Reply to Reviewer 2 Part B

Review of ‘Debris cover and the thinning of Kennicott Glacier, Alaska, Part B’ by Leif Anderson et al., under consideration for The Cryosphere

Thank you kindly for taking the time to review our manuscripts!

Part B of the Anderson et al trilogy aims to combine empirical relationships of surface properties and melt rates, based on the field measurements presented in Part A, with remote sensing observations of different surface features (particularly ice cliffs) in order to arrive at distributed estimates of melt rates for the period of observations. The analysis then uses these distributed melt values to address whether or not the identified melt hotspots can explain the thinning patterns evident at Kennicott. This is an important question as the glaciological community is trying to disentangle the influence of surface mass balance and ice dynamics for debris covered glacier evolution, and this is the first time the question has been addressed in Alaska, where surface debris is prevalent.

As such, this represents an important contribution to a current topic of research, and provides an answer to that question for Kennicott Glacier – these hot spots do have an important effect on the surface mass balance, but it is not plausible that they compensate for the overall melt reduction due to surface debris. The study could be more systematic to provide a definitive answer, and I have some criticisms regarding the empirical relationships presented in determining the distributed estimates of melt, as well as the difference in temporal scales between long-term elevation change (52 years) relative to single-year field measurements. However, although I have quite a few comments, no changes along these lines are likely to change the conclusions of the study. Rather, my principal concern is that the separation of this analysis from both Parts A and C reduces the strength and presentation of the entire analysis, while also leading to repetition of text and figures, as well as cross references. There are certainly gaps in the analysis because aspects have been included in Parts A or C rather than here, and some restructuring across the three manuscripts might improve the readability of all three. This is a choice for the authors and Editor to contemplate, but my opinion is that some consolidation would be beneficial.

Thank you kindly for taking the time to review these manuscripts!

Main comments:

My principal question in reading this manuscript was whether it should be standalone or integrated with Part A (and possibly C, which I have not reviewed). I appreciate that this is a difficult decision, and that Parts A, B, and C combined represent a substantial body of work.

Thank you for your thoughts on this we reply to all comments regarding structure in the main response.

There are certainly some advantages to be considered for maintaining separation between the manuscripts, but my impression is that A and B are both weaker manuscripts separated from one another. The field measurements were clearly collected for the purpose of deriving distributed melt rate estimates, which leaves Part A without a compelling conclusion or discussion.

It is worth noting that while it may seem that these are weaker manuscripts apart the question in our mind is: Do all three work together and complement one another.

We have version of the manuscript where we did combine Parts A and B and we simply found that there was too much in the manuscript, too many directions, too many diverse methods discussed.

A and B were presented in 2 parts so we could explore the in situ data alone. If we combine the manuscripts we will lose the analysis of the in situ data as well as the presentation of many measurements relevant for melt modelers. Additionally combing A and B means that the ice cliff detection method becomes a footnote instead of a primary contribution. There are other manuscripts that just present an ice cliff detection method.

We agree that Part A can be improved and we outline how that can happen in the proposed changes. We have more data we can add to Part A such that is takes more of a unique perspective but we excluded that data in favor of streamlining Parts A, B, and C. Please see our primary response about the structure of the manuscripts.

At the same time, Part B requires frequent reference to Part A, or blind faith of the part of the reader with regards to the methods and results of the field data. As a consequence, there is also quite a bit of repeated material to cover (for example, content from 4 figures in Part A is also displayed in Part B).

We appreciate this comment but we feel this issue is overemphasized. The reviewer finds that 1) you need to read Part A before understanding Part B and 2) that information is repeated between the two manuscripts.

Both of these issues would occur if Part A was accepted first and then Part B was submitted independently later. From Part A we pass on three datasets, and numerous observations to support the inferences made in Part B.

(for example, content from 4 figures in Part A is also displayed in Part B).

Content from A is in only in 4 panels in 2 figures (alaska context figure and curve fits to debris, sub-debris melt rate, and ice cliff backwasting). Which is perfectly acceptable going from one paper to the next. We can also leave out the data figure that is repeated but we feel that would require the reader to look back so we include what is really needed to understand Part B.

My instinct when reviewing Part A was that much of that material could be more meaningful if integrated with Part B, in supplementary information if not in the main text, and I now think that would greatly improve the readability of this manuscript.

We appreciate this view but when the manuscript was previously combined it was bloated and difficult to outline all the pieces needed for Part B with the detail needed.

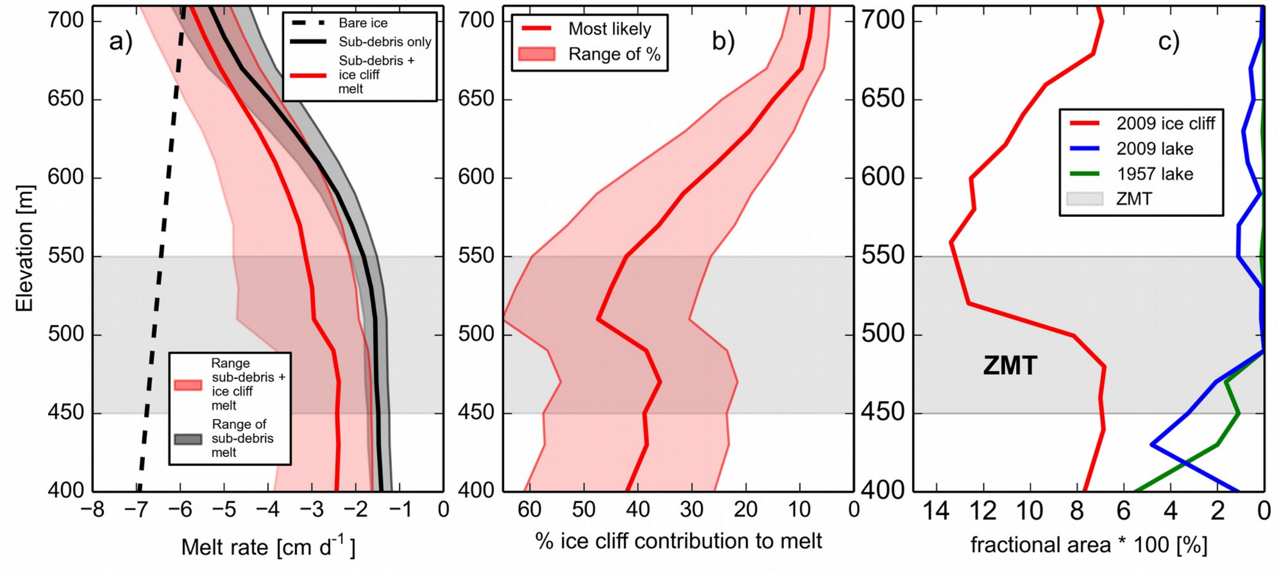
I like the ice cliff delineation method in its simplicity (although some details need to be clarified, below), but its transferability is not very clear. Often when developing/proposing a new method it is necessary to see how robust the method is, but in this case the method has clearly been developed specifically to map cliffs in this particular scene, in order to apply the empirical melt relationship. As such, it is a relatively small part of the story in deriving the distributed melt estimates.

Thank you for highlighting this. We present the new ice cliff detection method as a proof of concept for a rather tough case. Kennicott Glacier has more dense ice cliffs than any other debris-covered glacier that we know of. Slope threshold approaches do not appear to be as effective on this glacier.

We could include a more thorough comparison with other methods we feel that should be the focus of another manuscript.

At the same time, the maps of supraglacial ponds are not integrated very well into the story, while the exclusion of supraglacial streams (also mapped from satellite imagery) is a bit

Thank you for pointing this out. We actually feel that simple fact that ponds do not align with the zone of maximum thinning shows that they cannot explain the pattern. Please see Figure 10c.



It is also important to note that the red range in 10a and 10b are the extreme melt estimates.

We state this in the manuscript. It is not clear how lakes can control the zone of maximum thinning if they are not present there? We show that lakes are not present across this zone of maximum thinning from 1957 and 2009.

In Part A we also highlight that ice cliffs with ponds at their base do not backwaste at rates higher than ice cliffs with streams or ponds. We will emphasize that more here.

We can include a discussion of surface streams and do a back-of-the-envelope calculation for the amount of strain heat produced by the streams. We would be very surprised if streams would contribute enough melt to compensate for the melt reducing effect of debris, especially considering that ice cliffs would need to cover 90% of the debris-covered tongue to compensate for the effects of surface debris.

The study aims to address the role of melt hotspots in explaining thinning rates. For this to be a definitive analysis in this regard, I feel like this needs to be done in a more systematic manner, whereas the present analysis seems to focus exclusively on ice cliffs.

Thank you for raising this point. The important point that is emphasized in the manuscript is that this glacier has the highest concentration of ice cliffs of any that we know of, and is thus a good place to test if ice cliffs, especially can compensate for the melt reducing effects of debris.

The distribution of supraglacial streams seems to be a major gap in the analysis of this part of the study, especially as the properties of supraglacial streams are assessed in Part C. Streams are mentioned throughout the background as hot spots and possible factors contributing to melt, but are then completely neglected in the methods and discussion.

We think streams do have an important effect (and we discuss them at length in Part C). We are happy to discuss streams here in Part B. But we do address them in Part C where we do digitize many streams.

It is important to note that there are very few studies that actually address the effect of streams on debris-cover glacier mass balance and to date none have really quantified an effect. Here is where having space to describe our methods in Parts A and C is an advantage. We mention that cliffs with streams do not backwaste faster than cliffs without streams based on our measured 60 ice cliffs. This implies that they have a secondary effect on mass balance. We aren't aware that this has been noted before and we feel this is support for keeping Part A in its current form so we have space to point out these important observations based on the in situ data.

We can do the calculation here in this manuscript for the amount of strain heat produced by an extreme concentration of streams on Kennicott Glacier. We hypothesize now that while streams locally enhance melt they are largely inconsequential melt contributors when averaged across the glacier. See Section 4.2.1 for support of this argument here.

While streams appear to help maintain ice cliffs (see Part C) we are skeptical that they will be a major contributor to debris-covered tongue wide melt rates. We will address this in the revisions. If adding them to Part B helps in the continuity of the 3 manuscripts we are happy to add them.

I understand that the role of surface streams cannot be assessed quantitiavely in this manuscript, but neither can the role of ponds. Similarly, although internal ablation is usually regarded as negligible in this type of specific mass balance assessment, there have been suggestions that this is a non-negligible term for extensively debris-covered glaciers. It is exceedingly unlikely that this mechanism could lead to the debris-cover anomaly, but for completeness I think it should be considered numerically along with the cliffs and ponds.

We will explore back of the envelope estimates for englacial mass loss due to strain heating of water along with the surface streams. They could be interesting additions to the analysis.

The relationships between elevation and debris thickness, and between elevation and ice cliff backwasting, are very weak. In neither case does elevation appear to be a primary control of the property.

We are a bit confused about the elevation versus debris thickness statement because lower in this review you state: "In Part A it is clear that while elevation is a principal driver of debris thickness variability, there is considerable heterogeneity within any elevation bin." As we suggested in Part A it would be a nice idea to extrapolate down flowlines, but we felt like this was a methological development to focus on in later studies.

Despite the scatter we agree on, Figure 5 Part A (see below) shows that debris thickness does increase down glacier. We can argue about whether or not that is linear or not but the box-plots show the increase. Here is a reason why Part A standing alone is good for presenting the body of work. We are able to show the box plots that really reveal the thickening of debris down glacier.

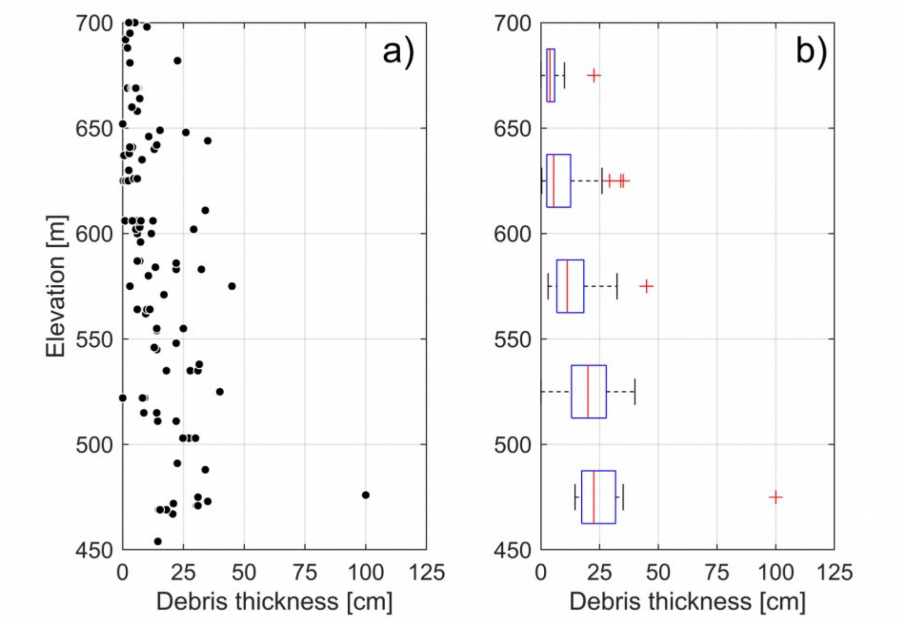


Figure 5. Pattern of debris thickness with elevation. a) In situ debris thickness measurements. b) Debris thickness boxplots in 50 meter elevation bins. Outliers are represented as +'s.

How different would your results be if you interpolated spatially (within flowlines or lithological units for example)? Did you consider alternative methods to provide a distributed debris thickness estimate?

We did consider different approaches but we felt that the number of ice cliffs and variable surface topography was a strong hinderace to using thermal infrared imagery. A full consideration of this would require an additional study.

For example, one could consider the radiance measured by satellite thermal infrared imagery as related to debris thickness, and use your measurements to constrain this relationship locally to upscale in space much more meaningfully.

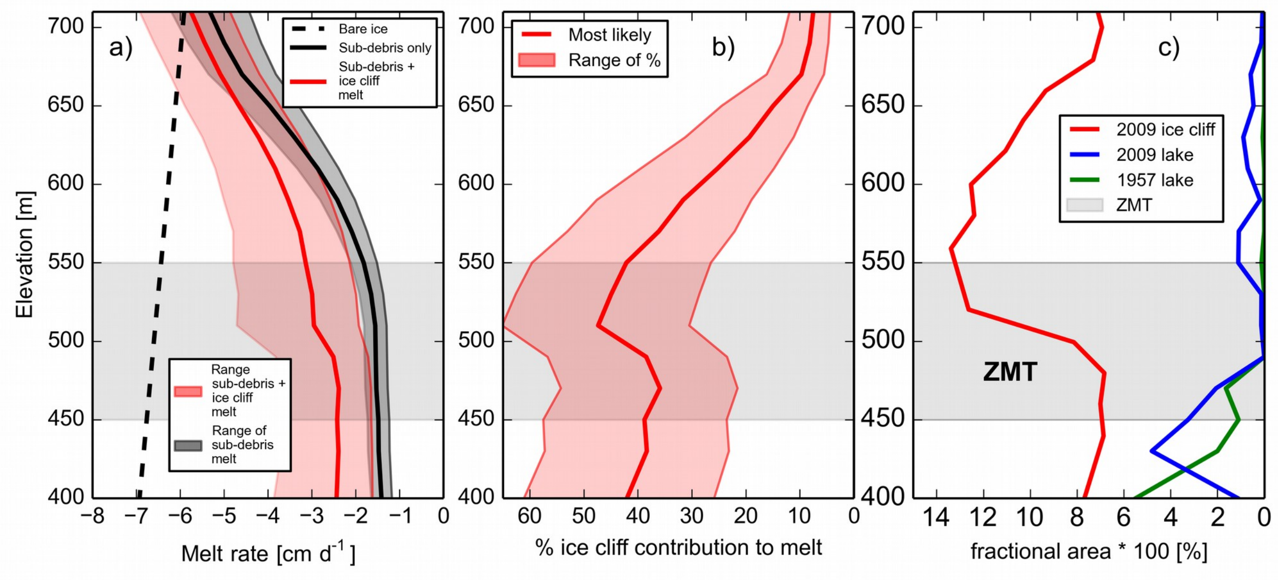
We did consider this approach but because of the number of ice cliffs in the study area we felt that this could be another manuscript in of itself, requiring significant justification and method development. We would need to remove the bare ice effects from each pixel before estimating debris thickness.

For ice cliffs, justification of a fit to elevation needs to be more explicit in this Part of the study.

We agree with this elevation does not need to be the main control here. We will explore other curve fits but a uniform distribution with elevation seems most defensible.

The conversion of backwasting rates into ice cliff melt misses a slope correction for the cliff area, which results in an underestimation of melt (see comment on L190).

The reviewer is correct and we have included an updated Figure 10 below. The change in the distributed melt estimates was minor but not negligible though.



This raises a difficult question that has not been carefully considered yet for debris-covered glaciers, which is that the real surface area can be 10-20% higher than the planimetric area. For this study comparing geodetic thinning observations with estimated melt, that is an important aspect to consider, as melt occurs relative to the real surface area. This effect is especially pronounced for ice cliffs, but is also crucial for the 'background' melt rate if the glacier surface is highly variable.

Having access to a 5 m DEM of the glacier we noticed this same effect when including the local surface for sub-debris melt rate. We agree and are happy to show this effect for sub-debris melt in a figure or two in the supplemental.

It is too bad that more recent geodetic difference data were not included in the analysis. At present it is not clear how the long-term thinning rate relates directly to the 2011 observations. Would it be possible to use the ArcticDEM datasets to derive a recent-period thinning pattern? This would be more meaningful for a comparison to the contemporary distribution of surface features. The long-term perspective is still useful for understanding dynamic changes, but comparing 50-year lowering rates to one-year melt patterns does not provide a definitive answer, especially for a clearly changing system.

We have these more recent geodetic estimates and we will include them in revisions. There is also laser altymetry data that supports the stable location of the zone of maximum thinning in the long run as well as the more recent decades.

Some rewriting is needed for readability and presentation standards in The Cryosphere. Although the ideas are well developed, some sections of the paper read as bullet points and/or the word choice has not been considered carefully, leading to some of my comments below.

We agree that there is some ambiguity in the writing that needs to be corrected. This review helped us understand where we can tighten the wording up considerably. Thank you kindly!

Detailed comments:

L23. Presumably this 19% of melt is for the debris-covered area?

Yes we will make sure the study area is defined correctly.

L25. Just a comment, for you to adopt or disregard: The literature has tended to use 'ponds' for these features as they are much smaller than supraglacial lakes on, e.g., the Greenland or Antarctic Ice Sheets.

We will switch to ponds. Helpful suggestion and we should be consistent with the literature.

L25. It would be nice to have the %areal coverage numbers here in the abstract, not just 'doubled'.

L27-27. This wording isn't very clear. By average melt rates you seem to mean the average for all surfaces, but as worded this seems to refer only to sub-debris melt. In the latter half of the sentence, do you mean that the overall melt relationship still follows an Ostrem-type relationship, even after accounting for cliffs and ponds?

We will clarify this as it is really important, making sure we emphasize which components we are including in the distributed melt.

Thank you for pointing this out. We will clarify this

Suggested change: Despite abundant ice cliffs and expanding surface lakes, average melt rates are suppressed by debris, with the primary control on elevation-band-average melt rates appearing to be debris thickness.

L34. These are broken sentences. The first half needs a reference even if it is now well understood.

We will join with a comma

L95. Has this happened progressively in the past 82 years, or primarily over some later period?

This is outlined in the citation, Rickman and Rosenkrans (1997). We will add a sentence describing the timing of expansion of lakes.

L98. Why is 'partially' here?

Good catch.

We will change this to: The presence of ice cliffs, surface lakes, and variations in debris thickness on debris-covered glaciers makes distributed estimates of mass balance difficult.

L108. Which spectral bands of the image do you use?

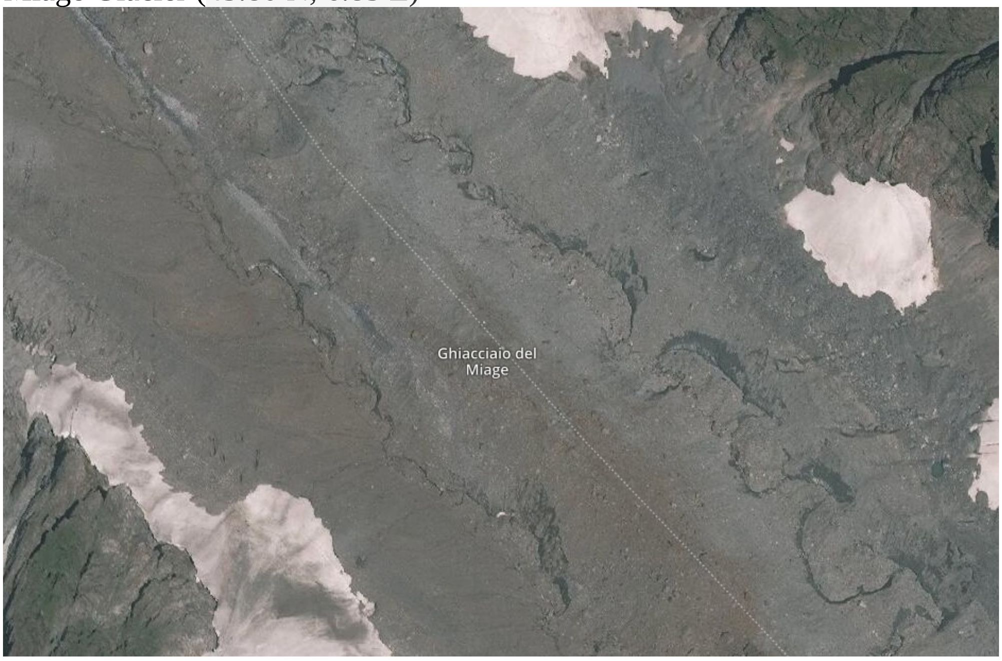
We will add:

On 109, before "The 2009 WV image…" add "We use the panchromatic band, which integrates radiance across the visible spectrum and provides the highest spatial resolution".

L113. This is true in certain conditions, depending both on debris lithology and meteorology. Such conditions may be prevalent for Alaskan supraglacial debris and melt seasons, but it is important to think whether such a method is transferable.

This is a good point. We looked into the universality of this observation that ice cliffs are generally darker than the surrounding debris cover using WorldView imagery in the Digital Globe's EVWHS viewer. We attach screenshots of Miage, Koxhar, and Lirung glaciers below. These are debris covered glaciers from the Alps and Tibet, which are found at different latitudes and lithologies than Kennicott Glacier. We find that, consistent with our Kennicott Glacier observations, ice cliffs on these other glaciers are generally darker than the surrounding debris cover. We can find examples where this is not true, for example along parts of the southwestern margin of Miage Glacier in the snapshot below. These regions may go undetected in the method we present here. We address this limitation in the manuscript (L281 in the original submission). To address this comment, we add at the end of this paragraph "This observation of darker ice cliffs is generally true for Kennicott and several other debris covered glaciers we examined, but this relationship should be verified before application to a different glacier of interest. There are situations (e.g., variable debris-covered and debris-free ice) where this method could detect darker regions that are not related to ice cliffs. Output should be examined to ensure that such conditions do not contaminate results". This language also partly addresses your L126-127 comment as well.

Miage Glacier (45.80 N, 6.85 E)



Koxkar (41.78 N, 80.1 E)

Lirung Glacier (28.23 N, 85.56 E)

L118. This is semantics perhaps, but I would argue that this whole workflow is your 'cliff detection method' (the name of step 2). It is a bit strange to have a 'cliff detection method' as the main step of a 'cliff detection workflow'.

Yes we need to use consistent terminology here and we will address it in revisions.

We have changed "detection" to "delineation" throughout.

L118. Is the histogram stretch just a linear min/max stretch? Is this histogram stretch applied to the image globally or locally within a patch? What spectral band(s) are used in either approach? I suppose that you also start with a debris outline (and glacier outline), which is important to acknowledge for a remote sensing approach.

We use a linear histogram stretch uniformly across the entire image. We do not clip the raster to the glacier, and include both on-ice and off-ice areas. The method ends up not being sensitive to this because we are tuning histogram stretch parameters to optimize performance on debris-covered ice, so it doesn't affect the stretch. We have added language at the end of the paragraph to clarify these points.

The step we call "detection" is the 'heart' of the ice cliff delineation method. The other steps are essentially pre- and post-processing. We have added these terms to the numbered steps to make this clearer.

Change to "1) processing: stretching the image brightness histogram to a suitable range for our ice cliff detection methods; 2) detection: applying an ice cliff detection method; and 3) post-processing: morphologically filtering of the detected ice cliffs (Fig. 5). We apply a linear histogram stretch uniformly across the image, including both the glacier and surrounding off-ice areas."

L123. Is the saturation stretch part of your step 1, or a separate part of step 2?

The saturation stretch is a pre-processing step (Step 1). However, we use different stretch values depending on which method we use (edge detection or adaptive binary), because the methods perform best with different exposure levels. We have added language to clarify this.

Change to "In pre-processing, we use separate saturation stretches (Fig. 5) for each method by applying the exposure function in the scikit-image package (skimage). The different methods perform best with different exposure levels, so we create two separate stretched orthoimages in pre-processing

L125. I see that the size of the moving window is included later as one of your parameters in the MC optimization. It might be good to give the reader a warning that that is the direction the methods are going, and that you will start with a description of the implementation of each approach first.

Thanks for this suggestion, we added language to foreshadow this progression.

Add at L120 after (Fig. 5): These steps introduce several processing parameters, which we select using a Monte Carlo optimization method. Below, we first present the processing steps, followed by our parameter optimization procedure.

L126-7. I would certainly not say that adaptive thresholding is insensitive to changes in debris cover and illumination, but it may be less sensitive. Ice cliffs are not uniform in surface character, and can appear both brighter and darker than the surrounding topography in different circumstances. It may be that these nuances are not so evident in the lower portion of Kennicott Glacier, but two particular cases would pose a major challenge for the ABT approach: 1) a population of ice cliffs with variable surface character (debris-free vs covered with fines) which will increase the spectral variance of the cliff population; 2) otherwise dark (potentially wet) debris. This is discussed later in the manuscript, but for a presentation of a new method, I think the accuracy and appropriate application needs some further advertisement/warnings.

We changed this to state that the approach is less sensitive to these variations than a global threshold. We added specific reference to this potentially problematic situation of alternating debris-free vs. debris-covered in the text we describe in response to your comment on L113. In response to that comment, we also added "warnings" that the output should be examined to look for spurious results, and that a user should verify that ice cliffs are in fact darker for their glacier of interest.

Change L126-127 to : "Because the brightness threshold varies across the image, the *ABT* approach is less sensitive to changes in illumination and debris color than a global threshold."

L133. Some of this content is Results

Yes, these can be considered results, but our purpose for including this language here is to motivate our last step in the processing procedure (morphological filtering). We have added a sentence to the start of this paragraph to make this intent clear.

Add to L132 start of paragraph: "The last step in our processing process is morphological filtering to remove spurious data".

L142. Was this 3% disagreement in total area? Can you provide a dice score for the two independent outlines? It is very possible to derive largely different cliff distributions but arrive with the same area.

Yes, the 3% was total area. WHA to look into Dice score.

L144-149. If I understand correctly, there are thus 6 parameters for the ABT and 5 for the SED implementations, with 4 shared between the two. How did this occur in practical terms? 2500 runs for each implementation, or 2500 runs used the same values for the shared parameters? More importantly, it is worth noting that with 5 parameters, 2500 runs results in an effective sampling of ~4.8x in each parameter (4.8^5 ~ 2500).

We added text to clarify this procedure. In each of the 2500 iterations, we select the value of every parameter at random using a uniform probability distribution across a set range of possible values for that parameter. The ranges were determined using operator judgement to cover all physically-meaningful values. This method allows searching a wider parameter space with fewer iterations than the approach you describe. Interactions between processing parameters across their entire possible ranges are captured. The parameter space is not sampled as systematically as you describe, but it is covered more broadly.

Change L148 sentence starting with "We ran…" to "We ran the ice cliff detection algorithm 2500 times with differing parameter choice. In each iteration, every parameter is randomly selected using a uniform probability distributions over that respective parameters range of possible values (Duan et al., 1992). This method allows us to efficiently test performance across a wide range of parameter values and is sensitive to interaction between selected parameters across their ranges."

L152. 'The origin' is a bit ambiguous here, as it is true for figure 5, with the x-axis as the negation of the true positive rate. So really this is ranked by distance from (1,0) in your optimization space, correct?

Thanks for pointing out this potential confusion. We added that we mean the origin on Figure 7, which you are right is complicated by the "1 – true positive" term. Perfect model performance is TP,FP=(1,0) like you say, this just becomes (0,0) on that plot due the "1-TP" term.

Change to "Euclidean distance from the origin on Figure 7, which defines…"

L153. Why did you choose to reduce the FP rate (at the expense of TP) from the optimal parameter set? Can you please provide a dice coefficient for this parameter set for each approach?

We will look into the dice coefficient approach. Thank you for pointing this out.

L157. Process observation (2) actually refers to melt rates, rather than backwasting rates.

Good catch we will correct this.

L158. The influence of lakes was noted earlier by Brun et al (2016) and Miles et al (2016), among others.

We will cite these works.

L163. In Part A it is clear that while elevation is a principal driver of debris thickness variability, there is considerable heterogeneity within any elevation bin. As your field measurements of debris thickness could not encompass the entire study area (that would not have been feasible), do you think they sufficiently characterise the unmeasured area (particularly the NW of the domain)? Have you tested the importance of debris thickness heterogeneity in your overall melt estimates? The subdebris melt relationship is not linear, so melt calculated with a mean thickness may not accurately approximate the mean melt rate.

This is a good point that is difficult to address for all folks working on debris-covered glaciers. Because we are trying to find out if ice cliffs can compensate for sub-debris melt this bias actually makes our estimate more generous.

Because many of our debris thickness measurements were taken at the top of ice cliffs though we suspect that our estimates underestimate debris thickness. We take our debris thickness measurements to be minimum estimates which means that sub-debris melt rate are likely to be even lower than what we estimate throughout this study, and through the study area.

L167. It is worth noting that you neglect internal ablation as well as other thermokarst processes (ponds, streams) in this computation for practical reasons.

Yes, we will note this.

L176. Please provide a goodness-of-fit for this empirical equation.

We will add this.

L180. It is interesting that as formulated, b_ice is the measured clean ice melt rate near the top of the study area, rather than the lapsed melt rate for each debris point. This is much more practical, but ignores the real melt suppression by the debris as a shortcut to a rate.

This is a good point and you laid out the reasoning, it is simply practical. At one point does emphasizing the details of physical processes (i.e. that bare ice melt rate increases a bit downglacier) get in the way of simple representation of the essence of sub-debris melt? The question really comes down to how important is it that bare-ice lapsed melt rates increase a little downglacier compared to the effects of debris thickening debris. On Kennicott Glacier it is clear to us that thickening debris is much more important than increasing energy for melt at lower elevations. The equation from Anderson and Anderson (2016) can easily be used with increasing bare ice melt rates if the user desires it.

In any case, I presume that the equation (as in Anderson and Anderson 2016) is based on the measurements presented in Part A? Please provide a goodness-of-fit measure.

Yes equation in Anderson and Anderson, 2016 is based on the data from part A. But note that the other fits from other glaciers in Anderson and Anderson (2016) are not necessarily taken across an elevation range. We will report an RMSE for the curve fits, optimizing for h_star.

L185. I am sceptical of this linear fit given the spread of observations in Part A – a goodness of fit would be expected to be very low.

We will provide additional curve fits for ice cliff backwasting and additional versions of Figure 10 to show the effect of this curve fit on our broader results. These figures can be placed in the supplemental.

If elevation is a secondary control, what might you presume is a primary control for the difference in backwasting rates?

We could address the scatter in backwasting rates more formally in Part A. We suspect that it is local debris thickness but a more formal analysis is needed. This could be a nice contribution to Part A. Because of the number of angles to look at the data and analyses in the full body of work we feel that multiple parts are justified.

L187. The similarity of backwasting rates for cliffs with/without lakes may be due to the observation type and period. Ponds and streams tend to incise thermoerosional notches, which can later collapse, thus enhancing the seasonal mass losses but not affecting what one would observe from the top of the cliff over a month or two. This is not a criticism of your work, it is just worth noting that this nonobservation doesn't mean a melt enhancement is not occurring.

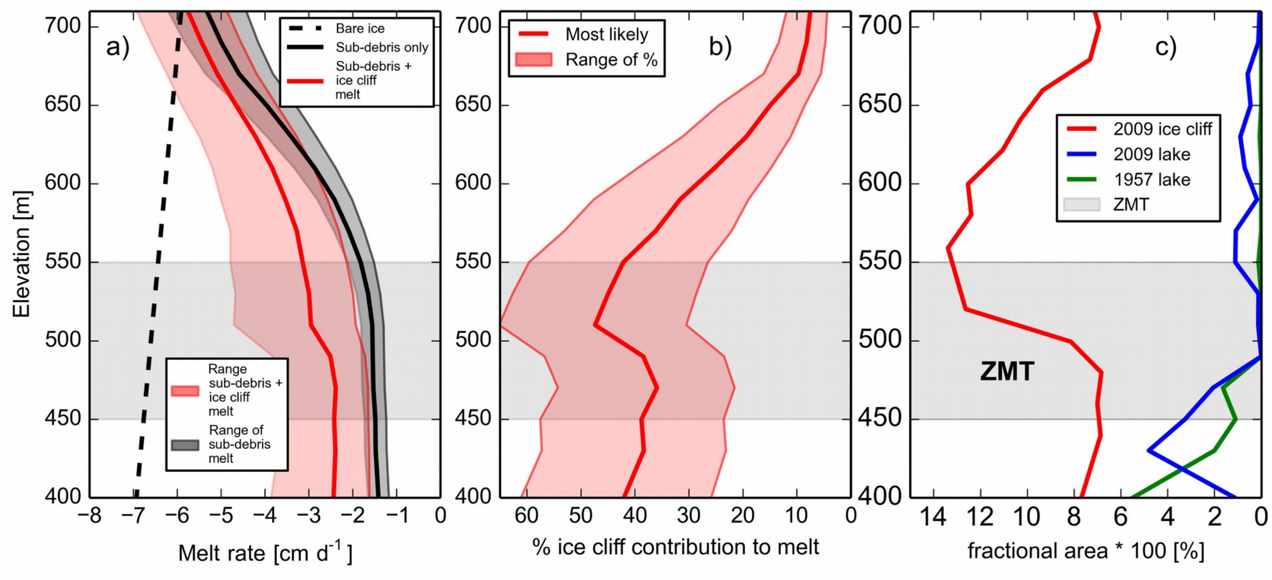
This is a very thoughtful comment. We agree. It is a really neat next line of research! We have in fact observed this very collapse phenomenon on a glacier in Switzerland!

L190. The correction in backwasting rates for cliff slope is correct, but it is also necessary to correct the cliff area from planimetric to surface area in order to correctly estimate melt from these inclined features, as melt occurs perpendicular to the surface. Thus

A_i =A/cos(theta)

With theta=40 degrees, this is a factor of 1.3 to all of the cliff-related melt calculations.

We will look into one more code but we think the reviewer is correct here. We have updated figure 10 with the correction (noting that the red bands show the extreme range discussed in the text):



L192. I think it's reasonable to apply a constant slope for the melt calculation, but why is this the 'most likely case'? Can you please provide some supporting information as supplemental figures, etc?

This a good point. The 'most likely estimate' based on our data collection and analysis.

L201. 'Most likely' is superfluous here; it is an estimate. It's nice that you provide bounds!

We will change to 'most likely estimate.'

L202. This 'best estimate' is using the parameter values already given in the text, correct?

Yes we will clarify this.

L213. I suppose you use the lapse rate (per timestamp? Hourly? Daily? Mean?) between the two stations for this estimation? It is notable that this lapse rate approach corresponded poorly to your on-glacier temperature observations (in the debris-covered area, Part A) – how do you think such an approach would correspond to the on-ice calculations?

This is good to clarify. Based on our understanding of melt factors (MFs) they are always relative to the temperature data used to tune the MF. And that off-glacier lapse rates are often a better representation of the lower 1 km of the atmosphere independent of the glacier. We could also just draw a vertical line in Figure 10a, assuming that temperature did not increase at lower elevations and the ice cliffs would still not compensate for the insulating effects of debris.

L218. Did you attempt to digitize ice cliffs in 1957 as well? The mention of ponds (and long-term change) is quite sudden, and should maybe be better integrated with the text.

We did not attempt to digitize ice cliffs, it would not be possible. We will better integrate this text but we feel this adds some good context to the work.

L219. 'insure' should be 'ensure'

we will change this.

L243. Should this be between 520 and 620 m? The fractional area is more meaningful than total area for understanding the cliff distribution.

Good catch we will discuss fractional area here.

L245. Importantly, this is of reported values.

We aren't exactly sure what is meant here but we could provide a more robust estimate of debris-cover from the other glaciers. We will search for this.

L247. N is too small to note any meaningful correlation between debris thickness and ice cliff coverage, unless you have a physical mechanism to implicate.

We agree that N is too small. But we do have a physical mechanism to describe shy this might be the case it is actually outlined more broadly in Part C. But we can reference that here and make it more clear in Part C as well. Emphasizing how debris has this effect.

This is another place where we feel that we need Part C. Here there is an interesting observation based on the data presented in Part B but we cannot explain this physical mechanism in Part B because we need to support the idea that the melt pattern follows Ostrem's curve before we can really get into the explanation of the physical mechanisms in Part C.

L255 and 265. I am still confused about the ice cliff melt rate distribution, for 2 reasons. First, it appears that at high elevations in the study area, it appears that your modelled ice cliff melt rates are lower than the modelled subdebris melt rates.

This is a good observation. And something we can improve in the manuscript. This is happening because are using empirical fits to the data. At the highest elevations where we measured ice cliff backwasting rate there were some ice cliffs associated with crevasses and thin debris that retreated slowly.

This is not plausible (or there would not be cliffs!). Second, the ice cliff melt rates in this region are also lower than for clean ice, which should be an approximate lower bound for ice cliff melt at all elevations: are nearly bare ice, but with surface debris well below the critical thickness (thus enhancing melt relative to bare ice, if we can neglect increased shading).

Also a good point. This also gets to the issue raised by this reviewer for our curve fits for backwasting rate with elevation. Perhaps a more physical way of representing the backwasting rate with elevation would be to pin the minimum melt rate of the ice cliffs to the bare-ice melt rate (6 cm/d) or use a uniform backwasting rate with elevation (at say the bare ice melt rate 6 cm/d).

While these changes will improve the analysis but we are confident, as this reviewer points out that these changes will not change the overall conclusions. Thank you for pointing this out.

L257. These rates correspond to a mean cliff enhancement factor of 1.72 relative to the mean subdebris melt rate, which I suppose is lower than anywhere else due to the thin debris.

That makes sense to us. This is a nice way to present the relative effects of the two.

L259. ‘Dominates’ is a strange term here. Certainly the reduced melt rate due to debris thickness is apparent, and debris thickness differences are more important than the difference in cliff density.

We can re-phrase this emphasizing that the melt rates follow more closely the sub-debris melt rate curve than the ice cliff backwasting curve (independent of the curve fit through the backwasting data).

L265. This appears to be a typo – the cliff melt rates are an order of magnitude lower than under debris?

We need to clarify the text here so it it more clear what we are referring to here. If you took all melt from ice cliffs in each elevation band and then calculated how much that lowered the entire area of the elevation band (ice cliffs + debris area) then ice cliffs lower the entire surface by the rates quoted in this sentence.

L271. This is again very disjointed to the rest of the analysis.

We will work to incorporate this better. We also feel that adding in streams to Part B would also feel disjointed and disrupt the flow of the manuscript.

Also, the low lake coverage in the upper ZMT makes a lot of sense as this area has steeper surface slopes in 2009 (Fig 2).

We agree with this.

L282. I appreciate consideration of the applicability and extension of this method to other sites/scenes. I think the biggest challenge for application to other scenes is that the tested parameter sets produced extremely variable results, and would need optimisation for every new site and image. There are also seasonality patterns to consider- cliffs often retain snow longer than the debris surface, for example.

We have not rigorously tested application of this procedure to other glaciers and scenes because our primary goal was to estimate ice cliff area on Kennicott Glacier for this study. It is plausible these processing parameters would perform well on other scenes of Kennicott Glacier, and perhaps for other glaciers as well. Illumination and sun angle can vary from scene-to-scene, and debris color can vary across different glaciers. However, Kennicott itself has several debris colors and textures, and the method does not appear to systematically differ in performance from one debris band to the next – this suggests the routine is not strongly sensitive to debris color. Varying illumination may change the ideal processing parameters, but the fact that the adaptive binary threshold is normalized by the brightness of pixels surrounding ice cliffs should mitigate sensitivity to this issue. That being said, you are correct that optimal performance could require training data (i.e., manually delineated ice cliffs) for a new scene to find optimal parameters. Manually digitizing ice cliffs in a few training areas is not incredibly time intensive, so we do not view this as a critical shortcoming of this method. We added text stating that the transferability of these processing parameters requires further investigation.

At L285 before “Using multispectral imagery…” add “The transferability of optimal processing parameters (both across time and space) requires further investigation.”

L324. The variability in observed backwasting rate is considerably stronger than any bias due to the observation location – the question is really where the mean lies.

We agree with this and will provide additional analyses to address the curve fit through the backwasting data.

L340. The potential distal effect of these features is conceptually well understood to be via internal ablation along englacial conduits, but is not possible to validate at present. See Benn et al (2001, 2012, 2017), Sakai et al (2002).

We will take a closer look at these papers. We agree that these effects are related to englacial conduits. We are wondering further how then englacial conduit melt out relates to the rest of the debris-covered glacier system. We believe that internal ablation though has a small effect on overall thinning. It isn’t clear what physically will produce the heat needed to cause the thinning rates observed from debris-covered glaciers.

L344. This section/paragraph feels orphaned. It is worthwhile to note that even accounting for the hypsometric distribution of cliffs, the spatial pattern still emulated Ostrem’s curve (just with different effective thicknesses), suggesting that this concept might be useful as a proxy for the altitudinal SMB pattern even where cliffs account for 40% of melt – just not directly comparable to stake measurements.

Nicely worded. We will work to incorporate this paragraph in, because this is really meant to be a transition and set up for Part C where we take the melt curve and compare it with a number of other analyses.

L353. ‘counter’ - should be singular

L356. ‘trend’ should be pattern

Table 3. I don’t think that Buri and Pellicciotti (2018) is the most appropriate study for Lirung Glacier for this purpose. Why the comment on EB below the table?

We will adjust the table accordingly.

Figure 2. It is not clear what the bars are in the upper left – is this the domain with supraglacial debris?

Oops that was dropped from the caption, yes it is the extent of continuous supraglacial debris transversely across the glacier.

Figure 3. This is a very nice conceptual summary! Can you include a pond or stream?

Yes we will add those.

Figure 4. In the caption for ‘c’, there is a reference to a ‘black line’ which corresponds to the ‘solid’ line I think.

Yes we will correct this.

Figure 5. This is a very nice summary of the method. Can you reproduce the same for the Sobel method to be included in the Supplementary Information?

Thank you. We have included a similar workflow figure for the sobel edge detection method in the supplements.

Figure 6. Panel (b) does not depict the bare ice area outlines as in (c).

We have added the bare ice cliff outline to panel b.

Figure 7. Nice depiction of the optimization. No colorscale is shown, though, and due to the different axis ranges, it is difficult to visualize the lowest Euclidean distance.

It is true that the different axis ranges make the Euclidean distance harder to visualize. There can be many more false positives than false negatives. In the limit, you can have as many false positives as you have pixels, whereas false negatives can only occur on ice cliff pixels. We chose the axis limits to show the range of possible outcomes rather than omit many data points to have the plot be at equal scale. We omit a color bar because there is already a lot going on with this figure, the colors are not crucial for understanding the figure, but rather facilitate visualizing distance, and we state the meaning of the colors in the caption.

Figure 9. See my comments in the text on line 255. Surely the lowest ice cliff melt rate (here 2.9cm/d) should correspond in space to the highest sub-debris melt rate (5.8 cm/d) – at the highest elevations. But then, the cliff melt rate should not be lower that the subdebris melt

rate – that makes little sense. This suggests to me that the linear parameterization of ice cliff melt with elevation may not be appropriate.

We agree and can adjust the 'most likely estimate' curve fits to reflect these physical realities.

Figure 10. Nice summary. Can you include a depiction of the cliff-only melt rates vs elevation in panel (a)?

This is a good idea. Yes we can, though we are somewhat hesitant to make the figure overly busy.

It is interesting that the cliff portion of melt is highest high in the ZMT, but still makes little difference in the mean melt rate profile.

Thank you kindly. It is precisely because of observations like this that we feel it is best to separate the manuscripts as we have. Such that there is space to discuss these items in more depth.

Also, it would be very meaningful to complete a version of panel (a) for the min and max melt parameterizations. Effectively these estimates are generous uncertainty bounds for your results.

The red bands are in fact the extreme uncertainty bounds. We need to better emphasize this in the caption and text because seeing that the red band as extreme bounds makes the analysis in section 4.2.1 much more compelling. Just need to clarify this for the reader.

Figure 11. Do you have a depiction of supraglacial streams (density or otherwise) to complete the picture?

We discuss streams in Part C and discuss feedbacks between the cliffs and streams there as we feel there is more room to discuss them there. We also note that we aren't sure what the actual effect of the streams are on ice cliffs. Because of this we feel that discussing streams in relation to ice cliff distribution as we do in Part C is a good idea as opposed to introducing them here. But this is a good item to consider as it would make for a more complete picture for Part B.

Figure 12. This is orphaned from the discussion and seems like an odd figure to close on.

Yes we see what you mean. We can move it to the supplement.

In addition to the changes described above we also propose these changes:

Part B: proposed changes

We feel that there is more than enough new material here for a stand alone paper, but in order to improve the manuscript we propose that we add these additional datasets/ideas to Part B:

- We will add additional text supporting the usefulness of our new ice cliff detection method. In the supplemental we will include additional satellite photos showing how ice cliffs tend to be darker than the surrounding debris so this method can therefore be applied on other glaciers. We will also compare our method with other approaches from other glaciers.

- We will present new DEM differences from 2007 to 2013. These dh/dt data show that the zone of maximum thinning remains in the same spot as for the period from 1957 to 2007.

We will also include additional laser altymetry data from 2007 that shows a similar thinning pattern.

-This will address one of the main criticisms from multiple reviewers.

- We will introduce back-of-the-envelope calculations of the possible effect of englacial melt, sub-glacial melt, melt under pond surfaces, and melt by streams. This will clear up any issues related to this manuscript not being comprehensive with regards to melt hotspots.
  - We will not include stream digitizations in this manuscript because we cannot possibly digitize all streams on the glacier surface (imagery is too coarse). The streams play more into the feedbacks in Part C. We will instead make arguments about the surface area coverage of streams and their plausible effect on surface melt.

- We will use a uniform curve fit through the ice cliff backwasting data. And also explore the effect of other curve fits, producing different figure 10a and 10bs which we will put in the supplemental and discuss in the main text.

- Add in the paragraph description that links each of the three papers and helps guide the reader through each manuscript.
- Make sure it is clear how generous the uncertainty estimates already are in this paper. One of the reviewers missed these error estimates completely.
- Emphasize the increasing importance of ice cliffs under thicker and thicker debris.