Response to the comments by Sam Herreid

The manuscript "Glacier shrinkage in the Alps continues unabated as revealed by a new glacier inventory from Sentinel-2" by Paul et al. presents a new glacier inventory that optimizes several criteria including consistency, a tight time-span of source data acquisition, precision, and complete coverage over the Alps. The authors provide a thorough and well written manuscript that I found easy and enjoyable to read. One unique element of this work is the lead author also authored an earlier study that derived a similar product from Landsat imagery acquired in 2003. I was surprised that, even with this degree of consistency, this new inventory was only marginally compatible for a comparison. This seems largely due to the inclusion of "new" glacierized area that was not previously mapped, yet the authors walk this back some in the discussion admitting that some "glaciers" might really be perennial snow and firn patches. Since other scientists will likely base their work on this inventory, I think it is most fitting that the expert authors here make decisions and interpretations that they are confident with. Below are mostly minor comments with some concerns regarding analysis subdivided by political boundaries, results that reflect the highly variable 'number of glaciers' and the buffer method error analysis.
We would like to thank Sam Herreid for the careful and constructive review!

Abstract
L23-24 "...national inventories have been used as a guide to compile a consistent update." How do several inventories generated from different countries by different analysists enable consistency?
Very good point, writing it this way is clearly misleading. It should mean 'consistent within each country' so with their earlier inventories. We have added this important point. As slightly different rules had been applied previously in each country, the new inventory is not alpine-wide consistent (with the related consequences for a comparison to 2003).

L23-27 It is odd that you develop the guiding datasets first and then explain methods to map glaciers from raw data. It's not perfectly clear what fraction of the new inventory is in fact new.
It is basically all new. The analysts only used the outlines of the previous inventories as a guide for the manual corrections. We have moved the 'Whereas' sentence forward to better reflect the logic of the workflow.

L24-25 You shouldn't have sentences beginning with "However" and "Whereas" back to back.
Agreed and changed.

L28-29 Is there a topographic result you can add to the abstract? This is the only sentence in the abstract referencing topographic information, and it doesn't explain why this is relevant to the study.
We have now added the median elevation and its dependence on location and aspect.

L57 Following the logic from L47-48 shouldn't this be ±3%?
Yes, this makes more sense and has been changed.

L59 Change "has" to "was"
Done.

L63 Change "representing" to "represents"

Done.

L72 "...in part to be compliant with the analysis in earlier inventors." Do you mean communicative? What are the other parts?
In part means that this is only one reason for incompliance. Apart from the mentioned requirement to be compliant with the interpretation of earlier national inventories, also the different interpretation due to the higher resolution sensor is a major issue. These are actually also the reasons for the missing detailed change assessment.

L90 What does "commissioning phase" mean?
It is related to the 3-6 month testing phase of the satellite with all its original settings. Image acquisition during this time is not performed according to the later nominal schedule.

L94 "...and distribute the raw outlines to the national experts for edition of wrongly classified regions" This sounds smart to use local knowledge, but to me it does not fit the description "consistent". To me a consistent method would be done by one person or one automated routine.
Yes, fully agreed. As mentioned above, we here refer to consistent on the national level. On the alpine-wide level this inventory is not consistent. However, we had several iterations of some outlines among the participating analysts that helped clarifying some major differences.

L96-97 "As a guide for the interpretation the analysts used the latest high-resolution inventory in each country" What does this mean exactly? Is the time stamp then some mix of 2015 and the range spanning the imagery used for the national inventories?
No, it just means that the outlines of the previous inventory have been used as a visual guide for the interpretation of debris-covered glacier parts and possible snow fields etc.

Study Region
L121 "With a total area of about 2000 km^2 in 2003" You say 2100 km^2 on L30 in the Abstract.
Yes, indeed. The value had been rounded to provide an easy to understand order of magnitude value to the overall mass loss (also the 1 m w.e. per year specific mass loss is rounded). We have clarified this in the revised ms.

L121-122 Please add a citation for "about 1 m w.e. . . . 2 Gt of ice per year."
Done (Zemp et al. 2015).

L126 "and the mean elevation is around 3000 m a.s.l., a unique value compared to other regions of the RGI." In which direction?
Both, up and down. It is the only region with the peak area distribution around 3000 m.

L130 ". . .many glaciers – large and small – become invisible under increasing amounts of debris" "Invisible" is imprecise language, "indistinguishable from optical data" is better.
It is not only from optical data, it also refers to field observations and geophysical methods. It can simply no longer be decided where a glacier is ending as everything might dissolve into a continuous ice-debris landform (e.g. ground ice, ice-cored moraines, rock glacier, peri-glacial forms). We have changed invisible to hidden.

L132 "...mapping their extent is increasingly challenging" Is this the same glaciers becoming more challenging or new glaciers becoming challenging?
Both, the original ones and possible new ones. It just becomes an increasing mess.

Datasets

L149-150 “only the required bands, no longer possible” I don’t understand this.

The policy of distributing Sentinel-2 scenes has changed, so that individual image bands could no longer be downloaded from this source. However, as we have done so, we think it is important to be precise here.

L161-164 I really like this level of detail you provide. Here you make clearer what “commissioning phase” means (question to L90).

Thank you. We thought it would be good to already mention it shortly in the introduction.

L193 Change “using” to “to use”

Done.

L194-195 “. . .due to the locally poor geolocation of the S2 scenes . . . the location of the ice divides was [change to: were] partly manually adjusted” Are you saying you moved correctly geolocated flow divided to agree with an erroneously geolocated S2 image? Please be clear what this implies about your new dataset (including the magnitude of these adjustments) and how you motivated your choices.

Yes, this was required but only locally, i.e. for a few dozen glaciers. The shifts imply that the new divides can only be applied to these Sentinel-2 scenes. Apart from this, it guarantees consistency with the glacier divides used in the previous inventories. This information has now been added.

Methods

L201-207 This is confusing to follow.

We agree and have rewritten it. It is now less condensed but hopefully clearer.

L208 “We followed the recommendation to select the threshold in a way that good mapping results in regions with shadow are achieved.” Where is this recommendation? Can you please add the threshold values used to Table 1 and cite that here?

It has been written in several publications. We have now cited Paul et al. (2015) as a more recent study. Theoretically we can provide most of the threshold values used, but practically this would be of limited help as digital numbers were up-scaled from 12 to 16 bit afterwards and a different and more complex pre-processing had been applied to glaciers in Austria. The effect of different threshold values (for th1 and th2) on the classification is illustrated in Fig. 3 and can be used as a guide for other regions.

L213 You mention several times “misclassified rock in shadow” which I believe means the dark rock becomes darker but is then classified as glacier, I’m confused by this.

It means that the rock in shadow is classified as ice because its value in the ratio image is the same as for dark glacier ice in shadow. This can happen at the edge of spectral separability when the SWIR band reflectance is becoming very small. We have added this for clarification.

L222 “contrast stretched” Is this just referring to how you assigned display colors? Probably unneeded information if so.

Yes, it means we have increased the brightness of the image for better display. Although a minor detail, we think this information is at least relevant for correctness.

L235 “All pre-processed scenes were provided in their original geometry for correction” This doesn’t make sense to me. What is the geometry of a pre-processed satellite image?

The Alps cover three UTM zones (31-33) and we have not re-projected them to a common geometry (UTM32) beforehand but provided them as is. The pre-processing only refers to the threshold selection for best mapping results.

L238 Dark bare ice (that you reference in the sentence before) is also an omission error
Yes, this is correct. We have added 'dark bare ice in shadow' to omission errors and written more precisely 'rock in shadow' for commission errors.

L239 And shadowed rock classified as glacier is also a commission error. Meaning, you're not wrong but it's slightly odd to not give these aforementioned quantities the same classification.
Yes, see above.

L240 I know this is a glacier definition interpretation question, but I really cannot imagine a 0.01 km^2 "glacier" internally deforming. The argument for including these patches is weak, in my opinion.
Fully agreed, but it stems from the original UNESCO classification for the world glacier inventory (WGI) that also included' perennial snow and ice', i.e. what we would call firn but might not densify to (and flow like) a glacier. The reason for including such features was related to the hydrological background of the WGI, i.e. it should contain all 'frozen water bodies on land'. As in steep terrain such small entities can indeed be 'real' glaciers (i.e. flowing and have a mass balance), we keep them here. For the wider discussion of this somewhat delicate topic we have now added the recent study by Leigh et al. (2019).

L246 "(see Paul et al. 2016)" do you mean "following Paul et al."?
Yes, we have now written '(cf. Paul et al. 2016)'.

L253 Regarding sub meter resolution data to inform debris cover delineation, I agree this can be very helpful, but I think it can also be less helpful. In some instances it is simply not clear, even if you were standing in the field, in these cases I think the high resolution images give us analysts a false sense of confidence. I would even argue that in some cases a lower 10-15 m resolution helps flow features stand out which is probably a better guide to finding the glacier margin than images that can resolve individual clasts. You make a similar counter argument to mine later at L321-323.
We fully agree but think that it really depends on the specific glaciers. Sometimes images with 10 m resolution provide the required details to identify such regions, and sometimes there is no benefit of the improved spatial resolution. The general issue that one can see different things at different image resolution is a general problem of more recent inventories and also a reason why the comparison to the 30 m resolution inventory from 2003 is problematic. Figure 4d illustrates a part of this interpretation variability for a difficult case.

L254 "we illustrate the strong glacier shrinkage from 1998" Strong relative to what? And I believe all citations so far have been to the 2003 former outlines, please add a citation for these 1998 outlines.
The strong shrinkage is from 1998 to 2016. The citation for the 1998 dataset is given in L249.

L265 "changes. . .are important." Hard to constrain importance, I would say "notable" or "visible"
Agreed and changed to 'visible'.

L270-284 The relevance of these two paragraphs is not clear to me. What information does your reader need from this to understand your study? I think it can be said much more concisely.

This is basically a description of the deviating glacier mapping conditions in Italy and how the additional scene selection has been handled. We think this is required to be transparent. However, we agree that the two sections can be shortened and have condensed the two sections to one.

L289 Can you please let us know how common these shadow errors were? In kmˆ2 preferably but also qualitative terms could be okay. I think it is useful information for others who will use this work as a guide for their own glacier inventories.

The regions in shadow have not been measured explicitly but this replacement only affects a couple (31) of very small glaciers (the largest is 0.13 km^2) existing in topographically favourable conditions (collectively they cover only 1.35 km^2). They do thus only show very small changes over decades and the 'old' outlines are likely more precise than digitizing something that is barely visible on the satellite image. We have added this explanation. We also cite now the study by Fischer et al. (2014), who have investigated the benefits of very high-resolution imagery for glacier mapping when it comes to such small entities.

L293 Do you mean "[Italian] alps"? also I don't think "i.e." is correct if you list the full set of 3.

Yes, corrected to 'Italian Alps'. When 'i.e.' means 'that is' (?) it is correct. Using e.g. ('for example') would be wrong.

L295 If you go into sub-region detail it would be helpful to have these regions labeled in Figure 1 and possibly summarized in a table.

We have now added the respective tile number (14) from Fig. 1. We provide some more details for Italy as the region is rather extended and scene selection was more complex than for the other three countries.

L300-304 I think this is an appropriate place to mention rock glaciers and either cite others regarding their definition/interpretation or use these data and possibly change since 2003 to make your own inference on this classification. Were you sufficiently convinced what you show in Figure 4c is in fact glacier? You defer to the previous inventory, but you clearly demonstrate a level of expertise in this subject and I think the readers will appreciate a more decisive call on these small, difficult glaciers.

The shown close-up is indeed a region on the edge. Whereas sharp ridges indicate that there should have been a glacier before, what remained in 2007 and also 2016 might no longer be classified as a glacier. The satellite image of the two larger glaciers (nr. 529 and 530 in the last Italian inventory) basically shows a debris-covered slope whereas the aerial image shows a trough filled with avalanche snow (seasonal snow as mentioned in the text). There might still be ice underneath, but calling this a glacier would likely only be justified when knowing that there has been a glacier before. For this inventory it is still included, but on page 179 of the book in the last Italian inventory (see http://sites.unimi.it/glaciol/wp-content/uploads/2019/02/5-Lombardia.pdf) it is written for both glaciers 'Almost totally debris-covered; in the CGI inventory it is labelled as "extinct" '. Hence, this is a very difficult case. We have now cited the study by Leigh et al. (2019) and added a short comment.

L317 I think something like "completely ablated" is more precise than "disappeared" but maybe only stylistic.

See above, disappeared should also include that they might still be there but that they are out of sight (e.g. buried under debris from rock-fall). In this case 'completely ablated' would be misleading, in particular when we do not know what has happened process-wise.

L325-326 "This glacier has thus strongly grown since 2003 due to a new interpretation..." This is incorrect language, change to "the interpreted glacier area strongly grew"
Fully agreed, it is not the glacier that has grown but the delineated extent (changed).

L336-339 I agree with your statement that political boundaries are meaningless in a scientific context, yet you go on to present your results per country. I understand that the source of some degree of your results are from national inventories, but my sense was this article is a pure research stitching together of these datasets. What is your motivation to partition your results by country?
It is basically only because we asked the national experts to do this. Despite the tough work, the work was not specially funded for some countries but cross-funded from other projects. Some national results thus help to highlight the contribution of the partners.

L341 I don't think "digital" is needed.
We agree that the statement would also work without the 'digital', but think that not all readers are familiar with GIS-based data processing. So for some readers this might be relevant information and we would thus like to keep it.

L360-362 Please provide the specific rational as to why a glacier specific comparison was not possible between glaciers that met the "point in polygon" check. One thing that makes this article so unique is that the 2003 inventory was made by you. This is the interesting point of unique consistency that you bring with this study but here it seems like you deflect from this. Has your personal definition of a glacier changed so much since the earlier inventory? There is a strange human element to this line of work/this study, which you later spend considerable effort attempting to constrain, yet this change in interpretation/definition of a glacier between the 2003 inventory and this one, with the same lead author, is surprising and interesting to me.
This might indeed come across a little bit strange, but there are a couple of reasons for it:
1. The interpretation is indeed partly different due to the better resolution of Sentinel-2. This led to small changes for some glaciers that partly resulted in a change of topology (e.g. two units are now connected or have split into several parts). Such changes can only be considered manually and this would be a considerable effort.
2. The 2003 inventory was also created by three different persons that have - despite all cross-checks - slightly different interpretations. It would be possible to directly compare extents on a glacier-by-glacier basis for the part that the first author has done, but this would only be a subsample of all glaciers and still suffer from point 1.
3. For Austria, the differences in interpretation are very large (see Fig. 10) and a direct comparison with the 2003 outlines thus not possible. The differences in Austria are mainly caused by (i) the improved spatial resolution of the Sentinel-2 images and (ii) by considering the previous national inventories generated from orthophotos and lidar data (Lambrecht and Kuhn 2007, Fischer et al. 2015) for consistency checks with the new inventory, whereas the 2003 outlines were derived from 30 m Landsat images without using existing datasets for consistency checks.

L382 ". . .we applied the buffer method. . ." Please add a citation for this
Added (Paul et al. 2017)

L390 I don't understand how it is computationally expensive to buffer the debris-covered areas (a smaller area than total glacier area) while it is computationally feasible to buffer all areas. I have applied buffers to all glaciers and debris cover on Earth on a typical laptop without outstanding computational demand. Is there a step or condition that causes the high computational cost?

There are at least three reasons for it (cf. Mölg et al. 2018): First, we do not get the debris-covered regions automatically (e.g. from a subtraction of the clean ice part from the corrected extents) as we have also corrected other regions manually (e.g. shadow, supra-glacial lakes). Second, application of the buffer method to a given set of outlines (polygons) is a very quick thing, fully agreed. But for the uncertainty assessment it is required to only use the outer perimeter of the debris (as it does not matter for debris enclosed by ice) and identifying this part of the perimeter is quite an effort as this is in general numerous bits and pieces rather than one closed polygon. Finally, there are different types of debris covered glacier areas. In some cases the debris boundary is clearly defined and can be accurately traced, sometimes it is rather a guess based on the analyst's knowledge (see Figs.4d and 11). Hence, we have decided to not perform this (statistical) assessment here but focussed on the multiple digitizing.

L391-392 I think Figure 4d is a good example of where this assumed realistic uncertainty estimate of ±2 pixels is not applicable. You say this depends on the degree of debris coverage which I agree with, could you apply a different buffer as a function of debris-covered area? Maybe it could help keep your above referenced computational cost low if only considering the very debris-covered glaciers.

Thank you for this suggestion. As mentioned above, apart from the degree of debris cover it is also the topological complexity of the locations with debris cover that increases computational cost. We have also no automated method to determine the 'very debris-covered glaciers'. The glacier to the NE of Glatscher da Gavirolas (named Glatscher da Fluaz) in Fig. 4d is also very challenging to digitize and can be about 100% larger than in our interpretation as the valley floor is to a considerable extent filled with debris-covered ice. In other words, the result obtained with a more complex buffer method (giving an uncertainty between 5 and 10%) would not reflect the real problem. We have now better described these uncertainties.

L396-397 Please watch your significant figures, the sum your report is slightly off. Further, do you have confidence in your results in Austria to 0.01 km^2?

The difference is due to the rounding of values, it is just 0.1 km^2. As discussed above, the 0.01 km^2 threshold is certainly the lowest sensible size for a glacier. This value is also valid for Austria, but here some entities being much larger (say 0.1 km^2) are included where we can certainly discuss if these are glaciers or perennial snow/ice features (e.g. persistent avalanche deposits). So it is less the minimum size that is problematic and more the different rules applied in each country to include certain features or not.

L399-402 I do not see the value in 'number of glaciers' based results. I am not convinced a different research group repeating your study will fine the same number of glaciers. If you disagree with this argument, I would at least suggest a minimum area of 1 km^2 to promote some degree of repeatability.

Even at that size it will likely be difficult due to some topological differences (see above L360). However, all values can be taken from Table 2. Apart from this, we 100% agree that the number of glaciers is science-wise the least relevant and has a very limited meaning. On the other hand, it is the most reported (e.g. the 215 000 glaciers in the RGI are widely known, but the total area?). So counting entities is highly important for communication, even if the number makes very little sense.

L415 “glaciers smaller than 1 km^2 can be found at all elevations” I see what you are trying to say but this statement is incorrect.
Do you mean that these smaller glaciers cover ‘nearly the full range of possible elevations’ rather than ‘all elevation’? This is fully agreed and changed accordingly.

L415-416 “indicating that their mean elevation [remove: does] only slightly depend [depends] on climate factors” This doesn’t make sense to me. I think you are trying to obliquely address the question: why aren’t some of these small glaciers bigger? I think there might be a statement you can make here but as it is I don’t think there is supporting evidence.
It simply means when small (and thus thin) glaciers are also found at low elevations (where temperatures are higher), they must be protected from the direct climate forcing (e.g. due to shadow, debris, avalanches, see above). Or in other words, their existence only slightly depends on climate factors (still they need some snow for nourishment). We can discuss if these entities can be called glaciers (e.g. do they flow or is it just a perennial snow/firn accumulation), but in most cases field investigations might be required to decide it (so for the time being we keep them included). As reviewer #1 had a similar comment, we have added a statement for clarification.

L417 “arrange around a climatically driven mean elevation which is around 3000 m a.s.l.” Why not compute the mean and add a fit mean line to Figure 6a? “...the largest glaciers are not those with the highest elevation range. . .” Yes, I actually think they are and your next statement contradict your previous statement “. . .and for the majority of glaciers the elevation range increases with glacier size” One very steep, small glacier is an outlier, not relation breaking evidence.
First point: When mean elevation is plotted against area this would basically give a straight line as values are normally distributed around the mean. So this does not make for an interesting plot and can also be described by a short text.
Second point: There are at least five glaciers with a higher elevation range than three larger ones. In this case it is a characteristic of the Mt. Blanc mountain range that is very high and very steep and does not allow building large glaciers. We would not mark it as an outlier as it applies to several glaciers and as the Mt. Blanc region is a regular part of the Alps.

L422 “This is typical for regions dominated by mountain and valley glaciers as these follow the given topography” This statement needs a citation and an example of regions where glaciers (or ice sheets) do not follow the given topography.
As mentioned in the following example, Plaine Morte glacier is an exception from the rule as it is a flat plateau glacier. This type and ice caps / ice fields do not follow this rule. We are not sure if there is a citation for this as it just follows from the geometry.

L423-424 “an exception from the rule” What rule? I would call it an outlier
The rule would be: The larger the glacier, the larger its elevation range, and we have now added this for clarification. This rule only applies to mountain and valley glaciers rather than plateau glaciers (or ice caps / ice fields).

L428 Can you please motivate your use of median here versus mean at L415.
There is no physically based motivation as all three elevations (mean, median und mid-point) carry about the same information. The non-physical motivation is that median elevations have been used widely to display the spatial variability.

L428 “largely driven by temperature, precipitation and radiation” Since you don’t specify glacier size this is in contradiction to L415-416.

We agree that this seems to be a contradiction at first glance, but the statement here refers to 'usual' glaciers rather than those existing under special conditions. The latter results in small glaciers at low elevations so that their (mean or median) elevations are also about normally distributed. There is thus no dependence of (mean or median) elevations on glacier size. To remove the large scatter of the small glaciers normal distribution and see the relevant signal, the map in Fig. 7 only shows glaciers larger than 0.5 km^2.

L428 Wouldn't topography be an important factor here?
Topography affects mass balance (and thus location) basically via the mentioned radiation receipt and (possibly increased) snowfall amounts. So topography is implicitly considered. However, to be clear we have now also added topography (in brackets).

L429 "temperature is rather similar at the same elevation over large regions" Needs citation
We hoped this is common knowledge but found a study that has now been cited (Zemp et al. 2007).

L431 Remove "amounts"
Done.

L434-435 Why are glaciers larger than 0.5 kmˆ2 les impacted by local topographic conditions?
The 0.5 km^2 threshold is arbitrary and has been selected to create a good display of all values in Fig. 7 (the inset shows the full sample). Otherwise, the portion of impacts from topography (e.g. shadow) decreases with increasing glacier size (for mountain and valley glaciers, not valid for ice caps).

L435 ". . .median elevations (around 2400 m a.s.l.)" L418 says 3000 m a.s.l.
We here refer to approximated minimum and maximum median elevations related to their dependence on location (Fig. 7) and mean aspect (Fig. 7, inset) rather than to the mean value for all glaciers.

L438 "b" not labeled in Figure 7
Indeed, thank you for spotting (changed now to Fig. 7, inset).

L439 "On average, glaciers. . .have median elevations that are about 400 m higher" Is this a mean of medians? I'm a little lost how you choose distribution middle metrics.
There are indeed quite a lot. Here we have not calculated mean values per aspect sector but just visually averaged the point clouds depicted in the inset plot of Fig. 7.

L440 "However, the scatter is high" It's hard to understand a "result" that is then walked back some qualitative amount. Can you draw on statistical tests to qualitatively inform signal form noise?
Also this value is only derived from a visual analysis of the scatter plot. There should not be a signal from noise separation as both values are equally relevant without impacting on each other. Indeed, the scatter plot is just reporting that there is some impact of radiation receipt left, but there is also a high variability for each aspect direction as spatially highly variable precipitation amounts exert a strong influence on absolute values. As a back-up for the visual analysis we have now calculated aspect-sector averages and standard deviations of mean elevation and added the related values to the plot.

L445: I am again missing the scientific argument for presenting result per country.
There is no scientific argument for it, it is more a political decision that 'comes with the territory' when national experts are contributing.

L456 “For a selection of 2873 glaciers present in both inventories. . .” According to this, the number of glaciers in 2003 was 65% of the number of glaciers now. I find this more concerning than a result. I am inclined to assume this new inventory is overly generous with a “glacier” classification and I would strongly encourage the authors to double check their confidence on what they are calling glaciers.

As mentioned above, people (incl. reviewers) are highly interested in the number of glaciers, even if this is a largely meaningless value ;) We have now written ‘for 2873 comparable polygon entities’ instead of ‘glaciers present in both inventories’ as this has also a topological component. For the most part, the excluded glaciers are very small and relate to not mapped entities in NE Italy (in 2003) and partly much larger glaciers in Austria due to another definition of glaciers (see Fig. 10). We have accepted this deviation for this study to include the work of the Austrian colleagues, but otherwise we agree that it also creates some trouble when going into details.

L460-461 Per country results should probably be in a table if you report some here.

This could be done but it would be biased by the different size-class distribution in each country as relative area changes have a size dependence. Strictly, it would be required to do this size-class dependent and this is somewhat beyond the scope of this inventory presentation (and would require a considerable extra effort as glaciers split and change size classes etc.). So if you don’t mind we would prefer going ahead with the aggregate numbers.

L462 “Reveals [should be: a] small shift” Can you please say more about this translation error, it doesn’t look linear(?) is it elsewhere? Should this be corrected?

It was a matter of big discussion with space agencies as the 90 m DEM they used for orthorectification of the Sentinel-2 scenes was not accurate enough, creating irregular pixel displacements. Additional, the set of ground control points (GCPs) used had some issues resulting in a small (about 30 m) more systematic shift. This caused the problems with applying the drainage divides used for the 2003 inventory also to the Sentinel-2 scenes. As far as we know, the problem with the DEM has been solved in the meantime but we had to use the images from the commissioning phase at that time.

L466-468 As stated above I’m concerned there is not a way to make a per glacier area change plot. In my opinion, the new outlines (yellow) in Figure 10 have a lot of unrealistic area: too narrow, too small, not following a flow pattern. For the glaciers that do intersect 2003 it is clear the outlines are of very high quality, but I think the 2003 interpretation of what is a glacier is in some cases better.

We agree to the latter but as explained above the Austrian inventory had included these ‘perennial’ features for consistency with earlier inventories. Most of them can be seen as water resources and have thus a right to be included, but many of them might not be glaciers in a classical sense (with flow, accumulation/ablation etc.). On the other hand, when carefully analysing the very high-resolution image available in Google Earth (from 13. Aug. 2015) for the region shown in Fig. 10 one can also argue that many of these features indeed consist of glacier ice and have been missed in the 2003 inventory. One can thus also argue that the 2003 inventory underestimated glacier area rather than the new one overestimating it. Both have their shortcomings (as illustrated in L476ff) and a related scatterplot would be dominated by the interpretation differences rather than real area changes. We think this is not meaningful. So apart from the differences in interpretation, here is also a warning that inventories that have been derived from sensors of different resolution can be incomparable. This problem has also been raised by Fisher et al. (2014) and Leigh et al. (2019) that we have added here. In short: Instead of forcing the comparison to work, we would prefer stating here that it can not work.

L478 “with a limited meaning on the basis of individual glaciers” These differences that disable a per glacier change analysis aggregate to the whole, I’m not sure how you can consider changes of the whole Alps or per country and not be able to consider individual glaciers.

As mentioned above, in part the per-glacier assessment has not been performed because automated methods cannot be applied due to the topological changes and the fact that ID assignments had to be performed manually. These difficulties can be avoided when only referring to aggregate values such as for Fig. 8 or the total. However, there is likely no error compensation but a ‘too small’ bias for 2003 and a ‘too large’ bias for 2015/16. Considering these biases, the relative area change rates will increase (see next point).

L480 “-15%” This value starts at -13.2% earlier in the manuscript, was bumped up to 15% at L459 and now is “even higher”. For the results section I think it’s best to present a consistent and confident value.

The -13.2% is only referring to the subsample of 2873 glaciers, the other values given are extrapolations that consider possible biases in the dataset (see point before). At first we consider the underestimation of area in the 2003 dataset (e.g. missing debris covered parts and missed small glacier), giving a larger 2003 area and thus a higher loss. In the next step we consider that with a different interpretation (e.g. perennial snow and ice excluded) the 2015/16 extent would be even smaller. This would further increase the area loss (exceeding 15%). So this is just an attempt to compensate for known biases in both datasets to present a more realistic value.

L493 “a detailed analysis” Where is this analysis? Or do you mean the overlay Figure only?

It just refers to Fig. 11. In the GIS we have zoomed in to the individual regions (taking also some of the outlines off) to better see where the differences in interpretation are located. All of these regions are visible in Fig. 11.

L495 “When excluding P1” what if P1 is the most trained expert?

Than all other participants are too conservative in their interpretation and P1 has to explain why the additionally considered regions under debris cover should belong to the glacier. But the text as written would still be correct.

L505 “digitized glacier extents increased by several per cent after consultation of very high-resolution satellite images” Is there a way this information could be added to Table 3?

This would basically be four additional tables showing lots of numbers. We can add them in a supplement but think they do not add much. Showing how the outlines have changed after the consultation would likely be more relevant. We have prepared a couple of examples for a supplement illustrating the change in interpretation and some critical regions at very high resolution.

L510-513 “If such regions have to be included...can be discussed...including these features in a glacier inventory or not is a (personal) methodological decision” This belongs in the discussion section.

We agree, but think it can also be mentioned here as otherwise we need to repeat the entire context for this statement in the discussion before making it.

L522-524 Check English here.

The sentence has been revised. (Now: “The comparison of topographic parameters (minimum, maximum and mean elevation, mean slope and aspect) as derived from the TDX and AW3D30 DEM revealed larger differences, in particular towards smaller glaciers.”)

L528 re-state acquisition gap dates.
Added.

L530 5 years of elevation change are within the uncertainty range? I would expect there to be a clear signal.
We have not corrected TDX for radar penetration and do not know the exact years of image acquisition as both datasets are merged products from several years. If the scenes over the Alps are from 2010/11 for AW3D30 and 2012/13 for TDX, we do not expect to see a clear elevation change signal.

Discussion
L535 missing citation here
We cite now Pfeffer et al. (2014) who show related histograms.

L536-537 Probably belongs in results section
We agree but think this is only data already presented in the results (though slightly differently aggregated).

L538-539 "However, for consistency with earlier inventories they have been included" Figure 10 makes it clear this is not true, consistency with earlier inventories would exclude non-glacier area. I think it's concerning that the authors walk back what is considered a glacier here in the discussion, in my opinion, we need this set of experts to make these (hard) decisions so the rest of the community can uses this product with confidence.
To the first point: We have changed the sentence to "with earlier national inventories" to be clear we are here not referring to the 2003 dataset.
To the second point: Fully agreed, it would be good when we can reach consistency here. The currently best suggestion we have on the table is that analysts can include in their (national) inventories whatever they want, but there should be a mark in the attribute table what the polygon is describing (e.g. glacier, perennial snow/firn, rock glacier, regenerated glacier, includes dead ice, etc.). This would allow the user of the datasets to later select them as required for a specific application. We are not yet there but something like this will hopefully be established soon.

L540 "precipitation amounts have a limited impact" I think this a contradiction to L431.
In L431 we talk about the impact of precipitation amounts on glacier location for those glaciers that do not benefit from special topographic conditions (e.g. the larger ones). Here we talk about the other 'glaciers', those found in topographically preferred locations that can be found at any elevation. There is thus only a limited impact of precipitation amounts.

L543-544 I don't think Figure 7 supports these claims.
The Fig. 7 inset does indeed not show glacier size but the map has been restricted to the larger glaciers. So 'modified by glacier location with respect to precipitation sources' is correct and 'modified by mean aspect' is also correct. The sentence has been rewritten to disentangle both observations. (Now: "Glacier mean elevation does not depend on glacier size but on glacier location with respect to precipitation sources, in particular for larger glaciers (Fig. 7). On top of this dependence is the variability with mean aspect (Fig. 7, inset)."

L546 "Widespread glacier thinning" L530 suggests no thinning
We here refer to a much longer time period (last 2-3 decades) than for the DEM comparison (2-5 years). (Now: "Widespread glacier thinning over the past decades and …")

L546-547 confusing wording
Rewritten in two sentences to be clearer.

L548 Confusing English in this sentence
Rewritten. (Now: "These separated parts can thus not be named 'regenerated glaciers' but they melt away as dead ice.")

L551-552 "merged their IDs. . .combined extent" This is unclear
Despite being two separated polygons, both have a common ID so that all topographic information is calculated for the combined ice masses rather than the individual polygons. (Now: … and used the same ID for both parts to obtain topographic information for the combined extent.")

L558-559 As stated above, interpretation/definition errors don't disappear at a wider scale.
Yes, this is correct but it allows automated computation as topology issues can be neglected. For this reason we have added the two further extrapolations that consider the interpretation biases (possibly too small/large glaciers in 2003/2016, respectively).

L565 Snow covering 20% to 30%, where did these numbers come from, I don't recall anything in the methods.
This has not been explicitly calculated but can be seen on the cited Figs. 9 and 11 (i.e. it is only a qualitative estimate).

L571 "change rates of identical size classes are compared" Do you mean hypsometry?
No, mean area change rates per size class (e.g. as shown in Table 2 of Paul et al. 2004).

L573 "revealed a large variability in the interpretation of debris-coved glaciers" Did it? I'm not sure if this was quantified.
It can be seen in Fig.11: All regions where the differently coloured lines are not on top of each other (so that only one colour is visible) are debris-covered glaciers. The larger the spread, the more difficult the interpretation has been.

Conclusions
L609 "DEM quality and co-registration" were these mentioned in the text?
Indirectly. In L246 we refer to the further processing that is described in detail by Paul et al. (2016).

Tables and figures
Table 3
Am I reading correctly that STD derived for n=4? Is glacier ID 4 two glaciers in the inventory? This sounds like an interesting result that most participants called in one. Is this a unique case, or common?
First point: Yes, the STD is for n=4.
Second point: For the sample of glaciers we had selected it was only this one. The split into two glaciers had been made based on the Sentinel-2 image only. During the digitizing and checking back with the higher resolution imagery it was revealed that both glaciers are still connected under debris cover. So when two parts had been digitized, both parts have been summed up.

Figure 3
The blue is a little misleading, possibly better as 4 sets of colored lines It's a little hard to tell what is error and what is true.

We can exchange the blue with another colour but actually the colours only indicate what is mapped with the related thresholds. There is yet no true or fase included. The blue colour is depicting (in b and c) the most conservative mapping. When lowering the respective thresholds, the yellow, red and green parts are additionally mapped. Panel d shows the consequence of the selected threshold th2=860, correct mapping of rock outcrops at the white arrow, and at the same time missed ice in shadow at the green arrow.

Figure 4
Are blue in a) and green in b) the same? Why different colors?
If blue in a) means light grey, yes these are the raw mapping results for both regions. Colours are different because analysts have chosen the examples for visualization independently. Not perfect but maybe acceptable.

Figure 9 and 10 Many of the small glaciers identified in these figures do not look like glaciers to me.
Fully agreed, at this scale they are hard to identify and some of the extensions might indeed only be perennial snow/firn. However, please check for both also the very high-resolution imagery available in Google Earth. Suddenly real glaciers materialize ...