

Answer RC1

François Doussot, Léo Viallon-Galinier, Nicolas Eckert, Pascal Hagenmuller

1 Specific comments

Comment 1

Predictors across elevation bands: The rationale for selecting predictors from multiple SAFRAN elevation bands is not fully clear. As most predictors relate to avalanche release, conditions at start-zone elevations should be more relevant than those near runout elevations. Taking certain variables taken from a single elevation (e.g., wind direction, persistent weak layers) and others from multiple elevations accentuates an inconsistency. Snowpack conditions at lower elevations are likely secondary influences (e.g., sufficient snow depth to cover ground roughness, snow available for entrainment, or weak layers for step-downs). Mixing these predictors with those from start zone elevations may weaken model performance and, as noted in the discussion, adds complexity due to elevation-dependent climate change effects. A more consistent approach that either directly targets start-zone elevations or more clearly justifying the inclusion of lower elevations would improve interpretability.

We thank the reviewer for this comment. We agree that the predictor selection across different elevations was insufficiently explained in the original manuscript and that some choices may appear inconsistent. In response, we revised the predictor set so that the variables directly related to avalanche release processes (e.g., wind speed) are now provided only at the representative release-zone elevation (2700 m). We also added a detailed justification for the variables retained across multiple elevations, explaining their potential influence on avalanche flow and runout, which is particularly relevant given that the dataset includes only avalanches that meet a predefined observation threshold. The corresponding explanation has been added to the Methods section:

Most predictors were selected to characterize snowpack conditions in the release area and therefore target avalanche release processes, such as wind slab formation or persistent weak-layer instability. Accordingly, variables directly related to release conditions (e.g., wind speed and strength-to-stress ratio) were provided only at 2700 m, which approximately corresponds to the mean elevation of the avalanche release zones. In contrast, a subset of variables was extracted at 1800, 2100, 2400, and 2700 m because conditions along the avalanche path may influence avalanche flow and runout altitude. This distinction is particularly relevant because the dataset includes only avalanches that meet a predefined observation threshold. Although these factors are expected to play a secondary role compared with release-area conditions, they may affect whether a released avalanche is recorded in the dataset. For example, sufficient snow depth or available transportable snow at lower elevations may be required along the path for an avalanche to reach the threshold. Solid and liquid precipitation were also provided at all elevations because precipitation can affect both avalanche release and avalanche flow. Wet-snow-related variables were similarly provided at all elevations because snowpack wetting generally progresses from lower to higher elevations. Accounting for this elevation-dependent evolution allows a more complete characterization of wet-snow conditions throughout the avalanche path. We retained variables with a plausible physical link to avalanche activity rather than applying a strict pre-selection procedure.

Gradient boosting methods are generally robust to the presence of weakly informative predictors, allowing physically relevant variables to be included without substantially degrading predictive performance. This approach preserves the physical interpretability of the model while allowing the algorithm to identify the most informative predictors among the provided features.

This new feature selection leads to minor changes in the evaluation section: while the annual and seasonal MAE are slightly higher compared to the last version of the manuscript, the "worst week" MAE is now significantly lower. These results are updated in the new version of the manuscript and do not require major modifications to the discussion.

Comment 2

Results by aspect: It would be interesting to report model performance by aspect (e.g., mean absolute error for each aspect) to better understand the skill of the model chain. Since the model predicts avalanches by aspect, this also raises the question of whether past and future trends differ by aspect (for example is there a stronger decrease on south-facing slopes than north-facing slopes?). While a full additional analysis may be beyond scope of this paper, briefly reporting performance by aspect and discussing potential aspect-dependent trends would add value.

Because of the valley topology, the number of avalanche paths varies substantially among aspect sectors. Eastern and western sectors contain relatively few avalanche paths, whereas northern and southern sectors each include 36 paths. We therefore focused our exploratory analysis on the comparison between northern and southern aspects.

Despite having the same number of avalanche paths, the two sectors exhibit markedly different avalanche activity. On average, since 2006/2007, northern aspects have experienced approximately 57 avalanches per year, compared with 32 on southern aspects. This difference in baseline activity should be kept in mind when interpreting aspect-dependent results, as avalanche counts may be affected by threshold effects, especially when avalanche counts are low.

As suggested, model performance was evaluated separately for each aspect sector using the same metrics as those presented in the manuscript. Performance was lower than for the aspect-aggregated model. For example, the annual mean absolute error (MAE) reached 33% for northern aspects and 34% for southern aspects, compared with 23% for the aggregated model. These results indicate greater uncertainty in aspect-specific predictions and, therefore, reduced confidence in aspect-level trend estimates. In addition, the model exhibited contrasting biases in the predicted annual number of avalanches between aspects, with a positive bias of 12% for northern aspects and a negative bias of 29% for southern aspects. These results indicate greater uncertainty in aspect-specific predictions and suggest that model errors are not homogeneous across aspect sectors. Consequently, comparisons of temporal trends between aspects should be interpreted with caution.

The Bayesian framework allows direct comparison of temporal trends across aspects via the posterior distributions of trend slopes. The results are discussed here based on the S2M reanalysis (not for the ADAMONT simulation). The posterior median estimates suggest a stronger decrease in the annual number of avalanches on northern aspects (-4.4% per decade; 95% credible interval: [-9.3%, 0.5%]) than on southern aspects (-3.1% per decade; 95% credible interval: [-8.4%, 2.4%]). However, both posterior distributions largely overlap (shown in Fig. 1), reflecting substantial uncertainty in the estimated trends. The posterior probability that the trend is more negative on northern than on southern aspects is 60%, indicating only weak evidence for an aspect-dependent difference in trends. Moreover,

this probability does not account for the differing biases and predictive performance of the aspect-specific models. Therefore, while the median estimates point toward a stronger decrease on northern aspects, the evidence remains weak and does not support a robust conclusion regarding aspect-dependent trends. Given the lower predictive performance of the aspect-specific models compared with the aggregated model, a dedicated analysis would be required to assess such differences more reliably.

While this additional analysis provides some insight into potential aspect-dependent differences, the associated uncertainties remain large and the results do not lead to conclusions that differ from those obtained with the aggregated analysis. To maintain the focus and conciseness of the manuscript, we therefore chose not to include this exploratory analysis in the revised version.

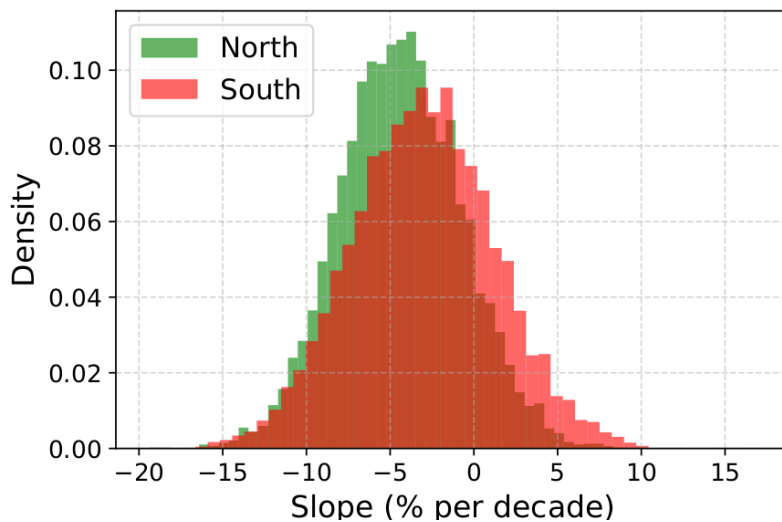


Figure 1: Computed slopes per aspect of the annual number of avalanches based on the S2M reanalysis

Comment 3

Underpredicting biggest cycles: This is an interesting result that leads to questions about whether it is a limitation of the physical models, the selected predictors, or the extreme value statistics models. This could deserve a bit more direct attention in the discussion.

First, the XGBoost model is primarily optimized to reproduce the more frequent conditions that dominate the observations. Such machine-learning approaches are therefore known to regress extreme predictions toward more common values, leading to an underestimation of the rare events, here the large avalanche cycles.

In addition, the proposed loss function implicitly accounts for uncertainty in avalanche release dates by distributing the observed avalanche counts among all potential release days. This generally improves model performance, as evidenced by the lower MAEs. A consequence of this approach is that intense avalanche cycles are partially spread over several days, which reduces the amplitude of daily peaks and tends to underestimate the highest daily avalanche counts.

Moreover, as discussed in Sec. 4.2, extreme precipitation is often due to eastern returns,

which are heterogeneous over the massif (and are maximum in the considered municipalities). They are smoothed by SAFRAN assumptions (massif-scale homogeneity), thereby reducing the intensity of the predictor signals. However, this is also the case in the training dataset, making it difficult to evaluate whether this limitation limits the statistical model's ability to reproduce the largest avalanche cycles.

These additional explanations have been added to the new version of the manuscript.

Comment 4

Length and repetition: Some sections are longer than necessary and occasionally repeat similar points across different sections the manuscript. Certain details could be reduced without affecting the core message, results, or interpretation. In my opinion streamlining the text would improve clarity and strengthen the paper overall, but the current level of detail is also acceptable if the authors prefer it as is.

We thank the reviewer for this general comment on manuscript structure and clarity. However, given the paper's structure and the need to maintain coherence among the methodological description, results, and discussion, it is difficult to substantially reduce certain sections without compromising the clarity of the overall narrative. In several cases, the level of detail was intentionally maintained to ensure that the methodological choices and their implications are clearly justified, particularly for readers less familiar with the modelling framework.

We also note that the comment does not refer to specific sections, which makes it challenging to identify precisely where reductions would be most appropriate. Some minor changes have been made to the new version of the manuscript to simplify and clarify the text.

2 Technical comments

Technical Comment T1

Line 70-73: Sentence is confusing and should be clarified.

This sentence has been changed by : *"At the scale of entire mountain ranges, biases associated with individual observers may be negligible because large-scale trends may still emerge despite local observational artefacts (Eckert et al., 2010d). However, such biases can become critical when the observation record is based on only a few observers."*

Technical Comment T2

Line 90: What metric did Castabrunet use for future avalanche activity?

It has been clarified in this version of the article: *"they used a Composite Index of avalanche activity derived from the EPA database and from the MEPRA hazard index (Durand et al., 1999). This synthetic indicator, which cannot be associated to a direct observable avalanche activity index, was computed at seasonal and annual time scales over large regions of the French Alps."*

Technical Comment T3

Line 133: Wind can also be a primary driver of snowpack variability by aspect, which I assume is also true for this study area.

Exactly, this remark has been taken into consideration in the new version of the article : *"and sometimes on wind direction"*.

Technical Comment T4

Line 167: "Strong" activity is vague. Please specify (e.g., higher avalanche counts).

This has been fixed in the revised text, using *"higher avalanche counts"*.

Technical Comment T5

Fig 2 and 3. Captions should clarify that these refer to the Haute-Maurienne subset, as "EPA dataset" could be interpreted as covering all of France.

This has been fixed in the revised text.

Technical Comment T6

Line 188: Elevation bands are not described as ranges. Are the bands centered around the reported values?

Thank you for pointing this out. The term "elevation bands" was potentially misleading, as SAFRAN provides meteorological and snowpack conditions at discrete reference elevations spaced every 300 m. To avoid ambiguity, we have replaced the term *"elevation bands"* with *"discrete elevation levels"*.

Technical Comment T7

Line 195: Briefly define/explain quantile mapping.

This has been fixed in the revised text: *"which is a statistical bias-correction method that adjusts the distribution of climate model outputs to match that of a reference dataset, which is here the SAFRAN reanalysis."*

Technical Comment T8

Line 201: Add a sentence describing the final weather forcing datasets (e.g., temporal resolution and how it is structured across elevation and aspect).

It has been added in the new version: *"Crocus outputs were produced at a 6-hour temporal resolution)". The spatial resolution (elevation, aspect and slope) is now explicitly provided ("here 1800, 2100, 2400 and 2700 m [...], N, E, S, W [aspect] and 40 degrees slope").*

Technical Comment T9

Line 206: Clarify whether "50 layers" refers to snow stratigraphy layers and whether "5 variables" refers to snowpack properties.

The “5 variables” has been deleted to avoid misunderstanding in the new version.

Technical Comment T10

Line 226: It would be worth noting in the discussion the limitations of using a single strength–stress ratio to represent persistent weak layer avalanches. These types of avalanche are among the most complex to model and forecast, and a simplified metric is likely to miss important processes. A brief acknowledgment of this limitation would strengthen the interpretation of results.

We agree with this remark, it has been added in the discussion (lines 472-479): *The choice of predictors could also be further refined. For example, we only used the strength-to-stress ratio to characterize persistent weak-layer avalanche problems. It provides a very simplified description of snowpack stability. More physically based metrics, such as the critical crack length (Gaume et. al, 2017), could potentially offer a more realistic representation of avalanche release processes and improve predictive skill.*

Technical Comment T11

Sect 2.3.2: It appears predictors are computed daily at 18:00. This should be explicitly stated.

It has been explicitly added in this section.

Technical Comment T12

Sect 2.4.2: This section does not clearly explain how the different metrics are combined into a loss function. While n_{min} and n_{max} are intuitive based on the observation windows, the role of ΔT is unclear. Even with Fig. 4, its interpretation and purpose in normalizing residuals are difficult to follow. Clarifying how ΔT functions within the loss formulation would improve readability.

You are right, in the description, we did not clearly relate the residual to the loss function. It is now corrected : *"This distance is then distributed uniformly across all days within ΔT by dividing it by the duration of the interval (in days). The resulting quantity defines the daily residual R used as the loss."*

Technical Comment T13

Line 276: Winter seasons are often labeled by the year they end, though conventions vary.

It has been corrected in the new version of the paper, in all the figures and in the text.

Technical Comment T14

Line 405: Sentence is unclear and should be revised.

This sentence has been clarified to define what is a zero-inflated count distribution : *"Indeed, the observed daily avalanche counts follow a zero-inflated count distribution, characterized by integer-valued observations with an excess of zeros (Young et al., 2022)."*

Technical Comment T15

Line 419: Please be more specific about the uncertainty in older observations. If the concern is potential underreporting in earlier periods, this could be stated more directly rather than described as generic uncertainty.

It has been clarified in the new version : *"These results suggest that avalanche reporting was not consistent throughout the 1958–2023 period in the Haute-Maurienne valley, with possible under-reporting during the earlier decades"*

Technical Comment T16

Line 560: Consider starting a new paragraph when transitioning to future results.

This has been fixed in the revised text

3 Additional modifications

In the previous version of the manuscript, a single model was trained using the S2M re-analysis dataset over the 2006/2007–2022/2023 winter seasons and then applied to the entire 1958/1959–2022/2023 period to estimate daily avalanche activity. However, this approach introduces a methodological inconsistency because the training period is also included in the reconstruction period. As shown by the leave-one-year-out cross-validation, the model exhibits systematic biases. When the model is evaluated on the data used for training, these biases are artificially reduced, which may, in turn, affect the estimated long-term trend.

To avoid this issue, we revised the methodology. For the 1958/1959–2005/2006 period, avalanche activity is estimated using a model trained on the 2006/2007–2022/2023 dataset, as in the previous version of the manuscript. For the 2006/2007–2022/2023 period, we use the results of the leave-one-year-out method as done in Sec. 3.1, ensuring that each year is predicted by a model that was not trained on that year.

This revised approach provides a clearer separation between model training and application, thereby avoiding potential in-sample bias in the trend analysis. The correction leads to moderate changes in the results, with generally weaker estimated trends than those reported in the previous version, as shown in Table 1. The discussion is still consistent with these new computed trends.

	Annual	DJF	MAM	Worst week
Previous version	-6.0%	-4.7%	-8.4%	-3.5%
New version	-4.7%	-1.5%	-8.4%	-1.7%

Table 1: Changes in the computed trends from the reanalysis. DJF refers to winter months (December, January, February), and MAM to March, April and May.