Review of "A data efficient machine learning model for autonomous operational avalanche forecasting" by Chawla and Singh

General Comments

The authors present a study that uses an avalanche occurrence dataset from The Great Himalaya Range in northwest India during the period 2010-2013 to train a Random Forest model. They then test this model on avalanche occurrence data from 2013/14 to 2017. Overall, the work is interesting and highlights the usefulness of a statistical model in avalanche forecasting from data sparse regions. The authors appear to apply the techniques appropriately and the use of Random Forest modelling, while not new in avalanche research, provides further evidence that it could potentially be incorporated as another tool in avalanche forecasting operations. However, there are a few issues that need to be addressed prior to publication.

First, 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.

Second, 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.

Finally, there are several items detailed below in the Specific Comments section that should be addressed to help clarify certain issues. Most importantly, I think more detail on the avalanche dataset is necessary (see comment below).

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.

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.

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.

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).

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.

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.

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

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.

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.

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

line 313: Can you discuss how your results actually compare to other studies and provide some interpretation on this rather than presenting the table?

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

<u>Technical Corrections</u>
Some of the sentence structure/wording is difficult to follow at times. Consider a grammar/language revision.

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.

<u>References</u>

Dreier, L., Harvey, S., van Herwijnen, A., and Mitterer, C.: Relating meteorological parameters to glide-snow avalanche activity, Cold Regions Science and Technology, 128, 57-68, 10.1016/j.coldregions.2016.05.003, 2016.

Guy, Z. M., and Birkeland, K. W.: Relating complex terrain to potential avalanche trigger locations, Cold Regions Science and Technology, 86, 1-13, 10.1016/j.coldregions.2012.10.008, 2013.

Marienthal, A., Hendrikx, J., Birkeland, K., and Irvine, K. M.: Meteorological variables to aid forecasting deep slab avalanches on persistent weak layers, Cold Regions Science and Technology, 120, 227-236, 10.1016/j.coldregions.2015.08.007, 2015.

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