Review of Past and future changes in avalanche problems in northern Norway estimated with machine-learning models

By Kai-Uwe Eiselt and Rune Grand Graversen

Summary

This paper uses a model chain to predict the past and future avalanche hazard in northern Norway. This work builds on previous work by the same authors, who developed Random Forest models to predict avalanche danger. The model chain they developed primarily consists of a dynamic downscaling of climate models in Norway for the past and future, which serves as input to the snow cover model SNOWPACK. Then, they build Random Forest models to predict avalanche days for several avalanche problems using meteorological variables (from the downscaled climate models) and snow instability variables from SNOWPACK. They show different historical trends in the frequency of avalanche days for different avalanche problems (e.g., wet, storm, wind, or persistent), as well as correlations with the Arctic Oscillation (AO). They conclude with projections of avalanche problems using climate projections (RCP4.5–8.5) for Norway, demonstrating similar results to those found in the Alps (Switzerland and France).

The paper is generally well written, well thought out, and is worthy of publication in The Cryosphere. The only major concerns I have regarding the methodology relate to the spatial aggregation of the downscaled climate simulations. In addition, more details should be provided concerning the SNOWPACK modeling for reproducibility purposes. It may also be beneficial to add a dedicated section in the discussion about the limitations and biases of their study, and how these affect their results (small one in the conclusion). There are a few punctuation issues across the text, and addressing them would enhance the flow of the manuscript.

We thank the reviewer for considering our manuscript and for the constructive review. We are grateful for the comments and provide a point-for-point response below.

As a general response regarding tenses, we prefer a style where our current work leading to the results presented in the article is referred to in present tense, while earlier work (by us or other authors) is referred to in past tense. We are of the opinion that this increases clarity and hope that it is acceptable to maintain this in our manuscript.

Major Comments:

1. Climate simulations tend to "smooth" extreme events due to their coarse resolution. In addition, the spatial aggregation of the climate simulations further enhances this smoothing effect, which is critical for avalanche problem types such as storm and wind slab. I believe this important bias needs to be addressed in the discussion, as it could affect the interpretation of the projected trends for these two avalanche problems. While the projected climate captures "thermal" avalanche problems such as persistent weak layers (PWL) and wet snow reasonably well, the projections for storm and wind slabs should be interpreted with caution. I also think that more information on the spatial aggregation would help the reader better understand this effect.

We agree with the reviewer that the spatial aggregation of the data is a major issue with our study. However, note that we differentiate only between the binary level non-AvD (danger level 1&2) and AvD (danger level 3&4&5), likely in itself representing a sort of "smoothing" of extreme-event-levels 4 & 5 by their combination with the more moderate level 3. Moreover, the avalanche warning regions in Norway are large (our five regions are 3000-9000 km^2), presenting a challenge to spatial aggregation and again likely representing in itself a "smoothing." We think this is a more general problem impacting machine-learning model accuracy, and we discuss this in the manuscript (lines 590-593). Following the suggestion in the reviewer's general comment, we will add a new section to the discussion ("6.4 Limitations") where we will include the points noted here as well as additional concerns (see the responses to the comments to line 257 and 597-598). We will also move our existing discussion of the limitations and concerns from the conclusions section (lines 590-624 in the manuscript) to this new section.

2. More details are needed concerning important parameters, parameterizations, and the simulation setup of the SNOWPACK model, in order to improve the reproducibility of this study.

We will expand the information given about our SNOWPACK setup and change the first paragraph in section 2.4 to:

"To obtain more detailed information on the snow cover, we run version 3.6.0 of the physics-based, multi-layer model SNOWPACK (Bartelt and Lehning, 2002; Lehning et al., 2002a, b). The model solves the governing conservation equations (for mass, energy, and momentum) within the snowpack and simulates the snow cover one-dimensionally. We run SNOWPACK with a 15 min computation time step, Neumann boundary conditions at the snow–atmosphere boundary, and the bucket scheme approach (Bartelt and Lehning, 2002) to simulate liquid water flow through the snow and soil. For atmospheric stability, the simplified Richardson-number correction is applied.

For soil evaporation the standard resistance approach is employed and fitted values are used for soil thermal conductivity."

However, regarding reproducibility, we note that the full SNOWPACK setup scripts are published on Zenodo (https://doi.org/10.5281/zenodo.17277192; see also the Code Availability section).

3. The figure sizes should be adjusted, as they are currently too small, and the font style does not match that of the manuscript.

We will adjust the figure sizes to increase readability. However, there is no requirement in *The Cryosphere* style guide (https://www.the-cryosphere.net/submission.html) for the font of the figures to match the font of the text (the requirement is only that the font of the figures be consistent). We are not familiar with such a requirement from other journals either and our previous article in *The Cryosphere* (Eiselt and Graversen, 2025, https://doi.org/10.5194/tc-19-1849-2025) uses the same font styles as here. Thus, we would like to maintain the font as is.

4. Several punctuation marks are missing throughout the text, which limits the flow and the comprehension of some sentences. I've highlighted a few examples below, but please check this consistently throughout the manuscript.

We are grateful to the reviewer for pointing out the punctuation issues and we will fix them in the next version of the manuscript.

Specific comments (line number)

Section 1 - Introduction

15: Already have impact the occurrence in the arctic, especially in mass movements. they are several references in the literature.

We will reformulate this to:

"Changes in climatic conditions, as observed for the 20th and as projected for the 21st century, impact the occurrence and character of natural hazards (Hock et al., 2019)."

Hock et al. (2019) is the chapter "High Mountain Areas" in the "IPCC Special Report on the Ocean and Cryosphere in a Changing Climate" which (among other things) gives a general assessment of the impact of past and future climate change on natural hazards in mountainous areas.

Section 2 - Data

115: Change apply to past tense “applied”

Our standard is to present our current work in present tense and earlier work (by us or other authors) in past tense. Since this “apply” here refers to our work for the current article, we would prefer to leave it as is.

158: What is slab snow??

This refers back to “new”: “*new* loose and slab snow,” both of which, as can be seen in, e.g., Fig. 3, are part of the avalanche problems. We will adapt this to “new loose and new slab snow,” to increase clarity.

160-161 : I think a ref to Figure 3 would be great here, as I struggle the get what the number means unless I look at Figure 3.

The reference to Figure 3 appears in line 157 and the text following this reference is meant as a discussion of this figure. However, we agree that a further reference to Fig. 3 is helpful here and will add it: “Figure 3 further shows that the general ADL…”

Figure 3: is ADL on the x axis the general? Please define.

This is correct. We are grateful to the reviewer for pointing out the oversight and will add this information in the figure caption. In addition, we will further improve the consistency by colouring the left y axis red and the right y axis black as in Figs. 1 and 2.

183: punctuation is needed to enhance the flow between danger and we.

Changed.

201: punctuation is needed to enhance the flow between conditions and Lind.

Changed.

205: too-strong is a bit vague for an amount of precipitation, or maybe it is about precipitation rate? Please clarify.

We will change this to “too-high amount of winter precipitation”

205-209: Not sure the relevance of these information to describe the dataset, it feels more like an introduction, or maybe as a part of the discussion to compare with the results.

We thank the review for bringing this to our attention and agree that this should not be part of the description of the data. It will be removed in the next version of the manuscript.

213: punctuation is needed to enhance the flow between cover and we.

Changed.

216: not sure if this is the right reference for key summary of SNOWPACK. This paper is an update status on snow cover modeling in avalanche forecasting including CROCUS and SNOWPACK.

We thank the reviewer for pointing this out. Indeed, Morin et al. (2020) gives a good summary of the key points of the *operational* setup of SNOWPACK, which does not necessarily correspond to our setup. Thus, we will remove this sentence. Please see our response to Major Comment 2 above for the new first paragraph of section 2.4.

219: punctuation is needed to enhance the flow between temperature and we.

Changed.

220-221: punctuation is needed to enhance the flow between (TSS) and we.

Changed.

226: Do you end up with 4 SNOWPACK simulations per warning region? Each simulations have the average grid cell for 4 elevation band? Is 20 the total number per warning region or the entire study area? A sentence that summarizes how many simulations per warning region is needed.

In line 227 we will reformulate and add.:

"…, assuming flat terrain, leading to four SNOWPACK simulations per warning region. This means that for our whole study area of Troms county 20 SNOWPACK simulations are performed…"

230-235: Maybe reduce these lines to one or two sentences, as it limits the comprehension of your methods. We assumed that it is included and it complicates for nothing this section.

We will shorten these lines somewhat. Unfortunately, we are unsure what the reviewer means with "We assumed that it is included,…" Please also refer to our response to the reviewer's next comment below.

257: why explain this? Either remove it or put it into the result.

We are somewhat confused by this comment. One of the major concerns the reviewer states in the general comment is the spatial aggregation of the data. In the mentioned lines we report that we have performed a sensitivity analysis specifically regarding the spatial aggregation of the data. That is, we have tested different ways of spatial aggregation, and this had little to no impact on our results. However, to be clearer, we will make this into its

own paragraph, partly reformulating it, and add another statement about a further sensitivity test we have performed in the meantime:

"To investigate the impact of the strong spatial aggregation on the prediction accuracy we have tested several different ways of spatial aggregation or selection of grid cells. We have generated the predictors for ten specifically wind- and snow-exposed grid cells per avalanche region (SNOWPACK was run for these grid cells specifically as well, see section 2.4), but this did not improve the performance of the machine-learning models. Moreover, we have tested taking the maxima/minima of the features for each individual elevation band (resulting in a much larger number of features), but the impact on model performance was again minimal."

Note that this paragraph will be moved to the new section on the limitations of the study (section 6.4).

258: based, use past tense .

As described above, we prefer a style where we use present tense for the work we do for the present article while earlier work is referred to in past tense. Also note, that in, e.g., line 256 we use past tense for something we have done ("We have tested…") but here we refer to something that did not become part of the paper (here this refers to the analysis based on a different kind of spatial aggregation), which is why we find past tense appropriate here.

Section 3 - Methods

264: you need to state at least the main analysis and parameter we should not need to read another paper.

We are unsure what the reviewer means with "main analysis" here. We recognise that we should have referred to Table E1 here, which lists the precise hyperparameter settings used for the random forest model. We will adjust section 3.1 in the following way:

"To establish the statistical linkage between meteorological data and avalanche danger we employ the widely used random forest (RF) model (Breiman, 2001), which 'grows' a number of decision trees (DTs, Breiman et al., 1984) that 'vote' on the final prediction result. Like Eiselt and Graversen (2025) we use the RF implementation from the Python library scikit-learn version 1.3.0 (https://scikit-learn.org/, last access 23 September 2025). One RF model is trained for each avalanche problem, resulting in four different RF models. The hyperparameter setups for the individual RF models are presented in Table E1 in Appendix E. The data were split into a test (winter 2020/21 and 2022/23) and training (remaining winters) data set."

We hope that from this the reader gets all necessary information from the paper itself and does not need to read our previous work.

265: do you have values or maybe a figure to show the imbalance and the effect of the algorithm.

We will add a figure about the class imbalance in the Supplement. We reproduce the figure here as Fig. 1.

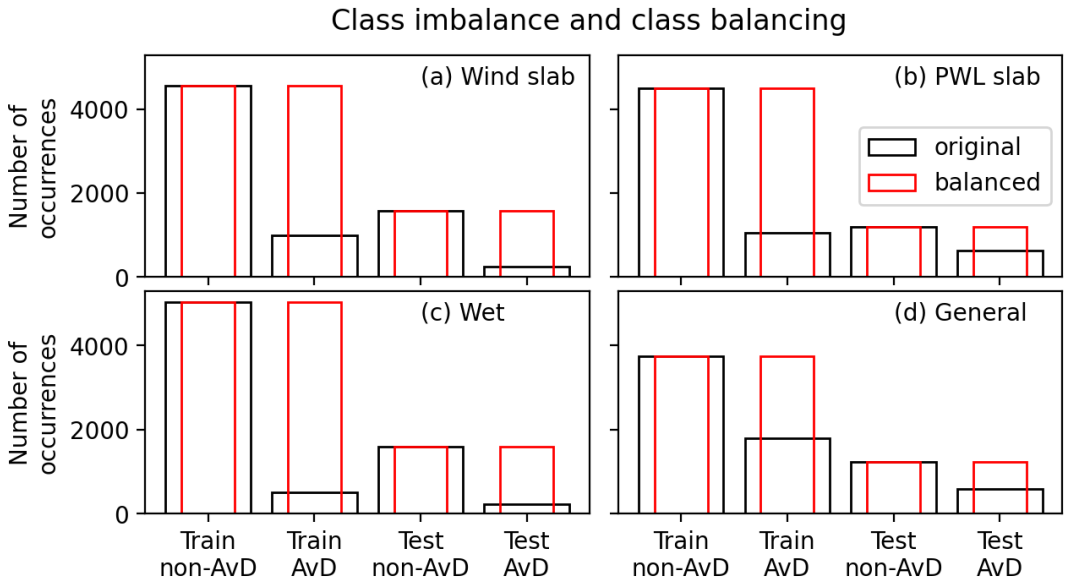

Figure 1: Class imbalance of the training and test data. The number of avalanche days (AvDs) and non-avalanche days (non-AvDs) is shown for the original imbalanced data in black and for the balanced data in red. The balancing was performed with the SMOTE algorithm (see section 3.2).

283: should the F1 score gives that?

We are not fully sure what is meant here, but the F1 score is just one way to aggregate precision and recall score. We wanted to consider several metrics that give more information about the prediction results of our model.

Figure 4: Please adjust the font to match the manuscript, and define what is general? Maybe remove true danger, as danger bring confusion between danger level and avalanche problem.

Regarding the font style please refer to our response to Major Comment 3 above.

We will remove “danger” from the axis labels since we agree with the reviewer that this is confusing. We thank the reviewer for pointing this out. Also, “General” refers to the “General ADL”. We will improve the consistency of this throughout the article.

Section 4 – Model Performance and features importances.

This section also results like section 5.

While it is true that this section also effectively contains results, it is more of an evaluation of the model and we wanted to frame our results mainly in terms of the title of the article, i.e., as the past and future avalanche problems in Troms.

310: the false alarm is also very high.

This is true, but since 100 % minus RC essentially is the false alarm, it would be redundant to state this.

313: please stick to one definition either problem or danger level.

We thank the reviewer for pointing this out and we will increase the consistency on this by using "problem" instead of "danger". We will change this throughout the manuscript (see the tracked changes version).

Section 5 - Results

Section 5.1.1 : please use past tense.

As mentioned earlier, we prefer a style where our current results are presented in present tense while earlier results are presented in past tense.

339 - 340 : please rephrase this sentence.

Without knowing what the reviewer's issue with this sentence is, it is difficult to accommodate this comment. Nevertheless, we will rephrase the sentence to the following:

"We expand the correlation analysis of Eiselt and Graversen (2025) between the general ADF and the AO index by considering the individual avalanche problems."

343 : maybe refer to the figure 8.

Added.

344 : be consistent with fig. Or figure.

We are unsure what the reviewer means here. If this refers to the apparent inconsistency between using "Figure 8" in line 340 and "Fig. 8" in line 343, we want to point out that the style guide to *The Cryosphere* (https://www.the-cryosphere.net/submission.html) says:

'The abbreviation "Fig." should be used when it appears in running text and should be followed by a number unless it comes at the beginning of a sentence, e.g.: "The results are depicted in Fig. 5. Figure 9 reveals that..."'

361 : was this define in the method section.

We are unsure what the reviewer is referring to here. We assume this comment to be a question similar to the following:

"Was this [the Monte-Carlo simulation method] defined in the methods section?"

If this is indeed the question, then the answer is no, this was not defined in the methods section, since we thought that a Monte-Carlo simulation to test the significance of the difference between two samples is a well-known method, that does not require a dedicated explanation. However, we will add a more detailed description in the appendix as follows:

"Appendix B: Monte-Carlo significance test

The significance of the difference between the avalanche-day frequency (ADF) predicted for the different periods within a given future climate scenario (Fig. 9 and Figs. S9–13 in the Supplement) is tested with a Monte-Carlo simulation which is described in the following: Let A and B be two periods (all periods comprise 20 avalanche seasons, i.e., 20 ADF values). The 'observed' statistic is calculated as the mean ADF of climate period A minus the mean ADF of period B. Then period A and B are combined and their ADF values are randomly permuted. The permuted values are subsequently divided again into two periods (A_{test} and B_{test}) and the difference between their mean values is calculated. If this mean difference between A_{test} and B_{test} is larger than the mean difference between A and B a counter is incremented. This procedure is repeated for a given number of permutations (here 100,000). Finally, the p value is calculated by dividing the counter by the number of permutations, giving the fraction of instances in which a random shuffling of the data produced a larger difference than the real difference of scenarios A and B, i.e., the observed statistic."

For transparency we note here that we noticed an error in another instance where we applied a Monte-Carlo simulation, which means that we will remove the sentence (lines 379-380): "However, the difference is statistically significant ($p < 0.05$) only in Indre Troms (based on a Monte-Carlo simulation)."

Section 5.2: there is way more reference to supplemental figures than figure 9, please put these into the text. Figure S9 has more references than figure 9. Or maybe the appendix, which is more accessible.

The frequent references to the supplementary figures only occur to support our statements regarding the statistical significance of the changes. It appears to us that the best way to accommodate the reviewer's comment is to simply remove most of the individual references, since we already state at the beginning of section 5.2 that the significance of the differences is shown in the supplementary figures.

Section 6 - Discussion

Section 6.1: how the precision of the model affects your results especially the PWL.

Partly in response to this comment as well as to reviewer #2, we have performed a further sensitivity analysis, where we trained additional random forest models by excluding different years as test data. We briefly describe this in the new section 6.4 on the limitations of the study and provide a more extensive analysis in a new text and new figures in the Supplement. We find that while the predicted absolute values of the ADF are different between the different random forest models, the changes over time are consistent across models. Thus, we are confident in our conclusions.

419 - 428: I think it might be worth it to discuss these factors between the development and the trigger of the PWL.

When it comes to the meteorological factors, we generally only consider short-term (up to 7 days) changes or maxima/means/minima, and thus we do not discuss the longer-term evolution between development and trigger of the PWL here. We will try to make this clearer in the text. We believe the longer-term evolution to be represented by the SNOWPACK-derived parameters and stability indices.

491: would it be better yrs instead of y.

Changed to “yr”.

497: it might also be warmer and thaw events stabilizing the snowpack.

We thank the reviewer for pointing this out. We will add it as a point in this line. The new paragraph will be:

“Three points may be noted regarding the potential linkage of the AO with the PWL slab AP: (1) The positive correlation of the AO index with temperature may imply thaw events that stabilise the snowpack, leading to fewer PWLs. (2) The generally negative correlation of the PWL slab ADF with the AO index (Fig. 10) may result from the concurrent higher wind slab and wet ADF, reflecting the fact that fewer weak layers persist for a long time as avalanches readily release due to frequent new snow and wind-drift loading. (3) The performance of the RF model in terms of predicting PLW slab ADF is low (Fig. 4), calling into question the robustness of the results regarding this AP (see also section 6.4).”

Section 7 – Summary and conclusions

593: why not write meteorological input as both are spatially aggregated for input to the rf's model.

This sentence will be moved the new section 6.4 on the limitations of the study and thus be slightly reformulated anyway.

597-598: I think this is rather concerning. it was also point out that SNOWPACK struggle to model artic snowpack, because of the high thermal gradient (Domine et al., 2019).

We are grateful for the suggestion of this article, and we will include it as a reference in our manuscript. However, we want to point out that Domine et al. (2019) test the SNOWPACK model for a location (Bylot Island) with a climate rather different from the climate in the county of Troms, due to the generally milder climate in northern Europe compared with North America. For example, according to Domine et al. (2019), the average temperature on Bylot Island is -14.5 °C while in Tromsø it is 2.6 °C, according to Hisdal et al., (2021). Other towns in Troms county are, e.g., Finnsnes with 3.4 °C, Harstad with 4.0 °C and Bardufoss with 1.0 °C. This indicates that the issues found by Domine et al. (2019) may not be as severe in Troms County as they are on Bylot Island. Similarly, the study by van Herwijnen et al. (2024) that we cite in the manuscript shows that the difference between the snowpack in Troms and in the Alps is smaller than that between Alaska and the Alps.

We note that the Norwegian Water and Energy Directorate (NVE) has recently published a new version of SNOWPACK specifically for the Norwegian conditions, but this was too late for our article.

We will include these points in our new section on the limitations of our study (section 6.4).

References

Domine, F., Picard, G., Morin, S., Barrere, M., Madore, J. B., & Langlois, A. (2019). Major issues in simulating some Arctic snowpack properties using current detailed snow physics models: Consequences for the thermal regime and water budget of permafrost. *Journal of Advances in Modeling Earth Systems*, *11*(1), 34-44.