



**University of  
Zurich<sup>UZH</sup>**

Department of Geography

# Predicting and Characterizing Backcountry Skiing Activity in the Swiss Alps using VGI Data

Master's Thesis  
July 2023

**Author:**

Leonie Schäfer  
16-716-763

**Supervisors:**

Prof. Dr. Ross Purves  
Frank Techel  
Günter Schmudlach

# Acknowledgements

Firstly, I want to thank Professor Ross Purves for his excellent guidance and for always providing me with new perspectives to look at my thesis. Ross once said that there is no such thing as “no result”, which encouraged me to keep going even when the results I obtained sometimes did not align with my expectations.

Secondly, I want to thank Frank Techel for sharing his valuable and helpful insights on avalanche research and particularly backcountry skiing, a field in which I was a complete novice at the beginning of this thesis. Although being often busy with avalanche forecasting, Frank always made the time to answer and discuss my questions.

I also want to express special gratitude to Günter Schmuldach for providing me with backcountry skiing data that was exceptionally well maintained and pre-processed, which made the further processing of the data much easier for me. This thesis would not have been possible without this data.

Lastly, I want to thank Michele Volpi, for taking the time to answer my many questions on machine learning.



# Abstract

In recent years, the popularity of backcountry skiing has grown significantly, and the rising use of GPS technology in mobile devices has progressively enabled research on backcountry skiing behavior. To date, only few studies focus on the spatio-temporal distribution of backcountry ski tours and particularly the prediction thereof. Knowing when and where individuals engage in backcountry skiing is a key aspect for assessing the avalanche risk that skiers encounter in the backcountry. This thesis presents pioneering work by using obfuscated Volunteered Geographic Information data of backcountry ski tours to predict backcountry skiing activity in the Swiss Alps. Furthermore, this study proposes a methodology to enrich geographically obfuscated GPS data with weather variables, such as precipitation, temperature, and sunshine duration. Through the implementation of a random forest algorithm, backcountry skiing activity could be predicted with a 77% accuracy by incorporating the avalanche conditions, weather conditions, free time variables, and popularity of the region as predictors. Popularity, the avalanche danger level, and the relative sunshine duration were shown to be the most important predictors for backcountry skiing activity. The probability of backcountry skiing activity is increased by a factor of 2.5 on a day with optimal weather conditions compared to a day with bad weather conditions, while the probability is decreased by a factor of 1.5 on days with high avalanche danger compared to days with low avalanche danger.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Research Gap . . . . .	2
<b>2</b>	<b>Background</b>	<b>4</b>
2.1	Avalanches . . . . .	4
2.1.1	Contributing Factors . . . . .	4
2.1.2	Avalanche Hazard, Danger and Risk . . . . .	7
2.2	Backcountry Skiing . . . . .	8
2.3	Leisure and Outdoor Recreation . . . . .	10
2.4	Volunteered Geographic Information . . . . .	12
2.5	Machine Learning . . . . .	13
2.5.1	Random Forest . . . . .	14
2.5.2	Imbalanced Data . . . . .	15
<b>3</b>	<b>Data</b>	<b>18</b>
3.1	Avalanche Risk Property Dataset . . . . .	18
3.1.1	Overview . . . . .	18
3.1.2	Obfuscation . . . . .	19
3.1.3	Pre-Processing . . . . .	20
3.2	Raw Danger Level . . . . .	20
3.3	Warning Regions . . . . .	21
3.4	Weather Data . . . . .	23
3.4.1	Air Temperature . . . . .	24
3.4.2	Relative Sunshine Duration . . . . .	24
3.4.3	Precipitation . . . . .	24
3.5	Digital Elevation Model . . . . .	25
<b>4</b>	<b>Methods</b>	<b>26</b>
4.1	Software . . . . .	27
4.2	Modelling Backcountry Skiing Activity - Conceptual Background . . . . .	27
4.3	Generation of Absence Data . . . . .	28

4.4	Data Enrichment with Predictor Variables . . . . .	31
4.4.1	Meteorological Features . . . . .	32
4.4.2	Popularity . . . . .	34
4.4.3	Raw Danger Level . . . . .	36
4.4.4	Holiday . . . . .	36
4.4.5	Day of the Season / Season . . . . .	37
4.5	Model Building . . . . .	37
4.5.1	Target Variable and Features . . . . .	37
4.5.2	Data-Specific Adjustments . . . . .	38
4.5.3	Algorithm-Specific Adjustments . . . . .	40
4.5.4	Verification and Validation . . . . .	42
<b>5</b>	<b>Results</b>	<b>44</b>
5.1	Characterization of Presence Data . . . . .	44
5.2	Presence vs. Absence Data . . . . .	48
5.3	Prediction . . . . .	52
5.3.1	Hyperparameters . . . . .	52
5.3.2	Skill Scores . . . . .	54
5.3.3	Variable Importance . . . . .	55
5.3.4	Partial Dependence . . . . .	57
5.3.5	Temporal Distribution . . . . .	60
5.3.6	Spatial Distribution . . . . .	62
5.3.7	Robustness . . . . .	71
<b>6</b>	<b>Discussion</b>	<b>74</b>
6.1	Model - Building, Performance, Usability . . . . .	74
6.2	Which Variables are the Best Predictors? . . . . .	77
6.3	Limitations . . . . .	79
6.4	Implications for Practitioners . . . . .	81
6.5	Implications . . . . .	82
<b>7</b>	<b>Conclusions and Further Work</b>	<b>85</b>
<b>A</b>	<b>Warning Regions</b>	<b>95</b>
<b>B</b>	<b>Skill Scores</b>	<b>97</b>

# List of Abbreviations

ARPD	Avalanche Risk Property Dataset
DHM	Digital Elevation Model
DT	Decision Tree
EAWS	European Avalanche Warning Services
CRS	Coordinate Reference System
FN	False Negative
FP	False Positive
GPS	Global Positioning System
KSS	Hanssen-Kuipers Skill Score
ML	Machine Learning
NA	Not Applicable
netCDF	Network Common Data Form
OOB	Out-of-Bag
PCF	Probability Change Factor
RCSD	Route Click Statistic Dataset
RDL	Raw Danger Level
RF	Random Forest
RQ	Research Question
SLF	WSL Institute for Snow and Avalanche Research SLF
$S_{rel}$	Relative Sunshine Duration
TN	True Negative
TP	True Positive
VGI	Volunteered Geographic Information
WR	Warning Region
WRC	Warning Region Code

# List of Figures

2.1	Avalanche triangle according to Fredston et al. (1994). . . . .	5
2.2	Example of a decision tree to categorize a skier. Black boxes show node splits, while coloured ovals show terminal nodes. The black oval at the top indicates the data input into the decision tree. . . . .	15
3.1	Distribution of yearly recorded tracks in the ARPD. . . . .	19
3.2	a. Warning regions, the smallest spatial units used to communicate avalanche danger in the avalanche forecasts in Switzerland. Each warning region is labelled with its warning region code (WRC). b. Greater climatic regions as defined by the first digit of the WRC. 1 – Westlicher Alpennordhang, 2 – Zentraler Alpennordhang, 3 – Östliche Alpennordhang, 4 – Wallis, 5 – Nord- und Mittelbünden, 6 – Zentraler Alpensüdhang, 7 – Engadin/ östlicher Alpensüdhang. Regions 8 – Jura and 9 – Mittelland are excluded. All warning regions, including names, can be found in Appendix A. . . . .	22
3.3	Operational station network SwissMetNet, which comprises approximately 160 automatic measuring stations, including approximately 100 automatic precipitation measurement stations. Basemap: swissopo. . . . .	23
3.4	Example data showing the interpolated, gridded meteorological variables used in this thesis, with a spatial resolution of 1 km. a. Temperature, b. Precipitation, and c. Relative sunshine duration for January 11, 2015. In b., ray-like structures are visible, which are artefacts of the RADAR image. Coordinates correspond to the Swiss Grid LV95. . . . .	24
4.1	Methodology of modelling and prediction. Numbers in brackets indicate the section in which the respective part of the process is discussed. . . . .	26
4.2	Examples of hypsometric curves for the warning regions a. 4232 – Südliches Simplon Gebiet and b. 2132 – Ybrig. To make different warning regions comparable, axes show the relative elevation (y) and area above respective elevation (x), rather than absolute values. Warning region specific information is provided in the top right corner. Red dots indicate the mean elevation of existing tracks. For better readability, only a sample of all tracks in these regions is displayed. . .	30

4.3	Hypsometric curve of warning region 1225 – Iffigen showing the mean elevation of presence tracks (red dots) and artificially generated elevation of absence tracks (blue triangles). To make different warning regions comparable, axes show the relative elevation (y) and the relative area above respective elevation (x), rather than absolute values. Warning region-specific information is provided in the top right corner. For better readability, only a sample of both groups is displayed. . .	31
4.4	Procedure to calculate meteorological variables for a given track on the example of $S_{rel}$ . Procedure is repeated for every track and meteorologic variable (Temperature, Precipitation, $S_{rel}$ ). a. - d. show different states of the calculation, 1. - 3. indicate different processes. . . . .	33
4.5	Application of <i>raster::mask()</i> with different spatial resolutions of $S_{rel}$ [%] (colour) and DHM25. The black cells show the polygonized, track-specific elevation belt emerging from DHM25 for an example track. Figures on the left show the $S_{rel}$ raster of the whole warning region overlayed with the elevation belt, figures on the right show the raster cells of $S_{rel}$ selected by the <i>raster::mask()</i> function overlayed with the elevation belt. a. $S_{rel}$ : 1 km, DHM: 200 m. b. $S_{rel}$ : 1 km, DHM: 1 km. c. $S_{rel}$ : 200 m, DHM: 200 m. . . . .	34
4.6	a. Histogram of warning region track densities. Each black horizontal line represents one warning region. Track densities are calculated based on all recorded tracks from season 2013/14 to season 2020/21. A smoothed density curve is indicated in black. b. popularity classification into five classes based on track density, ranging from 1 = high popularity to 5 = low popularity. The fraction of warning regions in the respective class is indicated on top of every column. . . .	35
4.7	Spatial distribution of popularity classes of Alpine warning regions in Switzerland. Popularity ranges from “1 = high popularity” to “5 = low popularity”. Touristic hotspots are displayed in blue. In three regions, no track was recorded in the observation period, therefore no value was calculated (indicated by NA). All warning regions, including names, can be found in Appendix A. . . . .	35
4.8	Data-specific model optimization process. The initial model incorporates all predictor variables with implementation 1 (4.4). The model is altered (from top to bottom, model 1 to model 8) by a stepwise introduction of implementation 2 for every predictor variable. If the altering yields a higher performance, it is accepted, otherwise it is discarded. In the last three alterations, the least popular warning regions and the least important predictor is excluded. . . . .	40
4.9	Confusion Matrix showing the relationship between predicted and observed entities. Green boxes show correctly predicted entities, red boxes show erroneous predictions. . . . .	42
5.1	Number of tracks per season in the observation period from 2013/14 to 2020/21.	45

5.2	Yearly evolution of weekly aggregated tracks for each season in the observation period between 2013 and 2021. Due to differences in season days (leap years), there are slight variations among the definition of weeks in each season. . . . .	46
5.3	Yearly evolution of weekly summarized tracks, stacked for all seasons to show the overall trend throughout the season. Note that the exact definitions of each week may differ slightly among seasons due to different amount of season days (leap years). . . . .	46
5.4	Frequency of weekdays throughout all seasons for presence tracks (yellow) and baseline frequencies (green). The baseline frequency for each day is 1/7 or 14.3%. . . . .	47
5.5	Frequency of raw danger levels throughout all seasons for presence tracks (yellow) and baseline frequencies (green). The baseline frequency is based on seasons 2012/13 to 2020/21, and is obtained from <a href="http://www.slf.ch">www.slf.ch</a> . . . . .	48
5.6	Overall track density per Warning Region for the Swiss Alps, 2013 – 2021. . . . .	48
5.7	Frequency distribution and boxplots of precipitation (a., b.), relative sunshine duration (c., d.) and temperature (e., f.) by class. In a., only precipitation values <2 mm and relative frequencies <4% are displayed, to avoid visual distortion owing to very high frequencies for 0 mm values (65% of absence tracks resp. 85% of presence tracks). For better readability, the y-axis for b. was logarithmically transformed, consequently only values >0 mm are displayed . . . . .	51
5.8	Frequency distribution of a. weekday, b. raw danger level, and c. popularity by activity group. Baseline frequencies are indicated by the black boxes. Baseline frequency of popularity emerges from Figure 4.6, baseline frequency of RDL emerges from <a href="http://www.slf.ch">www.slf.ch</a> . In c., absence frequency for level 5 is 0.08% and presence frequency for level 0 is 0.1%, therefore they are not visible. Presence frequency for level 5 is 0%. . . . .	52
5.9	a. OOB-error rate by class as produced by <i>randomForest::randomForest()</i> and b. cross-validation skill scores with increasing number of trees ( <i>ntree</i> ). . . . .	53
5.10	Skill scores obtained through cross-validation for different values of a. <i>mtry</i> , i.e., the number of randomly selected features for each split and b. <i>sampsiz</i> , i.e., the number of sampled data points for every tree. The sample size for both absence and presence class is the same. . . . .	54
5.11	Warning Regions with a. <10 (Model 6) and b. <15 (Model 7) total tracks from 2013/14 to 2020/21. . . . .	55

5.12	Variable Importance as measured by a. MeanDecreaseAccuracy and b. MeanDecreaseGINI. Data types for features in b. are listed next to the graph. For categorical features, the number of categories is expressed in brackets. Importance Metrics in this Figure emerge from a model with no hyperparameter fine-tuning or data-specific adjustments and are thus not comparable to other variable importance plots presented in this section. The following variable names are abbreviated: pop = popularity, rdl = raw danger level, SrelD = relative sunshine duration, precip = precipitation, temp = temperature, dos = day of the season, wrc = warning region code . . . . .	56
5.13	Variable Importance of best performing model (Model 5), representing a. presence class, b. absence class, c. weighted mean of both classes, and c. unweighted mean of both classes. The following variable names are abbreviated: pop = popularity, rdl = raw danger level, SrelD = relative sunshine duration, precip = precipitation, temp = temperature, dos = day of the season, wrc = warning region code. . . . .	57
5.14	Probability distribution of a. presence and b. absence class. . . . .	58
5.15	Partial dependence of predictor variables for presence class. The red line shows the mean of 20 iterations of the variable importance calculation, the transparent area shows the 95% confidence interval. . . . .	59
5.16	Mean daily probability for presence vs. daily proportion of presence tracks. The black curve indicates the local polynomial regression line with a 95% confidence interval. . . . .	61
5.17	Mean probability of presence vs. absolute number of presence tracks. Each point represents one day. The blue curve indicates the local polynomial regression line with a 95% confidence interval. . . . .	62
5.18	Sensitivity per warning region with best performing model (Model 5) for all test seasons. NA values are employed when calculation is not possible. . . . .	65
5.19	Specificity per warning region with best performing model (Model 5) for all test seasons. . . . .	66
5.20	Balanced Accuracy per warning region with best performing model (Model 5) for all test seasons. NA values are employed when calculation is not possible. . . . .	67
5.21	Hanssen-Kuipers Skill Score (KSS) per warning region with best performing model (Model 5) for all test seasons. NA values are employed when calculation is not possible. . . . .	68
5.22	Relationship between a. sensitivity, b. specificity, c. balanced accuracy, and d. KSS and total track count per warning region, season 2017/18. The blue curve indicates the local polynomial regression line with a 95% confidence interval. Plots for all other seasons can be found in Appendix B. . . . .	69



5.23	Histogram of a. sensitivity, b. specificity, c. balanced accuracy, and d. KSS values from all seasons and warning regions. Note that the bin width is 0.05 for all skill scores, but d. KSS has a greater range than the other skill scores, therefore it is represented by more bins. . . . .	70
5.24	Standard deviation of skill scores among warning regions. Standard deviations are calculated for each skill score and season. Boxplots contain standard deviations of all seasons. . . . .	70
5.25	Total tracks per season vs. skill scores. . . . .	73
A.1	Warning Regions used by the WSL Institute for Snow and Avalanche Research SLF. . . . .	96
B.1	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2013/14. . . . .	97
B.2	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2014/15. . . . .	98
B.3	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2015/16. . . . .	98
B.4	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2016/17. . . . .	99
B.5	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2018/19. . . . .	99
B.6	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2019/20. . . . .	100
B.7	Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2020/21. . . . .	100

# List of Tables

3.1	Example of obfuscated ARPD structure. . . . .	20
3.2	Number of tracks used for the analysis. . . . .	22
4.1	Most important R Packages with description and reference. . . . .	27
4.2	Example structure of presence and absence data before data enrichment with other attributes. . . . .	31
4.3	Overview over presence and absence data. . . . .	36
4.4	Target variable and predictors (i.e., dependent variables or features) with different data-specific implementations for modelling. . . . .	38
4.5	Feature discretization into classes. . . . .	39
4.6	Hyperparameters and tested values for best performance. . . . .	41
4.7	Uncertainties of meteorological variables that are introduced into the data to assess the robustness of the model against input uncertainty. . . . .	43
5.1	Mean and standard deviation for meteorological features by activity group. a. Precipitation, b. Relative sunshine duration, and c. Temperature. For relative sunshine duration and precipitation, additionally the relative frequencies for 0 mm (<10%) and >5mm (>90%) are given. . . . .	50
5.2	Skill scores of different model runs according to Figure 4.8. The standard deviations for each skill score is calculated from all different model runs. . . . .	55
5.3	Skill scores produced by the best performing model (Model 5) with permuted test seasons. For each test season, the complementary data is used for training. Mean and standard deviation of each skill score is provided at the bottom, as well as results from the correlation analysis. . . . .	72
5.4	Skill scores for a model with a known input uncertainty of meteorological variables compared to the best performing model (baseline). . . . .	72
6.1	Probability Change Factor (PCF) of good vs. bad weather, avalanche, and free time conditions. PCFs should only be applied to scenarios with RDL $\geq 0$ . . . . .	82

# Chapter 1

## Introduction

### 1.1 Motivation

Backcountry skiing has gained a lot of popularity in Switzerland during the last few decades (Lamprecht et al., 2014). Although reasons for this growth are not well understood, there seems to be a fascination to ski in untouched terrain, away from prepared slopes (Furman et al., 2010). Not only the solitude in the terrain, but also the additional thrill of being the first one to access an untracked slope of fresh snow tempts skiers to take disproportional risks in the backcountry. This fascination for untouched terrain is described by McCammon (2004) as the “scarcity trap”, one of six heuristic traps that frequently keeps backcountry skiers from making sound decisions. Having a look at accident statistics, it becomes evident that backcountry skiers<sup>1</sup> voluntarily put themselves at risk of serious injury or even death when entering avalanche terrain. Reasonably, a substantial amount of literature focuses on the avalanche risk and the analysis of backcountry skiing accidents, with the goal of gaining insights into the decision-making process of skiers and thereby minimizing the risk they take (e.g., Grímsdóttir and Mcclung, 2006; Schmudlach and Köhler, 2016). Many studies focus on the behavior of backcountry skiers with respect to terrain characteristics, snow conditions and avalanche forecast or demographic, group dynamics and skill level of the skier (Hendrikx et al., 2018; Johnson and Hendrikx, 2021; Mannberg et al., 2018; Winkler et al., 2021). There is plenty of literature on how various factors affect the risk of being exposed to an avalanche, however little attention has been paid on how these factors affect the actual backcountry skiing activity<sup>2</sup>. A reason for this could be that collecting extensive data of backcountry skiers is difficult as it requires data that is often sensitive in terms of the skier’s privacy (de Montjoye et al., 2013).

Techel et al. (2015), among others, pointed out that it is crucial to know when and where people go backcountry skiing (i.e., the exposure of elements at risk), to put avalanche accidents into

---

<sup>1</sup>Hereafter the terms backcountry skiers and (backcountry) recreationists are used interchangeably.

<sup>2</sup>The term “activity” is used in this thesis as a measure for whether people ski in a region or not, rather than the type of the activity itself.

context. Grímsdóttir and Mcclung (2006) defined the avalanche risk as the ratio of accidents to accidents and non-accidents, in other words, the probability of accidentally triggering an avalanche while travelling in avalanche terrain. Yet risk is often analyzed solely relying on accident statistics, overlooking the aspect of exposure or non-accidents (e.g., Winkler et al., 2021). To understand the importance of exposure, consider the following example: Many people feel uneasy to enter an airplane, which is understandable, considering some horrendous airplane accidents in the past. However, putting these accidents into perspective reveals a different picture. In 2017 for instance, approximately four billion passengers embarked on an airplane, while only eight people died in airplane accidents in the same year, which suggests that air travel is one of the safest modes of transportation in the world (Barnett, 2020). It becomes evident, that solely studying accidents without considering the exposure does not capture the full picture. Although there are a few studies in backcountry skiing literature that include exposure in the study of avalanche risk (e.g., Grímsdóttir and Mcclung, 2006; Pfeifer, 2009; Schmudlach and Köhler, 2016; Techel et al., 2015), it was not yet studied which factors contribute to high, respectively low activity.

By modelling the backcountry skiing activity, hence when and where people are in the field, the exposure can be quantified, and contributory factors can be identified. Additionally, it allows avalanche forecasters to evaluate the quality of their forecasts. Forecasters usually get feedback when an event (an avalanche or an avalanche accident) happens, which can help to evaluate whether the avalanche forecast was accurate. On the other hand, when they receive no feedback about avalanches or accidents, the question arises, whether this can be attributed to a low avalanche danger or to the fact that there were simply no people in the field who could possibly report an avalanche. Predicting the backcountry skiing activity will therefore help forecasters to evaluate the quantity of feedback they receive. For instance, when no avalanches get reported on a day where the activity is predicted to be high, it can be an indication of good avalanche conditions<sup>1</sup>. Contrarily, if the activity is predicted to be low and no avalanche gets reported, this does not necessarily indicate good avalanche conditions.

## 1.2 Research Gap

Skitouren guru<sup>2</sup> is a web service that supports backcountry skiers in the selection and planning of suitable ski tours by providing information about such factors. It assigns avalanche risks two times per day to thousands of ski tours in the alpine region. Users can search for ski tours based on several well-established criteria such as travel distance, elevation gain, difficulty and avalanche risk (Schmudlach, 2022b). Since 2007, Skitouren guru has collected over 6 million location points emerging from over 8500 GPS (Global Positioning System) trajectories that were

---

<sup>1</sup>The term “good avalanche conditions” is used in this thesis from the perspective of a backcountry skier, thus it refers to a low avalanche danger.

<sup>2</sup>[www.skitouren guru.ch](http://www.skitouren guru.ch)

uploaded by backcountry recreationists. This dataset has already been analyzed in cooperation with the WSL Institute for Snow and Avalanche Research SLF (SLF), for studies about avalanche danger levels (Techel et al., 2015), or the avalanche risks taken by backcountry skiers under different avalanche forecasts (Winkler et al., 2021). The same data is used in this thesis, except that all coordinate points are obfuscated to protect the privacy of the user. Woźniak et al. (2017) showed, that many people express privacy concerns when engaging in outdoor sports information sharing. Even though obfuscation is not a new phenomenon when it comes to location based services (Kachore et al., 2015), there are no studies that explore the usability of obfuscated GPS data in the backcountry skiing domain. Further, data on backcountry skiing has been only analyzed ex post (e.g. Schmudlach et al., 2018; Winkler et al., 2021) and was not yet used to predict backcountry skiing activity in the future.

These research gaps are tackled by using obfuscated, Volunteered Geographic Information (VGI) data of backcountry ski tours to predict backcountry skiing activity in the Swiss Alps.

Summarized, this thesis addresses two different research gaps: 1.) Predicting backcountry skiing activity and 2.) Doing so with obfuscated GPS data that emerge from VGI. This leads to the following research questions (RQ):

***RQ1:** How can backcountry skiing activity be modeled using obfuscated, user-generated trajectory data?*

***RQ2:** How do the different predictors, such as weather forecast, avalanche conditions, and free time, influence the backcountry skiing activity?*

***RQ3:** What are the most important predictors for backcountry skiing activity?*

## Chapter 2

# Background

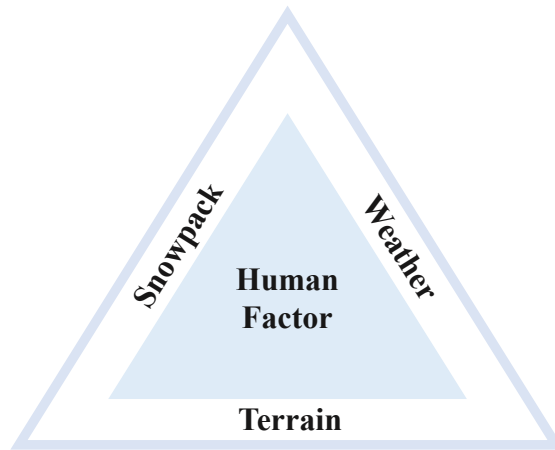
### 2.1 Avalanches

In Switzerland, on average 21 people die every year in avalanches. Nowadays, an overwhelming majority of the victims are recreationists and about 95% of accident avalanches are human-triggered (Schweizer and Techel, 2017). The behavior and decision-making of backcountry skiers is therefore strongly influenced by avalanches, thus it is important to discuss the most important aspects of avalanches and their formation.

Avalanches can be split into several distinct avalanche types. The two most important types are slab avalanches and loose snow avalanches. Loose snow avalanches start at a single point and spread out as they move down the slope, usually involving only surface snow. Oftentimes, loose snow avalanches get released naturally. Slab avalanches on the other hand have a distinct, broad fracture line, which is often initiated by a failure due to a weak layer at depth in the snowpack (McClung and Schaerer, 2006). Slab avalanches are far more dangerous than loose snow avalanches, and they account for more than 90% of deaths occurring in avalanches (SLF, 2023b).

#### 2.1.1 Contributing Factors

Fredston et al. (1994) used a triangle to graphically represent the factors contributing to an avalanche (Figure 2.1), which was later used and adapted by many different authors, avalanche forecasters and avalanche education programs (e.g., McCammon and Hägeli, 2007; Suter et al., 2007; Williams, 1998). The three physical factors presented in the avalanche triangle are snowpack, weather and terrain. As indicated by the triangle, these three factors interact with each other, however, the interactions are complex and much about avalanche release is still unknown (Schweizer et al., 2003). Simplified, weather and terrain both influence the development of a snowpack and the terrain characteristics determine whether an avalanche release is physically possible. In the center stands the “human factor”, which is the main contributor to avalanche accidents, as humans are the trigger for most of the accident avalanche releases (Schweizer and Techel, 2017).



**Figure 2.1:** Avalanche triangle according to Fredston et al. (1994).

## **Terrain**

Terrain is the only time independent avalanche triggering factor. The primary terrain requirement for an avalanche is a slope inclination between  $25^\circ$  and  $55^\circ$ . Slopes with an inclination below  $25^\circ$  usually do not produce enough shear stress and deformation for an avalanche release. Yet, there is no precise lower limit for slope inclination below which the terrain can be regarded as safe. For instance, wet avalanches can occur on slopes with inclinations below  $25^\circ$ . On steep slopes (above  $55^\circ$ ) on the other hand, continuous sluffing of loose snow prevents excessive snow accumulation and consequently avalanche formation (McClung and Schaerer, 2006). However, it is worth noting that there are different ways to measure the slope angle. Additionally, it is to mention that skiers in flatter terrain can get caught by an avalanche that was released in steeper terrain (remote-triggered or natural avalanches). A common rule of thumb for back-country recreation is a  $30^\circ$  threshold for avalanche formation. This rule is presented in many avalanche guides, and on some Swiss ski touring maps terrain steeper than  $30^\circ$  is specially colored (Schweizer et al., 2003). Terrain characteristics are also included in the avalanche forecast in the form of a “core zone”, which describes critical aspect and elevation for avalanche danger levels. As opposed to slope angle, elevation affects avalanche danger only indirectly. In elevations below the tree line for example, the snowpack is relatively stable due to temperature variations and little wind resulting in a generally lower avalanche danger (Harvey et al., 2012). Lastly, the terrain strongly influences snow redistribution through wind, which is discussed in the following section.

## Weather

Precipitation, wind, air temperature and radiation are the most important weather variables that directly or indirectly influence the stability of a snowpack. Generally, two different processes result from weather conditions: Snow accumulation and snow metamorphism that takes place after the snow is accumulated (McClung and Schaerer, 2006).

Precipitation is crucial, as it is primarily responsible for the formation of the snowpack. Once a snowpack is built, the type of precipitation (rain, wet/dry snow) determines how much weight and energy is added to the existing snowpack. Typically, new snow increases the avalanche risk, as it puts additional weight to the snowpack (Fredston et al., 1994; Harvey et al., 2012). Rain on the other hand adds water and heat to the snowpack, which favours wet slab avalanches as the water can lubricate a sliding surface. As precipitation is the primary requirement for snow accumulation, it is the strongest forecasting parameter for large and catastrophic new snow avalanches (McClung and Schaerer, 2006).

Drift snow occurs when snow gets redistributed by wind, and it often accumulates on the lee side of ridges or in gullies and notches, where wind speed decelerates quickly. Drift snow exerts an additional pressure on the snowpack, which can lead to an avalanche release. Additionally, snow particles become broken and abraded when they get deposited by the wind. This can lead to a tightly packed snowpack that can quickly produce slab-like textures, which is an additional factor favoring the formation of an avalanche. Even though drift snow accumulations are influenced by terrain, re-location by wind is often irregular, thus it can be challenging to detect. Also, wind influences the energy fluxes between the snowpack and the atmosphere, which impacts the snow metamorphism and consequently the strength of the bonds between layers in the snowpack (McClung and Schaerer, 2006; Schweizer et al., 2003).

Air temperature is critical as it determines whether precipitation falls as wet or dry snow, or as rain. After a snowpack is built, the air temperature, or more precisely the gradient of the air temperature, interacts with the snowpack, which can alter the surface layer of the snow and produce a weak layer. Also, when high temperatures cause rain, heat gets introduced into the snowpack. The temperature gradient inside the snowpack causes snow metamorphism. This can lead to a stabilization of the snowpack, as it strengthens the bond between particles (Schweizer et al., 2003).

Radiation can have two different effects on the snowpack, it can either warm or cool the snowpack, depending on the type of radiation (short- or long-wave) and the atmospheric conditions, such as cloud cover. Radiative cooling or/and warming can produce weak layers, or it can directly cause an avalanche (McClung and Schaerer, 2006).

## Snowpack

As outlined before, snowpack is a product of weather conditions and terrain. Weather conditions are directly connected to the snowpack as they initially produce it. Subsequently, the weather development after the snowfall determines the stratigraphy of the snowpack. Temperature fluctuations and radiation impact the snow metamorphism, which can produce weak layers



in the snowpack or strengthen the bonds and therefore the stability of the snowpack. When new snow is accumulated on top of a weak layer, a persistent weak layer can emerge, which is a typical avalanche problem, often referred to as old snow problem in avalanche forecasting (Techel et al., 2015). Precipitation generally adds weight and possibly energy to the snowpack, which impacts the stability. Besides weather variables, also the terrain influences the snowpack, as it determines where drift snow accumulates. Further, sun-exposed aspects receive more solar irradiation, which stabilizes the snowpack as a high temperature fluctuation leads to a stronger bond between new snow and old snow layers. Consequently, shaded areas are more prone to avalanches, because weak layers can persist over longer time periods when the temperature fluctuation is low (Harvey et al., 2012).

## **Human**

The human factor relates the snowpack instability to the actual avalanche hazard and includes aspects like ego, incorrect assumptions, or peer pressure (Fredston et al., 1994). According to Fredston et al. (1994), most accidents occur because victims either underestimate the avalanche hazard or overestimate their own ability to deal with it. It is therefore essential to incorporate the human factor into avalanche prevention. Because the human factor is closely related to backcountry skiing in general, it is discussed in more detail in Section 2.2, as it has been frequently studied in backcountry skiing literature.

### **2.1.2 Avalanche Hazard, Danger and Risk**

Avalanches play a critical role in the decision-making of backcountry skiers and therefore the activity prediction carried out in this thesis. This requires an unambiguous terminology for the different terms relating to avalanche danger. The definitions and ideas introduced in this section all follow Statham (2008), unless specified otherwise.

Avalanche danger and avalanche hazard are synonymous. Hereafter, the term avalanche danger is used, but it is interchangeable with avalanche hazard. Statham (2008) describes avalanche danger as a source of potential harm and a likelihood of 1.) the triggering and 2.) the size of an avalanche. It is important to acknowledge that the avalanche danger is independent from the avalanche risk. This means that the avalanche danger can be high, while the risk is low. An example that illustrates this in the context of backcountry skiing is when no individuals are in the field, but the avalanche danger is high.

In this thesis, the avalanche danger level (in this thesis referred to as raw danger level) will be frequently used. The avalanche danger level is a five-scale danger classification that is used to communicate the avalanche conditions. It is important to keep in mind that it describes the avalanche danger, and not the avalanche risk. Only when an element at risk is introduced, the avalanche risk can be assessed. The element at risk can be anything, for instance infrastructure, animals, or humans, depending on the context. In the context of this thesis, the element at risk is always assumed to be a backcountry recreationist, hence the term avalanche risk relates the avalanche hazard to individuals in the field. The avalanche risk can be defined as a combination

of 1.) the exposure of the element at risk (i.e., the backcountry skier), and 2.) its vulnerability to the avalanche danger. Vulnerability expresses the susceptibility to the impact of an avalanche hazard, which is specific to the element at risk. For instance, when a healthy adult stands beside an old person, they both may be exposed to the same avalanche danger, yet they are at different risks, as they have different vulnerabilities. The exposure on the other hand describes where and for how long an element at risk is in the terrain. For backcountry skiing, this is the single most important factor for the avalanche risk, as the exposure is the consequence of the actions a backcountry skier takes. The exposure is higher when a lot of individuals are in the terrain, as illustrated in Section 1.1 with the airplane example. Consequently, when analyzing the accident risk, the exposure, i.e., how many individuals are in the field, needs to be considered (Techel et al., 2015).

## 2.2 Backcountry Skiing

In this thesis, backcountry skiing is used as a collective term for snowboarding and skiing in remote areas. It can be defined as skiing or snowboarding in the backcountry, away from prepared slopes, where people primarily ascend under their own power. This stands in contrast to off-piste skiing, which takes place close to ski areas, where people often use ski resort facilities to ascend (Winkler et al., 2021). A typical backcountry skier is male, educated (university degree), has no children and is roughly aged between 26 and 40 (Hendrikx et al., 2022; Winkler et al., 2016).

The behavior of backcountry skiers has been researched in several parts of the world, including North America, Scandinavia, and the Alpine European countries, including among others Switzerland, France, and Italy. Even though some studies combined data from different geographic regions, there is a lack of comparison between them. Considerable differences in backcountry recreation and especially in avalanche warning systems between different geographic regions might influence the behavior of backcountry skiers, and as Hägeli et al. (2009) suggest, preclude the adoption of an existing decision aid to another geographic region. Other Alpine countries in Asia or South America are greatly underrepresented in literature.

There are mainly two different approaches to analyze the behavior of backcountry skiers: 1.) Survey-based studies, sometimes combined with GPS-tracking data, that focus on the decision-making process, demographics, and group-dynamics of the skiers (e.g., Furman et al., 2010; Hägeli et al., 2009; Hendrikx et al., 2022; Johnson and Hendrikx, 2021) and 2.) purely GPS data-driven approaches that analyze the skiing behavior in terms of terrain use and avalanche risk, without relating it to the individual demographic or decision-making processes (e.g., Bielański et al., 2018; Schudlach and Köhler, 2016; Techel et al., 2022). For a long time, field studies and investigations have been challenging due to the remoteness in which backcountry skiing naturally takes place. Within the last decade however, research has been enabled and driven by the more widespread use of GPS tracking technology (Bielański et al., 2018; Hendrikx et al., 2022).

As avalanches constitute the main danger for backcountry skiers, a great part of their decision-

making is based on minimizing the avalanche risk. Consequently, a vast majority of studies focus on avalanche risk and prevention. In the last 20 years, a paradigm-shift away from snowpack- and weather-related causes of avalanche accidents towards a mix of human-related factors occurred (Hendrikx et al., 2022). In his much-cited article about common heuristics in avalanche-related recreational activities, McCammon (2004) highlights the importance of the human factor in avalanche prevention. His most important finding is that people take higher risks under certain well-defined, human-related conditions. He concludes six heuristic traps that cause backcountry skiers to make erroneous decisions: People usually take higher risks when 1.) in familiar terrain (familiarity trap), 2.) in larger groups (social facilitation trap), 3.) the commitment to follow an initial decision is high and thereby overrules critical new information about avalanche danger (consistency trap), 4.) wanting to impress or be accepted by other people (acceptance trap), 5.) there is an overall positive impression of the leader (expert halo trap), and 6.) there is a chance to be the first one to ski an untracked slope of fresh snow (scarcity trap). To reduce the number of deadly accidents, avalanche education needs to address these six traps, says McCammon (2004).

Numerous studies have examined the human factor and its importance for the risk taken by backcountry recreationists (Atkins, 2000; Hendrikx et al., 2018; Johnson and Hendrikx, 2021; Maguire, 2014; McCammon, 2009). Hendrikx et al. (2022) conducted a study in the US and northern Norway that tracked the decision-making process of backcountry recreationists by using surveys combined with GPS data. They found that solo-travel, although commonly regarded as more risky behavior, represented a quarter of all observed tours. At the same time, only 6% of all solo-trips were undertaken by females. Also, groups of two or three males spent more time in complex terrain than any other group combination, while solo females avoided all complex terrain. However, it has to be noted that females are greatly underrepresented in backcountry skiing, and thus the sample size of female recreationists was comparably small. Not only gender or group dynamics, but also the self-assessed backcountry skills impact the behavior. Self-assessed experts spend more time in complex and more serious avalanche terrain, but they also have higher terrain management skills than self-assessed intermediates. Contrary to that, intermediates and experts seem to be similarly active with regard to the forecasted avalanche danger, with highest activities under moderate avalanche danger rating and lowest activities under low and high danger ratings in both groups. Even though the number of trips under a certain forecasted avalanche danger level does not differ within the two groups, Johnson and Hendrikx (2021) found that the terrain-use does differ, as skiers with higher avalanche education tend to adjust their terrain use towards less steep slopes when avalanche danger increases, while using more steep slopes on lower avalanche danger days.

The key role of the decision-making process is also highlighted by Hägeli et al. (2009), who showed that a decision-aid that combines avalanche conditions, terrain ratings, and elevation induces a more avalanche danger-sensitive behavior of amateur recreationists. This is in line with Sch mudlach and Köhler (2016), who combined avalanche forecast data with terrain data to calculate and later publish the avalanche risk of 625 popular routes on the Skitourenguru

website. A Route Click Statistics Dataset (RCSD) from Skitourenguru reveals that most of the users click on routes with “low risk”. Hence, the most prominent factor influencing the decision-making is unambiguously the avalanche danger. But besides the avalanche danger, there are other important factors as well. Furman et al. (2010) found in a survey-based study that recreationists, as suggested by McCammon (2004), are more likely to ski a slope when the slope is untracked, when there is a leader present in the group or when the slope is familiar. However, McCammon’s (2004) statement that men in the presence of women are more likely to ski hazardous terrain, was not supported by Furman et al. (2010).

Decision-making remains complex, and assessing the resulting risk requires knowledge about the exposure backcountry skiers face. To date, only few studies focus on the exposure, i.e. when, where and for how long individuals are in the terrain. Zweifel et al. (2006) made an attempt to quantify backcountry recreation by using a registration board and automated measuring stations to count backcountry skiers on designated routes. The results are promising, but only suited for small-scale studies. A larger scale study on backcountry activity assessment was not yet carried out.

## 2.3 Leisure and Outdoor Recreation

Activity prediction has been greatly left out in previous studies about backcountry skiing. However, there is research on the prediction of leisure activity, tourism in general and recreational traffic (e.g., King et al., 2014; Lingras et al., 2002; Mach et al., 2020). Additionally, there have been plenty of studies that examine the influence of weather conditions on outdoor recreation (Dwyer, 1988; Verbos et al., 2018), some of which include backcountry skiing.

Backcountry skiing is a recreational activity, often taking place on holidays and weekends, as it requires a half- or even full-day trip. Research on leisure and tourism is therefore well-applicable to backcountry skiing as well. Probably the most similar to backcountry skiing prediction is the prediction of skier days<sup>1</sup> in ski resorts (e.g., King et al., 2014; Riddington, 2002). Ski resorts exhibit a high demand of prediction techniques of skier days, because resorts must adapt their resort capacity planning operations to the changing travel patterns. Even though resort skiing is different from backcountry skiing, there are overlaps in the behavior of both types of skiers. The biggest difference is that in resorts skiers usually do not have to worry about avalanche danger, since resorts plan and prepare their slopes so that a minimum avalanche danger can be guaranteed. Hence the terrain use in resorts is not so much influenced by avalanche danger but by personal preferences, quality of the prepared slope, and or weather conditions (King et al., 2014). Other differences are that people in ski resorts consider resort-specific variables like the price of a ticket and the resort-facilities. However, as we investigate the predictive modelling of skier days and leisure activity in general, we see that many predictive variables among different

---

<sup>1</sup>A skier day in the skiing industry is standard metric defined as a single skier or snowboarder at one resort for any amount of time during one day.

fields are similar and can also be applied to backcountry skiing. King et al. (2014) subdivide the predictor variables for the skier day prediction into three dimensions, with the following, non-exhaustive variables:

- Weather Dimension: Current Snow Depth, New Snow Fall, Average Daily Temperature
- Economic Dimension: Unemployment Rate, Consumer Price Index, Gas Price Index
- Time Dimension: Day of the Week Indicator, Holiday Indicator, Day Number of the Season

The weather and time dimension can also be applied to backcountry skiing. The economic dimension however is disregarded in this thesis, because it is likely much less influential for backcountry skiing, as there is no direct financial aspect in the form of a ski resort fee to it. An additional factor that is not important for resort skiing but is crucial for backcountry skiing is the avalanche forecast, which is certainly an important predictor for backcountry skiing activity. Weather-related studies that focus on the impact on outdoor recreation have increased steadily since 2009 (Lee et al., 2015; Verbos et al., 2018). Rutty and Andrey (2014) found that virtually all skiers and snowboarders access a weather forecast when planning a tour, and that it can even deter them from ultimately going outside. They also found that winter recreationists are especially sensitive to conditions with precipitation, especially to rainfall and freezing rain. Precipitation is often a more palatable condition than for example temperature, as people can instantly feel if it is raining or not, but they might not sense slight variations in temperature. This is in line with Haugom and Malasevska (2019), who found that weather is among the most important attributes for Alpine resort skiers, with the most preferred conditions being sunny weather with no precipitation and an air temperature of  $-5^{\circ}\text{C}$ . It becomes apparent that precipitation is an important factor for not only resort skiing, but also backcountry skiing. In the same study, approximately a third of all participants stated that sun/cloud conditions were important, which is in line with tourist research revealing the sensitivity of tourists to good weather conditions, especially the sunshine duration (Wegelin et al., 2022). Dwyer (1988) used seasonality, day of the week, and weather data to explain fluctuations in the daily use of urban forest recreation sites. This is in line with Wegelin et al. (2022), who showed that the main explanation for fluctuation in touristic activity on Mount Rigi, a famous touristic mountain in Switzerland, was sunny weather followed by the day of the week and seasonality, as well as institutional factors such as vacation periods. The activity was increased by 133% on a sunny day, compared to a bad weather day, and by 80% on a Sunday, compared to a Monday. Similarly to the day of the week, “Activity Day” is a category of variables that is often used in recreational prediction and includes information about the day of the week (work day, weekend) and holidays (Verbos et al., 2018), similar to the “Time Dimension” in the skier day prediction as proposed by King et al. (2014). In conclusion of existing literature in the field of leisure and outdoor recreation, three types of predictors should be included in the prediction of backcountry skiing activity: 1.) Weather-related predictors, 2.) free time-related predictors,

and 3.) avalanche-related predictors.

Lastly, it has to be mentioned that research on leisure and recreation often relies on Volunteered Geographic Information (e.g., Nogueira Mendes and Pereira da Silva, 2018; Santos et al., 2022; Upton et al., 2015). The dataset used in this thesis is also user-generated and falls into the category of VGI, therefore it is important to shed some light on characteristics and issues associated with this specific type of data.

## 2.4 Volunteered Geographic Information

The term “Volunteered Geographic Information” (VGI) was first introduced by Goodchild (2007, p.212), who described it as the “engagement of large numbers of private citizens, often with little [...] formal qualifications, in the creation of geographic information”. Goodchild shaped the idea that humans build a sensor network, where each sensor (i.e., each human), is equipped with five senses, the intelligence to interpret and compile what they sense and the ability to move freely on the Earth’s surface. This immensely large network of sensors has great potential to be a significant source of geographers’ understanding of many different processes that occur on the Earth’s surface.

VGI was enabled by new web technologies developed in the early 2000s, collectively referred to as web 2.0, that allowed users to produce and share content online to the point where it became possible to construct websites that were almost entirely based on user-generated content (Goodchild, 2007). Especially since GPS technology was commonly included in cell phones in the early 2010s, the amount and availability of location-based data experienced a steep increase (Bielański et al., 2018; Hendriks et al., 2022). One of the most prominent VGI applications is Open Street Map (OSM), which is a web service that allows users to create geo-referenced information or contribute to and edit information contributed by other users (Goodchild, 2007). There are also many web services for a more specific target group, for instance skiers, bikers, or hikers, where users share geo-referenced information with their peers in a specific field. An example for such a target-specific web service is Skitouren guru, from which the data used in this thesis emerges.

Despite the great potential VGI data holds, we must be aware that the data is not produced by professionals, but mostly by untrained people. Therefore, the results may or may not be accurate (Goodchild, 2007; Senaratne et al., 2017; Zhang and Zhu, 2018). In contrast to purely scientific data, where the primary motivation for data collection and sharing is to achieve a certain research goal, there are a variety of reasons why non-professionals contribute to VGI. Epstein et al. (2015) conclude five different types of reasons, why people share their personal data online: 1.) The request for information, 2.) a desire for emotional support, 3.) seeking motivation or accountability from audience, 4.) motivating or informing the sharing audience, and 5.) communicating a certain impression to their audience. Additionally, while scientific data collection follows a precise approach (e.g., random, stratified or systematic), data collection through volunteers is oftentimes spontaneous and driven by opportunistic motivations which results in a spatial bias (Zhang and Zhu, 2018). It is important to be aware that there are

different motivations for VGI, as this possibly influences what people choose (not) to share. Therefore, we have to assume that user-generated content always contains a certain bias. A common issue also contributing to this bias is a phenomenon known as participation inequality. It describes fact that in online communities, 90% of the content is produced by only 1% of the users, while 90% of the users observe but never contribute (Nielsen, 2006). Zweifel et al. (2006) found that only a fifth of all backcountry recreationists were willing to use a registration board for the purpose of quantifying backcountry skiing activity, which confirms the participation inequality in the backcountry skiing domain. Additionally, the minority that contributes, may show a different behavior than users that do not contribute. For instance, Furman et al. (2010), Hägeli et al. (2009) and Hendrikx et al. (2022) pointed out that voluntary surveys on avalanche safety primarily attract participants who already have a special interest in avalanche safety, are more engaged and more aware of possible risks, which leads to a more conservative decision-making. Additionally, in the context of a safety survey, people might adjust their answers to more conservative behavior due to social compliance. On the contrary, when people share their content online, rather than in a scientific survey setting, the intent to impress peers by sharing exceptionally sophisticated and hazardous tours, can lead to a bias towards more reckless behavior. Unfortunately, VGI data shared in the setting of an online web service usually does not contain information about motivations behind information sharing. It is impossible to quantify this bias when no personal information about the contributing person is provided. It becomes evident that quality assessment is often impossible with VGI data, which reduces the credibility of this data source (Flanagin and Metzger, 2008).

Despite the limitations VGI data holds, Johnson and Hendrikx (2021) highlight the importance of citizen science techniques, as such detailed and abundant data about the behavior of backcountry skiers would be extremely difficult to obtain using a more traditional, survey-based approach.

## 2.5 Machine Learning

In this chapter, some basic concepts of machine learning (ML) are explained. Unless stated otherwise, the definitions and ideas follow Marsland (2015), who examined machine learning from an algorithmic perspective.

Machine learning is a technique where computers adapt their actions in order to make them more accurate, where accuracy is a measure of how well the chosen actions reflect the correct ones. This enables computers to learn from experience to perform a certain task, and by using this technique, to build an algorithm (Mitchell, 1997). During the learning process, an input is required, which tells the algorithm if it is getting better or not. Based on this input, ML algorithms can be classified into four different classes: 1.) Supervised learning, 2.) unsupervised learning, 3.) reinforcement learning, and 4.) evolutionary learning. Further, ML algorithms can deal with different types of problem set-ups such as classification, regression, clustering, anomaly-detection, and reinforcement learning (Alzubi et al., 2018). In this thesis, supervised learning is applied to solve a classification problem. Therefore, this is going to be explained in

more detail.

The basic principle of supervised learning is that the algorithm learns with a set of examples (training data), where the correct response is provided. In other words, each time the algorithm generates a response, this response is compared to the true response provided in the input data set, thereby determining the effectiveness of the actions taken by the algorithm. An interesting thing to notice is that if the algorithm would be fed with all possible combinations of input features, the algorithm would simply be an extensive look-up table. However, what distinguishes a machine learning algorithm from a look-up table is that the algorithm is able to produce correct outputs for inputs it has not encountered during the learning process. In other words, after learning from a given data set, the algorithm is ideally able to correctly solve the problem for a new, unseen dataset, with new, unseen input combinations. This is called generalization. In classification problems, each data point belongs to exactly one class, and the set of classes build up the whole possible output space. Hence the algorithm is trained to identify to which of the classes a data point belongs.

Machine learning tasks are usually solved by following the same set of steps: 1.) Data collection and feature selection, 2.) algorithm choice and parameter selection, and 3.) training and verification. An overview over how those steps are implemented in this thesis is provided in Chapter 4. In the following, an overview is given over the Random Forest algorithm, which is applied in this thesis to predict backcountry skiing activity.

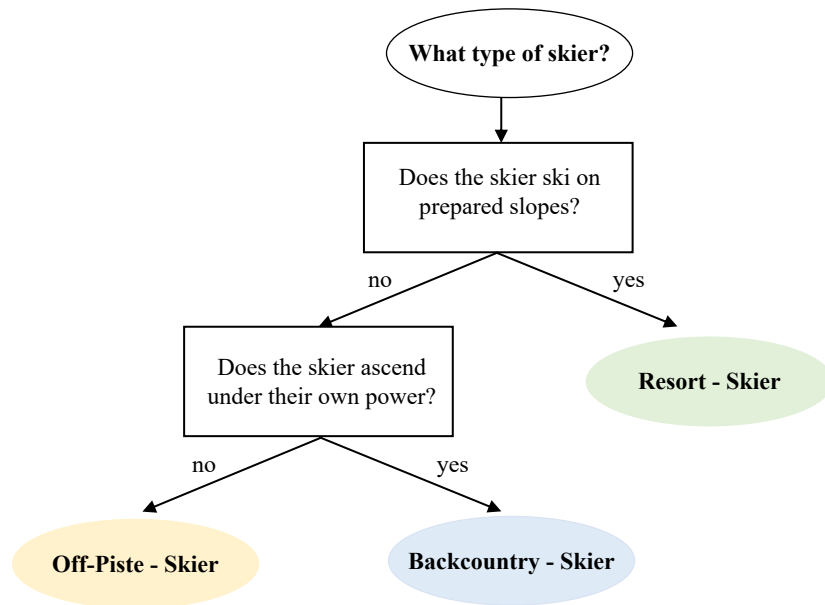
### 2.5.1 Random Forest

Random forest (RF) is an ensemble learning method that solves different types of problems based on the principle of a decision tree (DT). In the example of a classification problem, the DT classifies a data point (consisting of several input features) using questions about the attributes of the data item. The series of questions form a hierarchical tree, in which each node in the tree represents a question, and each child-node represents a possible answer to the question. An example of a decision tree is provided in Figure 2.2. To determine the best possible question (usually referred to as split) on a specific node, a measure quantifying the effectiveness of the split is required. The most common measure is the GINI impurity index, which quantifies the level of purity within each child-node. It reaches its maximum value when all training data points within one node belong to a single class, denoting a purity of 100%. A tree is generated by optimizing each node split in terms of purity of the child-nodes. The data point gets classified by following the tree from the top-node to a leaf-node (terminal child-node), which represents one class of the classification (Kingsford and Salzberg, 2008).

The idea of ensemble learning is that by combining lots of different learners, the result will get better than by using just one single learner. Because each tree is trained with slightly different data, it will produce slightly different results. A set of decision trees put together form a random forest. Every data point is run through all decision trees of a random forest, resulting in  $n$  classifications of  $n$  different trees. Ultimately, the data point gets assigned to the class for which most of the DTs voted (Marsland, 2015). To avoid correlation between the different



decision trees, RF uses a technique called “bagging”. This is where the randomness is introduced into the forest. Bagging means that each DT is trained with a randomly resampled training set. Additionally, for each node in a DT, a randomly selected set of attributes is used. This achieves more stable classification results, which are more robust to slight variations in the input data. Due to the Law of Large Numbers, RFs do not overfit, therefore the number of uncorrelated DTs can be increased for as long as the generalization error decreases (and longer, but this would not be sensible), in order to increase the accuracy (Breiman, 2001; Rodriguez-Galiano et al., 2015). Advantages of the RF algorithm are that it can deal with any data type. Further, it produces robust outputs, even for smaller datasets. Because the classification is based on decision trees, it is, in comparison to some other ML algorithms, relatively comprehensible how the algorithm classifies the data, as opposed to the black-box-nature of e.g., neural networks. Further, due to the bagging, the algorithm can calculate the accuracy on-the-go, by applying each of the decision trees to all the training data samples that have not been used for the given tree. This is called Out-Of-Bag (OOB) accuracy, respectively Out-of-Bag error. Also, since the different trees work independently, the random forest can run on as many different processors as available, which speeds up computational time almost linearly (Marsland, 2015).



**Figure 2.2:** Example of a decision tree to categorize a skier. Black boxes show node splits, while coloured ovals show terminal nodes. The black oval at the top indicates the data input into the decision tree.

### 2.5.2 Imbalanced Data

A dataset is technically considered imbalanced when the distribution among its classes is unequal. Even though there may well be more than two classes, this section focuses on data with only two classes. In the machine learning domain, a dataset is considered imbalanced when it

exhibits a severe imbalance in the order of 1:100 or higher. However, even when dealing with datasets that have imbalances smaller than 1:100, it is worthwhile to examine some of the issues that can arise due to the imbalance.

In a two-class classification setting, data imbalance leads to the presence of a majority class and a minority class. It has to be mentioned that the class imbalance does not necessarily have to be inherent to the nature of the phenomenon, but can also be a result of data collection, storage or time restrictions. Data that is imbalanced due to external factors is called extrinsic imbalance, which stands in contrast to intrinsic imbalance (He and Garcia, 2009).

The basic problem of imbalanced learning is that traditional machine learning algorithms often assume that the distribution among classes is balanced. When fed with imbalanced data, most algorithms fail to yield equally good performance in both the minority and the majority class (Krawczyk, 2016). Usually, the majority class reaches a high degree of accuracy, often near 100%, while the minority class remains undiscovered, ignored or assumed to be noise or outliers, which results in a low degree of accuracy (Ali et al., 2015). Unluckily, the minority class is often more important than the majority class, as imbalanced datasets often occur in domains like fraud detection (Makki et al., 2019), anomaly detection (Jiang et al., 2019) or medical diagnosis (Fotouhi et al., 2019). Misclassification in the minority class, for instance failing to detect cancer or fraud, are often more costly and severe compared to misclassification in the majority class. The wide range of applications and domains, where imbalanced data occurs, calls for solutions. Krawczyk (2016) summarises two main approaches to tackle the class imbalance problem:

### 1) Data-Level Methods

Methods that operate on the data-level make the data suitable for standard learning algorithms. The general goal is to balance the distribution by either oversampling the minority class or undersampling the majority class. The problem with this approach is that there is a possibility that either important information gets lost, or meaningless information is amplified (e.g., Chawla et al., 2002).

### 2) Algorithm-Level Methods

This approach focuses on the modification of existing algorithms to reduce the bias in imbalanced data settings. This can be achieved by penalizing errors differently, based on the class they belong to. This is called cost-sensitive learning. By assigning higher costs (or weights) to minority class errors, the bias towards the majority class can be alleviated (e.g., Zhou and Liu, 2010).

Using a hybrid approach by combining data-level and algorithm-level methods is highly popular and results in robust and efficient learners (Krawczyk, 2016; Krawczyk et al., 2014).

Not only learning, but also verification and performance assessment can be challenging with imbalanced datasets. While accuracy can be an adequate metric for balanced classes, it is less suitable for imbalanced classes. Accuracy can be defined as the ratio of all correct responses to the total amount of responses an algorithm produces. Consequently, it is inverse to the overall

error. In an imbalanced setting, the accuracy will be strongly influenced by the majority class, while the effect of the minority class goes unnoticed. The algorithm achieves maximum overall accuracy by always predicting the bigger class (majority class), simply because it outnumbers the errors in the minority class (Ebert and Milne, 2022). If always the majority class is predicted, the overall accuracy is high, while the minority class accuracy is low (Ali et al., 2015). A solution for this is to look at each class separately, by calculating the accuracy for both classes separately (i.e., sensitivity and specificity). By taking the arithmetic mean of both class accuracies, a balanced accuracy can be calculated, which is not influenced by class imbalance (Marsland, 2015). However, there are many settings in which balancing the success rate of both classes is not enough. Consider a tornado prediction or a cancer diagnosis: A false negative error, i.e., failing to detect cancer or to predict a tornado, is much more severe than a false positive error. In this case, balancing the error rate for both classes does not take into account that the errors in one class are more severe than the errors in the other class. In such cases, the performance metric must assess the gravity of the different error types and needs to penalize them accordingly (Ebert and Milne, 2022). It can be concluded, that it is essential to be aware of any potential class imbalance when training the algorithm, but also when verifying the predictions.

# Chapter 3

## Data

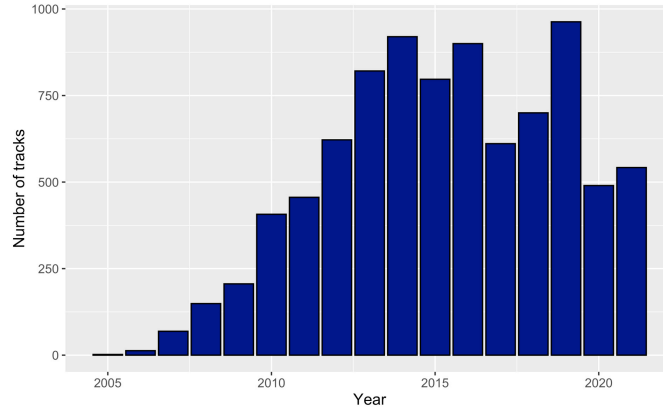
### 3.1 Avalanche Risk Property Dataset

#### 3.1.1 Overview

The core dataset of this thesis is the Avalanche Risk Property Dataset (ARPD) provided by G. Schmudlach (Schmudlach, 2022a). The dataset was pre-processed before it was utilized in this thesis. The methods discussed in this section were all carried out by the data owner. Pre-processing and further methods that were applied in this thesis are described in Section 3.1.3 and Chapter 4.

The ARPD comprises over 8000 GPS trajectories of backcountry ski tracks. Figure 3.1 shows that the first recorded track dates back to 2005, when only very few individuals were tracking and submitting their backcountry ski tour. In the following years, an increase in uploaded tracks can be observed, which is possibly attributed to an increased use of mobile devices capable to record a GPS track (Bielański et al., 2018; Hendrikx et al., 2022). The number of tracks continues to rise until 2013, where it stabilizes at approximately 750 tracks per year. For the analysis, seasons before 2013/14 are excluded due to the low number of tracks per season. This leaves a total number of 6052 tracks that were collected between 2013 and 2021. The trajectories were resampled in 10-meter intervals, which resulted in a total of 6.5 million transition points. Each transition point can be assigned to exactly one tour by a unique identifier, each tour encompasses several hundred transition points. The GPS dataset was intersected with a high-resolution digital elevation model (DHM) and with the avalanche bulletin for the respective day. The final dataset has more than 40 attributes, most of them describing terrain or avalanche conditions. After the calculation of terrain and avalanche attributes for each transition point, the coordinates have been intentionally obfuscated by the data owner. The obfuscation method is described in Chapter 3.1.2. A detailed documentation of the attributes and how they were calculated can be found in the ARPD User Manual (Schmudlach, 2022b).

The tracks were uploaded from contributors to the platforms Skitouren guru<sup>1</sup>, Gipfelbuch<sup>2</sup> or Camptocamp<sup>3</sup>. The data from all three websites is merged, and there is no information on which track was submitted to which website. According to Techel et al. (2015), camptocamp.org is mainly used by French- and Italian-speaking recreationists, while gipfelbuch.ch and skitouren-guru.ch are mainly used by German-speaking users. We must be aware that these two user groups might show different behaviors, which cannot be quantified as there is no information on which which tracks were submitted to which website. Additionally, we must be aware of the participation inequality bias that VGI generally holds, as described in Section 2.4.



**Figure 3.1:** Distribution of yearly recorded tracks in the ARPD.

### 3.1.2 Obfuscation

Although the data was originally submitted as trajectories, containing coordinate information, the data used for this thesis is obfuscated to maintain the privacy of the users. Transitions points do not have any coordinate information but only a Warning Region Code (WRC), which relates each point to a certain warning region (See Section 3.3). Therefore, the warning region is the only location information available for each track. However, plenty of attributes describing the terrain, for instance elevation, slope angle, and exposition allow the inclusion of the terrain and the broader spatial context (warning regions) into the analyses even though the exact location of each point is not provided. To illustrate the structure of the obfuscated data set, an example is presented in Table 3.1.

---

<sup>1</sup>[www.skitouren-guru.ch](http://www.skitouren-guru.ch)

<sup>2</sup>[www.gipfelbuch.ch](http://www.gipfelbuch.ch)

<sup>3</sup>[www.camptocamp.org](http://www.camptocamp.org)

**Table 3.1:** Example of obfuscated ARPD structure.

Identifier Attributes			Terrain Attributes			
Coordinates (x/y)	Track-ID	WRC	Elevation [m a.s.l.]	Slope [°]	Aspect [°]	...
obfuscated	234	1245	1911	32	97	...
obfuscated	234	1245	1924	35	95.5	...

### 3.1.3 Pre-Processing

For the further analysis, the ARPD is aggregated at the track level, resulting in a single data point for each backcountry ski track<sup>1</sup>. After this aggregation, each track consists of the following information and terrain attributes:

#### Information

- ID
- Date
- Year
- Weekday
- Warning region code (see 3.3)

#### Terrain

- Mean track elevation
- Raw danger level

A few tracks span two or more warning regions, as they took place at the borders of warning regions. This leads to multiple raw danger level (RDL) values corresponding to each warning region. To resolve this issue, only the warning region in which most of the track points were recorded (i.e., the statistical mode) was used. This means that the RDL value associated with the region in which most of the track is spent in is used as the representative danger level for the entire track

## 3.2 Raw Danger Level

The avalanche danger in Switzerland is described with a five-level danger scale ranging from low (1) to very high (5). This danger scale is consistent with the European Avalanche Warning Services (EAWS) (EAWS, 2023), and is referred to as raw danger level (RDL) in this thesis. It includes the following levels:

---

<sup>1</sup>Backcountry ski tours and backcountry ski tracks are two different terms that essentially describe the same act. However, track is used in a more technical sense, emphasizing on the GPS-tracking, while the tour denotes the act of backcountry skiing itself and is not necessarily related to GPS-tracking.

- 1 - Low
- 2 - Moderate
- 3 - Considerable
- 4 - High
- 5 - Very High

The basic principle is that with increasing RDL, there is an increase in terms of release probability, the density of potential triggering spots and the potential avalanche size (Schweizer et al., 2020; Winkler et al., 2021). Techel et al. (2022) claim that the five-level danger scale for avalanches is a strong simplification of a continuous, multi-dimensional feature, especially, since the avalanche danger does not increase linearly, but exponentially with increasing danger level. They show that avalanche danger can be more effectively assessed with a refined danger level, which was implemented into the Swiss avalanche bulletin in 2022. The newly introduced sub-levels give information on whether the avalanche danger is towards the bottom end (–), in the middle (=), or towards the top end (+) of the forecast level (Heggli, 2022; Techel et al., 2022). However, to make the raw danger level consistent throughout all seasons analyzed in this thesis, only the RDL, but not the refined danger level is used.

The RDL is included in the avalanche bulletin, which gets published every day at 5 pm in the winter season and is valid for 24 hours. In late spring and early fall, the bulletin gets published sporadically, according to the snow conditions. Besides the RDL, the bulletin includes the following sections (SLF, 2023a):

- Dangerous terrain in terms of elevation and exposition
- Avalanche problems
- A qualitative description of the avalanche danger
- Information about the snowpack
- Weather and measurements
- Recommendation for behavior in the terrain

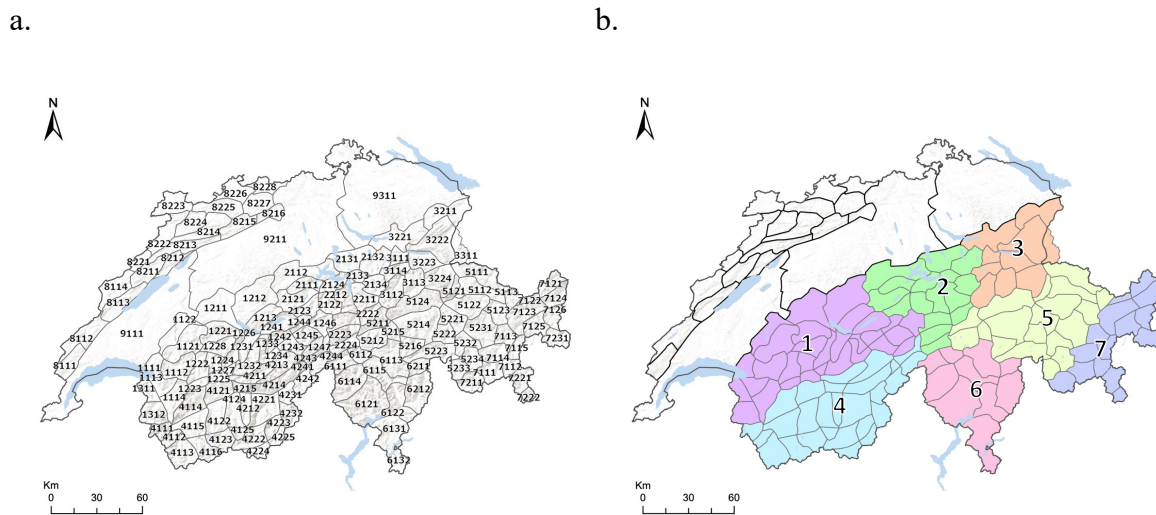
However, as most of this information is provided in text form, only the RDL is used for the analysis.

### 3.3 Warning Regions

The Warning Regions (WR), shown in Figure 3.2, are the smallest spatial units used to communicate avalanche danger by the avalanche forecasting service in Switzerland. Each warning region is identified by a unique warning region code (WRC), where the first digit of the WRC assigns each warning region to a greater climatic region. It is important to acknowledge that in reality, the borders between regions are not discrete but rather represent a gradual transition zone. This implies that there may be some degree of overlap or ambiguity in their delineation and respectively in the avalanche danger level assigned to each region. However, since the ARPD

is obfuscated, we cannot assess the proximity of a given track to the border of the respective warning region. Therefore, the analysis is done on the spatial level of the warning regions, without taking into consideration neighboring warning regions.

After pre-processing the backcountry ski tracks, each track is assigned exactly one WR. In total, there are 149 WRs (Figure 3.2a). The 149 WRs are divided into nine greater climatic regions. The greater regions “8 – Jura” and “9 – Mittelland” are excluded because the focus lies on Alpine regions as a vast majority of tracks take place there (Figure 3.2b). This leaves a total of 128 Alpine warning regions, with an average size of 200 km<sup>2</sup>, located in seven greater climatic regions. Of the total number of 6052 tracks (2013 – 2021), only 38 tracks lie in the regions “Mittelland” and “Jura” and are therefore excluded, leaving a total of 6014 tracks for further analysis (Table 3.2).



**Figure 3.2:** a. Warning regions, the smallest spatial units used to communicate avalanche danger in the avalanche forecasts in Switzerland. Each warning region is labelled with its warning region code (WRC). b. Greater climatic regions as defined by the first digit of the WRC. 1 – Westlicher Alpennordhang, 2 – Zentraler Alpennordhang, 3 – Östliche Alpennordhang, 4 – Wallis, 5 – Nord- und Mittellanden, 6 – Zentraler Alpensüdhang, 7 – Engadin/ östlicher Alpensüdhang. Regions 8 – Jura and 9 – Mittelland are excluded. All warning regions, including names, can be found in Appendix A.

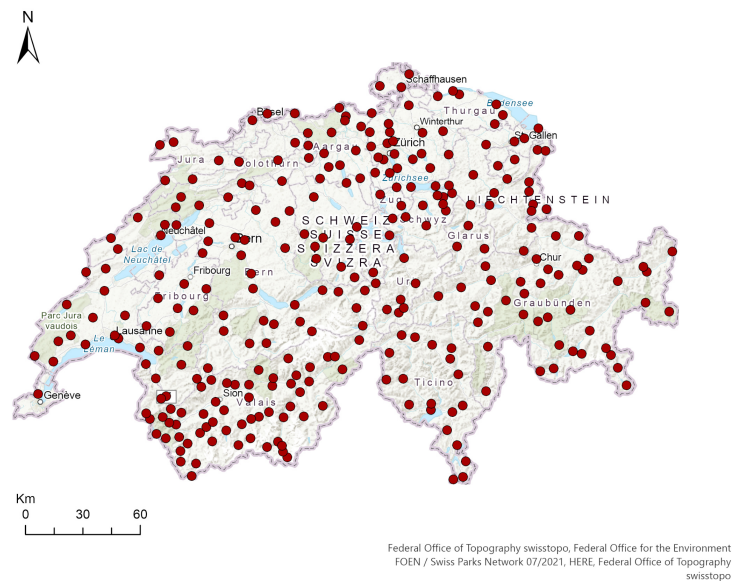
**Table 3.2:** Number of tracks used for the analysis.

Total tracks (2007 – 2021)	<b>8668</b>
Tracks since 2013/14	<b>6052</b>
Tracks since 2013/14 “Mittelland” and “Jura” excluded	<b>6014</b>



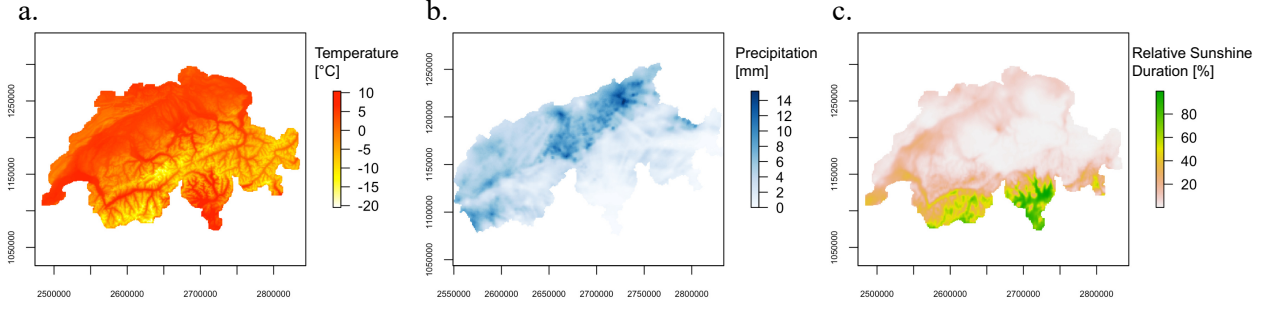
### 3.4 Weather Data

The weather data used in this thesis is provided by MeteoSwiss<sup>1</sup>. Even though the weather forecast rather than the actual weather measurement might determine the skiing activity, measurement data is used in this thesis due to simplicity and accessibility of the data. In Europe, while occasionally poor to medium-range accuracies may arise in weather forecasting, the overall accuracy remains high. Studies indicate that the correlation scores between forecasted and actual weather typically fluctuate around 80% (Rodwell et al., 2013). It is therefore assumed that the weather forecast of the evening before has a sufficiently high correlation with the observed conditions on the following day, which justifies the choice of the measurement data. The meteorological variables that will be used are temperature, relative sunshine duration, and precipitation. Wind data is not included in the analysis, as it would be too complex to analyze, and there is no access to data that captures the influence of wind in a meaningful spatial resolution. The meteorological variables were measured at the operational station network SwissMetNet (Figure 3.3) and further interpolated on a 1 km grid (Figure 3.4). An overview over the used variables is given in the next sections. A detailed documentation of each variable can be found in the documentation of MeteoSwiss Spatial Climate Analyses (MeteoSwiss, 2021e). All data is provided in a netCDF (Network Common Data Form) format, which is a standard format for grid-based climate data.



**Figure 3.3:** Operational station network SwissMetNet, which comprises approximately 160 automatic measuring stations, including approximately 100 automatic precipitation measurement stations. Basemap: swisstopo.

<sup>1</sup>[www.meteoschweiz.admin.ch](http://www.meteoschweiz.admin.ch)



**Figure 3.4:** Example data showing the interpolated, gridded meteorological variables used in this thesis, with a spatial resolution of 1 km. a. Temperature, b. Precipitation, and c. Relative sunshine duration for January 11, 2015. In b., ray-like structures are visible, which are artefacts of the RADAR image. Coordinates correspond to the Swiss Grid LV95.

### 3.4.1 Air Temperature

The daily mean air temperature 2 m above ground level is used for the analysis (MeteoSwiss, 2021a). For this data product, a deterministic analysis method has been employed, which specifically addresses challenges in high mountain interpolation, such as the non-linear temperature variations with topographic height and marked horizontal gradients. Nevertheless, interpolation inaccuracy is highest in Alpine regions and in the winter season, with a mean absolute error of 1.5°C. Large errors occur in inner Alpine valleys (up to 4°C), as those valleys are typical cold-pool environments, and the temperature gets systematically over-estimated (Frei, 2014). However, since backcountry ski tours usually do not take place in the valleys, this is not assumed to impact the analysis.

### 3.4.2 Relative Sunshine Duration

For the sunshine duration, the relative sunshine duration data product is used (MeteoSwiss, 2021c). The relative sunshine duration ( $S_{rel}$ ) is the ratio between the effective sunshine duration and the maximal possible sunshine duration with clear sky conditions determined for each calendar day. A sunshine period is defined as a period, where the direct solar irradiation exceeds 200 W/m<sup>2</sup>.  $S_{rel}$  is calculated using a statistical technique, which combines station measurements with high-resolution satellite-based clearness indexes. The median of the mean absolute error over a 10-year cross validation period is 10% for winter days, with an interquartile range of 6% - 14% (Frei et al., 2015). A relative, rather than an absolute measure, such as total irradiation, is chosen, because seasonal variations are filtered out in the relative sunshine duration calculation. This makes it easy to compare backcountry ski tracks that were recorded in different seasons.

### 3.4.3 Precipitation

CombiPrecip is a grid-based data product provided by MeteoSchweiz and is used as an approximation for precipitation in this thesis (MeteoSwiss, 2021d). CombiPrecip provides precipitation

fields, which are a geospatial combination of raingauge measurements and RADAR estimates. Detailed documentation of the interpolation and combination process can be found in the documentation of the CombiPrecip data product (MeteoSwiss, 2021e). The spatial resolution is 1 km<sup>2</sup> and the temporal resolution is 1 hour. The hourly rather than the daily precipitation sum is used because most backcountry ski tours take place in the early morning. There are four days within the observed period of 8 years, where no CombiPrecip data is available. For those days, the daily precipitation sums are used to fill the gaps (MeteoSwiss, 2021b).

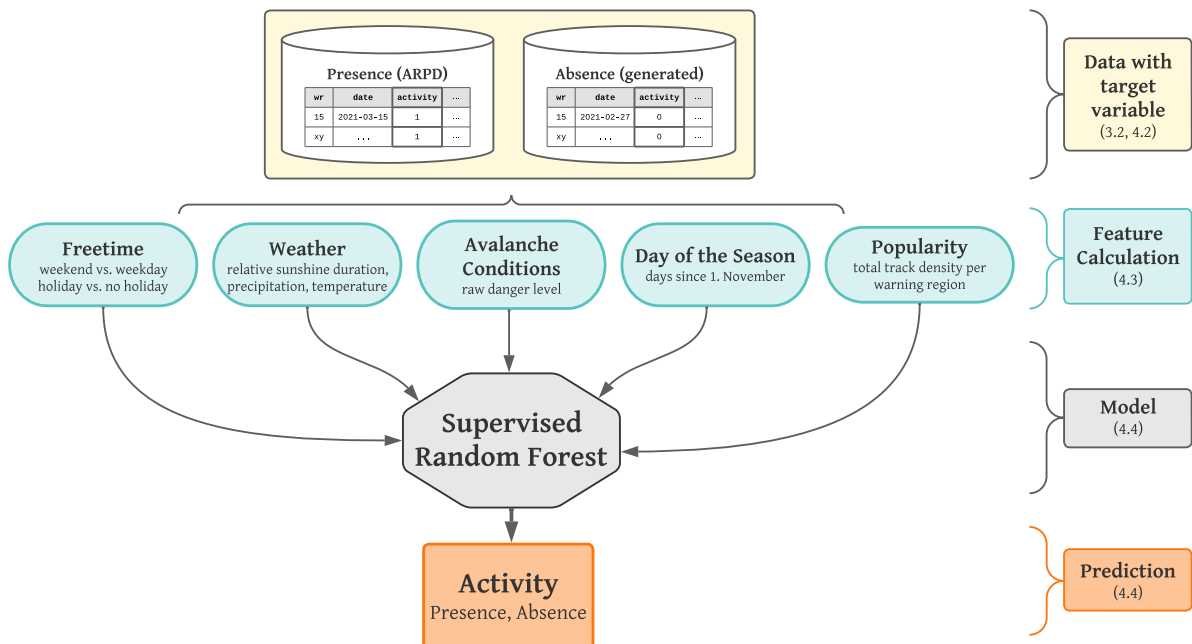
### 3.5 Digital Elevation Model

For the analysis, a digital elevation (height) model (DHM) with a 200 m spatial resolution was used. It originates from the DHM25 provided by the Federal Office of Topography, which is a raster-based elevation model of Switzerland with a 25 m spatial resolution (Swisstopo, 2005). The DHM25 is derived from the height information of “Landeskarte 1:25’000” (LK25), a national map of Switzerland. The mean accuracy of the DHM25 ranges from  $\pm 2$  m to  $\pm 3$  m in the Alpine region. The 200 m resolution of the DHM is used in this thesis rather than the finer 25 m resolution. The elevation information does not require a high level of precision, as it will be only used to specify a broader elevation band for each track. By opting for the 200 m resolution, a balance was achieved between computational efficiency and the accuracy required for the analysis.

# Chapter 4

## Methods

This chapter provides a detailed overview over the methods applied in this thesis. Since the goal of the thesis is to model and predict backcountry skiing activity, it is important to understand the conceptual idea behind the modelling process in order to follow the structure of this chapter. The conceptual background of the model is therefore introduced at the beginning of this chapter (Section 4.2). In Section 4.3, the generation of absence data is described, followed by the data enrichment of both presence and absence data in Section 4.4. Section 4.5 ultimately explains the modelling process in detail. A visual overview over the modelling process is provided in Figure 4.1.



**Figure 4.1:** Methodology of modelling and prediction. Numbers in brackets indicate the section in which the respective part of the process is discussed.

## 4.1 Software

All analyses described in Chapter 4 were performed using R version 4.2.2<sup>1</sup> and R Studio version 2022.12.0+353<sup>2</sup>. A list of the relevant R-packages used for the analyses is provided in Table 4.1. The notation for R-packages and -functions used in this thesis is *package::function()*, which indicates that *function()* is implemented (*::*) in *package*. The majority of visualizations are created with R and RStudio, however, some visualizations are produced with ArcGIS Pro version 2.9.6<sup>3</sup>.

**Table 4.1:** Most important R Packages with description and reference.

R-Package	Usage	Reference
<b>BAMMtools</b>	Natural Breaks Algorithm	Rabosky et al., 2014
<b>caret</b>	Random Forest Utilities	Kuhn, 2022
<b>dplyr</b>	Data Manipulation	Wickham, François, et al., 2023
<b>exactextractr</b>	Fast Raster Value Extraction	Baston, 2022
<b>ggplot2</b>	Visualization	Wickham, 2016
<b>hydroTSM</b>	Hypsometry Calculation	Zambrano-Bigiarini, 2020
<b>lubridate</b>	Date-Time Transformations	Grolemund and Wickham, 2011
<b>ncdf4</b>	netCDF Manipulation	Pierce, 2023
<b>randomForest</b>	Random Forest Algorithm	Liaw and Wiener, 2002
<b>raster</b>	Raster Manipulation	Hijmans, 2023a
<b>rgdal</b>	Spatial Data Manipulation	Bivand et al., 2023
<b>sf</b>	Spatial Data Manipulation	Pebesma, 2018
<b>terra</b>	Spatial Data Analysis	Hijmans, 2023b
<b>tidyr</b>	Data Manipulation	Wickham, Vaughan, et al., 2023

## 4.2 Modelling Backcountry Skiing Activity - Conceptual Background

The core dataset used to model backcountry skiing activity is the ARPD, which is described in Section 3.1. Because the data is obfuscated, therefore containing no exact location information, the spatial units of the predictions are the warning regions introduced in Section 3.3. The backcountry skiing activity has a daily resolution, hence the goal is to predict the activity per warning region and day. The 6014 backcountry ski tracks span over eight winter seasons and 125 warning regions. Hence, in a majority of warning regions, no more than five tracks were

---

<sup>1</sup>[www.r-project.org](http://www.r-project.org)

<sup>2</sup>[www.rstudio.com](http://www.rstudio.com)

<sup>3</sup>[www.esri.com](http://www.esri.com)

recorded each season. Accordingly, it rarely occurs that more than one track is recorded in a warning region on a specific day. This suggests that a multi-level daily prediction of the activity, e.g., ranging from “low activity” to “high activity”, is not possible for this type and quantity of data. For this reason, the target variable is treated as a binary variable, where 1 indicates activity (i.e., presence) and 0 indicates no activity (i.e., absence). The basic assumption for this approach is that there is backcountry skiing activity when at least one track is recorded. Consequently, all tracks in the ARPD serve as and are in the remainder of this thesis referred to as presence data or presence tracks. Owing to this methodology, the counterpart to the presence data, the absence data, needs to be generated since there exists no data providing evidence that there are no backcountry recreationists in the field. Therefore, an absence track is generated for each warning region and day where no track was recorded. The underlying premise is that the absence of data indicates the absence of individuals in the field, which indicates that there is no backcountry skiing activity (i.e., absence of evidence = evidence of absence). It must be noted that, following this methodology, there is no more than one absence track per day and warning region. On the other hand, in the presence data there is no such restriction, leading to some days where more than one track is recorded in the same warning region. However, since the absence data is generally much more abundant than the presence data, no duplicates are generated for the absence data.

For both the presence and the absence data, the same set of features (independent variables<sup>1</sup>) is calculated, which is used to predict the activity (target variable). The presence and absence data<sup>2</sup> together form the training data, which is further used in a supervised machine learning approach to predict backcountry skiing activity.

### 4.3 Generation of Absence Data

To create the absence data, all possible date – warning region pairs are generated. Spanning over 125 warning regions and eight winter seasons, this results in a total of 180'500 date – warning region combinations. No absence tracks were generated for the three warning regions, in which no track was recorded during any of the eight seasons and therefore a prediction is not possible. These warning regions are “1122 - Gruyère”, “4241 - Reckingen”, and “5215 - Val Sumvitg”. Furthermore, summer months between May 15 and November 15 are excluded. In a second step, all date – warning region combinations that already exist in the ARPD are excluded, leaving a total of 176'571 absence tracks. Consequently, the ratio between absence and presence data is approximately 30:1.

Every presence track is associated with terrain information, such as the elevation, from which an approximate location in the WR can be inferred. Naturally, absence tracks contain no terrain

---

<sup>1</sup>Hereafter, features and (independent) variables are used interchangeably.

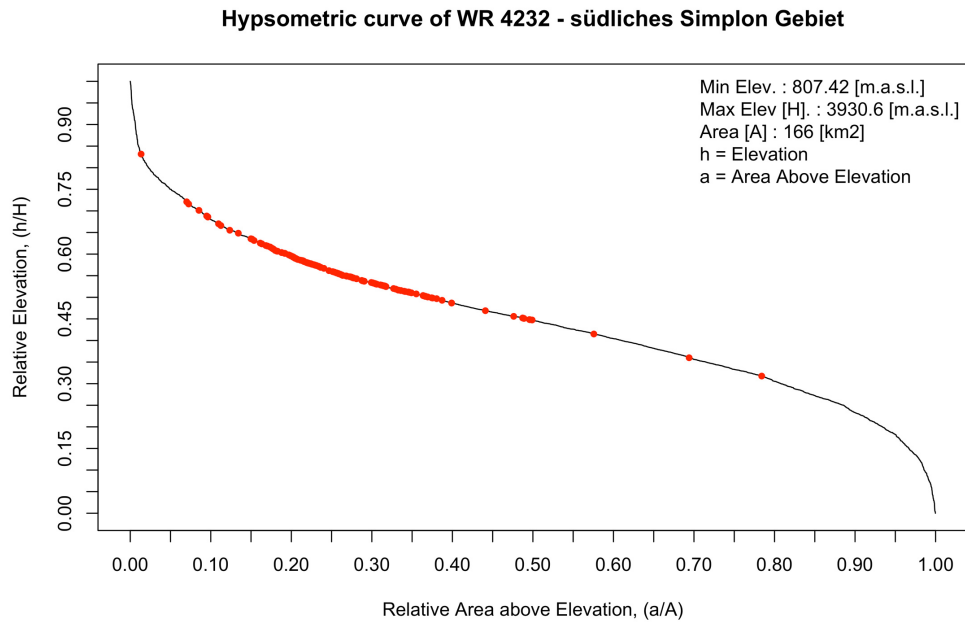
<sup>2</sup>Hereafter, presence (respectively absence) data and presence (respectively absence) tracks are synonymous and used interchangeably.

information, as they never took place. Therefore, every absence track requires an artificially generated reference elevation, which is used to join the meteorological variables to the absence tracks. This artificially created elevation is further referred to as reference elevation. The reference elevation is not generated randomly but depends on 1.) the warning region and 2.) the existing tracks in this specific warning region. To assess the topographic features of each warning region, a hypsometric curve is calculated for each warning region, using the DHM. The hypsometric curve shows the cumulative elevation frequency for a given region. Existing tracks (presence tracks) are plotted on the hypsometric curve to reveal in which elevation backcountry ski tours usually take place in this region.

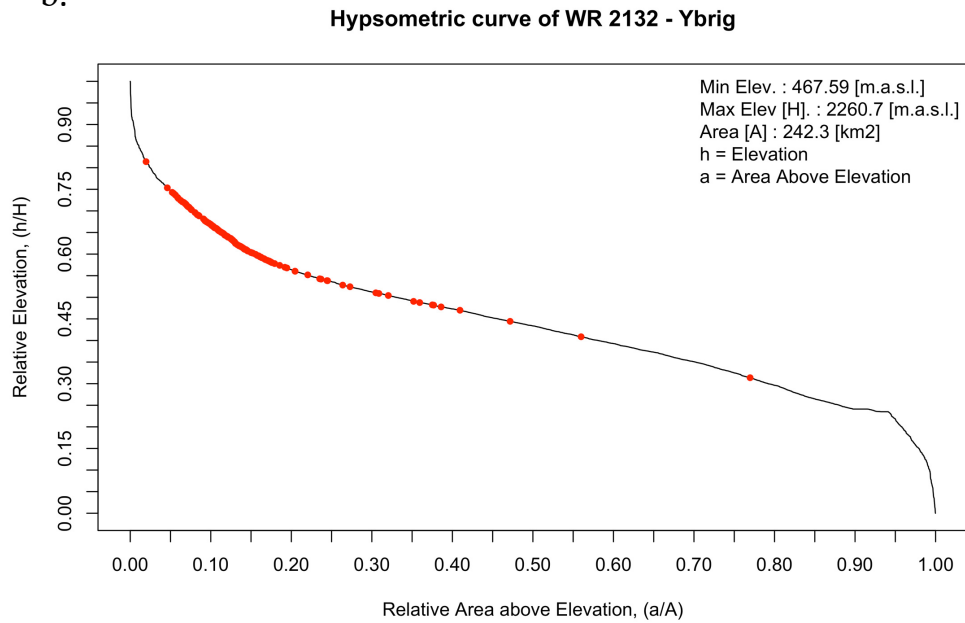
In Figure 4.2 two representative examples of hypsometric curves are given, which indicate that backcountry ski tours are often concentrated in a narrow elevation band specific to the warning region. It also shows that in generally lower elevated warning regions, backcountry ski tours often take place in the higher elevations of the regions (Figure 4.2b), whereas in higher elevated regions, backcountry ski tours can (but do not necessarily have to) take place in mid-range elevations (Figure 4.2a). This illustrates that it is critical to consider the presence tracks to generate a reference elevation for the absence tracks rather than taking for example the mean elevation of the warning region, or even a random value. An elevation belt is defined for every warning region, which spans from the 5% percentile to the 95% percentile of all tracks recorded in this warning region. Every absence track is assigned a random reference elevation that lies in the elevation belt of the respective warning region. An example that shows the artificially generated elevation of absence tracks compared to the elevation of real tracks is given in Figure 4.3.

After generating and processing the absence data, both presence and absence data is structured in the same way (Table 4.2), so that they can be enriched with additional features.

a.

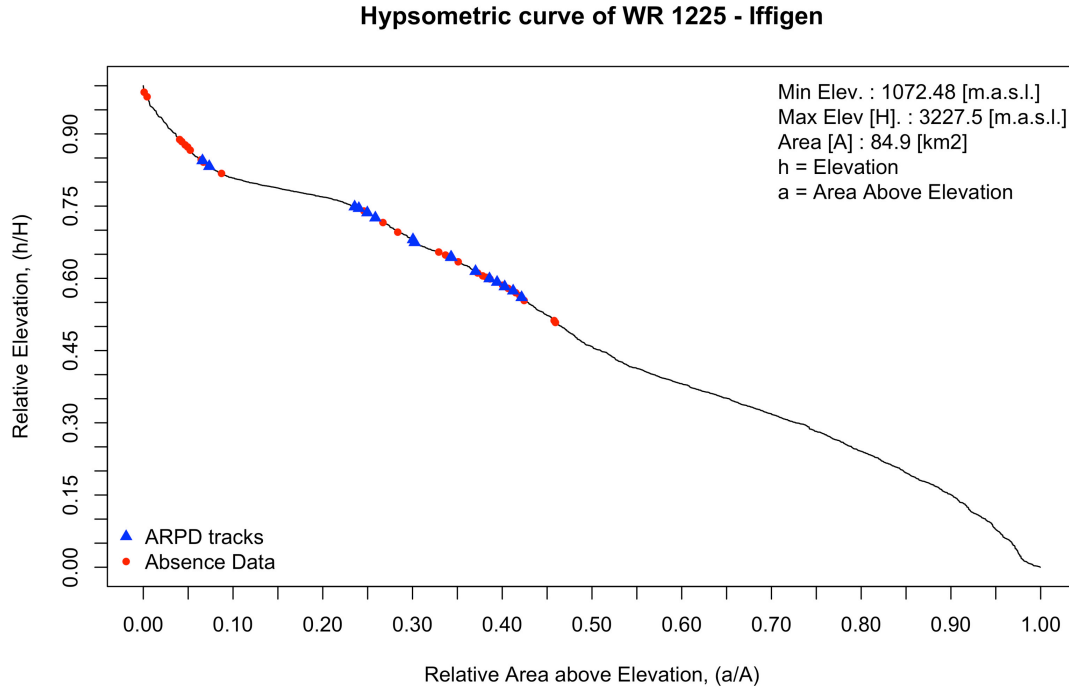


b.



**Figure 4.2:** Examples of hypsometric curves for the warning regions a. 4232 – Südliches Simpon Gebiet and b. 2132 – Ybrig. To make different warning regions comparable, axes show the relative elevation (y) and area above respective elevation (x), rather than absolute values. Warning region specific information is provided in the top right corner. Red dots indicate the mean elevation of existing tracks. For better readability, only a sample of all tracks in these regions is displayed.





**Figure 4.3:** Hypsometric curve of warning region 1225 – Iffigen showing the mean elevation of presence tracks (red dots) and artificially generated elevation of absence tracks (blue triangles). To make different warning regions comparable, axes show the relative elevation (y) and the relative area above respective elevation (x), rather than absolute values. Warning region-specific information is provided in the top right corner. For better readability, only a sample of both groups is displayed.

**Table 4.2:** Example structure of presence and absence data before data enrichment with other attributes.

Date	WRC	Reference Elevation	Other Attributes	Activity
2019-01-23	1245	1934	...	presence
2019-03-02	1222	2120	...	absence

## 4.4 Data Enrichment with Predictor Variables

This section illustrates how both presence and absence data is enriched with additional features that are later used to predict backcountry skiing activity. Meteorological features are described in Section 4.4.1. In the following, popularity (4.4.2), raw danger level (4.4.3), holiday (4.4.4) and day of the season (4.4.5) are discussed. The raw danger level is already present in the presence data, therefore, Section 4.4.3 is only relevant for the absence data.

#### 4.4.1 Meteorological Features

This section gives an overview over the meteorological variables. Essentially, they are all calculated using the same procedure. However, since precipitation, unlike temperature and  $S_{rel}$ , has a temporal resolution of 1 hour, some additional steps at the beginning are necessary. Therefore, this section is divided into the subsections “Precipitation” and “All Variables”.

##### Precipitation

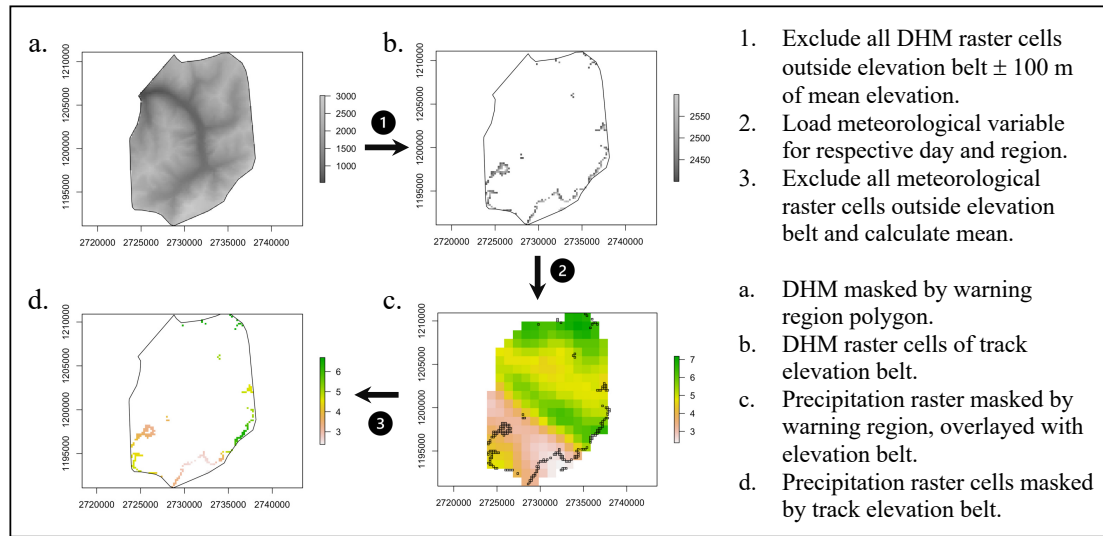
For precipitation, two data sources are available: CombiPrecip and the daily precipitation sum (Section 3.4.3). Because most backcountry ski tours take place in the morning between 7 am and 12 pm, it is crucial to differentiate between morning precipitation and the daily precipitation. This is particularly important on days where the weather in the morning is clear but there is a substantial amount of precipitation later in the day. Because the CombiPrecip data provides hourly temporal resolution, it is chosen over the daily precipitation sums. Daily precipitation sums are only employed on the few days, where CombiPrecip data is missing. On those days, only minimal or no precipitation at all occurred. Under these circumstances, the daily estimates are assumed to adequately represent the morning precipitation.

To calculate the precipitation in the morning, precipitation rasters for each time step between 7 am and 12 pm are summed up for each day. This results in only one raster per day, representing the morning precipitation. For the other meteorological variables (temperature and  $S_{rel}$ ), only daily measurements are available, making a differentiation between the morning and the whole day impossible. Therefore, for both temperature and sunshine duration, a daily average is used.

##### All Variables

Every track is enriched with three meteorological variables: Air temperature, precipitation, and relative sunshine duration. Exact locations of tracks are obfuscated, which makes it impossible to make an exact, location-based join of meteorological data. The only geographic reference is the warning region and the mean elevation (artificial reference elevation) for the presence (absence) tracks. Precipitation and temperature are meteorological features known to change with topographic height (Spreafico and Weingartner, 2005). The relationship between sunshine duration and altitude is not always straightforward, yet there are certain situations where there is clearly a strong variation with altitude. For example, in the presence of fog or low stratus, which is a common meteorological situation in Switzerland in the wintertime,  $S_{rel}$  can be reduced to 0% at ground level but might be near 100% above the stratus clouds or fog (Scherrer and Appenzeller, 2014). It becomes evident that for meteorological variables, not all locations in the warning regions are equally representative of a track that takes place in a specific elevation range. Particularly in the Alps elevation varies greatly, even at small distances. It is therefore crucial to take the elevation into account to ensure that meteorological attributes are representative of the track. To achieve this, meteorological attributes are calculated based on grid points lying in an elevation band  $\pm 100$  m of the mean elevation (artificially generated elevation) of the presence (absence) data within the respective warning region. This is done by applying the

procedure outlined in Figure 4.4 to all three variables.

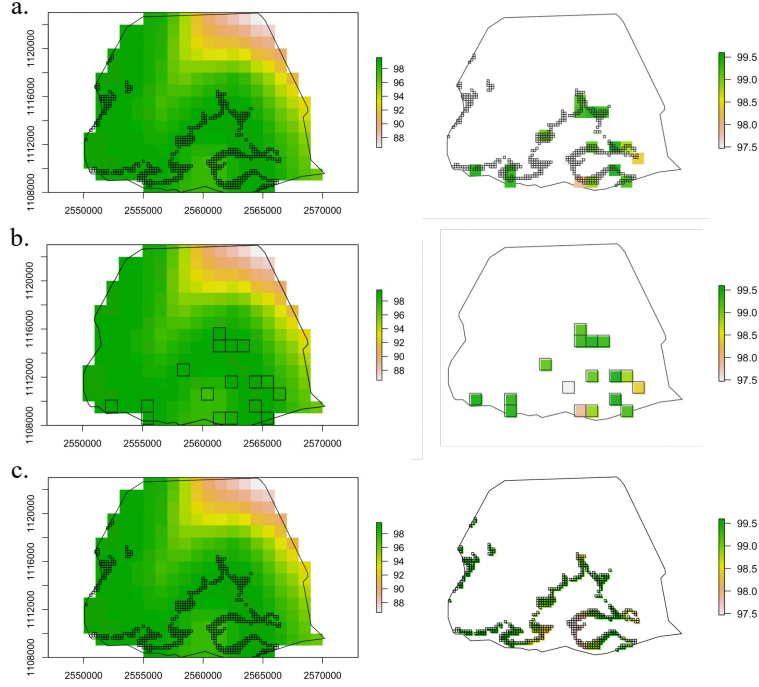


**Figure 4.4:** Procedure to calculate meteorological variables for a given track on the example of  $S_{rel}$ . Procedure is repeated for every track and meteorologic variable (Temperature, Precipitation,  $S_{rel}$ ). a. - d. show different states of the calculation, 1. - 3. indicate different processes.

Figure 4.4 can be summarized in three steps:

1. A 200 m resolution DHM is masked with the polygon of the warning region (Figure 4.4a).
2. All raster cells that do not fall within the 200-m elevation belt around the mean track elevation are excluded (Figure 4.4b).
3. The remaining raster cells are polygonised and used as a mask for the meteorological input variable (Figures 4.4b & d), which is precipitation in the example of Figure 4.4.

Step 3 is carried out using the function `raster::mask()`, which masks a raster (the meteorological variable) with a polygon (the polygonised elevation belt). In this function, a raster cell is only considered to be inside a polygon when the center of the cell lies inside the polygon. This can cause problems, when the raster and the DHM, which is used to create the elevation belt polygon, have different spatial resolutions. When a 1 km raster cell is overlaid by polygons emerging from a 200 m raster, most of the 1 km raster cells will not be masked by the polygon, even though the polygon touches the raster cell (Figure 4.5a). One solution to this problem is to use a DHM with the same resolution as the meteorologic input variable (1 km). However, this leads to an oversimplification of the topography and consequently the elevation belt (Figure 4.5b). Therefore, an intermediate step is required, in which the raster of the meteorological variable is resampled to match the spatial resolution of the DHM (200 m). Visually, the pattern does not change, but each 1 km raster cell is replaced by 5 x 5 raster cells with a spatial resolution of 200 m. The resampled cells masked by the elevation belt (Figure 4.5c) are further used to calculate a mean value that is representative of the track.

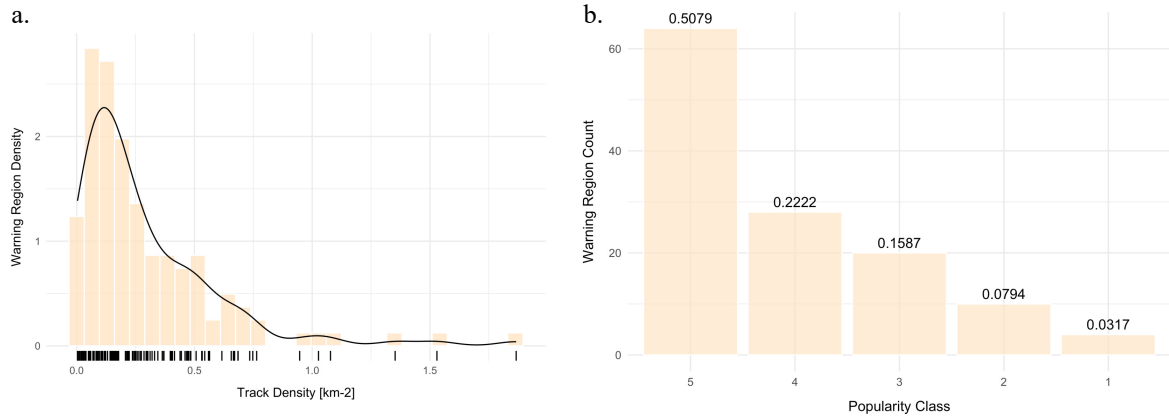


**Figure 4.5:** Application of `raster::mask()` with different spatial resolutions of  $S_{rel}$  [%] (colour) and DHM25. The black cells show the polygonized, track-specific elevation belt emerging from DHM25 for an example track. Figures on the left show the  $S_{rel}$  raster of the whole warning region overlaid with the elevation belt, figures on the right show the raster cells of  $S_{rel}$  selected by the `raster::mask()` function overlaid with the elevation belt. a.  $S_{rel}$ : 1 km, DHM: 200 m. b.  $S_{rel}$ : 1 km, DHM: 1 km. c.  $S_{rel}$ : 200 m, DHM: 200 m.

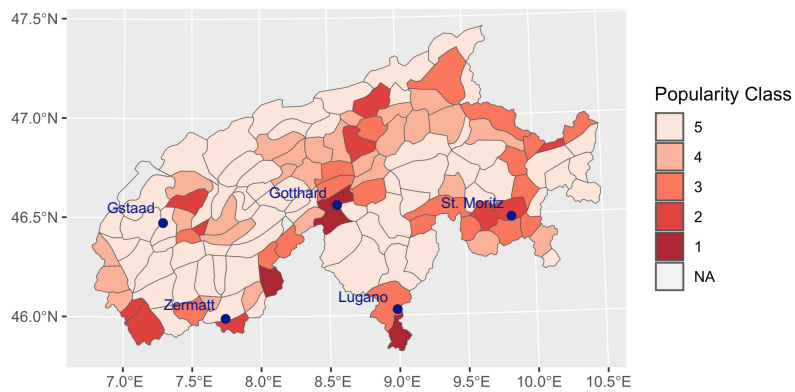
#### 4.4.2 Popularity

Techel et al. (2015) found that backcountry skiing is a spatially variable phenomenon, often concentrated near population centers. The proximity to a population center, often associated with accessibility, impacts the popularity of a region, which in turn influences the activity. The popularity is therefore an important, time-independent factor that needs to be considered. The basic principle to assess the popularity of a warning region is that a higher overall track density indicates higher popularity. The track density is calculated for each warning region, using all tracks within the eight-year period of the analysis. Track density values are further classified into five groups, using the Jenks natural breaks algorithm (`BAMMtools::getJenksBreaks()`). Figure 4.6 shows the density curve as well as the histogram of the five different popularity classes. Class 5 is the least popular and at the same time the biggest class, while class 1 is the most popular but smallest class. Only 3% or 4 regions out of all warning regions fall into class 1, while 50% fall into class 5. This shows that tracks are highly concentrated in only a few warning regions. In Figure 4.7 it is visible that popularity is to some degree spatially autocorrelated. Popularity hotspots are among others the Engadin valley around St. Moritz, parts of southern Valais including Zermatt and Gstaad, the central Alps including the Gotthard area, and southern Ticino around Lugano. All those hotspots are popular touristic regions

offering good accessibility and infrastructure, which is a possible explanation for the high track density. Less popular regions are situated at the northern Pre-Alps and generally the western Alps.



**Figure 4.6:** a. Histogram of warning region track densities. Each black horizontal line represents one warning region. Track densities are calculated based on all recorded tracks from season 2013/14 to season 2020/21. A smoothed density curve is indicated in black. b. popularity classification into five classes based on track density, ranging from 1 = high popularity to 5 = low popularity. The fraction of warning regions in the respective class is indicated on top of every column.



**Figure 4.7:** Spatial distribution of popularity classes of Alpine warning regions in Switzerland. Popularity ranges from “1 = high popularity” to “5 = low popularity”. Touristic hotspots are displayed in blue. In three regions, no track was recorded in the observation period, therefore no value was calculated (indicated by NA). All warning regions, including names, can be found in Appendix A.

### 4.4.3 Raw Danger Level

The raw danger level was already added to the ARPD by the data owner. Therefore, only the absence data needs to be enriched with RDL. For this, the avalanche bulletin<sup>1</sup> data provided by the WSL Institute for Snow and Avalanche Research SLF is used (F. Techel, personal communication, January 31, 2023). In lower elevated regions at the northern and southern edge of the Alps there are days in the beginning and the end of the season, but at times also during the season, where there is none or very few snow. On those days, no avalanche bulletin is issued, which leads to many days without a RDL value in the absence data. In the presence data as well there are a few days with no RDL value. In the R programming environment, no data is represented by NA, which is an abbreviation for “not applicable”. In the remainder of this thesis, NA is used to indicate no data values. How the NA values are further dealt with in the modelling process is described in Section 4.5. An overview over NA-values in presence and absence data is provided in Table 4.3.

**Table 4.3:** Overview over presence and absence data.

Presence Tracks	<b>6014</b>	
Presence tracks with RDL = NA	<b>7</b>	(0.1%)
Absence Tracks	<b>176'571</b>	
Absence tracks with RDL = NA	<b>25'514</b>	(14%)

### 4.4.4 Holiday

Holidays were queried through the Date.Nager API<sup>2</sup>, which provides holiday dates of over 100 countries. Included are the following Swiss holidays:

Weihnachten (25.12)	Stephanstag (26.1)	Neujahr (1.1)	Berchtoldstag (2.1)
Karfreitag	<i>Ostersamstag</i>	<i>Ostersonntag</i>	Ostermontag
Tag der Arbeit (1.5)	Auffahrt	<i>Auffahrtsbrücke</i>	<i>Pfingstsamstag</i>
<i>Pfingsonntag</i>	Pfingstmontag		

Holidays in *italics* are unofficial holidays that are adjacent to official holidays. They are either weekends, or workdays used as a “Brückentag”, which is a day between a weekend and a holiday that many people like to take off for an extended weekend. It has to be noted, that all holidays except for the Swiss national day (August 1) are regulated on a Cantonal<sup>3</sup> level. Therefore,

<sup>1</sup>[www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin.html](http://www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin.html)

<sup>2</sup><https://date.nager.at/api>

<sup>3</sup>Cantons are the political entities that form the Swiss Confederation. In total, there are 26 Cantons in Switzerland.

the holidays used in this thesis do not necessarily apply to all Cantons. The holidays used in this thesis are among the most popular in all Cantons. However, there are possibly holidays that are not included but might be important in specific Cantons, especially since religious denominations differ among Cantons.

#### **4.4.5 Day of the Season / Season**

The “day of the season” feature is an indicator of how advanced the season is. It is calculated as days since November 1 of each season. The values span from day 1 (beginning of season) up to day 200 (end of season). In other words, it is a date attribute without the year component and with a different starting date. This transformation of the date makes the temporal component of tracks of different seasons comparable. Additionally, a season attribute is included, which assigns each track to the season it took place in. This is done to see how (dis)similar the different seasons are, but is not included in the final model. A high influence of the season attribute on the prediction outcome would suggest that there are critical differences among seasons.

### **4.5 Model Building**

This section is structured as follows: In the beginning, an overview over target variable and features is given (Section 4.5.1). Further, two different approaches to improve the model are discussed: 1.) A data-specific approach (how is the data implemented?) (Section 4.5.2) and 2.) an algorithm-specific approach (how is the algorithm implemented?) (Section 4.5.3). In the last section (Section 4.5.4), the verification process is explained and discussed.

#### **4.5.1 Target Variable and Features**

To predict backcountry skiing activity a supervised random forest algorithm is applied. The target variable is further referred to as the activity variable: “Presence” indicates that the data point belongs to the presence data and therefore indicates activity, while “absence” indicates no activity. Table 4.4 gives an overview over the features used to predict the target variable. Some features are implemented in different ways, which is indicated by Implementation 1 and Implementation 2. The different implementations are more closely discussed in the next section.

**Table 4.4:** Target variable and predictors (i.e., dependent variables or features) with different data-specific implementations for modelling.

Target variable	Feature	Implementation 1	Implementation 2
Activity	Weekday	Nominal (7 levels), one hot encoded	Nominal (binary)
	Meteorological features	Ordinal (classed)	Numeric (continuous)
	Day of the season	Ordinal (classed)	Numeric (integer)
	Raw Danger Level	Ordinal, NAs excluded	Ordinal, NAs set to 0
	Precipitation	CombiPrecip-NAs excluded	CombiPrecip-NAs enriched with daily precipitation
	Holiday	Nominal (binary)	-
	Popularity	Ordinal	-
	Warning Region	Numeric (integer)	-

#### 4.5.2 Data-Specific Adjustments

Predictor variables (features) vary in their scale of measurement as well as in their number of categories. Meteorological features are continuous, while most other features are either ordinal or nominal. Some features are binary (e.g., holiday), while others can take on 200 different values (day of the season). Strobl et al. (2007) pointed out, that this type of data can lead to a bias in the variable importance measures of the random forest algorithm. Therefore, in one implementation some features are transformed to equalize the number of categories among features. For features in which there is no relationship among the different categories (i.e., nominal features), one hot encoding is applied. One hot encoding creates a new feature from every category of a categorical feature. The new features are binary representations of the categories of the old feature. It is usually applied for algorithms, which cannot directly deal with categorical, text-based features, as it transforms features from text-based to numeric features (Dahouda and Joe, 2021). However, since random forests can deal with text-based data, one hot encoding is used as a way to reduce and thus equalize the number of categories among features. In the first implementation, numeric features (precipitation, relative sunshine duration, temperature, and day of the season) are classified into different categories based on the Jenks natural breaks algorithm (Table 4.5). This allows the transformation from numeric to categorical features and simultaneously the reduction of categories. In the second implementation, they are treated as continuous numbers (meteorological features) or integers (day of the season). It has to be noted that the warning region code and the day of the season are implemented as an



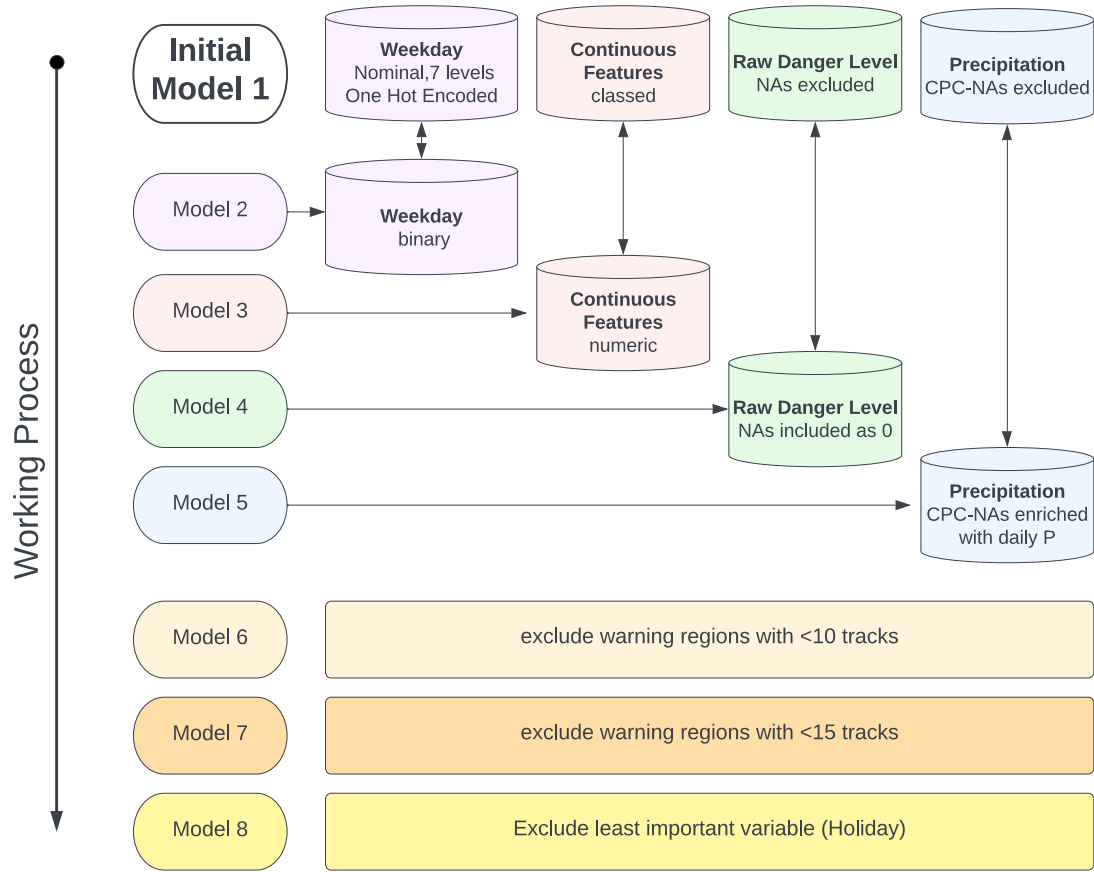
integers, since random forests cannot handle categorical features with more than 100 categories.

**Table 4.5:** Feature discretization into classes.

Feature	Range	Class
<b>Relative Sunshine Duration</b>	0 – 27%	3
	28 - 71%	2
	72 - 100%	1
<b>Precipitation</b>	0 – 1.8 mm	1
	1.81–7.1 mm	2
	>7.11 mm	1
<b>Temperature</b>	-30 – -9°C	-2
	-8.9 – -4°C	-1
	-3.9 – 0.3°C	0
	0.31 – 5.3 °C	1
	>5.31 °C	2
<b>Day of the Season</b>	0 – 61	1
	62 – 106	2
	107 – 151	3
	152 – 200	4

Tracks without a CombiPrecip value (i.e., NA) are excluded in a first approach, then enriched with daily precipitation sums in a second approach. The weekday feature is once treated as a nominal, one hot encoded feature with 7 levels and once implemented as a binary variable, which has only the two levels “workday” and “weekend”. The raw danger level is treated as an ordinal variable ranging from low (1) to high (5) avalanche danger. First, data that has no RDL value (i.e., NA), is excluded. In a second attempt, NA values are replaced with 0, which is not a real danger level used in practice but is used in this thesis to indicate an avalanche danger level smaller than 1. In an Additional approach, regions that contain a) less than 10 and b) less than 15 total tracks over all eight seasons are excluded, based on the assumption that a prediction is not sufficiently reliable when the data is very sparse. In a last step, the least important variable is excluded.

To assess the impact of different feature representations, a systematic approach is applied, which is visually described in Figure 4.8. First, an original model is built that implements all features according to Implementation 1 in Table 4.4. In the following models, the feature implementation is stepwise altered. If the altering yields a better model performance, it is accepted, otherwise it is discarded. The goal is to evaluate the effectiveness of each feature representation in enhancing the model’s predictive performance. The assessment of the predictive performance is discussed in Section 4.5.4.



**Figure 4.8:** Data-specific model optimization process. The initial model incorporates all predictor variables with implementation 1 (4.4). The model is altered (from top to bottom, model 1 to model 8) by a stepwise introduction of implementation 2 for every predictor variable. If the altering yields a higher performance, it is accepted, otherwise it is discarded. In the last three alterations, the least popular warning regions and the least important predictor is excluded.

### 4.5.3 Algorithm-Specific Adjustments

The random forest is built in RStudio using the *randomForest* package, which implements Breimann’s (2001) Random Forest algorithm. The *randomForest::randomForest()* function comes with a set of default hyperparameters that do not need to be adjusted. Even though leaving the hyperparameters to their default values is a common practice, Huang and Boutros (2016) emphasized on the importance of parameter tuning as a critical step in model fitting, which is outlined in this section.

First, the algorithm needs to be adjusted to handle the class imbalance. In the *randomForest* package there are two different parameters, which can be used to counteract class imbalance. With the *classwt* parameter, different weights can be assigned to the two classes. It has to be noted that in the documentation of the *randomForest* package, it does not become evident how exactly the *classwt* parameter is implemented into the random forest algorithm, therefore this

parameter has to be handled with care. The second parameter that can be used to balance classes is the *sampsize* parameter. It specifies how many data points from each class are used to build one decision tree. *Sampsize* can be either used with or without replacement. When used with replacement, each data point can theoretically be pulled from the sample several times. When used without replacement, each data point can only be used once for every tree. According to Strobl et al. (2007), trees should be built without replacement in order to produce an unbiased output. According to this study, sampling with replacement can aggravate minor input variations in the data, which can add an artificial bias to the data. However, both methods are applied and compared.

For the construction of the tree, the most common parameters to be adjusted are *mtry*, *ntree*, and *maxnodes*. *Mtry* is the number of features randomly sampled at each node as candidates for the split, *ntree* defines the number of trees to grow and *maxnodes* limits the maximum number of terminal nodes in each tree.

There exist built-in functions to find optimal hyperparameters for a random forest, such as *randomForest::tuneRF()*. However, they optimize with respect to the Out-of-Bag error estimate, which is insensitive to class imbalance. The OOB error estimate is calculated simultaneously to the random forest construction by using a tree to predict the outcome of all data that was not used to build this specific tree. It is an automatic byproduct of the *randomForest()* function. The mean of all error rates gives an OOB estimate of the generalization error (Breiman, 2001). Even though OOB-errors are often said to be unbiased, there are studies that suggest that OOB-errors can over-estimate the error (Janitza and Hornung, 2018).

Therefore, hyperparameters were optimized manually to find the value for which maximum performance is achieved. This is done using a stepwise increase of the respective parameter and is optimized for the Balanced Accuracy, which is calculated on an unseen test set, rather than an Out-of-Bag error. The balanced accuracy and other skill scores are discussed in the next section.

In the *sampsize* parameter, an equal number of presence and absence data points is used, which is a recommended practice for imbalanced data (Janitza and Hornung, 2018). According to Breiman (2001), random forests do not overfit with an increasing number of trees because of the Law of Large Numbers. In theory this means that higher *ntree* will always lead to a smaller or equally large error. Therefore, the optimization of *ntree* is treated with less priority.

Because *classwt* was found to perform radically worse than *sampsize*, it is discarded and therefore not optimized.

**Table 4.6:** Hyperparameters and tested values for best performance.

Hyperparameter	Values
<b>sampsize</b>	n : n, with n = c(50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000)
<b>mtry</b>	n = 1 – 12 (maximum of features)
<b>ntree</b>	n = 1 - 1000

#### 4.5.4 Verification and Validation

The data is split into train and test data. The test dataset consists of one whole winter season, the train data is the complement of the test data. Each model is trained with the train data that includes all data but the test data. The model is then applied to the new, unseen test data and a confusion matrix of predicted and observed values is created (Figure 4.9). Based on this confusion matrix, the following skill-scores are calculated and used for the performance assessment of the model:

		Observed	
		Presence (ARPD Tracks)	Absence (Generated)
Predicted	Presence	<b>True Positive (TP)</b>	<b>False Positive (FP)</b>
	Absence	<b>False Negative (FN)</b>	<b>True Negative (TN)</b>

**Figure 4.9:** Confusion Matrix showing the relationship between predicted and observed entities. Green boxes show correctly predicted entities, red boxes show erroneous predictions.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4.1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4.2)$$

$$\text{Balanced Accuracy} = \frac{(\text{Sensitivity} + \text{Specificity})}{2} \quad (4.3)$$

$$\text{Hanssen-Kuipers Skill Score (KSS)} = \frac{TP \times TN - FP \times FN}{(TP + FN) \times (FP + TN)} \quad (4.4)$$

(4.1) and (4.2) and are calculated according to Swets (1988). (4.3) follows from (4.1) and (4.2), and is used frequently when dealing with imbalanced classes (e.g., Bekkar et al., 2013; Marsland, 2015). Sensitivity (Specificity) are measures of how many of the presence (absence) data points were correctly predicted. The balanced accuracy is an advanced accuracy measure that is not influenced by class imbalance. The Hanssen–Kuipers Skill Score is also known as the Peirce Skill Score, which was first defined by Peirce (1884) and later by Hanssen and Kuipers (1965). It is a measure used in rare and severe event prediction, which is suitable for imbalanced classes, where the minority class is in the focus (Ebert and Milne, 2022).

After data- and algorithm-specific optimization, the test season is permuted to assess the sensitivity of the model to variations among test seasons. Additionally, the robustness of the model against the inherent uncertainty in meteorological measurements is evaluated. For this, intentional input uncertainties were introduced into the meteorological features (4.7). The data was manipulated using `stats::rnorm()`. This function generates random, normally distributed numbers. For the standard deviation, the uncertainty obtained from MeteoSchweiz documentation of the data products was used (MeteoSwiss, 2021a, 2021c), while the original values served as a mean, from which the generated values deviated. For CombiPrecip however, uncertainty is only provided as a RADAR scatter in dB, but not in millimeters. Therefore, an arbitrary value of 1 mm is chosen, based on the fact that most morning precipitation values do not exceed 1 mm. Because precipitation gets usually underestimated (MeteoSwiss, 2021d), uncertainties are always added to the mean, instead of using a normal distribution around the mean. Also, subtracting precipitation from the original value would result in many negative and thus non-sensical values.

**Table 4.7:** Uncertainties of meteorological variables that are introduced into the data to assess the robustness of the model against input uncertainty.

Variable	Uncertainty
Temperature	0.6°C
Precipitation	1 mm
S <sub>rel</sub>	10%

# Chapter 5

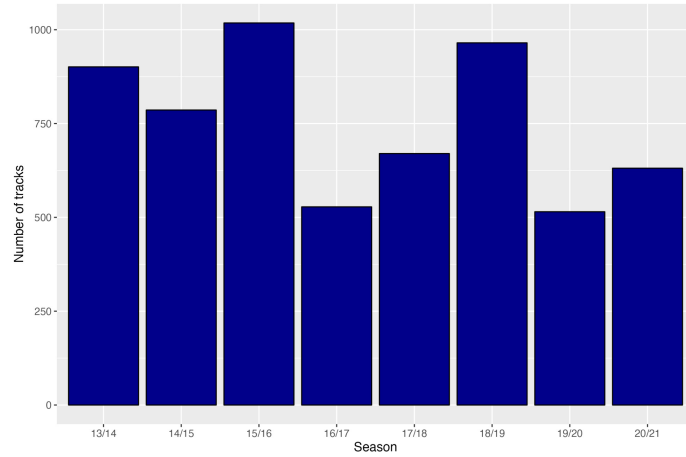
## Results

In this chapter, an overview over the results is provided. In the beginning, the presence data is characterized (5.1). Further, the backcountry skiing prediction is discussed in terms of the optimal algorithm setting (5.3.1), the skill scores (5.3.2), the influence of the different predictors (5.3.3 and 5.3.4), the spatial and temporal distribution of errors (5.3.6 and 5.3.5), and ultimately the robustness of the model (5.3.7).

### 5.1 Characterization of Presence Data

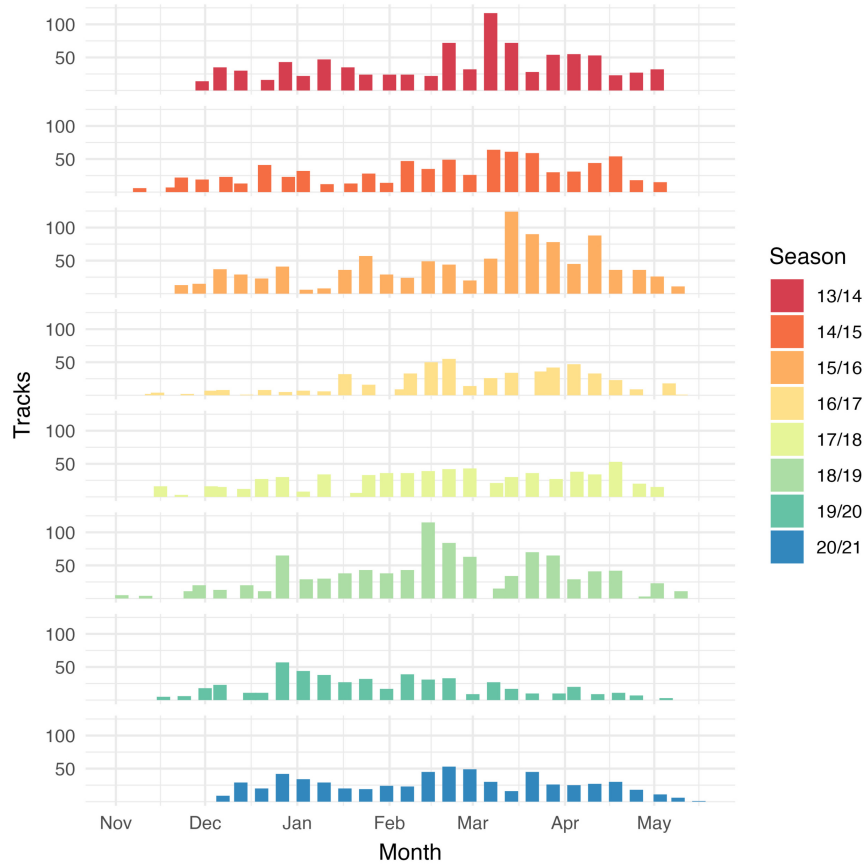
Figure 5.1 shows the seasonally recorded track number in the observation period spanning from 2013/14 to 2020/21. On average 750 tracks were recorded each season. However, there are substantial variations among the different seasons. During the season 2016/17, exceptionally few tracks were recorded, which can be attributed to the climatic conditions during this winter season. The whole winter was anomalously dry, which resulted in an extreme lack of snow. In parts of southern and western Switzerland it was the driest winter since 40, respectively 55 years. Additionally, the average temperature in the Alps deviated locally  $+1$  -  $+2$  °C from the norm temperature 1981-2010 (MeteoSchweiz, 2017b).

Further, the number of tracks per season decreased significantly in 2019/20 and 2020/21, which is likely due to the COVID-19 pandemic. It has to be noted that these seasons might show different patterns than the seasons not affected by the pandemic. However, analyzing and quantifying those patterns would go beyond the scope of this thesis.

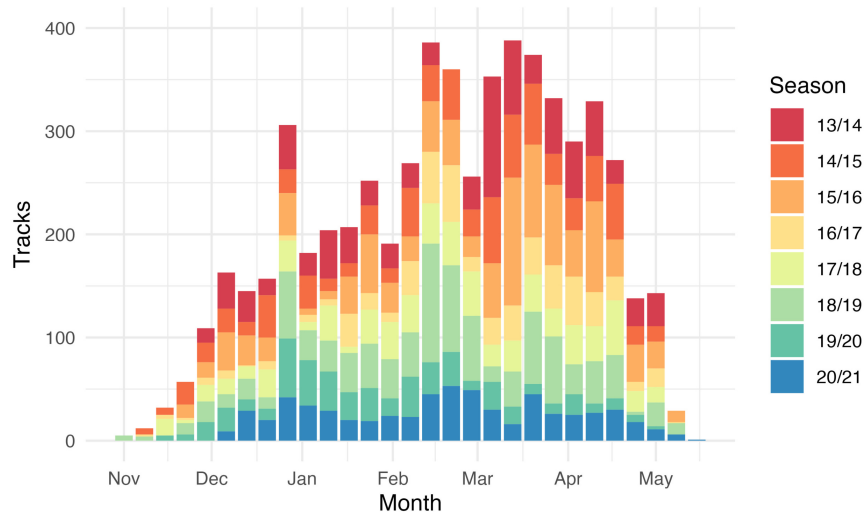


**Figure 5.1:** Number of tracks per season in the observation period from 2013/14 to 2020/21.

In Figure 5.2, the weekly evolution, i.e., the 7-day sum of total tracks, is visualized for each season. Figure 5.3 shows the same data as Figure 5.2, but the different seasons are stacked to visualize the trend over all seasons. The backcountry skiing season usually starts in November and ends in May. All seasons show a similar behavior, with an increase of total tracks from the beginning of the season until early spring, where in most seasons the maximum number of tracks were recorded. This is followed by a strong decrease in April and May, which denotes the end of the season. In the season 2019/20, the roughly constant number of tracks in January and February were followed by a steep drop in March. This coincides with the first COVID-19 lockdown in March 2020 in Switzerland. The season 2020/21 seems to be qualitatively comparable to other, non-pandemic seasons, but smaller in magnitude. The week between Christmas and New Year has exceptionally high numbers in all seasons except for seasons 2014/15 and 2016/17. December 2014 was characterized by above-average temperatures in all parts of Switzerland, which induced below-average snow conditions. Only in the last week of December 2014, a strong onset of winter with heavy snowfall took place, which possibly led to unfavorable skiing conditions (MeteoSchweiz, 2015). In the northern Alps, December 2016 was the second warmest, in the southern Alps the fourth warmest since the start of measurement in 1864. Additionally, in many regions it was the driest December ever measured in Switzerland (MeteoSchweiz, 2017a). This correlates with exceptionally low numbers of recorded tracks at the beginning of season 2016/17.



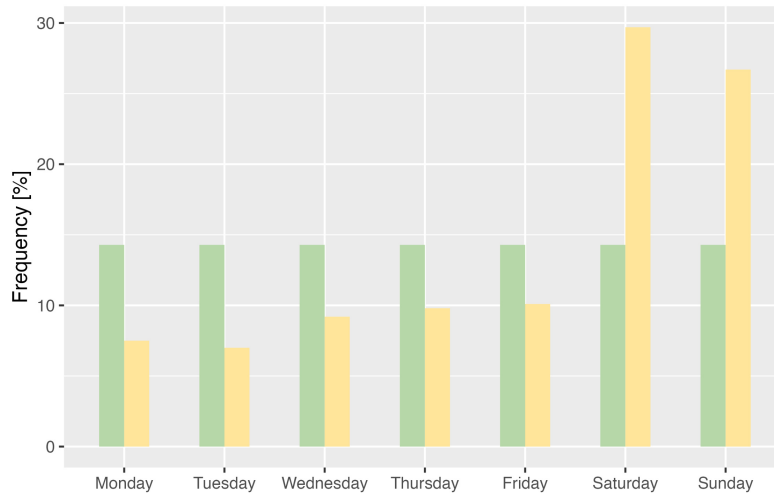
**Figure 5.2:** Yearly evolution of weekly aggregated tracks for each season in the observation period between 2013 and 2021. Due to differences in season days (leap years), there are slight variations among the definition of weeks in each season.



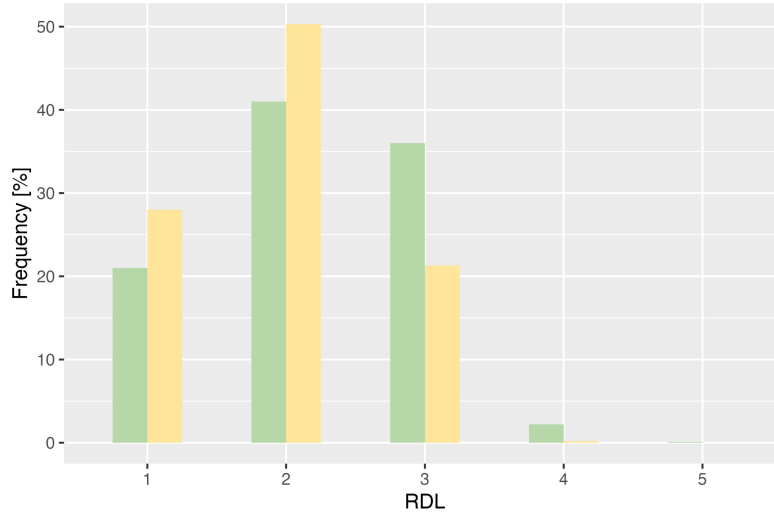
**Figure 5.3:** Yearly evolution of weekly summarized tracks, stacked for all seasons to show the overall trend throughout the season. Note that the exact definitions of each week may differ slightly among seasons due to different amount of season days (leap years).



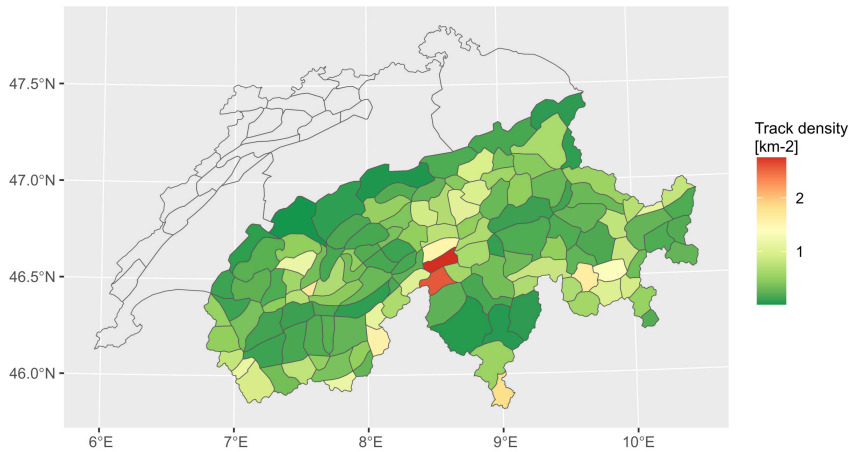
The total number of tracks over all seasons is nearly constant throughout the week and is significantly higher on the weekends (Figure 5.4). A majority of tracks (56.4%) were recorded on a weekend. This is approximately doubled compared to the baseline frequency of the weekend ( $2/7 = 28.6\%$ ). The distribution among workdays is almost uniform. Only a slight increase can be observed from Monday to Friday. Of the 6014 presence tracks, 377 tracks or 6.3% took place on a holiday. The number of total tracks per RDL roughly follows the baseline frequency of the avalanche danger levels (Figure 5.5). The most frequent baseline danger level (“2 – Moderate”) coincides with the most frequent danger level in the recorded tracks. The frequency of levels “1 – Low” and “2 – Moderate” is higher in the tracks than in the baseline, while levels “3 – Considerable”, “4 – High” and “5 – Very High” are less frequent in the tracks than in the baseline. No track was recorded for danger level 5, but instead, some tracks were recorded on days where no avalanche bulletin was published (level 0). Spatially, the tracks are concentrated in a few hotspots, which have been discussed earlier for the popularity calculation (Section 4.4.3). In Figure 5.6 the warning regions are colored according to the overall track density. The average size of a WR is about 200 km<sup>2</sup> and contains on average 44 tracks in total since 2013. The highest track densities are found in the regions “6111 – Bedrettal” and “2224 – südliche Urseren”, which are located near the border between the two climatic regions “2 – zentraler Alpennordhang” and “6 – zentraler Alpensüdhang”.



**Figure 5.4:** Frequency of weekdays throughout all seasons for presence tracks (yellow) and baseline frequencies (green). The baseline frequency for each day is  $1/7$  or 14.3%.



**Figure 5.5:** Frequency of raw danger levels throughout all seasons for presence tracks (yellow) and baseline frequencies (green). The baseline frequency is based on seasons 2012/13 to 2020/21, and is obtained from [www.slf.ch](http://www.slf.ch).



**Figure 5.6:** Overall track density per Warning Region for the Swiss Alps, 2013 – 2021.

## 5.2 Presence vs. Absence Data

The premise for the activity prediction is that absence and presence tracks, i.e., artificially generated and real recorded backcountry ski tracks, have different properties. This section provides a quantitative and qualitative comparative analysis of both classes in terms of the features used for the prediction. It has to be noted that comparison is exclusively carried out with relative frequencies rather than absolute counts because absence tracks greatly outnumber presence tracks.

In Figure 5.7, frequency distributions and boxplots are provided for each meteorological feature by class, Table 5.1 provides a quantitative overview. Precipitation values are generally low for

both absence and presence tracks, as it is only the 4 – hour mean sum in the track-specific elevation belt. In both classes, the majority of tracks obtained no precipitation at all. However, significantly more presence tracks received no precipitation (85%) compared to the absence tracks (65%). Contrastingly, the relative frequency of tracks receiving more than 5 mm of precipitation is 10 times higher in the absence (2.22%) than in the presence (0.21%) tracks. In the precipitation histogram (Figure 5.7a), it can be observed that the shape of the distribution is similar in both classes, but the absence class is shifted towards slightly higher precipitation. This is in line with the mean value of absence tracks (0.41 mm), which is about sixfold compared to presence tracks (0.07 mm).

For relative sunshine duration, a majority of the data lies either below 10% or above 90% in both classes (Table 9b). This is also reflected in the boxplots in Figure 5.7d, which show that both classes have a relatively high standard deviation. However, the two classes show opposing trends, which is visible in Figure 5.7c: For  $S_{rel}$  above 90%, the frequency of presence tracks exceeds absence tracks by 21%, whereas for  $S_{rel}$  below 10%, presence tracks are exceeded by the absence tracks by 22%. This is also reflected in the mean values, which differ by approximately 20%.

The temperatures for both classes exhibit a nearly normal distribution, as depicted in Figure 5.7e. Notably, the two classes show very similar means (Figure 5.7e). However, the standard deviation is about 1°C or 17% higher in the absence tracks, which means that presence tracks tend to concentrate within a narrower temperature range.

In Figure 5.8, the frequency distributions of weekday, raw danger level, and popularity are depicted. The distribution of weekdays in the absence data follows a uniform pattern, aligning with the baseline frequency for each weekday. As opposed to this, presence tracks exhibit a distinct concentration on the weekends, which was already discussed in Chapter 5.1.

6.3 (5.6)% of presence (absence) tracks took place on a holiday. The baseline frequency is on average 6.3%. It varies with the number of days in a season (leap year) and the specific dates of the holidays, as some holidays fall in or out of the season defined between November 15 and May 15. Hence the holiday frequency is approximately the baseline frequency and approximately 12% higher than the holiday frequency in the absence tracks.

For the raw danger level, the highest frequency can be observed in level “2 – Moderate” in both classes, which is in line with the baseline frequency discussed in Section 5.1. It is striking, however, that presence tracks are concentrated in levels 1 – 3, whereas absence tracks span a wider range from 0 – 4. Presence tracks exceed baseline frequencies in levels 1 and 2, while falling below in levels 3, 4, and 5. Absence tracks approximately align with the baseline frequency in levels 1, 4, and 5, while falling slightly below in levels 2 and 3. Danger level 0 exhibits a relatively high frequency in the absence tracks and is nearly nonexistent in the presence tracks. For popularity, absence tracks virtually follow the baseline frequency, while presence tracks are concentrated in the higher popularity levels 1 – 3.

**Table 5.1:** Mean and standard deviation for meteorological features by activity group. a. Precipitation, b. Relative sunshine duration, and c. Temperature. For relative sunshine duration and precipitation, additionally the relative frequencies for 0 mm (<10%) and >5mm (>90%) are given.

**a. Precipitation**

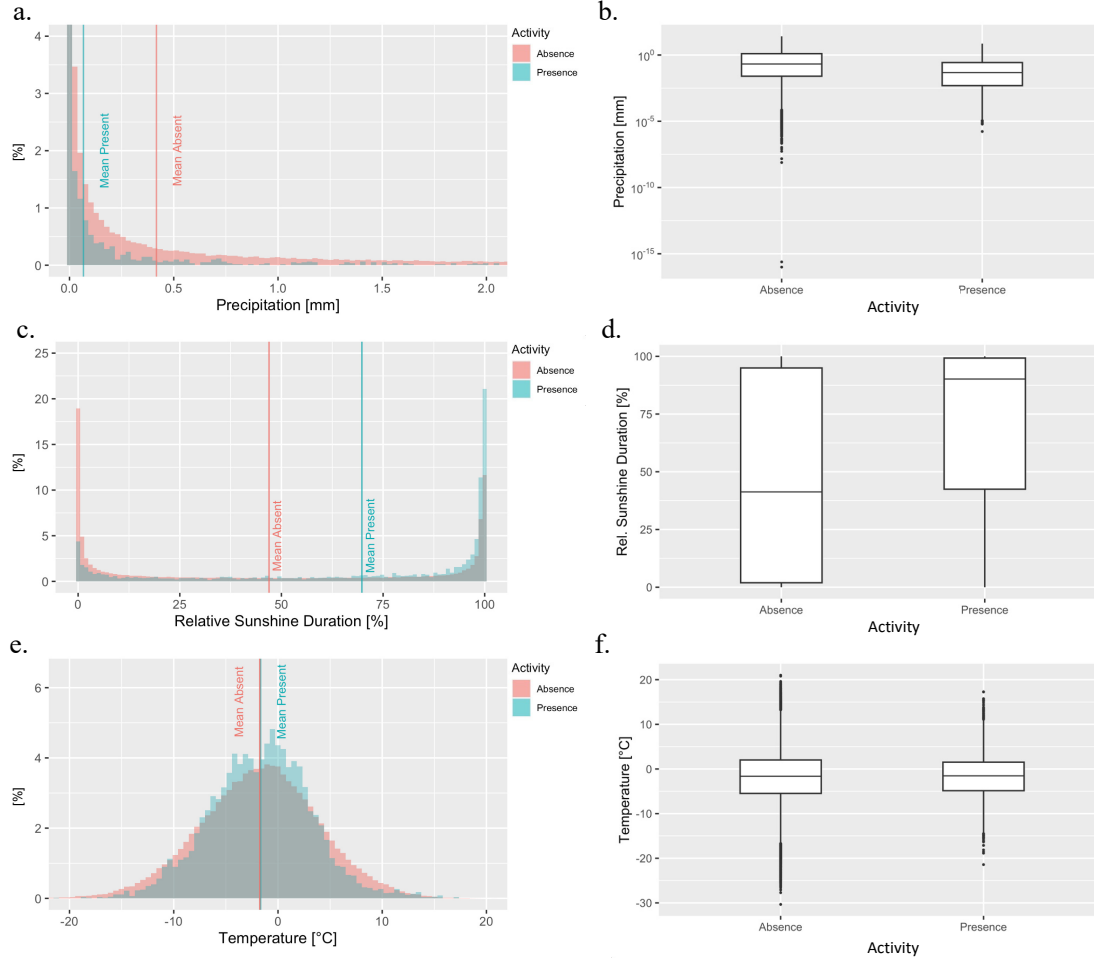
Activity	Mean [mm]	Standard deviation [mm]	0 mm	>5 mm
Absence	0.41	1.47	65%	2.22%
Presence	0.07	0.44	85%	0.21%

**b. Relative Sunshine Duration**

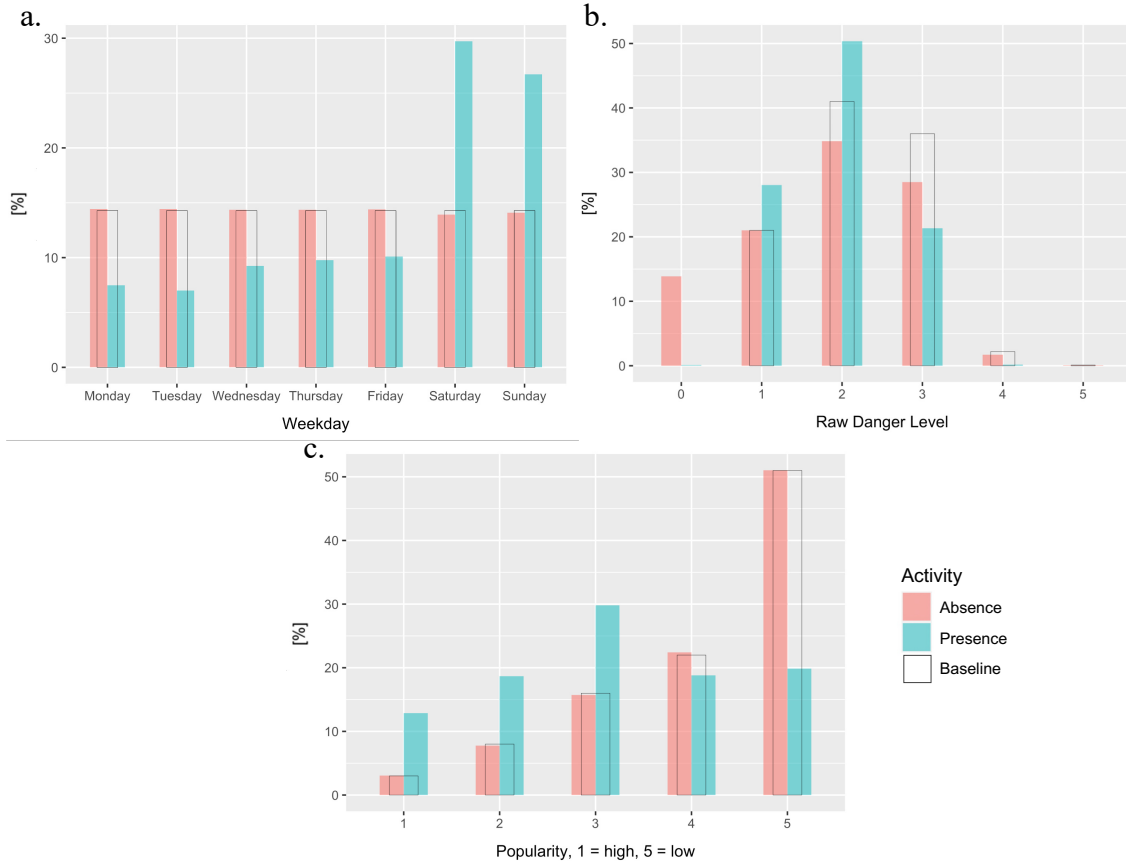
Activity	Mean [%]	Standard deviation [%]	<10%	>90%
Absence	46.95	41.29	35 %	29 %
Presence	69.80	35.87	13.2 %	50 %

**c. Temperature**

Activity	Mean [°C]	Standard deviation [°C]
Absence	-1.78	5.70
Presence	-1.67	4.85



**Figure 5.7:** Frequency distribution and boxplots of precipitation (a., b.), relative sunshine duration (c., d.) and temperature (e., f.) by class. In a., only precipitation values <2 mm and relative frequencies <4% are displayed, to avoid visual distortion owing to very high frequencies for 0 mm values (65% of absence tracks resp. 85% of presence tracks). For better readability, the y-axis for b. was logarithmically transformed, consequently only values >0 mm are displayed



**Figure 5.8:** Frequency distribution of a. weekday, b. raw danger level, and c. popularity by activity group. Baseline frequencies are indicated by the black boxes. Baseline frequency of popularity emerges from Figure 4.6, baseline frequency of RDL emerges from [www.slf.ch](http://www.slf.ch). In c., absence frequency for level 5 is 0.08% and presence frequency for level 0 is 0.1%, therefore they are not visible. Presence frequency for level 5 is 0%.

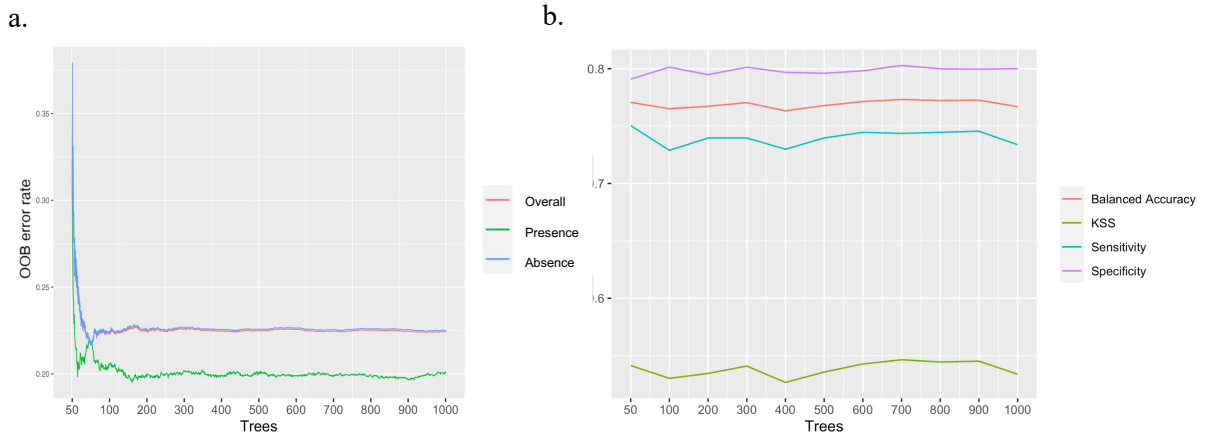
## 5.3 Prediction

### 5.3.1 Hyperparameters

The hyperparameters *ntree*, *mtry*, *sampsiz*e, and *maxnodes* were optimized by a using stepwise increase, as described in Section 4.5.3.

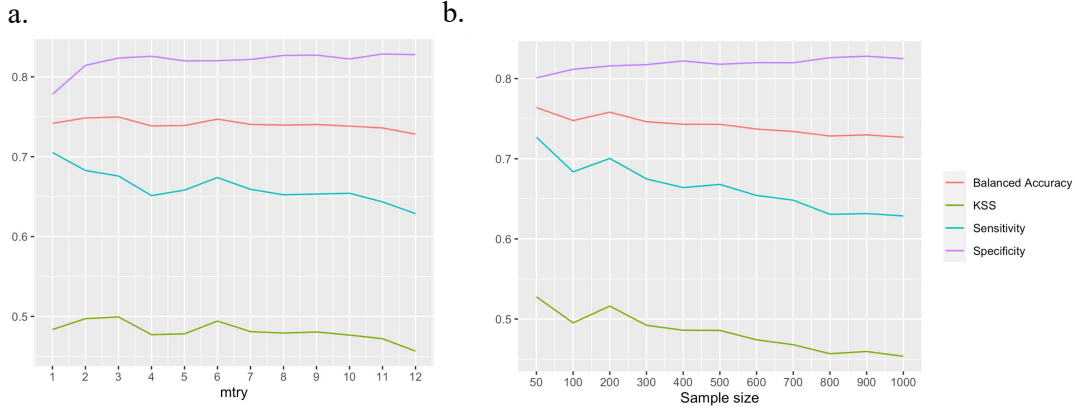
In theory, increasing the number of trees (*ntree*) will always result in a smaller or equally large error because of the Law of Large Numbers. For the most part, this could be confirmed with OOB-error estimates and cross validation. The OOB-error for both classes decreases rapidly with an increasing *n* and stabilizes at approximately  $n = 250$  trees and a OOB error rate of 0.225 (absence) and 0.20 (presence) (Figure 5.9a). In the cross-validation, similar results were obtained. For the cross-validation, skill scores rather than error rates were calculated, which means that high values, rather than low values indicate a good fit. The steep drop of the error rate, respectively increase of skill scores at the beginning is not captured by the cross-validation approach, since the minimal number of trees tested is 50 (Figure 5.9b). Sensitivity and KSS

slightly decrease from  $n = 50$  to  $n=100$ , while specificity increases. After approximately  $n = 500$ , all skill scores are relatively stable. Therefore, 500 trees were used for the final model.



**Figure 5.9:** a. OOB-error rate by class as produced by `randomForest::randomForest()` and b. cross-validation skill scores with increasing number of trees ( $ntree$ ).

For  $mtry$ , sensitivity decreases with increasing  $n$ , while specificity first increases, then stabilizes at approximately  $n = 3$  (Figure 5.10). The balanced accuracy and KSS slightly peak at  $n = 3$  and  $n = 6$ . However, the variations are minor. It has to be noted that  $mtry$  is also sensitive to the total number of features used. As the number of features varies with different feature implementations, the default value (square root of total number of features) is chosen. For the  $sampsiz$  parameter, skill scores focusing on the presence (Sensitivity, KSS) class generally decrease with increasing sample size, while specificity remains at a high level throughout all sample sizes. Therefore, a relatively small sample size of 200:200 is chosen. Sampling without replacement yields slightly higher performances, therefore it is preferred over sampling with replacement. An adjustment of the maximum numbers of terminal nodes ( $maxnodes$ ) yielded lower values in all skill scores than setting it to the default value. Therefore, the default value is employed, which means that there is no limit on the maximum number of nodes.



**Figure 5.10:** Skill scores obtained through cross-validation for different values of a. *mtry*, i.e., the number of randomly selected features for each split and b. *sampsiz*e, i.e., the number of sampled data points for every tree. The sample size for both absence and presence class is the same.

### 5.3.2 Skill Scores

An overview over skill scores calculated for different model runs are presented in Table 5.3. The indices of the different models correspond to the model implementations presented in Figure 4.8. All models hold the same algorithm-specific properties, with *mtry* and *maxnodes* set to the default value, *ntree* = 500 and *sampsiz*e = 200 for both classes. Furthermore, sampling without replacement was applied, as this produced slightly higher values.

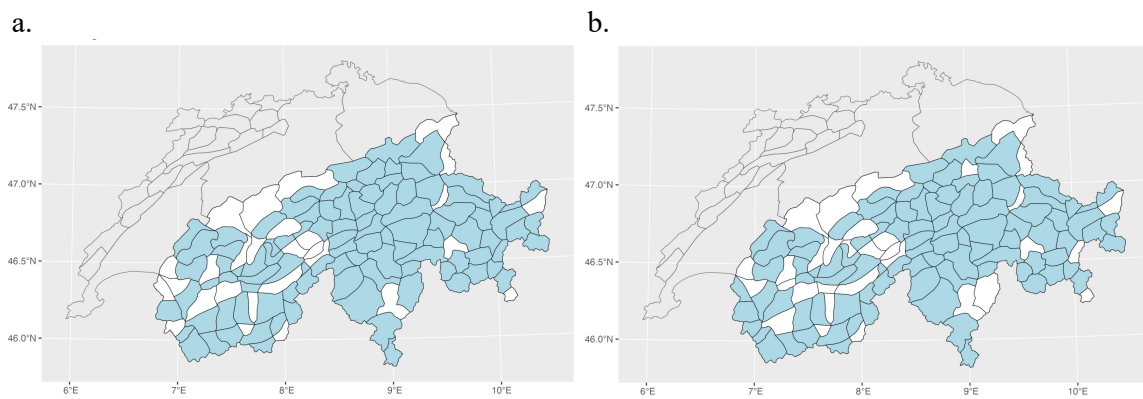
Model 5 yielded the highest performance throughout all skill scores. In this model, weekdays were treated as a binary variable (weekend / workday), meteorological (i.e., continuous) variables were treated as continuous variables rather than categorical variables, and NA values were included as 0 (RDL) respectively enriched with daily precipitation sum data (precipitation). It becomes evident that equalization of classes by either using one hot encoding or classifying continuous variables did not improve the model. Contrary to that, including NA values by replacing them with a dummy value (RDL) or with another data source (precipitation) did improve the model’s performance. However, it is worth noting that the performance improvement due to the inclusion of NA values is specifically observed in the presence tracks (i.e., sensitivity). In the case of absence tracks, the inclusion of NA values does not significantly affect the performance (i.e., specificity), as it remains relatively unchanged.

Further, excluding regions with only a small number of total tracks over the whole period (models 6 and 7; Figure 5.11) resulted in a deterioration of the model’s performance, as did the exclusion of the least important variable “holiday” (model 8). Generally, data-specific adjustments resulted in a greater variation in sensitivity (0.087) compared to specificity (0.01). The KSS exhibits a relatively high standard deviation of 0.093, while the standard deviation of the balanced accuracy lies between the values for specificity and sensitivity (0.046), as it is the geometric mean of the two.



**Table 5.2:** Skill scores of different model runs according to Figure 4.8. The standard deviations for each skill score is calculated from all different model runs.

Model	Sensitivity	Specificity	KSS	Balanced Accuracy
1 (Initial)	0.656	0.794	0.450	0.725
2	0.662	0.796	0.458	0.729
3	0.682	0.792	0.474	0.737
4	0.730	0.797	0.527	0.763
5	<b>0.743</b>	<b>0.800</b>	<b>0.543</b>	<b>0.771</b>
6	0.728	0.790	0.518	0.759
7	0.715	0.789	0.504	0.752
8	0.725	0.780	0.505	0.753
STDEV	0.087	0.01	0.093	0.046



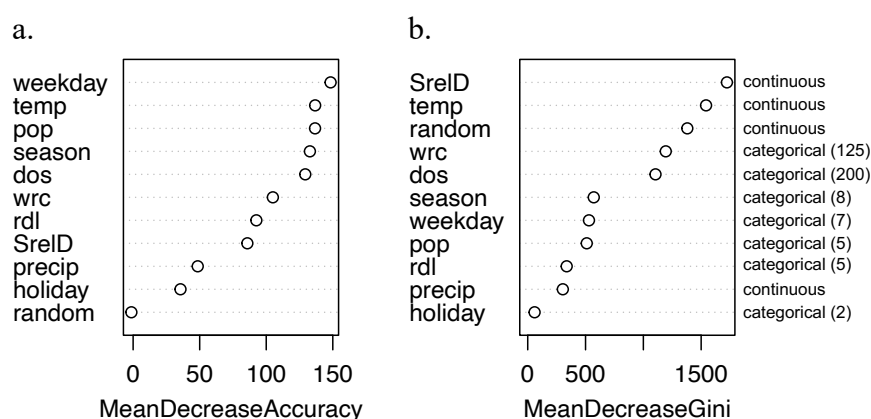
**Figure 5.11:** Warning Regions with a. <10 (Model 6) and b. <15 (Model 7) total tracks from 2013/14 to 2020/21.

### 5.3.3 Variable Importance

The *randomForest* package holds two different methods of variable importance measurement: Mean decrease in accuracy (MeanDecreaseAccuracy) and mean decrease in GINI impurity (MeanDecreaseGINI). The former expresses the mean decrease in accuracy the model suffers by excluding a given feature. The latter is a measure of how much a given feature contributes to the homogeneity, hence the purity, of a node in the resulting random forest (Breiman, 2001; Liaw and Wiener, 2002).

Strobl et al. (2007) pointed out that the MeanDecreaseGINI is biased and should not be used with data of varying types. According to this study, importance is over-estimated for categorical features with a high number of categories and numeric features, while under-estimated for features with only few categories or even binary features. This is due to the fact that features that offer a high number of possible splits (i.e., numeric and high-cardinality categorical features), are chosen more often for a split, and therefore their importance is over-estimated. The findings of Strobl et al. (2007) could be verified in this thesis. To uncover a potential bias, a random

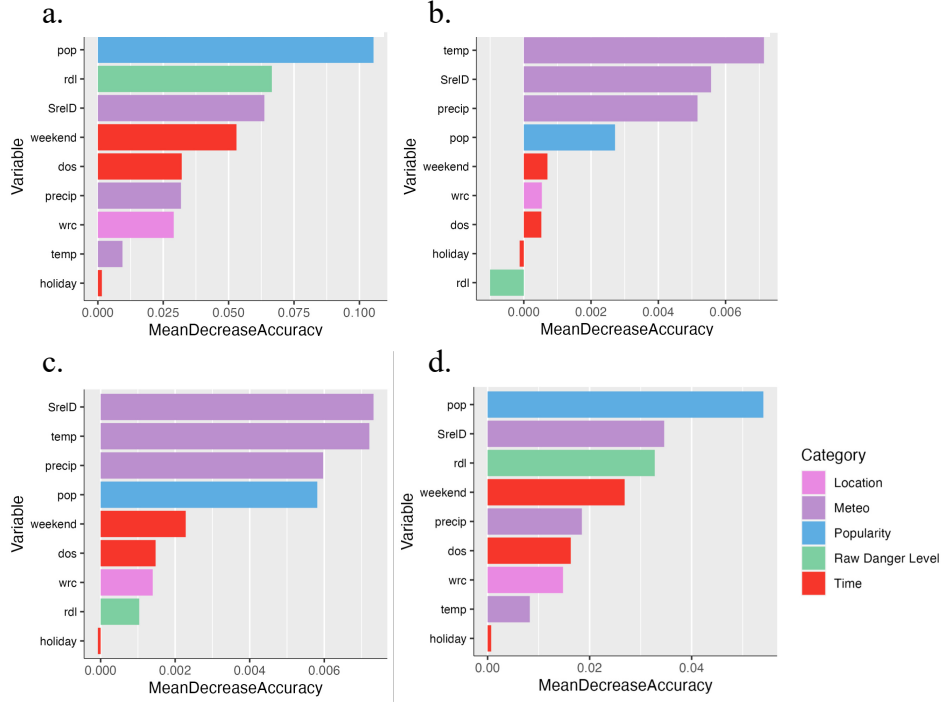
column was introduced. The random column contains randomly generated numbers between 0 and 1, it is therefore a continuous feature with a high number of possible splits. Ideally, this random column should have no influence on the prediction since there is no relationship between the target variable and the random column. The variable importance should therefore be 0 for the random feature. In Figure 5.12b however, it is visible that according to the MeanDecreaseGINI, the random column is the third most important feature. It becomes apparent that there is a tendency that continuous (meteorological features) and high-cardinality categorical variable (warning region code, day of the season) are the most important, while binary features and features with only few categories (holiday, raw danger level, popularity) are least important. The variable importance measured by the MeanDecreaseAccuracy does not seem to be subject to this bias, therefore it is preferred as a measure of variable importance.



**Figure 5.12:** Variable Importance as measured by a. MeanDecreaseAccuracy and b. MeanDecreaseGINI. Data types for features in b. are listed next to the graph. For categorical features, the number of categories is expressed in brackets. Importance Metrics in this Figure emerge from a model with no hyperparameter fine-tuning or data-specific adjustments and are thus not comparable to other variable importance plots presented in this section. The following variable names are abbreviated: pop = popularity, rdl = raw danger level, SrelD = relative sunshine duration, precip = precipitation, temp = temperature, dos = day of the season, wrd = warning region code

Figure 5.13 shows the variable importance calculated with the best performing model (Model 5). The variable importance differs by class and the weighted mean of both classes is strongly influenced by the majority class (absence), whereas the unweighted mean is strongly biased towards the presence class, since presence variable importance values are one to two orders of magnitude higher.

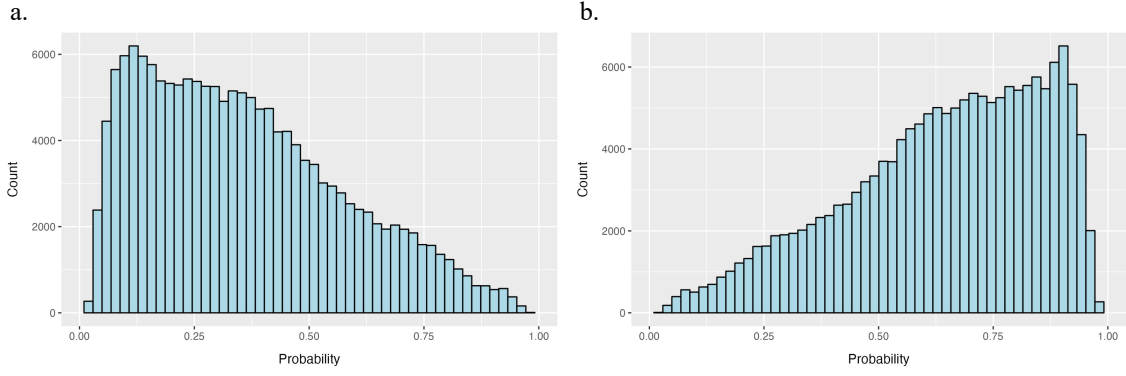
For the prediction of the presence class, the time-independent popularity variable is most important, followed by the avalanche forecast (RDL) and the relative sunshine duration. For the absence prediction, the three meteorological variables are most important, followed by popularity. In the absence prediction, the avalanche forecast has a negative importance, which indicates that it adds to a higher error rate compared to a model that does not include the avalanche forecast. In both classes, the importance of the holiday variable is approximately 0, which indicates that it has no importance for neither absence nor presence prediction.



**Figure 5.13:** Variable Importance of best performing model (Model 5), representing a. presence class, b. absence class, c. weighted mean of both classes, and c. unweighted mean of both classes. The following variable names are abbreviated: pop = popularity, rdl = raw danger level, SrelID = relative sunshine duration, precip = precipitation, temp = temperature, dos = day of the season, wrd = warning region code.

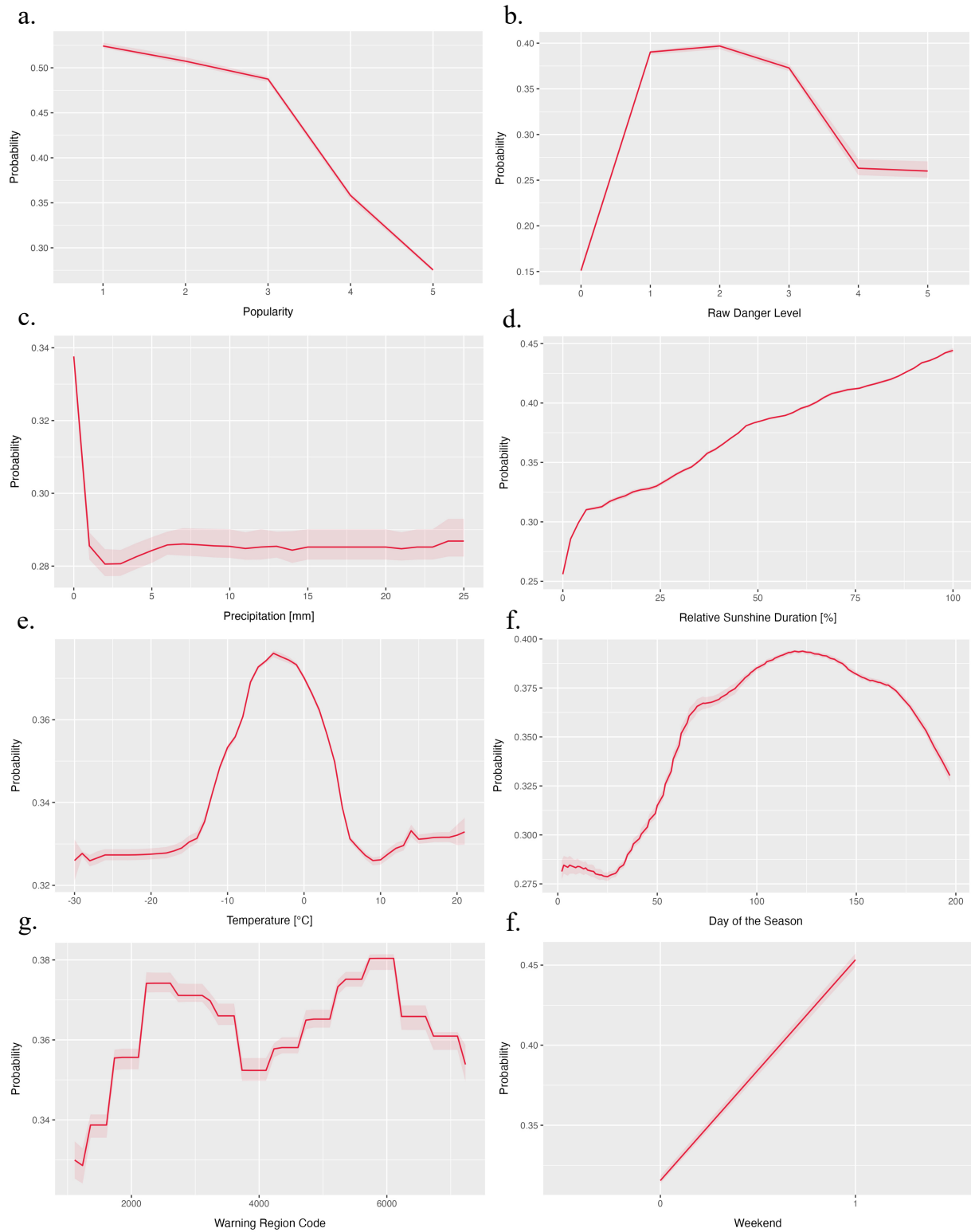
### 5.3.4 Partial Dependence

Partial dependence is a measure for the marginal effect a variable has on the predicted outcome of a model. The basic idea is to marginalize all features but one in order to reveal the relationship between the one feature that is not marginalized and the predicted outcome probability (Hastie et al., 2001). In other words, one feature is altered while keeping all other features constant to determine the influence each instance of a feature has on the probability for a given outcome. The resulting probabilities can be visualized in a partial dependence plot. In this case, the partial dependence is calculated for the presence class, as this is the class of interest. Because there are only two classes, the partial dependence for the absence class is inverse to the presence class. The y-axis of the partial dependence plot shows the probability for presence. However, the plots should be interpreted relatively rather than absolutely. Also, it is worth noting that due to the class imbalance, the probability distributions for both classes are fundamentally different. In Figure 5.14, it is visible that in the absence class, most probabilities exceed 0.5, while the opposite is the case in the presence class. Therefore, the mean probability in the presence class is below 0.5, which indicates that most of the data points are getting assigned to the absence class. This is the reason, why probabilities in the partial dependence plots for the presence class rarely exceed 0.5.



**Figure 5.14:** Probability distribution of a. presence and b. absence class.

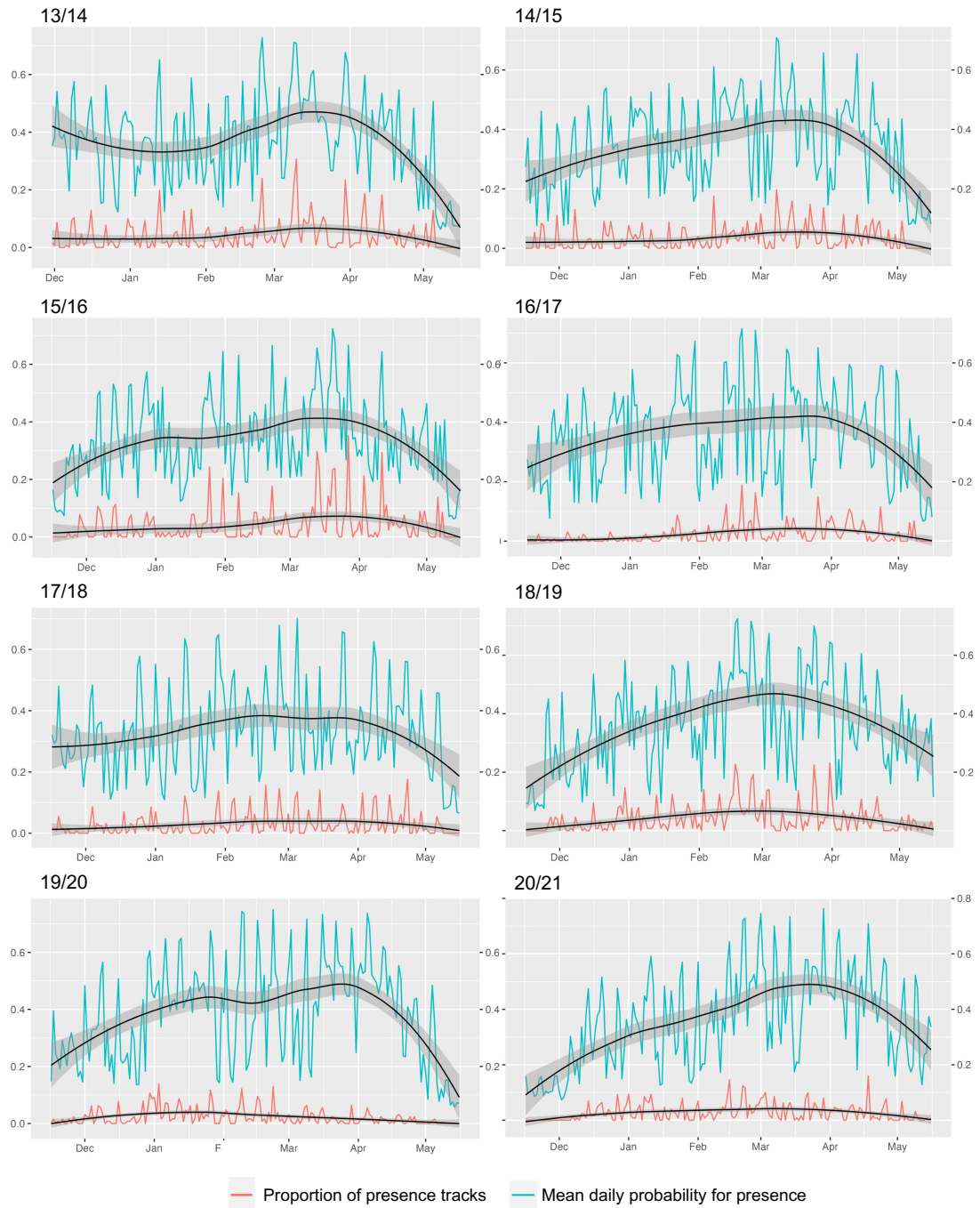
Figure 5.15 shows the partial dependence plot for each predictor variable. The partial dependence plots reveal that presence predictions are positively influenced by lower popularity values (high popularity) (a.). Further, presence tracks are associated with lower avalanche danger levels but are not likely to occur with danger level 0 (NA). As the danger level increases, presence tracks get less likely and the confidence interval widens. This can be attributed to the relatively rare baseline occurrence of higher danger levels, which leads to less reliable estimates and therefore higher uncertainties (b.). Further, the probability of a presence track is highest when no precipitation occurs and rapidly decreases with higher precipitation (c.). The probability increases almost linearly with increasing  $S_{rel}$  and is highest for a value of 100% (d.). The temperature plot shows that the optimum range for a presence track lies between  $-10^{\circ}\text{C}$  and  $0^{\circ}\text{C}$ . Presence tracks get less likely with extremely high (above  $8^{\circ}\text{C}$ ) and low (below  $-15^{\circ}\text{C}$ ) temperatures (e.). In the temporal dimension, presence is least likely at the beginning and the end of the season, and most likely in the middle of the season (f.). The partial dependence for the warning region variable has to be interpreted with care. As mentioned before, the warning region code was implemented as an integer, because random forests do not allow more than 100 categories for categorical variables. Therefore, not all warning regions displayed on the x-axis really exist. However, as the first digit of the WRC indicates one of the seven the climatic regions in the Alps, the partial dependence plot can be interpreted accordingly. The highest probabilities can be found in the Regions 2 – Zentraler Alpennordhang, 5 – Nord- und Mittelbünden, and 6 – Zentraler Alpensüdhang, while lowest probabilities are found in regions 1 – Westlicher Alpennordhang and 4 – Wallis (g.). It can be concluded, that a presence track is more likely to occur in the northeastern part than in the southwestern part of Switzerland. Ultimately, presence is more likely on a weekend, compared to a workday. Generally, it can be observed that the partial dependence plots qualitatively follow the distribution of values in the absence tracks (see Section 5.2).



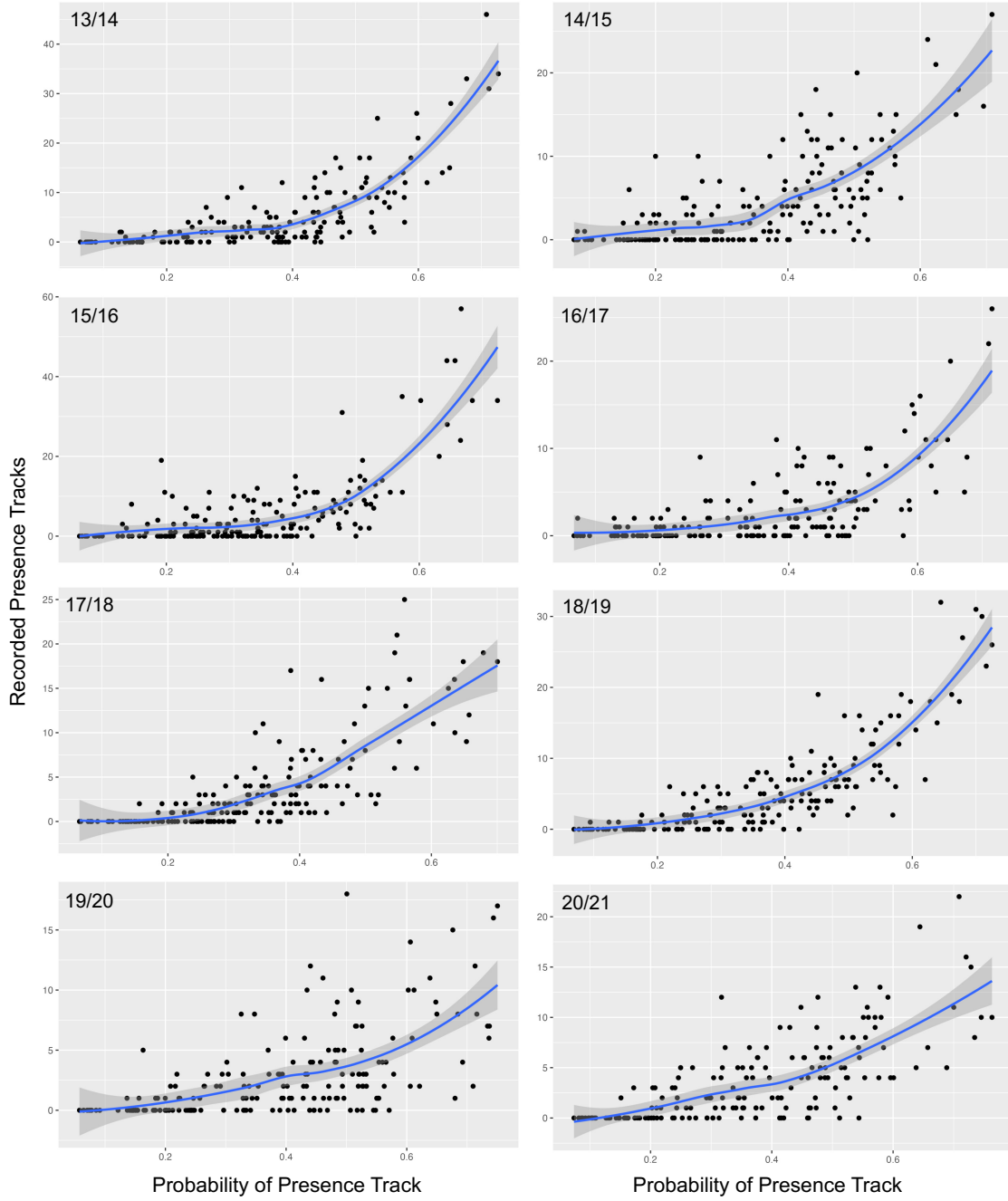
**Figure 5.15:** Partial dependence of predictor variables for presence class. The red line shows the mean of 20 iterations of the variable importance calculation, the transparent area shows the 95% confidence interval.

### 5.3.5 Temporal Distribution

In Figure 5.16, the mean daily probability for a presence track is shown for every season. In the same figure, the daily proportion of presence tracks is plotted. It can be observed that lower probabilities are found at the beginning and at the end of a season, while highest probabilities occur in the middle of the season around March. For the most part, this correlates with the number of presence tracks throughout the season. The regularly high probability peaks can be attributed to weekends, which consistently align with the peaks of presence tracks. Periods of anomalously few presence tracks, such as January 2016 approximately coincide with low probabilities. In March 2020 however, even though the number of presence tracks is exceptionally low, the probabilities are not particularly low. The reason for this is most likely the first COVID-19 lockdown, which greatly restricted the mobility. The probability therefore does not align with the track counts, since the COVID-19 restrictions are not included in the model. In Figure 5.17, it can be observed that there exists a distinct positive correlation between mean daily probability and daily presence track counts for all seasons. For some seasons, the positive correlation is more distinct, and the variations are relatively low (e.g., 2013/14 and 2018/19). In other seasons, the positive correlation is less pronounced and the variations are larger (e.g., season 2019/20).



**Figure 5.16:** Mean daily probability for presence vs. daily proportion of presence tracks. The black curve indicates the local polynomial regression line with a 95% confidence interval.



**Figure 5.17:** Mean probability of presence vs. absolute number of presence tracks. Each point represents one day. The blue curve indicates the local polynomial regression line with a 95% confidence interval.

### 5.3.6 Spatial Distribution

Figures 5.18 - 5.21 show the spatial distribution of the four skills scores sensitivity, specificity, balanced accuracy, and KSS for each season. The values were calculated using the best performing model (Model 5) for each season separately, using all seasons except the target season



as training data. Additionally, for every warning region the total number of presence tracks recorded in the target season are shown with circles of different sizes. Red triangles indicate that zero tracks were recorded in the target season. According to equations (4.1), (4.3), and (4.4), sensitivity, balanced accuracy, and KSS cannot be calculated when there are no presence tracks (i.e.,  $TP + FN = 0$ ), which is represented by NA. Figure 5.22 reveals the relationship between the total track count and the different skill scores for the example season 2017/18, which can be considered representative for all seasons.

### **Sensitivity** (Figure 5.18)

Sensitivity, i.e., how many presence tracks were correctly predicted (true positive rate), is consistently high in the central and northeastern part, as well as in the in the southeastern part of Switzerland (Engadin Valley and Ticino). In the Northwest, sensitivity is often in the lowest quartile, especially in the northern Pre-Alps. The overall pattern is for the most part consistent throughout all seasons. Higher sensitivity values tend coincide with regions with a higher number of presence tracks, while low sensitivity values can be found in regions where there are generally less recorded tracks. The positive trend between sensitivity and presence track number is relatively strong for sensitivity values below 1. However, the trend is undermined by sensitivity values = 1, which occur in regions with high track numbers, as well as in regions with only few tracks (Figure 5.22a). The standard deviation among the warning regions is approximately 0.4 and the highest of all skill scores, indicating high variations among warning regions. However, variations of the standard deviation among different seasons is low.

### **Specificity** (Figure 5.19)

Specificity, i.e., how many absence tracks were correctly predicted (true negative rate), is generally high and a majority of regions exhibit specificity values above 0.75. In regions where track occurrence is scarce or even absent, the specificity is maximized, while regions with a high number of presence tracks stand out with a comparably low specificity. This trend is also visible in Figure 5.22b, where a negative correlation between track count and specificity can be observed. The spatial pattern exhibited by the specificity is consistent throughout all seasons, and variations among different warning regions is comparably low (0.21), especially when compared to sensitivity, where the standard deviation of all warning regions is approximately doubled (Figure 5.24).

### **Balanced Accuracy**(Figure 5.20)

The balanced accuracy is the arithmetic mean of sensitivity and specificity, hence it is a combination of the two skill scores. High values are found in the center and northeastern part of Switzerland, however there is no clear spatial trend. There is also no clear relationship between track count and balanced accuracy (Figure 5.22c). Very low values however are exclusively found when less than 10 presence tracks were recorded. It can be observed that values exhibit a binomial distribution, with one peak at approximately 0.5 and one peak at approximately 0.7,

thus most values lie between 0.5 and 0.8 (Figure 5.23). This is in line with Figure 5.24, which shows that the standard deviation of the balanced accuracy is smallest of all skill scores, hence there is only little variation among different warning regions.

### **KSS (Figure 5.21)**

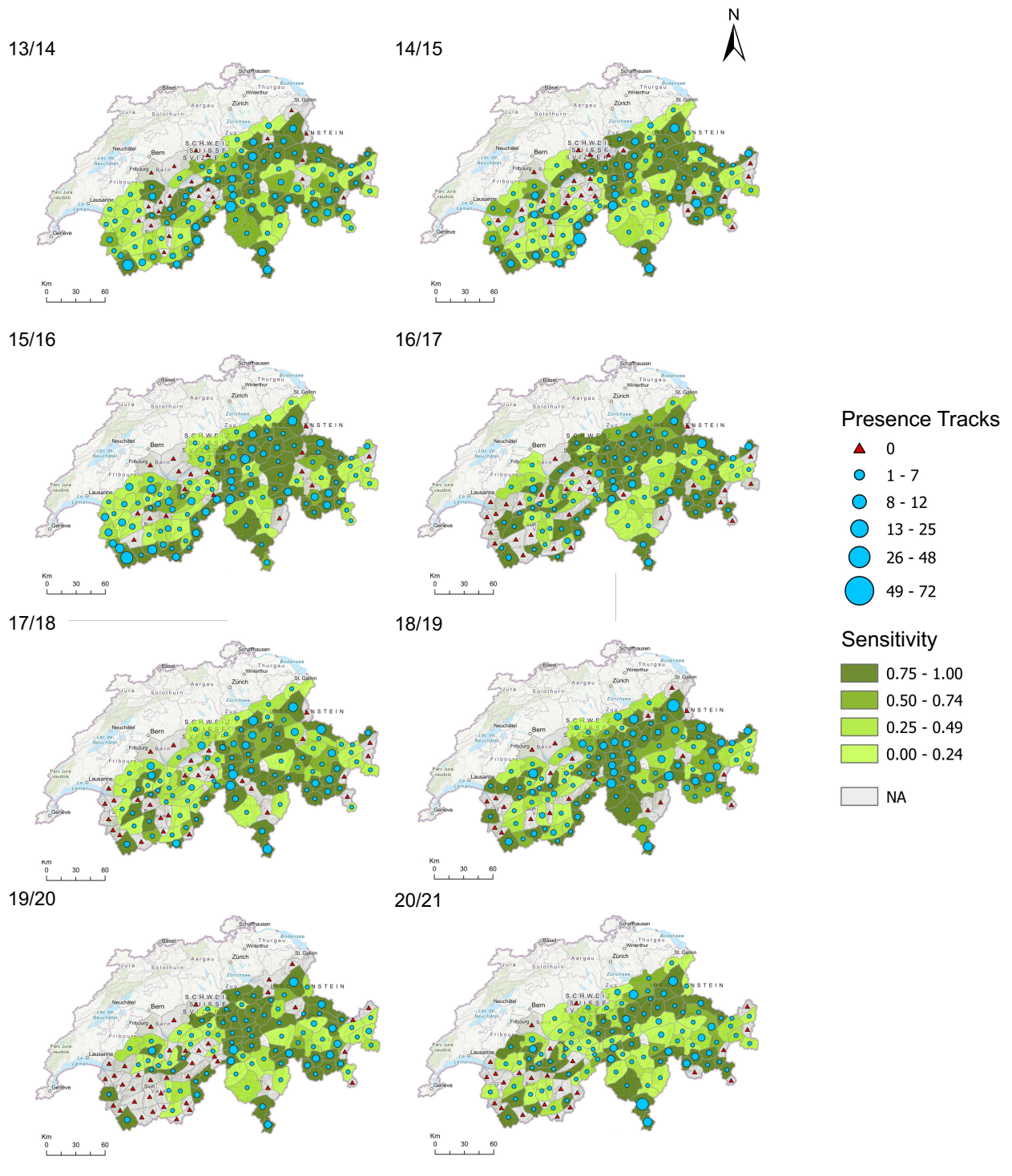
For KSS, there are some hot and cold spots (regions with high, respectively low values), where autocorrelation is present. However, the locations of the spots are not consistent throughout the different seasons and no general pattern can be perceived, similarly to the balanced accuracy. Hot and cold spots appear to be randomly distributed. Figure 5.22 does not show a linear correlation between KSS and the presence track count. Further, the variations among the different warning regions are relatively high, with a mean standard deviation of 0.31. The distribution of values (Figure 5.22d, 5.23d) is qualitatively congruent with the balanced accuracy (Figure 5.22c, 5.23c), only the range of values differs, as it is much wider for the KSS than for the balanced accuracy. This is due to the fact that both skill scores represent a standardized combination of errors from both classes. Analogous to the balanced accuracy, KSS values exhibit a binomial distribution, with one peak in the negative range and one peak at approximately 0.5. It has to be noted that KSS is the only skill score that can take values below zero. A KSS below zero indicates that the product of falsely predicted tracks ( $FN \times FP$ ) is bigger than the product of correctly predicted tracks ( $TP \times TN$ ). In other words, the model performs worse than a random prediction (Ebert and Milne, 2022). Figure 5.22 reveals that all negative values are concentrated in warning regions with less than five total tracks per season.

### **Presence Tracks**

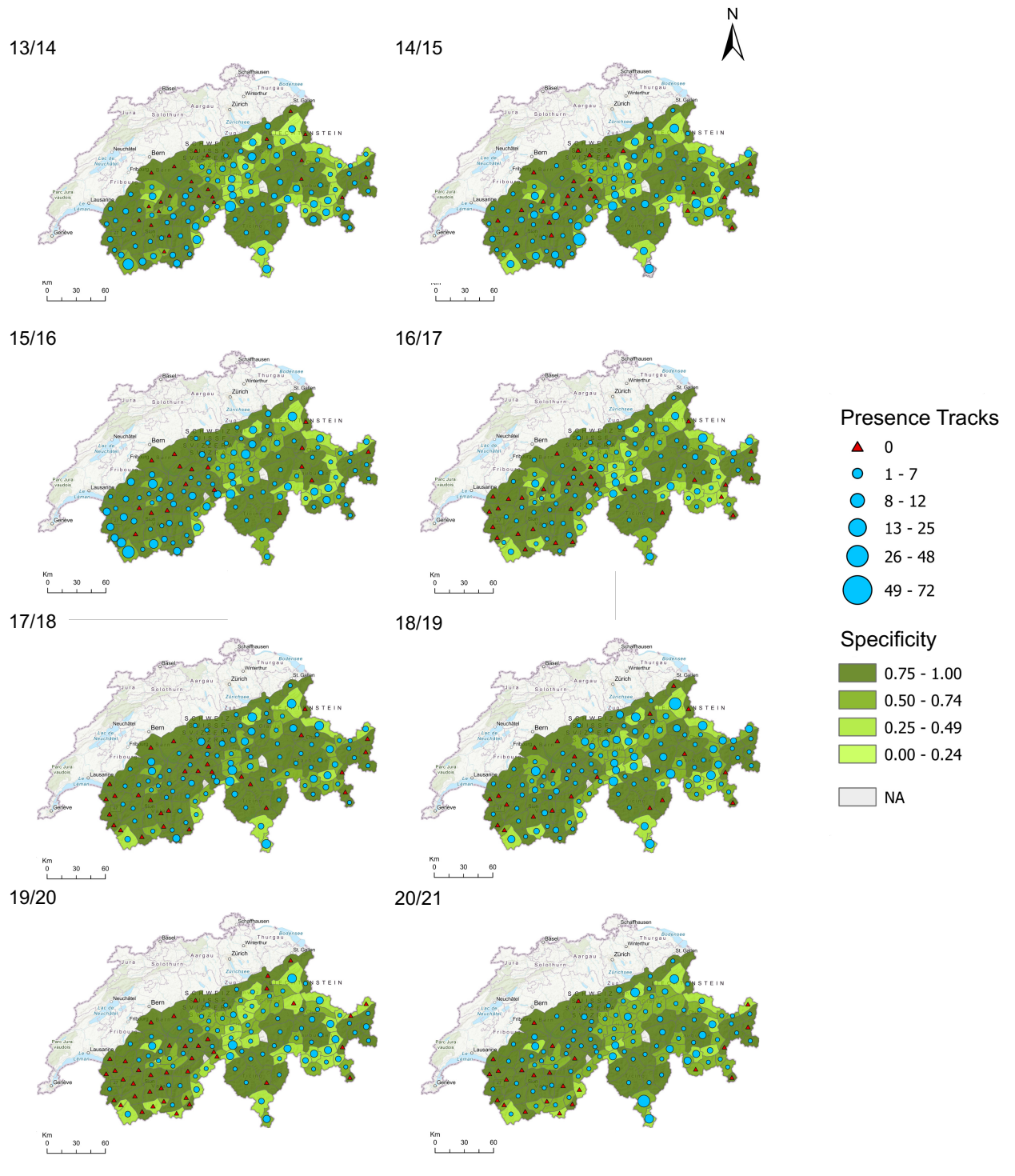
Recorded presence tracks are not uniformly distributed in space but concentrate in a few regions. Most of the tracks were recorded in the central and northeastern parts of the Alps. Less tracks were recorded in the western Alps. Numerous warning regions in the western Alps and in the northern Pre-Alps did not record a single track in at least one season. During the two COVID-19 seasons (2019/20, 2020/21), exceptionally few tracks are recorded in regions in the southwestern Alps (Valais), which are situated near the Italian and French border. Some regions with high track counts can be associated with touristic hotspots like “7114 – Zermatt” or “1222 – Gstaad”. Furthermore, the number of presence tracks impacts both sensitivity and specificity.

In conclusion, residuals in sensitivity and specificity have some spatial autocorrelation, while residuals of the balanced accuracy and the KSS seem to be more randomly distributed. Sensitivity and specificity are inversely affected by the total track number in each warning region. While a high number of presence tracks correlates with high sensitivity, it opposingly correlates with relatively low specificity values throughout all seasons. Put differently, the probability of detection in each class increases with increasing number of data points in the respective class. Therefore, the spatial autocorrelation of sensitivity and specificity is a consequence of the autocorrelation in the presence track distribution. In contrast to that, neither the residuals of the

KSS nor of the balanced accuracy show a clear spatial trend. Further, the standard deviation of the balanced accuracy among the different regions is very small, while bigger variations are exhibited by KSS.

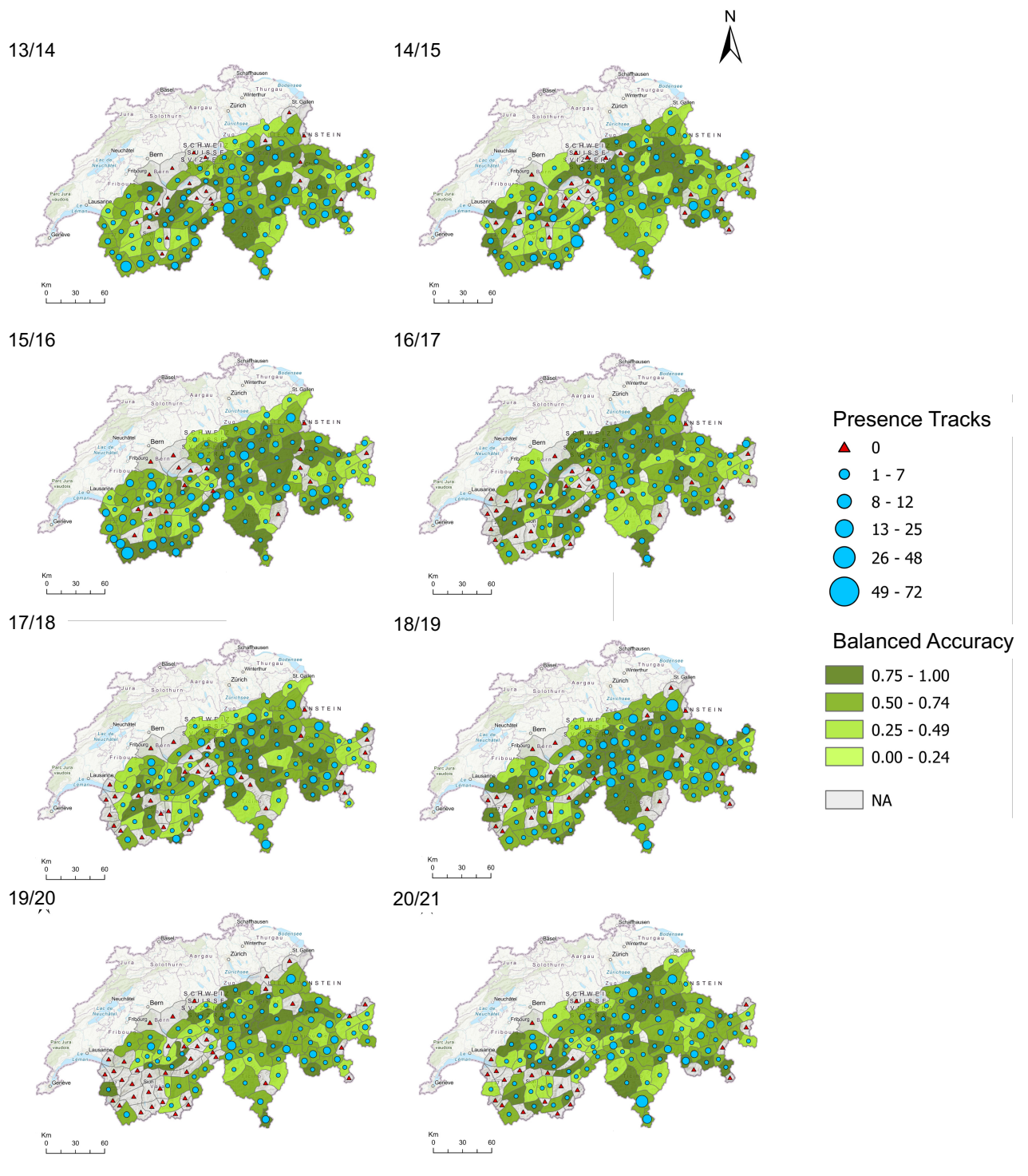


**Figure 5.18:** Sensitivity per warning region with best performing model (Model 5) for all test seasons. NA values are employed when calculation is not possible.

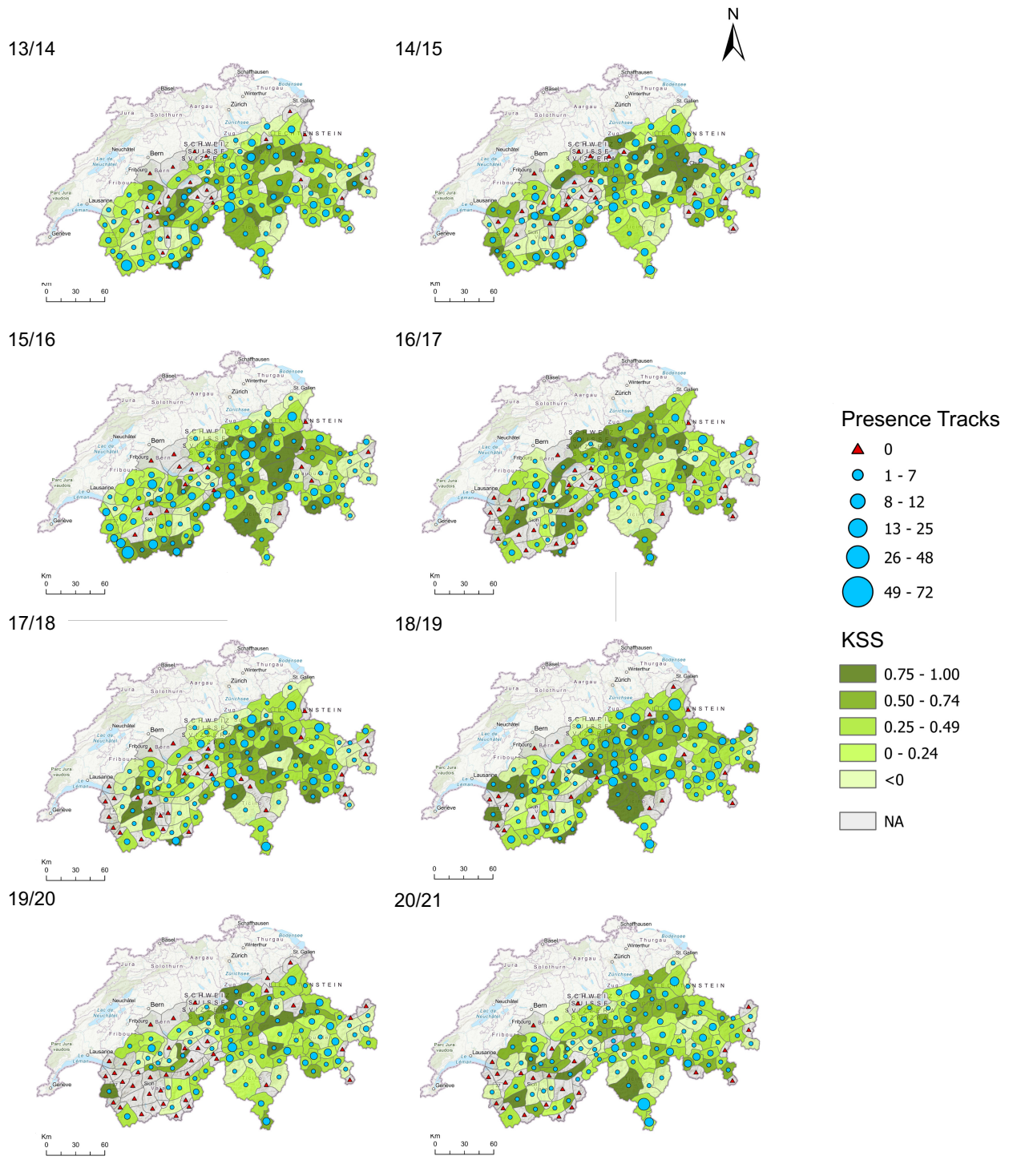


**Figure 5.19:** Specificity per warning region with best performing model (Model 5) for all test seasons.

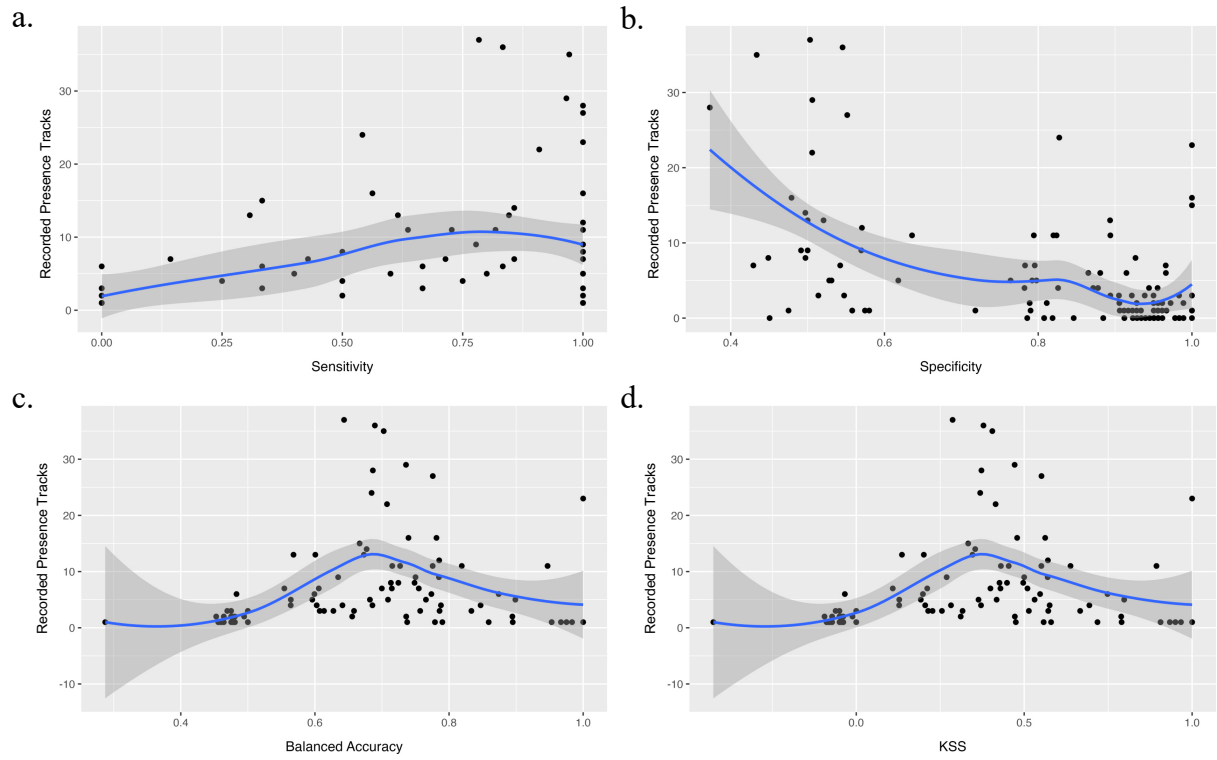




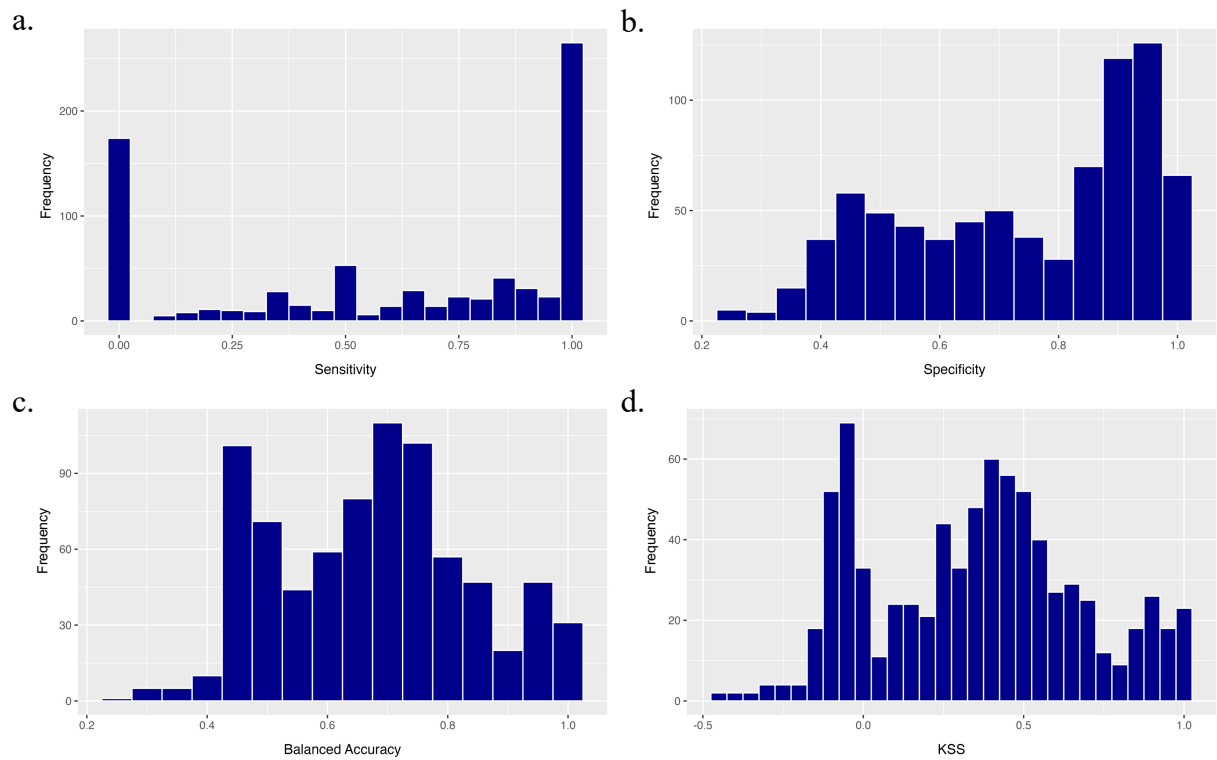
**Figure 5.20:** Balanced Accuracy per warning region with best performing model (Model 5) for all test seasons. NA values are employed when calculation is not possible.



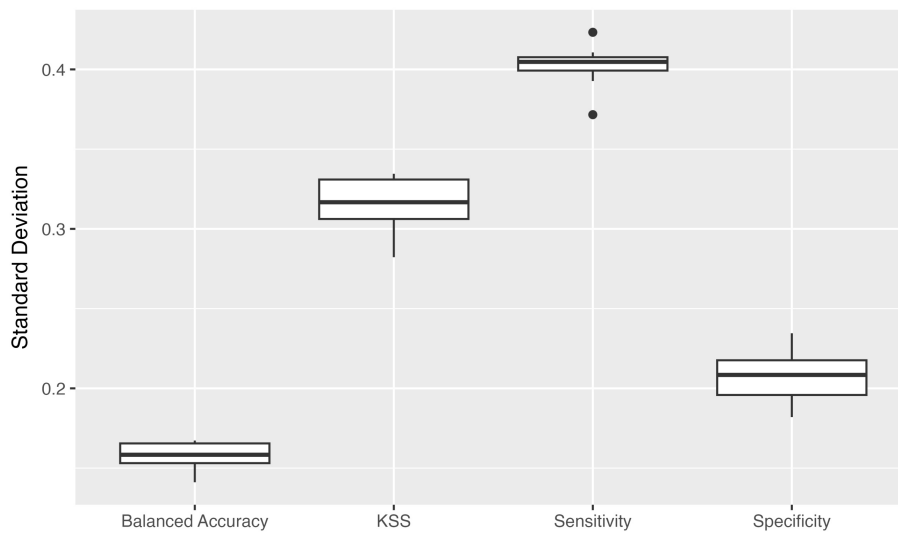
**Figure 5.21:** Hanssen-Kuipers Skill Score (KSS) per warning region with best performing model (Model 5) for all test seasons. NA values are employed when calculation is not possible.



**Figure 5.22:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy, and d. KSS and total track count per warning region, season 2017/18. The blue curve indicates the local polynomial regression line with a 95% confidence interval. Plots for all other seasons can be found in Appendix B.



**Figure 5.23:** Histogram of a. sensitivity, b. specificity, c. balanced accuracy, and d. KSS values from all seasons and warning regions. Note that the bin width is 0.05 for all skill scores, but d. KSS has a greater range than the other skill scores, therefore it is represented by more bins.



**Figure 5.24:** Standard deviation of skill scores among warning regions. Standard deviations are calculated for each skill score and season. Boxplots contain standard deviations of all seasons.



### 5.3.7 Robustness

The model was trained and optimized using season 2015/16 as a test set for validation. However, after the final model was fitted, the robustness of the model in terms of the used test season was assessed by permuting the test season. Table 5.3 shows the skill scores emerging from the different permutations. The standard deviation is less than 0.05 for all skill scores, indicating only small variations among test seasons. A correlation analysis was carried out to determine whether the number of presence tracks per season influences the skill scores. Even though the Pearson Correlation Coefficient is positive for all skill scores, indicating a positive correlation,  $R^2$  is relatively small (below 0.3), and p-values for a one-sided t-statistic are greater than 0.05. This suggests that there is no significant correlation between the number of tracks in each season and the performance of the model. A visual representation of skill scores and total track number is provided in Figure 5.25, which shows visually that there is no correlation between skill scores and track number.

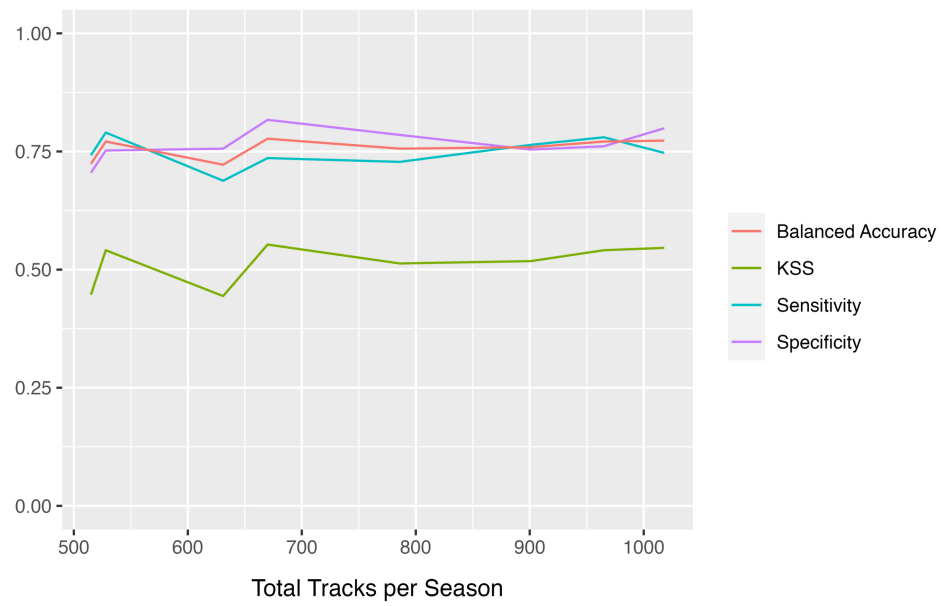
Lastly, in order to assess the robustness of the model against uncertainties in the input data, the model was trained with data that contained a known input uncertainty. The uncertainty was introduced into the meteorological variables, as they are believed to hold the highest uncertainty. Meteorological variables are measured by a sensor, then interpolated with a complex statistical method and further averaged over the elevation belt of each track. The uncertainty in other input features, such as weekday or the avalanche danger level, are considered to be very small. Table 5.4 shows the model performance in terms of the four skill scores obtained by applying the intentionally altered meteorological variables to the fitted model, as well as the deviation from the best scores. The results suggest that the deviations are minor (below 0.02) for all skill scores. However, sensitivity is stronger impacted than specificity, as the deviation from the best model is approximately 16 times higher for sensitivity compared to specificity (Table 5.4). KSS suffered the highest performance loss (-0.018), while the balanced accuracy was only moderately decreased (-0.008).

**Table 5.3:** Skill scores produced by the best performing model (Model 5) with permuted test seasons. For each test season, the complementary data is used for training. Mean and standard deviation of each skill score is provided at the bottom, as well as results from the correlation analysis.

Test Season	Sensitivity	Specificity	KSS	Balanced Accuracy	# Tracks per Season
<b>13/14</b>	0.764	0.754	0.518	0.759	901
<b>14/15</b>	0.728	0.785	0.513	0.756	786
<b>15/16</b>	0.743	0.800	0.543	0.771	1018
<b>16/17</b>	0.790	0.752	0.541	0.771	528
<b>17/18</b>	0.736	0.817	0.553	0.777	670
<b>18/19</b>	0.780	0.761	0.541	0.771	965
<b>19/20</b>	0.742	0.705	0.447	0.724	515
<b>20/21</b>	0.688	0.756	0.444	0.722	631
<b>Mean</b>	0.746	0.766	0.512	0.756	751
<b>STDEV</b>	0.032	0.034	0.043	0.022	195
<b>Pearson's r</b>	0.194	0.453	0.498	0.498	–
<b>R<sup>2</sup></b>	0.04	0.20	0.25	0.25	–
<b>p-value</b>	0.32	0.13	0.10	0.10	–

**Table 5.4:** Skill scores for a model with a known input uncertainty of meteorological variables compared to the best performing model (baseline).

	Sensitivity	Specificity	KSS	Balanced Accuracy
<b>Baseline</b>	0.743	0.800	0.543	0.771
<b>Meteo Uncertainty</b>	0.727	0.799	0.525	0.763
<b>Difference</b>	– 0.016	–0.001	–0.018	–0.008



**Figure 5.25:** Total tracks per season vs. skill scores.

# Chapter 6

## Discussion

This chapter analyzes the results of this thesis in terms of the research questions posed in Section 1.2. First, the model building and the usability of the model is discussed (6.1). Further, the most important variables for the prediction of backcountry skiing activity (6.2) and limitations of the model (6.3) are presented. Practice advice for decision makers in the avalanche forecasting domain is proposed in Section 6.4. Ultimately, the results are put into the context of literature, and possibilities for future research are introduced (6.5). The research questions are discussed in the following sections:

**RQ1:** *How can backcountry skiing activity be modeled using obfuscated, user-generated trajectory data?* → **Sections 6.1, 6.3**

**RQ2:** *How do the different predictors, such as weather forecast, avalanche conditions, and free time, influence the backcountry skiing activity?* → **Section 6.2**

**RQ3:** *What are the most important predictors for backcountry skiing activity?* → **Sections 6.2, 6.4**

### 6.1 Model - Building, Performance, Usability

#### Model Building

The challenges in the model building process were twofold: 1.) The construction of the training data and 2.) the model building process itself.

To construct the training data, the tracks needed to be enriched with additional attributes. While this was quite straightforward for attributes like weekday, day of the season, and holiday, it took more effort for the meteorological variables. Since the tracks were obfuscated, it was not possible to add variables purely by location. The only location information – the warning region – is a rather broad geographic context and simply taking the average of the whole region would not have been representative of a certain track, since meteorological variables can vary quickly in space, especially in Alpine regions (Scherrer and Appenzeller, 2014; Spreafico and Weingartner, 2005), and backcountry skiing usually takes place in a certain elevation band. As

discussed in Chapter 4, meteorological variables were added to the tracks by using the elevation information and the warning region. Meteorological variables usually come in a netCDF format, for which the R environment provides some useful packages. However, there were some issues with the netCDF files provided by MeteoSchweiz. For instance, the coordinate reference system (CRS) was not correctly stored for the precipitation data, which led to problems with raster computations<sup>1</sup>. Additionally, coordinates were stored differently for different meteorological variables, some storing the data starting from the top left corner of the map, others storing the data starting from the bottom left corner of the map. These issues have been ultimately solved by manually altering the CRS information and the way the netCDF files were read (top vs. bottom left corner).

Another challenge with the meteorological variables was that since the variables needed to be intersected with both the DHM and the warning region, it was computationally rather extensive, especially for the absence tracks, which were much more abundant than the presence tracks. This issue intensified for the precipitation data, for which some additional steps were necessary in order to sum up the hourly precipitation in the morning. It took a substantial amount of time to optimize the process and required implementing and comparing similar raster-functions of different packages, as there are considerable differences in terms of computational time among different implementations of the same function. In summary, the issues with the meteorological attributes primarily stem from technical aspects rather than conceptual ones.

The second set of challenges was related to the model building process. The main challenge was to deal with the imbalanced data, and to find a verification metric for the prediction that was not influenced by the class imbalance. This problem required a solution on the conceptual level (i.e., which verification measure is suitable) as well as on the technical level (i.e., how to adjust the algorithm to deal with imbalanced data). For the conceptual part, a solution was found through literature research, while for the technical part the best parameter setting was found by trial-and-error.

## Skill Scores

The best performing model was able to successfully detect 74% (80%) of the presence (absence) tracks. The model yielded a balanced accuracy of 77% and a KSS of 0.54, where 1 indicates a perfect score and 0 indicates an unskilled prediction. The balanced accuracy produced by the final model (77%) is slightly below the accuracy exhibited by prediction models that used weather data to predict (recreational) traffic congestion (85 – 95%) (Lee et al., 2015; Lingras et al., 2002; Liu and Wu, 2018). A study that resembles the problem set-up of this thesis, predicting recreational use of urban forest sites, yielded a 90% accuracy (Dwyer, 1988). Other

---

<sup>1</sup>Other users encountered the same problem, as an example from StackOverflow, a website for the online community of developers, shows. Reto Stöckli, who self-identified as the author of the data at MeteoSchweiz, confirms in the comment section that they have been experiencing difficulties with the netCDF files since they started using the new Swiss projection (LV95/CH1903+), and apparently, they have not found a solution yet. <https://stackoverflow.com/questions/71310611>

studies, e.g., studies on ski tourism prediction, used performance metrics only applicable to regression models (e.g.,  $R^2$ ), consequently they cannot be compared to the results of this thesis, because a classification approach was applied in this thesis (King et al., 2014; Riddington, 2002). Considering all limitations that will be described in more detail in Section 6.3, the level of performance is satisfactory, although being slightly worse than other models in related literature.

The challenges of the modelling process are manifold. One challenge was to find a suited skill metric to verify the predictions. Owing to the class imbalance, a simple estimate of the overall accuracy did not provide useful results, as the minority class was simply outnumbered by the majority class, as supported by existing literature (He and Garcia, 2009). It therefore required a measure that focuses on the minority class or at least is not influenced by class imbalance. Literature about rare and severe (weather) forecasting offers skill scores that fulfill this requirement, and that were ultimately chosen for this thesis: The balanced accuracy, a simple combination of sensitivity and specificity, and the KSS, a measure specifically tailored to rare and severe event forecasting (Ebert and Milne, 2022; Hanssen and Kuipers, 1965; Heierli et al., 2004; Marsland, 2015). While it is evident that false positives are preferred over false negatives for rare and severe events such as tornadoes, this inference is less straightforward when considering backcountry skiing. On the one hand it is ambiguous whether one error type is substantially more serious than the other, on the other hand it is questionable how rare backcountry skiing activity really is. Therefore, the balanced accuracy is possibly more suited as a verification measure for this problem. KSS and balanced accuracy show a very similar behavior among the different model implementations (Table 5.3), which suggests that they are equally as good in detecting the best performing model. However, as the requirements for the KSS are not necessarily given in the context of backcountry skiing, it should only be used to relatively compare different models, but not for making an absolute assumption on the performance of the model.

### **Spatial Dimension**

The model showed similar performance when applied to different seasons, which shows that it is not influenced by seasonal variations. The only exceptions to this are the two COVID-19 seasons (2019/20, 2020/21), which will be discussed later in more detail.

The residuals of both KSS and balanced accuracy are visually randomly distributed in space. This is an indicator of how well the model captures the relevant factors, as the structural problems of a model can be reflected by the autocorrelation of its residuals (Chen, 2016). Both KSS and balanced accuracy yield lowest values in regions where less than 10 tracks were recorded in the observed season. This suggests, that the performance of the model is low when presence tracks are missing.

For sensitivity and specificity on the other hand, certain spatial trends could be detected. Generally, more presence tracks are successfully detected in the northwestern part of the Alps, which can be explained with a higher frequency of total tracks in the region. Conversely, specificity is highest (often reaching values near 100%), in regions with small numbers of presence tracks

(e.g., western Alps). It becomes evident that the model reaches higher scores when predicting absence than presence. Given that the class imbalance problem was already addressed in the training process by using equally sized samples from both classes, it raises the question why the absence class still outperforms the presence class. The first thing to be pointed out is that even though the model was trained with artificially balanced classes, the validation data to which the model was applied, was not balanced, but contained the initial class imbalance. This led to a high number of false positives, which is a well-known problem in machine learning with imbalanced classes (e.g., Makki et al., 2019). Consequently, false negatives are less frequent than false positives while true negatives are more frequent than true positives, which is ultimately reflected in higher specificity compared to sensitivity. Yet, there is another factor that may be far more contributory to this problem. For the generation of the absence data, the assumption was made that “absence of evidence is evidence of absence”. In other words, it was assumed that if no track is recorded, there are no backcountry skiers in the field, which was represented as zero backcountry activity. The issue with this assumption will be discussed in more detail in Section 6.3.

### Temporal Dimension

The presence data revealed that backcountry skiing activity is a highly temporal phenomenon, with a typical weekly and seasonal cycle. The results show that this temporal dimension is well captured by the model. More activity is predicted on the weekends and in the mid-season, compared to workdays and beginning/end of the season. Furthermore, for periods of bad skiing conditions due to unfavorable weather (e.g., January 2016 or March 2019) low presence-probabilities were correctly predicted, which correlates with low presence track counts. Generally, probability for presence and actual recorded presence tracks are positively correlated throughout all seasons, which demonstrates the effectivity of the model. There is however also a time period where presence track counts do not correlate with the predicted probabilities. In spring and especially in March 2020, the first COVID-19 lockdown with very strong mobility restrictions led to a decrease in presence tracks but not in probability of presence, because the COVID-19 restrictions are not incorporated into the model. In these months, backcountry skiing conditions were theoretically good, but mobility was restricted to a minimum, which led to a significant decrease in presence tracks even though the model predicted otherwise. Additionally, in Figures 5.18 – 5.21 it is visible that presence tracks in warning regions near the French and Italian border experienced a strong decrease in track counts during the two seasons after the outbreak of the pandemic, while the decrease was less pronounced in regions more in the center of Switzerland. Mobility restrictions as experienced in Switzerland during the pandemic can however easily be incorporated into the model in the future.

## 6.2 Which Variables are the Best Predictors?

Variable importances differ among the two classes. For the presence class, popularity, avalanche conditions (RDL), and  $S_{rel}$  are most important. For the absence class, the meteorological

variables are the most important, whereas for the presence class only one ( $S_{rel}$ ) out of three meteorological variables ranks among the top three. A reason for this could be that even though there are weather conditions that are clearly associated with backcountry skiing activity or a lack thereof, the distinction between activity and no activity is not always straightforward. Days with bad weather conditions (i.e., high precipitation sums, very high/low temperatures) often exhibit only few presence tracks. Yet, there are bad weather days on which a track is recorded, as well as many days where the weather conditions are favorable, but no track is recorded. Even though only  $S_{rel}$  ranks among the most important predictors for presence, the findings of Haugom and Malasevska (2019) could be confirmed, as the results suggest that the optimum weather conditions for backcountry skiing are sunny conditions (high  $S_{rel}$ ), no precipitation and an air temperature of  $-5^{\circ}\text{C}$  (Figure 5.15).

Further, the results suggest that certain avalanche danger levels eliminate the possibility of a presence track with a very high certainty, for example when no avalanche bulletin is published (danger level 0) or when the avalanche danger is very high (danger levels 4 and 5). On such days, activity is very unlikely to occur. On the other hand, there are many days where the avalanche conditions are favorable, yet no presence track is recorded, similarly to days with good weather conditions. However, it is important to notice that even though all three danger levels (0, 4, and 5) exhibit low probabilities for backcountry skiing activity, the situation is different for low and high avalanche danger: While the lack of snow impedes the possibility of backcountry skiing on days with danger level 0, the avalanche danger seems to be what stops recreationists from going outside on days where backcountry skiing would be physically possible but the danger level is high or very high. While recreationists remain in a passive state on days without snow (the decision is made for them), they actively make the decision to not venture outside when the danger level is high. Consequently, the difference between absence and presence is more pronounced on days with danger level 0, as backcountry skiing is physically not possible, compared to danger levels 4 and 5, where some (but not many) people still go outside.

Even though some meteorological conditions decrease the likelihood of a presence track, a considerable number of tracks is still recorded on days with bad weather conditions, whereas hardly any tracks were recorded on days with critical avalanche conditions (due to the decision-making of the skier) or on days without an avalanche bulletin (due to the lack of snow). This suggests that the snow condition serves as a stringent criterion for individuals when deciding whether to venture outside, whereas the weather condition is a flexible criterion that is desirable but not essential. It has to be pointed out here, that the snow condition, rather than the avalanche danger itself is the stringent criterion, as on days with danger level 0, technically not the avalanche danger but the lack of snow is what stops recreationists from going outside.

Furthermore, the popularity is the most important variable for the presence class. This shows that activity is bound to popular and well-accessible regions, often near population centers. As the popularity is not dynamic over time, it would be worth excluding it from the model as a predictor and only focus on the most popular regions. This might increase the performance



of the model because it excludes regions where presence data is sparse and focuses on regions where data is more abundant, and the prediction thus more reliable.

The weekend variable is the fourth most influential for the presence class. This is due the majority of tracks being recorded on weekends, indicating a strong correlation between activity and weekends. Holidays on the other hand are among the least important predictors for both classes. A problem with the holiday variable is that for holidays that fall on a workday, the weekend variable is still 0, indicating that it is technically a workday, which it is not. The solution for this would be to set the weekend variable to 1 when a holiday falls on a workday, as it is not a workday. Another issue with the holiday predictor is that holidays differ between different political entities (Cantons) in Switzerland. As many recreationists head to other Cantons for backcountry skiing, it is difficult to relate warning regions to specific holidays. Additionally, some people work part time, hence they have an additional day off, which is not captured by the weekend or the holiday predictor. School holidays could also be incorporated into the model, though they are also dependent on the Canton, and are therefore harder to implement. For that, the holiday behavior of Swiss citizens could be studied, as people living in the French part of Switzerland possibly choose different destinations than people living in the German-speaking northwestern part of Switzerland.

Ultimately, it has to be noted that the variable importance of the absence class, and especially in comparison with the presence class, has to be interpreted with care, as there are some limitations to the absence data that will be discussed in the next section.

### 6.3 Limitations

Even though the model performs relatively well, there are some limitations to it. The main constraint lies in the data used for training. Interestingly, the obfuscation of the data was not the main issue, but rather the quantity of the data and the classification into absence and presence tracks. The underlying premise for the generation of the absence data was that absence of evidence is evidence of absence. But what was echoed by numerous scientists across various disciplines before seems to be applicable to this particular topic as well: Absence of evidence is not evidence of absence. Certainly, there have been many days where no track was recorded, yet people were backcountry skiing. Because such days were treated as absence tracks in this thesis, the input data for the model was initially flawed. The question is raised again: How rare are the presence tracks really? Even though the answer to this question is difficult to find, we can confidently say that they are not as rare as the data suggests, especially when considering the participation bias VGI data holds (Nielsen, 2006). This indicates that the data imbalance is caused by external factors (extrinsic imbalance) rather than inherent characteristics of the data (intrinsic imbalance) (He and Garcia, 2009). In other words, the imbalance in the data is likely not due to the nature of backcountry skiing, but due to the nature of data collection. The approach, where the absence data was generated as a complement to the presence data, would only work in a perfect setting in which every single backcountry tour ever carried out is recorded.

However, since a perfect setting is always a utopia, alternative methods are required. One approach would be to use data that describes the behavior of users on a website dedicated to backcountry skiing. The assumption that people do research on a route they plan for the next day is probably more valid than the assumption that absence of evidence is evidence of absence. Data that reveals how many times a backcountry skiing route is clicked on could possibly be used to approximate backcountry skiing activity for the next day, with the assumption that most people have a last look at the planned tour the night before. Such a dataset exists for the Skitouren guru website<sup>1</sup>, from which the data used for this thesis was obtained. The Route Click Statistics Dataset (RCS D) reveals the click statistic of the Skitouren guru website for many popular backcountry skiing routes in Switzerland. It could be used similarly to the ARPD to train a model.

A different approach would be to use remote-sensing data to detect backcountry skiing activity. Remote sensing data was already used in avalanche research to map the spatial distribution and size of avalanches (Hafner et al., 2021) or to measure snow depth (Bührl e et al., 2022). Zweifel et al. (2006) used satellite imagery to detect fresh tracks on the snow cover, from which backcountry skiing activity could be inferred. However, they argue that satellite imagery is costly and only applicable in conditions with optimal visibility. As an alternative to satellite imagery they suggest airborne photography, which is less impacted by clouds hindering visibility, but is also rather expensive.

Another limitation of the model is that the spatial resolution of the warning region is too small for the quantity of data used for training. On average, 750 tracks were recorded each season. Consequently, in many warning regions and on many days, zero tracks were recorded, which only allowed a binary activity classification. In practice, it would be more accurate to employ a probabilistic approach to estimate the likelihood of individuals being in the field rather than a binary classification. Another approach would be to expand the spatial entities used for aggregation to larger areas, for example the seven climatic regions (Figure 3.2). However, it is questionable to which extent spatially varying phenomena like avalanche or weather conditions can be representatively aggregated in larger areas. Alternatively, the focus could be shifted from all warning regions to only a fraction of warning regions that are known to be frequently visited. Yet, excluding warning regions with only very few presence tracks did not enhance the performance of the model.

Lastly, the model could potentially be improved by using additional meteorological predictors, such as wind or the amount of fresh snow. An additional predictor describing the overall mobility in the country could address for example local school holidays or COVID-19 restrictions. However, the primary limitation is assumed to stem from the labelling of the data with absence and presence, rather than from the choice of predictors.

---

<sup>1</sup>[www.skitouren guru.ch](http://www.skitouren guru.ch)

## 6.4 Implications for Practitioners

In this section, practice advice for decision makers is presented. It is especially relevant to avalanche forecasters, as well as to researchers studying backcountry skiing behavior and avalanche risk.

As discussed in the last section, a probabilistic approach rather than a binary classification should be applied. Therefore, this section gives an overview over some rules of thumb to estimate the probability of individuals being in the field under certain conditions. The rules of thumb were created by applying the model onto different scenarios. The scenarios include weather, avalanche conditions, and free time, as they have been shown to be the most influential dynamic predictors. The comparison is made between good and bad conditions. The terms “good” and “bad” denote conditions where the probability of activity is expected to increase (good scenarios), respectively to decrease (bad scenarios). By comparing two different scenarios, where all features except the feature of interest are kept constant, an empirical Probability Change Factor (PCF) can be calculated<sup>1</sup>. The PCF denotes the relative change in probability of activity due to the change of the given conditions, which is consistent with the relative change in probability revealed by the partial dependence plots (Figure 5.15). Therefore, the PCF can be determined by calculating the relative change in probability revealed by the partial dependence plot. Table 6.1 gives an overview over PCFs for the three different scenarios. It has to be mentioned that the PCF is only applicable for RDL greater than 0, as a RDL = 0 rules out presence with high confidence.

The PCF is 2.5 from bad to good weather, 1.5 from low to high avalanche hazard and 1.5 from workday to weekend. This is in line with Wegelin et al. (2022), who found that sunny weather, followed by the day of the week have the strongest impact on visitor numbers of a touristic mountain in Switzerland.

The partial dependence plot (Figure 5.15b) shows, that for RDL, the probability increases strongly from 0 to 1, and decreases again strongly from 2 to 4. Levels 1 and 2, as well as levels 4 and 5 are approximately equally likely. When two or more situations are combined in a way that the positive effect on the probability is amplified (i.e., combine good with good conditions), the factors can be approximately multiplied. This means that the probability is (2.5 x 1.5) times higher on a weekend with good weather compared to a workday with bad weather. Accordingly, the probability gets divided by the factor when two situations with balancing effects are combined (i.e., combine good with bad conditions). For example, the probability is increased by a factor of (2.5 / 1.5) on a weekend with bad weather compared to a workday with good weather. The PCFs suggest that weather has the biggest influence on the activity, which is inconsistent with the variable importance discussed in Section 6.2. It was shown that RDL is, after popularity, the second most important predictor. There are several reasons for this inconsistency. First, the PCF of weather (2.5) is a combination of multiple factors (Temperature, Precipitation, and

---

<sup>1</sup>The calculation follows the equation:  $\text{Probability}(\text{bad}) \times \text{PCF} = \text{Probability}(\text{good})$ .

$S_{\text{rel}}$ ), which reinforce each other, while the PCF for RDL is only based on one predictor. RDL exhibits a higher PCF than each weather variable individually, but not when they are combined. Secondly, we must consider the baseline distribution of the RDL. Danger levels 4 and 5, the highest danger levels, are also the ones that are issued least often. These levels are rare, not only in the presence data, but also in the absence data. As a result, the model faces challenges in distinguishing between absence and presence when the danger level is high, because of the limited number of both absence and presence tracks in the training data. Therefore, the PCF from RDL 2 to 4 is possibly underestimated by the model, due to the underlying baseline frequency. On the other hand, for danger level 0, which is employed when no avalanche bulletin is issued, the distinction between presence and absence is very clear, as hardly any presence tracks take place under these conditions, but a lot of absence tracks exhibit RDL 0. The probability increase from danger level 0 to 1 is 2.6, which is not included in the PCF of RDL. This explains why the variable importance of RDL is high, even though the PCF is smaller than the PCF for weather conditions.

For the other two scenarios (free time and weather), the baseline frequency is more uniform. Weekdays naturally occur equally often, and good/bad weather is also more or less balanced (strongly depending on the season). Therefore, the PCF for these scenarios is possibly more reliable. In conclusion, the RDL is an important predictor, but due to the baseline frequency and the danger level 0, its impact is more pronounced in lower danger levels compared to higher danger levels. The PCFs give an idea of the impact a scenario has on the probability for presence, however they have to be used with caution, as there are some limitations to them.

**Table 6.1:** Probability Change Factor (PCF) of good vs. bad weather, avalanche, and free time conditions. PCFs should only be applied to scenarios with RDL  $\geq 0$ .

Scenario		Definition	Probability Change Factor (PCF)
Weather	Good	$S_{\text{rel}} = 100\%$	2.5
		Precipitation = 0 mm	
		Temperature = $-5^{\circ}\text{C}$	
	Bad	$S_{\text{rel}} = 0\%$ Precipitation = 5 mm Temperature = $-10^{\circ}\text{C}$	
Avalanche	Good (low danger)	RDL = 2	1.5
	Bad (high danger)	RDL = 4	
Freetime	Good	Workday	1.5
	Bad	Week	

## 6.5 Implications

As Techel et al. (2015) pointed out, it is crucial to consider exposure when studying avalanche risk. It is necessary to know how many people are in the field in order to put avalanche accidents

or avalanche reports into context. Certainly, an avalanche accident is rated differently on a day where many people are in the field compared to a day where only very few people are in the field. Also, when no avalanche reports are received on a given day, this information holds less uncertainty the more people are in the field who could have possibly witnessed and reported an avalanche. Conversely, if there are no people in the field to potentially witness an avalanche, the uncertainty of the same information is higher. Zweifel et al. (2006) carried out one of the few studies in Switzerland, where they attempted to count backcountry skiers (hence backcountry skiing activity) on a daily basis. However, the study focuses on a few particular regions and analyses the activity *ex post*. This thesis provides a first experimental attempt to model and predict backcountry skiing activity *ex ante*, with the use of backcountry skier counts deduced from VGI data.

The results obtained in this thesis are in line with previous studies on the behavior of backcountry skiers. The probability of activity is increased on days with good weather conditions compared to days with bad weather conditions, which is in agreement with Haugom and Mala-sevska (2019), King et al. (2014), and Ruddy and Andrey (2014). Concerning the avalanche danger, the probability is decreased by a factor of 1.5 from danger level 2 to 4, which is a smaller decrease than found by Zweifel et al. (2006), but still confirms that people are more likely to go outside when the avalanche danger is lower. Further, the likelihood of activity is increased on the weekend compared to workdays by a factor of 1.5. Similar results are found by Techel et al. (2015), who found that the number of backcountry skiers is 2.3 times higher on weekends compared to workdays. It is worth noting that the PCF is derived from an estimation of the activity likelihood, rather than the absolute count of backcountry skiers. Consequently, the PCF values are only limitedly comparable to studies that focus on the increase in the absolute number of recreationists.

As described in Section 6.3, there are diverse limitations to the model. There is possibly a substantial misclassification error in the initial training data due to the conceptual idea used for the absence data, and presence data is generally too sparse in time and space for the detailed spatial level the predictions take place on. However, even with all the limitations and the fact that the data was obfuscated, a reasonably good result was obtained, where in approximately 77% of the cases the activity could be correctly predicted. As discussed in Section 6.1, this aligns with other recreational prediction models in literature (Dwyer, 1988; Lee et al., 2015; Lingras et al., 2002; Liu and Wu, 2018). Nevertheless, better results are expected with the use of more presence track data for training. Because the backcountry skiing data used in this thesis relies on people that voluntarily share information about their own tour, getting access to such data can be challenging. Especially trajectory data is highly sensitive (e.g., de Montjoye et al., 2013) and many people are not willing to share such data without a direct benefit, and rightfully so. Remarkably, the obfuscation of the data did not hinder the modelling process, which suggests that sensitive trajectory data may not be necessary for this type of prediction. This thesis shows that the date and the warning region of a recorded tour provides enough information to roughly estimate where and when people go backcountry skiing. However, it

also demonstrates that the biggest uncertainty lies in the nature of the VGI data. Assuming VGI data to depict reality is a risky assumption, given the variations in quality and largely unknown motivations of contributors (Goodchild, 2007; Senaratne et al., 2017). The participation bias, which is often inherent to user generated data, leads to the underestimation of actual numbers of recreationists in the field (Nielsen, 2006). Additionally, it is not clear, whether data obtained through VGI provides a representative sample for the behavior of backcountry skiers, as motivation may differ among individuals (Epstein et al., 2015), which has an impact on credibility the data (Flanagin and Metzger, 2008). Zweifel et al. (2006) found that only a fifth of all backcountry skiers were willing to use a voluntary registration board on site when carrying out a ski tour. Hence using a correction factor on the VGI data could possibly create a more accurate representation of individuals in the field. A correction factor would however only work for days and regions, where at least one track was recorded. When there is no data at all (which was often the case), activity cannot be inferred.

Lastly, this thesis showed that the modelling of backcountry skiing overlaps with leisure modelling in many regards. Popular factors studied in leisure prediction, outdoor recreation, and recreational traffic prediction, such as weather, free time, seasonality, and popularity/accessibility of the region were effectively used to predict backcountry skiing activity (Dwyer, 1988; King et al., 2014; Lee et al., 2015; Liu and Wu, 2018; Riddington, 2002; Rutty and Andrey, 2014; Verbos et al., 2018; Wegelin et al., 2022). The importance of weather in outdoor recreations as stated by Verbos et al. (2018) and Wegelin et al. (2022) could be confirmed. The only predictor exclusive to backcountry skiing is the avalanche danger, which could easily be incorporated into the model.

## Chapter 7

# Conclusions and Further Work

Although the model created in this thesis has some limitations, it provides a solid basis for backcountry skiing activity prediction, as it represents pioneering work in this research field. Additionally, this thesis provides an approach for effectively incorporating weather variables into geographically obfuscated data. It further allows 1.) an approximate estimation of whether backcountry recreationists are in the field (i.e., backcountry skiing activity) and 2.) an idea of which conditions enhance or diminish the probability for activity. The model reveals that both the relative sunshine duration and the snow conditions strongly impact backcountry skiing activity. Further, there is a clear positive correlation between probability of presence and actual presence tracks for all observed seasons, which demonstrates the ability of the model to predict backcountry skiing activity. It could be shown that the model is robust in terms of different validation seasons as well as to input uncertainty of meteorological features. However, a common saying in the machine learning domain is that “a model is only as good as the data it is fed”. This thesis shed light on the limitations of the backcountry skiing data used for the model, as well as on the conceptual idea, from which the absence data was generated. In the end, an outlook on possibilities for future research was given. Alternative data sources were presented, which could be investigated in the future to train an updated version of the model.

# Bibliography

- Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2015). Classification with class imbalance problem: A review. *International Journal of Advances in Soft Computing and its Applications*, 5(3), 176–204.
- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: An overview. *Journal of physics: conference series*, 1142, 012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>
- Atkins, D. (2000). Human factors in avalanche accidents. *International snow science workshop*, Big Sky, MT, USA, 46–51.
- Barnett, A. (2020). Aviation Safety: A Whole New World? *Transportation Science*, 54(1), 84–96. <https://doi.org/10.1287/trsc.2019.0937>
- Baston, D. (2022). Exactextractr: Fast extraction from raster datasets using polygons [R package version 0.9.1]. <https://CRAN.R-project.org/package=exactextractr>
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation Measures for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, 3(10), 27–38.
- Bielański, M., Taczanowska, K., Muhar, A., Adamski, P., González, L. M., & Witkowski, Z. (2018). Application of GPS tracking for monitoring spatially unconstrained outdoor recreational activities in protected areas – A case study of ski touring in the Tatra National Park, Poland. *Applied Geography*, 96, 51–65. <https://doi.org/10.1016/j.apgeog.2018.05.008>
- Bivand, R., Keitt, T., & Rowlingson, B. (2023). Rgdal: Bindings for the 'geospatial' data abstraction library [R package version 1.6-6]. <https://CRAN.R-project.org/package=rgdal>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Bührle, L. J., Marty, M., Eberhard, L. A., Stoffel, A., Hafner, E. D., & Bühler, Y. (2022). Spatially continuous snow depth mapping by airplane photogrammetry for annual peak of winter from 2017 to 2021. *The Cryosphere Discussions*, 2022, 1–37. <https://doi.org/10.5194/tc-2022-65>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, Y. (2016). Spatial autocorrelation approaches to testing residuals from least squares regression. *PLoS ONE*, 11(1), 1–19. <https://doi.org/10.1371/journal.pone.0146865>



- Dahouda, M. K., & Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access*, 9, 114381–114391. <https://doi.org/10.1109/ACCESS.2021.3104357>
- de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, 1376. <https://doi.org/10.1038/srep01376>
- Dwyer, J. F. (1988). Predicting daily use of urban forest recreation sites. *Landscape and Urban Planning*, 15, 127–138. [https://doi.org/https://doi.org/10.1016/0169-2046\(88\)90021-7](https://doi.org/https://doi.org/10.1016/0169-2046(88)90021-7)
- EAWS. (2023). Avalanche Danger Scale [Accessed: 2023-07-07]. European Avalanche Warning Services EAWS. <https://www.avalanches.org/standards/avalanche-danger-scale/>
- Ebert, P. A., & Milne, P. (2022). Methodological and conceptual challenges in rare and severe event forecast verification. *Natural Hazards and Earth System Sciences*, 22(2), 539–557. <https://doi.org/10.5194/nhess-22-539-2022>
- Epstein, D. A., Jacobson, B. H., Bales, E., McDonald, D. W., & Munson, S. A. (2015). From “nobody cares” to “way to go!”: A design framework for social sharing in personal informatics. *CSCW 2015 - Proceedings of the 18th ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 1622–1636. <https://doi.org/10.1145/2675133.2675135>
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72, 137–148. <https://doi.org/10.1007/s10708-008-9188-y>
- Fotouhi, S., Asadi, S., & Kattan, M. W. (2019). A comprehensive data level analysis for cancer diagnosis on imbalanced data. *Journal of biomedical informatics*, 90, 103089. <https://doi.org/https://doi.org/10.1016/j.jbi.2018.12.003>
- Fredston, J., Fesler, D., & Tremper, B. (1994). The Human Factor - Lessons for Avalanche Education. *Proceedings of the 1994 International Snow Science Workshop*, Snowbird, UT, USA, 473–487.
- Frei, C. (2014). Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. *International Journal of Climatology*, 34(5), 1585–1605. <https://doi.org/10.1002/joc.3786>
- Frei, C., Willi, M., Stöckli, R., & Dür, B. (2015). Spatial analysis of sunshine duration in complex terrain by non-contemporaneous combination of station and satellite data. *International Journal of Climatology*, 35(15), 4771–4790. <https://doi.org/10.1002/joc.4322>
- Furman, N., Shooter, W., & Schumann, S. (2010). The roles of heuristics, avalanche forecast, and risk propensity in the decision making of backcountry skiers. *Leisure Sciences*, 32(5), 453–469. <https://doi.org/10.1080/01490400.2010.510967>
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Grímsdóttir, H., & Mcclung, D. (2006). Avalanche risk during backcountry skiing - An analysis of risk factors. *Natural Hazards*, 39(1), 127–153. <https://doi.org/10.1007/s11069-005-5227-x>

- Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25.
- Hafner, E. D., Techel, F., Leinss, S., & Bühler, Y. (2021). Mapping avalanches with satellites – evaluation of performance and completeness. *The Cryosphere*, 15(2), 983–1004. <https://doi.org/10.5194/tc-15-983-2021>
- Hägeli, P., Haider, W., Longland, M., & Beardmore, B. (2009). Amateur decision-making in avalanche terrain with and without a decision aid: A stated choice survey. *Natural Hazards*, 52, 185–209. <https://doi.org/10.1007/s11069-009-9365-4>
- Hanssen, A. W., & Kuipers, W. J. A. (1965). On the relationship between the frequency of rain and various meteorological parameters (Vol. 81). Koninklijk Nederlands Meteorologisch Instituut Mededelingen. <https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubmetnummer/knmipub102-81.pdf>
- Harvey, S., Rhyner, H., & Schweizer, J. (2012). *Lawinenkunde. Praxiswissen für Einsteiger und Profis zu Gefahren, Risiken und Strategien*. Outdoor-Praxis. Bruckmann Verlag GmbH, München.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
- Haugom, E., & Malasevska, I. (2019). The relative importance of ski resort- and weather-related characteristics when going alpine skiing. *Cogent Social Sciences*, 5(1), 1681246. <https://doi.org/10.1080/23311886.2019.1681246>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Heggli, M. (2022, November 28). Subdivision of danger levels in the avalanche bulletin [Accessed: 2023-07-07]. <https://www.slf.ch/en/news/2022/11/subdivision-of-danger-levels-in-the-avalanche-bulletin.html%7B%5C#%7Dtabellement1-tab2>
- Heierli, J., Purves, R. S., Felber, A., & Kowalski, J. (2004). Verification of nearest-neighbours interpretations in avalanche forecasting. *Annals of Glaciology*, 38, 84–88. <https://doi.org/10.3189/172756404781815095>
- Hendriks, J., Johnson, J., & Mannberg, A. (2018). How do we really use terrain in the backcountry? a comparison between stated terrain preferences and observed backcountry travel behaviour. *Proceedings of the International Snow Science Workshop*, Innsbruck, Austria, 7–12.
- Hendriks, J., Johnson, J., & Mannberg, A. (2022). Tracking decision-making of backcountry users using GPS tracks and participant surveys. *Applied Geography*, 144, 102729. <https://doi.org/10.1016/j.apgeog.2022.102729>
- Hijmans, R. J. (2023a). Raster: Geographic data analysis and modeling [R package version 3.6-14]. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J. (2023b). Terra: Spatial data analysis [R package version 1.7-3]. <https://CRAN.R-project.org/package=terra>

- Huang, B. F., & Boutros, P. C. (2016). The parameter sensitivity of random forests. *BMC Bioinformatics*, 17(1), 1–13. <https://doi.org/10.1186/s12859-016-1228-x>
- Janitza, S., & Hornung, R. (2018). On the overestimation of random forest’s out-of-bag error. *PloS ONE*, 13(8), e0201904. <https://doi.org/10.1371/journal.pone.0201904>
- Jiang, W., Hong, Y., Zhou, B., He, X., & Cheng, C. (2019). A GAN-Based Anomaly Detection Approach for Imbalanced Industrial Time Series. *IEEE Access*, 7, 143608–143619. <https://doi.org/10.1109/ACCESS.2019.2944689>
- Johnson, J., & Hendrikx, J. (2021). Using citizen science to document terrain use and decision-making of backcountry users. *Citizen Science: Theory and Practice*, 6, 1–15. <https://doi.org/10.5334/CSTP.333>
- Kachore, V. A., Lakshmi, J., & Nandy, S. (2015). Location obfuscation for location data privacy. 2015 IEEE World Congress on Services, 213–220. <https://doi.org/10.1109/SERVICES.2015.39>
- King, M. A., Abrahams, A. S., & Ragsdale, C. T. (2014). Ensemble methods for advanced skier days prediction. *Expert Systems with Applications*, 41(4), 1176–1188. <https://doi.org/10.1016/j.eswa.2013.08.002>
- Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology*, 26(9), 1011–1013. <https://doi.org/10.1038/nbt0908-1011>
- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- Krawczyk, B., Woźniak, M., & Schaefer, G. (2014). Cost-sensitive decision tree ensembles for effective imbalanced classification. *Applied Soft Computing*, 14, 554–562. <https://doi.org/10.1016/j.asoc.2013.08.014>
- Kuhn, M. (2022). caret: Classification and Regression Training [R package version 6.0-93]. <https://cran.r-project.org/package=caret>
- Lamprecht, M., Fischer, A., & Stamm, H. (2014). Sport Schweiz 2014: Sportaktivität und Sportinteresse der Schweizer Bevölkerung (tech. rep.). Bundesamt für Sport BASPO, Magglingen.
- Lee, J., Hong, B., Lee, K., & Jang, Y.-J. (2015). A prediction model of traffic congestion using weather data. 2015 IEEE International Conference on Data Science and Data Intensive Systems, 81–88. <https://doi.org/10.1109/DSDIS.2015.96>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. <https://cran.r-project.org/doc/Rnews/>
- Lingras, P., Sharma, S., & Zhong, M. (2002). Prediction of recreational travel using genetically designed regression and time-delay neural network models. *Transportation Research Record*, 1805(1), 16–24. <https://doi.org/10.3141/1805-03>
- Liu, Y., & Wu, H. (2018). Prediction of road traffic congestion based on random forest. *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, IEEE, 2017, 2, 361–364. <https://doi.org/10.1109/ISCID.2017.216>

- Mach, L., Ponting, J., Brown, J., & Savage, J. (2020). Riding waves of intra-seasonal demand in surf tourism: analysing the nexus of seasonality and 21st century surf forecasting technology. *Annals of Leisure Research*, 23(2), 184–202. <https://doi.org/10.1080/11745398.2018.1491801>
- Maguire, L. (2014). The human behind the factor: a brief look at how context informs practice in recreational backcountry users. *Proceedings of the International Snow Science Workshop*, Banff, AB; Canada, 29, 942–948.
- Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*, 7, 93010–93022.
- Mannberg, A., Hendriks, J., Landrø, M., & Ahrland Stefan, M. (2018). Who’s at risk in the backcountry? Effects of individual characteristics on hypothetical terrain choices. *Journal of Environmental Psychology*, 59, 46–53. <https://doi.org/10.1016/j.jenvp.2018.08.004>
- Marsland, S. (2015). *Machine learning: An algorithmic perspective* (2nd ed.). CRC press.
- McCammon, I. (2004). Heuristic Traps in Recreational Avalanche Accidents: Evidence and Implications. *Avalanche News*, 68, 1–10.
- McCammon, I. (2009). Human factors in avalanche accidents: Evolution and interventions. *International Snow Science Workshop*, Davos, Switzerland, 27, 644–648.
- McCammon, I., & Hägeli, P. (2007). An evaluation of rule-based decision tools for travel in avalanche terrain. *Cold Regions Science and Technology*, 47, 193–206. <https://doi.org/10.1016/j.coldregions.2006.08.007>
- McClung, D., & Schaerer, P. (2006). *The avalanche handbook* (3rd ed.). The Mountaineers Books.
- MeteoSchweiz. (2015). *Klimabulletin Dezember 2014*. Bundesamt für Meteorologie und Klimatologie MeteoSchweiz.
- MeteoSchweiz. (2017a). *Klimabulletin Dezember 2016*. Bundesamt für Meteorologie und Klimatologie MeteoSchweiz.
- MeteoSchweiz. (2017b). *Klimabulletin Winter 2016/2017*. Bundesamt für Meteorologie und Klimatologie MeteoSchweiz.
- MeteoSwiss. (2021a). *Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD 1.2* (tech. rep.). Federal Office of Meteorology and Climatology MeteoSwiss.
- MeteoSwiss. (2021b). *Daily Precipitation (final analysis): RhiresD* (tech. rep.). Federal Office of Meteorology and Climatology MeteoSwiss.
- MeteoSwiss. (2021c). *Daily Relative Sunshine Duration: SrelD 1.0* (tech. rep.). Federal Office of Meteorology and Climatology MeteoSwiss.
- MeteoSwiss. (2021d). *Hourly Precipitation Estimation through Rain-Gauge and Radar: CombiPrecip 3.1* (tech. rep.). Federal Office of Meteorology and Climatology MeteoSwiss.
- MeteoSwiss. (2021e). *MeteoSwiss Spatial Climate Analyses: Documentation of Datasets for Users* (tech. rep.). Federal Office of Meteorology and Climatology MeteoSwiss.

- Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.
- Nielsen, J. (2006, October 8). The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities [Accessed: 2023-07-07]. <https://www.nngroup.com/articles/participation-inequality/>
- Nogueira Mendes, R., & Pereira da Silva, C. (2018). Looking back at recreational activities in protected areas using VGI from web-share services. *The 9th International Conference on Monitoring and Management of Visitors in Recreational and Protected Areas (MMV)*, 95, 140–142.
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Peirce, C. (1884). The numerical measure of the success of predictions. *Science*, 4, 453–454. <https://doi.org/10.1126/science.ns-%204.93.453-a>
- Pfeifer, C. (2009). On probabilities of avalanches triggered by alpine skiers. An empirically driven decision strategy for backcountry skiers based on these probabilities. *Natural Hazards*, 48(3), 425–438. <https://doi.org/10.1007/s11069-008-9270-2>
- Pierce, D. (2023). ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files [R package version 1.21]. <https://CRAN.R-project.org/package=ncdf4>
- Rabosky, D., Grundler, M., Anderson, C., Title, P., Shi, J., Brown, J., Huang, H., & Larson, J. (2014). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, 5, 701–707.
- Riddington, G. L. (2002). Learning and ability to pay: Developing a model to forecast ski tourism. *Journal of Travel & Tourism Marketing*, 13(1-2), 109–124.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804–818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- Rodwell, M. J., Magnusson, L., Bauer, P., Bechtold, P., Bonavita, M., Cardinali, C., Diamantakis, M., Earnshaw, P., Garcia-Mendez, A., Isaksen, L., et al. (2013). Characteristics of occasional poor medium-range weather forecasts for europe. *Bulletin of the American Meteorological Society*, 94(9), 1393–1405.
- Rutty, M., & Andrey, J. (2014). Weather forecast use for winter recreation. *Weather, Climate, and Society*, 6(3), 293–306. <https://doi.org/10.1175/WCAS-D-13-00052.1>
- Santos, T., Nogueira Mendes, R. M., Farías-Torbidoni, E., Juliao, R., & Silva, C. (2022). Volunteered geographical information and recreational uses within metropolitan and rural contexts. *ISPRS International Journal of Geo-Information*, 11(2), 144. <https://doi.org/10.3390/ijgi11020144>
- Scherrer, S. C., & Appenzeller, C. (2014). Fog and low stratus over the Swiss Plateau – a climatological study. *International Journal of Climatology*, 34(3), 678–686. <https://doi.org/10.1002/joc.3714>

- Schmudlach, G. (2022a). Avalanche Risk Property Dataset (ARPD) [data set, Accessed: 2022-09-27]. [https://info.skitouren guru.ch/download/data/ARPD\\_Manual\\_3.1.2.pdf](https://info.skitouren guru.ch/download/data/ARPD_Manual_3.1.2.pdf)
- Schmudlach, G. (2022b). Avalanche Risk Property Dataset (ARPD) User Manual (V3.1.2) [Accessed: 2023-07-11]. [https://info.skitouren guru.ch/download/data/ARPD\\_Manual\\_3.1.2.pdf](https://info.skitouren guru.ch/download/data/ARPD_Manual_3.1.2.pdf)
- Schmudlach, G., & Köhler, J. (2016). Automated Avalanche Risk Rating of Backcountry Ski Routes. Proceedings of the International Snow Science Workshop, Beckenridge, CO, USA, 1(2), 450–456.
- Schmudlach, G., Winkler, K., & Köhler, J. (2018). Quantitative risk reduction method (qrm), a data-driven avalanche risk estimator. Proceedings ISSW, 1272–1278.
- Schweizer, J., Mitterer, C., Techel, F., Stoffel, A., & Reuter, B. (2020). On the relation between avalanche occurrence and avalanche danger level. The Cryosphere, 14(2), 737–750. <https://doi.org/10.5194/tc-14-737-2020>
- Schweizer, J., & Techel, F. (2017). Lawinenunfälle Schweizer Alpen. Zahlen und Fakten der letzten 20 Jahre. Bergundsteigen, 98, 44–48.
- Schweizer, J., Jamieson, B., & Schneebeli, M. (2003). Snow avalanche formation. Reviews of Geophysics, 41(4). <https://doi.org/10.1029/2002RG000123>
- Senaratne, H., Mobasher, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science, 31(1), 139–167. <https://doi.org/10.1080/13658816.2016.1189556>
- SLF. (2023a). About the avalanche bulletin [Accessed: 2023-07-11]. WSL Institute for SnowAvalanche Research SLF. [www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin.html](http://www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin.html)
- SLF. (2023b). Avalanche Types [Accessed: 2023-07-11]. WSL Institute for SnowAvalanche Research SLF. <https://www.slf.ch/en/avalanches/avalanche-science-and-prevention/avalanche-types.html>
- Sprefaco, M., & Weingartner, R. (2005). The hydrology of switzerland: Selected aspects and results. Reports of the FOWG, Water Series, Berne.
- Statham, G. (2008). Avalanche hazard, danger and risk - a practical explanation. Proceedings of the 2008 International Snow Science Workshop, Whistler, BC, Canada, 224–227.
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics, 8. <https://doi.org/10.1186/1471-2105-8-25>
- Suter, C., Purves, R. S., & Harvey, S. (2007). Avalanche education with mlearning. GISRUUK 2007, 35.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. Science, 240(4857), 1285–1293.
- Swisstopo. (2005). DHM25 - Das digitale Höhenmodell der Schweiz (tech. rep.).
- Techel, F., Mayer, S., Pérez-Guillén, C., Schmudlach, G., & Winkler, K. (2022). On the correlation between a sub-level qualifier refining the danger level with observations and



- models relating to the contributing factors of avalanche danger. *Natural Hazards and Earth System Sciences*, 22(6), 1911–1930. <https://doi.org/10.5194/nhess-22-1911-2022>
- Techel, F., Zweifel, B., & Winkler, K. (2015). Analysis of avalanche risk factors in backcountry terrain based on usage frequency and accident data in Switzerland. *Natural Hazards and Earth System Sciences*, 15(9), 1985–1997. <https://doi.org/10.5194/nhess-15-1985-2015>
- Upton, V., Ryan, M., O'Donoghue, C., & Dhubhain, A. N. (2015). Combining conventional and volunteered geographic information to identify and model forest recreational resources. *Applied Geography*, 60, 69–76. <https://doi.org/10.1016/j.apgeog.2015.03.007>
- Verbos, R. I., Altschuler, B., & Brownlee, M. T. (2018). Weather Studies in Outdoor Recreation and Nature-Based Tourism: A Research Synthesis and Gap Analysis. *Leisure Sciences*, 40(6), 533–556. <https://doi.org/10.1080/01490400.2017.1325794>
- Wegelin, P., von Arx, W., & Thao, V. T. (2022). Weather myths: how attractive is good weather really for same-day visits to outdoor recreation destinations? *Tourism Recreation Research*, 1–13. <https://doi.org/10.1080/02508281.2022.2148076>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation* [R package version 1.1.0]. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data* [R package version 1.3.0]. <https://CRAN.R-project.org/package=tidyr>
- Williams, K. (1998). An overview of avalanche forecasting in north america. *Proceedings of the International Snow Science Workshop*, Sunriver, OR, USA, 161–169.
- Winkler, K., Fischer, A., & Techel, F. (2016). Avalanche risk in winter backcountry touring: status and recent trends in Switzerland. *International Snow Science Workshop*, Breckenridge, CO, USA, 270–276.
- Winkler, K., Schmudlach, G., Degraeuwe, B., & Techel, F. (2021). On the correlation between the forecast avalanche danger and avalanche risk taken by backcountry skiers in Switzerland. *Cold Regions Science and Technology*, 188, 103299. <https://doi.org/10.1016/j.coldregions.2021.103299>
- Woźniak, P. W., Fedosov, A., Mencarini, E., & Knaving, K. (2017). Soil, rock and snow: On designing for information sharing in outdoor sports. *Proceedings of the 2017 ACM Conference on Designing Interactive Systems*, 611–623. <https://doi.org/10.1145/3064663.3064741>
- Zambrano-Bigiarini, M. (2020). *hydroTSM: Time Series Management, Analysis and Interpolation for Hydrological Modelling* [R package version 0.6-0]. <https://github.com/hzambbran/hydroTSM>
- Zhang, G., & Zhu, A. X. (2018). The representativeness and spatial bias of volunteered geographic information: a review. *Annals of GIS*, 24(3), 151–162. <https://doi.org/10.1080/19475683.2018.1501607>

- Zhou, Z.-H., & Liu, X.-Y. (2010). On multi-class cost-sensitive learning. *Computational Intelligence*, 26(3), 232–257.
- Zweifel, B., Rätz, A., & Stucki, T. (2006). Avalanche risk for recreationists in backcountry and in off-piste area: surveying methods and pilot study at Davos, Switzerland. *Proceedings International Snow Science Workshop, Telluride, CO*, 733–741.





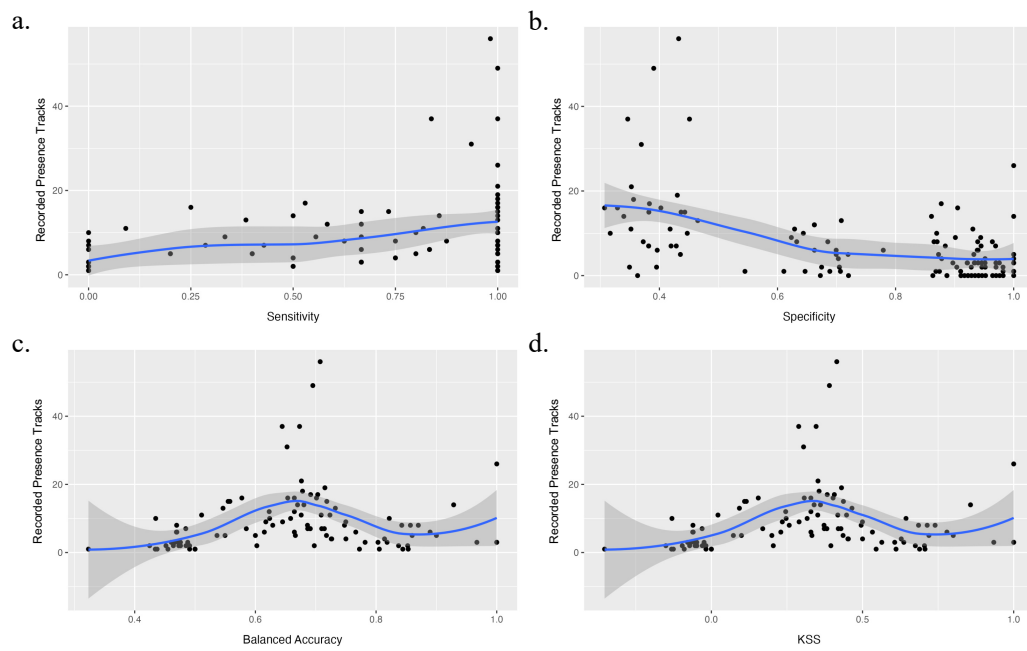
# Warning Regions



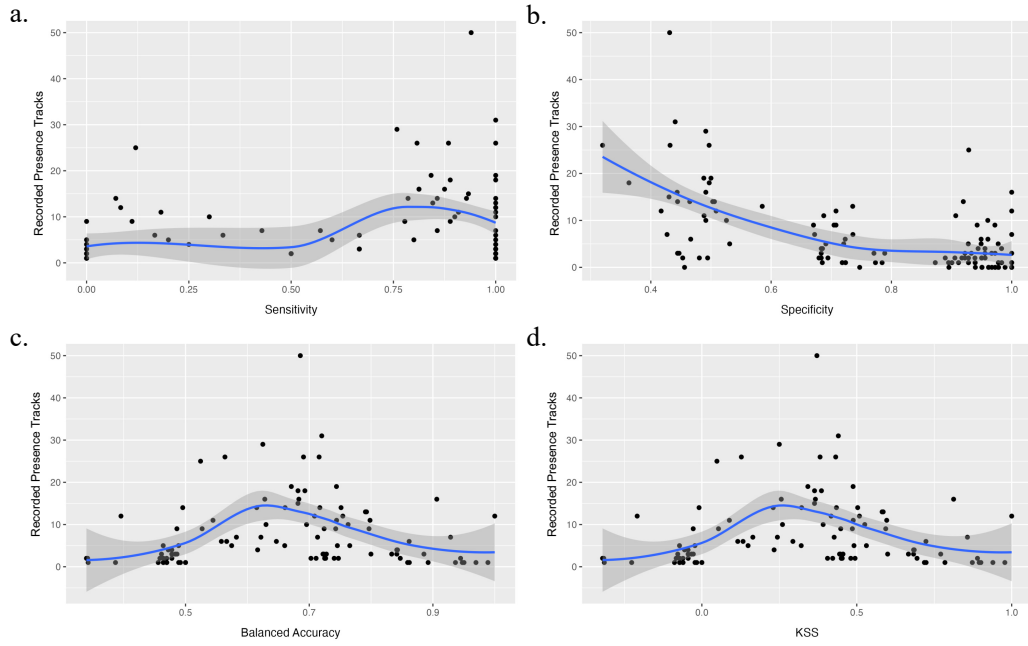
**Figure A.1:** Warning Regions used by the WSL Institute for Snow and Avalanche Research SLF.

## Appendix B

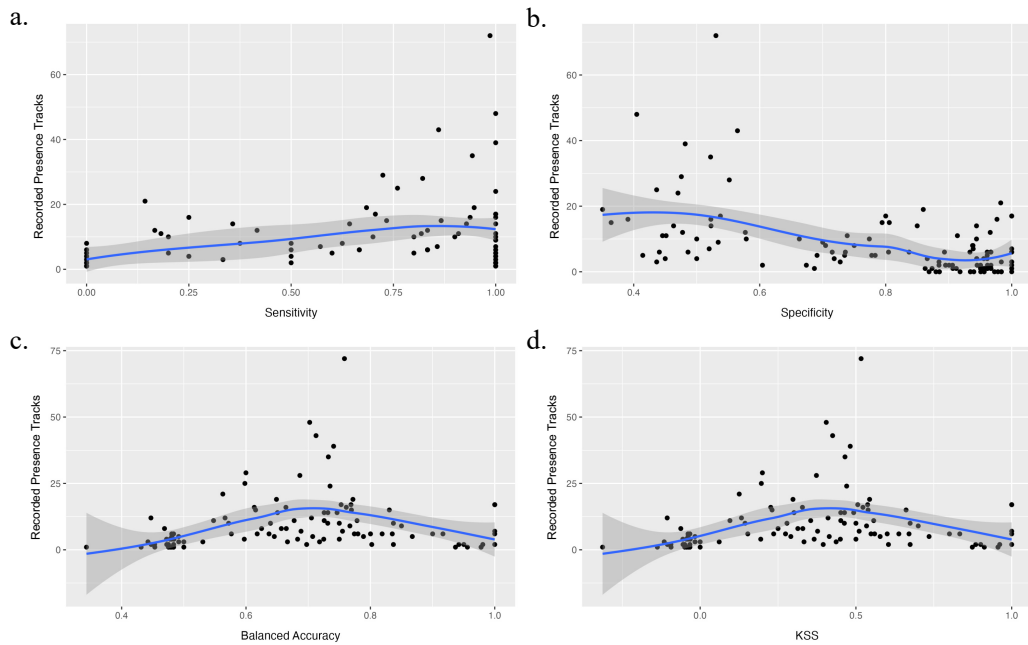
### Skill Scores



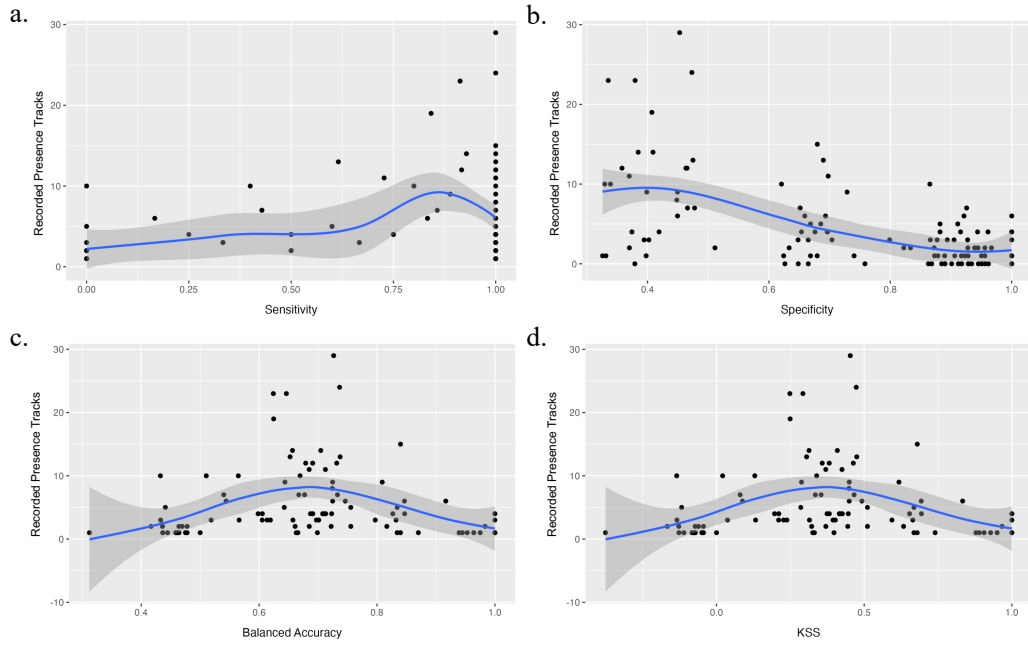
**Figure B.1:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2013/14.



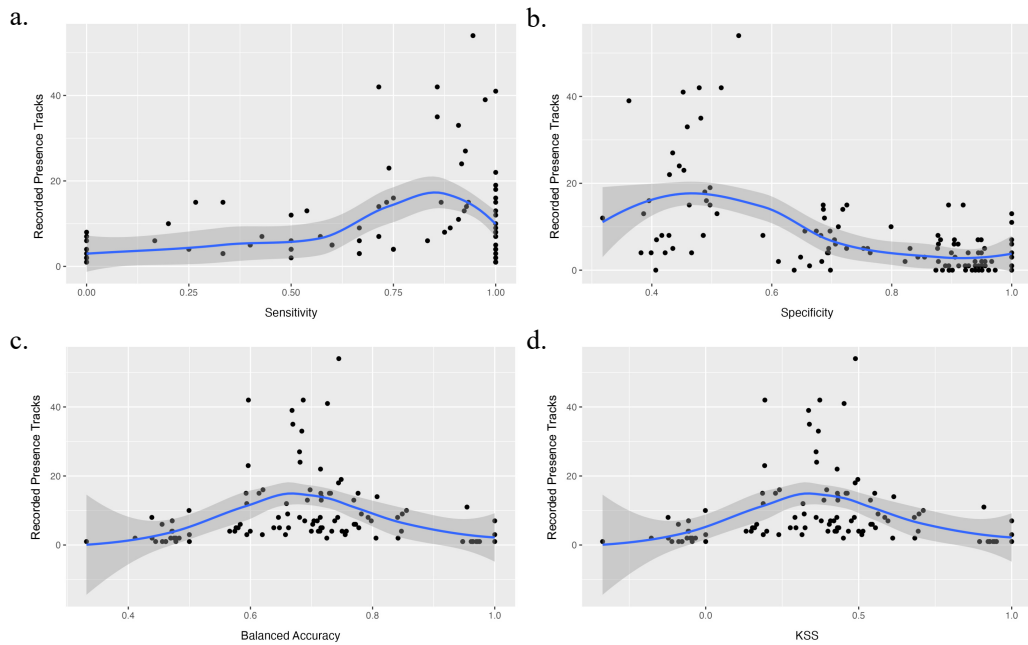
**Figure B.2:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2014/15.



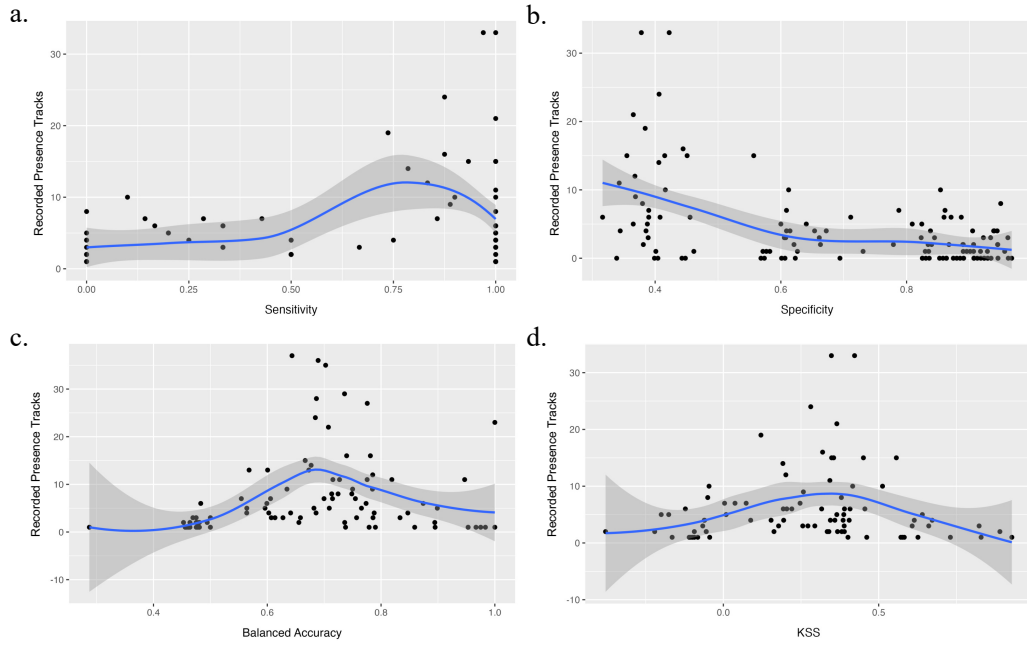
**Figure B.3:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2015/16.



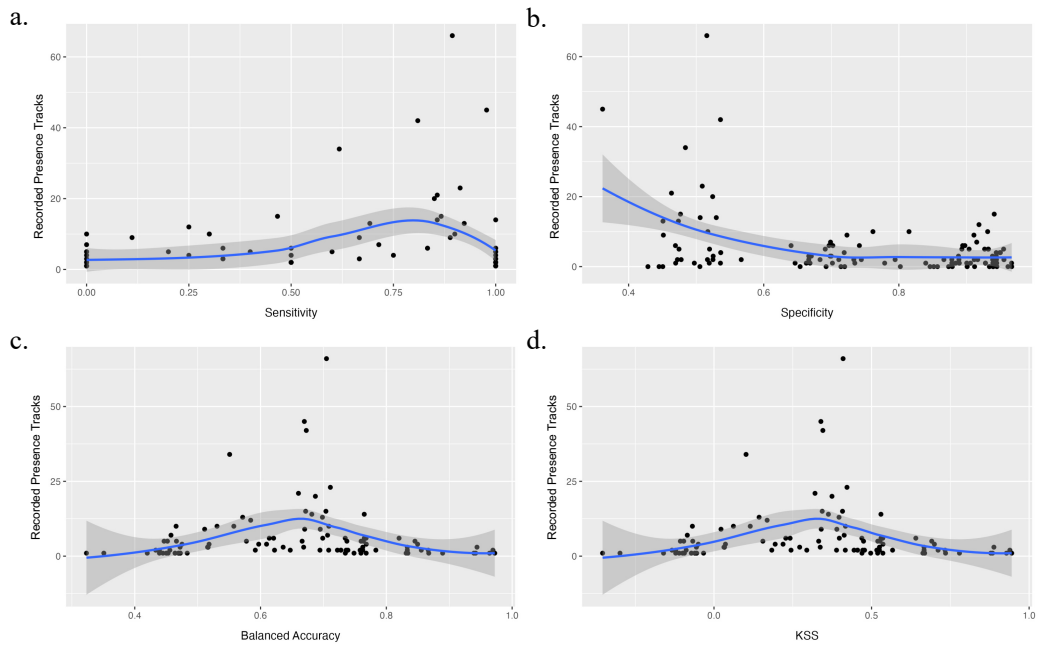
**Figure B.4:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2016/17.



**Figure B.5:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2018/19.



**Figure B.6:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2019/20.



**Figure B.7:** Relationship between a. sensitivity, b. specificity, c. balanced accuracy and d. KSS and total track count per warning region, season 2020/21.

## **Personal declaration**

I hereby declare, that the material contained in this thesis is my own original work. Any quotation or paraphrase in this thesis from the published or unpublished work of another individual or institution has been duly acknowledged. I have not submitted this thesis, or any part of it, previously to any institution for assessment purposes.

A handwritten signature in black ink, appearing to read 'Leonie Schäfer'.

Leonie Schäfer  
Zurich, 14.07.2023