Reply to Reviewer #2

We thank the reviewer for the positive feedback, insightful comments and constructive suggestions.

The present manuscript describes an analysis that takes advantage of a large existing dataset of observed and mapped avalanches to explore the relationship between avalanche danger levels and avalanche occurrence in the Davos region of Switzerland. The research is of high quality, and the results contribute valuable insights to the current discussion on avalanche forecasting practices and consistency. The various analyses described in the manuscript offers useful information on the role avalanche size in avalanche forecasting, and the recommendation on the number of expected avalanches at danger rating level High has the potential to be the starting point for making avalanche forecasting more objective by replacing the existing qualitative descriptors in the danger scale with more objective quantitative measures.
The manuscript fits well with the mandate of The Cryosphere journal, and it will offer great value for avalanche safety researchers and practitioners. However, despite the obvious strength of the research, I believe that the manuscript has a few weaknesses that should be addressed before the manuscript is published. My concerns mainly relate to the presentation of the dataset correction procedures and the qualitative description of the results. I hope that my comment below are useful for making the manuscript even more impactful.

GENERAL COMMENTS AND SUGGESTIONS

Correction procedure (Lines 109-140)

Given that the objective of your paper is to examine the relationship between avalanche danger ratings and avalanche activity, manually changing danger ratings based on observed avalanche activity prior to analysis seems risky. While I do not necessarily disagree with the approach, I have the following recommendations for making it more transparent for the reader:

To put the number of days with corrected the danger ratings into perspective, it would be useful to provide readers with counts and proportions of danger ratings prior to correction right at the beginning of this section. This information is currently only available for the corrected danger ratings (Table 2). Having this information upfront would help readers to understand how much of the dataset was modified.

We agree and will add the following Table to the revised manuscript.

Table A1: Frequency of danger levels before and after corrections. Also given are the changes per danger level.

Danger level	Number of days before corrections	Change of danger level					Number of days after corrections
		-2	-1	0	+1	+2	
1–Low	306	-	-	303	1	2	303
2–Moderate	1809	-	0	1765	32	12	1766
3–Considerable	1367	0	0	1310	57		1366
4–High	47	0	21	24	2	-	94
5–Very High	4	1	1	2	-	-	4

The description of the correction procedure refers several times to the fact that avalanche activity was 'unusually high' or 'unusually low'. However, you do never explicitly specify what your expectations regarding avalanche activity actually are and how you determined your thresholds (one exception is the recoding of moderate days with AAI > 1.0). Being more explicit about your criteria would make your procedure more transparent.

We agree and regret the confusion. The procedure and the criteria are as follows:

1. We evaluated all days with danger levels 4–*High* or 5–*Very High* and a value of the AAI ≤ 1 ("zero or unusually low").
2. We evaluated all days with danger level 3–*Considerable* and a value of the AAI > 13.6 (after the first correction step), which was the median AAI for the days with danger level 4–*High* or 5–*Very High* (line 120).
3. We evaluated all days with danger level 2–Moderate and a value of the AAI > 1, which was the median AAI for the days with danger level 3–*Considerable.*
4. We evaluated all days with danger level 1–*Low* and value of the AAI > 1.

I personally found the description of the correction procedure somewhat difficult to follow due to many details described in the text. I wonder whether a diagram (e.g., flow chart) showing which danger ratings were changed to what and for what reason would help the reader to better understand the magnitude of your changes and their potential impact on the subsequent analysis.

Thanks for the suggestion. We hope that we can address your concern with Table A shown above and the correction procedure described above.

The numbers in your description of the changes applied to danger with High and Very High danger ratings (Lines 111-116) do not seem to add up properly.

Thanks for checking. We will do as well and hopefully the procedure will become easier to follow with the addition of Table A.

Overall, you changed the danger rating in 122 of 3533 days, which only amounts to 3.5%. This seems like a rather small amount and my initially thought was that the correction procedure was unnecessarily complicated given that it will likely only have a minor impact on the analysis. However, it represents 9% of the days with avalanche activity, and, if I understand your descriptions correct, the number of days with danger ratings High and Very High changed substantially through the correction procedure. The number of days with a High danger rating was first reduced from 44 to 26 (-18) (Line 116) and then increased again to 94 (+68) (Line 141). This means that only 28% of the days with High danger ratings in the analysis dataset were originally assigned a High danger rating. Given that the High danger rating sample plays an important role in the subsequent analysis, I believe that the impact of the correction procedure on the nature of the dataset should be described more clearly.

The main effect of the correction procedure is that the median AAI for days with 4–*High* increased from about 10 to about 21. For days with 3–*Considerable*, the median AAI was 1 and did not change due to the corrections. Hence, the difference in avalanche activity between 3–*Considerable* and 4–*High* was already very prominent before the correction procedure (before and after the corrections: *U*-test, $p < 0.001$). With regard to avalanche size, the effects are less prominent. For instance, the finding that size 2 avalanches are the most frequent ones will not change due to the corrections.

A brief discussion of the potential effect of the correction procedure on the analysis results in the discussion section would further acknowledge its impact. I think it is important to explicitly mention that there is potential for a bit of a circular argument here: You corrected the danger rating levels based on avalanche activity expectations to later analyze exactly this relationship.

We agree and will add a short paragraph in the Discussion section along the lines described above.

Description of analysis methods
The section titled 'Data and methods' only includes descriptions of the derivation of the avalanche size, the danger rating dataset and the quality control and correction procedure but seems to completely skip a description of the actual analysis approach and statistical methods employed. This seems rather unusual. I believe the manuscript would benefit a short overview of the analysis approach that describes the measures used (e.g., avalanche activity index, proportion of days with avalanches, etc.) and how they relate to the components of avalanche hazard (e.g., snow stability, frequency of locations) in the methods section.

We agree and will complement the Data and Methods section as suggested.

Statistical support for qualitative descriptions
Much of the description of the observed patterns are rather qualitative with some statistical tests here and there. I am wondering whether some of the statement could be supported with statistical test statistics. I believe that this would considerable strengthen the power of the manuscript.

We agree and are fully aware that we did not provide many statistical test results. We did so since we believe that the analysis is simple and the data actually speak for itself. However, we will add some more statistical test results in the revised manuscript to better support some of the main findings.

Description of study area
On line 377, you provide recommendations about the number of expected avalanches at danger rating level 4-High (at least 10 per 100 km2), and on line 416, you suggest that the term "many avalanches" should mean on the order of at least about 10 avalanche per 100 km2. I believe that this is an interesting result. However, while you highlight that avalanche occurrence probability depends on scale as it is a combination of stability, its distribution within the forecast area and the size of the forecast area, it seems to me that the nature of the terrain in the forecast region would also have a substantial impact on the suggested number. I therefore wonder whether a more detailed description of the nature of the avalanche terrain in the study area (e.g., number of avalanche paths of different size, total extent of avalanche terrain) would offer valuable context for understanding the results and recommendations.

We agree that the type of terrain certainly affects avalanche activity. We will provide the proportion of avalanche terrain in the revised manuscript.

Insight into avalanche warning practices
In several places in this manuscript, you comment on the somewhat unexpected differences in the observed numbers of dry and wet avalanches at the same danger rating level (e.g., Lines 323-324, Lines 375-376). However, there is no explicit statement in the discussion or conclusion section that points out that these observations might indicate inconsistencies in forecasting practices. I think that a statement like this would fit nicely with the recent literature on avalanche forecasting inconsistency and further contribute to this research.

Thanks for that suggestion; we do actually mention it in the Conclusions (line 414), but will check whether we can include it in the Discussion section as well.

LINE-SPECIFIC COMMENTS
1. ***Line 95 – Figure 1*** *The exponential increase presented in Figure 1 seems to be the direct consequence of the classification criteria presented in Table 1. I wonder whether plotting the log of avalanche area versus avalanche size class would be more useful to highlight that the approach*

classifies avalanche in the spirit of the Canadian size classification. The number of avalanches per class shown in the chart do not add up to the total number of avalanches given in the caption.

We have deliberately chosen the length, as this measure is the one most practitioners in Europe can well relate to. Also, the length is included in the EAWS definition of avalanche sizes. In addition, we provide the median area per size class in Table 1.
Thanks for checking the numbers. There is indeed an error since for some of the very small avalanches a meaningful value of length could not be derived so that the total number reduces to 13,802.

*2. **Line 100** It might be useful to explicitly state that the weight of 0.81 is appropriate because it is highly likely that the avalanches without known triggers were likely natural avalanches.*

We will state that explicitly.

*3. **Line 191 – Table 3** It might be useful to add row percentages to the columns to better highlight the relationship between avalanche size distribution and danger rating.*

Thanks for the suggestion. We will consider adding the percentage values; by the way, they are already shown in Figure 3.

*4. **Line 192** The Kruskal-Wallis test only indicates whether there are any differences in the avalanche size distributions among all danger rating levels. You could follow-up with pairwise Wilcoxon rank-sum tests between adjacent danger rating levels to determine where exactly the differences are.*

Thanks for the suggestion. We have actually done so and considered all pairwise combinations, but missed to report the results. We will do so in the revised manuscript.

*5. **Line 196 – Figure 3** I think it would be best if the proportion scales in all charts would range from 0 to 1 and be styled the same.*

We agree and change the scale on the y-axis in Figure 3a.

*6. **Line 196 – Figure 3c** If I understand your analysis correctly, Fig. 3c depicts average or median number of avalanches of different sizes per day at different avalanche danger levels. A bar chart does not seem an appropriate way to display this information as bar charts are typically used to depict proportions. Populating the same layout grid with a series of box plots instead of vertical bars would represent not only the magnitude difference in number of avalanches between danger ratings and avalanche size, but also the range within the observations.*

Thanks for your suggestion. We prepared the box plots for the number of avalanches, but realize it would not be possible to provide all data in one graph as you suggest. Moreover, the information would be hard to depict since the scale on the y-axis would range from 0 to 250. Therefore, we prefer to keep Figure 3c as presented.

*7. **Line 209** Can the statement "The median as well as the 90-percentile avalanche length did not increase for the danger levels 1-Low to 4-High." be substantiated with a statistical test result?*

We will provide more details on the median and 90 percentile (see Figure below) The median length was 164, 154, 163 and 198 m for the danger levels 1–*Low* to 4–*High*, respectively. The corresponding 90 percentile lengths are: 527, 449, 448, 566 m.

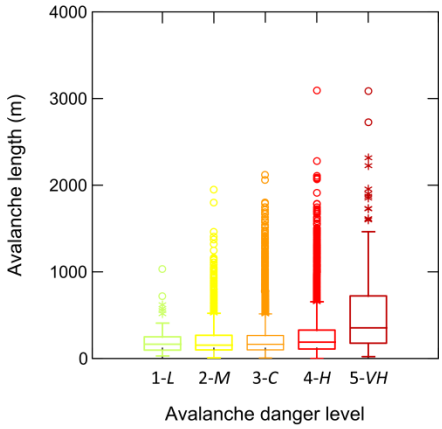

Figure A1: Avalanche length per avalanche danger level (N=13'745).

*8. **Line 217 and Figure 4a** It seems odd to combine avalanches with unknown snow conditions with the mixed category in the snow conditions analysis (Section 3.3) as these are very different categories. I think it would be better to leave these avalanches out of the analysis all together.*

Based on our experience the unknown conditions are often related to mixed conditions.

*9. **Line 227 and Figure 4d** The difference in the number of avalanches per day between dry and wet avalanches under avalanche danger level 1-Low shown in Fig. 4d seems minute. Can this statement be supported with a statistical test?*

As stated in lines 226-229 there were distinct differences between days with wet-snow avalanches and days with dry-snow avalanche on days when the danger level was 1–*Low* – though the number of cases is rather small:
On the 10 days with natural wet-snow avalanches, the number of avalanches was 18 and the total AAI was 15.2. Whereas on the 6 days with natural dry-snow avalanches, the number of avalanches was 8 and the total AAI was 0.71. As stated, the AAI was more than ten times larger.
On the other hand, in Figure 4d the average number of avalanches per day is shown, which is 1.3 for dry-snow and 1.8 for wet-snow conditions; this difference is not particularly large as you point out and statistically not significant.

*10. **Line 233** I assume that this discussion should be referring to the AVERAGE or MEDIAN number of avalanches per day.*

We will specify and add average in the revised manuscript.

*11. **Line 241 – Figure 4d** Same comment as for Figure 3c*

Please see our response to point #6 above.

*12. **Line 253** The statement "… size 4 avalanches were five times more frequent among the natural than the human-triggered avalanche." does not seem to be supported by Figure 5a.*

We agree that our statement was not adequate since it refers to the absolute number of size 4 avalanches. However, as there are three times more natural than human-triggered avalanches the proportions are 2.9 % vs. 1.6 % , still a significant difference (proportion test, p = 0.02). We will re-word the statement in the revised manuscript.

*13. **Line 256 – Figure 5c** Same comment as for Figure 3c and Figure 4d*

Please see our response to point #6 above.

*14. **Lines 268-273** Same comment as earlier regarding the average/median number of avalanches per day.*

Thanks for pointing this out, we will specify that the number refers to the average.

*15. **Line 293** I think it would be useful for the reader if the fact that no temporal trends in the avalanche size distribution were detected in the analysis dataset was included and substantiated in the initial description of the dataset.*

We may reconsider where to add this statement. We think it can be in either of the two sections.

*16. **Lines 316-324** It seems to me that some of the explanations of the data correction procedure described in this paragraph should be included in the methods section.*

As we will provide more details in the Methods section on the correction procedure (see above), we will revisit this paragraph and check where the explanations are needed.

*17. **Lines 340-346** In this section, you refer to the avalanche size distribution of human triggered avalanches, and the topic comes up again on Line 407. However, I did not find this explicit analysis in your manuscript. If I read your manuscript correctly, you only analyzed the number of human triggered avalanches under different danger ratings but not their size distributions. It seems to me that such an analysis would nicely complement your existing analyses.*

Harvey (2002) considered avalanches that caused damage to either people, infrastructure or forest, and reported avalanche size. Hence, we discuss his study to compare his results to our findings.
In fact, we have analysed the size distribution for human-triggered avalanches at the different danger levels. The distributions are very similar to the overall distribution shown in Figure 5a. For instance, at danger level 3–*Considerable* and 4–*High*, the proportion of size 3 avalanches is 0.74 and 0.73 (proportion test, p = 0.77). For size 4 avalanches the corresponding proportions are 0.14 and 0.19 (proportion test, p = 0.36). We will consider adding this information to the Results section.

*18. **Lines 355** Same comment as earlier regarding the average/median number of avalanches per day.*

Thanks, we will specify the number as the average in the revised manuscript.

*19. **Lines 378-381** The description of the results of Bründl et al (2019) is a bit confusing to me. What are the five frequency classes and how do they relate to the results presented in this paper?*

We will clarify the approach presented in Bründl et al. (2019).

20. ***Line 389*** *Add "and terrain choices." at the end of this sentence*

Thanks for the suggestion, we will add this additional factor.