Editor

Dear authors

I have now received the comments by the reviewers on your revised manuscript.

Please consider those and make according revisions.

I will then make a final decision.

Best regards,

Jürg Schweizer.

Dear Prof. Schweizer,

Many thanks for the consideration of our manuscript. We would also like to thank the reviewers for their comments that helped improve the manuscript in terms of scientific content and readability. There remained a few suggestions, which helped us refine some of the text. In response to the main comments we have conducted the following changes:

- We have added more details to the discussion to clarify the difference between cold and wet snow conditions and the implications for the mapping (F1 scores) and possible resulting biases
- We have added a figure to the SI showing an example of high backscatter values caused by avalanches remaining visible for > 6 months on Hispar Glacier

Our answers to each specific comment are further indicated in blue below and the line numbers indicated correspond to the revised Manuscript with tracked changes.

We think that the manuscript has been strengthened by these revisions, but none of our main results or conclusions have changed.

Thank you for your consideration of our revised manuscript.

Kind regards,

Marin Kneib and co-authors

Reviewer 1

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

Dear authors,

Thank you very much for the work you put into improving your manuscript. It reads a lot smoother and is noticeably clearer than the previous version. I have a few points to raise (still). Especially the first one is at the core of your work needing improvement before this may be published:

In your answer to my comments, you differentiate which statements regarding dry/wet snow refer to which processing step/process. There the difference is clearly stated and described while the manuscript is still lacking this clarity. Furthermore, in the same context, you need to elaborate and discuss the effect of snow wetness (detectability of dry or wet snow avalanches) on your results better and clearly put it in the context of previous work stating some of their findings explicitly, comparing and discussing. For example, 561 (1-2 months delay in activity) also points in the direction that it is easier to map the wet snow avalanches.

We apologise if that difference was not made clear enough in the manuscript. We have now added a sentence to the discussion to make this explicitly clear (L494-495):

'Therefore, while dry snow conditions lead to detectability limitations in Sentinel-1 images (Fig. 4), 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 (Eckerstorfer et al., 2022). '

We have now also further detailed the effects that snow wetness may have on the observed patterns (potential observational biases), L561-563:

'the detection at these lower elevations could be aided by the wetter snow conditions, leading to lower backscatter background values that are favourable for the avalanche detection (Eckerstorfer et al., 2022; Abermann et al., 2019).'

And L576-578:

'These relatively high values in Spring could partly originate from a bias in the avalanche detection, as low backscatter background values (wet snow) make it easier to detect avalanche deposits (Eckerstorfer et al., 2022; Abermann et al., 2019).'

Regarding the 1-2 months delay in avalanche activity that you refer to, we do not think that it could be caused by an observational bias as in this region (Everest) the monsoon (precipitation peak) occurs during the warm period (temperature peak), so if there is a bias in detection of events caused by dry to wet snow transition, there should be no delay.

You have corrected a lot of the passages with “spoken language”, but at the same time you have introduced some new ones with your corrections e.g., “was actually” or “are actually”. Please reexamine all newly inserted sentences and avoid spoken language in the whole manuscript.

Apologies for this, we have removed the term 'actually' L289 & L489.

229: The notation of the values in [] is not clear. Please explain what you mean in an easier to understand way and connect well to the sentences around.

We have now specifically mentioned ([min; max]) for the range of values to make this clearer. We have also indicated which parameters these correspond to: (T_S, T_V, T_O, T_{D1}, T_{D2} and T_{D3}). L230-232.

464: Please include in this discussion a sentence or two that all avalanches occurring on top of eachother within 6d/12d will be counted as one, additionally lowering the total number of detected avalanches.

This was already stated L473-475:

'Similarly, the detection of the avalanche events requires the previous deposits to have regained lower backscatter values for the signal to be visible, meaning that the surface of the deposit needs to have been smoothed by additional precipitation or melt for the next events to be visible at this location.'

We have made this even more explicit L475-477: 'We have observed this smoothing to require several weeks and even months before avalanches can be detected at the location of old deposits, while avalanche events are still occurring in the meantime and are therefore difficult to detect (Fig. S12d).'

Reviewer 2

The authors have clearly invested substantial effort in this study, covering three distinct regions characterized by diverse climatological and topographical features. The analyses conducted are thorough, delving into numerous details. However, for future work, I suggest streamlining the presentation of results to prioritize certain analysis over others. This approach would enable a more focused exploration of fewer topics, facilitating deeper, clearer analysis and discussion. Nevertheless, the manuscript has undergone significant improvements particularly in readability and structure. Figures 1, 3, and 4 have improved in terms of size or geographic context. Furthermore, both, the results and discussion sections have been streamlined, and the overall structure has improved through renaming the (sub)sections. Upon reviewing the revised script, I've provided some comments and suggestions for further improvement in the manuscript and supplementary material PDFs. These suggestions aim at refining the content further. Consequently, I recommend publication of the manuscript following the implementation of these suggested revisions.

We would like to thank Reviewer 2 for their thorough review and their very relevant and constructive comments.

1. Comparison to actual avalanche data: I understand the difficulty in finding ground truth avalanche records for comparison. However, a comparison to data from, e.g., the Mt. Blanc Massif would have strengthened the results from the semiautomated detection and should be considered for future work. For example, the ANENA (Association Nationale pour l'Étude de la Neige et des Avalanches) provides yearly publicily available avalanche reports from

2013 onwards, which might be worth considering. I appreciate the effort of the comparison to avalanche risk levels, although they are not the same as a comparison to ground truth data.

We agree that more validation data would have been welcome, but as mentioned in the first round of review, this is particularly difficult in HMA. For the Mt Blanc massif, we already noted in our previous response that 'the French historical avalanche maps (CLPA: "Carte des limites probables des avalanches", map of probable avalanche limits in English) do not cover the glaciers of the Mt Blanc massif due to a lack of data.' We can add here that even in non-glacierized regions these maps are generally too coarse to compare to (they also include not just the deposit areas but also the release and propagation areas) - see figure below. We have looked for the ANENA reports you refer to but only found reports of avalanche casualties, which are very sparse and give limited information on the spatial characteristics of the avalanche deposit (in general the location of the event is not even given). For this particular study we therefore considered such a comparison to be of limited value.

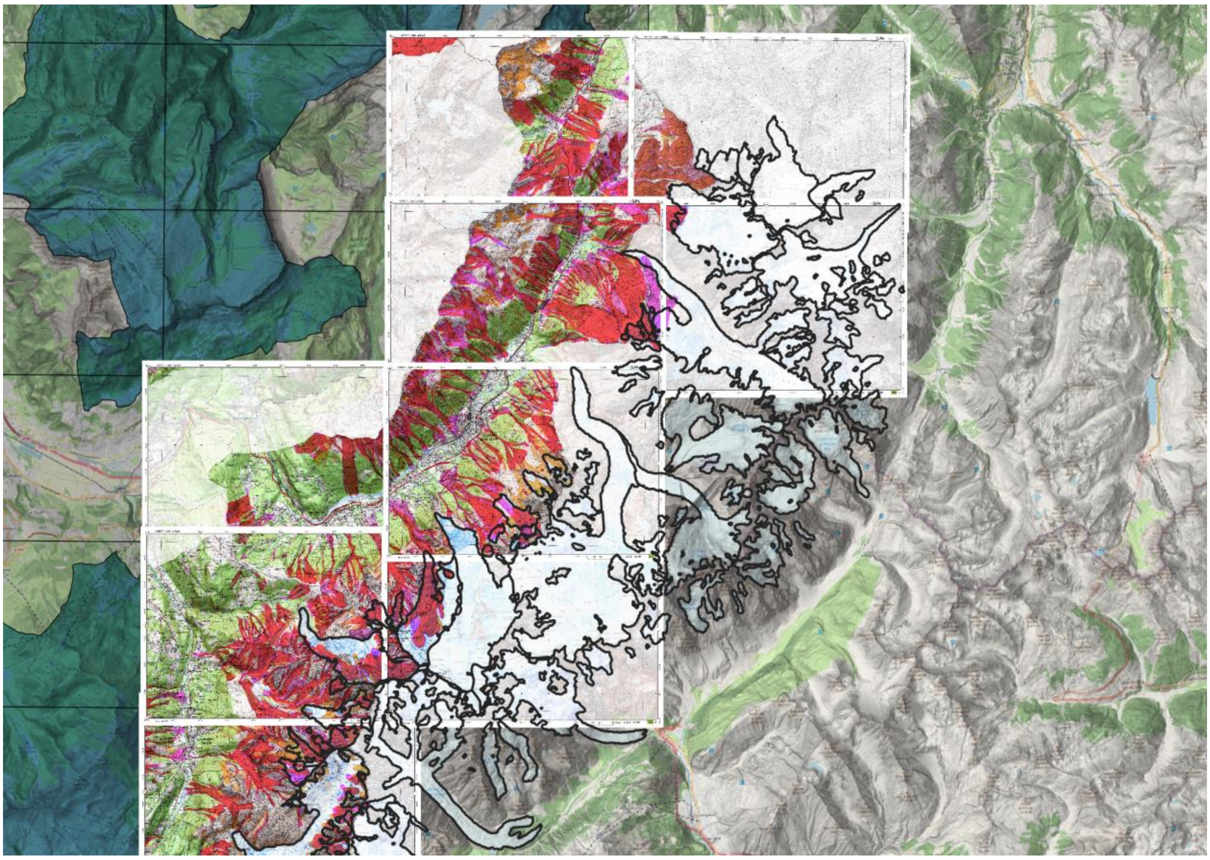

Figure R1: CLPA maps of the Mt Blanc region. CLPA couloirs in red, glacier outlines (our area of interest) in black. There is very little overlap between the two and the avalanche couloirs are relatively coarse.

2. Clarity about relative orbit: While Table 1 provides clarity regarding the images used, why not consider using daily images to enhance avalanche tracking. Of course, only images from same relative orbit can be compared, and coverage of different relative orbits are not always the same, but a smaller intervals between images could have offered a higher temporal resolution, potentially tracking avalanches and their transformations, including those affected by wind.

We thank the reviewer for sharing their idea: it would be a nice way to get more details on some of the events. However, the revisit times for most of these survey domains at relatively low latitudes remain limited to 6 and 12 days, even considering other orbits, as there is very little overlap between tracks. This is for example visible in these figures showing the different Sentinel-1 tracks:

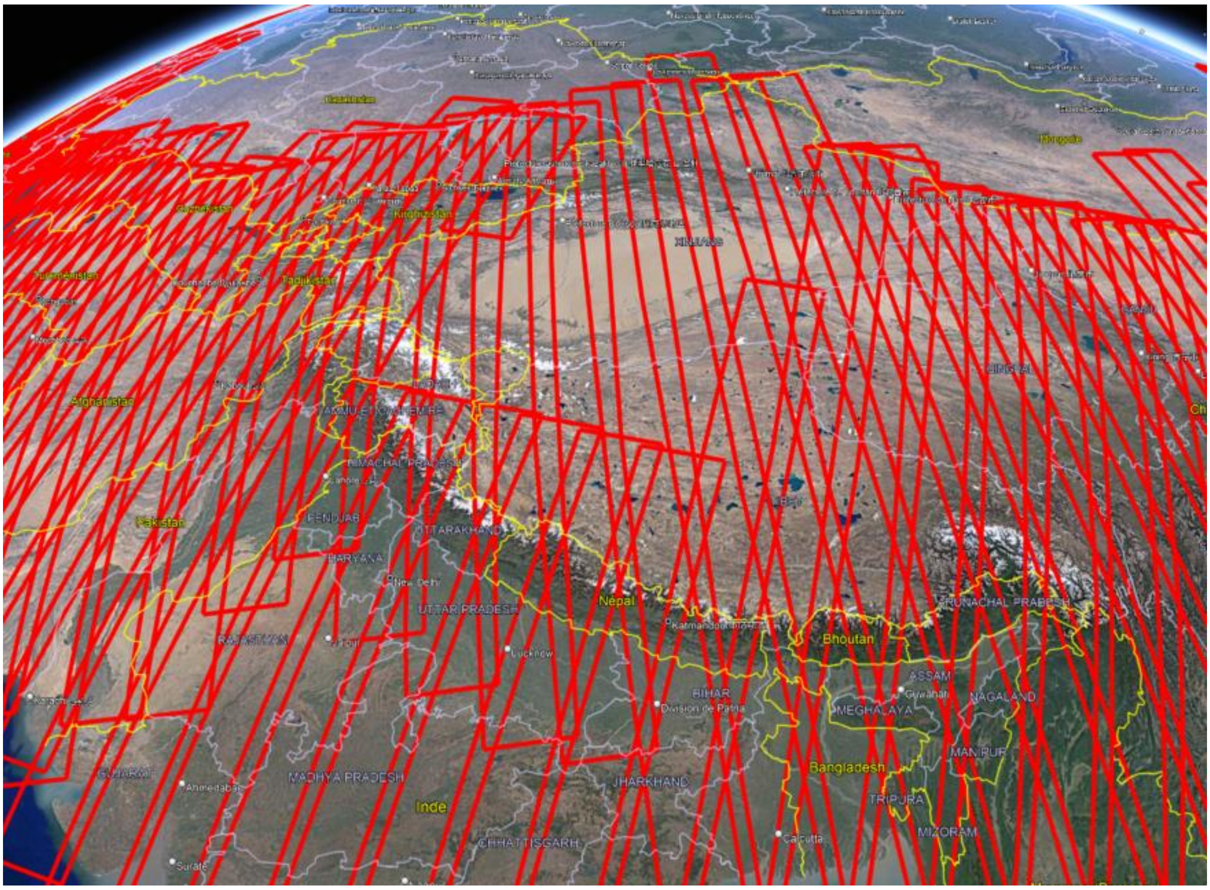

Figure R2: Sentinel-1 geographical coverage in 2020 over High Mountain Asia. Background image from Google Earth.

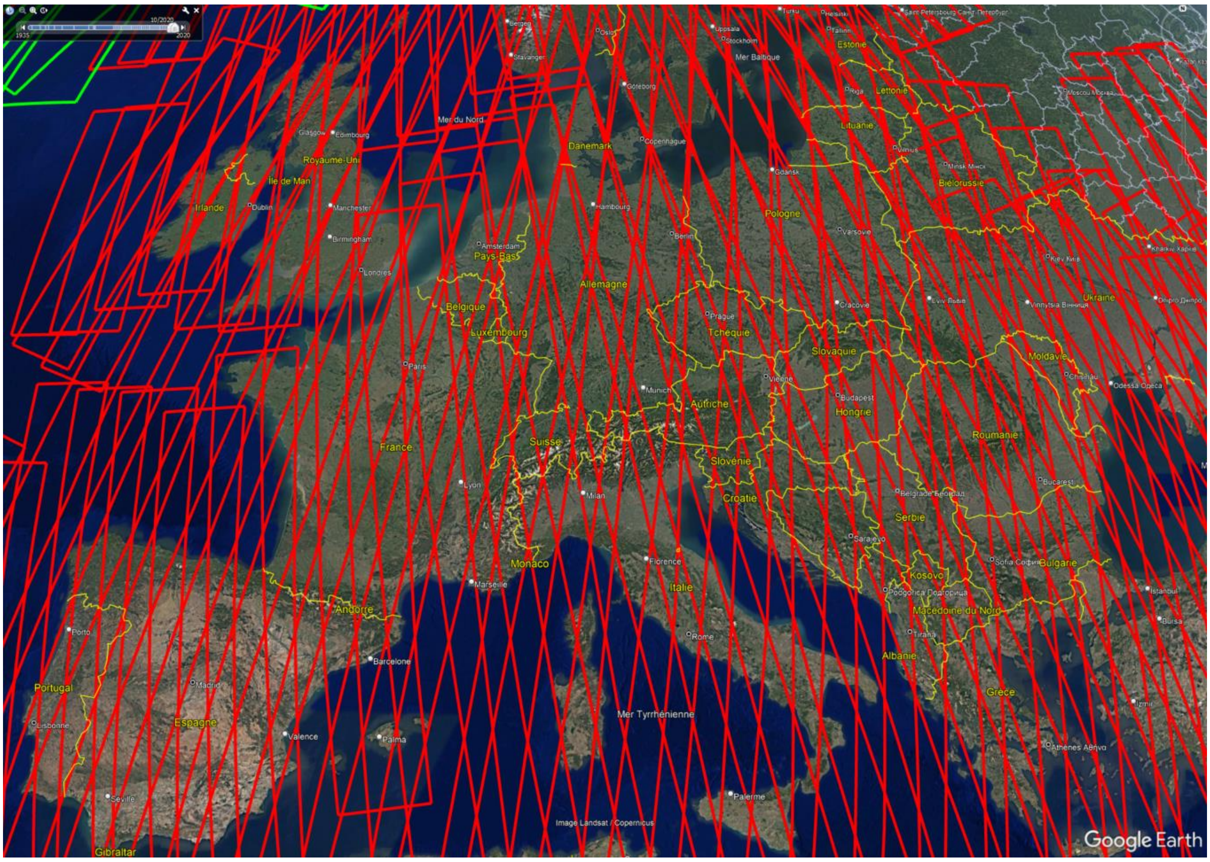

Figure R3: Sentinel-1 geographical coverage in 2020 over Central Europe. Background image from Google Earth.

3. Low F1 coefficient: The F1 score ranges as stated in the manuscript: “above 0.5 in 78% of cases (0.6 in 83% of the cases for the Hispar November-April scenes). The ascending scenes present in general lower F1-scores (lower than 0.5 in 92% of cases), particularly the May-October scenes of Everest for which the F1-score never exceeds 0.32. With an average F1-score of 0.46, the Everest descending NovemberApril parameter set is the most transferable, but still performs poorly (F1- score<0.4)” and “F1-score between 0.29 and 0.78”. I still am convinced that these values are rather low compared to articles using e.g., machine learning. For instance, Bianchi et al. (2021) achieved scores surpassing 0,66, while Hafner et al. (2022) reported 0,625 across diverse topographical regions. While a brief mention of the low values is provided in one paragraph of the discussion, I find that a critical and comprehensive analysis elucidating the underlying reasons for these comparatively lower values is lacking. Especially, the part: “The performance of such approaches is generally very good in dry snow conditions, with high precision (>0.7) and low false positive rates (<0.4), which correspond to F1-scores above 0.6-0.7 (Leinss et al., 2020; Eckerstorfer et al., 2019). 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).” needs to be modified. Please see my comments to this in the annotated PDF. A rectification of the low F1 scores should provide compelling justifications for the decision to refrain from employing machine learning methods in this work.

The high scores that you refer to for Hafner et al. (2022) are obtained for automated mapping with SPOT 6/7 optical images, this therefore is not comparable to our mapping with Sentinel-1 images - SPOT images provide high resolution optical images allowing a robust detection of avalanches, but need to be tasked and have therefore much less temporal and spatial coverage than Sentinel-1. Bianchi et al. (2021) did achieve higher scores with machine learning, although it is unclear from their study if these were impacted by the snow wetness changes. In any case, this agrees with our statement acknowledging the low F1-scores in certain conditions and saying that machine learning approaches provide a promising way forward to reduce false positive detections in situations of snow wetness changes (L498-501):

'Such false positive detections can be discarded manually based on size and texture considerations, which indicates that deep learning approaches based on convolutional neural networks, for example, offer a promising way to improve these classifications (Tompkin and Leinss, 2021; Waldeland et al., 2018; Yang et al., 2020; Bianchi et al., 2021; Kapper et al., 2023; Liu et al., 2021; Lê et al., 2023). '

Please see our answer below for details on the low F1 scores in wet to dry snow transition conditions.

4. Reasons behind the lower ascending and descending F1 scores: In the discussion, I suggest including a brief explanation of potential factors contributing to the lower F1 score observed in the ascending compared to the descending scenes.

We have added a brief explanation in the discussion (L503-506):

'Indeed, scenes unaffected by snow wetness changes (descending/morning acquisitions during the cold season) are well mapped regardless of the parameter set (Fig. 6). Ascending scenes, acquired in the afternoon, are more likely to be affected by snow wetness changes than descending scenes, acquired in the morning. This explains the lower F-1 scores for these scenes.'

5. Fig. 5: It is difficult to distinguish between the different lines for the manually and automatically detected avalanches as well as ascending and descending scenes. Is it necessary to separate ascending from descending, because earlier you mention that "The automated mapping generally underestimates the number and sizes of the avalanche deposits .." without distinguishing between the two. Combining the two would make the figure clearer to read.

We did not want to combine ascending and descending detections as the acquisition dates differ and as the temporal variability shows strong differences between the two, which are worth showing, also for Fig. S14-S16 of the SI. We tried separating figure 5 into four subplots instead of two but this did not improve the visibility of the lines.

6. Fig. 8b-e: In the caption you mention that the shaded black areas were excluded (masked out) from the analysis. Nevertheless, there seems to be an overlap with some dark blue color that I assume are detected avalanche deposits in the descending views. If you did not take the avalanches in the shaded area into account, these (descending) overlapped areas should be removed. Furthermore, the dark blue does not appear in the color legend. Moreover, the clarity of the figure could be enhanced by combining the ascending and descending views - considering that there is no reference of this categorization in the text specifically related to this figure, and no discernible difference appears to exist. There even seems to be some overlap in parts of the descending and ascending detected avalanches. Did you count them separately? Did you use them to confirm the detected avalanches?

Thanks for pointing this out, these small overlaps have now been removed. We have also specified what the mask corresponds to in the caption (L404-405):

'The shaded grey areas correspond to the intersection of the ascending and descending masks.'

The dark blue colour does appear in the legend at the bottom of the figure. We choose to separately treat descending and ascending orbits throughout the manuscript given that they indicated slightly different temporal patterns and that they corresponded to different acquisition dates - we therefore did count them separately. As a result we kept the separation in this figure as well. As described in the results/discussions, we counted them separately and chose not to use them as a confirmation although this is a good idea for future studies and we have now mentioned it in the discussion (L510-511):

'Although scenes acquired in the afternoon may help fill spatial and temporal gaps and could be used as a confirmation for some detections,'

We have also improved the readability of the figure by showing only contour lines spaced every 200 m and reducing the transparency of the masked areas.

Line-by-line comments

L14: seldom->seldomly

We respectfully disagree. Seldom is the right adverb, could be replaced with rarely but we've kept seldom here (L14)

L51: 'wind-blown from steep headwalls': I would remove this part as you are speaking of differences of avalanches in glacierized catchments versus off-glacier avalanches.

Done (L51)

L75: 'near real-time': this is still misleading, also in the article by Eckerstorfer et al. (2019) near-real time suggests the avalanche to be detected right after release.

We've removed this term (L75)

L79: increases -> increased

We use here the noun rather than the verb, and have therefore kept it as it is (L82).

L86-87: 'but they have been limited by the lack of large training datasets for this application.': The limits are rather due to the resolution of the Sentinel1 images, the low repeat-frequency in middle Europe, the detection of dry snow, the algorithm, shadow-effects and layover etc. There is at least one large dataset that could be used for training. PLease modify.

We have re-organized this paragraph to make the link clearer between the use of machine learning approaches and the limitations from Sentinel-1 images. We've also removed the limitation caused by the lack of training datasets (L87-94):

'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). 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; 2022; Hafner et al., 2021), or will not work in areas affected by radar shadow or layover.'

L87: remove 'high'

Changed to 'Sentinel-1 satellites have a repeat frequency of 6-12 days for low latitude regions (European Alps and HMA)'

L87: (6-12 days repeat cycles): please specify about which areas you are talking. When both Sentinel-1 satellites where running, repeat frequency was daily in northern latitudes

Changed to 'Sentinel-1 satellites have a repeat frequency of 6-12 days for low latitude regions (European Alps and HMA)' (L79-80)

L88: remove 'free of charge': repetition and the satellites are not free of charge

Done (L80).

L118: replace 'the same orbits for each survey domain' with 'only one orbit for ascending and descending track respectively being 6 or 12 days apart depending on the survey domain.'

Changed as suggested (L119-121): 'We used one orbit for the ascending and descending tracks, respectively 6 or 12 days apart depending on the survey domain, to guarantee that the incidence angles remained the same throughout the study periods.'

Figure 1: again, in a-c it would be nice to see the complete extent of the Sentinel-1 ascending and descending scenes, since there are only 2 it would have been easy to show them

As stated in our previous response: 'These regions are a small subset of Sentinel-1 scenes, and the boundaries of the Sentinel-1 tiles are not visible in this small subset. ' We show as an example a screenshot from the ASF datasearch showing in blue the different S1 tiles and our survey domain in the yellow polygon:

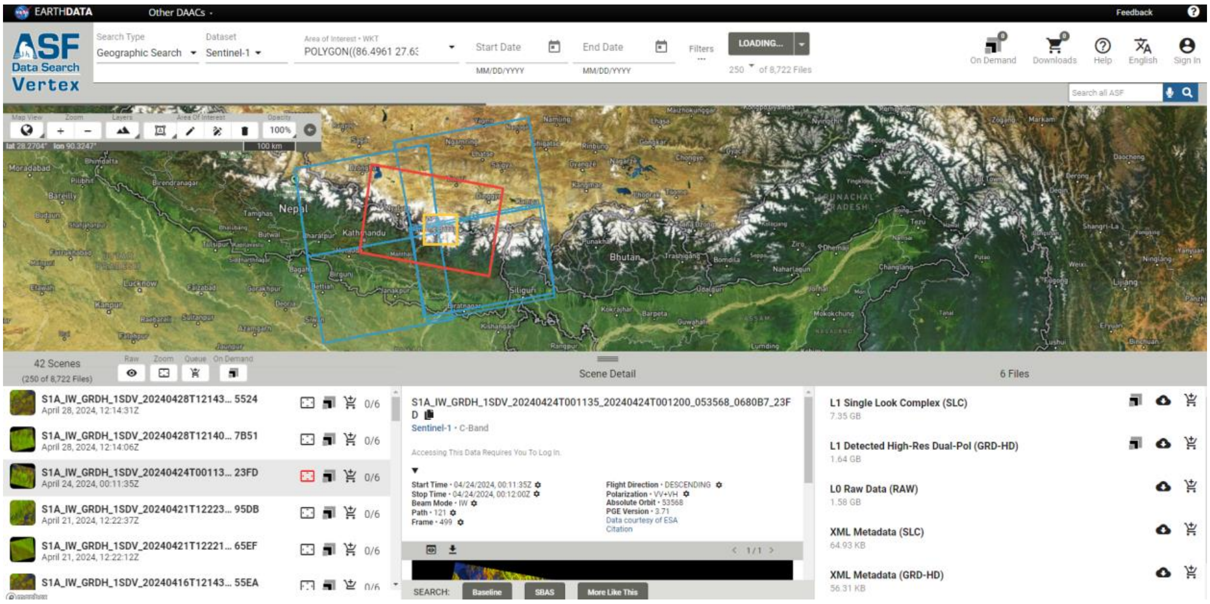

Figure R4: ASF data explorer for Sentinel-1 images over the Everest region (yellow polygon). Overlapping Sentinel-1 tiles are shown in red (ascending) and blue (descending).

Based on the scale difference, we do not see the added value of having the Sentinel-1 footprints displayed in figure 1.

L251: remove ‘All Sentinel-1 images were pre-processed in Google Earth Engine.’. Repetition, already mentioned in 3.1.

Done (L255).

L298: +/-: this still needs to be improved.

This will be modified in the final proofing stage.

L313-314: ‘linked to widespread snow backscatter increases likely due to wetness changes, especially during the May-October season’: i still would remove this, because it is an assumption that has not been checked and should be in the discussion

We agree in principle, but this is still required to justify the manual filtering of the false positive and false negative detections that comes right after: ‘Such false positive or false negative detections were manually removed or added based on considerations of shape, size and location, and this manual filtering was applied to all time series of all survey domains for the results presented in sections 4.2 and 4.3.’ (L320-322). We have therefore kept it as it is.

L328-329: ‘Pearson’s correlation coefficient’: please add formula, although it might be a common formula. maybe it fits inline

Considering that this metric is widely used and that any information about it can easily be found online, we consider that the reference to the original study is enough here (L333).

L402: ‘ interannual variability of deposit activity is not very strong’: I find it quite strong comparing e.g. 2017 and 2020.

Agreed, modified as (L413-414): 'There are pronounced seasonal differences (Fig. 9-11, S20-S22, Table S4) enhanced by the interannual variability of deposit activity (Fig. S19). '

L407-408: 'There are avalanches all year round over the Mt Blanc massif, but with a higher activity between January and July (Fig. 9).': It is very difficult to see this trend in Fig. 9. One has to search the rectangles with lighter red and then sort of count how many other rectangles there are for each column to be able to compare the numbers of avalanches over time.. I suggest to either improve the figure or rather refer to Fig. S23-26, in which this information can be seen more clearly.. The years, at least for the Mt. Blanc differ very much for each year (e.g. 2019 has very low numbers of avalanches compared to other years). Where are the same figures for the other 2 regions?

As suggested, we've now referred to Fig. S24-S28 (L420). These figures were meant to show the comparison with the avalanche forecast from Meteo France and as such we did not include these figures for the Hispar and Everest sites to not overburden the SI, given that the patterns for these 2 sites are also much clearer in Fig. 9-10.

L449: no -> little

Changed as suggested (L462)

L458: is -> are?

'Snow' is singular. We've kept 'is'. (L471)

L462-464: 'We have observed this smoothing to require several weeks and even months before avalanches can be detected at the location of old deposits, while avalanche events are still occurring in the meantime (Fig. S12d)': I find it very unlikely that it takes weeks to observe new avalanches on top of the old ones. I rather think, the opposite is the case. Usually, large wind speeds and precipitation can make it hard to detect avalanches within the 6 or 12 days window, removing traces of the avalanches before the next image is provided.

This is what we would have expected as well, but in some cases (serac falls? Large wet snow avalanches?) the high backscatter values from the avalanche deposit signal remain visible for long time periods. We have added the following figure (showing an avalanche deposit visible during 8 months in the Hispar region) in the SI and indexed it in the main text to highlight this:

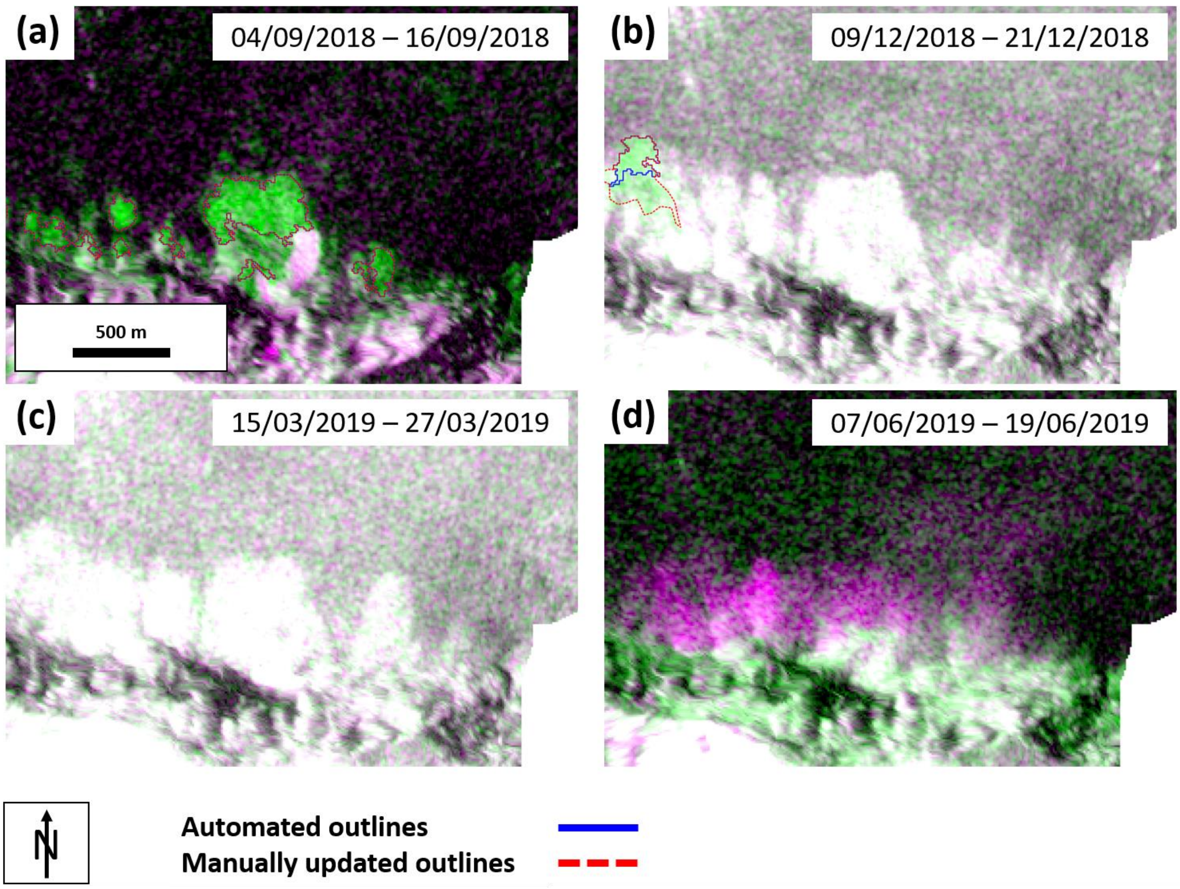

Figure S14: (a) Detection of avalanche deposits, with an older deposit visible in light purple, Hispar RGB composite 04/09/2018-16/09/2018. The high backscatter values from these avalanche deposits remained visible until 19/06/2019 (b-d). The different RGB bands range between -25 and -6 dB.

L476-480: This part needs to be modified. Currently, it sounds like it is easier to detect dry snow avalanches in Sentinel-1 images, and difficult to detect wet avalanches. PLease refer to Eckerstorfer te al. (2022): Performance of manual and automatic detection of dry snow avalanches in Sentinel-1 SAR images

Here the goal was to highlight the fact that 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. Apologies if this part is not clear, we have added the following sentence to make this clearer (L493-495):

'Therefore, while dry snow conditions lead to detectability limitations in Sentinel-1 images (Fig. 4), 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 (Eckerstorfer et al., 2022). '

L496-497: 'Our automated mapping therefore still requires manual edits,': so this would be a semi-automated mapping…

We've changed it to 'semi-automated' (L516)

L522-525: 'This difference is likely due to the detectability threshold, as well as the fact that recurring avalanches are likely to be missed if the surface roughness does not change between two events (Fig. S12c-d)': As mentioned above , I'would reformulate the sentence. High wind speed and new precipitation can mask avalanches within the rather large intervals of up to 12 days of Sentinel-1. Even 6 days can already be a long period if there are large wind speeds. I would also assume that avalanches on glaciers show differences than avalanches on snow.

Please refer to our response above and the figure we've added in the SI showing that high backscatter values from an avalanche deposit can remain visible for > 6 months. What you mention about the masking of deposits by wind/precipitation is however absolutely correct and we've added a sentence to make this clear (L477-478):

'In other cases, high wind speeds or new precipitation are likely to mask the deposits in the time interval of 6 to 12 days. '

L565: 'leading to high backscatter values that may reduce the detectability of avalanches, and especially slab avalanches (Fig. 4)': This needs to me modified: higher backscatter usually increases the detectability of avalanches

Here we referred to the high **background** backscatter values, not the backscatter values from the avalanches. We've modified the sentence to make this explicit (L587-589):

'This seasonality in avalanche activity could partly be explained by the presence of cold and dry snow at high elevations in the winter, leading to high backscatter background values that may reduce the detectability of avalanches, and especially slab avalanches (Fig. 4), in these upper reaches.'

Fig. S19: d-e -> d-f? F-g -> g-i?

Changed, thanks!