Reply to Reviewer #1

We thank reviewer #1 for the thorough and detailed review of our manuscript, which we greatly appreciate. In the following, we address each of the points raised. Black text indicates the reviewer's comments. The blue text shows our responses to the comments.

General comments

This paper investigates the potential to forecast natural dry-snow avalanches with snowpack simulations. Avalanche observation data are used to train and validate models that predict the probability of natural avalanche occurrence, as well as the probability of different avalanche sizes. Comparing to benchmark models that only considered the amount of new snow, they show improved model performance by adding snowpack stratigraphy and stability information, especially in regions prone to persistent weak layers. As a final step, model predictions were validated with regional avalanche danger ratings, which illustrated the potential for this model to support avalanche forecasting.

The study is interesting, thoughtfully designed, and well written (especially considering the complexity of the subject and data sets). The research is relevant to the avalanche research and forecasting community and fits well within the scope of NHESS. I recommend publication after addressing the following comments.

Specific comments

My main comment is that it could be clearer how aspect dependent information was applied in the model training and validation. Overall, the data and methods are presented very clearly, but understanding how aspect information was applied required a fair bit of extra effort which I think could easily be improved by providing more details. Some examples:

- Consistent terminology. It would help to explicitly describe values as "aspect-specific" throughout the manuscript when referring to AvD/nAvDs. "aspect-specific" and "slope-specific" are used interchangeably (e.g., lines 255, 336).

  We will keep the term "aspect-specific" only in the definition of AvD/navD (Eq. 1).

- 3.1.1 could provide a clearer description of how avalanche days were counted by aspect in the training data. My interpretation is that for each day of the study period Y is calculated 4 x (number of stations) times and then each value of Y is matched to the explanatory variables for the corresponding virtual slope simulation to build the training dataset. Are the flat field simulations discarded?

  The flat field simulations are not used. - To describe more clearly which data were used, we will add an explanation.

- Line 150 states the height of new snow was independent of aspect, which I assume means these variables were taken from flat field simulations. But if this version of SNOWPACK models snow redistribution, then shouldn't your profiles have different amounts of HN on different aspects and lead to a higher likelihood of natural avalanches on lee aspects?

  Conventionally, the height of new snow is measured in the flat field. Consistent with this definition, the height of new snow provided by SNOWPACK is the same for the flat field and the slope simulations. The new snow amounts hn1d and hn3d considered in our study are therefore indeed independent of aspect.

In addition, we also considered the thickness of precipitation particle layers as a further parameter in our analysis. For this parameter, which should capture the amount of recently fallen including snow transport by wind, we employed the snow redistribution module of SNOWPACK.

- Can you comment on the impact of not considering aspect in the AvD definition used in the validation part of the study (i.e., line 253)? The model predicts aspect-specific probabilities, but these are evaluated on whether avalanches were observed anywhere in the region. The relevance of this assumption should be justified.

We regret that we were not clear; this is a misunderstanding, as aspect-specific information was also used for the AvD definition applied on the validation data. To make it clearer where the AvD definition differs from Eq. (1), Section 3.1.1., we will rephrase the paragraph where we describe the definition for the validation data set.

- Given SNOWPACK's virtual slopes have limited verification studies, it could be interesting to briefly comment on whether this study finds they add value relative to flat field simulations only. Are there any insights or recommendations about virtual slopes to share with future researchers?

We will add a paragraph in which we describe previous work on using slope simulations for the purpose of avalanche forecasting.

Technical comments

- Introduction: I appreciate how the introduction clearly addressed the limitations of using avalanche observations for model verification. The objective and structure of the paper is also very clear.
- Line 107: Can you state how many avalanches were included in the AV3 dataset (as done for AV1 and AV2)?

We will include the number of avalanches in the AV3 dataset in Sect. 2.1.3.

- Line 120: It may be better to mention wind transport was enabled here rather than line 152. Also, can you describe whether this SNOWPACK setup determines snowfall with precipitation gauges or snow height measurements (since new snow height is an important variable in this study)?

We will shift the information about snow transport by wind and add information on the SNOWPACK setup.

- Fig. 1: This figure (and the entire Data section) is organized in a way that makes it very easy to understand the different datasets used in various parts of the study. Thanks.

We are happy to hear that Fig. 1 is helpful.

- Fig. 2: Can you specify a temporal period for number of avalanche days N(AvD). For example, over a specific study period, seasonal average, or something else?

The N(AvD) were observed during the three winters 2019/2020 – 2021/2022 as described in Sect. 2.1.1.

- Lines 145-146: Some technical notes on choosing variable names and ICSSG standards. Typically, HN3D would suggest a 3-day observation interval, which is different from the sum of observations made at 1-day intervals (due to settlement). Am I correct that in this case, you are summing HN1D values rather than using SNOWPACK to directly get a height of 3-day snowfall? Also, would it be more accurate to call the precipitation particle variable a "thickness" rather than a "depth"? I would interpret depth as the distance from the deepest PP/DF layer to the surface, but summing thicknesses could be smaller if there are other grain types above (e.g., RG/MF). If that is the case, then ICSSG symbol for thickness is D rather than z. Similarly, the standard symbol for grain type is F. I don't think changing these variables is essential, just something to consider.

  Yes, it is correct that hn3d refers to the sum of hn1d values over three days, which we also stated in the manuscript ("calculated as the sum of three consecutive hn1d-values», l.145). We agree that "thickness of precipitation particle layers" is a more appropriate term than "depth of precipitation particles" and will change the wording throughout the manuscript. Thank you for the comment regarding the ICSSG symbols. However, we prefer to stick to z_{pp} as we think that the definition in line 145 and the illustration given in Figure 3 is sufficient to understand the meaning of the variable.

- Line 148: Can you briefly describe the main inputs of sn38 and how these differ from the inputs to Punstable?

  We will add the definition of the natural stability index, which for each snow layer describes the ratio of shear strength to the shear stress exerted by the overlying slab.

- Fig. 1: It could be clearer here how many avalanche days were computed (i.e., one per station per day?) and how the aspect and elevation information was used.

  We will describe more clearly how many avalanche days were calculated per station, day and aspect.

- Line 181: Variables *st* and *asp* are defined but not used in manuscript.

  We will remove these abbreviations.

- 209: Variable *thr* is not defined and appears to only be used in the Appendix.

  We abbreviate the best-splitting threshold using thr. This abbreviation will be introduced.

- Line 208: The subsets are not just based on splits of the AV1 data, but also the snowpack data (i.e., critical grain type). Perhaps it's better to say "we split the training data…".

  We will change the wording according to your recommendation and specify the term "training data" in the previous paragraph.

- Line 220: Why was sn38 only used in the binary model and not the continuous model? It would help to list the x variables in the beginning of this subsection (since the variables used in the binary model aren't described either).

Due to the limited discriminatory power of sn38 in the binary approach, this variable was not considered further in the subsequent development of continuous models. We will include a sentence to clarify this.

- Line 228 and 243: The idea behind the BS+ metrics could be a bit clearer. Throughout the manuscript they are described as "minority class", "positive observed outcomes", "positive events", "when condition is fulfilled". I recommend a sentence to explain why these subsets are relevant in the methods and then choosing consistent terms that are more descriptive (e.g., only days with observed avalanches) to use throughout the manuscript.

  Thank you for this recommendation. We will describe the idea behind the BS+ score more clearly and use more descriptive wording in the revised version of the manuscript.

- Line 252: Why were aspect and elevation neglected when determining AvD here?

  Unfortunately, the region of Davos where we have aspect- and elevation-information does not cover the whole area (i.e. 5000 km^2) that we need to consider in the definition of an AvD. As the avalanche observations outside the region of Davos sometimes lacked information on aspect and elevation, we had to adjust the gap-check requirements. While aspect and elevation were considered for the avalanches from the region of Davos, for the two surrounding regions, this was not the case. A day was labeled as an aspect-specific AvD, if the AAI in the region of Davos was larger than 0.01 for the respective aspect and within an elevation band of ±250 m around the AWS WFJ, and if at least one natural dry-snow avalanche was observed within each of the two surrounding regions (1000 km^2 and 5000 km^2) regardless of aspect and elevation. The definition of an AvD was thus slightly adapted compared to Eq. (1) due to the lack of consistent information on aspect and elevation of the observed avalanches within the two larger surrounding regions. We will explain this more clearly in the revised version of the manuscript.

- Line 284: According to Table A2 the median is 12 cm not 13 cm.

  Thank you for pointing out this error. 12 cm would have been correct, as shown in Figure 4 and Table A2. We will change accordingly.

- Figure 4 is not cited anywhere in the manuscript.

  You might have overseen that Figure 4 is mentioned in line 286/7, which was probably due to the line break.

- Fig. 5: The 2020 season seems to stand out as anomalous in these figures, was there something unique about that season, such as the prevalence or absence of persistent weak layers?

  The 2020 season was a rather mild winter, with comparably few avalanche periods (Trachsel et al., 2020). This is also shown in Table 3. However, we have no conclusive information on whether persistent weak layers were of less concern during this winter compared to the two other seasons. Table 3 does not show evidence for either assumption.

- Line 316: Should this be dataset AV2 instead of AV3?

  Yes, thank you for pointing out this error, this should be data set AV2.

- 7: The HN3D model is presented before the Pcrit model in Fig 5, while it's the other order in Figs 7 and 8. Consistent ordering would be ideal.

  We will change the order of the plots in Figure 5, with the P_{crit}-model being shown before the hn3d-model.

- Lines 371-380: This paragraph is a little confusing because not all the values discussed are shown in Fig. 9b, which appears to be due to whether the Pdeep < 0.77 cases are included or not. Which case is more relevant to present here? The text could be clearer about which case is being discussed and which case is shown in the figure.

  We will restructure this section to make it easier to understand.

- Line 395: I am particularly interested in how the extreme cases of widespread versus no activity impact the results at danger level 3. Since the models are fit to these extreme cases, they do not capture the "natural avalanches possible in certain areas" conditions experienced at danger level 3. However, it seems the models produce a desirable result with a wide range of P(AvD) observed at considerable danger in Fig. 8, which speaks to the range of conditions and uncertainty about natural avalanches experienced at considerable danger.

  To develop the models for the prediction of natural avalanche activity, we put a lot of effort into the quality of the target variable data (AvD vs. nAvD). By cross-checking the activity in smaller regions (250 km^2) with the activity in larger regions (1000 and 5000 km^2) and retaining only the data points at the two ends of the stability spectrum (widespread vs. no activity), we aimed at reducing the error contained in the visual observations of local observers. To overcome this binary approach, we fitted continuous functions describing the probability for an AvD, P(AvD), forcing the models to be close to 0 at nAvD-data points and close to 1 at AvD- data points. With the very broad definition for danger level 3 (considerable), some of the validation data points are closer to the conditions of AvD data points in terms of P_{crit} or hn3d, and others may be closer to the conditions of the nAvD training data points. This explains the wide range of P(AvD)-values obtained at danger level 3 (considerable). Techel et al. (2022) also demonstrated the capability of P_{crit} to capture the differences in stability regarding sub-levels describing variations within danger levels.

- Line 446: While not published in a peer-reviewed journal, Bellaire & Jamieson (2013) estimated avalanche size from simulated profiles in their 2013 ISSW paper (Fig. 2 is similar to your Fig. 6a).

  Thank you for pointing this out, we were not aware of this study and will provide reference to this work in the revised version of the manuscript.

- Fig. 10: This caption could provide higher-level description of what is shown, as it is difficult to understand without also reading the text. Also, what exactly do the contour lines show? Steps in density distributions?

  Yes, the contour lines indicate steps in density distributions. More specifically, the respective outermost contour line represents the (0, 0.1] percentile interval, and the innermost contour line the (0.9, 1.0] interval. We will add an explanation in this regard in the caption to Figure 10.

- Fig. 11: Why are the pink bars for Rutschblock (T2) only shown for low danger and not others?

  Thank you for pointing this out. This was in fact an error which will be corrected.

- Conclusion: Given the stated objective of the paper was to "*investigate whether the instability model developed by Mayer et al. (2022) applied to one dimensional SNOWPACK simulations can be used to predict natural dry-snow avalanches*", I think this question can be more directly answered somewhere in the conclusions to summarize what was learned and how others could apply the model.

  To answer this question more directly, we will add a sentence at the beginning of the Conclusion section.

References

Bellaire, Sascha, and Bruce Jamieson. "On estimating avalanche danger from simulated snow profiles." In Proceedings of the International Snow Science Workshop, Grenoble–Chamonix Mont-Blanc, pp. 7-11. 2013. https://arc.lib.montana.edu/snow-science/item.php?id=1740

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.

Techel, F., Mayer, S., Pérez-Guillén, C., Schmudlach, G., and Winkler, K.: On the correlation between a sub-level qualifier refining the danger level with observations and models relating to the contributing factors of avalanche danger, Natural Hazards and Earth System Sciences, pp. 1911–1930, https://doi.org/10.5194/nhess-22-1911-2022, 2022.

Trachsel, J.; Zweifel, B.; Techel, F.; Marty, C.; Stucki, T., 2020: Schnee und Lawinen in den Schweizer Alpen. Hydrologisches Jahr 2019/20. WSL Ber. 101: 76 S.