

Response to Reviewers – Round 2

Editor

We thank the editor for his comments on our manuscript and for the suggested changes. We have incorporated most of them into the updated version of the manuscript. Regarding the footnotes we have taken the shorter links into the text and the longer links (as suggested) into the references. The other footnotes are either incorporated directly into the text or removed. We have taken another look at the journal's guidelines and attempted to correct instances where we did not follow them. We also thank the editor for the suggestion of the “avalanche day” which we have now implemented in the manuscript. This has led to the removal of the “binary-case level” as we now only distinguish between non-avalanche day (non-AvD) and avalanche day (AvD). The former binary-case frequency (BCF) is now the avalanche-day frequency (ADF). Consequently, some reformulations and updates of the figures were required, which are now implemented in the updated version of the manuscript.

We have attempted to separate results and comparison to earlier work (section 5.1) slightly more, hoping this has increased the clarity.

Finally, we note for transparency that we found an error in our code, which required updating Figs. 7-9 as well as part of the supplementary material. However, this only marginally affected the results (i.e., the correlations of AO and ADF) and had no impact on our conclusions. The error is also corrected in the online published version of the code.

Reviewer #1

The manuscript with the title "Predicting avalanche danger in Northern Norway using statistical models" has been revised. The authors have seriously discussed the suggestions and considered the recommendations. This has considerably improved the quality of the content of the manuscript. From my point of view as a reviewer, the manuscript now fulfills the requirements of a scientific article.

Technical corrections

The footnotes often include a date. However, the date format does not comply with the guideline for EGU publications (see Mathematical notation and terminology):

Date and time: 25 July 2007 (dd month yyyy), 15:17:02 (hh:mm:ss). Often it is necessary to specify the time if referring to local time or universal time coordinated. This can be done by adding "LT" or "UTC", respectively. If needed when referring to years, CE (common era) and BCE (before the common era) should be used instead of AD and BC since CE and BCE are more appropriate in interfaith dialogue and science.

We thank the reviewer again for considering the updated manuscript and the kind response. We are grateful for the notice about the incorrect date format in the footnotes. The dates (now to be found in the references; see our response to the editor above) should now correspond to the guidelines.

Reviewer #2

In this study, several machine learning models for predicting avalanche danger levels have been developed. The authors have made significant improvements in the revised version of the manuscript. However, the results' writing, structure, and presentation could be further modified to enhance readability. I recommend publishing this paper after addressing the following suggestions:

We thank the reviewer for considering our updated manuscript and for the recognition of the improvements.

I suggest a review of the English and the structure of the paper. The models' development, performance, and a discussion of the results of the models with earlier studies could be separated.

We acknowledge that our English is not perfect as we are not native speakers. We have read through the whole manuscript again and attempted to carefully improve the English. We have separated the presentation of the results from the comparison to other studies in section 5.1 and hope that this is now more acceptable.

However, we once again want to explain our choice of integrated presentation of results and discussion: In our opinion, the separation of results and discussion often either necessitates the redundant restatement of at least some results, or the reader is constantly referred back and forth between results and discussion. We recognise that the separation between results and discussion can in some cases add clarity and structure, and we appreciate the suggestions, but we here opted for the first approach and would like to keep this way.

· I still find the introduction quite long and suggest that some parts could be moved to other sections (please see the previous review's comments).

We have (also following the suggestions of the editor) removed and shortened some parts of the introduction (see especially lines 116-131 in the new version).

· Section 4: The features are described sometimes using abbreviations and others with full names followed by abbreviations in parentheses. This should be presented consistently. Additionally, the comparison of the main important features of the Random Forest models with previous studies should be moved to a discussion section instead of being included in the description of the model development and optimization.

We thank the reviewer for pointing out the inconsistency in this section. However, we believe that we were mostly consistent as in fact the only time we used the full name was for the parameter `nsw7` and otherwise we only used the formulation "parameters related to ..." followed by the abbreviations in parentheses. We have now used a similar formulation ("parameter associated with...") followed by the abbreviation also for the parameter `nsw7` (now lines 353-354). For transparency we mention that this comment also made us aware of an inconsistency in the naming conventions in and between Table 2 and Fig. 3 which is now corrected. We thank the reviewer again.

Since we do not include a separate discussion section (see our response above for another attempt to explain our approach), we would like to keep the comparisons with earlier studies where they are.

· Section 5: The results are mixed with discussions and comparisons to previous studies, which makes the reading difficult. For instance, the model evaluation in Section 5.1 starts with a comparison of the overall performance of the model (balanced or unbalanced?) with previous studies, followed by the presentation of the performance of the two models. Furthermore, the distinction between the balanced and unbalanced models is unclear. The performance of these two approaches is sometimes

presented together in the main text (e.g., Table 4) and, at other times, separately in the supplementary material. Also, what is the advantage of applying the balanced strategy if the overall performance of the balanced model is lower (Table 4)? A comparison of the overall performance of both methods and per danger level will help to understand why developing models using both strategies. Which model has been used for Hindcasting avalanche danger (balanced or unbalanced)?

Please see our response above for another attempt to explain our approach to not separate results and discussion. We have attempted to improve the structure of section 5.1.

Moreover, there seems to be a misunderstanding about the balanced and unbalanced “model”. We only trained *one* model based on balanced data as described in sections 3.2 and 4; no model based on unbalanced data was trained. The balanced and unbalanced performance referred to in the text, tables, and figures relates to the balanced or unbalanced *test data*. We recognise that while this was made clear in Figs. 5 and 6 as well as Table 5, in Table 4 and parts of the text it is somewhat ambiguous. We apologise for this and have changed it to make it clear that we are referring to the model performance with respect to the (un)balanced test data. Furthermore, we have added more clarification (that the balanced test data have been oversampled using the SMOTE technique to equalise class frequency) in the brief introduction to section 5 (now lines 375-383).

· Supplement: The results of the artificial neural networks are only shown and mentioned in the supplementary material. I suggest removing this if it is not more discussed or compared with the performance of the Random Forest models.

We have removed the part about the ANN from the supplementary material and now only briefly mention at the end of section 5.1 (now lines 449-451) that we have trained an ANN (as presented in Sharma et al., 2023) on the same training-test data setup as our RF model and obtained a similar performance.