Editor's comments

«Does combining modelled snowpack stability with machine learning help with predicting avalanche activity?"

submitted by Viallon-Galinier, Hagenmuller and Eckert to *The Cryosphere*

The paper describes a study combining weather data, modelled snow data and modeled snowpack stability data with observations of avalanches employing the random forest method to forecast avalanche activity (avalanche/non-avalanche days) for 24 for elevation and aspect segments.

The manuscript is easy to follow; the research objectives are relevant and doubtless challenging. The manuscript is a valuable contribution to the field of numerical avalanche forecasting.

As reviewers have pointed out – and I share their concerns – it is questionable whether the avalanche dataset selected is fully suitable to reach the goals of the study. With less than 3000 avalanches in 58 years observed in 110 avalanche paths, about every second winter an avalanche is observed in a specific path. Whether this frequency is suitable to predict daily avalanche activity in 24 aspect and elevation segments remains to be shown.

This said I do neither oppose the use of the dataset and nor question the study overall, I simply invite the authors to reflect whether the question put in the title can be adequately answered and how well the results can be generalized.

On the title, for instance, my recommendation is not to word it as a question, or at least put the question in a way that it can be answered. For instance, to answer the question you would need to consider combinations other than stated in the title. However, you only explore one method. I suspect you primarily wonder about the value of modeled stability information (compared to, for instance, weather and snow data only), in particular in view of your goal to forecast for 24 aspect and elevation segments. By the way, stability information, mostly not modelled, has been used in several previous studies, even back in some early work on numerical forecasting in the 1990ies – and was shown to be important.

Below you will find some more specific comments:

Lines 13-14: I think it would be valuable to show how the model would perform with a simple target variable (not considering 24 aspect and elevation segments). Moreover, I think "cutting-edge" should not be related to data here.

Lines 20-21: I recommend rewording. Long-term and short-term are somewhat oddly used here. I suppose you refer to hazard assessment in the context of hazard mapping vs. avalanche forecasting. Hazard mapping and avalanche forecasting are both mitigation measures having a long-term and a short-term effect, respectively.

Line 24: I am not sure whether you focus on forecasting or prediction.

Line 34: Please cite the corresponding peer-reviewed publication rather than a magazine article (Schweizer and Jamieson, 2007).

Line 53: Some early models that used stability information were described by Schweizer et al. (1994) and Schweizer and Föhn (1996).

Line 58: Schirmer et al. (2009) used output variables from a numerical snow cover model as input for statistical forecasting.

Line 65: It has been shown that the type of data (input variables) can at least be as important as the method applied. For instance, regardless of the method, with meteorological data as input only, the performance is always limited (e.g., Schirmer et al., 2009)

Lines 66-67: Suggest rewording to better reflect the aim.

Lines 67-68: Suggest rewording so that the three different types of input data become more obvious.

Line 79: In your replies to the reviewers' comments, you point out many uncertainties related to the dataset. Hence, I wonder whether you can state that the avalanche observations are particularly reliable.

Lines 97-98: As you describe above, EPA includes primarily large natural avalanches in selected path. It is questionable whether these data describe the overall avalanche activity in the area. There is much more potential avalanche terrain in the area and almost always when there are large or very large avalanches, there will be many small and medium-sized avalanches as well. On the other hand, I suspect, there are also numerous days when no very large natural avalanches occurred, but still many small to medium-sized natural avalanches at higher elevations. Hence, what is observed and recorded in EPA cannot provide a complete picture of natural avalanche activity. All these additional smaller natural avalanches are very likely also relevant for operational avalanche forecasting. Hence, in conclusion, I suggest you reword and specify the statement.

Line 133: I recommend you add "for a given depth" or "a given layer interface"

Line 146&147: I recommend replacing humid by wet or moist.

Line 156: As far as I remember, Reuter et al. (2022) have recently applied the a time-dependent index to describe natural failure initiation.

Line 160: It is confusing to the reader that changes in snow depth are associated with stability indices. Hence, I suggest regrouping the input variables.

Line 166: I agree that the model resolution you select is more demanding. However, the question remains unanswered whether this additional sophistication represents an added value given the overall rather poor performance of the model from an operational point of view.

Lines 174-175: I suggest using more descriptive variable category names, for instance, weather, snow and stability. Certainly, "Bulk" does not seem appropriate to me. Also, in Table 1, you refer to derivates, which seem to include snow depth changes (though you only refer to dry snow indices). Please clarify.

Lines 219-221: I suggest you also mention that the recall is, commonly in the forecasting context, called the probability of detection (POD) or hit rate. This would ease comparison with previous work.

Lines 221-222: I suggest you also mention that the false positive rate is called the false alarm rate (FAR).

Line 236: Random classification is usually associated with the diagonal in the ROC diagram.

Line 254: elevation

Lines 257-258: The example you provide is probably one of the most prominent avalanche winters in your dataset. There were three, well-known major avalanche cycles in late January and

February in 1999. It would certainly be interesting to know the performance in such a "catastrophic" winter. In addition, are these model results obtained with the model trained with all input variables?

Lines 262-263: Please adapt the caption in Figure 4a to the description provided here.

Line 270: It may be worth mentioning that sensitivity (75.3%) and specificity (100-23.6%) are similarly good, which is nice, but that due to the unbalanced dataset the precision is low (3.**3**%).

Lines 278-279: The text here, in the revised manuscript, is unchanged, but Figure 5 (previously Figure 4) changed, i.e. the importance scores are different now. Are all changes in Figure 5 reflected in the text?

Line 280: In Figure 5 it is indicated that there are 25 variables from the dry snow stability group whereas in the text you refer to 30.

Line 294: Here you denote snow depth and its derivatives as "bulk", in contrast to the variable categories described in Table 1. Please clarify and/or improve Table 1.

Lines 313-314: What means "e.g. 2021"? By the way, both studies are published now. Please update the references (Mayer et al., 2022; Pérez-Guillén et al., 2022).

Line 336. I am not aware of any country who explicitly specifies the avalanche danger in 24 aspect and elevation segments.

Lines 347ff: It seems odd to me not to consider snow depth and in particular new snow depth as meteorological variables. In any case, you cannot conclude that meteorology is irrelevant and that this contrasts with other studies that probably all considered new snow depth. Moreover, it is doubtful that the alleged difference stems from modelled vs. measured new snow depth. Most readers would probably agree that new snow depth (precipitation) is a meteorological driver of avalanche danger. Hence, it seems that your conclusion depends on your choice of categories and cannot be generalized.

Line 353: Isn't new snow depth the significant variable rather than snow depth?

Lines 559-360: I am not sure how this last sentence refers to the importance of new snow depth.

Lines 371-373: As far as I remember van Herwijnen et al. (2016) and van Herwijnen et al. (2018) showed that dry-snow avalanche activity was correlated with snowfall on times scales of 2 or more days, while for wet-snow avalanches the correlation was shorter and with energy input. Hence this seems to be in contrast with what you describe. However, I agree with your statement in the following sentence.

Line 383: Previously you referred to 2779 events, not 2518. Please clarify.

Lines 384-285: Please refer to my previous comment on "the representative screenshot of the overall avalanche activity". In addition, I am not sure I understand why in Haute Maurienne the scarcity of reported avalanche events is not a problem.

Line 412: I am not sure I understand what you mean here with "the interest of physics".

Lines 419-420: Please reword.

Line 429: In my understanding "avalanche prediction" means to predict the exact time and location of a single avalanche event. I think so far, we rather forecast avalanche activity.

Lines 434-435: I suggest rewording the statement: the combination proves to be valid.

Line 435: I guess new snow depth was a significant variable, not snow depth.

Line 436: As far as I remember, Jamieson and co-workers have shown that stability indices, though derived from measurements (not modeled), were valuable inputs for forecasting (e.g., Zeidler and Jamieson, 2004).

Line 439-440: While I agree that snow cover models and thereof derived stability information is valuable, I recall that your model is not that great in identifying avalanche prone situations. The precision is low. Please consider rewording.

Line 443: I suggest deleting "cutting-edge". The random forests method has become a rather standard tool over the course of the last decade. In addition, I am not sure you can call the data cutting edge.

Line 446: Similarly, not convinced the stability indices are that cutting-edge.

Davos, 5 January 2023
Jürg Schweizer.

References

Mayer, S., van Herwijnen, A., Techel, F. and Schweizer, J., 2022. A random forest model to assess snow instability from simulated snow stratigraphy. The Cryosphere 16(11): 4593-4615.

Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., van Herwijnen, A., Olevski, T., Obozinski, G., Pérez-Cruz, F. and Schweizer, J., 2022. Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland. Nat. Hazards Earth Syst. Sci., 22(6): 2031-2056.

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

Schirmer, M., Lehning, M. and Schweizer, J., 2009. Statistical forecasting of regional avalanche danger using simulated snow cover data. J. Glaciol., 55(193): 761-768.

Schweizer, J. and Föhn, P.M.B., 1996. Avalanche forecasting - an expert system approach. J. Glaciol., 42(141): 318-332.

Schweizer, J. and Jamieson, J.B., 2007. A threshold sum approach to stability evaluation of manual snow profiles. Cold Reg. Sci. Technol., 47(1-2): 50-59.

Schweizer, M., Föhn, P.M.B., Schweizer, J. and Ultsch, A., 1994. A hybrid expert system for avalanche forecasting. In: W. Schertler, B. Schmid, A.M. Tjoa and H. Werthner (Editors), Information and Communications Technologies in Tourism, Innsbruck, Austria, 12-14 January 1994. Springer Verlag Wien, New York, pp. 148-153.

van Herwijnen, A., Heck, M., Richter, B., Sovilla, B. and Techel, F., 2018. When do avalanches release: investigating time scales in avalanche formation. In: J.-T. Fischer et al. (Editors), Proceedings ISSW 2018. International Snow Science Workshop, Innsbruck, Austria, 7-12 October 2018, pp. 1030-1034.

van Herwijnen, A., Heck, M. and Schweizer, J., 2016. Forecasting snow avalanches by using avalanche activity data obtained through seismic monitoring. Cold Reg. Sci. Technol., 132: 68-80.

Zeidler, A. and Jamieson, J.B., 2004. A nearest-neighbour model for forecasting skier-triggered dry-slab avalanches on persistent weak layers in the Columbia Mountains, Canada. Ann. Glaciol., 38: 166-172.