Emma Sørset

Site-specific probabilistic avalanche forecasting

Specialization Project, Industrial Mathematics, TMA4500
Supervisor: Ingelin Steinsland
December 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematics

NTNU

Acknowledgements

I would like to sincerely thank my supervisor, Ingelin Steinsland, for her valuable guidance and patience in addressing my questions and challenges throughout this process. I also thank the Norwegian Public Roads Administration, particularly Eivind Schnell Juvik, for giving me the opportunity to work on this thesis, offering expertise on avalanche-related topics, facilitating access to essential data, and showing great enthusiasm for the project's potential.

Abstract

Avalanche hazard is traditionally described as a combination of the probability of an avalanche release and the expected avalanche size (Thumlert, 2024). Forecasting the probability of an avalanche release is thus a critical component of mitigating the physical risks of avalanches. This project compare probabilistic forecasters for avalanches and examine how they perform on different danger levels, and if they perform differently when the danger levels are included in the model, for a section of county road 7900 in Holmbuktura, Lyngen, Norway. The suggested models are fit on weather variables such as daily precipitation, temperature, wind, and snow depth.

We have trained and evaluated several statistical models using data collected from seven winter seasons (2017–2024). The proposed models include cross-validated climatology, generalized linear models (GLMs), decision trees, random forests, and k-nearest neighbors, all for a binary case. The models are evaluated using metrics such as the Brier Score, Brier Skill Score, and reliability- and sharpness diagram. Additionally, their performance was compared to regional avalanche danger forecasts provided by `Varsom.no`.

Our results indicate that random forest models outperform other methods, demonstrating high reliability, sharpness, and robustness across seasons. The evaluations are based on leave-one-season-out cross-validation. Compared to regional danger level forecasts, the statistical models offer improved site-specific predictions, particularly for days with high avalanche risk (danger level 4). They all achieved a Brier Skill Score between 0.36 and 0.57 relative to the regular climatology, and for all methods, the score was better when the models where fitted to the danger level. The value of creating separate models for each danger level is greatest for danger level 4, i.e., when the probability of avalanches is highest. The findings suggest that these models can lay a foundation for a framework to enhance operational decision-making by providing more localized and precise avalanche risk assessments.

Contents

1 Introduction **6**

2 Background **10**

2.1 Avalanches 10

2.1.1 Avalanche warning 12

2.1.2 Avalanche mitigation 14

2.2 Cross validation 14

2.3 Probabilistic forecasting 15

2.4 Evaluating forecasts 15

2.4.1 Brier score 16

2.4.2 Brier skill score 17

2.4.3 Evaluation of Statistical Properties 17

2.5 Forecasting methods and statistical models 18

2.5.1 Climatology 18

2.5.2 Generalized Linear Methods 19

2.5.3 Tree regression 21

2.5.4 Random Forest method 23

2.5.5 Method of Nearest Neigbors 23

3 Study area, data and exploratory data analysis **25**

3.1 Location 25
3.2 Data 25
3.2.1 Radar Detections 25
3.2.2 Danger Level Warning data 28
3.2.3 Weather Station Data 28
3.2.4 Interpolated weather observations 29

4 Modeling **32**

4.1 Explanatory variables 32
4.2 Model fitting 33
4.3 Model evaluation 33
4.4 The Models 34
4.4.1 Climatology 34
4.4.2 GLM 34
4.4.3 Simple Tree 35
4.4.4 Random forest 35
4.4.5 Method of nearest neighbors 35

5 Results **37**

5.1 GLM-based model 37
5.2 Simple Tree 39
5.3 Random Forest 43

5.4 KNN 45
5.5 Climatology 48
5.6 Comparison of the models 49
5.7 Case study of two months 51
5.7.1 February 2019 51
5.7.2 April 2021 52

6 Discussion **54**

7 Conclusion **56**

A Appendix **i**

1 Introduction

A snow avalanche, hereafter referred to as an avalanche, is a complex natural phenomenon occurring in steep mountain slopes covered with snow, posing significant risks to human life, infrastructure, and the environment. Understanding avalanches is crucial for multiple reasons. It is essential not only for urban planning and infrastructure development but also for individuals who engage in mountain sports during winter(Mitterer et al., 2016). In this project, we focus on one-day ahead forecasts of the probability of an avalanche release to improve operational decisions in avalanche mitigation.

Due to the complexity of the phenomenon, avalanches can be challenging to predict, as the snow and weather change continuously and vary greatly (Lome et al., 2024). If the amount of snow is small and the snowpack is stable, the probability for an avalanche to release is low. On other days, the situation is the complete opposite. If there is a lot of snow and an unstable snowpack, there is a great risk of avalanches, and it is recommended not to be in areas where an avalanche might release (Nes, 2018). However, the days between these two situations are the most common. These are also the days when it is most difficult to predict whether or not an avalanche will happen (Eivind Schnell Juvik, personal communication).

In this project, we consider a section of county road 7900 (FV7900) that goes from Seljelvnes in Balsfjord to Jøvik in Tromsø (Wikipedia, a). The road goes along the shoreline on the western side of the Lyngen Alps. It is the only road connection to the townships Jøvika and Olderbakken, with a total of 80 inhabitants (SSB). The section in question is shown in Figure 1 and is the 2006m long section of FV7900 along Holmebuktura. Climatic conditions, which vary throughout the season, play a significant role in avalanche activity, influencing both the snowpack and other risk factors. These variations are particularly pronounced in areas like Lyngen, where seasonal changes are significant due to the polar night and a coastal arctic climate.

The combination of steep mountainsides, like in Figure 3, and snow poses a significant risk of avalanches. The Lyngen Alps is a highly recognized playground for alpine-sport enthusiasts (Friflyt), who voluntarily expose themselves to the risk of avalanches but with the possibility to choose if the conditions are within their acceptable risk limits. The people living in the Lyngen Alps do not have the same possibility. They must use the roads and be in avalanche terrain to get to school, work, and other daily life activities. As a part of the studies at the Norwegian University of Science and Technology (Norges Teknologiske og Naturvitenskapelige Universitet, NTNU), one has to reflect on the sustainability relevance of the topic concerning the United Nations Sustainable Development Goals. Goal nr 9 is "Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation" (United Nations). This project aims to contribute to a resilient

infrastructure.

From an avalanche safety point of view, it would be a good decision to close an exposed road on all days when there is a chance of an avalanche. However, this is not sustainable for people living in areas where all roads are exposed to avalanches at some point, as they will become completely isolated. Thus, one is forced to evaluate risk versus cost.

At the road section along Holmbuktura, this risk-to-cost threshold has been set at avalanche danger level 4 (Eivind Schnell Juvik, personal communication). For danger level 1, the road is considered safe. However, most days have avalanche danger levels 2 and 3. Danger level 1 is meant to indicate that the probability of an avalanche release is $< 10\%$ and that the expected potential avalanches are too small to pose a danger to human lives and infrastructure. Danger level 4 indicates a $> 70\%$ chance for an avalanche to happen (Eivind Schnell Juvik, personal communication). To reduce the risk for the inhabitants of Lyngen without increasing the societal cost of closing the road, we examine if statistical models can contribute to the safety assessment of the daily avalanche situation.

A radar-based warning system has been implemented to mitigate the risk of the road section. This system detects avalanches and closes the road using traffic light signals. The radar has been detecting avalanches since 2017 and provides a good data foundation for testing different models, as it provides high-quality avalanche observations. Therefore, it is a good case for training, validating, and comparing models.

The European Avalanche Warning Services (EAWS) defines site-specific avalanche warning as assessing the likelihood(e.g., probability) of an avalanche releasing and endangering people or infrastructure in specific avalanche paths (EAWS, 2022a). In Norway, the avalanche warning is given daily from the 1st of December to the 31st of May (NPRA). It is a collaboration of The Norwegian Water Resources and Energy Directorate (Noregs vassdrags- og energidirektorat, NVE), The Norwegian Meteorological Institute (Meteorologisk Institutt, MET), and The Norwegian Public Roads Administration (Statens Vegvesen, NPRA). The Norwegian mainland is divided into 24 avalanche regions. Every day, the avalanche group, consisting of three avalanche forecasters, "snøskredvarslere," and one avalanche metrologist makes and publishes a forecast for each of the 24 avalanche regions in addition to Svalbard. They base their forecasts on information of the previous weather situation, meteorological forecasts, and reports that are handed in from around the country (Varsom.no, a). They use this information to publish a general warning for each region, giving the most widespread or dangerous type of problem and at which altitudes and cardinal directions they most likely will be present. This is published every day at `Varsom.no` and in the Varsom phone application (Varsom.no, b). The expert avalanche forecasts provide regional assessments of danger levels but lack the specificity required for road safety decisions, particularly in critical sections like Holmbuktura. Since 2020, NPRA has been responsible for natural hazard

preparedness across the entire public road network in Norway (Orset and Frekhaug, 2024). This includes avalanches, and for critical road sections like Holmbuktura, they provide site-specific warnings and other mitigation measures(Orset and Frekhaug, 2024).

This project is built on and is motivated by the work of Sigbjørnsen (2024). The study aimed to predict a daily avalanche danger count for the road section in Holmbuktura and studied the potential for using satellite observations together with weather observations and forecasts. Sigbjørnsen (2024) found that the satellite observations did not give extra predictive power. We base our choice of explanatory variables on the findings of Sigbjørnsen (2024) and use simplified versions of his models. These models are cross-validated over several seasons to evaluate performance and compare their influence on operational decision-making.

We are using Sigbjørnsen (2024) results regarding the variables chosen in our models and simple versions of his models that we cross-validate over several seasons to evaluate our results and compare how the models perform differently to influence an operational decision. Similar studies as the one of Sigbjørnsen (2024) have explored various methods for predicting avalanches. Hennum (2016) applied Random Forests and logistic regression to avalanche data from Senja, concluding that random forests were more effective for predicting avalanche activity. Buser (1983) investigated the use of nearest-neighbor methods for predicting avalanche releases, and Davis et al. (1999) applied tree analysis to examine the relationship between weather variables and avalanches. Further has site-specific avalanche forecasting been claimed to succeed in real situations by Lome et al. (2024) and (Krogh et al., 2024), but without any evaluation in a statistical framework. Despite a large amount of research, there is a lack of studies with a systematic approach to and assessment of the mathematical methods regarding site-specific probabilistic forecasting of avalanches. By choosing a set of already tested variables and processes, we thoroughly explore, analyze, and compare the model's performance, the influence of the danger levels, and possible contribution to operational decisions.

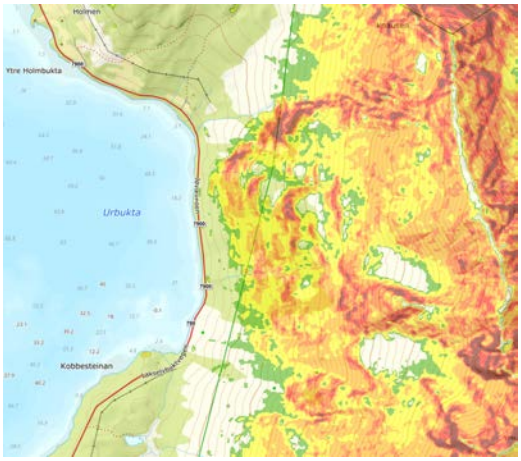
Our models are fit both to a dataset containing observations for all days in our observation period, without knowing the forecasted danger level, and independently to each danger level. The models assesses the probability of an avalanche being released in Holmbuktura daily. To investigate the research question, we propose four different statistical regression models to predict the likelihood of an avalanche release given certain relevant weather conditions, with a cross-validated climatology as a fifth model. A thorough evaluation is conducted using leave-one-season-out cross-validation. The evaluation metrics being used are the Brier score, Brier skill score, and reliability and sharpness diagrams. Further, we introduce and compare the models with the regional expert avalanche forecast.

Chapter 2 provides a fundamental introduction to avalanche dynamics and forecasting,

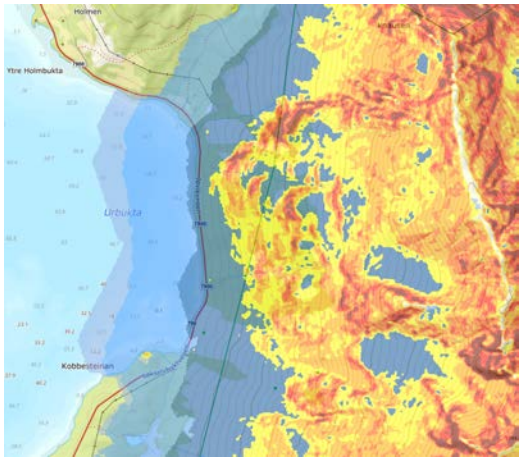
along with the mathematical theory underlying the models. Chapter 3 introduces and explores the chosen road section and the data being used. In Chapter 4, the methods are described in detail. The results are presented in Chapter 5 and further discussed in Chapter 6. Finally, Chapter 7 concludes the project with a summary of the findings and offers suggestions for future work.

Figure 1: The red line shows the avalanche-exposed section of county road 7900 in Holmbuktura (vegdata.no).

Figure 2: Informational poster in Holmbuktura(Fylke, 2024).



(a) The map shows the steepness of the slope. A darker color indicate a steeper slope.



(b) Steepness of the slope and potential avalanche zone. A darker blue indicate a higher frequency of avalanches affecting the area.

Figure 3: The maps show the steepness of the slope and potential run-out zones for Holmbuktura. The detailed color explanations are given in Figure A1 and A2 in Appendix A.

2 Background

2.1 Avalanches

If not otherwise indicated this subsection is based on EAWS reports If not otherwise indicated, this subsection is based on EAWS article *Typical Avalanche Problems* (EAWS, 2022b). An avalanche is, by the European Avalanche Warning Services (EAWS, 2024), defined as rapidly moving snow masses that have a volume exceeding 100 m^3 and a minimum length of 50 meters. For an avalanche to happen, certain criteria must be fulfilled. Firstly, there must be snow. Secondly, the slope must be steeper than 30°. Finally, a weak layer must be present (Nes, 2018). A weak layer is, by the EAWS, defined as a "snowpack layer in which the crystals are poorly bonded and the layer has the potential to fail," (EAWS, 2024). As displayed in Figure 3, FV7900, our study road, is exposed to avalanches, as the entire road lies within the run-out zone of potential avalanches, except for two smaller sections.

Avalanches are split into two different classes: slab avalanches, as seen in Figure 4a, and loose snow avalanches, shown in Figure 4b. In a slab avalanche, a slab of snow suddenly releases at once due to a weak layer in the snowpack. Slab avalanches are highly dangerous, and one has to wait for the snowpack to stabilize for the danger to decrease unless active

(a) Slab avalanche bca.

(b) Loose snow avalanche (ortovox).

Figure 4: Slab avalanche vs. loose snow avalanche.

measures like bombing the snowpack are possible (Wolfgang Fellin, personal communication March 2024). In a loose snow avalanche, the cohesion between the snow particles in the fresh snow is lost. A loose snow avalanche starts at one point and widens like a fan(EAWS, 2024). These avalanches are often small and unlikely to bury a person or a car .

The climatic situations that might lead to avalanches are divided into five different categories . These are new snow, wind slabs, persistent weak layers, wet snow, and gliding snow. The different kinds of problems will trigger various types of avalanches, and although a short description of the various problems will be provided, based on EAWS (2022b), is this not something that will be considered in the mathematical models of this project.

New snow is related to current or recent snowfall. The amount of fresh snow on the already existing layer, the state of the current snowpack, and the weather conditions are the deciding factors for the criticality of the problem. New snow can lead to both dry loose snow avalanches, and slab avalanches .

The second mentioned problem is wind slabs. This occurs when the snow has been transported by the wind, either dry snow on the ground or windy and snowing simultaneously. Wind slabs typically lead to dry slab avalanches but are only located in some areas, contrary to new snow problems, which generally are "everywhere" .

The third avalanche problem is persistent weak layers. A weak layer is a layer in the

snowpack with weak cohesion between the snow crystals. This occurs in cold weather when faceted crystals, depth hoar, or surface hoar crystals are covered with new snow, and the new snow stabilizes internally but not with the previous top layer. This problem can last a few days to months and sometimes the entire winter season .

Fourthly, we have wet snow problems. When the snow is heated or it is raining, there can be a significant amount of water in the snow. This reduces the cohesion in the snow and thus increases the danger of avalanches. Wet snow can lead to both wet loose snow avalanches and wet-snow slab avalanches.

The last type of avalanche problem is gliding snow avalanches. This is when the entire snowpack glides on the ground, mainly on grassy slopes or smooth rock faces. Glide snow avalanches are particularly difficult to predict.

Spatially, an avalanche has 3 zones. It is the starting zone, which is the area where an avalanche releases. The slope in this area must be larger than 30 degrees. The second area is the avalanche path. This is the "way" that the avalanche "travels" down the mountain. The third zone is the avalanche deposit area, also known as the run-out zone. Each of the regions provides different risks. One recon that the maximum distance an avalanche travels is at an 18-degree angle from the release point.

2.1.1 Avalanche warning

As said in the introduction, the avalanche warning service in Norway is a collaboration between NVE, MET, and NPRA. From December 1st to May 31st, they publish daily forecasts for certain regions, including Lyngen. In the case of an avalanche danger of 4 or 5, they also publish warnings for the unprioritized avalanche region. The avalanche warning regions are big, as shown in Figure 6, and the forecasted avalanche danger level is based on the most common or most dangerous problem in every region. The avalanche warnings are based on meteorological data from previous weeks, snowpack conditions, and expert observations, with the information communicated through the platform Varsom (NPRA). This warning is based on the EWAS matrix shown in Figure 5.

Within the avalanche forecasting field, one separates the forecast type into 3 different spatial scales. Synoptic scale ($> 10^4 km^2$), meso-scale ($> 10^2 km^2$) and micro scale ($< 1 km^2$) (McLung, 2007). The avalanche warning regions in Norway are of a synoptic scale and represent the highest risk or most widespread problem. Holmbuktura lies in between the definition of meso- and micro-scale. The smaller the area, the more complex the avalanche warning gets, as one needs more precise information (McLung, 2007). In 2022,

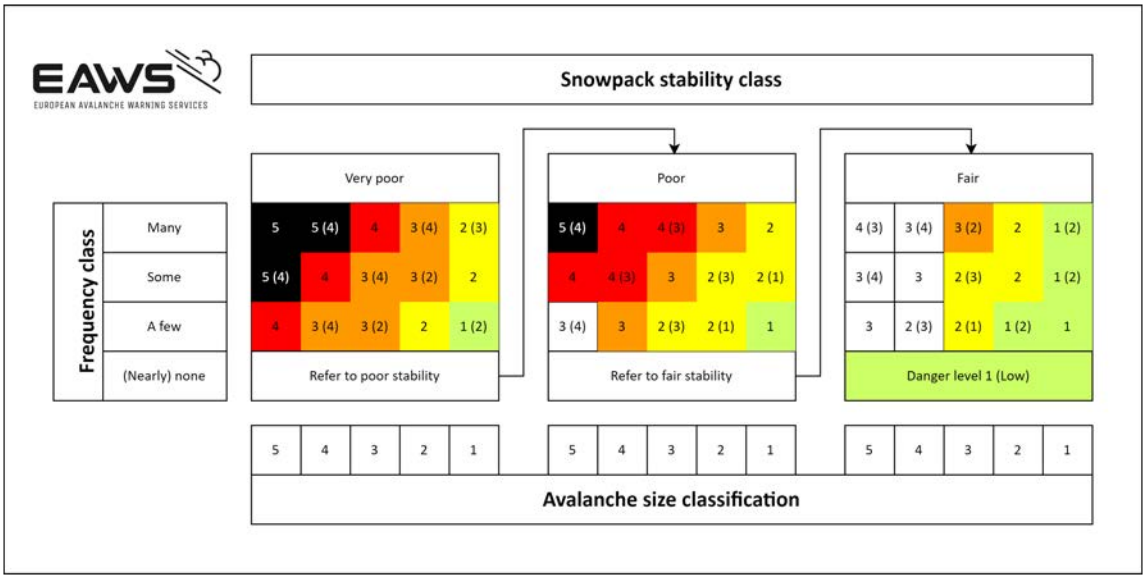
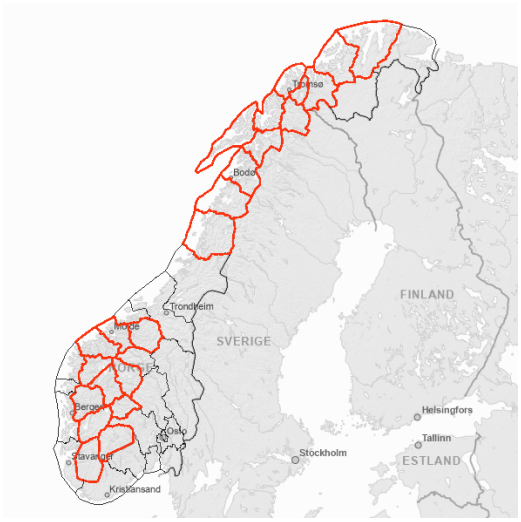
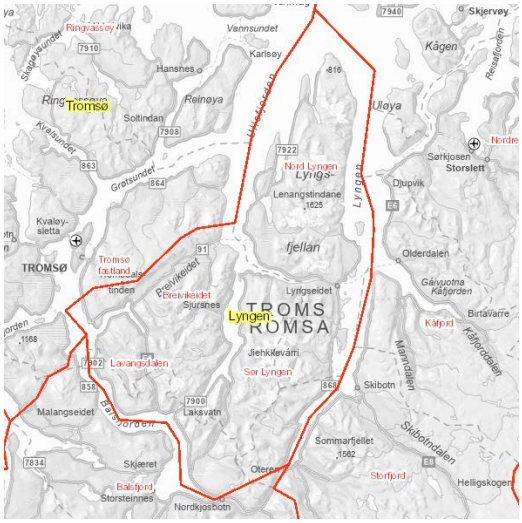


Figure 5: The EAWS Matrix is a tool used for determining the avalanche danger level based on snowpack stability, its frequency distribution, and avalanche size (EAWS Matrix).



(a) All warning regions on the Norwegian mainland



(b) The lyngen region.

Figure 6: Red lines indicate regions where the avalanche danger is forecasted for the entire season. Black outlines are the regions where the warning is only published on days with danger levels 4 and 5. Spitsbergen (Svalbard) and Jan Mayen are not included in these figures and are not among the prioritized regions (senorge.no).

the general assembly of EWAS developed a new set of recommendations for site-specific avalanche warnings (Jaedicke et al., 2024). They state that one should use the release and run-out probability when assessing the avalanche danger(EAWS, 2022a). This highlights the relevance of this project.

2.1.2 Avalanche mitigation

Avalanche mitigation can be divided into permanent and temporary measures. These can again be divided into active and passive measures. Exampels of permanent measures are snow nets, tunnels, planting forests, or, as in Holmbuktura, a traffic light system or a permanent warning sign. Within temporary measures, we find actions such as avalanche forecasts, artificially triggering avalanches, or closing or evacuating areas when the danger is too high. Traffic lights, warning signs, and avalanche forecasts are passive measures; triggering avalanches or physically blocking them are active measures (Wolfgang Fellin, personal communication, May 2024).

It could be an idea to build a road tunnel for FV7900 for the stretch of Holmbuktura, but tunnels are expensive and not in line with Tromsø City Counsils Transport Plans avalanche mitigation program, which rather aim to use cost-effective measures like radar surveilance and avalanche forecasting Krogh et al. (2024). A combination of a well-calibrated radar system and accurate forecasters is a more affordable measure, in line with the Transport Plan.

2.2 Cross validation

Cross-validation (CV) is a method evaluate models and balance using as much as possible data for training the model, and avoid using the same data for training and evaluation. CV is done by dividing the training data into K parts. This can be done either by randomly selecting K equally sized groups. After dividing the data, one group is taken out of the training set, and the method is fitted for the remaining (K-1)/K parts of the training data. The fitted model is then tested on the part that was taken out and evaluated. This is done in turn for each of the K groups into which the data was divided. This method of CV is called K-fold cross-validation (Hastie and Tibshirani, 2009, pp. 241–245). For this project, one season has been left out at a time, e.g., a leave-one-season-out cross-validation (LOSO-CV) method has been chosen, as it, in addition to reducing the risk of over-fitting, also gives insight into the prevalence of seasonal changes.

2.3 Probabilistic forecasting

Forecasting is the process of making predictions about the future based on past and present data (Wikipedia, b). Gneiting and Katzfuss (2014) defines a probabilistic forecast as a forecast in the form of a distribution over future quantities or events; hence, when one uses predictions and probability distributions to make the forecast (Gneiting and Katzfuss, 2014, p. 126). In this project, we are suggesting and examining different models that aim to predict the probability of an avalanche release, hence comparing probabilistic forecasts. The goal for each model is to predict whether or not an avalanche will be released on a given day, represented as a binary event y_i, where $y_i = 0$ indicates no avalanche and $y_i = 1$ indicates an avalanche. E.g., are we looking at binary responses, thus do we have a bernoulli variable, with probability mass function

$$f(y; p) = p^y(1-p)^{1-y} \quad \text{for } y \in \{0, 1\}. \tag{1}$$

Here, y is the event of no avalanche (0) or avalanche (1) and p is the probability for an avalanche event that we are estimating in order make a probabilistic prediction. Therefore, we consider probability mass functions where the following conditions must hold:

$$\sum_{y_i \in \{0,1\}} f(y_i \mid x) = 1, \tag{2}$$

$$f(y_i \mid x) \geq 0 \quad \forall y_i \in \{0, 1\} \tag{3}$$

Here, Y_i are the possible binary outcomes, and x is the vector of explanatory variables. The probability distribution of the forecast is of interest because it allows us to quantify the uncertainty of the predictions. In the past few decades, probabilistic forecasting has gained popularity and has been applied in various fields Gneiting and Katzfuss (2014, p. 126).

Avalanche occurrences are difficult to predict with accuracy due to their dependence on multiple interacting factors such as weather conditions, snowpack characteristics, and terrain features, as described in Chapter 2.1. As a result, forecasts should also be probabilistic to account for this uncertainty.

2.4 Evaluating forecasts

An ideal forecast is a forecast that perfectly predicts the right event (Gneiting et al., 2007). As this is unlikely, the aim is to make best possible forecast. Certain diagnostic tools might be useful for evaluating forecasts and improving skills. When evaluating the calibration, one also evaluates the performance of the forecaster.

One needs diagnostic tools to assess the quality of the calibration of a probabilistic forecaster and, thus, its performance.

2.4.1 Brier score

Scoring rules give a numerical score to probabilistic forecasts. They are an attractive way to summarize predictive performance measures, simultaneously addressing sharpness and calibration (Gneiting and Katzfuss, 2014). Gneiting and Katzfuss (2014) states, "One can view a scoring result as a negatively oriented penalty that one wishes to reduce as much as possible.". The most favorable forecast is thus the one with the lowest scoring result. The score is denoted $s(F, y)$, where F is the predicted distribution of the probabilistic forecast when the event $y \in \mathbb{R}$ materializes (the realized value (in a binary case, $y \in \{0, 1\}$ where 0 is no avalanche and 1 is avalanche). The Brier score (BS) is defined as

$$BS = \frac{1}{T} \sum_{t=1}^{T} (f_t - y_t)^2 \tag{4}$$

where f_t is the predicted probability of an event and y_t is the event at time t (Mason, 2004). BS is a proper scoring rule that distinguishes between calibration and refinement components. More precisely, the BS measures the mean squared difference between the predicted probability f_t assigned to the possible outcomes for the event at time t and the actual outcome o_t at time t across all times $t \in \{1, \ldots, T\}$ in a set of T predictions. The BS is negatively oriented; a lower score implies a better forecaster. (Jolliffe and Stephenson, 2011, p. 145).

Although the BS is a proper scoring rule, ensuring that the forecaster with the best score provides predictions closest to the true observations—it does have some weaknesses. For instance, a forecast that consistently predicts a 50% probability for an event will always achieve a BS of 0.25, regardless of the actual outcomes. This does not imply that a forecast with a worse BS cannot provide useful forecasting information. It is essential to analyze the tendencies of the forecast; for example, if a forecaster consistently underpredicts the probability of avalanche activity, it might still effectively indicate periods of higher or lower avalanche danger. Such tendencies can provide valuable input for operational decision-making.

2.4.2 Brier skill score

The Brier skill score (BSS) compares a forecast's relative skill to a reference forecaster's skill. It is defined as

$$BSS = 1 - \frac{BS_{forecast}}{BS_{ref}}. \tag{5}$$

Here $BS_{forecast}$ is the BS of the forecast we are suggesting and BS_{ref} is that of a reference forecaster. If the $BSS \geq 0$, the forecast has a better skill than the reference forecast. If $BSS = 0$, the suggested forecaster and the reference forecaster have equal skill. If the BSS is negative, the forecast is misleading, and the reference forecast is to be preferred (European Centre for Medium Range Weather Forecasts, 2018). The BSS is thus postitive oriented, the higher the score, the better the forecaster, relatively to the reference (Jolliffe and Stephenson, 2011).

2.4.3 Evaluation of Statistical Properties

When one wishes to evaluate a forecast system, one must use a statistical approach. For a binary probabilistic forecast to be statistically correct, one must have a sufficient amount of results and check that the predicted event occurs at a rate similar to the predicted probabilities (Jolliffe and Stephenson, 2011, p. 139). For example, should it, for 20% of the days where a 20% probability of an avalanche is predicted, be an avalanche. If the observed proportion of events significantly differs from the predicted probability, the forecast is statistically incorrect (Jolliffe and Stephenson, 2011, p. 139). One condition for the validity of uncalibrated probabilistic forecasts for the occurrence of an event is thus consistency between the predicted probabilities and the observed frequencies. This consistency is required for users who want to decide based on an objective quantitative risk assessment. This property of statistical consistency is called reliability. Reliability is just a necessary but not a sufficient measure for the usefulness of a probabilistic forecaster (Jolliffe and Stephenson, 2011, p. 139).

The reliability can be evaluated with a reliability diagram. A reliability diagram shows the relation between the proportions of the occurred event and the predicted probability for the event. A reliability diagram for a perfectly reliable forecaster is a 45-degree diagonal line where the forecasted probabilities match the observed frequencies exactly. To construct a reliability diagram, one bin the forecasted probabilities and calculate the observed proportion of the event for each bin. The mean of the forecasted probabilities is then plotted along the x-axis with the corresponding portion of events on the y-axis. To

introduce an indicator of the uncertainty of the reliability, a 95% confidence interval for each bin is calculated. The confidence interval is based on the assumption that each bin can be modeled as a binary event with the probability for success equal to the mean of the predictions in the bin and the number of trials equal to the number of observations in each bin. The value for the confidence interval is then divided by the number of observations in the bin.

A second condition is thus that a forecast system must be able to separate between situations under which an event occurs with lower or higher frequency than the climatological frequency. This is called resolution (Jolliffe and Stephenson, 2011, p. 126). The more distinct the observed frequency distributions for various forecast situations are from the full climatological distribution, the more resolution the forecast system has. A maximum resolution is obtained when reliable forecast probability distributions have zero spread, i.e., they are concentrated on single points. This corresponds to a perfectly reliable deterministic forecast system (Jolliffe and Stephenson, 2011, p. 141). For a binary event, this means that the predicted probability for the event is 0 for all occasions when the event does not occur and 1 for all occasions when the event does occur. A reliable forecast system's resolution is identical to sharpness (Jolliffe and Stephenson, 2011, p. 142). Sharpness refers to the concentration of the forecasted distribution and does not take the true observation into account. Sharpness is a highly attractive quality as it refers to the precision of the forecast. The sharpness diagram is a tool to evaluate the sharpness of a forecast. The sharpness diagram shows the frequencies for the forecasted probabilities divided by the total number of forecasts. A forecast system is called sharp if it is able to predict probabilities that differ from the observed frequency. A forecast system with little sharpness will have an increased frequency for values close to the regular climatological frequency. These systems will not add a significant value compared to the regular climatology (The MET Office).

.

2.5 Forecasting methods and statistical models

The methods used in this project are now introduced.

2.5.1 Climatology

Climatology, denoted as ρ_c, is the proportion of occasions that a given event occurs (European Centre for Medium Range Weather Forecasts, 2018). The regular climatology is thus

given by

$$P_{clima} = \frac{\sum_{i=1}^{n} y_i}{n}, \tag{6}$$

for an event y_i observed over n periods. Climatology is the most widely used reference forecaster when calculating the BSS explained in section 2.4.2 (Mason, 2004). The regular climatology is per construction reliable, but it is not useful as it does not provide any forecast information apart from the proportion of occasions that the event occurs. The regular climatology is reliable per construction, but it is not useful as it does not provide any forecast information apart from the portion of the event that occurred. The regular climatology can take the role as baseline. All proposed models should improve compared to the climatology.

2.5.2 Generalized Linear Methods

Generalized Linear Models (GLMs) provide a flexible framework that unifies various regression approaches for instances where the response variable does not follow a normal distribution. However, GLMs rely on the assumption that the effect of the covariates can be modeled using a linear predictor (Fahrmeir et al., 2022, p. 283).

Linear regression models and regression models for non-normal response variables share common properties that can be summarized within the GLM framework:

1. The mean, $\mu = E(y)$, of the response variable y is connected to the linear predictor $\eta = \mathbf{x}'\beta$ by a response function h or a link function $g = h^{-1}$.
2. The distribution of the response variables can be expressed in the form of a univariate exponential family.

A GLM describes the relationship between the mean of a response variable Y and an independent variable x. It consists of three components: a random component, a systematic component, and a link function (Casella and Berger, 2022, p. 591). The response variables $Y_1, \ldots, Y_n$ form the random components. These are assumed to be independent and follow a distribution from the exponential family.

An exponential family of probability density functions (pdfs) or probability mass functions (pmfs) is defined as:

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) (Casella and Berger,\ 2022, p.\ 111).$$

The model represents the systematic component, where the function of the predictor variables x_i relates to the mean of Y_i. The link function $g(\mu)$ connects the random and systematic components. Specifically:

$$g(\mu) = \eta, \quad \text{where } \mu_i = E[Y_i], (Casella and Berger,\ 2022, p.\ 591).$$

The distribution function for the univariate exponential family can also be reformulated as:

$$f(x|\boldsymbol{\theta}) = \exp\left(\frac{y\theta - b(\theta)}{\phi} w + c(y, \phi, \theta)\right), (Fahrmeir\ et\ al.,\ 2022, p.\ 317).$$

For a Poisson distribution, the density is given by:

$$f(x|\lambda) = \frac{\lambda^y \exp(-\lambda)}{y!}, \quad y = 0, 1, \ldots$$

(Casella and Berger, 2022, p. 111), which can be reformulated as:

$$f(x|\lambda) = \exp\left(y \ln(\lambda) - \lambda - \ln(y!)\right), \quad y = 0, 1, \ldots$$

This formulation shows that the Poisson distribution belongs to the exponential family. As it is univariate, it os compatible with the GLM framework. Using regression with a Poisson-distributed response rather than a logistic response offers the advantage of incorporating avalanche counts, thus utilizing more detailed information. Poisson regression is a type of GLM used to model count data. The response variable y_i represents the number of events occurring within a fixed period or space and follows a Poisson distribution:

$$y_i \sim \text{Poisson}(\lambda_i),$$

where λ_i is the expected event occurrence rate. The rate parameter λ_i, which is the expected value of y_i, is related to the linear predictors $\mathbf{x}_i$ through the log-link function:

$$\log(\lambda_i) = \mathbf{x}_i^\top \beta,$$

which can be rewritten as:

$$\lambda_i = \exp(\mathbf{x}_i^\top \beta).$$

Here:

- $\mathbf{x}_i$ is the vector of predictor variables.
- β is the vector of coefficients for the predictors, forming the systematic component of the GLM.

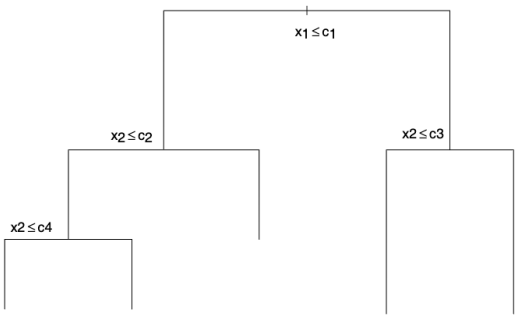


Figure 7: Example of a simple binary decision tree (Copied from (Tutz, 2012), p. 318).

The log-link function ensures that the rate parameter λ_i remains positive (Fahrmeir et al., 2022, p. 308).

When fitting a Poisson GLM, the model estimates the rate parameters $\lambda = (\lambda_1, \ldots, \lambda_n)$. The probability mass function is given by:

$$f(y|\lambda) = P(Y = y),$$

and the probability for $P(Y \geq 1)$ is computed as:

$$P(Y \geq 1) = 1 - P(Y = 0).$$

2.5.3 Tree regression

Tree analysis has its roots in automatic interaction design (AID). The most commonly used type of regression tree is classification and regression trees (CART) (Tutz, 2012). The method is, in principle, simple and uses binary recursive partitioning to divide the feature space into a set of rectangles, as shown in Figure 8. A model is then fitted to each rectangle. Unlike traditional regression models that estimate parameters, tree models produce a binary tree that visualizes the partitioning of the feature space, as demonstrated in Figure 7 (Tutz, 2012). In CART, Classification refers to the prediction of underlying classes, and by treating the classes as the outcomes of a categorical variable, classification can be treated within the general frame of regression (Tutz, 2012).

In a setting with a one-dimensional response variable y and a two-dimensional predictor vector $\mathbf{x}$, the tree is constructed as follows:

- **Step 1:** Choose a predictor and a split point to divide the predictor space into two

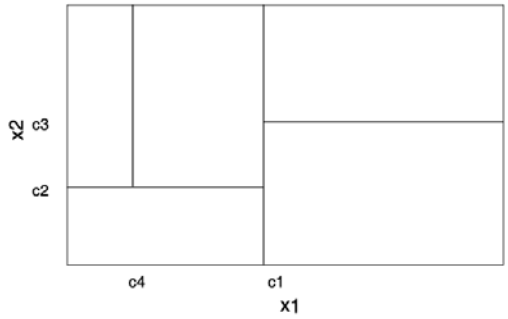


Figure 8: Example of partitioning the feature space (Copied from (Tutz, 2012), p. 318).

regions based on a splitting rule. For example, in Figure 8, the split yields:

$$\{x_1 \leq c_1\}, \quad \{x_1 > c_1\}.$$

- **Step 2:** Next, split one or both of the obtained regions using the same criteria for predictor and split-point selection. In Figure 8, this corresponds to the splits at c_2 and c_3.
- **Step 3:** Continue splitting until a stopping criterion is met. For example, Figure 8 shows the resulting partitioning of the feature space into five regions, which corresponds to the upside-down tree shown in Figure 7.

This process produces a regression region with a piecewise constant fit:

$$\hat{\mu}(\mathbf{x}) = \sum_{i=1}^{5} \gamma_i I((x_1, x_2) \in R_i),$$

where $\hat{\mu}(\mathbf{x})$ is the predicted value for a given input with parameter vector $\mathbf{x}$, and $\gamma_i$is the response value assigned to region R_i, and $I((x_1, x_2) \in R_i)$ is the indicator function.

Although it becomes challenging to visualize partitions like those in Figure 8 when there are more than two predictors, the tree structure shown in Figure 7, can still be used, and a procedure similar to the one described above can be followed.

When constructing a tree-growing algorithm, one has to decide upon the criteria used to determine the best split and when to stop splitting to avoid overfitting. In standard splitting, the data at a node A is divided into two groups, A_1 and A_2, based on a single variable. For variables measured on a continuous or ordered scale, the split is defined as:

$$A_1 = A \cap \{x_i \leq c\}, \quad A_2 = A \cap \{x_i > c\},$$

where c is the threshold value. The number of possible splits depends on the scaling of the variable and the dataset $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$. For continuous variables, there are $n-1$ possible thresholds. While splits into more than two groups are possible, this project focuses exclusively on binary trees. The splitting criteria can vary depending on the approach. Test-based splits is a method where Potential splits are tested against no split, with the best split chosen iteratively until an optimal result is achieved. The stopping criterion can be the maximum number of splits, branches, or leaves, or it can be a numerical criterion to prevent overfitting. By carefully managing the splitting process and stopping criteria, tree models can balance predictive accuracy and generalization to unseen data.

2.5.4 Random Forest method

Random forests (RF) are among the most efficient classifiers proposed in recent years (Tutz, 2012, p. 458). RFs are ensembles of decision trees, all trained on the same dataset but perturbed or modified by some random mechanism to ensure diversity among the trees (Fahrmeir et al., 2022, p. 73). According to Breiman (2001, p. 6), a random forest (RF) is defined as a collection of regression or classification predictors $f(\mathbf{x}, \theta_l)$, $l = 1, \ldots, L$, where θ_l are independent and identically distributed random variables. For regression tasks, the RF predictor $\hat{f}_{RF}(\mathbf{x})$ is the average of the individual tree predictors $f(\mathbf{x}, \theta_l)$. For classification, the final prediction is the class with the highest average percentage (majority vote) across the trees.

Each tree in a random forest is grown using a random mechanism to introduce variation and prevent the creation of identical trees (Fahrmeir et al., 2022, p. 73). The final RF predictor is obtained by aggregating the predictions of all individual trees in the ensemble. Both empirical evidence and theoretical studies suggest that random forests offer improved predictive performance compared to a single decision tree. Furthermore, RFs address the issue of overfitting, which is a common problem with standalone decision trees (Fahrmeir et al., 2022, p. 73).

2.5.5 Method of Nearest Neigbors

The method of nearest neighbors (kNN) is a non-parametric algorithm that estimates the response of a given input x by averaging the responses of its k-nearest neighbors in the input space. More formally, kNN identifies the k observations x_i in the training set T that are closest to x, based on a predefined distance metric, and computes the estimate $\hat{p}(x)$ as:

$$\hat{p}(x_j) = \frac{1}{k} \sum_{x_i \in N_k(x_j)} y_i,$$

where x_j is the parameter vector for the observations which response we want to predict, for the is the $N_k(x_j)$ represents the neighborhood of x_j, defined by the k closest points x_i in the training sample, and $\hat{p}(x_j)$ is the probability that $y = 1$ (Hastie and Tibshirani, 2009). A distance metric quantifies the similarity or proximity between observations in the feature space to determine the closeness between data points. Common distance metrics include Euclidean, Manhattan, and Minkowski distances. For this project, the Euclidean distance is applied to scaled parameter values to ensure that all features contribute equally to the distance calculation, e.g., that all columns in the parameter matrix have the same mean and standard deviation.

For kNN, the error on the training data is expected to increase as k becomes larger, and it should be exactly 0 when $k = 1$, as the prediction is based on the closest data point. However, evaluating kNN models on an independent test set provides a more reliable comparison of different parameter choices and configurations (Hastie and Tibshirani, 2009).

To assess the performance of kNN models, metrics such as the Sum of Squared Errors (SSE) and the coefficient of determination (R^2) are used.

3 Study area, data and exploratory data analysis

3.1 Location

Holmbuktura is technically the rocky area by the shoreline between Ytre Holmbukt and Kvitbergsbukta, and is located at $69°29'24.36814''\text{N}, 19°43'59.13648''\text{E}$. The highest peak on the ridge above Holmbuktura measures 1170 masl. For the remainder of this project, Holmbuktura will refer to the entire east face of the mountain between Fugldalen and Holmbukt.

3.2 Data

Our response and parameter variables have gathered data from four different sources. First, we have data from avalanches detected by a radar detector in Holmbuktura. Then, as our parameter variables, we have the danger level set by avalanche experts, precipitation and snow measurements from a weather station in Ytre Holmebukt, and interpolated values of the observed weather from several nearby weather stations.

3.2.1 Radar Detections

During a test run of the radar in Holmbuktura in 2017 and 2018, all the relevant avalanche releases were detected, and the false alarm rate was 3.4 percent (Meier, 2018). The radar in Holmbuktura is delivered and operated by Wyssen (Eivind Schnell Juvik, personal communication) on behalf of NPRA and is a long-range avalanche radar (LARA®). The radar is located at the bay's north end, as shown in Figure 2. It has a range of up to 4000m, and the opening angle is $90\,\text{deg} \times 15\,\text{deg}$. This means that one radar covers the entire mountainside in Holmbuktura. LARA® is a Doppler radar, which means that it uses the principles of the Doppler effect to detect moving masses. After a server has processed the data, it returns the information about when, where, the size, and the duration of an avalanche.

The radar has been operating since February 2017 (Eivind Schnell Juvik, personal communication) and has continuously detected avalanches since then. The data has been received in two parts. The first period runs from February 2nd, 2017, to April 26th, 2023. The second part covers detections from October 15th, 2023, to June 9th, 2024.

The official period for publishing daily avalanche warnings in Norway is from December

Figure 9: The area covered by the radar is marked in green. The green circle at the top left corner of the marked area indicates the position of the radar. The numbers along the road indicate the position of the traffic lights (vegdata.no).

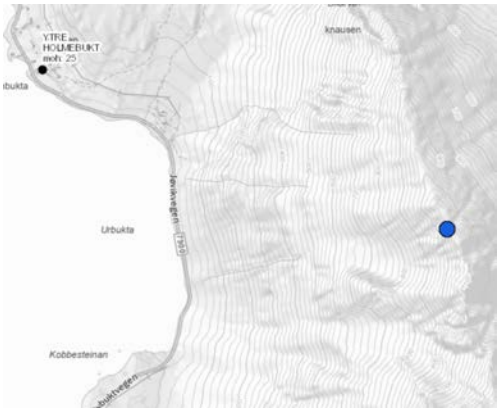


Figure 10: Weather information positions. The blue circle shows the geographical point used for the interpolated data of observed weather, while the black circle displays the position of the MET weather station (xgeo.no).

1st to May 31st. Since we are using the avalanche warning as one of our parameter values, we will only look at avalanches that occur during this period for every season. Since we wanted a dataset containing the same parameters for all observations, we only used the parameters included in radar detections for both periods. When combining the datasets and removing the dates outside of the period together with the parameters that are only given for one period, we end up with the 7 parameters and a total of 550 avalanche detections. A weakness of this dataset is that we lose the possible information about false positive detections from the radar, which increases the uncertainty of the final results. This is done as it was only provided for the dataset for the first period. These data have been restructured so that there is one row for each day, between December 1st and March 31st for every season, containing the number of avalanches detected every day.

Figure 11 displays the autocorrelation function for the number of avalanches using data from all seasons from 2017 to 2024. It's clear to see that the correlation of the number of avalanches is limited to one day, and even here, it's a low correlation. In Figure 12, the ACF for just one season, 2021-2022, is shown, and it is clear that the time dependence of the number of avalanches is also short within one season.

In Table 1, the frequency of the number of avalanches within one day is displayed, and it is clear that there is a majority of days with no avalanches at all and that the frequency decreases as the number of avalanches increases. Table A3 in Appendix A displays the frequency of the number of avalanches within one day for every season. The day with the most avalanches was March 30th, 2023, with 37 detected avalanches in one day. We will

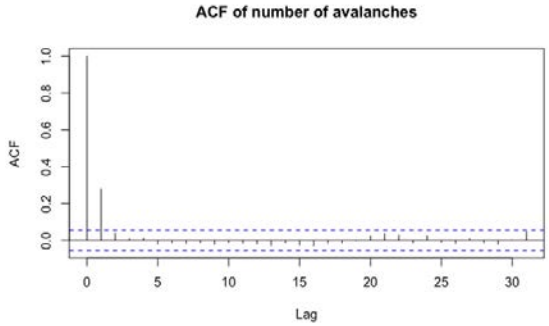


Figure 11: Auto correlation function of the number of avalanches. The blue dotted line indicates the level of significance.

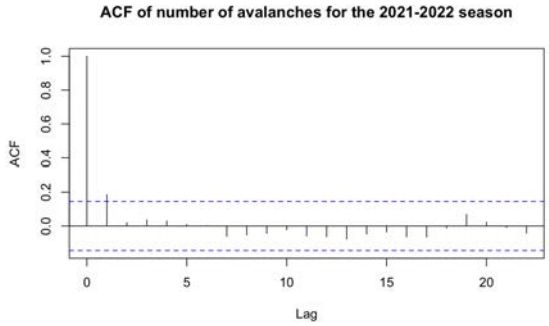


Figure 12: Auto correlation function of number of avalanches for the 2021-2022 season. The blue dotted line indicates the level of significance.

Number of avalanches	0	1	2	3	4	5	6	7	8 or more	total
Frequency	1095	84	40	18	13	5	5	3	13	550

Table 1: Frecuency of number of avalanches per day.

Season	17/18	18/19	19/20	20/21	21/22	22/23	23/24	total	with ava	portion
1	39	29	1	42	7	5	26	149	4	0.027
2	85	90	110	66	92	104	115	662	38	0.057
3	51	58	69	69	78	66	36	427	118	0.276
4	7	4	3	4	5	7	3	33	21	0.636

Table 2: Frequency of the danger levels for each season, along the frequency and portion of days with avalanche activity for each danger level.

consider a binary case and only examine events in which no avalanches were released, or one or more avalanches were released. Thus, we ignore the number of avalanches except when training the GLM model.

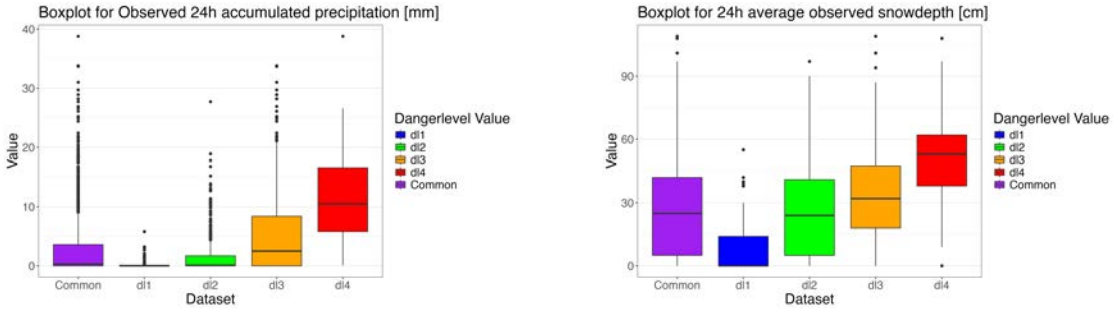
3.2.2 Danger Level Warning data

The regional avalanche warnings published on `varsom.no` can be accessed through an API. The API gives access to observations done by experts and others who have registered avalanche activity and other avalanche-relevant observations, in addition to the regional forecasted danger level. For this project, only the forecasted avalanche danger will be used.

In Table 2, the frequency of the avalanche danger levels for each season and in total are displayed. It is clear that the distribution for every year follows the same trends, with most days at level 2, second-most at level 3, then 1, 4, and finally, are there no days at level 5. When referring to dangerlevel, "dl" or "DL" followed by the danger level number is often used.

3.2.3 Weather Station Data

Observed weather data from the Ytre Holmebukt weather station, which MET monitors is also used. The location of the station is shown in Figure 10. This weather station gives information about 35 different weather variables, giving various information about precipitation, snow, and the type of weather (Norsk klimaservicesenter, 2024). Based on the work of Sigbjørnsen (2024), we have limited our choice of variables only to include the observed snow depth (in cm) and the 24h accumulated amount of precipitation (precipitation on December 2nd is the accumulated amount from 6 am December 1st to 6 am December 2nd). Figure 13 shows the box-plots for the 24 hours accumulated precipitation (13a, and the average snowdepth (13b). These show that it is a significat difference in the amount of observed precipitation among the danger levels. It is also a clear tendency of



(a) 24 hours accumulated precipitation measured at the MET weather station Ytre Holmebukt.

(b) Average snow depth measured at the MET weather station Ytre Holmebukt.

Figure 13: Box-plot of weather parameters from the MET weather station.

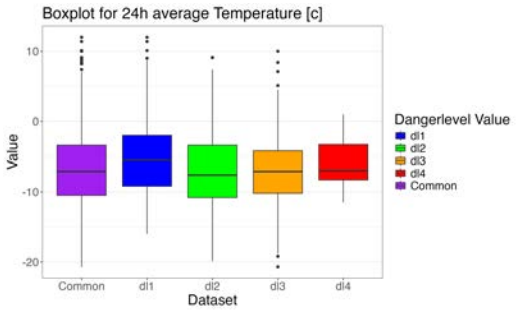
a bigger snowdepth for higher dangerlevels, but it is a bigger range and more overlap for each danger level.

3.2.4 Interpolated weather observations

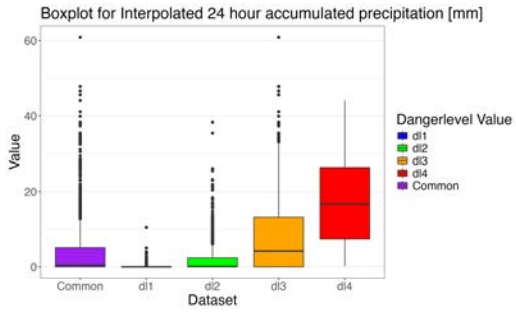
The online service `xgeo.no` allows users to access interpolated values of the observed climate situation from MET-operated weather stations. The data is interpolated over grid cells measuring 1 by 1km (Eivind Schnell Juvik, personal communication).

We have chosen one grid cell that covers the elevation at which most avalanches are released (Eivind Schnell Juvik, personal communication) and retrieved the data variables for that cell. There are a total of 52 available variables. (NVE, 2012). Based on the work of Sigbjørnsen (2024) and Baissa (2024) and advice from Eivind Schnell Juvik, it has been chosen a smaller set of variables new snow depth, snow depth, 24h accumulated precipitation, temperature, and wind direction and speed. These are explained in Table A4. Figure 14 shows the box-plots for the for the temperature and interpolated 24 hour accumulated precipitation. The temperature does not indicate big differences between each danger level. The interpolated 24 hour precipitation follows a similar patern as the observed precipitation. In Figure 15, the box-plots for the interpolated snow-variables are shown. The new snow depth shows a clear difference between the danger-levels. The interpolated snowdepth does not show the same difference between the danger levels.

Based on the windrose in Figure 16 and advice from Eivind Schnell Juvik, the wind data has been modified to one number for each day, which is the average wind speed for the cardinal directions: S-SW-W-NW-N.

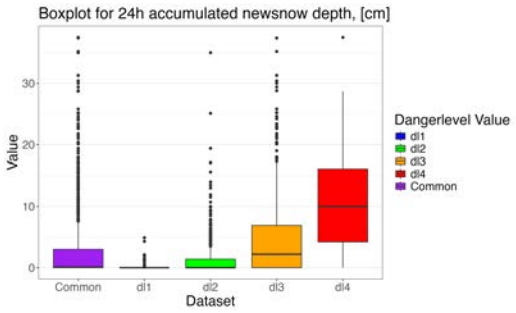


(a) Box-plot of temperature for all datasets.

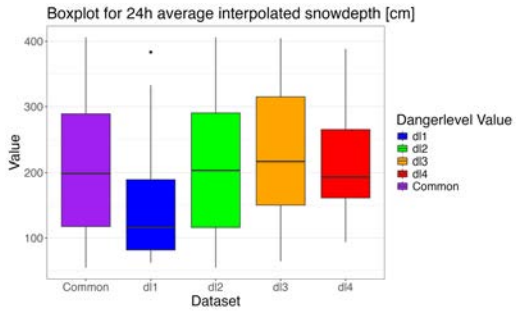


(b) Box-plot of the interpolated value of 24 hours accumulated precipitation for all datasets.

Figure 14: Box-plot of temperature and interpolated value of the 24-hour accumulated precipitation.



(a) Box-plot of 24h accumulated newsnowdept for all datasets.



(b) Box-plot of 24h average assumed snow depth for all datasets, based on interpolation of nearby stations, adjusted for terrain height.

Figure 15: Box-plot for interpolated snow parameters

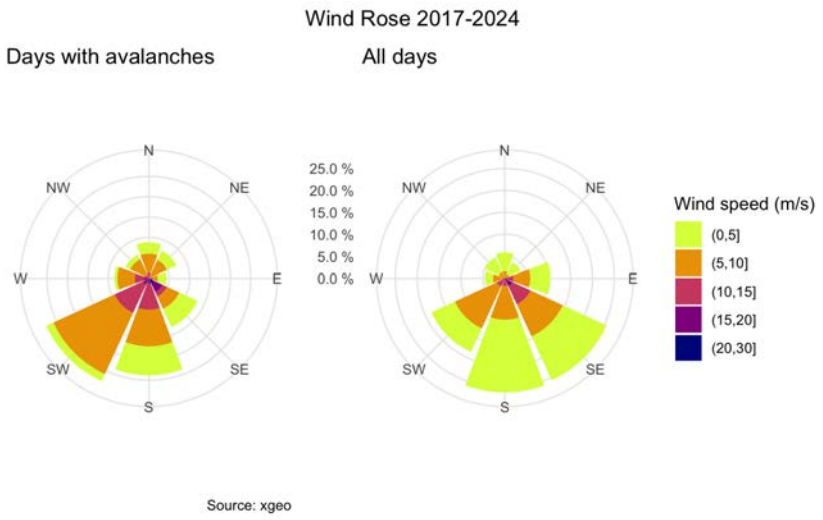


Figure 16: Windrose for the winter seasons from 2017-2024. The left windrose shows the wind situation for days with avalanches, and the right windrose shows all days.

4 Modeling

Five different methods have been chosen. Each resulting model is evaluated by its skill to predict avalanches as a univariate binary response of whether or not an avalanche is released. To minimize uncertainty in the model and maintain simplicity, the explanatory variables are the observed weather variables for the same day, not the weather forecast for the day being predicted, which would be more like a realistic situation. The parameter selection is based on the statistical inference done by Sigbjørnsen (2024). There are 1271 unique observations in the dataset, after excluding those with incorrect or outlier data, that have been used when fitting the models. The observations are from December 1st to May 31st for 7 winter seasons.

Five different types of methods have been used, as explained in Table 3.

Method Type	**Description**
Climatology	Basic statistical summaries of past data, cross-validated per season.
GLM	Generalized Linear Models with a Poisson distribution.
Binary Tree	Decision tree for classification.
Random Forest	Ensemble method using multiple decision trees.
kNN	K-Nearest Neighbors regression.

Table 3: Overview of the five modeling methods used in the analysis.

4.1 Explanatory variables

All models have been fitted using 7 explanatory variables. These are explained in introduced in Chapter 3, section 3.2.3 and 3.2.4. A detailed overview of the explanatory and response variables is given in Appendix A Table A4.

For all methods, except for the climatology, the data has been structured in two different ways for two separate fittings. The first dataset structure, hereafter denoted as "Common-dataset," includes all explanatory variables, for all days. The models are then fitted without any information about the danger level. The models fitted on the common dataset are denoted "common model" or "common models." In the second structure, the data has been partitioned into four groups based on the four danger levels, explained in the next pharagrap. These are denoted as danger-level datasets or dl datasets and still contain all explanatory variables. The models are fitted independently for each dl dataset before being

assembled to fit the original dataset before being evaluated. These models are called dl models.

For the models that are trained separately by danger level, the dataset is divided into $n = 4$ subsets, where each subset corresponds to one of the four avalanche danger levels ($D \in \{1, 2, 3, 4\}$). Specifically:

$$\text{Group } i = \{x \in \text{Dataset} \mid \text{DangerLevel}(x) = i\}, \quad \text{for } i \in \{1, 2, 3, 4\}.$$

Each subset is used to train a model independently, allowing the danger level to contribute indirectly as an explanatory variable. This is also explained in Table A1 in Appendix A.

The covariance matrix for the dataset including all days, used for the models trained on all days without information about the danger level, $\mathbf{X}_{\text{common}}$ is a 1271×7 - matrix, illustrated in equation 7 where each column $\mathbf{x}_j$ represent the vector of observation of parameter j ,$j \in \{1, 2, \ldots, 7\}$ for each day.

$$\mathbf{X}_{\text{common}} = \begin{bmatrix} x_{11} & x_{12} & \cdots \\ x_{21} & x_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{7}$$

The covariance matrix for each dl-dataset, $\mathbf{X}_{\text{dl}_i}$, is a $n_i \times 7$ - matrix, where $i = (1, 2, 3, 4)$ represent the danger level, and n_i is the total number of days with each danger level, which can be seen in Table 2. As for the common dataset, each column in $\mathbf{X}_{\text{dl}_i}$ represent the daily observations for each explanatory variable.

4.2 Model fitting

We will provide more details on the process of fitting each model. Still, to avoid extensive repetition, we first present the general structure that is common to all models.

4.3 Model evaluation

All models are validated using LOSO-CV. This has been done for both the common dataset and each danger-level dataset. This applies to all models except climatology, where only the danger-level datasets have been used but with LOSO-CV.

The models have been evaluated by their BS. The BS is calculated from the `brier()$bs` function from the `modelMetrics` (Hunt, 2020) package and BSS relative to the simplest version of the climatology, using the BS and base `R` functions. The BSs and the BSSs have been calculated for each year, each danger level, and the complete set. The standard deviation for the BSs over the years for each method has also been calculated.

Further, is the reliability diagram with the 95% confidence interval for each model and the corresponding sharpness diagram plotted.

4.4 The Models

4.4.1 Climatology

In addition to the regular climatology described in Chapter 2.5.1, has a LOSO-CV climatology based on dl-datasets been constructed. This is for the remainder of the project, referred to as climatology, and it must not be confused with regular climatology. The climatology has been made in the same way as the other models by using LOSO-CV on the dl-datasets. To do this, base `r` functions have been used.

4.4.2 GLM

Five models based on GLM have been fitted. The four dl models have been combined, leaving us with two models when evaluating. The models have been fitted using all explanatory variables and a log link function. The response variable has been the daily avalanche count.

Based on the results from Sigbjørnsen (2024), there have not been used interaction terms between the explanatory variables and linear parameters. The resulting systematic component is shown in Equation 8

$$\eta_i^m = \beta_1^m x_{i1} + \beta_2^m x_{i2} + \ldots + \beta_7^m x_{i7}. \tag{8}$$

Here, $\beta_l^m, l \in (1,2,3,4,5,6,7)$ is the regression parameters and $x_{i1}, x_{i2}, \ldots, x_{i7} i \in \{1, n\}$ are the explanatory variables, m indicates the dataset ($m \in \{\text{common}, dl_j\}, j = 1,2,3,4$ and n is the number of observations in each dataset.

The model is fitted using the `glm()` function, realized with the `predict()` function, and

processed using `dpois()`, all from base `R`. This gives us a model that predicts the same binary event as the other models.

4.4.3 Simple Tree

Five simple tree models are fitted, four of which are fitted on the datasets divided by danger level and later combined, giving two final models to evaluate. The simple binary decision trees use the sum of squared errors (SSE) as a splitting criterion. Hence, it selects the split that minimizes the SSE. As the mean of squared errors (MSE) for a binary variable is the BS, a splitting decision based on the SSE corresponds to a splitting criterion based on the best possible BS.

The trees have been fitted with the `rpart()` function from the rpart package in `R` (rpart). The response variable used when fitting the trees is a binary response that is 0 if no avalanche has been released and 1 if one or more avalanches have been released.

4.4.4 Random forest

Five RF-models are fitted, four if which are fitted independently on the dl datasets, and then combined, giving two final rf models to evaluate. The models are fitted using the function `randomForest` from the library randomForest (randomForest). The number of explanatoy variables considered at each split is 2, which is the default value when having 7 parameters. The default is `p`/3 where p is the number of explanatory variables (randomForest). Each model grows 1000 trees.

4.4.5 Method of nearest neighbors

Five kNN are fitted, four of which are fitted independently on the dl datasets and later assembled, giving two final models to evaluate. For the common dataset, k is decided by plotting the Brier score for all values of k from 1:40 and choosing the one where the curve flattens. The best model is selected using the default summary from the trainControl function in the caret package. The mean squared error and the R-squared are then used to choose the best model. Using the resulting model, the predictions for the year taken out are calculated. The kNN default summary uses the SSE and R^2 to evaluate the best model when training. The value for K was chosen separately for the different datasets but remained the same for all years. This is done by iterating through different values for k for

the entire dataset (common and dl-s) (not divided per year) and then calculating the BS for each k-value. The k-value where the BS curve flattens has then been chosen.

5 Results

First, the results for each model are presented, along with some evaluation results. Then, the remaining evaluation results of the models are presented together. After this, the results for two months, February 2019 and April 2021, are presented to illustrate the models.

5.1 GLM-based model

The GLM model, described in Chapter 4.4.2, is fitted with the data presented in Chapter 3.2 for all datasets described in Chapter 4.1. The resulting fitted models can be utilized to predict the probability of avalanches, described in Chapter 2.1. The BS, described in Chapter 2.4.1, is calculated for each fold, as explained in Chapter 2.2. Further, the BSS described in chapter 2.4.2 and the Reliability and sharpness diagram explained in Chapter 2.4.3 for each fitted model are calculated and constructed.

For the common model, the interpolated and observed snow depth, the new snow depth, and the wind parameter show consistently high statistical significance. For the dl1-model, no parameters have a p-value below 0.05, and for the remaining dl models, the temperature, wind intensity, and new snow depth are the dominant parameters.

The resulting predictions, along with the climatology, are shown in Figure 17 for the 2017-2018 season and Figure 18 for the 2019-2020 season. These seasons correspond to the best and worst BS, respectively, for the common model. For the dl-model, predictions, and climatology for the 2019-2020 season are shown in Figure 19, while Figure A12 presents the results for the 2023-2024 season, representing the worst and best BS, respectively. Predictions and climatology for the remaining seasons, for both the common and dl-models, are provided in Appendix A. It is noticeable that the probabilities decrease towards the end of each season for both the common and dl-models across all years.

The reliability plots are presented in Figure 21 and sharpness- diagrams in Figure 22. Figure 21a shows that the common model is reliable for low probabilities, consistently overestimates the avalanche probability for higher levels, and gets less reliable for higher predictions. The uncertainty of the reliability increases slightly with the predicted probability. The sharpness diagram, shown in Figure 22a, shows a higher frequency of low predictions, e.g., that the model predicts a low probability for an avalanche release more often than a high probability. It also shows an increased frequency similar to regular climatology, suggesting the model lacks significant forecast information. Figure 21b shows that the dl-model is better and slightly more reliable for low probabilities but loses skill for

higher predictions. The sharpness diagram, shown in Figure 22b, does, however, indicate a higher sharpness, indicating that it might be some forecasting value despite the lack of reliability.

The BSs of the models for each year are presented in table 4. From Table 4, one sees the season with the worst BS is the 2019-2020 season for both models and that the season with the best BS is the 2017-2018 season for the common model and the 2023-2024 season for the dl-model. Both models have a large spread in BS, indicating that they are significantly affected by seasonal variations.

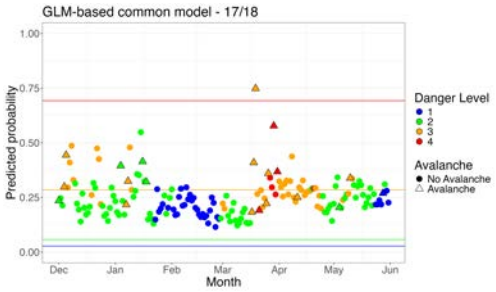


Figure 17: GLM-common model predictions for the 2017-2018 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

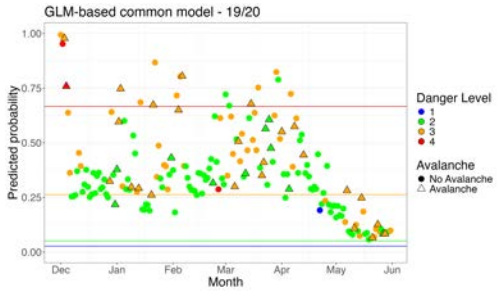


Figure 18: GLM-common model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

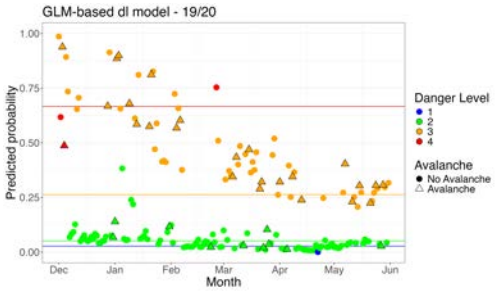


Figure 19: GLM-dl model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

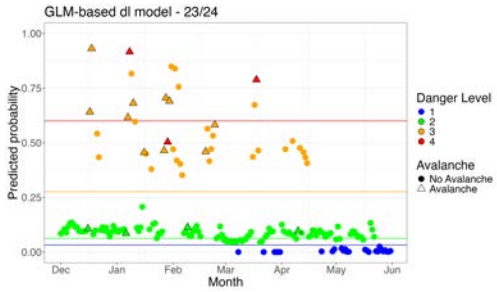
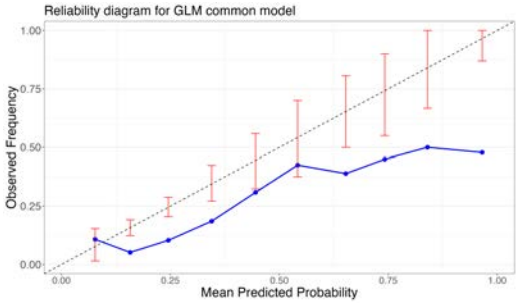
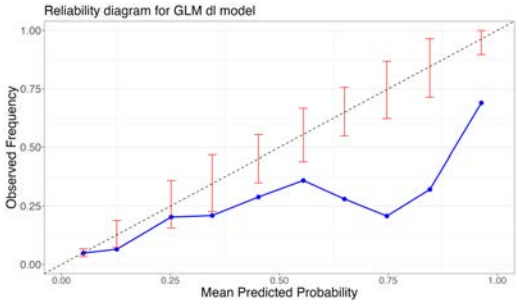


Figure 20: GLM-dl model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

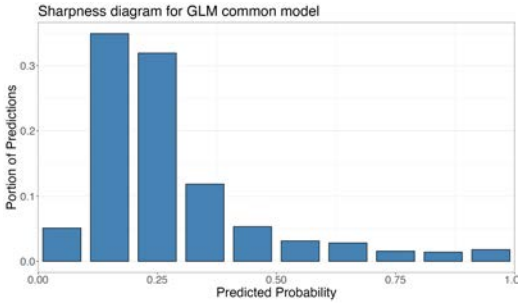


(a) Common model

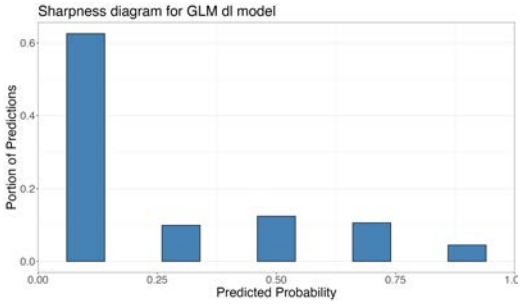


(b) DL-model

Figure 21: Reliability diagrams for the GLM-based common-model and dl-model.



(a) Common model



(b) dl-model

Figure 22: Sharpness diagrams for the GLM-based common-model and dl-model.

Season	17/18	18/19	19/20	20/21	21/22	22/23	23/24	total
Common	**0.1069**	0.1085	0.2046	0.1257	0.1256	0.1393	0.1094	0.1316
dl	0.1535	0.1280	0.1636	0.1634	0.1213	0.1609	**0.0762**	0.1382

Table 4: BSs for the GLM-models. The red number highlights the BS for the worst-performing fold, and the bold number highlights the best.

5.2 Simple Tree

The Simple Tree model, described in Chapter 4.4.3 is fitted with the data presented in Chapter 3.2, for all datasets described in Chapter 4.1. This gives parameters and a fitted model that can be utilized to predict the probability of avalanches, described in Chapter 2.1. The BS, described in Chapter 2.4.1, is calculated for each fold, as explained in Chapter 2.2. Further, the BSS, described in chapter 2.4.2, and the reliability and sharpness diagram, explained in Chapter 2.4.3, for each fitted model are calculated and constructed.

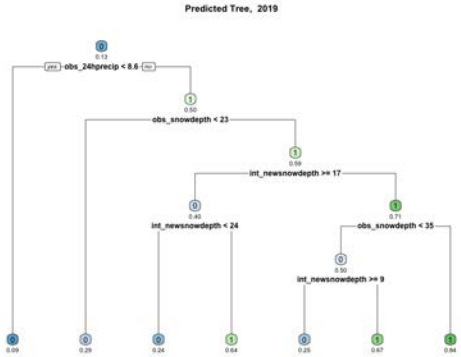
For the common model, the observed and interpolated 24 hourly accumulated precipitation are ranked as the most important variables. For dl1, no variables show any importance. For the remaining dls, new snow depth and precipitation rank as the variables with the highest importance. Figure 23a, and 23b shows the resulting trees for the model fit to the 2019-2020 and 2023-2024 seasons respectively. These illustrate that the variable's importance rankings differ between the folds.

The resulting predictions, along with the climatology, are shown in Figure 24 for the 2019-2020 season and Figure 25 for the 2023-2024 season. These seasons correspond to the worst and best BS, respectively, for the common model. For the dl-model, the predictions and climatology for the 2019-2020 season are shown in Figure 26, while Figure 27 presents the results for the 2023-2024 season, representing the worst and best BS, respectively. Predictions and climatology for the remaining seasons, for both the common and dl-models, are provided in Appendix A. Unlike the GLM models, the tree models do not exhibit a decreasing trend in predictions toward the end of each season.

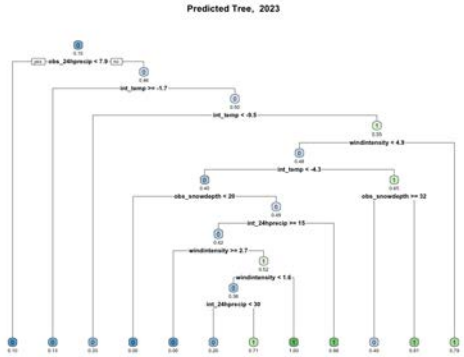
The reliability plots are presented in Figure 28 and sharpness-diagrams in Figure 29. Figure 28a indicates that the model is unreliable, as it fluctuates up and down and does not follow the ideal shape for a reliable forecaster. The sharpness diagram, shown in Figure 29a, shows a higher frequency of low predictions and a sharp model but also that the model does not cover the entire specter of probabilities. Figure 28b shows that the dl-model is more reliable for low probabilities than the common model but rapidly loses skill for higher predictions. The sharpness diagram, shown in Figure 29b, does indicate slightly less sharpness compared to the common model, but also that it covers the entire range of probabilities more evenly.

The BS for the models for each year are presented in table 5. As one can see, the BS for the 2019-2020 season is higher than that for the remaining seasons. The dl-model also scores better than the common model for most years and in total.

The BSs for the models for each year are presented in table 5. From Table 5, one sees the season with the worst BS is the 2019-2020 season for both models and that the season with the best BS is the 2023-2024 season. Both models have a large spread in BS, indicating that they are significantly affected by seasonal variations. It is also not consistent which model has the best BS.



(a) Tree for season 2019-2020, common-model



(b) Tree for season 2023-2024, common-model

Figure 23: Example of decision trees for two years.

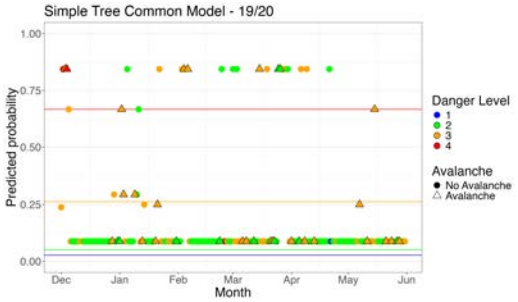


Figure 24: Common tree-model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

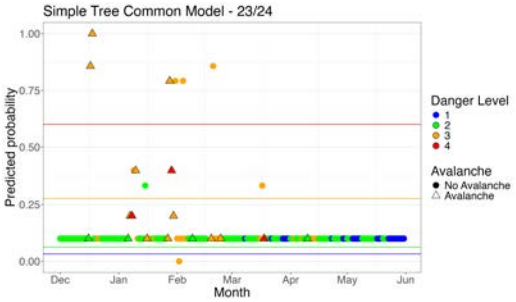


Figure 25: Common tree-model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

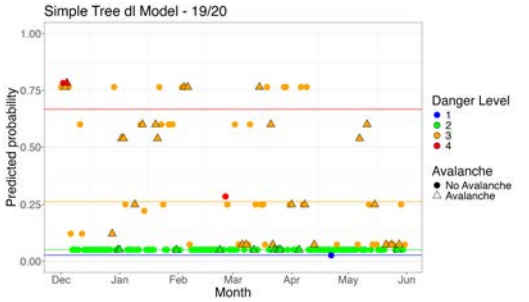


Figure 26: dl-Tree model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

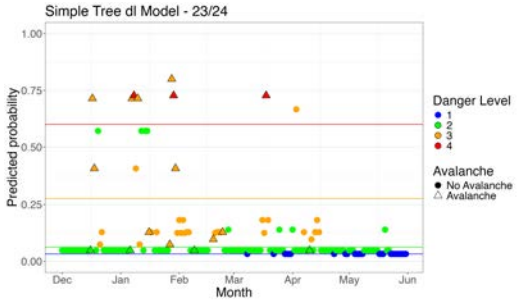
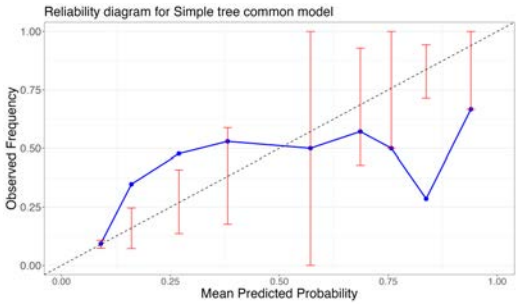
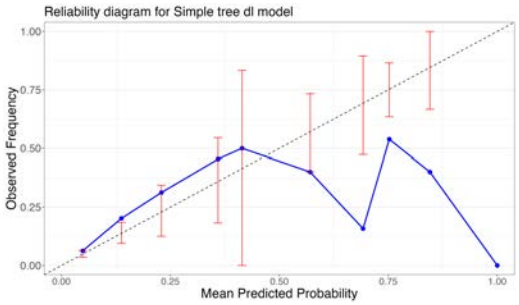


Figure 27: dl-Tree model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

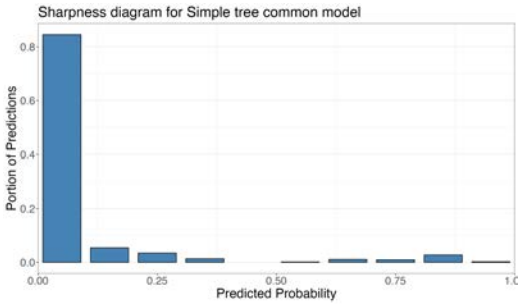


(a) Common model

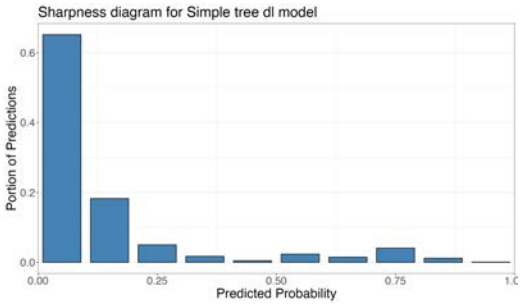


(b) dl-model

Figure 28: Reliability diagrams for the common Tree model and dl-Tree model.



(a) Common model



(b) dl-model

Figure 29: Sharpness diagrams for the common Tree model and dl-Tree model.

The BS for each season is given in 5. The BS for the 2019-2020 season is worse than all other years.

Season	17/18	18/19	19/20	20/21	21/22	22/23	23/24	total
common	0.0871	0.1282	0.1877	0.1121	0.1398	0.1040	**0.0774**	0.1196
dl-data	0.0890	0.1228	0.1685	0.1068	0.1580	0.1078	**0.0596**	0.1162

Table 5: BSs for the Tree-models. The red number highlights the BS for the worst- performing fold, and the bold number highlights the best.

5.3 Random Forest

The RF models, described in Chapter 4.4.4 is fitted with the data presented in Chapter 3.2, for all datasets described in Chapter 4.1. This gives parameters and a fitted model that can be utilized to predict the probability of avalanches, described in Chapter 2.1. The BS, described in Chapter 2.4.1, is calculated for each fold, as explained in Chapter 2.2. Further, the BSS, described in chapter 2.4.2, and the reliability and sharpness diagram, explained in Chapter 2.4.3, for each fitted model are calculated and constructed.

Precipitation variables are the most important predictors of avalanches across all models. Both 24-hour interpolated and observed precipitation are of large importance. The interpolated snow variables are moderately important, and the observed snow depth ranks even lower. Temperature has a moderate influence, while the wind parameter has low importance. Variable importance differs between danger levels. For example, in dl4, temperature and new snow are the most important predictors, while precipitation plays a minor role.

The resulting predictions, along with the climatology, are shown in Figure 30 for the 2019-2020 season and Figure 31 for the 2023-2024 season. These seasons correspond to the worst and best BS, respectively, for the common model. Similarly, the predictions for the dl-model, along with the climatology, are shown in Figure 32 for the 2019-2020 season and Figure 33 for the 2023-2024 season, representing the worst and best BS for the dl-model, respectively. The predictions and climatology for the remaining seasons, for both the common and dl-model, are provided in Appendix A. Overall, the plots do not reveal any major differences between the common and dl-models, apart from a slightly larger spread within each danger level for the common model.

The reliability plots are presented in Figure 34 and sharpness- diagrams in Figure 35. Figure 34a indicates that the model is reliable for relatively low predictions and loses reliability when the predicted probability exceeds 60%. The sharpnes diagram, shown in Figure 35a shows a higher frequency of low predictions, and indicate a sharp model, and an even decrease in the prediction frequency with increasing predictions. Figure 34b shows that the dl-model is relatively reliable for all predicted values, despite a minor fluctuation,

as it follows the development of a perfectly reliable forecaster. The sharpness diagram, shown in Figure 35b, also indicates a higher sharpness for the dl-model than the common model.

The BSs of the models for each year are presented in table 6. The BS for the 2019-2020 season is higher than for the remaining seasons, and the 2023-2024 season has the lowest BS for both models. Although a higher total BS, the dl-model seems to have more desirable forecaster skills than the common model.

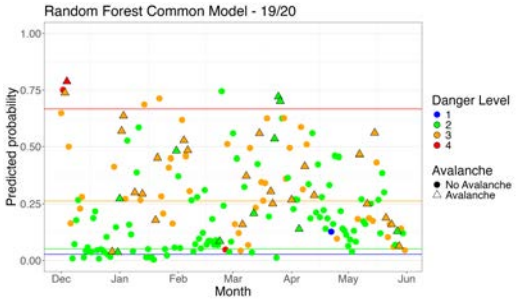


Figure 30: Common-RF model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

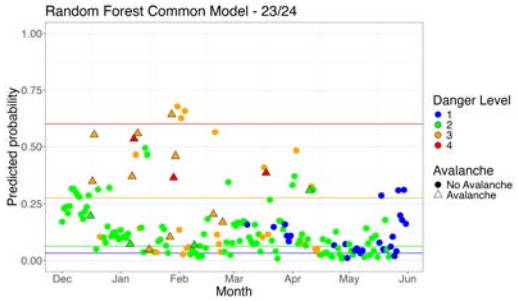


Figure 31: Common-RF model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

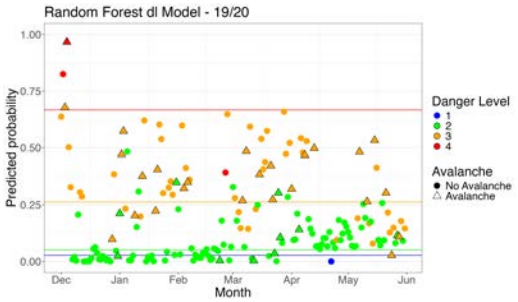


Figure 32: dl-RF model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

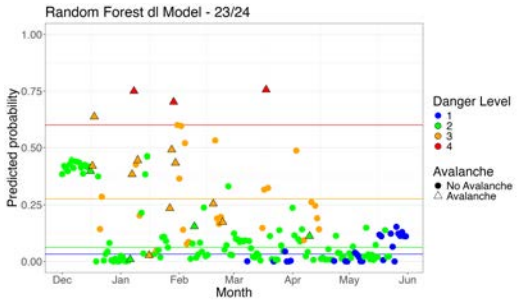
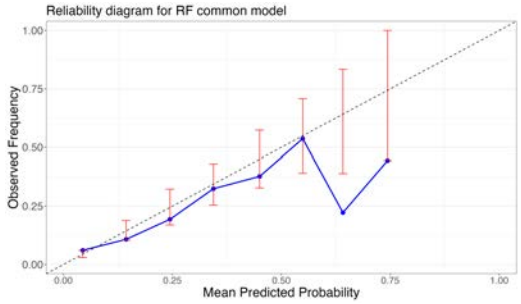
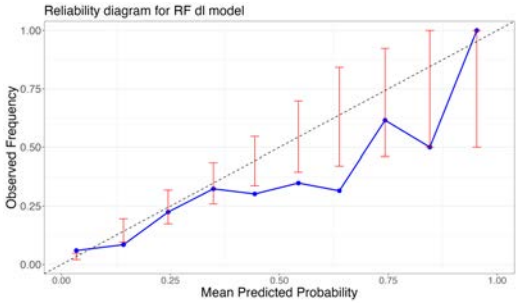


Figure 33: dl-RF model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

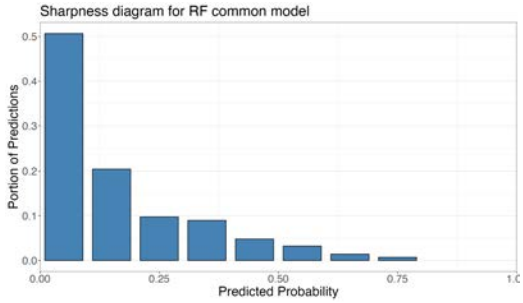


(a) Common model

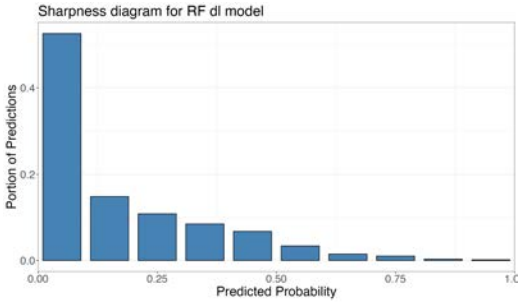


(b) dl-model

Figure 34: Reliability diagrams for the common RF model and dl-RF model.



(a) Common model



(b) dl-model

Figure 35: Sharpness diagrams for the common RF model and dl-RF model.

Season	17/18	18/19	19/20	20/21	21/22	22/23	23/24	total
common	0.0830	0.1029	0.1574	0.1163	0.1276	0.1116	**0.0803**	0.1114
dl-data	0.0904	0.1095	0.1525	0.1146	0.1361	0.1058	**0.0776**	0.1124

Table 6: BSs for the RF-models.he red number highlights the BS for the worst- performing fold, and the bold number highlights the best.

5.4 KNN

The KNN models, described in Chapter 4.4.5 is fitted with the data presented in Chapter 3.2, for all datasets described in Chapter 4.1. This gives parameters and a fitted model that can be utilized to predict the probability of avalanches, described in Chapter 2.1. The BS, introduced in Chapter 2.4.1, is calculated for each fold, as explained in Chapter 2.2. Furthermore, the BSS (Chapter 2.4.2), as well as the reliability and sharpness diagrams

(Chapter 2.4.3), are calculated and constructed for each fitted model. The chosen values for K, the number of neighbors, for each dataset are given in Table 7.

The resulting predictions, along with the climatology, are presented in Figure 36 for the 2019-2020 season and Figure 37 for the 2023-2024 season. These seasons correspond to the worst and best BS, respectively, for the common model. For the dl-model, predictions, and climatology for the 2019-2020 season are shown in Figure 38, while Figure 39 displays the results for the 2023-2024 season, representing the worst and best BS, respectively. Predictions and climatology for the remaining seasons, for both the common and dl-models, are provided in Appendix A. Overall, the plots show no major differences between the common and dl-models, except for a slightly larger spread within each danger level in the common model.

The reliability plots are presented in Figure 40 and sharpness- diagrams in Figure 41. Figure 40a indicates that the common model is reliable for relatively low predictions but with an increasing uncertainty for predictions over 50%. The sharpness diagram, shown in Figure 41a, shows a higher frequency of low predictions and a relatively high portion of predictions coinciding with the regular climatology. Figure 40b shows that the dl-model is unreliable for all predicted values above 25%. The sharpness diagram, shown in Figure 41b, indicates a higher sharpness for the dl-model than the common model.

The BSs of the models for each year are presented in table 8. The BS for the 2019-2020 season is higher than for the remaining seasons, and the 2023-2024 season has the lowest BS for both models.

Data	common	dl1	dl2	dl3	dl4
K-value	16	3	23	6	3

Table 7: K-values for each dataset

The reliability plots are shown in Figure 40, and sharpness diagrams in Figure 41.

Figure 40a shows that the common model is close to the ideal reliability line for low probabilities. It also seems unable to identify subgroups of parameter values that give a higher probability than about 60%. The sharpnes diagram, shown in Figure 41a explaines this as there are a higher frequency of low precictions which make sense as ther are a majority of days with no avalanches, however is it noticable less sharp than previous models. Figure 40b shows that the dl-model has a larger prediction range but seems to overestimate the release probabilities and generally lies below the reliability line. The sharpness diagram, Figure 41b, also shows a high portion of predictions for low probabilities and seems sharper

than the common model.

The BS for each season is shown in Table 8. This indicates that the model performs the worst in the 2019-2020 season, closely followed by the 2021-2022 season, while the 2023/2024 season has the best score. The common model scores best for all but one year, which corresponds well with the reliability plot.

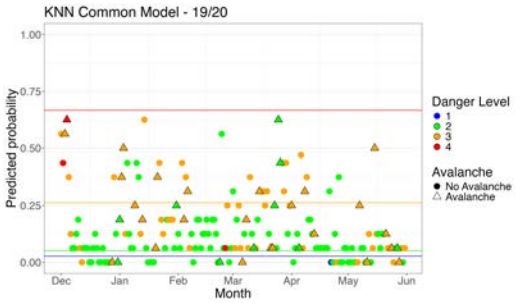


Figure 36: Common-KNN model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

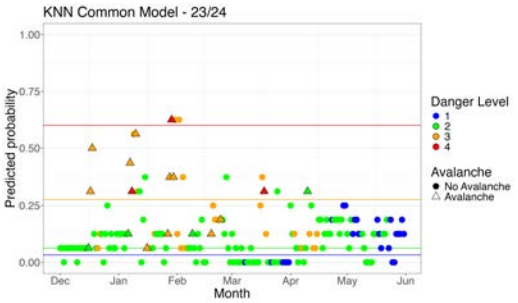


Figure 37: Common-KNN model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

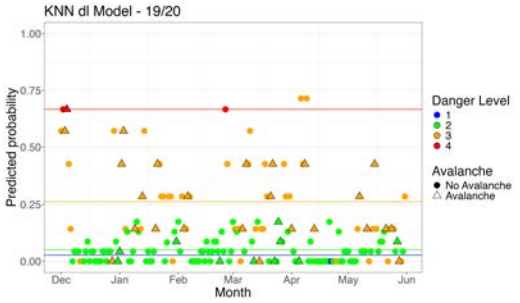


Figure 38: dl-KNN model predictions for the 2019-2020 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

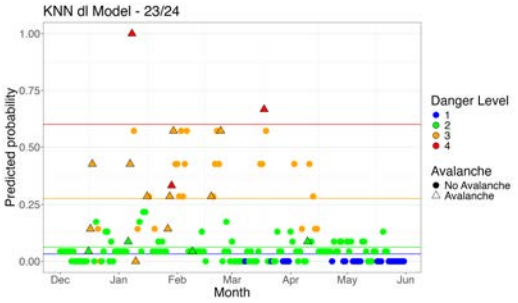
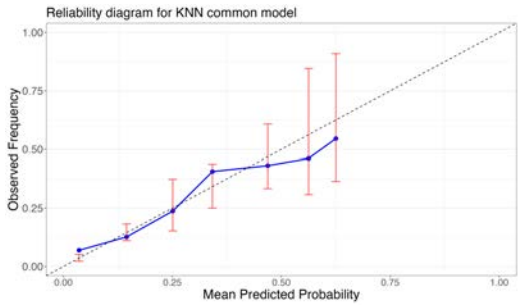
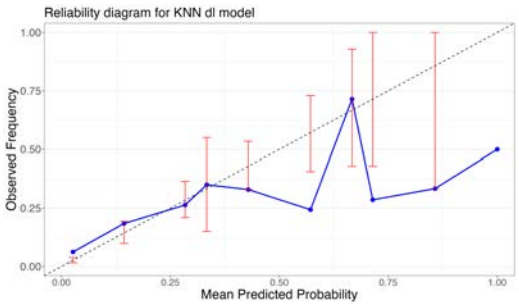


Figure 39: dl-KNN model predictions for the 2023-2024 season. The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9.

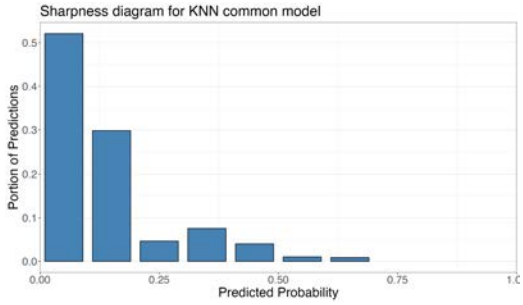


(a) Common model

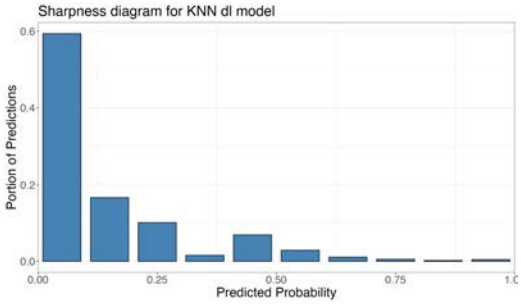


(b) dl-model

Figure 40: Reliability diagrams for the common KNN model and dl-KNN model.



(a) Common model



(b) DL-model

Figure 41: Sharpness diagrams for the common KNN model and dl-KNN model.

Season	17/18	18/19	19/20	20/21	21/22	22/23	23/24	total
common	0.0910	0.1081	0.1538	0.1096	0.1301	0.0988	**0.0723**	0.1092
dl-data	0.1014	0.1126	0.1597	0.1117	0.1414	0.1159	**0.0743**	0.1168

Table 8: BSs for the KNN-models. The red number is the worst BS for each model, and the bold number is the best.

5.5 Climatology

The climatology model, described in Chapter 4.4.1, is fitted. The Brier Score (BS), introduced in Chapter 2.4.1, is calculated for each fold following the procedure outlined in Chapter 2.2. Additionally, the Brier Skill Score (BSS) and the reliability and sharpness diagrams, discussed in Chapters 2.4.2 and 2.4.3, respectively, are computed for each fitted model. The resulting probabilities of an avalanche event for each danger level across all years are presented in Table 9.

The reliability diagram is shown in Figure 42. The regular climatology, presented in Chapter 2.5.1, is perfectly reliable. The climatology model indicates signs of non-reliability, which might indicate an inconsistency in the predicted danger level between the different seasons. The sharpness diagram does not show any sign of sharpness; hence, the climatology does not contribute to any forecasting value, which is sensible as it is constructed without any information apart from the observations of avalanches. Table 10 presents the BS for the climatology. This shows that the 2019-2020 season also has the least accurate predictions for climatology and that the 2023-2024 season has the best BS.

-	17/18	18/19	19/20	20/21	21/22	22/23	23/24
1	0.0273	0.0333	0.0270	0.0093	0.0282	0.0278	0.0325
2	0.0572	0.0577	0.0507	0.0621	0.0526	0.0591	0.0622
3	0.2846	0.2656	0.2626	0.2765	0.2779	0.2909	0.2762
4	0.6923	0.6552	0.6667	0.6207	0.6071	0.6154	0.6000

Table 9: Climatology separated on danger level and cross-validated per season.

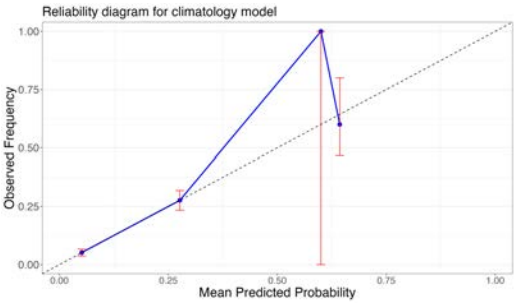


Figure 42: Reliability diagram for climatology model.

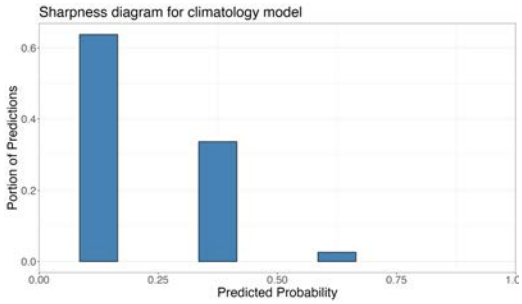


Figure 43: Sharpness diagram for climatology model.

Season	17/18	18/19	19/20	20/21	21/22	22/23	23/24	total
Brier	0.0920	0.1067	0.1443	0.1031	0.1305	0.0950	**0.0648**	0.1050

Table 10: BSs for the Climatology-model. The red number highlights the BS, and the bold number highlights the best.

5.6 Comparison of the models

Table 11 presents the BSs for the models trained without the danger level and for each danger level. Table 12 presents the BSs for the models trained on each danger level sepa-

rately.

danger level	1	2	3	4	total
GLM	0.0571	0.0901	0.2147	**0.2237**	0.1317
Tree	**0.0303**	0.0618	0.2261	0.3038	0.1168
Random forest	0.0340	0.0635	**0.2016**	0.2568	0.1114
KNN	0.0313	**0.0585**	**0.2016**	0.2809	**0.1092**
Regular climatology	0.0395	0.0613	0.2179	0.4740	0.1221

Table 11: BSs for models trained on the common dataset. The red number highlights the worst-performing model for each danger level, and the bold number highlights the best.

danger level	1	2	3	4	total
GLM	0.0472	0.0572	0.2884	0.2329	0.1382
Tree	**0.0268**	0.0579	0.2289	0.2309	0.1162
RF	0.0309	0.0615	0.2128	**0.2032**	0.1124
KNN	0.0328	0.0565	0.2263	0.2895	0.1168
Climatology	**0.0268**	**0.0543**	**0.2011**	0.2448	**0.1050**

Table 12: BSs for models trained on dl-dataset. The red number highlights the worst-performing model for each danger level, and the bold number highlights the best.

The Brier Score (BS) for the 2019-2020 season is the highest across all models, while the 2023-2024 season achieves the best BS for most models. This suggests that the climatic conditions vary between seasons, with some seasons presenting a more complex climatic situation. The standard deviation of the BS for all models is shown in Table 13. This shows that the Tree models have the highest standard deviation and, thus, are the least robust towards seasonal changes and that the KNN model has the lowest standard deviation for the common models. In contrast, the RF model has the lowest standard deviation for the dl models.

Method	GLM	Tree	RF	KNN	Climatology
Common	0.0344	0.0371	0.0266	0.0246	-
dl	0.0323	0.0379	0.0256	0.0275	0.0261

Table 13: Standard deviations for BSs.

The Brier skill score (BSS) is calculated using the simple climatology (average over all days) as a reference forecaster. The brier score for the reference is given in the last row in

Table 11. The BSS for the models trained without knowing the danger level is presented in Table 14 and for the models trained on the danger levels in Table 15. The tables show that all models are better than the simple climatology for danger level 4, which is important as that is the danger level with the highest occurrence of avalanches and also the danger level where it's most likely that the road section will be closed.

Danger level	dl1	dl2	dl3	dl4
GLM	-0.4459	-0.4685	0.0145	**0.5293**
Tree	**0.2327**	-0.0082	-0.0377	0.3610
Random forest	0.1397	-0.0348	**0.0750**	0.4598
KNN	0.2077	**0.0457**	0.0749	0.4090
Regular climatology	0	0	0	0

Table 14: BSSs for common models. The reference forecaster used is regular climatology.The red number highlights the worst-performing model for each danger level, and the bold number highlights the best.

Danger level	dl1	dl2	dl3	dl4
GLM	-0.1980	0.0681	-0.3234	0.5100
Tree	**0.3214**	0.0553	-0.0502	0.5143
Random forest	0.2175	-0.0030	0.0233	**0.5724**
KNN	0.1689	0.0782	-0.0385	0.3909
Climatology	**0.3214**	**0.1143**	**0.0773**	0.4851

Table 15: BSSs for dl-models. The reference forecaster used is regular climatology. The red number highlights the worst-performing model for each danger level, and the bold number highlights the best.

5.7 Case study of two months

Two months, February 2019 and April 2021, have been selected to study the models' performance in more detail.

5.7.1 February 2019

The predicted probabilities for an avalanche release for each day in February 2019 are given in Figure 44 and 45. The avalanche danger for each day is not provided directly in the plot

but can be read from the plot climatology as it follows the danger level. The dl-models seem better tuned for the first two weeks when no avalanches were released, but they also seem to under-predict the probabilities for the first day with an avalanche event, where the common models seem to catch the change in the situation. The BSs for all models are given in Table 16. Here, the GLM-models score the best both among the common models and the dl-models, and the Tree scores the worst. The BSs are overall high.

The common models and the dl models both show the same tendencies, but by closer inspection, the common model shows an interesting feature. On the 12th day of February, the climatology plot indicates that the danger level is 2. Still, all models show an increase in the predictions, and avalanche activity is detected for this day. Two days later, the danger level is raised to 3. On the 20th day, the avalange danger is set down to 2 again, but all models gives a predicts a relatively high probability for avalanche activity, and it does occur, this also goes for day 26. There are indications of the same tendencies when studying the dl-models, but they are not as clear.

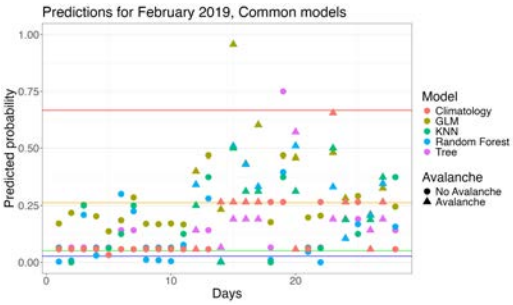


Figure 44: Predicted probabilities for avalanches by the common models for February 2019.

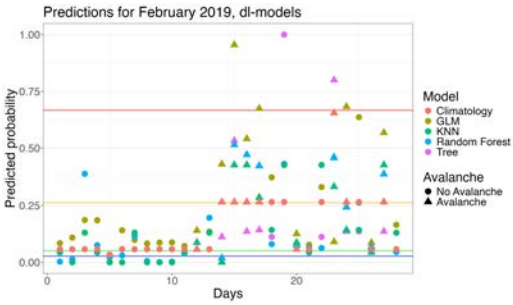


Figure 45: Predicted probabilities for avalanches by the dl-models for February 2019.

Model	GLM	Tree	Random forest	KNN	clima
common	**0.1682**	0.2569	0.1951	0.2077	-
dl	**0.1832**	0.2775	0.2195	0.2405	0.2266

Table 16: BSs for all models for February 2019. The red number highlights the worst-performing method, and the bold number highlights the best for each data group.

5.7.2 April 2021

The predicted probabilities for an avalanche release for each day in April 2021 are given in Figure 46 and 47. As for the plots for February, the danger level for each day is not

directly given but can be read from the plot of the climatology. The common models seem to provide a low overall probability for avalanches, but they all give the highest probability of an avalanche on the only day with danger level 4, where an avalanche actually happened. The Simple Tree changes significantly between the dl-dataset and the common dataset and performs better in the dl-model. The BSs for all models are given in Table 17. Here, the GLM model performs the best among the common models and the climatology among the dl models. The simple Tree performs the worst among the common models, and the GLM model performs the worst among the dl models. The Brier scores are lower overall than in February.

When comparing the common and dl models, the most noticeable difference is the tree models. For the common model, it never forecasts a higher probability than 40%, while the dl-model has several days with a predicted probability over 75%. The common model follows the dangerlevel indicated by the climatology and seems to correctly predict a lower probability for the danger level most days between the 8th and 14th day.

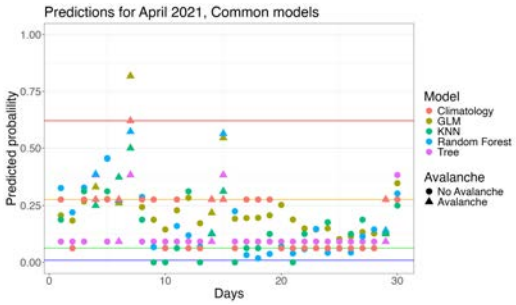


Figure 46: Predicted probabilities for avalanches by the common models for April 2021.

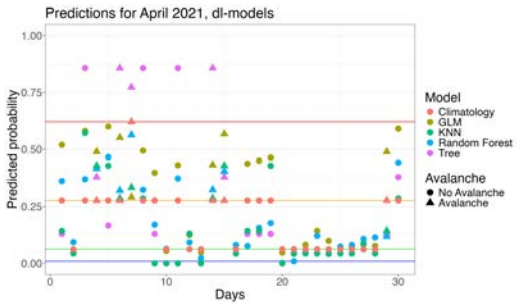


Figure 47: Predicted probabilities for avalanches by the dl models for April 2021.

Model	GLM	Tree	Random forest	KNN	Climatology
common	**0.1200**	0.1318	0.1201	0.1259	-
dl	0.1447	0.1378	0.1228	0.1271	**0.1193**

Table 17: BSs for all models for April 2021. The red number highlights the worst-performing method, and the bold number highlights the best for each data group.

6 Discussion

All models seem to give a relatively wide range of predictions compared to our initial working hypothesis, which was that all probabilities would be close to 0 or, on occasion, extremely high. However, both the common and the DL models seem to be able to detect changes that correspond well with the expected danger level.

Statistical inference on the variables is not prioritized in this project, but we find it useful with a brief comment. For the GLM-based model, the new snow depth, the observed snow depth, wind, and temperature show strong statistical significance. For the tree and random forest models, it is observed and interpolated precipitation, and then new snow depth that has the highest importance, while the other parameters vary. This corresponds well with the result of (Sigbjørnsen, 2024). The wind was, however, not included as a parameter suggestion for that project and might be beneficial to investigate further as it plays an important role in snow distribution.

The common GLM model shows indications of unsharpness, suggesting low forecasting skill, though it somewhat follows the optimal reliability curve. Its performance varies significantly, performing worst for danger levels 1 and 2 but best for danger level 4. The dl GLM model demonstrates greater sharpness but reduced reliability and never scores highest among the dl-models. Both GLM models underpredict avalanche probabilities in late-season months, regardless of actual avalanche activity. Dividing the season into two parts or analyzing by month could address this issue.

The common tree model is sharp but unreliable, with the highest BS standard deviation, indicating seasonal sensitivity. The dl tree model shows slightly better reliability for low predictions but remains inconsistent and non-robust, with the worst standard deviation among dl-models. The RF models outperform others in forecasting characteristics, aligning well with the goal of supporting operational avalanche forecasting. Both models display reliability, sharpness, and robustness with low seasonal variation sensitivity. The dl RF model effectively forecasts high probabilities, further demonstrating its potential for decision-making in avalanche-prone areas.

The common kNN model is relatively reliable, while the dl kNN model is less reliable, possibly due to reduced neighbor availability at higher danger levels. Both models exhibit moderate sharpness, though the common kNN model often predicts values similar to the climatology. The dl KNN model is sharper but could benefit from techniques like repeated CVs or stratification, this is however outside the scope of this project.

As stated in the introduction, danger level 1 corresponds to an up to 10% probability

for avalanches. Under this definition, some days with danger level 2 should likely be downgraded. Further, there is a significant gap between avalanche proportions for danger levels 2 and 3, as well as levels 3 and 4. Danger level 4 implies a $> 70\%$ risk of avalanches, and the climatology for all years ranges between 60% and 69%. Three models, the GLM, Tree, and RF models, outperform climatology for this danger level, with RF-dl showing the strongest results. This highlights the potential of RF models to complement expert forecasts by offering more localized and precise assessments for road closure decisions. A useful forecaster must outperform regular climatology, which lacks forecasting information. Across all danger levels, some models demonstrate better performance than climatology, indicating their ability to provide meaningful forecasting insights.

One working hypothesis was that the information added by including the danger level in the models would overrule other variables. This has not happened to the expected extent. Overall, both the GLM model and the KNN model score better in terms of BS and reliability for the common models, while the Tree and the RF methods perform better for the dl-dataset. One reason for this might be that the dl-models are only trained on observation from the same danger level and thus lose the information lying in possible cases of doubt about the danger level—E.g., when expert forecasters were uncertain about whether the avalanche danger level was 2 or 3.

The observed tendencies of common models during the detailed case studies of February 2019 and April 2021 suggest that these models can detect shifts in avalanche risk before avalanche danger experts recognize such changes. This capability could be critical for operational decision-making, allowing for earlier warnings and more proactive risk management for road safety. Alternatively, these findings may indicate a bias in expert assessments, potentially influenced by recent avalanche activity. In either case, common models could serve as valuable tools for guiding decision-making to determine when further evaluation of conditions is warranted or when maintaining the current danger level may be appropriate.

This study does not account for inaccuracies in radar detections, interpolated and observed weather variables, or changes in the danger level forecasting strategy. Future work could address these limitations by incorporating a more detailed analysis of past weather conditions and exploring additional variables, such as wind redistribution patterns, to refine the models. Such improvements could enhance the operational value of the forecasts. It might also be useful to analyze seasonal changes in the weather if one is going to provide a day-by-day risk measure based purely on the models presented in this project, as different climatic factors dominate in different seasons.

7 Conclusion

This section summarizes the study's findings and provides suggestions for further work.

The results of this study align closely with the findings of Sigbjørnsen (2024), showing that the RF models outperform the other models overall. They exhibit high reliability, sharpness, and robustness across different seasons. Additionally, all models perform significantly better than regular climatology for danger level 4 and, in most cases, for other danger levels. These findings suggest that the models contain valuable forecasting information that can contribute to operational decisions regarding road safety.

To further improve model performance, several avenues can be explored:

- Conducting a more comprehensive statistical inference on the parameters.
- Introducing time-dependency to better account for temporal patterns.
- Comparing the current models with ensemble models to assess potential gains in performance.

Expanding the study to include an analysis of snowpack stability could also provide valuable insights. Furthermore, incorporating the expected size of avalanches as a response variable would be beneficial, as not all avalanches reach the road.

For the authorities responsible for road management, it could be valuable to identify specific avalanche paths that are more exposed and investigate how different climatic situations influence these paths. This could enhance site-specific forecasting and improve mitigation measures.

References

Kebene Amanuel Baissa. Snow avalanche hazard mapping using spatio-temporal bayesian models with inlabru. Master's thesis, Norwegian University of Science and Technology, 2024.

bca. Spring and summer avalanche hazards, 2024. URL https://backcountryaccess.com/no-no/blog/p/spring-and-summer-avalanche-hazards. Accessed 2024-25-09.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324. URL https://link.springer.com/article/10.1023/A:1010933404324.

Othmar Buser. Avalanche forecast with the method of nearest neighbours: An interactive approach. *Cold Regions Science and Technology*, 8(2):155–163, 1983. doi: https://doi.org/10.1016/0165-232X(83)90006-X.

George Casella and Roger L. Berger. *Statistical Inference*. Brooks/cole Cengage Learning, Belmont, CA, USA, 2022.

Robert E. Davis, Kelly Elder, Daniel Howlett, and Eddy Bouzaglou. Relating storm and weather factors to dry slab avalanche activity at alta, utah, and mammoth mountain, california, using classification and regression trees. *Cold Regions Science and Technology*, 30(1–3):79–89, 1999. doi: https://doi.org/10.1016/S0165-232X(99)00032-4.

EAWS. Site-specific avalanche warning, definitions and recommendations, 2022a.

EAWS. Typical avalanche problems. *Avalanche problems, approved by General Assembly of EAWS, Davos, 2022*, 2022b. URL https://www.avalanches.org/wp-content/uploads/2022/09/EN_EAWS_avalanche_problems.pdf.

EAWS. Glossary, 2024. URL https://www.avalanches.org/glossary/. Accessed: 2024-09-14.

EAWS Matrix, 2022. URL https://www.avalanches.org/standards/eaws-matrix/. Accessed: 2024-16-11.

European Centre for Medium Range Weather Forecasts. Section 12.b statistical concepts - probabilistic data, 2018. URL https://confluence.ecmwf.int/display/FUG/Section+12.B+Statistical+Concepts+-+Probabilistic+Data#Section12.BStatisticalConceptsProbabilisticData-BrierSkillScore(BSS). Accessed: 2024-11-18.

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian D. Marx. *Regression: Models, Methods and Applications, 2nd Edition*. Springer Berlin, Heidelberg, Germany, 2022.

Friflyt. Lyngsalpene, 2024. URL `https://www.friflyt.no/topptur/lyngsalpene`. Accessed: 2024-11-15.

Troms Fylke, 2024. URL `https://www.tromsfylke.no/tjenester/samferdsel/skredberedskap/skredvarslingsanlegg/holmbuktura/`. Accessed: 2024-09-09.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151, 2014. doi: https://doi.org/10.1146/annurev-statistics-062713-085831.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268, 2007. doi: https://academic.oup.com/jrsssb/article/69/2/243/7109375.

Robert Trevor Hastie and Jerome Friedman Tibshirani. *The Elements of Statistical Learning*. Springer New York, NY, 2009. doi: https://doi.org/10.1007/978-0-387-84858-7.

Anders Asheim Hennum. Data-driven avalanche forecasting - using automatic weather stations to build a data-driven decision support system for avalanche forecasting. Master's thesis, Norwegian University of Science and Technology (NTNU), 2016. URL `https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2392436`. Accessed: 2024-16-12.

Tyler Hunt. *ModelMetrics: Rapid Calculation of Model Metrics*, 2020. URL `https://CRAN.R-project.org/package=ModelMetrics`. Accessed: 2024-12-06.

Christian Jaedicke, Fabiano Monti, Paola Dellavedova, Lukas Stoffel, Sergio Azzeralo, and Antoni Molné. Appliation and evaluation of the EAWS guidelines for site-specific avalanche warning. In *Proceedings of the International Snow Science Workshop*, pages 7–11, 2024.

Ian T. Jolliffe and David B. Stephenson. *Forecast Verification: A Practitioners Guide in Atmospheric Science*. John Wiley I& Sons Ltd, 2011. URL `http://danida.vnu.edu.vn/cpis/files/Books/Forecast%20Verification%20-%20A%20Practitioners%20Guide%20in%20Atmospheric%20Science.pdf`.

Kaja Krogh, Mariia Pihlainen, Martin Stefan, and Arne Kavli Berg. Avalanche operation for road protection in Grøtfjord, Tromsø. In *Proceedings of the International Snow Science Workshop*, pages 828–832, 2024.

Kristin Brandtsegg Lome, Hallvard Skaare Nordbrøden, and Nils Arne Kavli Walberg. Site specific awarning Pollfjellet AKA T̈he Avalanche Machine; the significance of meteorological analysis. In *Proceedings of the International Snow Science Workshop*, pages 1–6, 2024.

Simon J. Mason. On using "climatology" as a reference strategy in the brier and ranked probability skill scores. *Monthly weather review*, 132:1891–1895, 2004.

D. M. McLung. The Elements of Applied Avalanche Forecasting Part II: The Physical Issues and the Rules of Applied Avalanche Forecasting. *Natural Hazards*, 26:131–146, 2007. doi: https://link.springer.com/article/10.1023/A:1015604600361.

L. Meier. Radarmålinger av snøskred ved fv. 293 holmbuktura: Resultater fra testmålinger i 2017 og 2018. Report, Statens vegvesen, July 5 2018. URL `https://vegvesen.brage.unit.no/vegvesen-xmlui/handle/11250/2672819`. Accessed 2024-01-12.

Christoph Mitterer, Gebhart Walter, Harald Riedl, Hermann Brugger, Johann Seiwald, Jürg Schweizer, Michael Winkler, Patrick Nairz, Paul Dobesberger, Paul Kössler, Paul Mair, Peter Paal, Peter Plattner, Robert Horntrich, Rudi Mair, Sigfried Sauermoser, Simon Rauch, Walter Würtel, Walter Zörer, and Werner Beikircher. *Ausbildungshandbuch der Tiroler Lawinekommissionen*. Amt der Tiroler Landerregierung, Innsbruck, Tirol, 2016.

Christer Lundberg Nes. *Skikompis*. Friflyt, 2018.

Norsk klimaservicesenter. Stasjonsiformasjon, 2024. URL `https://seklima.met.no/stations/`.

NPRA. Skredvarsling, 2024. URL `https://www.vegvesen.no/fag/teknologi/geofag/skred/skredvarsling/`. Accessed: 2024-11-15.

NVE. Api.nve.gts, 2012. URL `https://api.nve.no/doc/gridtimeseries-data-gts/#multipointtimeseries`.

Knut Inge Orset and Martine Holm Frekhaug. NATURAL HAZARD PREPAREDNESS IN THE NORWEGIAN PUBLIC ROADS ADMINISTRATION (NPRA). In *Proceedings of the International Snow Science Workshop*, pages 252–256, 2024. doi: https://arc.lib.montana.edu/snow-science/objects/ISSW2024_P1.29.pdf.

ortovox. Avalanche basics, 2024. URL `https://www.ortovox.com/uk/safety-academy-lab-snow/01-avalanche-basics/avalanche-knowledge`. Accessed 2024-25-09.

randomForest. *randomForest: Breiman and Cutlers Random Forests for Classification and Regression*, 2024. URL `https://cran.r-project.org/web/packages/randomForest/index.html`. Accessed: 2024-12-06.

rpart. *Recursive Partitioning and Regression Trees*, 2023. URL https://cran.r-project.org/web/packages/rpart/rpart.pdf. Accessed: 2024-12-06.

senorge.no, 2024. URL https://www.senorge.no/. Accessed: 2024-05-09.

Benjamin Sigbjørnsen. Site-specific probabilistic forecast for avalanches. Master's thesis, Norwegian University of Science and Technology, 2024.

SSB. Befolkning, etter grunnkrets, statistikkvariabel og år, 2024. URL https://www.ssb.no/statbank/table/04317/tableViewLayout1/. Accessed: 2024-11-15.

The MET Office. Reliability and sharpness diagrams, 2024. URL https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/interpret-reliability. Accessed: 2024-17-12.

Scott Thumlert. The four key components of the physical hazard posed by snow avalanches across terrain. In *Proceedings of the International Snow Science Workshop*, pages 77–81, 2024.

Gerhard Tutz. *Regression for Categorical Data*. Cambridge University Press, 2012.

United Nations. The 17 goals, 2024. URL https://sdgs.un.org/goals. Accessed: 2024-11-15.

Varsom.no. Slik lager vi snøskredvarsel, 2024a. URL https://www.varsom.no/snoskred/snoskredvarsling/slik-lager-vi-snoskredvarsel/. Accessed: 2024-11-15.

Varsom.no. Skredvarsling, 2024b. URL https://www.varsom.no/snoskred/snoskredvarsling/om-snoskredvarslingen/. Accessed: 2024-11-15.

vegdata.no, 2024. URL https://vegkart.atlas.vegvesen.no/#kartlag:geodata/@684656,7715985,13/hva:hva%5B0%5D%5BabsoluteIntervals%5D=false&hva%5B0%5D%5Bid%5D=883. Accessed: 2024-08-09.

Wikipedia. Fylkesvei 7900, 2024a. URL https://no.wikipedia.org/wiki/Fylkesvei_7900. Accessed: 2024-11-15.

Wikipedia. Forecasting, 2024b. URL https://en.wikipedia.org/wiki/Forecasting. Accessed: 2024-10-13.

xgeo.no, 2024. URL https://www.xgeo.no/index.html?p=fag. Accessed: 2024-10-09.

A Appendix

Background

The color scale for the steepness maps presented in Figure 3 is presented here:

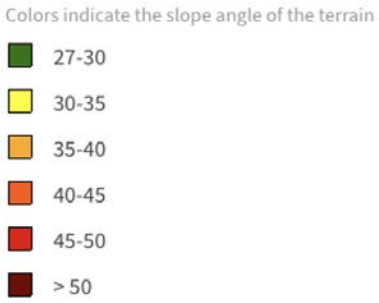


Figure A1: figure
Color scale for Figure 3a.

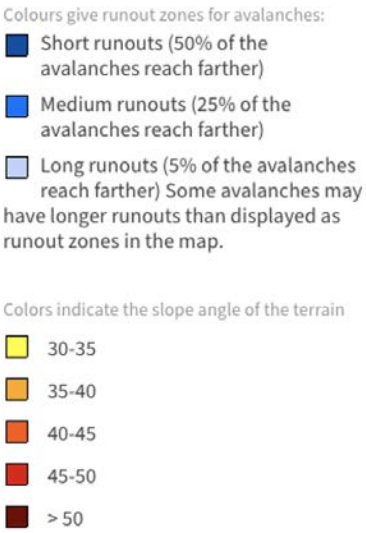


Figure A2: figure
Color scale for Figure 3b.

Data

Group	Subset Definition	Description
Group 1	All data points with danger level 1	Days where danger level = 1
Group 2	All data points with danger level 2	Days where danger level = 2
Group 3	All data points with danger level 3	Days where danger level = 3
Group 4	All data points with danger level 4	Days where danger level = 4

Table A1: Division of the dataset by danger levels.

Danger Level	February 2019	April 2021
1	1	-
2	16	14
3	10	15
4	1	1
5	-	-

Table A2: Frequency of danger levels for February 2019 and April 2021.

Radar detection data

Num ava/season	0	1	2	3	4	5	6	7	8 or more
17-18	162	10	3	1	2	2	0	1	1
18-19	155	18	2	4	1	0	0	0	2
19-20	148	18	9	4	2	1	0	0	1
20-21	156	12	9	1	2	1	1	0	0
21-22	149	13	6	6	4	0	1	0	3
22-23	159	7	7	1	2	0	1	2	3
23-24	166	6	4	1	0	1	2	0	3
total	1095	84	40	18	13	5	5	3	13

Table A3: Frequency of daily avalanche count per season.

Dataset

Parameter	Variable name	Unit	Description
New snow depth	int_newsnowdepth	[cm]	Interpolated depth of the new snow the last 24h
Temperature	int_temp	[°C]	Interpolated value of average temperature last 24h
Precipitation	int_24hprecip	[mm]	Interpolated value for the total amount of precipitation for the last 24h based on observations
Snowdepth	int_snowdepth	[cm]	Interpolated value for the snow depth at 06:00 based on observations
Observed snow depth	obs_snowdepth	[cm]	measured snow depth at the weather station at Ytre Holmebukt
Observed precipitation	obs_24hprecip	[mm]	measured amount of precipitation at the weather station Ytre Holmebukt
Danger level	dangerlevel	1,2,3,4,5	Regional forecasted avalanche danger level
wind	wind	0,1	Binary variable that is 1 if there was wind from directions that might increase the avalanche danger, 0 if not
Wind intensity	windintensity	-	Average wind speed from directions that might increase avalanche danger
Number of avalanches	num_avalanches	0,1,2,3...	number of avalanches this day (response for GLM)
Avalanche occurred	avalanche_occured	0,1	A binary variable for whether or not avalanche activity has been detected, 0 if no, 1 if yes (response)

Table A4: Description of explanatory variables and response variables.

Results

The red, yellow, green, and blue line displays the values for the predicted climatology for the corresponding season, presented in Table 9 in all remaining figures.

GLM

Table A5: P-values, GLM

Coefficient	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**
(Intercept)	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16	2.86e-13	< 2e-16
int_newsnowdepth	1.30e-14	3.21e-09	3.53e-06	5.06e-11	1.04e-12	2.06e-05	5.26e-11
int_24hprecip	0.27530	0.56001	0.27029	0.49630	0.72170	0.914605	0.98728
int_snowdepth	1.08e-05	1.22e-06	1.21e-07	5.38e-06	7.22e-09	1.79e-05	0.08602
int_temp	3.81e-06	0.01595	0.00243	0.00065	5.08e-05	0.154758	0.03201
obs_24hprecip	0.00597	0.05856	0.05392	0.26750	0.02490	0.000881	0.00554
obs_snowdepth	1.21e-12	9.94e-10	< 2e-16	2.15e-09	< 2e-16	5.87e-05	3.97e-08
windintensity	5.42e-08	0.00151	1.98e-06	1.03e-07	5.60e-05	0.110790	6.90e-12

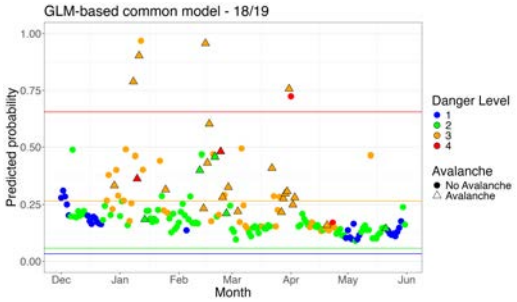


Figure A3: Common GLM model predictions for season 2018-2019.

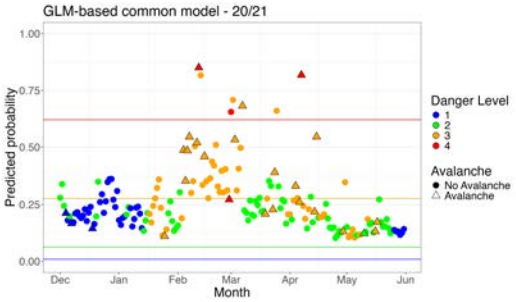


Figure A4: Common GLM model predictions for season 2020-2021, Poisson.

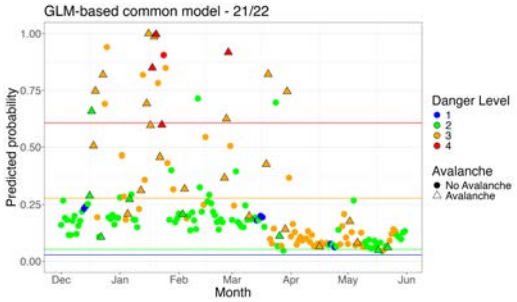


Figure A5: Common GLM model predictions for season 2021-2022.

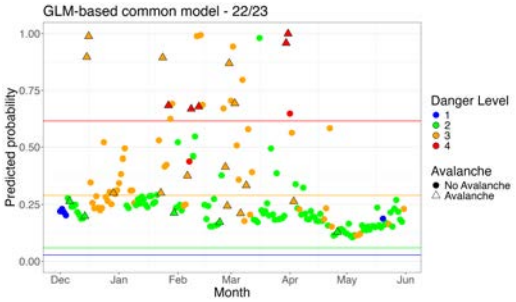


Figure A6: Common GLM-model predictions for season 2022-2023.

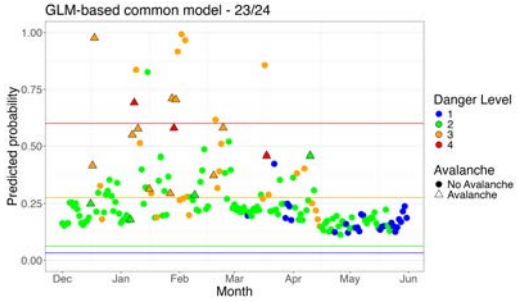


Figure A7: Common GLM model predictions for season 2023-2024.

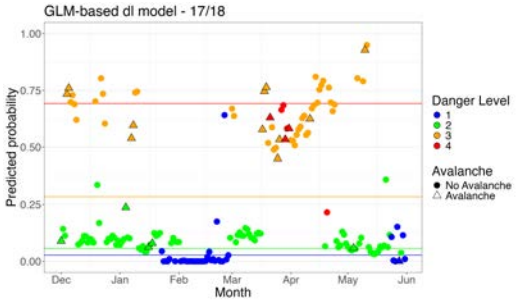


Figure A8: dl GLM model Predictions for season 2017-2018.

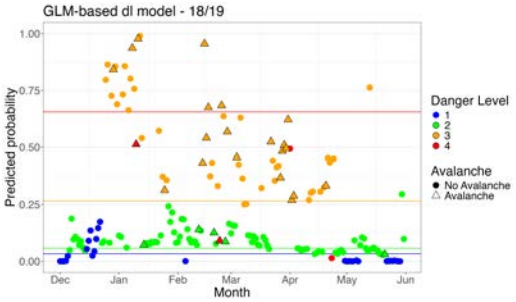


Figure A9: dl GLM model predictions for season 2018-2019.

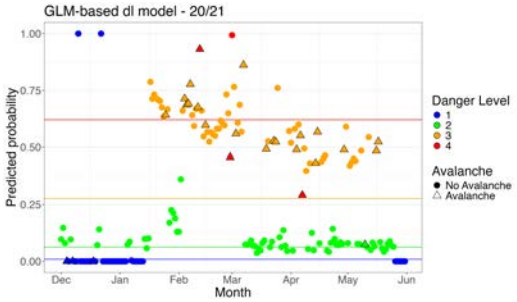


Figure A10: dl GLM model predictions for season 2020-2021.

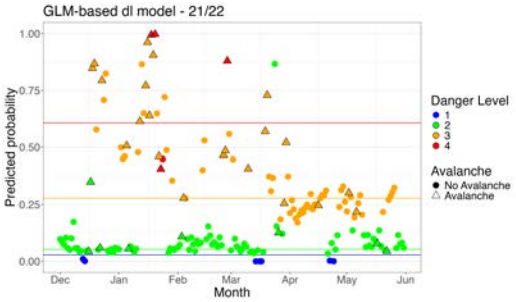


Figure A11: dl GLM model predictions for season 2021-2022.

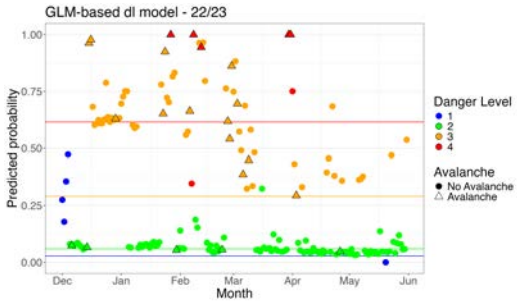


Figure A12: dl GLM model predictions for season 2022-2023.

Tree

Table A6: Variable Importance Scores

Variable	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**
obs_24hprecip	38.03	44.70	41.00	36.34	37.85	38.54	37.63
int_24hprecip	34.60	36.73	34.41	31.30	32.18	31.44	33.05
int_newsnowdepth	22.35	25.37	28.34	20.22	21.75	22.84	22.38
obs_snowdepth	8.08	8.26	8.42	3.33	5.42	7.70	9.06
int_snowdepth	5.27	13.32	2.00	7.37	4.36	5.87	3.01
int_temp	7.66	3.69	2.37	10.70	5.70	11.26	9.96
windintensity	1.02	1.59	0.67	0.20	1.49	0.66	11.91

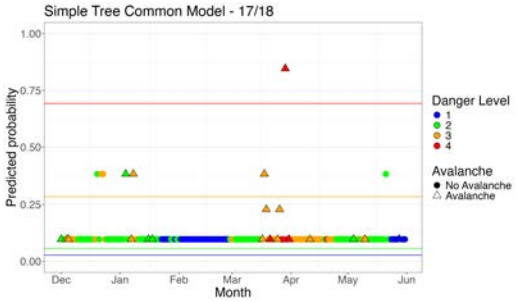


Figure A13: Common Tree-model predictions for season 2017-2018.

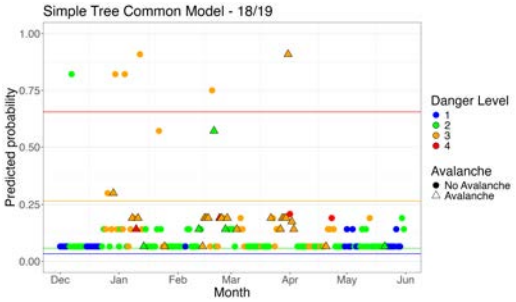


Figure A14: Common tree model predictions for season 2018-2019.

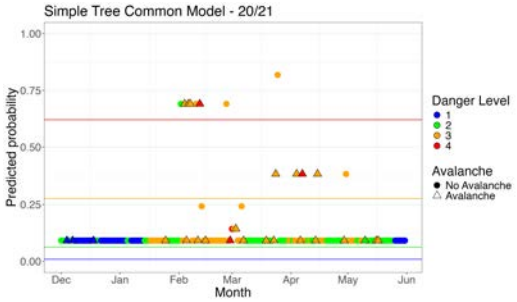


Figure A15: Common tree model predictions for season 2020-2021.

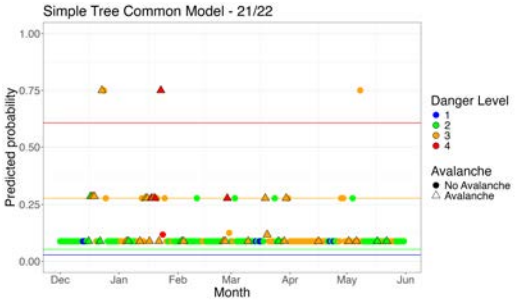


Figure A16: Common tree model predictions for season 2021-2022.

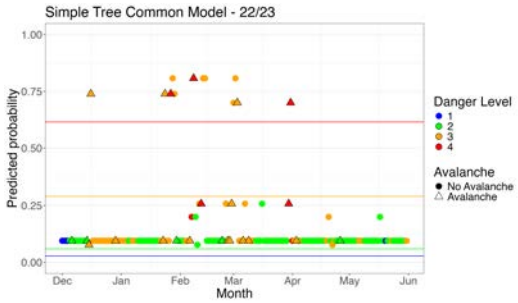


Figure A17: Common tree model predictions for season 2022-2023.

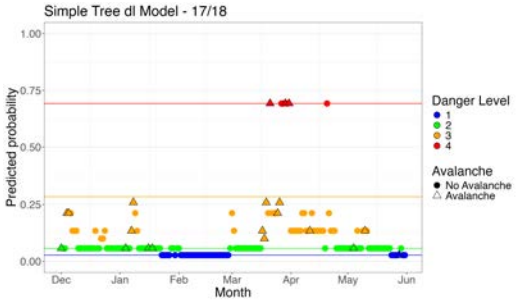


Figure A18: dl tree model predictions for season 2017-2018.

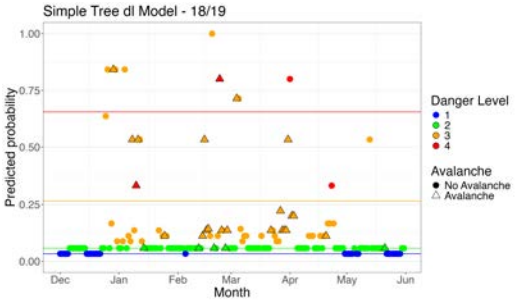


Figure A19: dl tree model predictions for season 2018-2019.

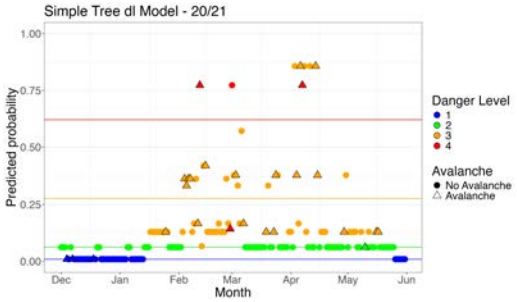


Figure A20: dl tree model predictions for season 2020-2021.

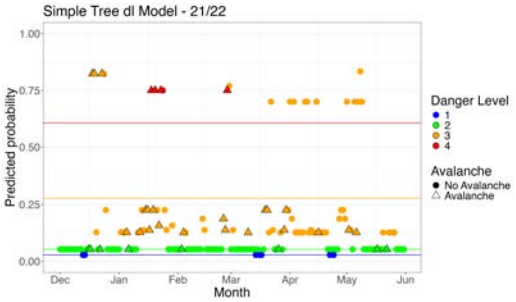


Figure A21: dl tree model predictions for season 2021-2022.

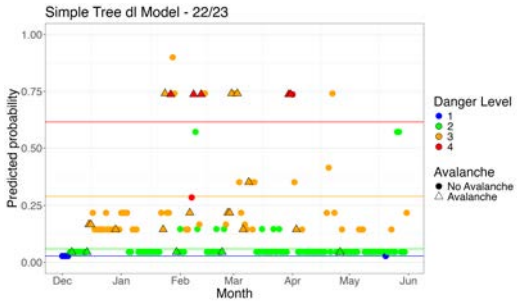


Figure A22: dl tree model predictions for season 2022-2023.

Table A7: Variable metrics, RF

Variable	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7
int_newsnowdepth	No	0.0220	0.0268	0.0458	0.0159	0.0194	0.0262	0.0211
	Yes	0.0215	0.0325	-0.0116	0.0356	0.0228	0.0200	0.0244
int_temp	No	0.0099	0.0118	0.0106	0.0081	0.0074	0.0072	0.0089
	Yes	0.0049	0.0023	-0.0036	0.0031	-0.0037	0.0085	0.0001
int_24hprecip	No	0.0527	0.0528	0.0537	0.0592	0.0483	0.0521	0.0472
	Yes	-0.0395	-0.0455	-0.0392	-0.0069	-0.0427	-0.0349	-0.0349
int_snowdepth	No	0.0109	0.0140	0.0117	0.0105	0.0102	0.0087	0.0157
	Yes	-0.0079	-0.0110	-0.0161	-0.0138	-0.0142	-0.0114	-0.0173
obs_snowdepth	No	0.0106	0.0116	0.0136	0.0138	0.0113	0.0097	0.0154
	Yes	0.0154	0.0098	0.0234	0.0111	0.0197	0.0234	0.0111
obs_24hprecip	No	0.0489	0.0462	0.0469	0.0578	0.0466	0.0540	0.0431
	Yes	0.0202	0.0165	0.0299	-0.0005	0.0270	0.0235	0.0219
windintensity	No	0.0121	0.0107	0.0118	0.0097	0.0078	0.0096	0.0125
	Yes	-0.0061	-0.0110	-0.0121	0.0003	-0.0033	0.0027	-0.0140

Random Forest

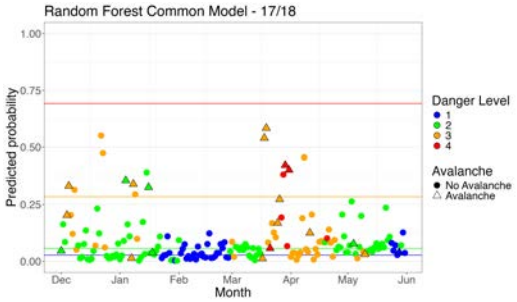


Figure A23: Common RF model predictions for season 2017-2018.

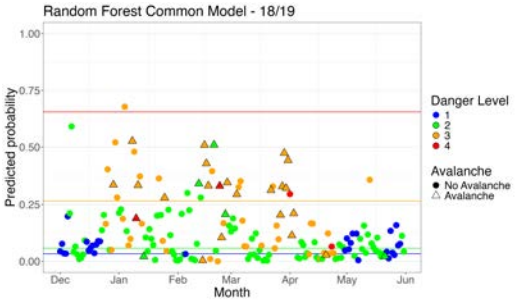


Figure A24: Common RF model predictions for season 2018-2019.

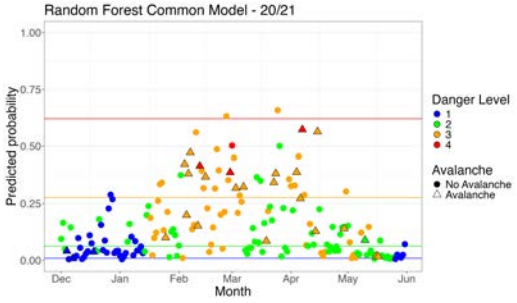


Figure A25: Common RF model predictions for season 2020-2021.

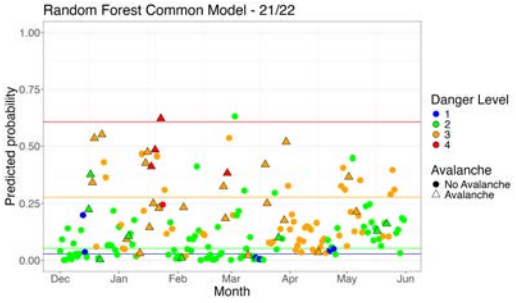


Figure A26: Common RF model predictions for season 2021-2022.

x

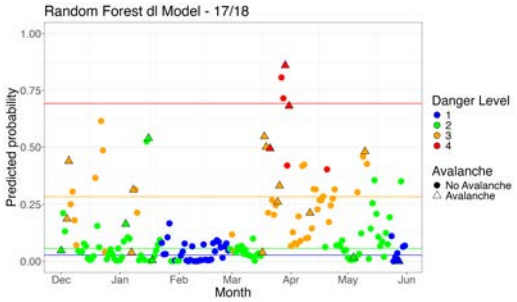


Figure A28: dl RF model predictions for season 2017-2018.

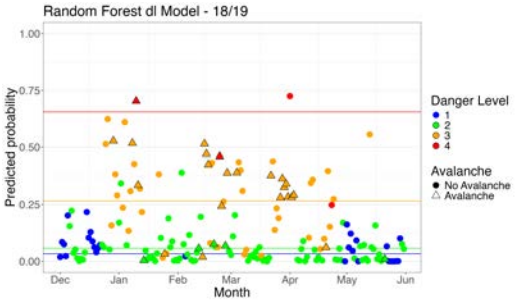


Figure A29: dl RF model predictions for season 2018-2019.

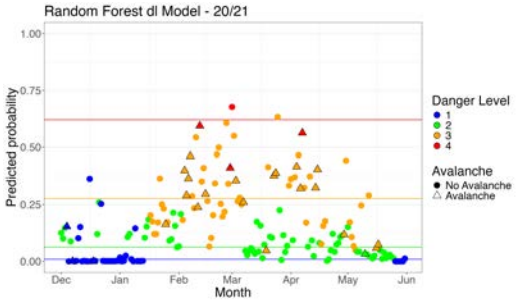


Figure A30: dl RF model predictions for season 2020-2021.

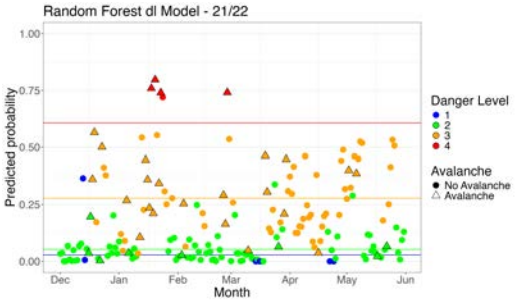


Figure A31: dl RF model predictions for season 2021-2022.

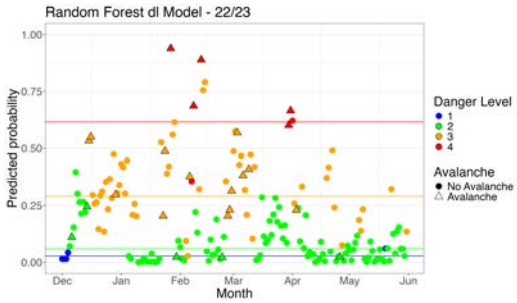


Figure A32: dl RF model predictions for season 2022-2023.

KNN

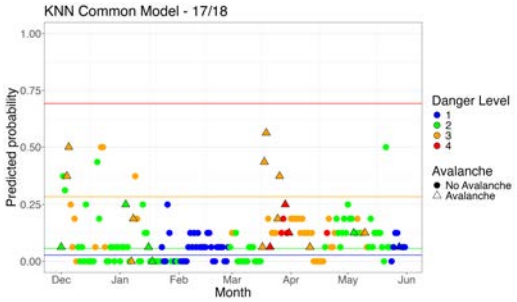


Figure A33: Common kNN model predictions for season 2017-2018.

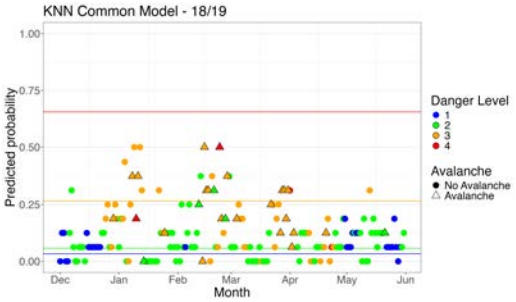


Figure A34: Common kNN model predictions for season 2018-2019.

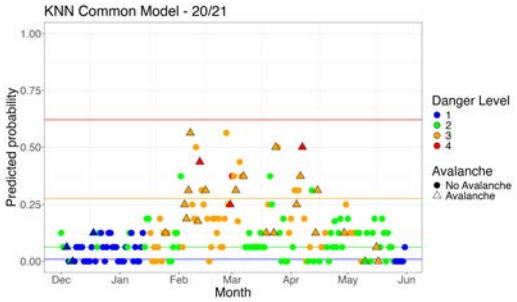


Figure A35: Common kNN model predictions for season 2020-2021.

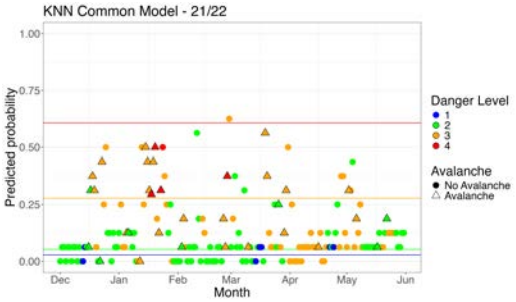


Figure A36: Common kNN model predictions for season 2021-2022.

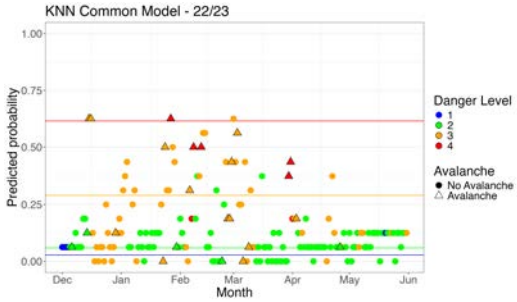


Figure A37: Common kNN model predictions for season 2022-2023.

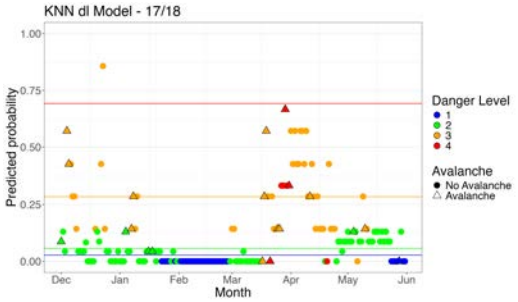


Figure A38: dl kNN model predictions for season 2017-2018.

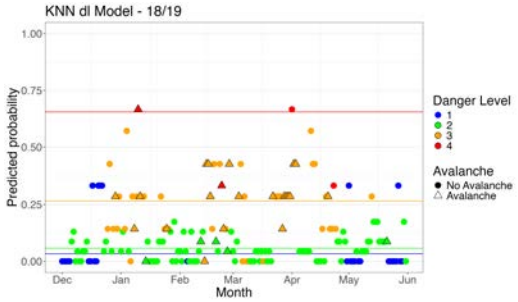


Figure A39: dl kNN model predictions for season 2018-2019.

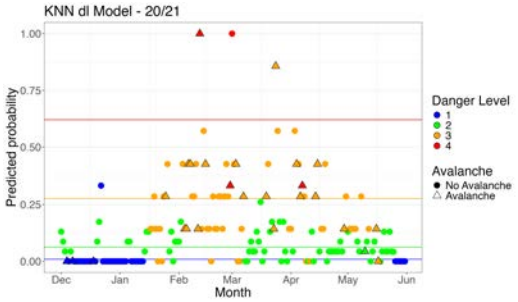


Figure A40: dl kNN model predictions for season 2020-2021.

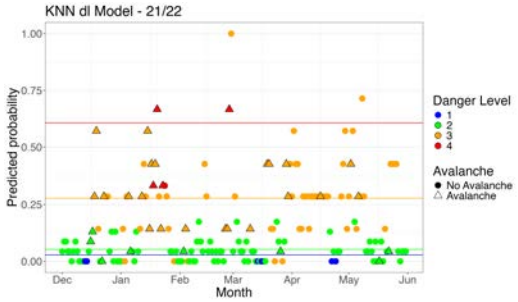


Figure A41: dl kNN model predictions for season 2021-2022.

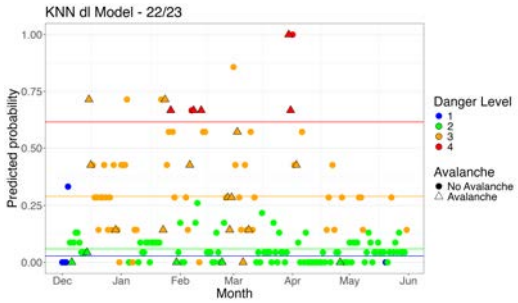


Figure A42: dl kNN model predictions for season 2022-2023.