Reviewer 1

Dear authors,

Thank you for the interesting work combining avalanche detection with glacier mass balance. I agree with you that remote sensing has great potential for avalanche identification that goes beyond what has been done so far. However, I have identified several points that need to be addressed before publication.

We would like to thank Reviewer 1 for his/her high quality review, and his/her very relevant and constructive comments.

Below you find a collection of the most relevant and general comments followed by specific comments referring to lines in your manuscript:

- In your processing chain for the Sentinel-1 data you state that you have averaged the polarization modes VV and VH. Various studies however have observed a difference in backscatter based on the polarization of around 5 dB (e.g., Liu et al., 2022). Why did you still choose to treat the two polarizations the same and did you check if treating polarizations differently would affect your results? (see also comment to 147ff)

  We averaged the VV and VH polarizations in order to reduce the radar speckle, as has been done in previous studies (e.g. Leinss et al., 2020). However, you are correct in the sense that VV and VH have a difference in backscatter and therefore should in theory be treated separately. We had conducted some initial tests to check if it really made a difference for the automated mapping of avalanches, but results were inconclusive. Some studies have indicated that VH polarization is more suited for avalanche detection (Hafner et al., 2021), but this difference is likely most important for low incidence angles (e.g. Fig. 7 of Tompkin and Leinss, 2021), which were regions that were in part masked out when applying our shadow/layover and brightness masks (Section 3.1). When calibrating our method separately to the VV and VH RGB images for the Mt Blanc massif for the period 2019-2020, we find generally lower values than those obtained with the averaged polarizations (Table R1). This indicates that our approach is better suited to the combined VV and VH polarizations. We will now mention this interesting point in the discussion section and add Table R1 as Table S7 to the SI:

  'Future method developments could also benefit from separating the VV and VH polarizations, particularly for regions of the SAR images with low incidence angles (Tompkin and Leinss, 2021). While in our case we obtained better results by averaging the two (Table S7), other machine learning-based approaches would likely benefit from the additional information provided by the two polarizations (Liu et al., 2022).'

  Table R1: F1-scores obtained for the calibration of our method to VV and VH RGB triplets for the period 2019-2020 over the Mt. Blanc massif.

Polarisation	*Path*	*Season*	*F1-score calibration*

VV	Descending	November-April	0.29
		May-October	0.40
	Ascending	November-April	0.47
		May-October	0.31
VH	Descending	November-April	0.17
		May-October	0.18
	Ascending	November-April	0.14
		May-October	0.30
(VV+VH)/2	Descending	November-April	0.56
		May-October	0.56
	Ascending	November-April	0.54
		May-October	0.49

- The (description of the) process of manual avalanche mapping for parameter estimation followed by automatic detection but then again extensive manual corrections is confusing. How good was your automatic detection, why did you need to manually correct in the first place? How much time is gained by including the automatic mapping compared to going fully manual? Additionally, it is not always clear from which processing step the results presented in the result section came from.

We apologise for the lack of clarity in the overall description of our approach. It is important to note that developing a mapping approach for avalanches was only one of the focus points of this study, and that our main objective was the study of the spatio-temporal variability of these events for the three studied regions, as mentioned in the introduction (L96-99):

'To this end we (1) calibrate and evaluate our automated mapping approach at each site and assess its transferability to other sites, (2) extract the size-frequency characteristics of avalanches at various spatial scales over a period of five years and (3) evaluate the implications for the glacier mass balance.'

While the fully automated approach did lead to relatively high scores, as described in sections 4.1.2 (Table 2) and 4.1.3, there were still a number of false positive and false negative detections, which we feared may influence the final results related to the spatio-temporal variability. That is why the manual refinement was applied. Considering that 36% of the avalanche area was removed and 41% was added in this second manual step which mostly affected the descending and May-October scenes, but also that the mapping is considerably simplified when the regions of interests are already partly mapped, we consider that applying the automated approach first still reduced the mapping time by at least half (more for morning and winter scenes, which required less manual updates).

We will make sure to clarify this point in the discussion section:

'Our automated mapping therefore still requires manual edits, although we consider that applying the automated mapping approach and then updating the outlines by hand has reduced the mapping time by at least half relative to a fully manual mapping, more if only morning scenes were to be considered.'

We will also make sure to explicitly state what set of outlines is used in the different results sections:

'Here, we first compare our manually derived outlines with high-resolution Pléiades images and evaluate the results and transferability of the automated mapping approach (Section 4.1). We then use the manually updated set of outlines to obtain the characteristics of avalanche deposits (Section 4.2) and their spatio-temporal variability (Section 4.3) for all three survey domains.'

Furthermore, in the methods we will put the paragraph on the application of the mapping to the entire Sentinel-1 time series after the paragraph on the comparison with Pléiades, and explicitly state the reason for the manual editing:

'We required highly accurate maps of avalanche deposits for the analysis of their spatio-temporal characteristics of avalanche deposits.'

- There is a lack of clarity in your discussion regarding the detectability of dry snow avalanches (line 440ff, 456ff). You state that "the performance of such approaches is generally very good under dry snow conditions […] most difficulties for periods with wet snow conditions", this contradicts (other) previous research (e.g., Eckerstorfer et al., 2022, Abermann et al., 2019). But you also contradict yourself by writing later that "cold, low density snow […] is likely to be missed by this method, which likely also explains the upper elevation limits to avalanche detections, especially during the cold season." Please carefully reexamine the passages and literature concerning this topic and adapt your discussion.

  These apparently contradictory statements refer to two different things:

  - When avalanches are manually detected in Sentinel-1 scenes in dry snow conditions, they are usually also well mapped by the automated approach, as indicated by the high F1-scores, and there are less false positive detections than in wet snow conditions.
  - However, when we make the comparison with Pléiades optical images (Fig. 5d), we can also note that cold snow avalanches are not always detectable in

the Sentinel-1 images, as they do not always change the roughness of the snowpack much, as mentioned in the discussion section 5.1 (L453-456):

'Our comparison of the Sentinel-1 with the Pléiades avalanche deposit outlines indicate that avalanches detected with Sentinel-1 are of relatively large size (>4000 m^2 deposits) with high surface roughness, which limits the detectability to avalanches with high enough snow temperatures to form granular deposits (Steinkogler et al., 2015), or which are formed from cohesive wind slabs (Fig. 5d) or that entrain rock or ice debris, for instance from serac falls (Fig. 5c)'

As such, these observations are in line with Eckerstorfer et al., (2019; 2022) and Abermann et al., (2019).

However, we do understand that this difference may not be obvious in the current version of the manuscript and will therefore move the paragraph concerning the comparison with the Pléiades images at the beginning of this discussion section, and highlight the fact that the discussion of the second paragraph only applies to the comparison of the automated versus manually delineated outlines in the Sentinel-1 images:

'The performance metrics obtained from our automated mapping approach compared to the manual detections in the Sentinel-1 outlines are similar to that of other studies following similar threshold-based approaches'

- Concerning the structure of your manuscript: There are 21 sections, subsections and subsubsections which make it hard for the reader to orient. Please try to simplify and keep an eye on descriptive headings for your sections.

  Such a structure and number of sections is consistent with other articles published in The Cryosphere. We felt that it was better to have more subsections with explicit titles rather than larger sections with less obvious structure. We do however agree that some of the headings could be more explicit. We will make the following changes:

    ○ '3.3 Parameter calibration' instead of 'Threshold calibration'
    ○ '4.1 Sentinel-1 avalanche mapping' instead of 'avalanche mapping'
    ○ '4.1.1 Comparison of Sentinel-1 and Pléiades manual detections' instead of 'Sentinel-1 avalanche mapping potential.'

- You have included many figures, please keep only those illustrating your main points/results (see also specific comments).

  In the specific comments reviewer 1 suggests removing Fig. 9a, Fig. 10, 12 and 13 from sections 4.2 and 4.3 of the results related to the characteristics and spatio-temporal variability of detected avalanche deposits, and questions the readability/relevance of some parts of figures 1, 4 and 5 which are more related to the evaluation of the mapping approach. For the latter, we do agree that some elements may be confusing and may impact the overall readability of the overall figure. As per our answers to the specific comments, we will therefore:

    ○ Move panels e and d of figure 1 to the SI

Regarding the removal from the main text of Fig. 9a, Fig. 10, 12 and 13, which describe the characteristics and spatio-temporal variability of detected avalanche deposits, we note that while a portion of our manuscript is meant to focus on the methodological aspects of avalanche mapping with Sentinel-1 images, the other main objective was the study of the spatio-temporal variability of these events for the three studied regions, as stated in the introduction (L96-99):

'To this end we (1) calibrate and evaluate our automated mapping approach at each site and assess its transferability to other sites, (2) extract the size-frequency characteristics of avalanches at various spatial scales over a period of five years and (3) evaluate the implications for the glacier mass balance.'

We feel that these figures are important elements highlighting these characteristics and spatio-temporal variability and therefore do illustrate our main points/results. However, we also understand the reviewer's concern related to the overall high number of figures in the manuscript and to the readability of figures 9-13. We will therefore:

- ○ Improve the readability of Fig. 9-13 following the reviewer's specific comments
- ○ Move Fig. 10 to the SI along with Fig. 4, which will bring the number of figures down from 13 to 11

- There is a lot of spoken instead of written language used, for example "this is even more true", "indeed", "ultimately". Please stick to scientific written language, especially for starting and connecting your sentences.

  We will replace 'This is even more true' with 'This is particularly the case' and remove or replace the adverbs 'ultimately' and 'indeed'.

- Lastly, you have a mix of methodology, discussion and results in those three sections. Could you please go over this again and try to disentangle. If you want to keep discussion with results you could also do a section "Results and Discussion".

  We will make sure to clearly separate results from discussions, following specific comments from both reviewers. Specifically, we will:

  - Re-organize and streamline section 4.1.1 to stick to the comparison between Pléiades and Sentinel-1 outlines
  - Remove sentences repeated from the methods in sections 4.2 and 4.3

Detailed comments with reference to line numbers:

39: You make it sound like rock(fall) contributes to mass balance. Debris cover may indirectly contribute by preventing surface melt, but since you want to help calibrate glacier mass models it is weird to mention rockfall with snow/ ice avalanches as if it's the same process. I assume they are not calibrated the same way in the models. Not being able to

differentiate between those mass movements in my opinion is a limitation of the chosen method (which should be mentioned in the discussion/limitations).

If we define the mass balance of a glacier as everything that comes in minus everything that comes out of a glacier, then we would argue that rockfalls do contribute to the glacier mass balance. It is true however that this is something very distinct from snow/ice avalanches (although you could imagine that these snow/ice avalanches may also transport with them some amount of rocks). We will therefore keep this term here but will add more details about the limitations that it poses in the discussion:

'It is also noteworthy that this mapping approach with Sentinel-1 will likely not differentiate large rockfalls on glaciers from snow avalanches, which could explain some of the activity in the summer and autumn in the Mt Blanc massif. '

70ff: The citations on previous avalanche work in the introduction do not account for most recent work while you mention most of the recent work in the discussion/436ff. Could you please try to paint a more complete picture of recent work in the introduction and not bring "new" work only in the discussion without mentioning it before.

We will make sure that the new work also appears in the introduction. Specifically we will:

- refer to the work by Sartori and Darbiri (2023) and Guiot et al. (2023), as suggested by Reviewer 2.
- refer to recent work on the use of machine learning approaches for the automated mapping of avalanches:

  'More recently a number of studies have also trained machine learning approaches to improve the mapping of avalanches (Tompkin and Leinss, 2021; Waldeland et al., 2018; Yang et al., 2020; Bianchi et al., 2021; Kapper et al., 2023; Liu et al., 2021), but they have been limited by the lack of large training datasets for this application.'

- directly refer to the work by Hafner et al. (2022, 2023) for the use of optical images to detect avalanches.

84: Could you please quickly explain how validity was proven and what challenges were found?

We will reformulate this sentence as follows:

'The validity of these approaches has been demonstrated by quantifying the overlap between outlines from Sentinel-1 images and those obtained from high-resolution optical and field observations (Leinss et al., 2020; Hafner et al., 2021).'

For the limitations, we refer to the end of the paragraph, where we wrote:

'There remain limitations to these approaches, especially as they fail to detect smaller events (<4000 m^2) or have a high rate of false detections in the case of transitions from wet to dry snow that also result in increasing the SAR backscatter (Eckerstorfer et al., 2019).'

101: To me it is unclear how large your three study areas are. Please specify.

The size of the study areas is indicated for both ascending and descending scenes in Figure 1 (see also response to comment below). We will make sure to specify this in the caption (see response to next comment).

Figure 1: Connected to 101: are the numbers in the upper right corner your added up area of interest? Or is it the percentage of the added up?

The percentage given corresponds to the portion of glacier area covered either by ascending or by descending scenes. We selected the Sentinel-1 images so that the ascending and descending scenes always covered the entire survey area. We will make sure to specify this in the caption:

'The numbers in the upper right corner indicate the total area of interest covered by the ascending and descending scenes, respectively, and the third number indicates the percentage of glacierized area covered by ascending or descending scenes.'

Please add an overview map so that every reader can get an idea where your areas of interest are located. Additionally, you need to explain RGI 6.0 at first mention and as you cannot assume that every reader knows what it is.

Agreed, we will add an overview map. And we will specify here what the RGI 6.0 stands for and cite the source:

'Randolph Glacier Inventory (RGI) 6.0 outlines (RGI Consortium, 2017) are shown in black'

(d) and (e): these two graphs are quite hard to read, and I am not sure the content is essential for understanding your work. Maybe replace them with the overview map.

We will move these panels to the SI and put the overview map instead.

128: It would be interesting to know if all Pleiades imagery was taken on a day Sentinel-1 was acquired also, or (and if how long) the time gap was?

Agreed, we will specify this here:

'The winter and August Pléiades scenes were acquired on the same day as a Sentinel-1 acquisition, while the July scene was acquired two days before the nearest Sentinel-1 acquisition.'

147/172: Here you mention a 500m AND a 450m high pass filter? Did you filter twice? Or if these sentences describe the same operation why is the kernel size different? Additionally, a 450/500m kernel is quite large- did you perform an investigation of the effects of different filter kernel sizes on visibility of avalanches and results?

Apologies for this misleading part. These are indeed two different filters, but the 45 px one is actually a low pass filter (not a high pass, as we wrongly wrote). We will make sure to correct 'high pass' to 'low pass' there.

We did check the influence of kernel sizes on the visibility of avalanches. We actually needed our kernel to be at least of the same order of magnitude as the largest avalanche

events (up to 10^6 m^2), to be able to smooth out all avalanche deposits, even the largest ones. This will be specified in the text:

'We selected this kernel size to be able to smooth even the largest avalanche deposits'

This is consistent with the overall objective of this additional filtering step:

'The idea of this additional step was that an avalanche event results in a spatial discontinuity in the backscatter, if not with the image before, at least in the current image.'

147ff: You have averaged the polarization modes VV and VH. Various studies however have observed a difference in backscatter based on the polarization of around 5 dB (e.g., Liu et al., 2022). Furthermore, the different polarization modes do not provide information about the same objects (e.g., HV results in a higher noise floor, see also Howell et al., 2019) and a combination of this dual-polarization data should hence be treated as an index rather than a 'normal' dB scale image as suggested in Nagler et al. (2021). Why did you treat the two polarizations the same and did you check if treating polarizations differently affects your (manual and automatic) avalanche detections.

Please refer to our response to the corresponding general comment.

154ff: How much of your area of interest was masked out when excluding shadow and layover? Could you please specify how much of your area was at the end classified glacierized and considered for analyses (and consequently is the area you are referring to in the rest of your manuscript).

We will specify it here:

'As a result, 35%, 28% and 43% of the considered area was masked out for the Everest, Mt Blanc and Hispar regions, resulting in a total area available for mapping of 492, 140 and 762 km^2, respectively (Fig. 1).'

156: You cannot assume the reader to know what RGI 6.0 is, please explain this abbreviation.

We will specify here what this abbreviation stands for.

Figure 3(e): Given recent work comparing manually mapped outlines (Hafner et al., 2023), I am wondering if the blue outline is "more correct" and how dependent it is on the operator (see also comment to 224/186).

This is a valid question and we acknowledge that manual outlines will always hold some bias, no matter how expert the operator is. This is the reason why we compared the outlines from multiple independent operators to make a first estimate of the underlying uncertainties (Fig. S1-S2, Sections 3.3, 4.1.1). However, in the absence of ground truth, as is the case for these three regions, the manual outlines had to be used as a reference for the automated mapping approach. We do agree that this point deserves more attention and we will:

- mention in the results the F1-score of the different operators:

'The outlines of the operators had F1-scores between 0.54 and 0.66 relative to the outlines of the operator who derived the entire manual dataset for all three sites. We also directly compared the manual outlines from this operator with the consensus outlines from the other three operators, which were the outlines for which at least two operators agreed (Kneib et al., 2021).'

- add some explanations in the discussion and refer to the work by Hafner et al. (2023) on this topic:

'The performance metrics obtained from our automated mapping approach compared to the manual detections in the Sentinel-1 outlines, have a wide range of values (F1-score between 0.29 and 0.78) depending on the season and acquisition time. For most scenes, the F1-score was actually similar to those obtained by manual outlines from independent operators (Table S1, Hafner et al., 2023).'

Figure 4: What exactly do you want to show with this?

We found this plot interesting as it showed the sensitivity of the automated mapping to the threshold values used. We however agree that it is not a major figure of this manuscript and will move it to the SI.

224/186: Could you please give a number for the agreement between the experts (e.g., Intersection over Union, IoU) to understand how much uncertainty is introduced if one operator manually corrects predictions (see also comment to Figure 3).

We agree that this is a crucial point, and the results of this comparison are given in the Results section L277-285. We will however specify the F1-score (rather than IoU, to be in line with the metric used in the rest of the study), which was calculated in Table S1, there as well:

'The outlines of the operators had F1-scores between 0.54 and 0.66 relative to the outlines of the operator who derived the entire manual dataset for all three sites. We also directly compared the manual outlines from this operator with the consensus outlines from the other three operators, which were the outlines for which at least two operators agreed (Kneib et al., 2021).'

234: How did you determine a match between Pleiades and Sentinel-1? How much overlap was needed?

We are unsure whether you refer to a spatial or temporal overlap, so we answer for both:

Spatial overlap

For this comparison and all the calibration and validation of the methods, we directly made a pixel-by-pixel comparison, which meant that we did not need to consider the overlap of individual avalanche events. We will specify it here:

'We compared on a pixel-by-pixel basis the Sentinel-1 outlines that occurred over given periods in the summer and in the winter with manually derived outlines of avalanche

deposits from high resolution (0.5 m) Pléiades orthoimages over part of the Mt Blanc survey area'

And in Section 3.3:

'We used the F1-score, also known as the Dice coefficient, as a metric to quantify the goodness-of-fit of the automated delineation on a pixel-by-pixel basis (Dice, 1945)'

Temporal overlap:

To be able to compare deposits from a Sentinel-1 RGB pair with deposits from a Pléiades image, the Pléiades image needs to have been acquired as close as possible to the second Sentinel-1 image of the pair. This was the case for most Pléiades acquisitions, and we will specify this in the data section 2:

'The winter and August Pléiades scenes were acquired on the same day as a Sentinel-1 acquisition, while the July scene was acquired two days before the nearest Sentinel-1 acquisition.'

There is of course a risk that the deposits detected in the Pléiades image are anterior to the first image of the Sentinel-1 pair so for all the examples that we presented for this qualitative comparison (Fig. 5) we also checked for older Sentinel-1 deposits. We will highlight that these were just qualitative examples in section 4.1.1:

'The qualitative comparison of the manually derived Sentinel-1 deposits with the Pléiades deposits detected over time periods of ~1 month in the winter and summer seasons gives more insights on the potential of Sentinel-1 images to identify particular deposits (Fig. 5). '

4.1.1: This chapter would fit better into the discussion, except for some specific results.

Agreed, we will re-organize and streamline this section to stick to the direct comparison of Pléiades and Sentinel-1 outlines:

'The comparison of the manually derived Sentinel-1 deposits with the Pléiades deposits detected over time periods of ~1 month in the winter and summer seasons gives more insights on the potential of Sentinel-1 images to identify particular deposits (Fig. 5). It indicates locations of very good agreement, usually for large deposits with a lot of surface texture (Fig. 5b). But there are also false positive detections, for example caused by the opening of crevasses (Fig. 5c), as well as false negatives (Fig. 5d), that could reach large sizes (up to 60000 m^2, Fig. 5e). The comparison of the aggregation of one year of Sentinel-1 manual outlines with all the deposits identifiable in the end-of-summer Pléiades scene above 2700 m a.s.l results in a F1-score value of 0.47, with a majority of false negatives (Fig. 5). A large amount of undetected Pléiades deposits are actually smaller than the Sentinel-1 detectability threshold of 4000 m^2. Nevertheless, excluding them does not change the comparison (F1-score value of 0.49) between the Pléiades and aggregated Sentinel-1 deposits.'

264: As mentioned in the general comments, discussion is mixed into the results, for example here as you try to give reasons for results.

As mentioned above, we will remove here any redundancy with the discussion section 5.1.

Figure 5: Even though it is known that avalanche deposits, especially large ones remain visible for a long time, it seems weird to me to compare avalanches from 1.11.2019 Sentinel-1 imagery to 9.8.2020 Pleiades. I would be very careful with this comparison as I believe a comparison of what was visible around the same time is a lot more plausible. If you compare "everything with everything" it becomes hard for the reader to follow. Hence, could you go over your analyses again and leave those not very relevant out.

In this figure, we compared the aggregation of avalanche deposits over one full year. For Sentinel-1 this was easily conducted by taking all outlines since 01/11/19 and until 09/08/20. For Pléiades, it appeared to us that in an end-of-melt season image (09/08/2020), avalanche deposits that had not already completely melted were well visible (rougher, darker surface) and likely corresponded to the accumulation of all avalanche deposits of the previous year, thus the comparison with the aggregated Sentinel-1 outlines. The comparison between the 2 (Fig. 5a) actually shows a relatively good agreement, which comforts our hypothesis and gives an interesting perspective on the mapping of on-glacier avalanche deposits. We however also agree that it doesn't fit so well with the other results shown in figure 5, so we will:

- move panel 5a to the SI.
- shift the first paragraph of 4.1.1 at the end of the second paragraph to give it less importance

276: As mentioned in the general comments, methodology is mixed into the results, like here where you mention (again) how you compared operators. The overall IoU (comment to 224) would be very nice here in addition to the consensus numbers. Additionally, how do you define an avalanche event and how do you separate it? Hafner at al. (2021) found that avalanches in Sentinel-1 might be detected in more than one blob, so just taking connected pixels might be problematic/ lead to one avalanche being counted twice or more times.

We will indicate here the F1-scores for the comparison between operators. As mentioned above, for this entire study the comparisons of outlines are made on a pixel-basis as we are not so much interested in the events, but rather in the spatial and temporal variability of avalanches. We will make sure to clarify this in the methods:

'A single operator performed the manual detection, and to account for biases in the delineations, we compared on a pixel-by-pixel basis these outlines with those of four other operators for 4 scenes (2 ascending and 2 descending) covering the Mt Blanc region and 4 scenes covering the Everest region (Kneib et al., 2021; Table S1, Fig. S1-S2). '

We will remove from these results the repetition of the description of the methods and indicate the F1-scores of the different operators:

'The manual outlines of the three external operators had F1-scores between 0.54 and 0.66 relative to the outlines of the operator who derived the entire manual dataset for all three sites (Table S1, Fig. S1-S2). '

298: The references to S9b &Co are of different style than in other places and seem to be a mixture between page number and Figure caption. There is quite a few of those throughout your manuscript, please check them all and correct them (e.g., 388, 391, 474).

These are references to figures in the Supplementary Information. In this case, panel b of supplementary Figure 9. They follow the style recommendations of the journal The Cryosphere.

312: What do you mean by referencing to S11-13? Is this supposed to be page numbers? Maybe it should be refereeing to a chapter (then it should be the chapter numbering)? There is quite a few of those throughout your manuscript, please check them all and correct them (e.g., 377, 388, 391).

As explained above, these are references to figures in the Supplementary Information.

302: Removing 36% and adding 41% is quite a lot. I already mentioned this in the general comments, could you please add a section to the discussion where you discuss the benefits of your approach despite it not being transferable and needing quite a bit of manual work.

Please refer to our answer to the general comments.

312/Figure 6: It is unclear for which variables correlation was calculated exactly. Please make this clear.

We will specify this in the figure caption:

'The Pearson's correlation coefficients characterizing the correlation between the validation set and the outlines from the automated mapping approach are indicated in blue (ascending) and red (descending).'

4.1.3: Did you use the automated mapping prior to manual corrections for this analysis? Or after the step described in 302?

This section still refers to outlines from the automated mapping approach. We will specify this at the start of the results section (see response to comments above) and in the caption of Figure 7:

'F1-score obtained when applying different sets of parameters to sets of images for which they were not calibrated, without any manual modifications. '

Figure 7: What is N-A and M-O for the Ascending and Descending?

This stands for November-April and for May-October. We will specify this in the caption:

'N-A and M-O stand for the November-April and May-October periods, respectively'

347: What is R2?

This stands for coefficient of determination. We will spell it here.

Figure 8: It is not clear at first sight that the legend in (a) is true for all panels. Please move the legend outside the panels as it applies to all.

We will change 'Total size of avalanche events' to 'Normalized area of all avalanche events' for the legend of panel c for clarity. But the legend of panel a only applies to panels a and b, which we find to be self-explanatory as it is. We will therefore keep it this way:

'Figure 8: Size distribution of avalanche events at the three different sites, with (a) and without (b) normalisation. (c) Normalized area of all avalanche events expressed as a function of the surveyed area segmented in 200m elevation bins. '

359ff: Here you are mixing methodology and results again, please disentangle.

We will remove the first three sentences that are repeated from the methods.

362: What is meant by activity is not clear, I assume it is repeated occurrence of an avalanche in the same place. Since it is not possible to detect whole outlines in Sentinel-1, how did you determine deposits to be "the same", in other words how much overlap did you require?

Following the comment above we will remove this sentence that is repeated from the Methods. The definition is given in Section 3.6 of the Methods, which we will modify for clarity:

'The union of all avalanche pixels over time indicates individual deposits affected by more or less avalanche activity. We estimated the influence of avalanches on a given glacier, independently for ascending and descending orbits, with two metrics: area affected by avalanches and avalanche activity. The area affected by avalanches is estimated by taking the union of all individual avalanche deposits, and expressed relative to glacier area. The avalanche activity is calculated for each pixel as the number of deposits affecting this pixel over a given time period. It is then calculated on a per-deposit basis by taking the maximum activity and on a per-glacier or per-elevation band basis by taking the area of the glacier affected by avalanches divided by glacier area or area of elevation band, respectively.'

365: You are using Change detection with a D and D-i image, with change appearing green in your data. The green hue vanishes when moving forward in time even if the deposit remains well visible/ the backscatter of a single image high. Did you analyze backscatter separately or how did you come to that conclusion?

This is a very nice way of summing up our approach. What we meant here was that when there is a new avalanche on an old deposit, it doesn't necessarily lead to an increase in surface roughness (and therefore backscatter), which means that the mapping approach does not work to detect this new avalanche. This is shown by examples in Fig. S9c-d.

Figure 9: I do not see a benefit in (a), the graph is hard to read. I believe a simple table could be a better choice for bringing your point across. Furthermore, the choice of color makes the differentiation between ASC and DESC hard (it is also not color blind safe). Additionally, areas detected in both ASC and DESC cannot be identified. To get the full picture, it would

also be nice to be able to see the area that was masked as outside the glacierized or in radar shadow/layover.

For a more general answer on the value of Figs. 9a, 10-13, we refer the reviewer to our answer to the general comment on this matter. More precisely for this panel (a), we feel that it is particularly important to summarise the avalanche activity of the different individual deposits in each survey area in terms of number of events per unit area, and such information would be difficult to fit in a table. This panel also fits well with the other, more qualitative panels, that indicate the number of repeated avalanches on different avalanche deposits. We will however put a log scale for the y axis to make it more readable. We will also indicate the count in number of deposits per square kilometres to make it more comprehensive. To improve readability of the maps, we will show the descending deposits in blue, indicate in the caption that the ascending and descending deposits overlap (this will ensure that it is colour blind safe) and indicate in the figure the shadow/layover and glacier mask.

Figure 10: This figure is hard to read. Overlapping circles cannot be distinguished and the absolute values of circles of the same size in (a), (b) und (c) differ, conveying that values are the same, though they are not. I would remove that Figure.

For a more general answer on the value of Figs. 9a, 10-13, we refer the reviewer to our answer to the general comment on this matter. We still find this figure interesting as it specifically targets the quite loose definition of what is an 'avalanche-fed' glacier. However, given the reviewer's concern about the high number of figures in our manuscript, we will shift it to the SI. To improve the readability we will put the circle edge colour in black so that overlapping circles are visible. We will also scale the circles to the same size using a log scale to make them comparable between regions.

4.3: Please carefully reexamine for mixture of results with discussion.

We will remove any explanation of the linkages between avalanche activity and meteorological variables for the Mt Blanc massif.

Figure 11/12/13: I would suggest displaying only one of those in the manuscript and moving the other two to the Appendix. Furthermore, you should not use the same color range for number of avalanches and for avalanche area. Additionally, the number of avalanches is discrete (I assume you are displaying per acquisition day), while the way you display it implies a continuous scale. Lastly, the x-axis is the same for all panels, I think they would be better readable/comparable if the time would only be displayed once at the bottom of all three panels. In (c) you could improve readability by adding a thin grey line at 0° Celsius. Additionally, you should indicate data gaps in Fig 12 the same for all panels and not once white and once black.

For a more general answer on the value of Figs. 9a, 10-13, we refer the reviewer to our answer to the general comment on this matter. For Fig. 11-13, we believe that the comparison of the avalanche activity at the three sites is one of the main results of this paper and will therefore keep all three figures in the main text, to allow visual comparison between regions. We will however improve the readability by:

- using a different colour for area and number
- using a discrete colour scale for the number of avalanches and specify in the caption that these numbers are 'for each Sentinel-1 pair'
- using one x-axis for all three panels
- adding a thin grey line at 0°C
- indicating the data gap in T & P data in white
- adding the region title at the top of the figure

440ff: "the performance is generally very good in dry snow conditions"- this contradicts findings by Eckerstorfer et al. (2022) who found a low Probability of detection for solely dry snow avalanches and you also contradict yourself in 456ff where you state that "cold, low density snow avalanches are likely to be missed". I suppose cold low density snow avalanches make up a good proportion of avalanches occurring under dry snow conditions. You also state that a rough surface is detected by its backscatter, generally wet snow avalanches tend to have a rougher surface. Could it be that the changes in overall snow wetness, wet to dry or dry to wet from D-I to D are one of the main drivers affecting detectability. For example, in Figure 11 it seems that the avalanche activity was very low for rain on snow (after a period of low temperatures) events which are generally known for critical avalanche situations and remained low until temperature conditions were stable again over a period.

Please refer to our response to the general comment for the first part of the comment.

You are absolutely correct that wet snow conditions lead to false positive detections in the case of a dry to wet snow transition. We will explicitly state this in the discussion:

'The few studies that targeted extensive periods rather than a specific event also encountered the most difficulties for periods with wet snow conditions, leading to extensive false positive detections which had to be removed manually in situations of dry to wet snow transitions (Eckerstorfer et al., 2019)'

447: "such approaches", please be specific and precise.

We will specify 'such machine learning approaches'

469: "we therefore recommend"- therefore refers to previous reasoning and discussion which is absent here. Please elaborate why you believe morning scenes are better suitable before giving a recommendation. Following your argumentation- wouldn't it be possible to mitigate the effects of snow wetness changes (caused by a rise of temperatures during the day) by comparing morning to morning and evening to evening scenes?

On the contrary, 'therefore' refers to the limitations of the automated approach described in the previous paragraph:

'Indeed, scenes unaffected by snow wetness changes (descending/morning acquisitions during the cold season) are well mapped regardless of the parameter set (Fig. 7). '

And yes, as indicated in Table 1 (acquisition time) the scenes acquired on the same orbit are always taken at the same time of the day, so choosing morning-to-morning scenes is therefore the obvious choice. We will specify this here:

'For future implementation of SAR detection of avalanches, we therefore recommend prioritising the use of morning-to-morning scenes'

472: I am a bit skeptical of you being so sure about a good manual check/ correction. Could you please elaborate a bit on why you are so sure to be able to detect (true) false positives/ false negatives without additional information.

This is a good point. We will mention that even with manual edition, these scenes will likely lead to more uncertainties:

'Although scenes acquired in the afternoon may help fill spatial and temporal gaps, it is important to note that they will require additional work to separate actual avalanche events from false positive detections caused by snow wetness changes. This is a difficult task leading to higher uncertainties for the mapping, and will likely not considerably change the long-term spatio-temporal patterns of avalanche activity (Fig. 11-13). '

480: Didn't you exclude all deposits smaller than 4000 m2?

Good point. We will remove the second part of this sentence and remind the reader that deposits smaller than 4000 m^2 were filtered out i.e.:

'The size distribution of avalanches with Sentinel-1 RGB pairs reaches a maximum around 4000 m^2 (avalanches smaller than 4000 m2 have been filtered and therefore not considered in this study).'

491: How did you determine an overlap? See also comment to 362.

Please refer also to the response to comment L. 362.

509-522: Not all discussion is related and relevant to your work, I suggest to significantly shorten this section. Additionally, I wonder if avalanches being more concentrated at lower elevations is mostly related to wetter snow conditions and better detectability (see also Eckerstorfer et al., 2022, Abermann et al., 2019). See also comments to 440ff.

We will remove the discussion on slope control on avalanching. You also have a good point regarding the detectability at lower elevations, which we will indicate here:

'In addition, the detection at these lower elevations could also be aided by the wetter snow conditions (Eckerstorfer et al., 2022; Abermann et al., 2019). '

538: You contradict yourself here regarding to 440ff.

We've addressed this comment in the General Points.

5.3: Generally, since one of your research questions is to "evaluate implications for the glacier mass balance" I expect you to be a bit more specific and elaborate on this a bit more.

Please see response to specific comments below.

556: Did they use whole outlines for parametrization or just deposits/part of the avalanche as can be detected from Sentinel-1?

These studies have calibrated their parametrization in a qualitative way, based on the general shape and extents of avalanche deposits visible in relatively coarse optical images. We will specify this:

'Such calibration has been conducted in a qualitative way based on comparing the deposits from the model and the general shape and extents of deposits in a few optical images.'

562: This sections content does not really fit the chapter headline.

On the contrary this section addresses two important points in relation to glacier mass balance:

- Glaciers with avalanches have steep headwalls, but glaciers with steep headwalls do not necessarily have avalanches. Therefore one cannot only use topographic arguments to characterize a glacier as 'avalanche-fed'. We will simplify this sentence to improve readability:

  'At the glacier scale, we could therefore show that the presence of steep slopes within the glacier catchments is a clear necessary condition for avalanches to occur (Fig. 10, S14; Hughes, 2008; Laha et al., 2017), although not a sufficient one'

- Avalanches are well correlated with precipitation, indicating that there is little to no snow retention from surrounding headwalls at the scale of ~1 month, which is something important to consider for the rerouting of the snow.

We will move this paragraph to the start of 5.3 for the link with the calibration of the avalanche parametrization to be made clearer.

5.3/6: I am missing a section with a throughout discussion of the limitations of your method. That would for example include that even though your methodology can detect the approximate area and frequency of avalanches, it does not give you any information on the mass of snow involved.

We argue that in 5.3 this point is apparent in the first sentence of paragraph 1:

'While the Sentinel-1 images do not give any indication on the volume or mass of the redistributed snow, we obtained from these products key information related to the spatial extents of the avalanche deposits and the spatio-temporal variability of the avalanche activity'

We will add a similar sentence at the end of the conclusion:

'While it does not give any information on the mass redistributed by avalanches, our approach enables the mapping of avalanche deposits over long time periods at the scale of a small mountain range'

573/578/589: "we successfully established a semi-automated framework"- with the manual identification of thresholds, the automatic detection and the extensive manual correction (requiring extensive domain knowledge) it remains unclear to the reader how large this benefit is. What is the gain (e.g., time) compared to full manual mapping and how much of your methodology may be reused and saves whoever uses it time (this is linked to limitations, see comment to 5.3/6). Furthermore, are there ways to translate the avalanche area into mass without additional measurements (e.g., Hynek et al., 2023) that in your case are not available.

We have mentioned above that applying the automated approach reduces mapping time by at least half, more if only morning scenes are considered. Another major output of this study is a quality-controlled dataset of 16302 avalanche deposits in data-scarce regions, which constitutes an excellent training dataset for future method development. We will add this point to section 5.1:

'In the end, this study resulted in a manually checked dataset of 16,302 avalanche deposits. This dataset will be highly beneficial for the training of future mapping approaches.'

Regarding the translation of avalanche area into mass, we believe that using avalanche outlines to calibrate simple avalanche redistribution parametrizations is the most straightforward way to get to mass, as direct volume-area scaling without accounting for the actual precipitation amounts is unlikely to work. We agree that the best would be direct measurements of this contribution, and are working on such approach on Argentière Glacier combining high-resolution DEMs and field measurements (similar to Hynek et al., 2023), but such efforts are extremely time and resource consuming and will only ever work for 1 or 2 sites so for larger scale modelling, such combined approaches of remote sensing & modelling are better suited.

We will highlight these elements at the end of the conclusion:

'While still requiring manual checks, this approach considerably reduces the mapping effort, and the large dataset obtained will help train future mapping approaches, and calibrate mass redistribution parametrizations to be applied in the surface mass balance routines of glacio-hydrological models. '

Could you please give an outlook on what is still needed for including avalanches into glacier models large scale with your methodology.

This will be indicated at the end of section 5.3:

'Once calibrated, such avalanche redistribution parametrization can be coupled to the mass balance routine of a glacier model, for a more accurate representation of accumulation processes (Bernhardt and Schulz, 2010; Ragettli et al., 2015; Quéno et al., 2023).'

References:

Abermann, J., Eckerstorfer, M., Malnes, E., and Hansen, B. U.: A large wet snow avalanche cycle in West Greenland quantified using remote sensing and in situ observations, Nat. Hazards, 97, 517–534, https://doi.org/10.1007/s11069-019-03655-8, 2019.

Liu, C., Li, Z., Zhang, P., Huang, L., Li, Z., and Gao, S.: Wet snow detection using dual-polarized Sentinel-1 SAR time series data considering different land categories, Geocarto International, 37, 10 907–10 924, https://doi.org/10.1080/10106049.2022.2043450, 2022.

Eckerstorfer, M., Oterhals, H. D., Müller, K., Malnes, E., Grahn, J., Langeland, S., and Velsand, P.: Performance of manual and automatic detection of dry snow avalanches in Sentinel-1 SAR images, Cold Regions Science and Technology, 198, 103 549, https://doi.org/https://doi.org/10.1016/j.coldregions.2022.103549, 2022.

Hafner, E. D., Techel, F., Leinss, S., and Bühler, Y.: Mapping avalanches with satellites – evaluation of performance and completeness, The Cryosphere, 15, 983–1004, https://doi.org/10.5194/tc-15-983-2021, 2021.

Hafner, E. D., Techel, F., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice, Natural Hazards and Earth System Sciences, 23, 2895–2914, https://doi.org/10.5194/nhess-23-2895-2023, 2023.

Hynek, B., Binder, D., Citterio, M., Larsen, S. H., Abermann, J., Verhoeven, G., Ludewig, E., and Schöner, W.: Accumulation by avalanches as significant contributor to the mass balance of a High Arctic mountain glacier, The Cryosphere Discuss. [preprint], https://doi.org/10.5194/tc-2023-157, in review, 2023.

Howell, S. E., Small, D., Rohner, C., Mahmud, M. S., Yackel, J. J., and Brady, M.: Estimating melt onset over Arctic sea ice from time series multi-sensor Sentinel-1 and RADARSAT-2 backscatter, Remote Sensing of Environment, 229, 48–59, https://doi.org/https://doi.org/10.1016/j.rse.2019.04.031, 2019.

Nagler, T., Schwaizer, G., Keuris, L., Rott, H., Luojus, K., Moisander, M., Small, D., Metsämäki, S., Malnes, E. and Eckertorfer, M.: SEOM S1-4Sci Snow: Development of Pan-European Multi-Sensor Snow Mapping Methods Exploiting Sentinel-1, Final Report, Deliverable 4.2, https://eo4society.esa.int/wp-content/uploads/2021/06/S14SciSnow.D4.2_v1_2_FR.pdf, 2021.