Summary of major revisions

Frank Techel, and co-authors

We sincerely thank editor Pascal Haegeli (ph) and reviewers Florian Herla (fh) and Christoph Mitterer (cm) for their extensive, thoughtful, and constructive feedback. Their comments helped us to significantly improve the clarity and focus of our manuscript.

Summary of Major Revisions

Given the request by the editor (ph) to simplify and restructure the manuscript, we have made several major changes, which we summarize below. Please also refer to the marked-up manuscript (track changes PDF) for detailed edits. We had already responded to the original reviewer comments in our earlier response, and have addressed all of their points in this revision, except for the two listed at the end of this reply. We use the abbreviations *fh*, *cm*, and *ph* to indicate points raised specifically by individual reviewers.

While large parts of the manuscript have been revised considerably for clarity and structure (see track changes PDF), the overall storyline, methodology, key findings, and their interpretation have not changed, or only marginally.

- **We clarified the scope and objectives** of the study, especially emphasizing that our comparison is limited to the *danger level* component of the public avalanche forecast, and that the results apply to the specific context of Switzerland. (ph, fh) This included changing the title, abstract, and introduction (60-61).

- **We simplified the structure of the manuscript** (ph), particularly in the Methods and Results sections including the following points:

  - We removed the parallel analysis of nowcasts and forecasts (ph). We now exclusively analyze predictions in forecast mode (150-151).

  - We removed the use of GPS-derived non-events for the human-triggered avalanche analysis and retained only the analysis introducing and relying exclusively on the reference distribution approach, as suggested by reviewer ph. While we believe that GPS data provide a valuable additional perspective on user exposure and complement the reference-based approach, we followed the reviewer's recommendation in order to simplify our manuscript. We briefly return to this point in the Discussion (456-463).

  - We restructured the *Data and Methods* section (Section 3) making it clearer how data and methods are linked to the two research questions (ph). For instance, Section 3.2.4 now describes explicitly the analysis and steps needed to address research question 2 (ph). Restructuring this section included revising and simplifying Figure 2 (fh, ph, cm).

- **We improved the explanation and justification of key methodological choices**, particularly in Section 3.2 (Methods), in response to reviewer requests for increased clarity and transparency (fh, cm, ph).
  - We restructured the Methods section to establish a clearer and more logical link between the analysis steps and the two research questions (ph). For instance, the steps and additional methods needed to address research question 2 are now listed in Section 3.2.4 linking them closer with this research question. (ph)
  - We expanded our explanation for choosing $\Pr(D \geq 3)$ as the danger-level model output used in the evaluation, and contrasted this with other possible outputs (Figure A1 in Appendix) (cm, ph). We added a supporting Figure A1 in the Appendix showing model behavior and decision rationale behind the use of $\Pr(D \geq 3)$. (cm)
  - We clarified the rationale for using relative event ratios in Section 3.2.3. (ph)
  - We clarified the logic and implications of our binning strategy, including why binning was necessary to enable comparison between continuous model predictions and discrete danger levels. (fh, cm)
  - We added a new explanatory figure illustrating the binning concept and approach (Figure 5a) (fh, cm) and showing all bin thresholds per model to increase transparency in Figure 5b. (fh)
- **We enhanced the clarity of figures and tables**, especially Figures 1, 2, 6, and 7, and Table 1, by improving layout, labeling, and captions. (fh, cm, ph) For instance, we now show the median model-predicted values for reference distributions and event data directly in Figure 6a-d.
- **We shortened and refocused the Discussion**, aligning and linking it more directly with our two research questions. (ph) We now begin by interpreting the key findings in relation to the two research questions (Section 5.1), moved some explanatory content to the Methods section (L244–249, Figure 4), and reduced the discussion on the potential advantages of GPS tracks over the reference distribution, as well as the limitations of the reference approach, to the Limitations section (L481–501).

Although not requested by the reviewers but we took the opportunity to **add a third forecasting season (2024/25)** to the dataset. This addition, which we discussed with the editor (ph) prior to implementation, increased the sample size and robustness of the analysis, but did not alter the overall story or key findings, nor did it increase the complexity of the manuscript.

Two suggestions by the reviewers were not implemented in this revision:

- We did not perform a simulation experiment with degraded forecast quality, randomized danger levels, or alternative binning approaches for the model predictions, as proposed by reviewer fh. While we agree that these are valuable avenues for future work, they fall outside the scope of the present study and would require additional design considerations and extensive methodological description. Nonetheless, as a partial step in this direction, we now use bootstrap sampling to derive 90%-confidence intervals for the event ratios (261-262, 276-278).
- We did not include a case-by-case analyses or time series for specific regions or periods (as suggested by reviewer cm and as originally outlined in our responses to reviewers fh and cm), in order to maintain focus on the evaluation, to avoid

expanding the manuscript further, and to follow the recommendation by reviewer/editor ph to concentrate on the main story. We agree, however, that this is a promising extension for future work.

We believe these revisions have strengthened the manuscript considerably.