Reply to Erich Peitzsch
June-, 2021

Thank You for presenting a detailed review of the manuscript. Your comments will improve the manuscript greatly.

1. Reply of General Comments

1.1 Restructuring

The manuscript requires further restructuring. The Introduction is thorough and clear up until Section 4 where methods are mixed with model tuning results, and it is difficult to keep track of methods and results. I suggest having distinct Results sections on model performance and model generalization that are separate from the description of the implementation of the model (i.e. methods). This should make the results clearer to the reader.

Presently section 4 of manuscript includes description of methods used for performance measures, cross-validation and model generalisation. As suggested, these can be included in a separate methods section. Accordingly, following restructuring changes can be made in the revised manuscript:

1. Section 4.2 [Performance measures] can be moved into methods section.

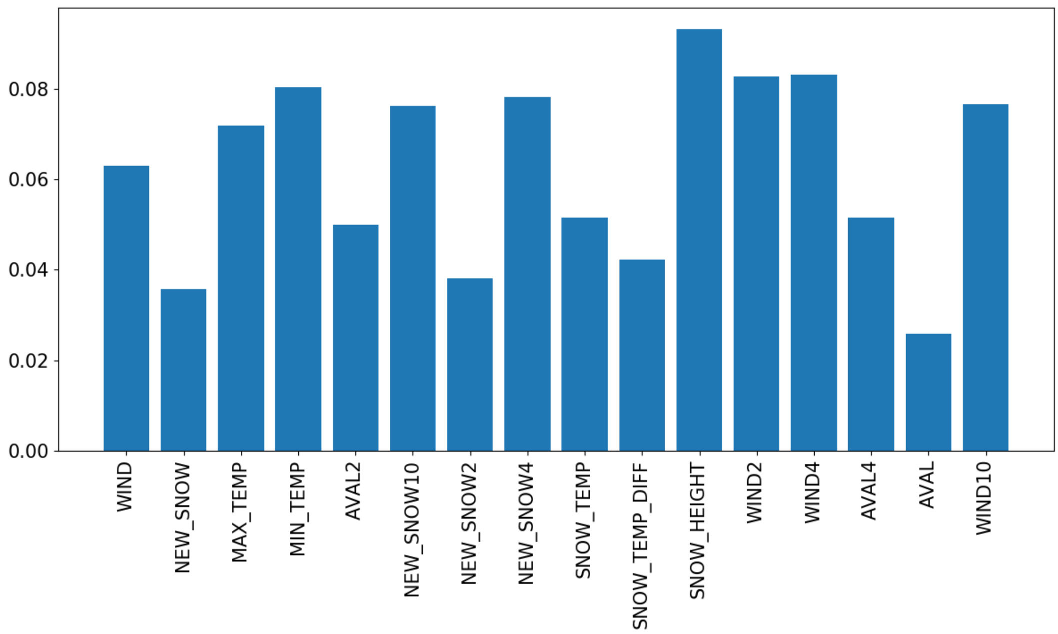
2. From Section 4.3 [Model training and hyper parameters training] the details of hyper parameter training can be moved into a sub-section under methods section.

3. From Section 4.4 [Model Generalisation] the details of how the model was trained and tested will be moved to a sub-section of methods section. These details will only be referred from the section on model generalisation.

1.2 Interpretation of the physical process described by overall ensemble

The authors state that the advantage of the ensemble RF model is due to the ability of the model to provide information on the stochastic nature of the process. The authors provide evidence of individual trees detailing the line of reasoning, but I don't see interpretation of the physical processes described by the overall ensemble. The authors detail the predictive capabilities of the model by presenting POD, FAR, and HSS scores and provide examples of specific trees within the ensemble (which is good), but not how the overall ensemble provides insight into the physical processes at play. Presenting the overall variable importance and describing what this means for avalanche occurrence, in general, in this region would be helpful. The two specific examples provided in Figures 8 and 9 are a nice way to visualize individual days, but a clearer explanation of how the overall ensemble model describes the general processes at play and enhances avalanche forecasting is necessary.

We have computed the importance score for each training feature, presented in the bar graph below.



Some observations:

1. SNOW_HEIGHT has the highest contribution in avalanche formation followed by the variable for cumulative snow-fall of past few days.

2. NEW_SNOW has low contribution, yet the cumulative snow fall history is important(NEW_SNOW10, NEW_SNOW4).

3. AVAL has low contribution, yet the cumulative avalanche history is important (AVAL4 , AVAL2).

From 2 and 3, we may believe that variables showing past 4 day snow instability (AVAL4) and 4 day snow fall are more valuable for avalanche forecasting than the variables showing immediate instability and snowfall. The weather and snowpack history of past few days contributes in complicated ways to increase hazard. We will analyse this in future work.

2. Specific Comments

line 82: It would be very useful to incorporate how RF models have already been used in the avalanche literature in this section (e.g.(Mitterer and Schweizer, 2013; Guy and Birkeland, 2013;Marienthal et al., 2015; Dreier et al., 2016)). Currently, there are no references to how RF has been used in avalanche research, specifically. In addition, results from your study compared to these studies should be interpreted in the Discussion.

These references will be included and discussed in our revised manuscript.

lines 98-111: Classification trees and even RF (see comment above) have been used enough in avalanche research that this detailed description probably isn't necessary. Consider providing a succinct sentence overview of the process and directing the reader to Breiman, 2001 and other avalanche studies that use these techniques.

This can be done in the revised manuscript, however referee 2 asked us to rewrite this clearly. Hence we propose to retain detailed description (though in a revised manner as suggested by refree 2)

line 115-137: I appreciate the technical detail and description of the C4.5 algorithm RF method as I am not familiar with this specific RF approach. However, I suggest condensing all of Section 2 to a broad and succinct overview with appropriate references for the reader and then referring the reader to a Methods supplement which contains the specific equations and more detail. Lines 138-144 already provide a start to the succinct summary of RF.

As suggested, the Section 2 can be further condensed and shifted under 'Methods' section.

line 163: The avalanche occurrence dataset is very important in this study. Can you please describe in more detail how avalanches are observed? I assume that all avalanche observations are derived using infrasonic and seismic sensors since line 165 states "all necessary input data for the proposed model may be recorded using automatic devices". Is there some sort of manual observer network or at least some manual validation of the infrasonic and seismic signal? Also, how much of the study area is instrumented? In other words, what proportion of the area are you able to detect avalanches on a daily time step? The quality of the avalanche dataset will likely have a great influence on the results. So, some discussion here of the quality control and/or limitations of the avalanche dataset would be useful, particularly the signal validation component (i.e. false alarms from the infrasonic and seismic sensors).

All the snow-meteorological and avalanche observations used in this work were observed and recorded manually using traditional instruments and visual observations. The details of snow-met observatory is given in line 154. The observatory is manned by trained persons. The avalanche occurrences in the remote sections of study area are reported by the villagers and Indian Army troops manning the border outposts. It is likely that some occurrences might have passed unreported. As rightly pointed out that this has an implication on the quality of dataset and in turn on the results. But this limitation about avalanche occurrence observations is well recognised by avalanche research community. However, the point that we wanted to convey by line 162-166 is that all these features can also be recorded automatically as the technology for the same exists. If this technology is adopted, it may remove dependency on manual presence to observe and record data (snow-met as well as avalanche occurrences). In other words, the data requirements of proposed model can be met with commercially available automatic systems. At this juncture, the issue of data quality is beyond the scope of the subject matter of this paper, as the paper focuses on the RF as a technique for decision making, interpretability of results in terms of physical process responsible for the event, and the potential it carries to be used as an autonomous tool. However, in order to avoid any confusion and make the point more clear, the paragraph (line 162-166) will be re-written in revised manuscript.

line 170-172: Are these seasons similar in input variables? If so, was this tested statistically? I suspect that if the variables in the seasons in the training dataset are significantly different than the test dataset, then this would adversely affect the model results? Can you provide some insight on this? Also, please provide sample size (n=?) for the training and test datasets, respectively.

These seasons are not similar in all input variables. The testing dataset (Dec 2013- Mar 2017) (n=485) carries higher temperature and lower standing snow than the training dataset (Dec 2010- Mar 2013) (n=364). This was tested statistically. We believe this does not affect the model performance. Our model gives the conditional probability of an avalanche given the input variables, the relation of avalanches on the factors should stay the same even if the distribution of those factors changes during the seasons. For example:

Number of days in training dataset with snow-fall > 20cm was: 36
Number of days in testing dataset with snow-fall > 20cm was: 26

We observe that snow-fall days were fewer in the testing dataset, despite it being of longer duration than training dataset. Yet the probability values of an avalanche occurring when snow-fall > 20 cm were found to be 0.55 for training dataset and 0.5 for testing dataset which are comparable.

line 273-275: This statement is confusing to me as currently written. Did you use the temperature values from lines 269-270 as thresholds? What are the “temperature bounds” in Table 6? It is unclear. Also, this should be presented in the methods, not the results. See General Comments re: restructuring

We filtered the dataset using the following rule:

$-2.75^{\circ}C < MIN\_TEMP \leq -0.75^{\circ}C$ AND $MAX\_TEMP \geq 1.75^{\circ}C$

The bounds on MIN_TEMP and MAX_TEMP have been called temperature bounds and have been adopted from Figure 9. In this filtered dataset we found a greater percentage of avalanche days than the original dataset. The increase in avalanche hazard may have been caused by other factors which occur more frequently within these temperature bounds. We did a univariate analysis to rule out such factors: if the distribution of a variable in the filtered and unfiltered datasets differed, it might be contributing in increasing the hazard when the temperature bounds are satisfied. Therefore such variables were further analysed and results compiled in Table 6.
The above analysis is justified in ‘results’ section and not ‘methods’.

line 277: Again, what exactly is the “temperature bound rule”?

A day satisfies the temperature bound rule if the MAX_TEMP and MIN_TEMP recorded that day satisfy the following:

$-2.75^{\circ}C < MIN\_TEMP \leq -0.75^{\circ}C$ AND $MAX\_TEMP \geq 1.75^{\circ}C$

Table 6: Please list the sample size of the full dataset (avalanche and non-avalanche days). Also, I assume that this is the dataset from the RF model output using N=5000, and not the original observed avalanche dataset. Please clarify this in the caption and text.

This is the complete dataset (training + testing) from Dec 2010 – Mar 2017. The size (n) of samples used for each estimate has been mentioned in the table.

Including the training dataset from which we derive the rule may bias the results to support the rule. Therefore we have made an additional table using only the testing dataset as well. The table is consistent with all the conclusions drawn from Table 6. We have given this table in appendix section of this reply.

line 284: Please explain how the model is able to infer the formation of a melt-freeze crust? As I understand it the model provides the probability of avalanche occurrence based on the input parameters? Can you really extend this to mean that a M/F crust formed since this is not a snowpack model? Assuming the “temperature bound rule” uses the values presented earlier, it’s not clear to me how we can infer the formation of a melt-freeze crust from these data/trees? Please provide a clearer line of reasoning/evidence to support this claim.

A day satisfies the temperature bound rule if the MAX_TEMP and MIN_TEMP recorded that day satisfy the following:

$-2.75^{o}C <$ MIN_TEMP $\leq -0.75^{o}C$ AND MAX_TEMP $\geq 1.75^{o}C$

If a day satisfies this rule then it has a sub-zero night temperature and above-zero temperature during the day. Melting is caused during day when temperature is greater than zero. The melted water on surface then refreezes again during night to form a crust.

Formation of M/F crust explains why the decision tree in Figure 9 shows high avalanche probability. Deduction of M/F crust from input features requires knowledge of physics not present in model, but in this situation the reasoning used by model is consistent with reasoning based on physical effects: M/F crust and higher snow density. The model was able to infer the effect of these physical effects without using information about them directly as we have not included the M/F crust or snow density in the input features set.

lines 291-304: This section seems out of place and should probably be placed in the Introduction.
This will be corrected in revised manuscript.

line 313: Can you discuss how your results actually compare to other studies and provide some interpretation on this rather than presenting the table?
Comparison using skill scores is provided in the table. We have also provided discussion on descriptive forecasts and data efficiency etc. vis-à-vis other models.

line 329: How can the model account for “situations not encountered before” if they don’t exist in the training dataset?

By “situations not encountered before” we mean situations to which the model generalises from the training set e.g: there may be no situations during the training when snow-height > 3m, but we can expect the model to generalise to such situations. But model may not generalise to all future situations. This may be due to very few examples near similar to that situation, poor model bias, missing data etc.

3. Technical Corrections

lines 104 and 107: change “till” to “until”.

line 126-129: font size seems different

line 218: HSS is spelled out in Table 4, but should be done so in the main text body as well.

Figure 5: Please include “results from test dataset” in caption.

These will be corrected in revision.

4. Appendix

Statistic	**Unfiltered Dataset**	**Dataset Filtered by Temperature Bounds**	**Dataset Filtered by SNOW_HEIGHT > 1.0m**
Proportion of Avalanche days	0.18 (n = 485)	0.33 (n = 57)	0.32 (n = 164)
Mean Snow Height	0.74m (n = 485)	1.13m (n = 57)	1.37m (n = 164)
Proportion of Avalanche days when NEW_SNOW > 0 m	0.3 (n = 169)	0.51 (n = 29)	0.45 (n = 78)
Proportion of Avalanche days when NEW_SNOW > 0.1 m	0.35 (n = 48)	0.61 (n = 18)	0.43 (n = 48)
Proportion of Avalanche days when NEW_SNOW> 0.2 m	0.34 (n = 55)	0.58(n = 7)	0.37 (n = 27)

Table: A revision of Table 6 given in the manuscript, using only the testing data.