Author Response to Referee #2 (Anonymous)

We appreciate the helpful comments of Referee #2. All comments were well taken and we provide detailed responses below.

For ease of differentiation between reviewer comments and author responses, we have changed the font, indented, and italicized the *reviewer comments*. Please find our responses below.

> *Consistency is not the same as significance. Many caveats, uncertainties in resolution and signal detection, and fundamentally unverified links to physical processes mean that this paper is ultimately a concept and database that falls short on presenting meaningful analysis or testable hypotheses. At the end of this detailed paper, with 24 figures, multiple tables, and additional supplemental material, one is left with no new scientific insights about the forcing or nature of cryospheric changes in SE Peru. Despite a lot of detail and many qualitative statements about relatively small variations, it is not convincingly clear whether the changes are real or attributable to methodological limitations, resolution, and/or interpretation errors. The hypotheses posed are not testable with the given dataset of images. The richness of the geospatial and temporal variability is not fully exploited in getting at more understanding of processes.*

We appreciate this comment and insight and we regret that the reviewer feels that "this paper is ultimately a concept and database that falls short on presenting meaningful analysis or testable hypotheses". While it is true that this paper is data-rich and analysis-light when compared to some other papers, we find that there is also a need for a paper such as ours. Importantly, when equal guidelines as stated by the reviewer are applied to the Cryospheric or Geomorphologic Community, very few manuscripts would get published. We emphasize that some of the decline rates, descriptions and process-linkages (e.g., lake-area growth and glacial decline) provided by us are instrumental in formulating physical-based interpretations.

Both glacial and lake area data are still remarkably scarce throughout the Cordillera Vilcanota. Even area measurements of the Quelccaya Ice Cap, which has been well studied, do not provide as complete of a record as they could of the last 20-30 years. Other studies, such as Salzmann et al. (2013) provide some measurements of the QIC and the Cordillera Vilcanota, but proceed to pursue more in-depth climate analyses. We believe the data in our paper complements their data (and those of other studies in this area) by providing additional information on glacial area trends between their data points. We emphasize that our updated time series of glacial-area changes show significant variability throughout the past ~30 years. Decline rates can of course be calculated between two area measurements, however, we argue that decline rates can be skewed by the end-member glacial areas, which themself depend on the specific image that is used to obtain the area measurement. As the reviewer will see in our updated manuscript, even when obtaining areas from images which appear snow free, the glacial areas still vary slightly (ours overlap within our 1-sigma error bars).

Additionally, while we realize that for some of our years we could only obtain one area measurement, therefore not knowing whether it truly is a minimum or not, our decline rates for 1988-1999 and 2000-2010 incorporate not just two area measurements, but multiple area measurements, thereby reducing the error associated with potentially using an area measurement that was not the minimum area for that year. The density of our dataset enables us to produce decadal retreat rates and correlation with lake-level changes that would otherwise not be as possible or as accurate with many fewer images. We do note that our trends for the QIC are similar to those of Salzmann et al. (2013); their decline in area from 1985-2009 is 23 % while ours from 1988-2010 is 24 % (22 % from 1988-2009 – we no longer have a 1985 measurement to use in these comparisons). However, we agree with the reviewer that there is always room for improvement and that additional spatiotemporal analysis could be employed. In fact, we have tried several different approaches prior to synthesizing the manuscript, but we were not able to provide significant more insight as in the revised manuscript.

Additionally, in response to reviewers #1 and #2, we have laid out some of the problems we were facing when generating a dense time series. We are now confident in our glacial area measurements, particularly those for which multiple measurements for a particular year exist, which enables us to more accurately obtain an estimate of the minimum glacial area for that year. For those years where only one area measurement exists, we are confident that the minimum area for that year falls within the lower error bound (1-sigma) of that area measurement.

Importantly, our study provides uncertainties for both regressions and in area measurements, which has not always been the case in previous studies. Some studies do report uncertainties in glacial-area delineation, which is useful, however, in addition to estimating errors based on the number of perimeter pixels used in our areas, we have also included uncertainties for the regressions/decline rates.

> *This paper is a descriptive one that unfortunately lacks analytical rigor. Many associations are made without statistical significance. The resulting inferences on causation are without impact. The final conclusions are not novel. Because the methods themselves are not convincingly linked to process (i.e. snowlines to ELA; glacier area changes to mass balance, or even to actual ice rather than transient snow).*

In the revised manuscript, we have added new analysis to link lake-area changes to glacial-area changes. Our methodology involves glacial-area changes and not volumetric measurements, because these data are less dense and volumetric changes derived from digital elevation models are often associated with large uncertainties. Importantly, our database and dense time series allows to distinguish between decline rates in the 1990’s and 2000’s. However, we would like to point out that we have generated DEMs from various stereo-satellite and radar sources and are in the process of estimating their uncertainties (which are unfortunately in some cases large).

> *Furthermore, though internally consistent in image processing to track changes, the dataset does not conform with previous (and ongoing) Peruvian glacier and lake inventories.*

We do not understand the referee’s comment with regards to “not conform[ing] with previous (and ongoing) Peruvian glacier and lake inventories”. We feel our methods do conform with these inventories. Discrepancies between the different inventories arise through use of different band ratios (NDSI, TM3/TM5, TM3/TM4 etc.), different thresholds on these band ratios to identify snow/ice, and different filtering techniques and kernel sizes. Even within Peru alone, there is no “standard”. Racoviteanu et al. (2008) used the NDSI in the Cordillera Blanca, while Salzmann et al. (2013) used a ratio of TM4/TM5 in the Cordillera Vilcanota. Between these two methods, Albert (2002) determined a 1.9 % difference in accuracy. Between the ratio TM3/TM5, which we used, and the TM4/TM5 ratio, he determined only a 1 % difference. Additionally, not all studies report the threshold they used with their band ratios, and so we followed the TM3/TM5 methodology of Svoboda and Paul (2009) to follow an established method and threshold. We did include a closing filter, which other studies have not. We respectfully disagree with the reviewer’s comment that we have not conformed to other Peruvian glacier inventories.

This is true for the lake inventory also. We pursued the methodology outlined by Huggel et al. (2002). The identification of thresholds for the NDWI values to identify the lakes is subjective, and depends on each individual image. Lake sediment content varies widely in glacial areas, and this causes difficulty in identifying a consistent threshold across all images.

Based on these factors, we disagree with this comment. Our methodology does conform with other studies, and we have shown that our data is in-line with previous publications. We agree, our results cannot be compared exactly due to differing glacial extents and potential differing dates of images, however, our results and trends align with those of other studies.

It is not clear how this product will help inform methods of climate change adaptation, the authors' more broadly stated goal.

What we intended by the use of this statement was that in order to adapt to climate change, one first has to understand how the glaciers and lakes are behaving. If one does not know more specifically how the glaciers and lakes are changing, how can one adapt to climate change? Our study provides more detail on glacial and lake area changes over recent decades. Importantly, our time series indicates that there is more inter-annual variability than has previously been recognized. These trends in glacial and lake areas can aid understanding of their behavior, and as such, ultimately, with a better understanding of glacier and lake changes, can help inform methods of climate change adaptation. However, we will revise this statement.

Our responses to the specific comments are below:

Title: Caveats and lack of multiple images per year limit the actual use of all the images. In fact, many of the early MSS scenes are not used in rate calculations or regressions. If all regression curves do not access MSS, why include it?

We previously did not include the MSS imagery in the regressions because they were taken from images with much lower spatial and spectral resolutions. Additionally, previously we included area measurements where classifications were problematic, and those that had potential snow. Where classifications look reasonable and where we have – after visual inspection – higher confidence in the imagery and lack of snow in the area, we have decided to include the MSS areas in the revised manuscript to give the reader a sense of past glacial areas, while the error bars indicate to the reader our lower confidence in the specific area measurement itself. As a result, the MSS area measurements extend the time series back to 1975 for all but two of the glacial IDs. In the revised manuscript we have included these MSS-derived areas in our regressions for each individual glacial ID for the whole time period, to gain the most long-term decline rate possible for each.

And thus how accurate is the title? None of the images date back to 1975 for rate calculations. Moreover, the most consistent and data rich time slice is 1988 to 2010 (coverage by Landsat TM/ETM+).

This is no longer the case as we now do include some area measurements derived from 1975 imagery, and as such, decline rates for each ID, although unfortunately not all. For example, QIC imagery dates back to 1980, not 1975. It is true, however, that the most data rich and consistent time slice is 1988-2010, but this cannot be avoided based on the imagery available. Additionally, all of our lake area time series date back to 1975. In summary, in our revised version of the manuscript we do include some of the MSS imagery from 1975 in the glacial area time series and in the rate calculations where the areas exist.

Methods: The authors claim to delimit their study area according to Morales Arnao (1998), but should rather use the existing glacier inventory framework, also tied to protocol of the WGMS.

We assume the referee is referring to the fact that we are looking at larger glacierized masses rather than separating each glacier into individual entities based on their drainage basins? This is true, and we will separate the glacierized masses into individual glaciers and pursue analysis on the individual glaciers in a future publication. However, here, we have chosen to aggregate them according to their main units, i.e., including them as part of the larger continuous mass to which each glacier belongs. We have done this for several reasons: most importantly, we aimed at obtaining regional decline rates rather than decline rates based on individual glaciers alone (which can be influenced by other factors such as aspect or elevation, amongst others). We do not want to include results for all glaciers individually per catchment area as our glacier results are only a part of this study, alongside the investigated lake area changes. Some of the lakes investigated are within watersheds which include multiple glacier drainage basins. So, for several reasons, we have chosen to show results for the main glacierized area units, rather than the glaciers individually. Quelccaya Ice Cap, as an example, is generally studied and its area reported as an ice cap, rather than being separated into the main drainage basins. Paul et al. (2013) mentions that this relates to purpose, with hydrologists preferring to use individual drainage basins in an ice cap, while glaciologists often keep the ice cap as one single unit. We understand that in the case of ice caps this is different, however, for the purpose of our study, we chose to keep the individual glaciers as part of their larger units.

With regards to delimiting the study area, as we have mentioned in our paper, different studies have used different extents of, for example, the QIC or the Cordillera Vilcanota. We have been unable to find a study which specifically mentions which glaciers are incorporated in the Cordillera Vilcanota, or specifically which glaciers are part of the QIC and which are just very nearby, such as on the curved extending arm to the north west of the main part of the ice cap. As a result, we have chosen to report as specifically as possible, which glacial areas we include in our measurements, so that this is transparent at least in our study.

> *Note that the paper refers to the inventory using 2 different references; are these the same? Ames et al., 1989; Hidrandina, 1988?*

We have cited both references, because one is the Spanish original (Hidrandina, 1988) including maps, while the Ames et al., 1989, is an English translation of part of the original work. While they are somewhat similar in content, we continue to cite both references.

> *Resolution: Resampling to 15 m with data that have resolution >15m is not gaining more accuracy. Why did the method not broaden to least resolved?*

We appreciate this comment and have revised our method section to make this clearer. We emphasize that we are not resampling to a higher resolution to gain more accuracy. Instead, we have used bilinear resampling to align all imagery to the same spatial extent with the same spatial resolution and projection. We have resampled the 30-m ASTER SWIR to 15-m only to match the resolution of the visible bands so that the bands could be used in conjunction with each other. However, when calculating uncertainties, we have always used the lowest resolution of any band used, i.e. for our glacier classifications which use ASTER bands 3, 4, and 1, while band 4 was resampled to the same resolution as bands 1 and 3, for our uncertainties we use the pixel resolution of band 4, 30 m, the lowest resolution of the bands used. We do not claim to add more accuracy, only that we are resampling the data to be able to merge different datasets with different spatial resolutions. We emphasize that this is a common procedure and has previously been used (e.g., Racoviteanu et al., 2008). We have explained this better in our revised manuscript.

> *Glacier change detection: The authors seemingly make a case for having a more authoritative coverage of imagery. However, they admit to the issue of snow cover. How is it that they discern images (or even subsections) without snow cover?*

We did not claim to have a "more authoritative coverage of imagery". We of course have access to the same imagery as other studies, however, we have processed more images than other studies. For the discussion paper, we processed 144 images and previously worked with 77 of these images for the glacial time series. In the revised manuscript, we have discarded 20 of them due to inappropriate snow cover, but added 13 additional images (12 of which have been used to add to the glacial area time series), which we found since publication of the discussion paper.

As for discerning images and subsections without snow cover, this is done visually, and as such is subjective. We have visually inspected each image (as authors in other studies have done) to determine whether or not snow is present. We admit that we have previously determined images to be not usable for the entire study area, but then have failed to exclude these glacier areas from the time series (see response to reviewer #1). Nevertheless, this has now been addressed, and the large number of images and decline rates based on several images make the declines rates more robust.

Supplemental material adds more details but not clarity. Fig SM C12: what is the base image date? If the snowlines are plotted, again with one date, why are the ranges given for multiple dates?

We have re-evaluated our snowline section, however, we are thankful for this comment. In the revised manuscript, we have made this figure clearer. In this figure in the Discussion Paper, the 3 smaller images at the top of the figure included their individual dates and snowlines. We then increased the size of the middle image from the top panel where the snowline was highest, and we overlaid the snowlines of the other two images on top of that, so the change in snowline through time in relation to each other was visible. We assumed the reader would recognize that the main image was taken from the panel at the top, and the snowlines from the three top images overlain. The median elevations for each snowline obtained from each of the three images are reported in the main image. All information in the main image stems from the three smaller images. We initially thought this was clear enough, however, we have made it clearer in the revised manuscript.

The authors claim a robust image analysis, but accuracy is not verifiable. Intra-annual changes on the scale of 19% imply not that the ice mass is actually that dynamic, but rather that snow-ice detection as this scale and using images is problematic. Ultimately more statistical significance needs to be quantified.

In response to this we refer to our responses to reviewer #1 and to our response two comments above, specifically that we acknowledge that an error occurred in our interpretation of glacial areas and we have re-evaluated our image time series to address this. As we have mentioned in response to reviewer #1 and #2, we are most confident in our area measurements for those years where multiple images exist, as by obtaining glacial areas on multiple images for those years, we are best able to hone in on the minimum glacial area (likely those with no additional snow, even though the other images used didn't appear to have snow obstruction either, but sometimes it is difficult to ascertain this).

Obviously, for those years with only one measurement of glacial area, we would much prefer to have multiple measurements to hone in on the best estimate of glacial area (specifically, the minimum glacial area) for each of those years. However, for some years the imagery just is not available to do this. We have also compared the images we have used to those used in other studies in this region. Some of the previously used images in other studies we have now chosen not to classify, as in some cases there appears to be the possibility of snow. We take this as evidence of the subjective nature of identifying potential snow in the imagery, and how classifying multiple images for a year (where possible) can be extremely beneficial. For those years where only one measurement exists, we find that our measurements are no more or less accurate than other studies that also only provide one measurement for a given year.

We are now confident in our updated images used, and the reason we highlight the intra-annual changes is to support the need to include dates of imagery used when area measurements are reported. Our revised manuscript continues to contain a dense glacial area time series with some images removed (20) and some new images added (12) as compared to the previous version. We argue that the decline rates derived from a denser time series, including regression uncertainty and uncertainties of area measurements, are a significant step forward. We do not claim to report significant intra-annual changes in the revised manuscript and we emphasize that our glacial-area changes from year-to-year overlap within their 1-sigma error bars.

p 582, L16): The authors claim: "While not all images are suitable for glacier classification (local/regional snow cover or clouds obscuring outlines), our study classified as many images as possible to gain as much information as possible on how the glacierized regions behave on an annual as well as a decadal time scale." This is vague.

We have rewritten this to say: "Not all images are suitable for glacier classification (local/regional snow cover or clouds obscuring outlines). In order to gain as much information as possible on how the glacierized regions behave on an annual as well as a decadal time scale, we have classified all images where ice is visible at the glacial boundary, no snow patches exists beyond the extent of the glacierized region, and cloud cover does not occlude any part of the glacial boundaries." We hope this is less vague.

> *They try to explain that they take care for accuracy by selecting image dates closest to minimum extents in the year. How do they know this? This needs further explanation, and this assumes that such a minimum exists, and is uniform across all years. Also, they then acknowledge that multiple images are often not available, so that decline rates are based on one image. Thus they can not get around the sticky problems. To the extent they make a novel contribution, a more rigorous assessment of these errors in attribution and lack of data is needed.*

We are not "selecting" image dates closest to minimum extents in the year. We only determine the minimum extent once we have obtained glacial areas using all usable images for one year. We then identify the minimum glacial area of all glacial areas within one year, and highlight that as a "minimum area". We realize that it may not be the absolutely minimum area for that year, however, where we have multiple images, we are confident that we are approximating the minimum area for that year. Multiple measurements for one year do overlap within their 1-sigma bounds. Importantly, we do not claim that the minimum is "uniform across all years". In fact, our data with multiple measurements from one year show that dates/months of minimum areas vary, depending on several regional climatic and local factors.

However, it is true that these data are not perfect and that there remains some ambiguity that the reviewer refers to as a "sticky problem" where we only have one data point per year. In these cases, we have done what other studies have also done: obtained an area measurement using one image for those years. In our study, we have included uncertainties in the regression. It is likely, based on what we observe during years with multiple area measurements, that the minimum area for these years where we only have one data point is within the 1-sigma error bars of the measurement that we have. Additionally, our decline rates are not based on one image because we do have multiple measurements over time. We look at retreat rates taken from >2 measurements and not just two end members, as has previously been done by other studies. The fact that we do this helps to reduce errors that may exist if an area measurement is higher one year due to its nature as being an only measurement that may not be the true minimum for that year.

> *Why were the lake classifications edited by the glacier classification (p 582)? Why was glacier assumed more accurate?*

This is a misunderstanding on the behalf of the reviewer. We do not edit the lake classifications with the glacier classifications, rather, we edit the glacier classifications with the lake classifications. This is to edit out the lakes that were erroneously classified by the glacier classification methodology, which often had problems distinguishing between the spectral signatures of snow, ice, and high sediment content lakes.

Neither classification was assumed to be more accurate. Given that the lakes need to be removed from the glacier classifications in any case, it made more sense to use the lake classification to do this with, rather than manually editing out each lake that was erroneously classified.

> *Moreover, if the value of this paper is in methods of lake delineation, it would be more appropriate to quantify the performance of the NDWI/hillshade algorithm. As it is, the authors simply say it "performed well." Based on what?*

We have not devised the methods of lake delineation that we have used. We have relied on previous studies to delineate lakes (e.g., Huggel et al., 2002; Bolch et al., 2008). We have added to the Bolch et al. method, and in this case, we elaborate on this in Figure 5 of the manuscript. In this figure we have also indicated what we refer to as "performed well", where we indicate which lakes are "well classified" (based on if the classified outline corresponds closely to the visual lake outline) and which lakes are not well classified. As such, the classifications "performing well" are based on visual identification of the lake outlines compared to the imagery. We have compared some of the automated lake classifications to manually-clicked outlines of the same lakes and find variable results which depend on the nature of the lake outlines themselves (for example, whether sediment content is high or low at the boundaries of the lake), but overall had a reliable match. Upon classification, the classifications of each lake were visually examined, and if the classified outlines did not match up visually with the lake outlines, those lake outlines were removed from that classification. We have modified this statement accordingly in the revised manuscript.

Geospatial data analyses: More could be done to analyse patterns and explore other controls on lake area and glacier mass variability. The authors don't explore reasons for the patterns shown. Many questions remain.

As previously stated, we have revised our interpretation of lake-area changes and have added additional analysis. Prior to submission of the first version of this manuscript, we have experimented with different geospatial methods – but none of them could explain the spatial patterns. We therefore present an analysis that explains some of the patterns, but we acknowledge this and do not claim to have found the source of all lake-level changes in the central Andes.

Why does the CV have 5x more area recession rate than QIC?

This is because of its significantly greater size than that of QIC. In the Discussion Paper we have reported both normalized and non-normalized recession rates. In the revised manuscript, non-normalized recession rates are now in the Supplementary Material. We include them as the non-normalized rates are more comprehensible, but dominated by size, and due to this we also include the normalized recession rates.

Why do smaller glaciers recede faster? This is not a novel observation, but is consistent (Rabatel et al., 2013), and probably relates to hypsometry. This would be easy to explore further, and should be. It would be important to document the hypsometry of the ice masses as well as their location. Snowlines are interesting, but more relevant in explaining the relative recession rates might be mass above the snowline. The authors direct only a qualitiative analysis, pointing to relative size of glacier cluster (on Fig. 9).

We are currently pursuing hypsometric analyses of the glacial areas and intend to present a figure reflecting this in the revised manuscript. We have also revised our snowline section and have calculated AARs for the QIC for 3 years; 1988, 1998 (which appears to be an extreme end member), and 2009, as well as AARs from records presented in the literature.

Lakes: What is a 'characteristic' proglacial lake? There is no discussion of the metrics considered, despite having a large # of lakes (n=50). What is mean, mode size? How is connectivity to glaciers established?

We have removed the term 'characteristic' in describing proglacial lakes. With regards to the metrics, we previously stated on page 595 that "41 out of 50 (or 82%) are less than 2 km^2, and only 3 out of 50 (or 6%) are larger than 5 km^2. However, we have included more metrics than just this in our updated version of the manuscript. Regarding establishment of connectivity to glaciers, on page 596 we said "Using the SRTM DEM, we delineated the watersheds for each of the lakes, and identified whether they were connected to glacial regions or not". These watersheds were delineated using standard-GIS flow accumulation procedures. If any of the resulting watersheds overlapped with any glacial areas, the lake of that watershed was considered to have "connectivity to glaciers". We shall explain this better in the revised manuscript.

Sibinacocha is arguably not a characteristic proglacial lake. Lake area increase is better explained by the fact that the dam was completed in 1996, and not related to a step change in glacier melt contribution. E.g. http://www.gmisa.com.pe/versioningles/web/energia_egemsa.htm.

We appreciate this comment and have revised our manuscript accordingly. The reason to include Lake Sibinacocha was to have a larger lake included in the analysis, but we realize that lake area post-1996 is likely controlled by hydropower management.

Connectivity to glacier melt is presumed by proximity only (since it is not explained otherwise). Yet ground water could filter and recharge lakes.

Please see our earlier response (two comments before): Connectivity to glacier melt is not presumed by proximity but is based on watershed coverage, as in, connectivity is presumed if any glacial area is within the extent of each given watershed. Ground water absolutely could filter and recharge lakes, and it is incorporated to some degree by use of watersheds.

P 587, L21: "different melting processes including GLOFs" are invoked to explain spatial and temporal differences in lake area changes. What are different melting processes? GLOFs: are there records or examples to show relationship to lake area changes? What other processes like evaporation, groundwater infiltration could be involved rather than simple melt-refilling? Here the reality on the ground needs more consideration and fundamental context explored (i.e. bedrock lithology, soil permeability); the whole region proximal to QIC is an elevated till plain, with large sections of poludified soils. None of these factors are evaluated.

We agree with the reviewer that the different melting processes are unclear. It is true that there are a large number of processes that can control lake-area behavior, stretching from glacial-melt processes to evaporation, groundwater infiltration, seepage, and other processes. We have revised this in the manuscript, and removed reference to GLOFs in relation to lake area changes as we have been unable to find concrete examples for this area.

With regards to evaluating other potential factors influencing lake area changes, we do not intend to give a closed water budget for this area, but instead first-order factors controlling these lake-area changes. The reviewer is correct in stating that there are several additional factors that influence lake levels, including bedrock lithology and soil permeability, but also permafrost processes and vegetation cover. It has been our intent to examine lake area changes purely on a first order level, with catchment area dictating the amount of water.

Fig. 17: The link of lake to glacier area here is also not fully justified, although a direct hydrologic connectivity is implicit in posing such a close annual match of annual rate change. QIC: sits on an ignimbrite plateau; fracture flow is possible, as is re-routing of meltwater channels along the receding ice margin. Some lakes are formed in contact with ice, while others are in basins of varying degree of drainage (i.e. Buffen et al., 2009, Quaternary Research 72, 157-163).

It is true that some lakes are formed in contact with ice, some on top of the ice, and others in basins of varying degree of drainage. However, all of the lakes that we investigate are investigated in relation to their drainage basin (catchment area) and all lakes are pro-glacial lakes. The results we illustrate in Figure 17 are those for all lakes that are within QIC watersheds. We are integrating all flow within the catchment, therefore, in this case it is not important whether a lake is formed on the ice or beside the ice or slightly away from the ice.

Fig 23: this figure is difficult to read. A summary table indicating size of lake, proximity to glaciers, and relative growth would be more helpful.

We find this to be an efficient way of presenting the large amount of data. The reviewer is correct in noting that it may not be easy to read, but the figure also contains the results of 50 individual lakes and their changes. We are considering the possibility of including a table with these results in the Supplementary Material.

> *The limitations in methodology prevent interpretation for small lakes, by the authors own admonition in discussion. Yet they also tell us that the majority of the dataset is comprised of small lakes. So how many of the actual lakes are large enough to actually discern a meaningful size change? This is something the authors should compute, given known resolution, and a reasoned depiction of uncertainty. Again, this all requires more explicit quantification and presentation of results; a table would be far superior. Ultimately, are the average % changes statistically significant? Without better quantification of both error and uncertainty, this is not clear, and therefore nor are any interpretations of process or causation. Fig. 24: here we see traces of small lakes, but without any indication of uncertainty based on the image resolution. Trends should have error bars. Lakes 35 and 26 are new as of 1988 ice edge. Again the edge of the QIC is not a smooth grade where melt should relate to lake size, given the steep, columnar ignimbrite cliffs over which ice cascades.*

We understand the referee's comment regarding this issue, however, the problem lies in how we phrased this in the discussion, not in the actual lake results. It is true, the majority of the dataset is comprised of small lakes, however, we see strong lake area changes in some of the small lakes, and minimal lake area changes in some of the large lakes. Due to the classification methodology and the need for varying thresholds per image, we have found that the lake area measurements do fluctuate beyond what is likely a signal, however, we use uncertainties to address this issue. (1-sigma errors of the lake area measurements have been calculated in the same format as those of the glacier area measurements, using Equation 1. We have mentioned this in the revised manuscript.) The areas of the majority of lakes do not vary significantly beyond their uncertainties, and by this, we meant they do not show clear signals or trends. Instead, we should have said "do not show significant lake area changes (growth or decline) beyond their uncertainties". Those lakes that we interpret as having clear signals are those where an area increase or decrease is visible beyond their 1-sigma uncertainties.

We are thankful to the reviewer's comments and have revised our statement in the revised manuscript. We also have included 1-sigma uncertainties on all lake area graphs, where previously there were some small lake area graphs (Figure 24) where error bars were omitted. We also provide examples of lakes which grow/decline beyond their uncertainties, in addition to those which do not, in both the manuscript and the Supplementary Material, so this can be more easily understood.

> *Snowlines: Like the lakes and glaciers, snowlines face problems of significance. What are the actual vertical uncertainties, given previous undisclosed methods? The authors equate ELA with snowline, and from outset describe their attempts "proved unsuccessful" in applying methods from other regions. How so? What was metric of success?*

With regards to equating ELA with snowline, we have not done this in the general fashion suggested. As we have mentioned on page 588-589, Rabatel et al. (2012) suggest that the highest altitude reached by the snowline over the course of the entire ablation season may provide an estimate of the ELA for that year, although it likely is an underestimation. We used this to obtain our best estimates of the ELA using the snowlines that we classified, only suggesting that the highest snowline reached for a given year represents a possible minimum estimate of the ELA. Our vertical uncertainties have been modified to only include those in the higher altitude direction given that Rabatel et al. suggest it is still likely an underestimate.

With regards to our attempts proving unsuccessful in applying methods from other regions, we refer to our response to reviewer #2, who also commented on this. In summary, we expected to easily be able to classify the snowlines if they were visible in the imagery, and also via use of previously disclosed methods. However, what we mean by “proved unsuccessful” is that we were unable to classify these snowlines using any of the previously disclosed automated algorithms; only MEMSA worked for all images, where the other methods did not. These snowlines would have been easily classifiable by hand, as visually, the snowline is obvious. However, we wanted to use an algorithm as manual classification is subjective and variable depending on the classifier even when the outline appears obvious (Paul et al., 2013), but algorithms previously used by others to classify the snowline in other regions did not work on all of our imagery. Because of this, we used MESMA, which worked well in most cases. Specifically, by “well” we mean that an overlay of the classified snowline corresponded closely with the snowline visible in the imagery. This represented our metric of success. We have significantly revised our snowline section in the updated manuscript, and now focus on the snowlines for 3 years spanning 21 years.

> *There is a problematic circularity to the justification of this study that also undermines significance of interpretation. The authors use previous studies to "validate" their measurements (P598), yet also claim previous work suffers an unspecific amount of uncertainty.*

We understand the reviewer’s comment with regards to this “circularity”. However, in mentioning this, it was not our intention to specifically compare our measurements to those of previous studies because of these problems. We would have liked to validate our measurements in reference to those of other studies, but by pointing out the various problems and uncertainty involved in the other measurements, we came to the conclusion that we couldn’t do this. We specifically discuss a potential reason for why our measurements are higher than the earlier measurements, and while we would have liked to validate our results, the problems we mention render this not possible. We have clarified this in the revised manuscript.

> *P598 L14: this critique of previous studies is not substantiated by any evidence: "most elevation measurements likely nave some uncertainty...large at these altitudes." Why? And how is this same critique not equally applied to this paper?*

This statement refers to the fact that earlier elevation measurements based on barometric methods have had larger uncertainties associated with then. Daily weather changes lead to rapid air-pressure changes. We have removed this statement in the revised manuscript.

> *Fig. 18: the dates of publications are listed for other snowline estimates, but this does not accord with the observation, or basis. For example, field work was carried out prior to pubs of Thompson and Mercer, and Hidrandina (1988) as a publication used aerial photography from the 1960s.*

The reviewer is correct and we are fully aware of this. We could not obtain the exact years of these snowline estimates, so we assumed that they were from around the time of publication and used that date. While multiple field seasons are mentioned in these publications and we don’t know which the measurement derives from, we have now decided to pick the field season year closest to the time of publication to use for the given equilibrium line altitude measurement. This equates to ~1976 for Mercer and Palacios (1977), ~1978 for Thompson (1980), and given that the QIC glacial areas in the WGI are from 1962, we now use 1962 for the Hidrandina (1988) estimate. We have updated this in the manuscript accordingly.

> *Broader Impact: The authors say their work can "be used by those seeking to develop methods to adapt to climate change in this region." How? What specifically would be used for adaptation, and how? Who would be able to access and utilize these data? This vague statement is unsubstantiated and appears to be a throw-*

away line without much thought. More coordination with previous and ongoing Peruvian glacier and lake inventories is encouraged.

As we have mentioned in a previous comment, our intended interpretation in writing this comment was that in order to adapt to climate change, one first has to understand how the glaciers and lakes are behaving. In the context of our study, we specifically refer to a dense time series of glacial areas throughout the Cordillera Vilcanota, which we have used to obtain more accurate (and more reliable) retreat rates than previously reported using two images alone. We have also established and compared roughly decadal retreat rates between the late 1980s and the 1990s, and the first decade of the 21st century, and we have linked retreat rates with lake area changes. We have revised this statement to read: "These data can be… used to gain a more in depth understanding of recent glacial and lake area changes throughout the Cordillera Vilcanota, given that documenting and understating past changes is a first step in preparing for future changes."

Technical corrections:

P578, L9: dryer should be spelled drier.

This has been corrected.

P589: much of the content of the Results is more appropriate to include in discussion.

We have revised and streamlined the manuscript and taken this comment into consideration.

Two separate ones are cited in text/listed for the Glacier Inventory, with different years: Ames et al., 1989; Hidrandina, 1988.

We appreciate this observation, however, we have explained in a previous response why we continue to cite both.

It would be more accurate to say Thompson and Hastenrath began research in QIC in 1974, and not 1963 (referencing B. Morales Arnao, 1998, USGS report 1386-I).

We have changed this to say 1974.

Too many figures; recommend cutting down.

We have removed several of the figures in the revised manuscript.

List of References Used:

Albert, T. H.: Evaluation of Remote Sensing Techniques for Ice-Area Classification Applied to the Tropical Quelccaya Ice Cap, Peru, Polar Geography, 26(3), 210–226, 2002.

Ames, A., Dolores, S., Valverde, A., Evangelista, P., Javier, D., Gavnini, W., Zuniga, J. and Gomez, V.: Glacier Inventory of Perú, Part 1, S. A. Hidrandina, Huaraz, Perú, 1989.

Bolch, T., Buchroithner, M. F., Peters, J., Baessler, M. and Bajracharya, S.: Identification of glacier motion and potentially dangerous glacial lakes in the Mt. Everest region/Nepal using spaceborne imagery, Natural Hazards and Earth System Science, 8, 1329–1340, 2008.

Hidrandina, S. A.: Glacier Inventory of Perú, Consejo Nacional de Ciencia y Technología, Perú, 1988.

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.

Mercer, J. H. and Palacios, O. M.: Radiocarbon dating of the last glaciation in Perú, Geology, 5, 600–604, 1977.

Paul, F., Barrand, N. E., Baumann, S., Berthier, E., Bolch, T., Casey, K., Frey, H., Joshi, S. P., Konovalov, V., Le Bris, R., Mölg, N., et al.: On the accuracy of glacier outlines derived from remote-sensing data, Annals of Glaciology, 54(63), 171–182, 2013.

Rabatel, A., Bermejo, A., Loarte, E., Soruco, A., Gomez, J., Leonardini, G., Vincent, C. and Sicart, J. E.: Can the snowline be used as an indicator of the equilibrium line and mass balance for glaciers in the outer tropics?, Journal of Glaciology, 58(212), 1027–1036, 2012.

Racoviteanu, A. E., Arnaud, Y., Williams, M. W. and Ordoñez, J.: Decadal changes in glacier parameters in the Cordillera Blanca, Peru, derived from remote sensing, Journal of Glaciology, 54(186), 499–510, 2008.

Salzmann, N., Huggel, C., Rohrer, M., Silverio, W., Mark, B. G., Burns, P. and Portocarrero, C.: Glacier changes and climate trends derived from multiple sources in the data scarce Cordillera Vilcanota region, southern Peruvian Andes, The Cryosphere, 7(1), 103–118, 2013.

Svoboda, F. and Paul, F.: A new glacier inventory on southern Baffin Island, Canada, from ASTER data: I. Applied methods, challenges and solutions, Annals of Glaciology, 50(53), 11–21, 2009.

Thompson, L. G.: Glaciological investigations of the tropical Quelccaya ice cap, Peru, Journal of Glaciology, 25(91), 69–84, 1980.