Answer to Karl W. Birkeland (RC2)

Léo Viallon-Galinier Pascal Hagenmuller Nicolas Eckert

General comments

> In this paper the authors present a method using random forests to predict natural avalanches running to the valley bottom in the French Alps. Their methods appear to be solid, and the question they are trying to answer is important. In comparison to previous research, the novelty of their approach is that they make their predictions at the spatial scale of specific elevations and aspects. The paper is generally well-written and clear. I believe this research makes a valuable contribution, but I also feel there are issues that should be addressed prior to publication.

We thank K. W. Birkeland for this detailed and useful review. We answer point by point to the different issues raised below.

> Here are a few of the major issues that I believe should be addressed:

> - It would be helpful for the reader to better understand the spatial characteristics of the starting zones of the approximately 110 avalanche paths in the study area. Looking at Figure 1, it appears that most of the starting zones will have either a NW or a SE aspect. I am not sure about the distribution of the starting zone elevations. A Figure like Figure 2 (which shows the distribution of avalanche events by aspect and elevation) should be created for the avalanche path characteristics. In fact, it would be useful to pair this new Figure with Figure 2 so the reader could assess the effect of the avalanche path characteristics on the number of avalanches in each elevation/aspect zone.
> - Along these same lines and again looking at Figure 1, I assume that the elevations and aspects of the avalanche starting zones are not evenly distributed in the 24 classes (three elevation and eight aspect categories). How does this affect the analyses? I understand that the authors would like to use the 24 elevation/aspect categories used in avalanche forecasts, but I wonder if it is appropriate to use all 24 categories for a dataset that appears to be unbalanced in the distribution of avalanche starting zone characteristics? How is this affecting their results?

Avalanches are observed only if they are located in pre-defined avalanche paths and reach the observation line. However, information reported contain the elevation of departure of the avalanche and the aspect of the area where it started. This may not be directly linked to avalanche path information as a path may be globally south-oriented but sides may look south-east or south-west, or even East or West for avalanche paths with large departure zones. We use the information related to the precise recorded avalanche as soon as it is available (most of the time), that is why we presented the data for the observed avalanches rather than for the avalanche paths.

Moreover, as the valley turn, we have globally north and south facing paths in Lanslevillard and more South-East and West facing path in Bessans for instance. However, even though a wide variety of aspects is represented in terms of avalanche paths, this does not ensure that all paths are equivalent. Some may have more forest than other, or different vegetation influencing susceptibility to avalanches, some may be steeper than others, etc. We cannot ensure an equal representation of all possible aspects and elevations with similar conditions.

However, as we point out in the description of the dataset, we believe that due to the high number of observed avalanche paths and the steepness of the slopes, the recorded EPA avalanche activity is a good proxy of overall avalanche activity of the Haute-Maurienne valley. Hence, when the goal is to predict the avalanche activity of Haute-Maurienne, the use of a realistic avalanche activity, including unbalance between elevation and aspects seem relevant. This means that the model may not be directly transferable to other areas. We will introduce this in the discussion: *we here train the model with the Haute-Maurienne data. Some climatological or terrain features may lead to a predicted avalanche activity specific to the Haute-Maurienne area, especially with a higher senitivity of* *certain aspects or elevations (eastern crests during easternly returns). The model may not be transferable directly to other areas without a new calibration.*

> - Another issue is the inclusion of both dry and wet snow avalanches in the same analysis. This was also pointed out by the other reviewer. Since we know that the avalanche release mechanisms for these two primary categories of avalanches are quite different, as are the meteorological factors that lead to instability, why are these included in the same analysis? Perhaps this is because both wet snow stability indices and dry snow stability indices are included? Wouldn't it be better to split all the avalanches into "dry" and "wet" categories, and then proceed with the analysis on each of these two subsets of the data?

The EPA being an observation of avalanche deposits, with remote observations from valleys, it provides few information on processes in starting zone. In particular, although the deposit is often described as mainly dry or mainly humid, the wetness of the snowpack in the starting zone is not reported. It is therefore difficult to classify between dry and wet avalanches based on information reported in the dataset. More generally, classification between dry and wet avalanches is not always obvious, especially during the progressive wetting of the snowpack (from top to bottom) during spring or when dry snow falls over a wet snowpack.

By removing these litigious situations and using snow cover modelling, it remains possible to define two subsets of wet and dry snow. However, we do not agree on the need of splitting a priori the two types of avalanche processes. We use the tree-based RF model. If the wetness of the snowpack is a critical factor to identify the situation, it should be selected during the optimization process as one of the top split in the tree directly by the model, especially as we provide relevant indicators to identify if the snowpack is rather dry or wet, such as the height of wet snow or the mean liquid water content. Then, the two branches will analyze different characteristics depending on the situation (dry or wet). The model should then be able to deal with different situations. In the dry snow this is also required as we have to identify situations where a persistent weak layer is involved from situations where only the new snow have to be considered, for instance. We thus do not think that a split between dry and wet situations would help the classification, even though we know that it is a common approach in avalanche community.

We nevertheless tested to focus on the wet snow situations, as it is closer to the analysis done by forecasters. We extracted the situations for which the snowpack is mainly wet from the whole dataset (both for non avalanche days and avalanche days), based on snow cover modelling. The performance on the resulting model, focused on wet snow was not better than the full model. We then do not pursue in this path.

We will introduce a paragraph in the discussion section to summarize this.

> - The other reviewer also mentioned another issue I believe needs to be addressed. The dataset does not include all avalanches that occurred, but rather it consists predominantly of avalanches running to the valley floor. I assume these are almost all quite large avalanches. Can you provide a range of the size of the avalanches? Are they all Size 3 (on the Canadian or the U.S. destructive scale) or larger? Or perhaps size 4 or larger? What effect do the authors believe that this bias toward large avalanches has on their results?

The observation network was designed at the end of the 18th century when avalanche sizes were not yet normalized. Hence, the avalanche size is not explicitly used. The minimal size is indirectly defined for each avalanche path by the position of the observation threshold that should be crossed by avalanches to be recorded by the observer. Empirically, we can imagine that no size 1 avalanche are recorded. Avalanches of size 2 may be recorded, especially if an accident is related to this avalanche or if the avalanche reached high altitude infrastructure and most of the recorded avalanches may be of size 3 or more. However, this was never evaluated. We thus prefer not to give an indication that will not be properly supported. Moreover, due to the high number of avalanche paths in Haute-Maurienne and the steepness of the slopes, we believe that EPA provides a good overview of avalanche activity, as we point out in the description of the dataset: *Besides, the steep topography of Haute-Maurienne reduces the effect of the threshold of observation as most of the avalanches reach the valley floor, providing a representative screenshot of avalanche activity of avalanches reaching low altitudes.*. Then, we do not expect a bias on the results, as least as long as the model is not applied on other areas. We will develop this idea in the discussion.

> - While the authors reference some of the more recent work on predicting avalanches with random forests, I feel like they might want to also reference some early work that attempts to better predict avalanche activity using the statistical techniques available at that time. These older papers had more the more modest goal of trying to predict avalanche days (without elevation/aspect of the starting zones), but

> they were a first step in this direction. This does not have to be a comprehensive review at all, but just a sentence or two with some references would be nice to see. Some older examples exist of researchers using discriminant analysis (examples: Bovis, 1977; Foehn and others, 1977), nearest neighbor techniques (example: Buser, 1983), and binary regression trees (example: Davis and others, 1992). Also, who was the first to use random forests for this type of work? Perhaps one of the authors who you already reference?

Thanks for pointing this lack. We will add a paragraph to the introduction to shortly summarize the history of machine learning and avalanches. The pioneering works were performed by Bois and Foehn in the 70s, with linear methods, while the first use of classification trees were by Davis et al in the late 1990s and random forest models were firstly used in the 2010s (e.g. Mitterer et al, 2013).

> - Finally, one thing that perplexes me about this research is why new snowfall is rated so low in importance (Figure 4). This is completely different than prior research, which typically rated snowfall as the most important factor for dry avalanche release. Why do the authors believe this is the case? Is it because the “snow depth and variations” class is capturing this essential information? Or is it because of this information is captured (fully or partly) in some of the stability indices? Or is it the mixing of the dry and wet snow avalanches into one dataset? It might also be related to the fact that the dataset consists of only large avalanches. What do the authors think?

The variables we use contain a lot of redundancy and correlations. The random forest select the variable that allow the best separation into two groups at each step. We observe here that post-processed variables such as snow depth variations or new snow depth on 24, 72, 120h seem to be slightly more relevant than bulk snowfall the given day (or on 3 days), This is not contradictory with previous studies as it does not mean that snowfall do not contain relevant information. It just means that other variables are more relevant for the information related to new snow.

We have several possible ways of explanation. The first is that contrarily to most of the previous studies, we use large-scale modelled meteorological information rather than locally observed meteorology. We know that the Haute-Maurienne massif experience some heterogeneous meteorological conditions, especially during easterly return events, for which modelled meteorological information may not be fully representative. The second one is that most of the meteorological information we consider (precipitations and wind) are identical for all aspects and elevations and temperature are identical for all aspects and highly correlated between elevations whereas we know that the snowpack are generally quite different. The snowpack variables are able to summarize the history of past conditions that have built up the snowpack while meteorological information is not at the correct time scale for this. We will add a paragraph in the discussion to discuss this result on meteorological variables: *Meteorological information is not sufficient by itself. Contrarily to many other studies [e.g. Buser et al., 1989; Mayer et al., 2022], we do not use observed meteorological information but large-scale modelled information [Durand et al., 2009]. Thus, the meteorological information is uncertain and nearly identical for all aspects and elevations while underlying snowpack are generally significantly different. We then did not expect a good prediction at high spatio-temporal resolution with only meteorological information.*

> Despite the above comments, I believe this is valuable research and is deserving of publication once the authors address or respond to these issues.

> I have also attached an annotated PDF, which includes corrections to some typographical errors, as well as further suggestions and suggested wording changes.

We gathered our answers to the attached comments below.

> I hope the authors find my comments and suggestions useful.

> Karl Birkeland

> Some possible older references (the authors may have other/different older references they wish to cite):
>
> Bovis, M.J. 1977. Statistical forecasting of snow avalanches, San Juan Mountains, Southern Colorado, U.S.A. Journal of Glaciology 18(78), 87-99.
>
> Buser, O. 1983. Avalanche forecast with the method of nearest neighbors: An interactive approach. Cold Regions Science and Technology 8, 155-163.

> Davis, R.E., K. Elder, and E. Bouzaglou. 1992. Applications of classification tree methodology to avalanche data management and forecasting. Proceedings of the 1992 International Snow Science Workshop, Breckenridge, Colorado, 123-133 (available at: https://arc.lib.montana.edu/snow-science/item.php?id=1245).
>
> Foehn, P.M.B. and others. 1977. Evaluation and comparison of statistical and conventional methods of forecasting avalanche hazard. Journal of Glaciology 18(78), 375-387.

Attached comments

We only detail hereafter the main comments of the PDF. We will take into account all the detailed suggestion in the attached PDF in the revised version.

> Page 1: I would suggest re-wording the title to make it more direct, while keeping the same meaning. I'm also not sure that "snow physics" is appropriate... perhaps it would be more accurate to state that you are really using modeled stability indices? Another thing that is important to emphasize throughout the paper is that we are talking about predicting large avalanches in this study (that go to the valley floor). Given all this, one suggestion would be: Does combining modeled stability indices with machine learning help with predicting large avalanches?

We propose to change the title to *Does combining modelled snowpack stability with machine learning help with predicting avalanche activity?*. We believe that the method presented here is not specific to large avalanches. Moreover, in the specific case of Haute-Maurienne, even though EPA observation dataset record avalanches that reach the valley floor, due to the specific geography of this area and the large number of observed paths, it is representative of the overall avalanche activity.

> Page 4: Are all the avalanches in the database naturally triggered? Or are there also some artificially triggered avalanches?

Avalanches reports are based on the observation of avalanche deposits. Observers have few information on the origin of the avalanches. The observation network was designed to give an overview of natural avalanche activity. Hence, natural avalanches are natural ones. However, we cannot ensure that no triggered avalanches are present in the dataset.

> Page 11: This is interesting. I am curious why snowfall does not look important in Figure 4, and rainfall also does not look important. It seems that snowfall should be among the most important variables, and the other snowfall variables (snow depth and variations in depth) are important. I will be interested to see how this is explained in the Discussion.

We answered in the general comments and will include further discussion on this point.

> Page 12: This is a surprising finding since new snow is often one of the most important variables for discriminating between avalanche and non-avalanche days.
>
> I find this to be really surprising since others in the past have had some reasonable results (though definitely not perfect) with only looking at meteo variables such as new snow, wind and temperatures.
>
> I am hoping you will discuss this in your discussion section.

We answered in the general comments and will include further discussion on this point.

> Page 14: I understand the other measures, but I think it would be helpful for the reader if you more explicitly explained "threshold" and what a value of that threshold means in the context of this Table. What is a "good" threshold, or is there such a thing? Larger numbers or smaller numbers or ??
>
> I read through some of the explanation in Section 2.6.3, but it was still not clear to me so I went and did my own work to try to better understand it.
>
> Now that I understand it better, perhaps your wording is OK. But, you could have another look and see. I would like it if there was something even in this figure legend that told us how to interpret these threshold values. Clearly they don't line up the same as the AUC values for the different sets of variables.

There is no good or bad threshold here, we reported this value as an information, as it traduces some details of the behavior of the model. We will adapt the legend to make it clear that, contrarily to other values, this is not a score, but an additional information.