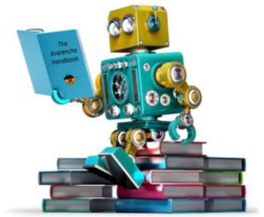
# How close are we to automated avalanche forecasting?
# Lessons from testing machine learning methods in Norway and Canada

Simon Horton[1], Karsten Müller[2], Pascal Haegeli[1], Rune Engeset[2]
[1]Simon Fraser University, Burnaby, Canada
[2]Norwegian Water Resources and Energy Directorate, Norway



## Introduction

We want to promote discussion in the avalanche community about the benefits and challenges of using machine learning and AI for avalanche forecasting, as these methods are become widely implemented in our daily lives. Guikema (2020) discuss considerations for AI applications in natural hazard risk analysis, providing a framework to discuss where we currently stand. We draw from our experiences applying machine learning methods to predict regional scale avalanche hazard in Norway and Canada.

## AI for natural hazards

Guikema (2020) identify the following settings where AI methods have been successful for natural hazard applications:

- Large clean data sets.
- Data are representative of future situations.
- Relationships between the data will be the same in the future.

But natural hazard applications often face the following challenges:

- Training data are not sufficiently large or representative.
- Unbalanced data sets that lack meaningful validation possibilities.
- Results and uncertainties that are difficult to communicate to decision makers.

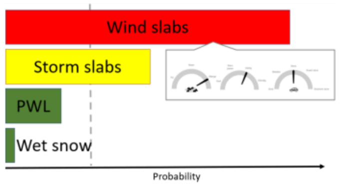
**Contacts**

Canada: simon_horton@sfu.ca
Norway: kmu@nve.no

## Approaches used in Norway

**Our motivation**

- Support avalanche forecasters and automate tedious tasks. AI can assist in data aggregation and filtering (Müller et al. 2018)
- Provide avalanche forecasters with draft forecast(s), e.g. probability of different avalanche problems
- Provide semi-automated forecasts at higher spatial and temporal resolution



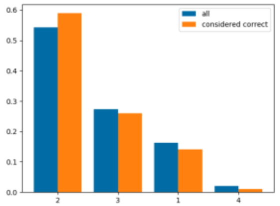

*Possible outcome: Probabilities for various avalanche problems for a given day and region.*

**Expected outcome**

AI is often good at generalizing but shows poor performance in predicting outliers/extremes (i.e. high or very high danger). Thus, we expect AI to be reliable in predicting common scenarios, while the human-forecaster needs to focus on potential extremes and rare cases, which an AI will ideally flag as anomalies/outliers.

**Our challenges**

Data preparation is a huge and important task, since the results will heavily depend on the quality of training data. Data imbalance is a problem when training AI-models (see figure to the right). We try to remove poor or plainly wrong data (e.g. precipitation readings, input errors from observations, forecasts that in retrospect proved to be wrong) to avoid learning wrong correlations (e.g. our observers provide feedback on the correctness of a forecast: 5% were judged to be too high or low, 33% were judged as correct and 62% did not get feedback).




*Distribution of danger levels in our training dataset. While 54% of the forecasts were assigned DL 2 only 2% were assigned DL 4. Skew is even higher if only samples considered correct are used.*

**Our approach and preliminary results**

We are training AI on our historic forecasts and gridded snowpack and weather models. The focus initially was on shallow learning algorithms, but tests with neural networks have shown promising results.

- We used regional statistics over gridded snow/weather models to automate weather forecasts, reducing the work time for the meteorologist (and improving their consistency)
- We trained a clustering algorithm (k-NN) to match current/expected conditions to historic forecasts. Thus, providing a template for the forecast.
- Classic ML (CART, RF, SVM) trained to predict problems based on historic forecasts results in high F1-scores (> 0.9) when trained to predict (e.g. moderate danger level or typical problems such as wind slabs). However, it proved random (useless) for rare events (i.e. high DL). Here a simpler model based on weather prediction alone might be superior.
- Planned: Since the order of events is important we intend to test recurrent neural networks (RNN)
- Planned: We envision an ensemble approach that combines different AI models. If they agree in the outcome a higher reliability can be expected than when they all suggest different outcomes.

## Approaches used in Canada

**Our motivation**

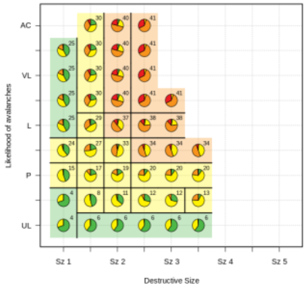
We apply machine learning methods to better understand the hazard assessment process and identify influencing factors in forecasting decisions.

**Methods**

We explore relationship in large operational data sets of hazard assessments from Canada including:

- Relationships between danger ratings and avalanche problems (e.g. size, likelihood, type)
- Relationships between problem types and weather and snowpack conditions.

We use "transparent" model approaches like decision trees to understand how forecasters factor relevant variables into their decisions.



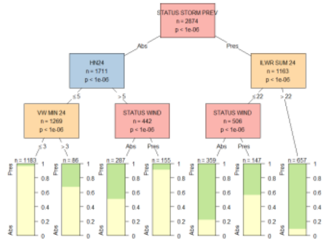

*Probability of danger rating given the size and likelihood of an avalanche problem.*

**Results**

The decision trees reveal both intuitive results (e.g. danger increases with size and likelihood, storm slab problems exist when there's new snow) but also unusual factors (e.g. different decision rules for different locations, seasons, individuals/agencies, and relationships between problems with unexpected snowpack properties). These inconsistencies in operational data sets suggest caution should be taken when using this data for training predictive models.

**Next steps**

We are presenting our results to forecasters to facilitate learning about their processes. This could inform expert discussions about what factors *should* be applied in forecasting decisions and move towards improved forecasting procedures. It could also provide insights into how we can better prepare representative training data sets for future machine learning approaches.




*Decision tree showing whether a storm slab avalanche problem is present based on whether one was present the previous day, recent weather conditions (snowfall, wind, and incoming longwave radiation), and whether there is a concurrent wind slab problem.*

## Challenges and recommendations

**Representative data sets**

- Our data sets are relatively small and incomplete. Forecasters already have challenges interpreting this data, and many assumptions are needed when preparing data for statistical analysis. We need to create meaningful datasets that include relevant variables in a way that is suitable for analysis.
- Weather observations (class III) are generally more consistent than snowpack and avalanche observations (class II and I). Human assessed categories such as problems and hazard, as well as model-derived data, are generally more consistent but have additional uncertainties.
- The primary dependent variables when training models is human assessments of danger and problems, which are subjective and prone to biases and inconsistencies. AI is better suited for clearly defined responses (e.g. avalanche activity in a specific path). New avalanche data sets from remote detection and satellite imagery provide new opportunities for consistent dependent variables.
- Can we move towards international standardized data sets for statistical analysis? Standardized weather observations were critical for improved weather prediction.

**Meaningful validation**

- Studies on the accuracy of human avalanche forecasters is generally in the 70-80% range, meaning the bar for models to provide assistance is lower than in other fields.
- Like other natural hazards, our data is unbalanced, meaning high-impact events are generally rare (e.g. danger levels 4 and 5 are much less frequent and levels 1-3). This makes meaningful verification difficult.
- Can models actually help predict anomalies and low-frequency high-consequence situations that forecasters have difficulty predicting, or are they better suited to identify common situations and provide a baseline forecast?

**Communication to decision makers**

- How should we implement and test predictive models in operational settings?
- Some methods are more transparent (e.g. decision trees) than "black-box" methods like neural networks. What's the trade-off between predictive skill and transparency?
- How do we provide quality control and avoid feedback of model errors and biases when training models with operational data on an ongoing basis?
- Collaborative and iterative development between forecasters and researchers is critical moving forward.

## References

Clark, T.: Exploring the link between the Conceptual Model of Avalanche Hazard and the North American Public Avalanche Danger Scale, MSc Thesis, Simon Fraser University, 2019.

Eckerstorfer, M., Malnes, E., Müller, K.: A complete snow avalanche activity record from a Norwegian forecasting region using Sentinel-1 satellite-radar data, Cold Regions Science and Technology, 2017.

Guikema, S.: Artificial Intelligence for Natural Hazards Risk Analysis: Potential, Challenges, and Research Needs, Risk Analysis, 2020.

Horton, S., Towell, M., and Haegeli, P.: Examining the operational use of avalanche problems with decision trees and model-generated weather and snowpack variables, Nat. Hazards Earth Syst. Sci. Discuss., in review, 2020.

Müller, K., Engeset, R.V., Landrø, M., Humstad, T., Granan, E.B., Thorset, H.: Avalanche Problem Solver (APS) – a decision support system for forecasters (part 1), Proceedings, International Snow Science Workshop, 2018.