Reply to Reviewer #1

We thank reviewer #1 for the constructive review of our manuscript, which we greatly appreciate. In the following, we address each of the points raised. Black text indicates the reviewer's comments. Below we describe (in blue) how we will address each comment when revising the manuscript.

General comments

This study examines future avalanche activity in the Swiss Alps with a modelling approach that combines downscaled climate projections, a detailed snow stratigraphy model, and machine learning models. The models predict the number of days with dry and wet snow avalanches under different climate scenarios at 7 high-elevation locations, finding an overall decrease in dry snow avalanches and an increase in wet snow avalanches. These trends are interpreted by showing how the snowpack structure will change at different elevations.

The overall quality of the manuscript is high, with a clear presentation of the methods and results and interesting discussion and conclusions. The study design combines several state-of-the-art methods to investigate a significant topic that fits well within the scope of The Cryosphere. I recommend publication after addressing the following comments.

Specific comments

- **Climate warming versus climate change:** From what I understand "climate warming" primarily refers to the temperature component of "climate change", as opposed to the entire climate system (precipitation, feedback loops, etc.). Were the authors deliberate in how they used each term throughout the manuscript (e.g., "climate warming" in the title suggests the study focuses on the response to warming temperatures)?

  Thank you for pointing out the differences between these notions. We used the term "climate warming" as we consider rising temperatures to be the primary driver of all other changes. However, based on your comment, in the revised version of the manuscript, we will consistently use the term "climate change".

- **Explanation of climate models and data:** The climate data used for the study should be explained in more detail. The SNOWPACK model is thoroughly explained, but there is comparatively little information about the EURO-CORDEX model and CH2018 datasets, especially for readers unfamiliar with these products. Please briefly introduce and define GCMs and RCMs and explain the differences between the three RCP scenarios. Listing the specific GCM/RCM models in Table B1 has minimal meaning without explaining what they are and how they differ. Also, specifically for this study, it would be valuable to comment on how well the CH2018 datasets resolve fluctuating weather systems over a season and whether they are appropriate for predicting realistic snowpack stratigraphies. Do they produce smoothed average values, or do they capture realistic storms interspersed with high-pressure weather? For example, warming-related intensification of heavy snowfall is mentioned in line 38, would this be reflected in the

data? The importance of resolving extreme weather events is addressed in the discussion section, but commenting on these aspects earlier would help readers interpret the results.

We will provide more detailed information on the climate scenarios and underlying models in the revised version of the manuscript. In particular, we will point out that the CH2018 data set does provide transient daily time series of meteorological variables rather than only average values. The CH2018 climate scenarios were statistically downscaled from the coarser RCM simulations of the EURO-CORDEX ensemble to automatic weather stations (AWS) using quantile mapping (QM). This means that, for example, if the daily precipitation sum on a given day in the corresponding RCM grid cell equals the 70th percentile of the reference period climate simulations, then the daily precipitation sum at a particular station is set to the 70th percentile of the daily precipitation observed at that station during the reference period. In principle, this implies that the downscaling preserves realistic weather patterns and the downscaled data capture realistic storm events that are present in the RCM simulations if those events are represented in the reference period. However, QM assumes that the relation between conditions in the grid cell and the station is stationary, which might not hold true under changing climatic conditions. Moreover, the QM transfer function is subject to uncertainties due to the limited number of years considered for the calibration, in particular for high and low quantiles. Furthermore, the extremes that have not been observed in the reference period are mapped using a constant extrapolation of the biases for the 1st and 99th percentiles. As you suggested, we will comment on these aspects in the methods section of the revised manuscript.

- **Quantile mapping method:** Please explain the basic principles behind quantile mapping to provide an overview of its function and application. Was the goal bias correction, downscaling, elevation adjustment, etc., and why was it chosen over alternative statistical methods?

  Quantile mapping (QM) is an established statistical method used to transform the distribution of modelled data so that its statistical properties align more closely with observed data. When applied in a climate modelling context, QM involves building a transfer function that relates quantiles from the modelled distribution of a meteorological variable to the corresponding quantiles in the observed distribution for a given reference period. The transfer function is then applied to bias-correct a transient time series of the simulated variable. QM is thus primarily a bias-correction method but can implicitly serve as a downscaling method if the observed reference data refer to a higher spatial resolution than the climate model (Feigenwinter et al., 2018).

  We used QM to transfer the CH2018 climate scenarios for the SwissMetNet stations to the higher-elevation IMIS stations. Herein, the multivariate quantile mapping approach from Cannon (2018) enabled us to do a bias-correction while restoring intervariable relationships, a clear advantage over alternative statistical methods for bias correction. The CH2018 scenarios for the SwissMetNet stations were themselves obtained by applying the QM approach to the EuroCordex RCM simulations, thus simultaneously bias-correcting and downscaling this data set (see previous answer). We will provide missing information in the revised version of the manuscript.

  We preferred QM over other methods like the delta change method (Michel, 2021) or weather generators (Peleg, 2019) because it can generate transient time series while remaining computationally efficient.

- **Figures:** The overall figures quality is high, but some figures are confusing because they mix multiple data types and axes into a single graphic. It would be worth examining whether any subfigures should be split or omitted. Examples where data/axes do not fit with the rest of the figure include the left column in Fig. 3, the right column in Fig. 4, the mix of subfigures in Fig. 5, and the validation column in Fig. C2. Also, several axes label relative differences as fractions, but the manuscript text discusses them in terms of percentages. Perhaps it would be easier to interpret if the axes were labelled with percentage values (i.e., -10% instead of -0.1)?

  We agree that some figures include a lot of information. However, with the comprehensive figure captions we provide, we think there is no need to change the figures. It does not seem unusual to us to have several panels with different axes in one illustration.

- **Impact of validation findings:** 3b suggests the model chain has a systematic bias towards underpredicting wet avalanche days at low elevations and overpredicting at high elevations. Are there any known reasons for this and could it be corrected? Furthermore, how does this impact the results? Fig. 3 and 4 show distinct trends in wet avalanche activity by station elevation, but are these results exaggerated due to this bias?

  Thank you for this interesting remark on Fig. 2b (not 3b we suppose). As the underlying model chain (downscaling procedure + SNOWPACK + random forest "blackbox" models) is quite complex, it is hard to tell why wet-snow avalanche days are underpredicted at lower elevations. Interestingly, the validation of the simulated snow height shows a similar systematic bias towards underpredicting snow height, with higher relative differences for lower stations (Appendix Fig. C2 h). Lower snow depths also imply fewer days with snow cover and thus fewer potential wet-snow avalanche days. Yet, this relation is not sufficient to explain observed biases, as the validation does not show the same trend for the dry-snow avalanche days (Fig. 2a). Appendix Figs. C1 8-11 reveal that snow depth at the four lowest stations is particularly underestimated after the peak of snow depth in March / April, which is a relevant time period regarding wet-snow avalanche activity. This underestimation of snow depth at the end of the season can either originate from an underestimation of snowfall (Appendix Figs. C1 15-18), or from an overestimation of melt rates due to differences in the input variables and SNOWPACK settings between the simulations based on the climate scenarios and the simulations based on measurements. For example, incoming shortwave radiation in the climate scenarios is obtained from quantile mapping based on shortwave radiation measurements from a nearby SwissMetNet station, as it is not measured at the IMIS stations. This uncertainty in radiation might also contribute to errors in the simulation of the ablation period and thus influence the parameters used as input for the wet-snow random forest model (e.g. the liquid water content). Yet, finding the ultimate reason for the observed trend in the deviations between simulations driven with climate data versus measurements would require more in-depth station-specific analyses considering all uncertainties inherent in the model chain, and goes beyond the scope of our paper.

  In this context, we have noticed that the manuscript is missing a paragraph on the SNOWPACK settings used in the simulations based on measurements for validation. We apologize for this oversight and will add this paragraph in the revised version of the manuscript.

- **Spatial/frequency distribution:** The discussion acknowledges that the model only addressed one component of avalanche risk: the likelihood of triggering. While the importance of destructive size is discussed, the spatial/frequency distribution of avalanches is also a commonly recognized

component. The European Avalanche Warning Services defines avalanche danger based on snowpack stability, frequency distribution, and avalanche size. Similarly, the conceptual model of avalanche hazard defines the "likelihood of avalanches" as a function of "sensitivity to triggers" and "spatial distribution." The spatial component of hazard should be acknowledged and discussed. For example, the results show some elevation trends that may relate to frequency distribution, but in general the machine learning models primarily predicted likelihood of triggering. Could the methods be adapted to further examine this, either with spatially distributed snowpack models (e.g., by aspect), or perhaps training machine learning models to predict distribution?

We agree that the frequency distribution of snow instability is an important element for defining the avalanche danger level in avalanche forecasts. Avalanche forecasts are issued for a certain region. Therefore, it is important to know how many avalanches are to be expected, and of which size. In our study, we simply assess the triggering potential given a certain snow stratigraphy. In other words, there is no spatial component. It is beyond the scope of this study to estimate how the spatial distribution would be affected.

We will address the spatial component of hazard in the discussion of the revised manuscript. In our view, one crucial part in obtaining reasonable projections of future avalanche activity is the preparation of meteorological input data that represent the conditions at the resolution of interest. To obtain spatially distributed projections of avalanche activity, high-resolution gridded input data to drive the model chain is necessary. As such projections containing all variables required to drive SNOWPACK were not available, we focused on obtaining climate scenarios for seven stations at different elevations to obtain a first estimate on the trend of future avalanche activity in the Swiss Alps. In a next step, these projections could be refined by analyzing results for different aspects (see also answer to the editor).

- **References:** The manuscript is well cited, but perhaps the reference list is a little long at 5 full pages. Also, I think Verfaillie et al. (2018) is very relevant to this study, and perhaps the recently published review by Eckert et al. (2024).

  Thank you for mentioning these studies which we will cite in the revised version. As in our view there are no useless references in our reference list, we would like to keep all existing references.

Technical comments

- Line 110: Appendix B is mentioned before Appendix A. Consider reordering.

  Thank you for noting this. We will reorder the appendices.

- Line 178: How was three-day sum of new snow height calculated and was it done the same way for SNOWPACK output and AWS measurements?

  The three-day sum of new snow height provided by SNOWPACK is calculated as the sum of the three consecutive values of the height of new snow within 24 hours. The AWS measurements provide snow depth only and do not measure new snow height. Hence, for this case, SNOWACK

was driven with snow depth measurements, and new snow height was calculated taking into account settlement of the snowpack. For the simulations driven with the downscaled climate model output, SNOWPACK was forced with precipitation and the height of new snow was again computed considering settlement. We will add this information in the revised version.

- Line 181-183: It seems odd to introduce a figure in an appendix before a figure in the main body. I would consider reordering (assuming Fig. 2 is more relevant than Fig. C1).

  We will reorder the two sentences.

- Line 198: "and with AWS measurements" would better explain relative differences than "or with AWS measurements".

  As the sentence refers to two different types of forcings ("SNOWPACK simulations either forced with the downscaled climate models or with AWS measurements"), we believe that "or" is the better choice in this case. To enhance clarity, we will rewrite the sentence to "To validate the complete model chain, including both dry- and wet-snow avalanche activity models, we calculated the seasonal number of avalanche days (AvDs) using SNOWPACK simulations. These simulations were forced with either downscaled climate models or AWS measurements, and we then computed the relative difference between the resulting seasonal numbers of AvDs."

- 3: Do you have any idea why the ensemble of climate models resulted in more spread for wet avalanche days than the dry avalanche days? Was the wet avalanche machine learning model more sensitive to a specific input?

  Thank you for the interesting remark. To answer this question with confidence, a detailed sensitivity analysis comparing both avalanche activity models would be necessary. As this would go beyond the scope of our study, we can only speculate here. The most important meteorological parameter for the dry-snow model is precipitation. For the wet-snow model, on the other hand, radiation and temperature are explicitly included as input features, and precipitation still plays a major role as it drives the length of the snow season and thus the potential length of the wet-snow avalanche season. The dependence of the wet-snow model on the combination of different meteorological parameters can thus be assumed to be stronger compared to the dry-snow model, leading to larger overall variability.

- Line 265: An aside comment… Could you split the dry snow avalanche days by weak layer type to count the number of days with avalanches on persistent versus non-persistent grain types? What other information about the avalanches could be derived from snowpack models (avalanche problems, aspect-trends, etc.)?

  Identifying the weak layer grain type would indeed be an interesting next step. However, this is not as straightforward as one might think, for example, we would have to decide whether to consider all layers exceeding the instability threshold, or just the deepest one or the one with the highest instability value. We therefore did not include this in our analysis. As mentioned above, analyzing trends for different aspects would be an interesting next step. Moreover, analyzing the depths of the weak layers identified by the dry-snow instability model could yield interesting information on changes in potential avalanche size.

- Line 297: "and despite" instead of "as despite".

  We intended to express causality with the use of "as" in this case.

References

Eckert, N., Corona, C., Giacona, F. et al. Climate change impacts on snow avalanche activity and related risks. Nat. Rev. Earth Environ., 5, 369, https://doi.org/10.1038/s43017-024-00540-2, 2024.

Verfaillie, D., Lafaysse, M., Déqué, M., Eckert, N., Lejeune, Y., and Morin, S.: Multi-component ensembles of future meteorological and natural snow conditions for 1500 m altitude in the Chartreuse mountain range, Northern French Alps, The Cryosphere, 12, 1249–1271, https://doi.org/10.5194/tc-12-1249-2018, 2018.

References

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Climate Dynamics, 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

Feigenwinter, I., Kotlarski S., Casanueva A., Fischer A.M., Schwierz C., and Liniger M. A.: Exploring quantile mapping as a tool to produce user-tailored climate scenarios for Switzerland. Technical Report MeteoSwiss No. 270, MeteoSchweiz, 2018.

Michel, A., Sharma, V., Lehning, M., and Huwald, H.: Climate change scenarios at hourly time-step over Switzerland from an enhanced temporal downscaling approach. Int J Climatol., 41: 3503–3522, https://doi.org/10.1002/joc.7032, 2021

Peleg, N., Molnar, P., Burlando, P. and Fatichi, S.: Exploring stochastic climate uncertainty in space and time using a gridded hourly weather generator. Journal of Hydrology, 571, 627–641. https://doi.org/10.1016/j.jhydrol.2019.02.010, 2019