MACHINE LEARNING AND PATTERN RECOGNITION FOR PUBLIC AVALANCHE FORECASTING

Andrew Schauer[1]*, John Sykes[1,2]

[1] *Chugach National Forest Avalanche Center, Girdwood, AK, USA*
[2] *Simon Fraser University Avalanche Research Program, Burnaby, BC, Canada*

ABSTRACT: Avalanche forecasters in the U.S. and Canada use the five-level North American Public Avalanche Danger Scale (NAPADS) as a tool to communicate avalanche risk to backcountry recreationists. Forecasters consider the avalanche problem type, distribution, sensitivity, size, and likelihood to assign an appropriate danger rating for a given forecast period. While there is ongoing progress in standardizing this decision-making process, there remains a large degree of subjective judgment based on professional experience. This project develops two novel decision support tools for assigning daily danger ratings. The first incorporates weather station data, previous avalanche forecasts, and reassessed avalanche forecast data to identify established patterns in public avalanche forecasts and predict daily danger ratings. We assess the performance of Extreme Gradient Boosting (XGB) and Super-Organizing Maps (SOM) and compare the predicted danger ratings using the machine learning models to those issued by human forecasters. We find 73-79% agreement between the models and professional forecasters, with the XGB model outperforming the SOM. The second tool uses a national record of avalanche forecasts to identify relationships between avalanche size, likelihood, and assigned danger rating, fully implementing the workflow described in the Conceptual Model of Avalanche Hazard (CMAH), and clearly defining the relationship between avalanche size/distribution/sensitivity and predicted danger rating. The final products of this research are (1) a machine learning tool that uses weather and avalanche variables to predict the avalanche danger rating for an avalanche center in Southcentral Alaska, and (2) an online interactive tool to visually guide the user through the CMAH workflow, linking avalanche size and likelihood to the predicted danger rating using data from 24 avalanche centers across the U.S. Both of these forecasting tools incorporate visual representations of uncertainty in the predicted danger rating based on historical records. These tools aim to assist backcountry avalanche forecasters in predicting avalanche danger ratings and quantifying uncertainty in the assigned rating. We believe this work will leverage modern machine learning methods and archived forecast data to capture the institutional knowledge of avalanche centers, which can support the decisions of newer forecasters and help seasoned practitioners identify personal biases in their forecasts.

KEYWORDS: Machine Learning, Decision Support, CMAH, NAPADS

1. INTRODUCTION

Public backcountry avalanche forecasting requires an assessment of the size and likelihood of a specific type of avalanche problem(s), and assigning an appropriate danger rating for the forecast period. Statham et al. (2018) introduced the Conceptual Model for Avalanche Hazard (CMAH), which defined a clear framework to identify the avalanche problem type, and then assess the sensitivity, distribution, and size of expected avalanche activity. This established a consistent workflow and terminology that many avalanche practitioners use today. This terminology has been incorporated into the North American Public Avalanche Danger Scale, but there remains a large amount of subjectivity in connecting avalanche size and likelihood to an appropriate danger rating in North America. This is in contrast to the European Avalanche Warning Service, which has established the EAWS matrix as a decision support tool to connect avalanche size, likelihood, and danger rating for regional avalanche forecasts (Müller et al., 2025). This matrix style of generating daily danger ratings is more prescriptive, but also has the potential to be more objective across different forecasters and forecast regions.

While the CMAH framework may be relatively new, the practice of monitoring weather and snowpack properties to predict size and likelihood of avalanches has been in place for decades (e.g. Atwater, 1954; LaChapelle, 1966). Until fairly recently, this process has largely involved incorporating manual field observations and numeric weather models to make decisions based almost entirely on professional experience and expert judgement. However, modern tools are quickly becoming available to support these decisions with model output. The majority of the emerging tools use a combination of process-based snowpack model output, numerical

* Corresponding author address:
Andrew Schauer
email: andrew@chugachavalanche.org

https://doi.org/10.15788/1790098992

weather output, weather station observations, and machine learning algorithms to detect unstable layers within the snowpack (e.g. Herla et al., 2024; Mayer et al., 2022), estimate the size and/or likelihood of avalanches (Mayer et al., 2023, van Herwijnen et al., 2023) and predict avalanche danger (Perez-Guillen et al., 2025).

Models are approaching accuracy rates similar to human forecasters (Techel et al., 2025), with roughly 66-75% agreement between models and human-forecasted danger rating (Perez-Guillen et al., 2025), 75% probability of detecting weak layers within a snowpack (Herla et al., 2024), and 69% recall/75% specificity for detecting avalanche and non-avalanche days (Mayer et al., 2022) compared to testing data. Davis et al. (2026) investigated reassessed human-issued avalanche forecasts for U.S. avalanche centers and found the maximum predicted danger rating matched the maximum reassessed danger rating for 84% of the days on record across all avalanche centers, with rates varying between 68% and 97% for individual avalanche centers. Techel et al. (2024) compared assessments between two different forecasters on the same day, and found agreement rates of 87% in Switzerland, and 76% in Norway.

With model accuracy approaching that of human forecasters, avalanche centers are increasingly incorporating model output in daily workflows. However, the ability and success of implementing these models depends on (1) accuracy of numerical weather output in the forecast area; (2) the capacity to run a model chain for an area of interest; and (3) the ability to display the model output in a way that is both informative and easy to understand.

In an effort to address the gap in connecting size, likelihood and danger rating in the CMAH workflow and provide tools for avalanche centers facing the model limitations described above, we introduce two tools that are computationally inexpensive and easy to incorporate in an avalanche forecasting workflow:

1. A web-based user-interface that navigates the CMAH workflow, and overlays the defined avalanche size/likelihood over a danger rating matrix where each cell displays a distribution of previously assigned danger ratings for the specified combination of size and likelihood. This is based on a national dataset incorporating forecasts from across the U.S.

2. A machine learning based decision-support tool that predicts the daily danger rating based on user-specified weather and avalanche parameters. Whereas the CMAH dashboard is not location-dependent, this tool is developed using weather and avalanche data from the CNFAC Turnagain Pass zone and is only applicable for that location.

Both of these tools incorporate uncertainty into the output display, showing a distribution of predicted danger ratings for the user-defined conditions. These can help improve consistency, reduce bias, and add historical context for challenging forecast scenarios.

2. METHODS

2.1 CMAH Dashboard

The CMAH Dashboard tool (cmah-dashboard.onrender.com) features three sliders where the user defines the sensitivity, distribution, and expected size for an avalanche problem (Fig. 1a). These three inputs feed two output displays. The first (Fig. 1b) is a sensitivity/distribution display as defined in Statham et al. (2018). The second is a color-coded matrix of NAPADS danger ratings corresponding with combinations of avalanche size and likelihood (Fig. 1c).

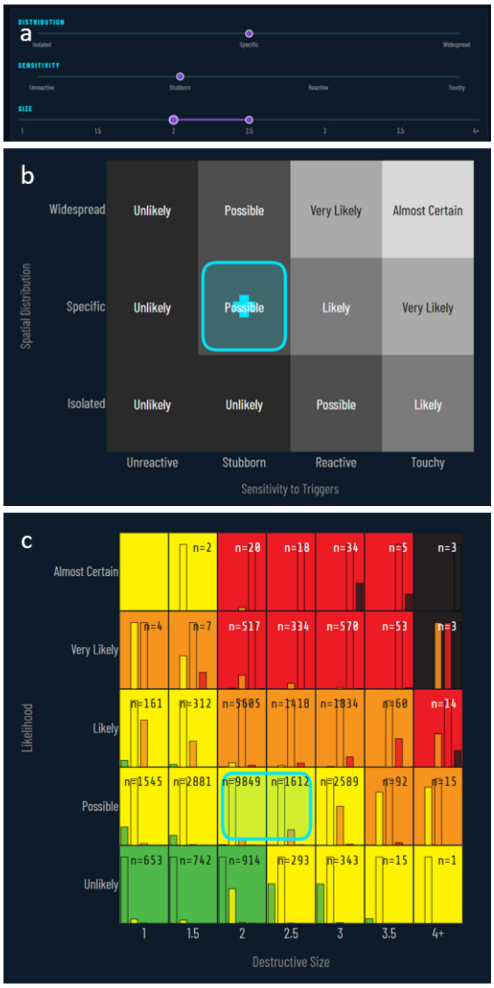

Figure 1: Output displayed on the CMAH dashboard. The user adjusts the sliders to specify sen-

sitivity, distribution, and size for the primary avalanche problem (1a). This defines the area of a box that is overlain on the sensitivity/distribution/likelihood matrix (1b). The lower plot (1c) shows the size/likelihood/danger rating matrix. Danger rating histograms are generated from the NAC dataset.

The histograms displayed on the size/likelihood danger matrix are built using data from the U.S. Forest Service National Avalanche Center (NAC). This dataset includes 32,518 forecasts over 14 seasons from 24 avalanche centers across the US. The number of forecasts in the record varies among avalanche centers, with the Bridger-Teton Avalanche Center (BTAC), Colorado Avalanche Information Center (CAIC), and Northwest Avalanche Center (NWAC) accounting for the majority of the forecasts on record (Fig. 2). Most daily forecasts include more than one avalanche problem. When there are multiple avalanche problems on a given day, we record the assigned danger rating for the size and likelihood defined in Problem 1. The cells in the matrix are color-coded based on the danger rating with the highest count for that cell.

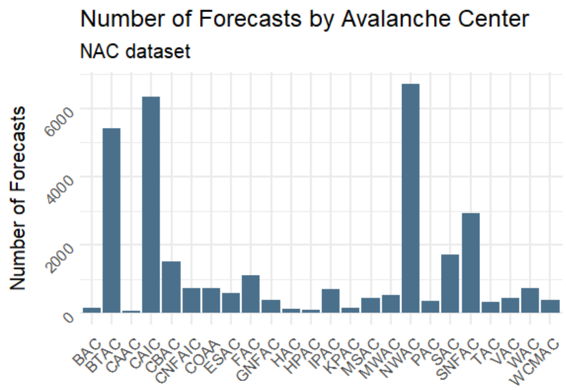

Figure 2: Number of forecasts recorded by each avalanche center in the NAC dataset. BTAC, CAIC, and NWAC account for the majority of all forecasts, but the dataset includes data from 24 different avalanche centers.

2.2 Danger Rating Prediction Tool

For the danger rating prediction decision-support tool, we compare performance of two different types of machine learning models – an Extreme Gradient Boosting (XGB) model (Chen and Guestrin, 2016) and an ensemble of Super-Organizing Maps (Wehrens and Buydens, 2007). These models were trained using avalanche forecast and weather station data from the Chugach National Forest Avalanche Center (CNFAC) Turnagain Pass forecast zone from December 2009 through April 2025.

The CNFAC has been reassessing forecasts since Nov. 2021. This involves a daily assessment in which the forecaster on duty assigns an observed danger rating for the previous forecast period. For the 2021/22-2024/25 seasons, we use the reassessed danger ratings when there is a discrepancy between the predicted and reassessed values.

In addition to the avalanche forecast data, we also use weather station data taken from three weather stations on Turnagain Pass: the Sunburst weather station is situated above treeline at 1,158 m (3,800 ft), Seattle Ridge at 731 m (2,400 ft), and Turnagain Pass SNOTEL at 573 m (1,880 ft).

Avalanche danger rating data is treated as an ordered categorical variable. We treated the avalanche problem type as a binary variable using one-hot encoding (Cerda et al., 2018), effectively including 11 binary variables where each variable represents one avalanche problem type (including one variable for 'none' and one for 'springtime diurnal' per the standard practices of CNFAC). For a given day, each avalanche problem listed in the forecast discussion has its own set of 11 binary avalanche problem type variables. The 24 weather station variables are treated as continuous numeric variables.

The models are trained on 80% of the data set, with 20% of the data set aside as a testing dataset. Since the avalanche danger dataset is highly autocorrelated, we use a block cross-validation sampling strategy (Roberts et al., 2017), randomly selecting entire seasons to be withheld as test data, rather than individual days to designate the test and training groups.

Extreme Gradient Boosting

We trained a separate Extreme Gradient Boosting classifier (Chen and Guestrin, 2016) for each elevation band. The target was the issued NAPADS rating, replaced by the next-day reassessment when one existed. Alpine and treeline models used the five danger classes (Low–Extreme). Below-treeline days recorded as No Rating (typically insufficient low-elevation snow) were excluded from BTL training and evaluation, because that label is determined by snow cover rather than the weather and problem features available to the model. Extreme danger occurred only once in the record (18 February 2010) and is absent from the test seasons.

Predictors combined (i) same-day avalanche-problem type, likelihood, and expected size for up to three problems, (ii) the previous day's danger rating and problem characterization, (iii) 24 daily weather variables from Sunburst, Seattle Ridge, and the Turnagain Pass SNOTEL, and (iv) calendar month and two era indicators marking when problem type (17 November 2012) and likelihood (1 January 2018) entered the archive. Problem type was one-hot encoded (Cerda et al., 2018). This feature set matches the intended operational use, in which the forecaster first specifies today's avalanche problems, then the model returns a danger-class probability distribution. It is not a weather-only nowcast.

Hyperparameters were selected by randomized search with chronological cross-validation inside the training window. The final models used a softmax multi-class objective so that each prediction is a probability vector over danger classes (Fig. 4). Early stopping used the 2022/23 season, which is inside the training window, and never the held-out test seasons.

Super-Organizing Map Ensemble

Super-Organizing Maps (SOM) are a special application of Self-Organizing Maps (Kohonen, 1998) that allow categorical, binary, and continuous variables to be applied as separate layers to test and train the model. The algorithm organizes observations with multiple dimensions by analyzing the dataset and creating a grid of bins, or 'nodes' that covers the extent of the variability within the dataset, then assigning each observation to the node that best represents that observation. In this case, that means that days with similar weather and avalanche forecast variables should be grouped in similar nodes. A dataset may be represented by any number of nodes. In this case we train 16 different SOM models with square arrays (i.e. 5x5 to 20x20), and hexagonal geometry resulting in a suite of models with 25 to 400 nodes.

To use the SOM ensemble for prediction for a given day, we feed the models all of the variables except the danger ratings for each elevation. Each model assigns that day to the node which most nearly represents the given conditions, and then predicts the danger rating by taking a count of the danger rating assigned to each day within that node. For example, if a day matches a node in which 25 days are assigned 'Moderate' and 35 are assigned 'Considerable', the model would predict the danger rating for that day to be 'Considerable'. We then tally up the 'winning' danger rating that each of the 16 SOM models assign, and use the winner as the predicted value for that day.

Model Performance

We measure model performance with a weighted F1 score for each elevation band (Sasaki, 2007). We also include F1 scores for a persistence forecast baseline (yesterday's danger rating) so that model predictive skill can be interpreted relative to the strong day-to-day autocorrelation in danger ratings.

2.3 Forecast Reassessment

We use the reassessment dataset to estimate the baseline accuracy for a human forecaster as another means of assessing model performance. We calculate simple summary statistics for the five seasons of data, reporting the accuracy rates at each elevation band, and also the percent of days where the reassessed danger rating matched the predicted danger ratings for all elevation bands.

3. RESULTS

3.1 Model Performance

The XGB model had higher performance statistics in at all three elevation bands (Fig. 3). Both models had the highest scores in the alpine elevation band, and both performed better than the persistence baseline.

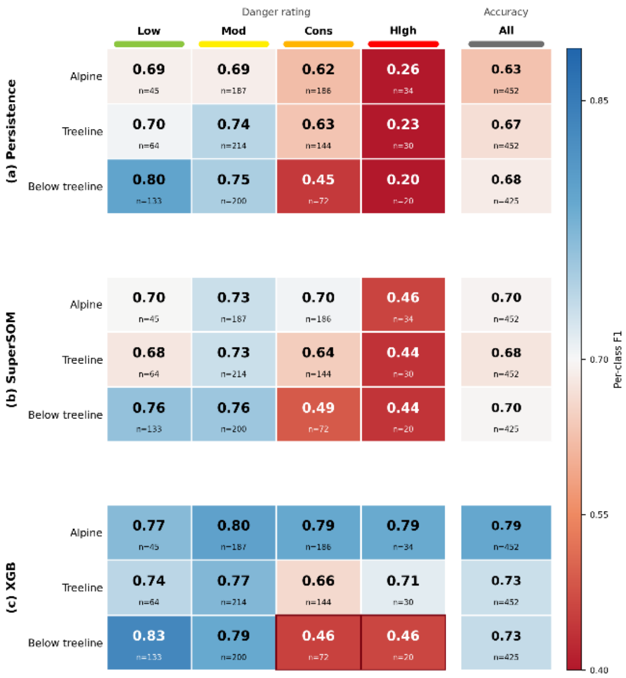

Figure 3: F1 scores by elevation band and danger rating for the persistence forecast (a), SuperSOM ensemble (b), and XGB model (c).

3.2 Reassessment data

For the 341 human forecasts with reassessment data, we found that the human-predicted danger ratings matched the human-reassessed danger rating for 93% of the days in the alpine, 91% at treeline, and 92% below treeline. The reassessed danger ratings agreed with the predicted danger ratings at all elevation bands for 83% of the days.

4. DISCUSSION

The XGB model performed similarly to daily CNFAC forecast/reassessment agreement rates (73% to 79% weighted accuracy for XGB vs. 83% daily agreement for human forecasters). This accuracy rate is consistent with the previously established rates in the U.S. (Davis et al., 2026; Logan and Greene, 2023) and in Europe (Techel et al., 2024).

Since we trained the model on a dataset that did not have reassessed values available for the majority of days, model performance is limited by the accuracy of the data. Since it is not possible to train a model that is accurate 100% of the time, we add to the utility of these tools by building in a way of displaying uncertainty. Rather than simply displaying the suggested danger rating for the day, we provide a histogram that shows the likelihood distribution for the danger rating given the current conditions (Fig. 4).

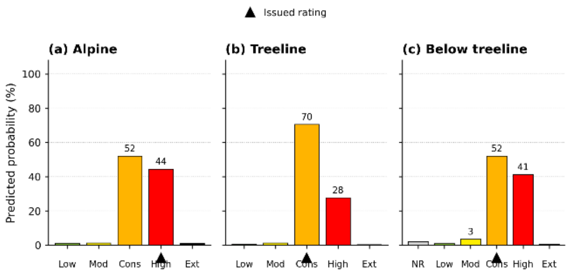

Figure 4: Model-predicted danger ratings for Nov. 30, 2023. The colored bars represent the predicted likelihood for each danger rating at each elevation band. For example, the model predicts a 52% likelihood of Considerable danger in the Alpine, 70% likelihood at Treeline, and 52% likelihood Below Treeline. The actual human-issued danger ratings for that day are marked by the black triangle at the bottom of each chart. For this day, the human-issued danger rating in the alpine did not correspond with the maximum likelihood danger rating from the model output.

We similarly display the uncertainty in the CMAH tool by showing the danger rating distribution for days with the specified size and likelihood (Fig. 1c). This approach visually represents the uncertainty inherent in assigning a danger rating, and displays it in a way that is useful to avalanche forecasters.

5. CONCLUSION, LIMITATIONS, AND FUTURE WORK

We used existing avalanche forecast records to develop two decision-support tools to help improve consistency and reduce subjectivity in assigning a danger rating.

The CMAH tool helps forecasters work through the CMAH workflow interactively, and visually displays the connection between size, likelihood, and danger rating. The tool was developed using data from 24 avalanche centers across the U.S., so the application should be useful to a wide geographic range.

The danger rating prediction tool draws on previous weather and avalanche forecast records for a CNFAC forecast zone to recognize patterns in the historical data, and generate a probabilistic danger rating based on similar days within the record. It is important to recognize the geographic limitations with the danger rating prediction tool we introduce here. The model was trained using weather and avalanche forecast data from Turnagain Pass, and has not been validated for other forecast zones. While this is a logical next step for model development, it would not be appropriate to apply this model to another forecast zone without first assessing its performance in a new location.

Introducing predictive models into the forecasting workflow can bias decision-making if the model output replaces, rather than checks, an independent assessment. The danger-rating prediction tool was therefore built so that the forecaster first specifies today’s avalanche problem type, size, and likelihood; the model then returns a probability distribution over danger classes given those inputs and the associated weather. We recommend that forecasters complete their own CMAH-based rating before consulting the tool, and use the output only as a consistency check against historical practice. Used this way, the model can surface patterns in the archive and flag days that diverge from how similar conditions have been rated, without becoming the source of the rating itself.

While these tools apply the CMAH elements as well as measurable weather metrics, it is important to note that neither tool incorporates Travel Advice into the danger rating assessment. This key component of NAPADS is often the determining factor when forecasters are faced with a scenario in which size, distribution, and likelihood may correspond with different danger ratings. While these tools are able to display some of the uncertainty inherent in assigning a danger rating, there still remains a large degree of subjectivity that relies on expert judgement to assign an appropriate danger rating.

Arguably the most useful feature of both of these tools is that ability to display the uncertainty in assigning a danger rating. These visual displays can provide context to help inform avalanche forecast discussions, thereby improving the way we can communicate risk to the public.

ACKNOWLEDGEMENTS

The authors used the AI tools Claude and Grok to assist in generating code used to design the CMAH dashboard, building and training machine learning models, and generating plots. All AI output was reviewed for accuracy and edited manually as needed.

REFERENCES

Atwater, M. Snow Avalanches. Scientific American, 190, 1, 26-31. 1954.

Cerda, P., Varoquaux, G., & Kégl, B. "Similarity encoding for learning with dirty categorical variables." Machine Learning, 107, 1477–1494. 2018.

Chen, T., & Guestrin, C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. DOI: 10.1145/2939672.2939785. 2016.

Davis, E., Miller, Z., Savage, S., Glass, D., Logan, S., Greene, E., Reardon, B., Donnelly, C., Schauer, A. Assessing Public Backcountry Avalanche Forecasts. Proceedings of the International Snow Science Workshop, Whistler, BC, CAN. 2026.

Herla, F., Haegelin, P., Horton, S., Mair, P. A large-scale validation of snowpack simulations in support of avalanche forecasting focusing on critical layers, Natural Hazards and Earth System Sciences, 24, 2727-2756, https://doi.org/10.5194/nhess-24-2727-2024. 2024.

Kohonen, T. The Self-organizing map. Neurocomputing, 12, 1-6. 1998.

LaChapelle E. Avalanche forecasting- A modern synthesis. IAHS Publication 69: 350-356. 1966.

Logan, S., & Greene, E. Assessing the Colorado Avalanche Information Center's Backcountry Avalanche Forecasts. Proceedings of the International Snow Science Workshop, Bend, Oregon, pp. 518–525. 2023.

Mayer, S., Techel, F., Schweizer, J., van Herwijnen, A. Prediction of natural dry-snow avalanche activity using physics-based snowpack simulations, Natural Hazards and Earth System Sciences, 23, 3445-3465, https://doi.org/10.5194/nhess-23-3445-2023, 2023.

Mayer, S., van Herwijnen, A., Techel, F., Schweizer, J. A random forest model to assess snow instability from simulated snow stratigraphy, The Cryosphere, 16, 4593-4615, https://doi.org/10.5194/tc-16-4593-2022. 2022.

Müller, K., Techel, F., & Mitterer, C. The EAWS matrix, a decision support tool to determine the regional avalanche danger level (Part A): conceptual development. Natural Hazards and Earth System Sciences, 25, 4503–4525.DOI: 10.5194/nhess-25-4503-2025. 2025.

Pérez-Guillén, C., Techel, F., Volpi, M., van Herwijnen, A. Assessing the performance and explainability of an avalanche danger forecast model, Natural Hazards and Earth Systems Science, 25, 1331-1351, https://doi.org/10.5194/nhess-25-1331-2025. 2025.

Roberts, D. R., Bahn, V., Ciuti, S., et al. "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure." Ecography, 40, 913–929. 2017.

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, 663-691. 2018.

Sasaki, Y. The truth of the f-measure. Proceedings of the workshop on machine learning and data mining in pattern recognition (MLDM) Leipzig, Germany. 2007.

Techel, F., Purves, R.S., Mayer, S., Schmudlach, G., Winkler, K. Can model-based avalanche forecasts match the discriminatory skill of human danger-level forecasts? A comparison in Switzerland, Natural Hazards and Earth System Sciences, 25, 3333-3353, https://doi.org/10.5194/nhess-25-3333-2025, 2025.

Techel, F., Lucas, C., Müller, K., Pielmeier, C., & Morreau, M. Unreliability in expert estimates of factors determining avalanche danger and impact on danger level estimation using the Matrix. Proceedings of the International Snow Science Workshop, Tromsø, Norway, pp. 264–271. 2024.

Van Herwijnen, A., Mayer, S., Pérez-Guillén, C., Techel, F., Hendrick, M., Schweizer, J. Data-driven models used in operational forecasting in Switzerland, Proceedings of the 2023 International Snow Science Workshop, Bend, OR, U.S.A., pp. 321-326. 2023.

Wehrens, R., and Buydens, L.M.C. Self- and Superorganizing maps in R: the kohonen package. Journal of Statistical software, 21 (5). 2007.