Reply to Reviewer #3

We thank Christoph Mitterer 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 reviewers' comments. The blue text shows our responses to the comments.

Summary

The authors introduce and investigate the predictive skills of two new models: an avalanche day predictor and an avalanche size estimator. Both are based on snow cover model results and observed avalanche activity and especially designed and valid for natural dry-snow avalanche events. The models are trained and tested on individual data sets. The training data consists of two large data sets of avalanche observations and snowpack simulations using the 1-D physics based model SNOWPACK. The validation data sets include avalanche observations, avalanche danger level assessments and snow cover simulation results.
The model for the avalanche day predictor focuses on a RandomForest (RF) model based on derivates of SNOWPACK variables presented by the authors team very recently (Mayer et al., 2022) and is trained using a 3-years data set of avalanche observations data covering the entire Swiss Alps. The estimator model is trained with observed data only, but includes a very large data set of avalanche observations covering 30 years for the Swiss Alps.
To test the performance of both models, the authors test their novel approaches against a very common benchmark model (Height of the three-day sum of new snow; hn3d) and validate their findings against a fully independent data set of avalanche observations. Large parts of the interpretation and discussion is done by comparing the results to a 21-years data set of regional avalanche danger levels assessment.

Results show good predictive performance results for characterising days with natural dry-snow avalanche activity; especially when natural dry-snow avalanche activity was driven by shallow snowpacks consisting predominantly by persistent weak layers. Compared to the simple benchmark model, performance is very similar and increases slightly when the new approach is combined with the benchmark model. The results for the avalanche size estimator when tested again the independent data set are also encouraging. Finally, both approaches are compared to the regional avalanche danger levels.

Evaluation

The approaches are not fully novel but connect skilfully recent advances with a large data set that represents the currently available golden standard within the avalanche research community. Research objectives are very clear and concise; methods are well designed, but also how the authors addressed the objectives are mostly well described and easy to grasp.
Language is concise and the manuscript is well written. Major findings are very relevant to the avalanche research and forecasting community. The content fits very nicely into NHESS. There are a few parts of the discussion and interpretation of results that need revision. Having addressed the below stated comments, I recommend publication.

General comments

I have the following general comments:

- In the Abstract (Line 1) and Introduction (Lines 19-21) you give the reader the feeling that you would like to tackle both, very local (path scale) and regional avalanche forecasting. When reading the full manuscript, it becomes obvious that you address regional avalanche forecasting (e.g., Section 3.1.1 or the fact that you compare and discuss results to the regional avalanche danger level). Please be more specific in that case and drop the connection to the local avalanche forecasting.

We will make it clearer in Abstract and Introduction that we are aiming at predicting regional-scale avalanche activity.

- The authors compare their avalanche day predictor model to the conventional natural stability index on a 38° steep slope (sn38) which was one of the few indices developed within SNOWPACK to better assess natural avalanche activity (Lehning et al., 2004). The new model outperforms the sn38 (Figure 4), but the authors do not really discuss why this is the case. They state that "The poor performance of sn38 is in line with other studies (Schweizer et al., 2006; Jamieson et al., 2007)." To my knowledge the first study tackles sk38 only, thus skier-triggered scenarios and not spontaneous avalanche activity. The second study compares the natural stability index (sn) based on measurements to natural avalanche activity in the surroundings of the study plot – which is a significant different approach to the one presented (modelled vs. measured). In fact – to my knowledge the only qualitative investigation on the performance of sn38 is given in Lehning et al. (2004). There the authors show reasonable results.
  So, it remains difficult to set the presented low performance skills into context. Therefore, it would be very interesting and valuable to tackle in more detail the question, why sn38 has such a low performance compared to the avalanche day predictor model. Both, the conventional and the novel approach, are heavily parameterised by snow density and almost rely on the same concepts: the most important variable for the instability model by Mayer et al. (2022), the viscous deformation rate, shares the identical input parameter as the natural stability index, namely natural shear strength – which in turn is parametrised using snow density. 4 out of the 5 most important features building the RF model rely on snow density. It would be very beneficial for the community if the authors could e.g. use the Weissfluhjoch data set to shed some light into this topic. I know that this represents large efforts, but I believe it would give even more impact to presented results.

We included the natural stability index sn38 to evaluate the new approach based on the output of the instability model in direct comparison with such a benchmark model. The index sn38 is defined as the ratio of weak layer shear strength divided by the shear stress due to the overlying slab. This criterion seems well-suited to model natural avalanche activity from a physical point of view. However, the parametrization of this simple criterion within SNOWPACK has some weaknesses: While the shear stress can be simply calculated from the load and thus only inherits the errors from estimating precipitation mass based on measured snow depths, shear strength is a rather complex microstructural parameter. The current SNOWPACK parametrization of shear strength is based on density and grain type (Jamieson and Johnston, 2001), which may not be sufficient to capture the influence of microstructure as also pointed out by Richter (2020). In particular, the evolution of the SNOWPACK shear strength over time only depends on density if grain type does not change.
In their comparison of modelled sn38 values with forecasted danger levels, Lehning et al. (2004) pointed out a tendency of the natural stability index to indicate maximum weaknesses close to the ground which they attributed to an underestimation of the increase in shear strength due to its simplified parametrization within SNOWPACK. An underestimation of the increase in shear strength could also be an explanation for the low precision of the sn38, which led to many false alarms in our study. We thus agree that your question may warrant a

detailed analysis of sn38 but this would be out of the scope of the present study, as field data including observed snow strength would be necessary.

Finally, you are correct when stating that Schweizer et al. (2006) analyzed the skier stability index sk38 rather than the natural stability index. We included a reference to this paper as they stated “the natural stability index is a poor predictor for spontaneous releases”. But as they did not prove this statement within their study, we will delete this reference in the revised version of the manuscript. Instead, we will refer to Reuter et al. (2022) who also demonstrated a rather poor performance of the sn38 and showed that using time derivatives of this index has a higher predictive power. This is in line with Jamieson et al. (2007) who analyzed sn38 based on field measurements and also concluded that critical values of stability indices are less useful than their trends.

- Interconnected to the comment above: How and why is the instability model suited to predict natural avalanche activity, even though it is heavily trained on data that mostly represents skier-triggered avalanche activity?

  It is correct that the instability model predicts the probability of (potential) instability, related to human-triggered avalanches. However, the input features of the instability model describing the weak layer (e.g. grain size) and the overlying slab (e.g. ratio of the mean slab density and the mean slab grain size) are also related to the release of natural avalanches as mentioned in line 55 f. While not mentioned in Mayer et al. (2022), we would like to point out that often natural avalanches were reported in the vicinity of the “unstable” profiles used for the training of the instability model. And lastly, we think that our results are in line with our understanding of snowpack stability: according to Mayer et al. (2022), profiles are classified as unstable if pcrit > 0.77, and – in this study – we find natural avalanches are likely if pcrit > 0.82. Thus, the latter is a subset of the profiles classified as unstable from a human-triggering perspective.

- The discussion regarding the comparison to the regional avalanche danger levels is nice but needs in a few points a much broader approach: The statement that danger level 3-Considerable needs sub-levels could also be reversed in the fact that the Swiss forecaster need to train themselves more in order to transfer the overlapping parts into the neighbouring classes instead of increasing the level of discretization. Can you comment on that please. Figures 10-11 are very important but touched very shortly. I would appreciate more details here.

  We agree that our results are based on how the avalanche danger levels are assigned in the Swiss avalanche bulletin. However, the broad range of what is considered danger level 3-Considerable, shown in Figure 10, seems not to be a Swiss bias but rather inherent to the broad definition of the danger levels. This broad range, from a rather low proportion to a rather high proportion of locations where natural avalanches may initiate, is also mirrored in the EAWS matrix (EAWS, 2023), a look-up table assisting forecasters with danger level assessments in Europe. The EAWS matrix suggests for level 3 (considerable) combinations ranging from “Many locations with very poor stability (which corresponds to natural avalanches) exist. Avalanches can reach size 2.” and “(Nearly) no locations with very poor stability but some locations with poor stability (human triggered avalanches are typical for this class) exist.” The first definition has a tendency towards level 4 (high), which is reflected both in the figure and in the EAWS matrix. This broad range of 3 (considerable) also mirrors what is described in Swiss avalanche forecasts (Hutter et al., 2022), again in line with the EAWS matrix, and what can be seen, when comparing it to actual observations, stratified by sub-levels as used in Switzerland (as in Techel et al., 2022).

Specific and technical comments

- 1 (Lines 48-49): Why don't you address all danger levels here? In fact, at danger level 3-Considerable the definition mentions: In certain situations, some large, and in isolated cases very large natural avalanches are possible.

  On purpose, we only addressed the respective danger levels for which natural avalanches are generally expected (4 (high) and 5 (very high)) or not expected (1 (low) and 2 (moderate)). At 3-Considerable, the range is much wider, including situations when avalanches are primarily triggered by additional load as a skier and situations when natural avalanches are typical (e.g., EAWS, 2023).

- 1.1. Line 96: Counter for Table Numbering is not in sequential order. You mention Table 3 before you mention Table 2 in the text.

  Thank you for pointing this out. We will remove the first reference to Table 3, as we want the table to remain close to the Results section. Thus, Table 2 will be referred to before Table 3.

- 2 It would be very helpful to introduce a new habit when using SNOWPACK simulations, namely placing the INI-Files of the model runs into the Appendix.

  We will publish the INI-Files of SNOWPACK runs together with the data used to build the new models.

- 1.1 Definition of avalanche days and non-avalanche days (Lines 197-198): Does that mean that your training data set has no AvD due to a size 4 avalanche?

  No, the values and ranges indicated represent the median and the interquartile range (IQR). The latter includes only the 25-75%-range, thus, larger values, as for instance size 4 avalanches as the largest avalanche, exist.

- 3.1.2. Avalanche size estimator: Could you please specify in a little more detail, why you have chosen exactly this approach?

  As our instability model is based on assessing the stability of a simulated profile, we focused on fracture depth as an indicator of avalanche size, since we thought it to be feasible to estimate fracture depth from a simulated profile.

- 4.2 Line 376: Figure 9e does not exist.

  Thank you for pointing this out. It should read Figure 9b.

References

EAWS, 2023: EAWS Matrix. https://www.avalanches.org/wp-content/uploads/2022/12/EAWS-Matrix-EN.pdf

Hutter, V., Techel, F., and Purves, R. S.: How is avalanche danger described in textual descriptions in avalanche forecasts in Switzerland? Consistency between forecasters and avalanche danger, Natural Hazards and Earth System Sciences, 21, 3879–3897, 2021.

Jamieson, J. B., and C. D. Johnston, Evaluation of the shear frame test for weak snowpack layers, Annals of Glaciology, 32, 59– 68, 2001.

Jamieson, J. B., Zeidler, A., and Brown, C.: Explanation and limitations of study plot stability indices for forecasting dry snow slab avalanches in surrounding terrain, Cold Regions Science and Technology, 50, 23–34, 2007.

Lehning, M., Fierz, C., Brown, R. L., and Jamieson, J. B.: Modeling instability for the snow cover model SNOWPACK, Annals of Glaciology, 38, 331–338, 2004.

Mayer, S., van Herwijnen, A., Techel, F., and Schweizer, J.: A random forest model to assess snow instability from simulated snow stratigraphy, The Cryosphere, 16, 4593–4615, 2022.

Reuter, B., Viallon-Galinier, L., Horton, S., van Herwijnen, A., Mayer, S., Hagenmuller, P., & Morin, S.: Characterizing snow instability with avalanche problem types derived from snow cover simulations. Cold Regions Science and Technology, 194, 103462, 2022.

Richter, Bettina. Improving numerical avalanche forecasting with spatial snow cover modeling. Dissertation. ETH Zurich, 2020.

Schweizer, J., Bellaire, S., Fierz, C., Lehning, M., and Pielmeier, C.: Evaluating and improving the stability predictions of the snow cover model SNOWPACK, Cold Regions Science and Technology, 46, 52–59, 2006.

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, 2022.