig. 13. ~~Normalized vertex~~ Vertex distance distributions for algorithm (blue) and spectral (red) vertices, as compared to a manual control dataset, normalized to the maximum distance. This indicates generally ~~tighter~~ closer agreement between the algorithm and manual datasets than between the spectral and manual datasets.

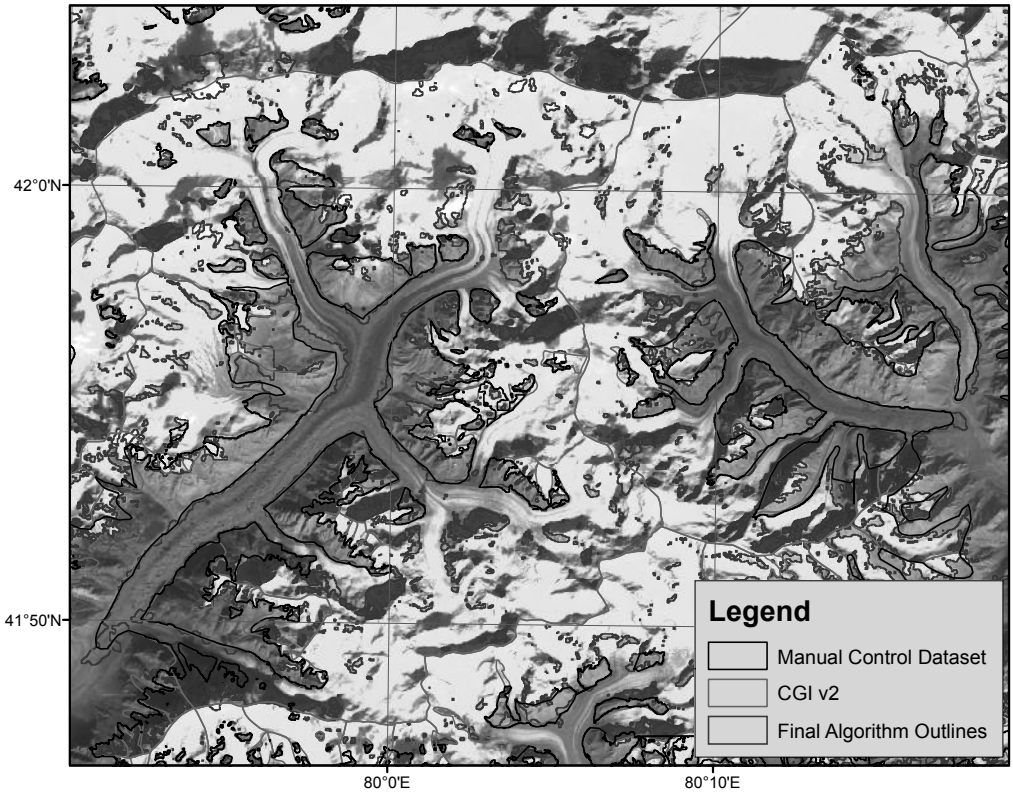

Fig. 14. Algorithm outlines (~~orange~~purple) compared to the control dataset (black) and the ~~RGI v4.0~~ CGI v2 (red). Illustrates high fidelity in overall debris-tongue length between the three datasets, although the algorithm outlines exhibit noise along the edges of debris tongues.

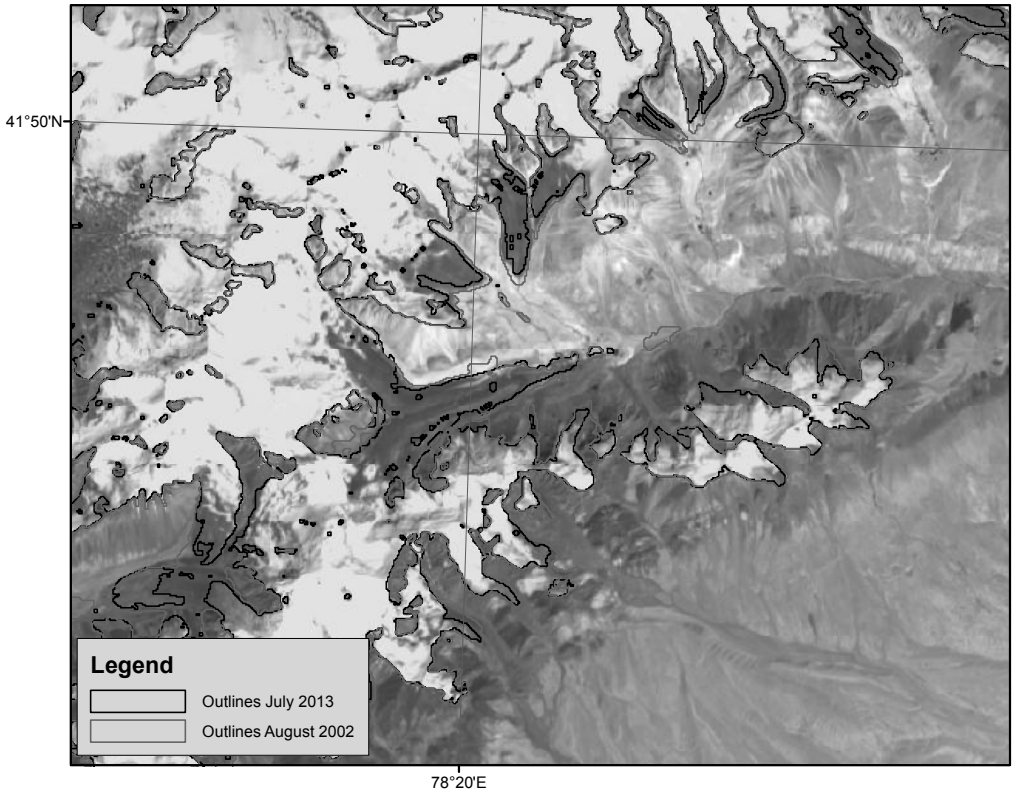

Fig. 15. Algorithm outlines for ~~September~~ July 2013 (black) and algorithm outlines for ~~October~~ August 2002 (red), showing small retreats in glacier areas, particularly at the debris tongues. ~~Also shows an example~~ Vicinity of ~~commonly misclassified river sand~~the Akshiirak glacierized massif, central Tien Shan.