General comments:

This paper addressed past and future avalanche frequencies in northern Norway using the SNOWPACK model and the random forest (RF) model. The target avalanches were not only general problems but also wind slab, persistent weak layer slab, and wet snow avalanche problems. The past avalanches were investigated mainly with consideration of their linkage to the Arctic Oscillation (AO) index, and the avalanche frequencies were well correlated with the AO index. The future dry-snow avalanches would be estimated to decrease, while the wet-snow avalanches would increase until mid-century. The topic and results are valuable for the scientific community. The introduction provided a nice review of the global warming impact on avalanches.

However, I have a concern about the originality of this study. I agree with the authors that this work presents an original case to show future avalanche problems in Norway; however, the other aspects of originality seem limited. The random forest model used had mainly been developed in the authors' previous work. The linkage between avalanches and the AO index had also been found in the authors' previous work. The future estimations, including their procedure, are similar to those in previous works, such as Mayer et al. (2024). I feel that the originality of this work would be insignificant for "The Cryosphere", even though the differences in locations themselves are valuable to the scientific community.

We thank the reviewer for considering our manuscript and for the detailed points. We provide a point-for-point response to the comments below.

We appreciate the reviewer's concerns regarding the novelty of our work, as we feel it is important to address that clearly in the manuscript. The random forest model in our previous work was only developed to be applied to the general avalanche danger and here we develop new models for the individual avalanche problems. In fact, since the danger levels for the individual avalanche problems were not directly available, it was necessary to convert the raw data (size, sensitivity, distribution) from Varsom into danger levels for each avalanche problem and to analyse the problems to find out which of them to focus on (this analysis is presented briefly in section 2.1; see also Fig. 3). Also, we changed our procedure for the preparation of the features to attempt to include more of the underlying variability potentially influencing avalanche danger (as described in section 2.5). Moreover, we utilised a new and different model optimisation procedure (section 3.3). We note that in the manuscript we clearly indicated our departure and further development relative to earlier work. However, we will attempt to make this even clearer, e.g. already in the abstract. Further, while the linkage between avalanche danger in northern Norway and the AO was established in our previous work, again, this was only done for the general avalanche danger, while here we investigate the connection with individual avalanche problems.

Regarding the future projections and the similarities with the work of Mayer et al. (2024), we recognise that our work is conceptually similar, although our application is for a different region having fundamentally different climate and meteorology than that of Mayer et al. (2024). However, as we point out in the article, the data situation in Norway is substantially different from Switzerland and it required considerable additional developments (as detailed in the article) to arrive at our results. Also, we think that it is interesting and encouraging in itself that our results agree with Mayer et al. (2024) and this should not be used as an argument that our work lacks novelty.

The utilization of the RF model also seems problematic. From my understanding, the authors estimated the avalanche-day frequency (ADF) by cumulating the daily 1/0 output from the RF model. However, this procedure might lead to a biased ADF because the RF model was not optimized by minimizing the error of the ADF. Actually, the sum of predicted AvD for wind slab avalanches is 440, while that of true AvD is 245 (Fig. 4), indicating a mean bias towards overestimation in the ADF. I suppose the RF model should be a regression type, rather than a binary type. I recommend confirming the RF model's reproducibility regarding ADF by comparing it to the observation.

We thank the reviewer for pointing out the bias in the model predictions. We will include a new figure in the Supplement that shows the comparison of the distribution of true and predicted AvDs and non-AvDs to make this even clearer. We appreciate the suggestion of using a regression-type RF model, but this would involve a whole new feature aggregation and selection as well as model optimisation procedure and a fully new analysis since the ADF is a seasonally aggregated metric, while our metric (AvD) is a daily metric. Hence, this is outside the scope of this study and is left for future work. However, we have conducted a new sensitivity analysis to account for the bias described by the reviewer. Since this sensitivity analysis also accommodates the following comment, we describe it further below.

This may be related to the above problem, but I am also concerned that the authors did not consider uncertainties arising from the RF model. Seeing Fig. 4, the RF model may produce a very large uncertainty in its projection. For example, the RF model incorrectly predicts general avalanches with probabilities of 36% in AvD predictions and 17% for non-AvD predictions (Fig. 4). I am not certain, but the uncertainty range is comparable to or more than that of climate models. Furthermore, the authors converted AvD/non-AvD from avalanche danger level simply by a threshold (Section 2.1), which also causes uncertainty. However, the authors show no data to discuss this kind of uncertainty arising from the conversion. These problems would change the results of statistical tests for linear trends in past and future avalanche frequencies (Figs. 6, 7, 8 ,9), and if so, the authors' conclusion may be changed. Authors should quantitatively demonstrate the uncertainties associated with past and future projections arising from the RF model, and these uncertainties should be considered in the statistical analysis. This point is crucial for ensuring the reliability of the RF models' estimation.

We appreciate the concerns of the reviewer about uncertainty, and, as mentioned above, we have conducted a new sensitivity analysis to account for this uncertainty. However, please first note that our purpose here is to show the *tendency* of the development of avalanche danger and we are interested in the *change* of the ADF and not its absolute values, and this is especially true when it comes to the linear trends, the AO-ADF linkage, and the future development. As a general note, this is also what is typically done in climate model studies of the future (i.e., the focus on the *anomaly* instead of on the absolute value) since climate models struggle with representing the real-world climate state. This problem is hence not specific to our study and constitutes a general issue with model-based research on climate change. Thus, we believe many of the concerns of the reviewer, while definitely not irrelevant, are at least somewhat overstated. Accordingly, our new sensitivity analysis also focuses on the *tendencies* and not the absolute values. We utilise the fact that when training our random forest model we excluded two years (winters ending in 2021/23) as test data and train three additional models excluding different years (winters ending in 2018/22, 2019/24, 2020/22). We perform the historical and future-projection ADF predictions with the new random forest models and compare the results (including those from the random forest model analysed in detail in the manuscript) to investigate the robustness of our conclusions. While, as expected, the absolute ADF values vary between the models, the tendencies (linear trends, AO-ADF correlation, future development) are similar across models, increasing our confidence in our conclusions. We will include the description of this new sensitivity analysis in our new section 6.4 on the limitations of our study and add a more detailed analysis as well as several new figures to the Supplement to show the comparison between the different random forest models.

Finally, we are unsure how we should "show data" to discuss the uncertainty arising from our *convention* of using the AvD for danger level 3 and larger. Again, this is just a *convention* to facilitate our study on the *tendency* or *change* of the avalanche danger. Please note that (as also noted by reviewer #3) the danger levels do not necessarily correspond to a probability of avalanche occurrence across time and space and, hence, a quantitative uncertainty analysis appears inappropriate here. For convenience, we reproduce our justification for our AvD/non-AvD convention from our earlier work (Eiselt and Graversen, 2025) here:

"Furthermore, the ADF appears related to avalanche activity, since Pérez-Guillén et al. (2024a) in a case study in the Swiss Alps using an automated seismic avalanche detection system found that on days with no avalanche, the mean ADL was 1.9±0.8, while on days with at least one avalanche, it was 3.2±0.5, hence providing a clearly binary appearance. Similarly, in an investigation of Swiss backcountry GPX tracks as a proxy for non-avalanche events, Techel et al. (2024) found that for non-events the median probability of ADL ≥ 3 was only 0.14, while for events it was 0.58. Hence, on a day with ADL 3 or 4, avalanche events are likely, while they are unlikely on days with ADL 1 or 2, justifying our definition of AvD and non-AvD."

Specific comments:

L41: "RPCs" seems to be a typo instead of "RCPs".

We thank the reviewer for noting this oversight.

L46: You need to define the abbreviation NorCP here.

Again, thank you for spotting this oversight.

L55: From my understanding, Lazar and Williams (2008) assessed a potential avalanche period very simply based on air temperature exceeding 0 °C or not. Although I do not want to treat authors' opinions carelessly, I disagree with this.

We thank the reviewer for this assessment, and we agree with it. We will add a caution about this simplified approach by Lazar and Williams (2008) in our text, changing the beginning of the paragraph to:

"To the authors' knowledge, Martin et al. (2001) was the first study investigating the change of avalanche activity under changing climatic conditions based on a statistical linkage between meteorological parameters and avalanches. By implementing constant positive perturbations of temperature and precipitation in their statistical model, they found for a study area in the French Alps that while new-snow avalanches declined, wet-snow avalanches increased in frequency (at least relatively). Lazar and Williams (2008) were likely the first to employ future emission scenarios to investigate future development of avalanche activity or danger, although they simply defined their avalanche periods based on a temperature threshold."

L105–121: These contents are better moved to Section 2.

We thank the reviewer for this suggestion. However, we view this as a motivation and a very short introduction to our study and thus would like to keep it in the introductory section of our article.

L133: A dual abbreviation definition of ADL.

We will remove the definition here.

L136: What are the active avalanche problems?

This was our way of indicating that these problems are the ones relevant for that specific day. We will change “active” to “relevant”.

L140: What are distribution and sensitivity?

We will extend this part of the sentence to:

“…, the latter being derived from the spatial distribution of hazardous sites and the sensitivity to triggers (Müller et al., 2016a, b; Statham et al., 2018; Müller et al., 2023).”

We hope this makes clear what is meant by distribution and sensitivity.

Figure 3: Is the left axis showing the number of avalanche days? What is the avalanche problem frequency?

We will change the caption of this figure. Please note that further changes in the caption were necessary to accommodate a comment by reviewer #1:

“Average danger level per avalanche problem (AP; red) and the number of days on which the AP was identified by the forecasters (black) in northern Norway. The average danger level was calculated only for the days the specific AP was identified. The ADL (avalanche danger level) on the x axis refers to the general avalanche problem. The data cover the period from winter 2016/17 to 2024/25 for the general avalanche problem and 2017/18 to 2024/25 for the other APs.”

We will also change the figure slightly: The red colour will be moved to the left to be consistent with the other figures and individual avalanche problems on the x axis will be reordered according to the frequency of occurrence.

Section 2.4: Please describe the model settings for soil.

We will add the settings we have used although we did not change the default.

Section 2.4: How did you calculate liquid water content (LWC)? LWC is very important for wet avalanches (Fig. 5). Furthermore, local LWC exceeding 5% is very important for wet-avalanche predictions (Wever et al. 2016). This point should be taken into account.

We are unsure what the reviewer exactly means here. We use the standard LWC output from SNOWPACK. The parametrsiation used to simulate liquid water flow through the snow is the bucket scheme. We will add this information to section 2.4

Section 2.4: Please describe how you obtain daily snowpack variables. The original output of the SNOWPACK model is generally hourly data, but you use daily avalanche data.

Please note that this is precisely described in section 2.5, the whole purpose of which is the description of the procedure we utilise to obtain the features that represent the input for the random forest model. More information is also given in Table B2 in the appendix which lists the features derived from the SNOWPACK output and gives a brief description of each of them.

L218: How do you prepare long-wave radiation data?

We do not use any long-wave radiation data as this is not required by the model if one gives the surface temperature. We provide the link to the SNOWPACK documentation where this is stated: https://snowpack.slf.ch/doc-release/html/requirements.html

L218: You used the net short-wave radiation. So, you mean that the albedo depends on a land surface model implemented in a meteorological model? If so, does this affect the SNOWPACK simulation? The snowpack calculation is very sensitive to the short-wave radiation.

Yes, this is a weakness of our data input as we do not have the incoming and outgoing components of the net radiation available. However, given the large-scale spatial aggregation of our data, we caution against overinterpreting the influence of this on the SNOWPACK calculation.

L220: The linear model should be described in the Appendix or Supplement.

We will add a short description of the model as well as the model equation and a new plot showing the relationship of predicted and true TSS to the Supplement.

L226: How did you calculate precipitation, wind, and relative humidity? A simple arithmetic mean is generally inappropriate for these variables.

The data were spatially averaged by simple arithmetic mean (see section 2.5). While this may be “inappropriate” as the reviewer states, we also performed several sensitivity tests, trying different aggregations as described in the lines below. Since this did not have any noticeable impact on our results, we decided to stay with the arithmetic mean approach.

L228–235: These lines should be described in the Appendix or Supplement.

In response to reviewer #1 as well as the last comment below we will add a dedicated section on the limitations of our study (section 6.4). We will move these lines there. Please note that the point of including these lines was to anticipate a comment such as the previous one and to defend our choice of the averaging methodology we use.

Section 2.5: This content is too hard for readers without a background in the RF model. Can you merge this content into Section 3?

Maybe there is some kind of misunderstanding here and we are unsure what the reviewer means. The content of section 2.5 in itself has nothing to do with the random forest model and is simply a description of how we aggregate the raw input data from NORA3, NorCP, and SNOWPACK to obtain daily values.

L273: What are min_samples_leaf, min_samples_split, max_depth, n_estimators, and max_features?

We thank the reviewer for pointing out our oversight of not explaining the hyperparameters. We will change this sentence to:

"This variation mostly derives from the hyperparmeters min_samples_leaf (MSL) and min_samples_split (MSS), representing the number of samples that remain at a leaf and after a split, respectively. The other hyperparameters (see Table E1 in Appendix E) appear to have a much smaller influence (not shown). We note that MSL and MSS have a similar effect, both determining the number of samples at the leaves of the DTs in the RF. In fact, MSL is the "finer" tuning parameter and MSS has no impact if MSS ≤ MSL. Given these hyperparamerter dependencies we deviate from the model optimisation and feature selection procedure conducted in earlier work (e.g. Pérez-Guillén et al., 2022, Hendrick et al., 2023, Eiselt and Graversen, 2025); that is, we do not perform a randomised grid search over all the different hyperparameters, but instead only test different values of MSL, while holding the other hyperparameters constant (Table E1 in Appendix E)."

Note that, as described in this new formulation, we have made the decision to change the procedure slightly and now only vary MSL since MSS has no effect if it is smaller than MSL. This required an update of almost all figures, since we re-optimised and re-trained the random forest model. However, this resulted only in small changes in the details of the analysis, which may be seen in the tracked changes document.

We will also change our description of the random forest in section 3.1:

"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."

L285: You mean leave-one-out cross-validation? However, your procedure is not the leave-one-out cross-validation, but the k-fold cross-validation, actually. Leave-one-out cross-validation is a method in which a single independent data point is excluded from the training data. In this study, a single independent data is a 1/0 in a day, not a year.

We thank the reviewer for pointing this out and we will change it in the text, clarifying that one "fold" corresponds to one winter. We were rather referring to the fact that we always leave one winter out, i.e., it is a leave-one-winter-out validation.

L286: I do not understand why five years of training data are available even though you have Norwegian avalanche bulletin's data from 2017/18 to 2024/25.

We apologise for the lack of clarity here: As is typical for machine-learning training, we spilt our data in a training and test data set. We exclude two seasons (winter 2020/21, and 2022/23) as test data, leaving five years of data for training. We will add this information in section 3.1.

L509–512: This is also problematic from the viewpoint of the applicability of RF models to future climate. Does the RF model linearly increase the wet-snow ADF by increasing air temperature (or liquid water content) if only there were enough snowpack? However, one of the necessary conditions for wet avalanches is a high liquid water content, locally exceeding 5% (Wever et al., 2016). Satisfying this condition, wetting of an initially below-freezing snowpack is important (Mitterer et al. 2011). Capillary barriers or melt–freeze crusts are also key phenomena. Therefore, the authors need to confirm whether the models' behavior in linearly increasing wet-snow ADF by increasing air temperature is really appropriate in Norway.

We thank the reviewer for these suggestions and for pointing out the important references. However, since the random forest model methodology is rather a black box and (as can be seen in Table D1 in Appendix D) grows 500 decision trees of depth 50 which is impossible for the human mind to comprehend, it is infeasible to perform the kind of investigation suggested here. Please note that there may be a misunderstanding here: The random forest model does not predict the ADF itself. It rather predicts if a certain day is an avalanche day or if it is not and it uses thresholds to do this. Thus, a random forest model is certainly not linear. The only thing we can say is that the temperature is an important feature that has some explanatory power for the wet avalanche problem. In preliminary work after the manuscript was submitted, we attempted to use the Shapley value methodology (as was done by Pérez-Guillén et al., 2025, for Switzerland) to get more information on how the individual features affect the random forest model prediction, but this only tells us that generally higher values of the feature t_max (i.e., daily maximum temperature) make it more likely that a wet avalanche day is predicted, while lower values make it less likely. Thus, such a detailed analysis may be possible with a much simpler decision tree model, maybe of depth 3 or 4, meaning most of the features we use would need to be excluded. Something like this was done by, e.g., Mitterer and Schweizer (2013). But given the severe spatial aggregation of our data and the large areas of our warning regions we are unsure of the utility of such exact thresholds. However, it may be an interesting avenue for future research.

References:

Mitterer, C. and Schweizer, J.: Analysis of snow-atmosphere energy balance during wet-snow instabilities and implications for avalanche prediction, The Cryosphere, pp. 205–216, https://doi.org/10.5194/tc-7-205-2013, 2013.

Pérez-Guillén, C., Techel, F., Volpi, M., and van Herwijnen, A.: Assessing the performance and explainability of an avalanche danger forecast model, Nat. Hazards Earth Syst. Sci., 25, 1331–1351, https://doi.org/10.5194/nhess-25-1331-2025, 2025.

L590–637: These lines should be described in Section 6.

We thank the reviewer for the suggestion. The reviewer #1 had a similar comment, and we will move these lines to our new section “6.4 Limitations”.

References:

Mayer, S., Hendrick, M., Michel, A., Richter, B., Schweizer, J., Wernli, H., and van Herwijnen, A.: Impact of climate change on snow avalanche activity in the Swiss Alps, The Cryosphere, 18, 5495–5517, https://doi.org/10.5194/tc-18-5495-2024, 2024.

Wever, N., C. Vera Valero, and C. Fierz (2016), Assessing wet snow avalanche activity using detailed physics based snowpack simulations, Geophys. Res. Lett., 43, 5732–5740, doi:10.1002/2016GL068428.

Mitterer C, Hirashima H, Schweizer J. Wet-snow instabilities: comparison of measured and modelled liquid water content and snow stratigraphy. Annals of Glaciology. 2011;52(58):201-208. doi:10.3189/172756411797252077