Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2019-379-AC1, 2020

Natural Hazards and Earth System Sciences
Discussions
Open Access

Interactive comment on "Data efficient Random Forest model for avalanche forecasting" *by* Manesh Chawla and Amreek Singh

Manesh Chawla and Amreek Singh

zmfzmj123@gmail.com

Received and published: 19 February 2020

We are grateful for the careful attention, the manuscript has been revised according to the comments suggested here:

Responses to specific comments:

Line 31: Revised. Line 35: Revised(gaume et al. added)

Use of words risk, hazard done according to definitions.

Line 80: Revised (Citations at the start of any sentence are formatted). Section 3: We will soon add this information in a revised manuscript. Line 232: Citation added. Line 231: ????

Printer-friendly version

Discussion paper

Figure 5âĂŘ6 and Table 5 appear to present the same information, either the figures or the table could be removed: True, but the numerical values from Table 5 have been used to find values in Table 6 to demonstrate operational performance.

Line 425: Example was selected in march month , temperature shows increasing trend in this month mostly. We will add a trend parameter in future work. Line 441: Sample size given for each probability estimate derived in Table 8.

Table 3: Formulas corrected. Figure Numbers have been corrected.

We have cited avalanche climate literature in discussion section, following citations were added: (Rewritten discussion section added in supplements [discussion.pdf])

Hägeli, P., McClung, D. M. : Avalanche characteristics of a transitional snow climateâĂŘColumbia Mountains, British Columbia, Canada. Cold Regions Science and Technology, 37, pp. 255–276, https://doi.org/10.1016/S0165-232X(03)00069-7 , 2003.

Haegeli, P., McClung, D. M.: Expanding the snowâĂŘclimate classification with avalancheâĂŘrelevant information: Initial description of avalanche winter regimes for southwestern Canada. Journal of Glaciology, 53, 266–276. https://doi.org/10.3189/172756507782202801, 2007

Mock C.J., Birkeland K.W. : Snow Avalanche Climatology of the Western United States Mountain Ranges, Bulletin of the American Meteoroligical Society , https://doi.org/10.1175/1520-0477(2000)081<2367:SACOTW>2.3.CO;2, 2000.

Shandro , B., Haegeli, P.: Characterizing the nature and variability of avalanche hazard in western Canada, Natural Hazards and Earth System Sciences, 18, pp. 1141–1158, https://doi.org/10.5194/nhess-18-1141-2018 , 2018.

Sharma, S.S., Ganju, A.: Complexities of avalanche forecasting in Western Himalaya âĂŤ an overview, Cold Regions Science and Technology, 31(2), pp. 95-102, https://doi.org/10.1016/S0165-232X(99)00034-8 , 2000.

Thumlert, S., Bellaire, S., Jamieson, B.: Relating avalanches to largeâĂŘscale ocean–atmospheric oscillations. In International Snow Science Workshop , Banff Canada, 29 September ,pp. 481–485 , 2014.

The following citations were added to literature survey in introduction: (Rewritten introduction section added in supplements [intro.pdf])

Choubin, B., Borji, M., Mosavi, A., Sajedi-Hosseini, F., Singh, V.P., Shamshirband, S.: Snow avalanche hazard prediction using machine learning methods. Journal of Hydrology, 577, p.123929, https://doi.org/10.1016/j.jhydrol.2019.123929 ,2019.

Davis, R.E., Elder, K., Howlett, D. and Bouzaglou, E.. 1999. Relating storm and weather factors to dry slab avalanche activity at Alta, Utah, and Mammoth Mountain, California, using classification and regression trees. Cold Reg. Sci. Technol., 30(1–3), 79–90

Gaume, J., Schweizer, J., Herwijnen, a Van, Chambon, G., Reuter, B., Eckert, N., Naaim, M.: Earth Surface Evaluation of slope stability with respect to snowpack, Journal of Geophysical Research. 119(9), pp. 1783 – 1799, https://doi.org/10.1002/2014JF003193, 2014.

Rahmati, O., Ghorbanzadeh, O., Teimurian, T., Mohammadi, F., Tiefenbacher, J.P., Falah, F., Pirasteh, S., Ngo, P.T., Bui, D.T.: Spatial Modeling of Snow Avalanche Using Machine Learning Models and Geo-Environmental Factors: Comparison of Effectiveness in Two Mountain Regions, Remote Sensing, https://doi.org/10.3390/rs11242995 , 2019.

Rubin, M.J., Camp, T., Herwijnen. A.V., Schweizer. J.: Automatically Detecting Avalanche Events in Passive Seismic Data, 11th International Conference on Machine Learning and Applications, Boca Raton , FL, USA, 12 – 15 Dec. 2012, https://doi.org/10.1109/ICMLA.2012.12, 2012.

Schirmer, M., Lehning M., Schweizer, J.,: Statistical forecasting of regional avalanche danger using simulated snow-cover data. Journal of Glaciology, 55(193), 761 – 768, https://doi.org/10.3189/002214309790152429, 2009.

Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., Kelly, J: A conceptual model of avalanche hazard. Natural Hazards, 90(2), 663–691, https://doi.org/10.1007/s11069-017-3070-5 , 2018.

Thuring, T., Schoch, M., Herwijnen, A.V., Schweizer, J.: Robust snow avalanche detection using supervised machine learning with infrasonic sensor arrays, Cold Regions Science and technology, 111, pp. 60-65, https://doi.org/10.1016/j.coldregions.2014.12.014, 2015.

Technical Corrections Line 120âĂŘ125: The reader would benefit from an explicit explanation of the random forest terms nonâĂŘleaf node, leaf node, and child node. Done [The supplement intro.pdf ends with this explanation]

Line 176: Corrected.

Line 179: Is this sentence describing the snow climate or the meteorological climate? If it is meant to describe the snow climate, please review Haegeli and McClung (2007) and Sharma & Ganju (2000) for snow climate classifications. The continental snow climate exhibits colder temperatures, more frequent periods of clear skies and less snowfall, which produces a thinner snowpack that is conducive to the formation of depth hoar and persistent weak layers (McClung and Schaerer, 2006).

The region has continental snow climate classification according to Mock and Birkeland scheme. Additional data has been visualised in figure 3b [supplement attached as climate.png] to give an intuition about the regions climatology. The figure gives monthwise median values for precipitation , temperature and sunshine duration of approximately 25 years (Nov-1993 to April-2017)

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-379, 2019.

Data efficient Random Forest model for avalanche forecasting

Manesh Chawla[1], Amreek Singh[2]

[1]Snow and Avalanche Study Establishment, Manali - 175103, India
[2]Snow and Avalanche Study Establishment, Chandigarh - 160037, India

Correspondence to: Manesh Chawla (zmfzmj123@gmail.com)

Abstract.

Fast downslope release of snow (avalanche) is a serious hazard to people living in snow bound mountains. Released snow mass can gain sufficient momentum on its down slope path to kill humans, uproot trees and rocks, destroy buildings. Direct reduction of avalanche threat is done by building control structures to add mechanical support to snowpack and reduce or deflect downward avalanche flow. On large terrains it is economically infeasible to use these methods on each hazard site. Therefore forecasting and avoiding avalanches is the only feasible method to reduce hazard, but sufficient snow stability data for accurate forecasting is generally unavailable and difficult to collect. Forecasters infer snow stability from their knowledge of local weather, terrain and sparsely available snowpack observations. This inference process is vulnerable to human bias therefore machine learning models are used to find patterns from past data and generate helpful outputs to minimise and quantify uncertainty in forecasting process. These machine learning techniques require long past records of avalanches which are difficult to obtain. In this paper we propose a data efficient Random Forest model to address this problem. The model can generate a descriptive forecast showing reasoning and patterns which are difficult to observe manually. Our model advances the field by being inexpensive and convenient for operational forecasting due to its data efficiency, amenable to automation and ability to describe its decisions.

1 Introduction

In snow bound areas avalanches cause loss of life and property worldwide. Avalanche deaths are estimated at 250 per year (Schweizer et al., 2015). Government and private agencies are funded to reduce avalanche risk for important activities and property e.g. road/rail transport, construction, border patrolling etc. This effort has led to development of several techniques to reduce avalanche risk. Avalanche hazard mapping is done to estimate the long term hazard at each slope in a region (Choubin et al., 2019; Rahmati et al., 2019). The map is used to plan active risk reduction methods e.g. building control structures, modification of nearby terrain or use of explosives to trigger avalanches in controlled way (Fuchs et al., 2007). Using active techniques at each hazard site is economically infeasible therefore avalanche forecasting is practised to reduce avalanche exposure. Individuals can use information in forecast to minimise short term risk in snow bound areas. Avalanche forecasting aims to identify the locations of snowpack weakness, their spatial distribution and sensitivity to triggering (Statham et. al. 2018). Observing snowpack stability at a high spatio-temporal resolution over large terrain is a difficult problem. Therefore stability at most risk sites is deduced using secondary observable data e.g. meteorological and snowpack parameters from a similar representative site, terrain parameters of the site, expected changes to snowpack by imminent weather etc. Snow stability shows high variance with respect to terrain features (Gaume et al., 2014). Deduction process for snow stability from secondary data has not been mathematically formulated therefore forecasters need to rely on their intuition of local terrain and snowpack patterns to estimate stability and collect more information to minimise uncertainty (LaChapelle, 1980; Schweizer et al., 2008; McClung and Schaerer, 2006). Numerical and statistical models are important tools for adding objectivity to this process.

1

Fig. 1. Introduction rewritten

Printer-friendly version

Discussion paper

6. Discussion

The model gives acceptable forecast accuracy of triggering probability by fresh snowfall or other natural causes. In 51 warnings it detected 25 out of an average of 29 avalanche days per winter [Table 6]. On average half of total warnings of natural triggering are true. This precision is reasonable given the difficulty of forecasting natural avalanches. The false alarms can indicate un-triggered snow instability. Descriptive forecast can provide more information about nature of these instabilities and their probable locations.

Consider the rule of figure 9 example:

-2.75 < MIN_TEMP <= -0.75 AND MAX_TEMP >= 1.75 then avalanche probability = high (> 90% natural triggering probability).

This rule seems to predict melt avalanches. Such a simple yet effective rule in terms of temperature only is difficult to find for a forecaster. We checked this hypothesis by additional data mining. Statistics from a filtered database containing only days which satisfy these bounds are compared to same statistics from the original unfiltered database [Table 8]. Other features correlated to the temperature bounds may be causing hazard. To rule this out we made a simple univariate analysis, where variables with significantly different distributions in filtered and original sets were analysed. Of these we believe only snow height is a variable leading to significant changes in hazard levels. To analyse effect of snow height, we applied another filtering to get data where snow height is greater than the mean snow height of temperature filtered data. Statistics from these three datasets are compared in Table 8.

The data mining results in Table 8 show that snowfall when the rule is satisfied leads to higher triggering probability. This is due to combination of factors: formation of melt freeze crusts and higher density of fresh snow at higher temperatures (Statham et al., 2014; Meløysund et al., 2007). The fresh snow bonds poorly with crust and due to its higher density it is also more likely to slip from crust. When rule is satisfied and no snowfall occurs, the triggering probability is higher than days when mean snow height is much higher. This suggests significant melting instability.

The model inferred the effect of a critical snowpack structure (melt freeze crust) from meteorological data. Capturing more complex properties of these structures e.g. persistence and strength require further feature engineering. Effect of persistent snowpack structures and climatic oscillations on avalanche activity has been analysed in detail by many researchers (Laternser and Schneebeli, 2003; Hägeli and McClung, 2003; Thumlert., et al. 2014). The resulting characterisations of avalanche climates can be used to derive relevant indexes to forecast (Haegeli and McClung, 2007; Shandro and Haegeli, 2018). Example above demonstrates that the model can be expected to account for these complex effects using simple and relevant extracted features.

An explanation of data efficiency is that decision trees model such reasoning and ensemble accounts for the different causes of avalanches. Variables involved for avalanche hazard are different for various situations therefore in avalanche datasets the important variables involved in causing hazard vary across the sample space. Nearest neighbour models are unable to adapt to this variation in feature importance as they use the same distance metric to forecast in every neighbourhood of sample space. Trees in ensemble accord importance to different features hence this method can account for the differences in important variables. The trees trained with splitting features matching the important features for input day give higher probability outputs than other trees.

Prediction is made using parameters which can be measured automatically too. Therefore such models can use data from dense sensor grid to improve performance. If additional parameters are required to improve forecasting process, only a few records of these new parameters are required for training an updated model. Therefore data efficiency of a model implies

14

Fig. 2. discussion section rewritten

Printer-friendly version

Discussion paper

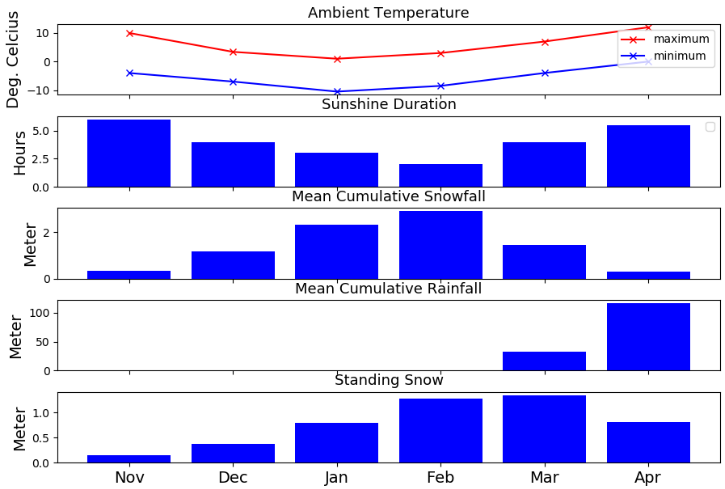

Fig. 3. climate figure