

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/nhess-2021-341-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2021-341

Karsten Müller (Referee)

Referee comment on "Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland" by Cristina Pérez-Guillén et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2021-341-RC2>, 2022

Review of Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland

General comments

The paper presents the development of a machine-learning model capable of assessing the avalanche danger level based on input data from automatic weather stations and a snowpack model in the Swiss Alps. The models are trained using a large data set of forecasted danger levels and a filtered subset of "re-assessed" danger levels from local nowcasts.

Compared to previous studies the presented paper uses a much larger and well-refined data set. The trained machine-learning models achieve performances comparable to human forecasters throughout the region of the Swiss Alps. Previous studies did either have either poorer performance or were more limited in their spatial extend.

The topic is of scientific interest and value for avalanche researchers, forecasting services and stakeholders. The topic is within the scope of NHESS.

The authors present their study in a clear manner. The manuscript is well written and structured.

The abstract provides a good summary of the goals, methods and conclusions of the presented study.

Tables and figures are of high quality and readability contributing to the good overall impression of the paper. The language is precise and understandable.

The paper is long. However, it combines the field of avalanche forecasting and machine-learning using the Random Forest algorithm and needs to (and does) explain both concepts to the reader potentially being unfamiliar with one or both of them. I therefore only have minor suggestion on how to shorten it - see specific comments.

It is not clear from this paper how you apply or intend to apply the model in a forecasting setting since it is trained and run on input data measured and modeled at an automatic weather station. I also miss a discussion on the how to apply the models in an operational setting and the expected benefits in supporting the human avalanche forecaster - see specific comments.

Specific comments

I-171 Your models are trained on station data. That means they require a measurement and a subsequent SNOWPACK model output to be applied. Thus, RF#1 and RF#2 as described in this paper only provide a hindcast or nowcast.

In order to be used operational your models need be run with input data from weather prediction models and the corresponding output from SNOWPACK at the location of IMIS stations. As far as I can see this is not addressed in your paper. Please add or reference information on how this is or could be done. I expect that the transition from the spatial resolution of the weather model to the station site (especially in mountainous terrain) poses some scaling issues which might have an effect on performance/accuracy. This should be addressed in the discussion e.g. in connection to section 6.3.

I-207 Why do you only filter by elevation and not by aspect? I assume you do not filter by aspect because most (all used?) IMIS stations are on a flat field and thus cannot be assigned an aspect. Please add a short explanation.

I-216 It seems legitimate to use the most recent winter seasons as test data. However, it should be ensured and stated that these do not exhibit any special avalanche conditions not or barely seen during previous winters - have you considered/tested a random draw from all data with an equal amount from each month as an alternative? If yes, what was the effect on model accuracy.

I-275 "Note that this last step..." - what do you mean by this sentence? It is not clear to me to which "last step" you refer and what the effect on model performance is. Could you clarify?

I-355 While the section "Exemplary case studies" is useful for the reader in order to get an overview over potential model outcomes in relation to published avalanche forecasts, it is not necessary for the understanding of the paper. Considering that the paper is already very long, I suggest to move this section and Fig.8 to the Appendix or provide it as supplementary material.

I-328 What is the "daily averaged accuracy"? Is it the average of the predictions from RF#1 and RF#2 or is it the average of the results from all stations within a forecasting

region with regard to Dforecast for that region?

I-405 The last two sentences in this section should be revised. I understand it such that performance was lower because the danger levels (1 and 3) - that have highest prediction performance - are less common in these regions. However, I had to read it several times to understand what you mean.

It would also be interesting to know if you could identify common traits for stations/sites that had a high accuracy (e.g. >0.8): specific elevations, typical snow or weather conditions?

I-540 see comment for I-171

I-573 Your features include several stability indices and information on weak layers. Does that mean the provided stability information from SNOWPACK is not good enough to detect/predict persistent weak layers or the stability related to them?

I-591 Could you discuss the intended operational application of the models and their main benefits to the human forecaster in more depth. I could imagine that the models would be useful in deciding when to increase or decrease the danger level and to assess the spatial or temporal extend of a given danger level.

I-602 It would also be interesting to know in the discussion what your expectations on model performance are. I would argue that your results are as best as it can get. You state that a human forecaster has an average accuracy of 76%. You use the assessment by the human forecaster as your labels. Thus, the model inherits human mistakes and biases. For RF#2 these biases are somewhat corrected for or at least replaced by biases or mistakes in human assessed nowcasts.

I-603 It is not clear from you paper that your model "predicts" avalanche danger. I read it that your model can be used to validate or quality control a published forecast once data has been measured at an IMIS station.

I-610 see comment for I-573

Technical comments

I-141 "...which jointly account for more than 75% of the cases." Change to "which jointly account for 77% of the cases."

Fig.3 ideally the y-axis of the DL proportion [%] plot for Dforecast would have the same maximum value - currently these are 50% and 40%.

I-311 "...the two models...", missing "s"

I-318 remove one "particularly"

I-422 spelling "Eq. 1"

I-463 Split this sentence in two.

I-474 "...only the 10%..." - remove "the"

I-581 Change to "..., predicting high probabilities for both danger levels."

I-587 remove one "the" and the end of the line