University of Zurich[UZH]

Department of Geography

GEO 620 - Master's Thesis

Analysis of Planned Route Trajectories to Gain Insights into Route Planning Behaviour for Backcountry Ski Tours

Department of Geography
Geocomputation
University of Zurich

Date of Submission: 31.07.2018

Author
Christoph Schönenberger
12-707-196

Supervisors
Prof. Dr. Ross Purves
Stephan Harvey
Frank Techel

Faculty Representative
Prof. Dr. Ross Purves

Contact

Author

Christoph Schönenberger
Brunnenweid 61
CH-5643 Sins
ch.schoenenberger@outlook.com

Supervisors

Prof. Dr. Ross Purves
University of Zurich
Department of Geography - Geocomputation
Winterthurerstrasse 190
CH-8057 Zurich
ross.purves@geo.uzh.ch

Stephan Harvey
WSL-Institut für Schnee- und Lawinenforschung SLF
Flüelastr. 11
CH-7260 Davos Dorf
harvey@slf.ch

Frank Techel
WSL-Institut für Schnee- und Lawinenforschung SLF
Flüelastr. 11
CH-7260 Davos Dorf
techel@slf.ch

Zusammenfassung

In den letzten Jahren hat eine Verschiebung des Forschungsschwerpunkts im Lawinenbereich von geophysikalischen Faktoren zum Faktor Mensch stattgefunden. Verschiedene Forscher nutzten Bewegungsdaten, um Einblicke in die Entscheidungsfindung und das menschliche Verhalten in Lawinengebieten zu gewinnen. In dieser Arbeit wurde eine neue Art von Bewegungsdaten analysiert, welche bisher vernachlässigt wurde: Geplante Bewegungsdaten. Ziel dieser Arbeit war es, durch die Analyse geplanter Routenverläufe neue Erkenntnisse über den Planungsprozess von Skitouren zu gewinnen.
Die in dieser Arbeit analysierten Daten stammen von White Risk, einer webbasierten Lawinenpräventionsplattform. Neben E-Learning-Funktionalitäten beinhaltet White Risk ein kartenbasiertes Tool zur Planung von Skitouren. Die geplanten Touren können in einer Smartphone-Anwendung heruntergeladen und bei Skitouren als Orientierungshilfe verwendet werden. Da die geplanten Routen auf einer Karte gezeichnet werden, fehlt ihnen die zeitliche Komponente, wie sie aus herkömmlichen Bewegungsdaten bekannt ist. Diese fehlende Komponente schränkt die anwendbaren Methoden aus der computergestützten Bewegungsanalyse (eng. Computational Movement Analysis (CMA)), die normalerweise auf Bewegungsdaten angewendet werden, stark ein.
Vor der Hauptanalyse durchliefen die Daten eine gründliche Aufbereitung, bei der die Daten gefiltert und mit Sekundärdaten angereichert wurden. Die Filterung diente der Beseitigung von Testrouten, die bei weiteren Analysen nicht berücksichtigt werden sollten. Die Analysen umfassten einen allgemeinen Überblick über die Daten in Bezug auf zeitliche und räumliche Verteilung, Partizipationsungleichheit und Lawinenverhältnisse. Es wurden zudem Routenattribute und verschiedene Einflüsse darauf untersucht. Schliesslich wurde die Ähnlichkeit zwischen verschiedenen Routen sowie zwischen Routen und GPS-Tracks anhand der Fréchet-Distanz berechnet. Dies ermöglichte die Anwendung von DBSCAN-Clustering auf den geplanten Routen.
Die zeitliche und räumliche Verteilung der geplanten Routen stimmte mit Ergebnissen bestehender Arbeiten mit ähnlichen Daten überein. Die Ergebnisse zeigten, dass für Tage mit erheblicher Lawinengefahr mehr Routen als erwartet geplant wurden. Weiter wurde gezeigt, dass die Routenattribute durch die Auflösung von Sekundärdaten (Digitale Höhenmodelle), Partizipationsungleichheit und Lawinengefahr beeinflusst werden. Es wurde festgestellt, dass regelmässige Nutzer längere und detailliertere Routen in höheren Lagen planen. Sowohl die Ähnlichkeit als auch das Clustering wurden stark durch das Gelände, in welchem die Routen geplant wurden, beeinflusst. Das Clustering wurde zusätzlich durch Parametereinstellungen beeinflusst. Dennoch konnten durch das Clustering der geplanten Skitouren Hauptrouten und ein grober Routenkorridor identifiziert werden.

Abstract

In recent years, the focus of avalanche research has shifted towards the human factor in avalanche processes. Various researchers have exploited movement data to gain insights into decision making and human behaviour in avalanche terrain. In this thesis, a new type of movement data was analysed, which has been neglected so far: Planned movement data. The aim of this thesis was to gain new insights into the planning process of backcountry tours through the analysis of planned route trajectories.
The data analysed in this thesis originated from White Risk, a web-based avalanche prevention platform. Besides e-learning capabilities, White Risk includes a map-based tool to plan backcountry tours, which can be downloaded to a mobile phone application and used as a guide during backcountry tours. Since the planned routes are drawn on a map, they lack a temporal component as it is known from conventional movement data. This severely restricts applicable Computational Movement Analysis (CMA) approaches, which are usually applied to movement data.
Before the analysis, the data underwent thorough preprocessing, during which the data was filtered and enriched with secondary data. The filtering served the removal of test routes, which should not be considered for further analyses. The analyses encompassed a general overview of the data in terms of temporal and spatial distribution, participation inequality and avalanche conditions. Route attributes and various influences thereon were investigated. Finally, the similarity in-between routes and between routes and GPS-tracks were calculated using the Fréchet distance. This enabled the application of DBSCAN clustering on the planned routes.
The temporal and spatial distribution of planned routes was consistent with findings of existing works on reported backcountry activity. The results revealed that more routes than expected were planned for considerable avalanche danger. Furthermore, it was shown that the route attributes are influenced by the resolution of secondary data (DEMs), participation inequality and avalanche danger. It has been found that heavy users plan longer, more detailed routes at higher elevations. Both similarity and clustering were strongly influenced by the terrain in which the routes were planned. Clustering was further affected by parameter settings. Nonetheless, it was possible to identify major backcountry routes and a rough route corridor by clustering planned backcountry routes.

Acknowledgements

The end of this thesis also marks the end of six years at the GIUZ. It was a very eventful time with much learned and many great people which I got to know. This thesis would not have been possible without some people I would like to thank at this point.

First and foremost, I would like to thank my three supervisors, Ross, Stephan and Frank. I appreciated it very much to have three such great experts at my side. The fruitful discussions with you and your valuable inputs have helped me to make the most of this thesis.

Furthermore, I'd like to thank my parents. Thank you for your continuous support during all of my studies. I cannot express how grateful I am to you. Such a successful study experience would not have been possible without you.

Katharina, I don't know if I would have made it through this thesis without you. They say a sorrow shared is a sorrow halved. However, with you it was rather quartered. Thank you for taking my mind off my thesis when I needed it. Thank you for not talking about the thesis when I was fed up with it. Thank you for keeping my mood up during this time. Off to new adventures!

Thank you Thomas for sharing your experiences with me and confirming to me that I am on the right track. You were always there for me when I had questions or when I was in doubt again. Thanks also for proofreading the thesis.

Thanks also to Oliver, who also proofread the thesis.

Many other people have accompanied me through my studies and enriched them. It was great to experience this time with you and I hope to stay in contact with you for many years to come.

Contents

List of Figures

List of Tables

List of Abbreviations

5NG	5 Nearest GPS-Tracks
5NN	5 Nearest Neighbours
BEV	Bundesamt für Eich- und Vermessungswesen
CDTW	Context-based Dynamic Time Warping
CGI	Contributed Geographic Information
CMA	Computational Movement Analysis
DEM	Digital Elevation Model
DTW	Dynamic Time Warping
EDM	Edit Distance on Movement
EDR	Edit Distance on Real sequence
ERP	Edit distance with Real Penalty
GIScience	Geographic Information Science
HMM	Hidden Markov Model
IGN	Institut national de l'information géographique et forestière
KDD	Knowledge Discovery in Databases
kNG	k Nearest GPS-Tracks
kNN	k Nearest Neighbours
LCSS	Longest Common Subsequence
LIP	Locality In between Polylines
MBR	Minimum Bounding Rectangle
NWED	Normalized Weighted Edit Distance
OSM	Open Street Map
OWD	One-Way Distance

POI	Points of Interest
SAC	Swiss Alpine Club
SLF	WSL Institute for Snow and Avalanche Research SLF
swisstopo	Federal Office of Topography
UCC	User-Created Content
UGC	User-Generated Content
VGI	Volunteered Geographic Information

Chapter 1

Introduction

Avalanches are one of the most significant natural hazards in alpine areas. Disruptions of important transport corridors and destruction of infrastructure can cause damage in the millions. Furthermore, the lives of many recreationists are threatened by avalanches. In Switzerland, an average of 25 people die in avalanche accidents every year (SLF, 2018c). Techel and Zweifel (2013) noted that over 90 % of these fatal accidents occur during recreational activities such as freeriding or ski touring. In 95 % of all avalanche accidents, the avalanche was human-triggered (Harvey et al., 2002, p. 449). These facts illustrate the need for a thorough understanding of avalanche processes, as well as human behaviour in avalanche-prone terrain to support the prevention of avalanche accidents.
For a long time, researchers have mostly focused on geophysical aspects of avalanche processes, such as weather, terrain or snow cover (McClung, 2002, 111). However, in recent years the focus has increasingly shifted towards the investigation of the human factor in avalanche accidents. In contrast to geophysical aspects, the human factor is not clearly defined. It encompasses various aspects which influence decision-making processes, e.g. risk propensity, attitudes, etc. With the rise of positioning technologies, movement data is made increasingly available for various analyses, also in the field of avalanche research. Hendrikx et al. (2013) and Haegeli and Atkins (2016) pioneered the analysis of movement data to gain insights into human behaviour in avalanche terrain.
In 2013, the WSL Institute for Snow and Avalanche Research SLF (SLF) released White Risk 2.0, a web-based avalanche education platform (Harvey et al., 2013). This platform includes a tool to plan backcountry ski tours with a map interface. To date, the data from this trip-planning tool has not been investigated in depth. Therefore, this thesis seeks to explore the planned route trajectories gathered by White Risk to obtain a better understanding of the planning process of backcountry tours.

1.1 Problem Statement

To date, little research has utilized movement data for avalanche research. Existing literature has mainly focused on decision-making during backcountry skiing. No attempts were made to gain deeper insights into the planning phase of backcountry skiing, even though it can potentially improve avalanche prevention. A reason for this might be the lack of appropriate data. The analysis of planned route data

from White Risk has the potential to enahnce understanding of the backcountry trip planning processes and, thus, broaden general knowledge about the human factor in avalanche accidents and improve avalanche accident prevention.
With its structural resemblance to 'real' movement data, planned movement data should be processable using methods from Computational Movement Analysis (CMA). However, there is a lack of research on the processing of planned movement data. Even though planned movement data is closely related to 'real' movement data that is commonly analysed, there are differences in the data structure. These differences impact the applicable methods and require detailed attention. Hence, there is a need for evaluation of applicable CMA methods for planned route trajectories.

1.2 Aim and Scope

The overall aim of this thesis is to analyse planned movement trajectories gathered with White Risk and therewith gain better insights into the planning processes of backcountry tours. To achieve this aim, several processing steps will be taken. Therefore, this thesis is subdivided into multiple parts. First, differences between planned movement data and 'real' GPS-tracks need to be assessed. Based on these, appropriate analysis methods can be identified. Secondly, the data needs to be preprocessed in order to apply the identified methods. Furthermore, the preprocessing serves to reduce the amount of data to be processed and to remove undesirable data, such as test data by users. Third, the accuracy of the preprocessed tours needs to be assessed and results generated. Finally, these results can be put into context and conclusions about the planning processes of backcountry ski tours can be drawn.
In this thesis, only data from the Swiss Alps will be considered. Even though White Risk allows for the planning of trips worldwide, data outside of Switzerland and data in non-mountainous areas within Switzerland will be neglected. Route planning data from White Risk outside of Switzerland is rather sparse and data from non-mountainous areas is considered irrelevant to this research.
Furthermore, general decision-making theory and decision-making during backcountry tours will be omitted. It is unknown whether users who planned a route executed it after all. Thus, no statements about the decision-making during tours can be made and only the planning processes of backcountry ski tours will be addressed. As a new dataset was used in this thesis, preprocessing the data occupied a large share of the work conducted and therefore constitutes a significant part of this thesis.
In this thesis, no new CMA approaches will be developed for planned movement data. Instead of a qualitative or quantitative analysis of approaches, the existing approaches that can potentially be applied to the data at hand to generate meaningful outcomes will be evaluated. Only some of these will eventually be used for the data analysis. Nonetheless, potential future research directions for the computational analysis of planned movement data will be suggested.

1.3 Significance

As mentioned in the previous section, the aim of the thesis is to extend current knowledge about the planning process of backcountry tours, to expand on the hu-

man factor in avalanche research. It is hoped that the successful achievement of this aim will illuminate a current dark spot in understanding of the human factor in avalanche processes. On one hand, this might help to improve prevention measures for avalanche accidents in the early stage of recreational activities. On the other hand, the research community might obtain a deeper understanding of how users plan backcountry tours.
Furthermore, the extensive data preprocessing should accelerate future research on data from White Risk. This should enable researchers to delve deeper into the human factor and planning processes of backcountry tours based on this data.
The general focus of this thesis was not the application of CMA methods on planned route trajectories, but the usage of such with this data. However, the assessment of applicable CMA methods for planned movement trajectories should build the groundwork for further research on this type of movement data.

1.4 Thesis Structure

The remainder of this thesis is structured in six chapters, as follows. Chapter 2 covers existing literature relevant to this work. The literature is divided into three sections: Computational Movement Analysis (CMA), Volunteered Geographic Information (VGI) and avalanche research. Research gaps will be identified based on this literature and research questions will be posed. Chapter 3 will first describe the primary and secondary data used in this work and how it was gathered or generated. Second, the study area for large-scale analyses and the three smaller study areas for computationally-intensive analyses will be outlined. Chapter 4 will include all methods that were applied to the data described in chapter 3. This includes the preprocessing of the data, as well as the influences of specific data characteristics (e.g. raster resolution and participation inequality) and CMA approaches. Chapter 5 will then outline the results obtained from the application of the methods in chapter 4. These results will be discussed in chapter 6. The discussion should help to answer the research questions posed in chapter 2. Furthermore, the issues and limitations met in this thesis will be addressed. Chapter 7 will provide a short summary of the thesis, outline its contributions and suggest further research directions.

Chapter 2

State Of Research

The aim of this chapter is to give an overview of the state of research in fields related to this work. A review of the current literature will anchor this thesis within a greater body of work and provide the necessary background information which is relevant to the analyses conducted within the scope of this thesis. Furthermore, a thorough analysis of existing literature enables the finding of research gaps, which are yet to be addressed by the research community.
In order to achieve this aim, literature from three research branches is synthesised into three sections. The chapter is structured as follows. First, a general overview of CMA will be presented (Section 2.1), which is then narrowed down to specific analysis steps, namely segmentation & filtering and similarity & clustering. As the data analysed in this work is a form of movement data, many of the applied methods will originate from CMA. Therefore, a thorough understanding of the available approaches is crucial and enables the selection of appropriate methods. Second, the VGI phenomenon will be discussed, including the issues and specific areas of application related to this work (section 2.2). The data from White Risk can be seen as a form of VGI. VGI come with several implications, which might affect subsequent analyses. Hence, these implications need to be considered and accounted for. Lastly, avalanche research will be addressed with a focus on the human factor (section 2.3). As stated in chapter 1, the general aim is to gain insights into the planning phase of backcountry tours. This planning phase can be considered part of the human factor in avalanche research. Thus, a general overview of avalanche research and specifically the human factor is necessary. The research gaps identified will be outlined in section 2.4 and research questions based on these will be presented in the subsequent section 2.5.

2.1 Computational Movement Analysis

Laube (2014) states that CMA draws concepts and methods from three research areas. These are Geographic Information Science (GIScience), computer science and statistics. He defines computational movement analysis as follows:

> *"Computational Movement Analysis is the interdisciplinary research field studying the development and application of computational techniques for capturing, processing, managing, structuring, and ultimately analysing data de-*

> *scribing movement phenomena, both in geographic and abstract spaces, aiming for a better understanding of the processes governing that movement."*
>
> (Laube, 2014, pp. 4–5)

The application of CMA is highly dependent on the data that is used. This data may differ in terms of objects that are tracked as well as technology that is used for their tracking. This influences the choice of appropriate abstraction, conceptual modelling, and formalisation which is needed in order to computationally analyse such data (Laube, 2014). Laube (2009) describes six conceptual movement spaces which can be used to model movement data. These movement spaces are depicted in figure 2.1. Objects that can relocate to an arbitrary position without any restrictions (e.g. birds) move in an Euclidean homogeneous space (a). If the movement is obstructed by environmental features or objects, we talk about constrained homogeneous space (b). For visualisation purposes, it might make sense to model time as an additional dimension. Such a three-dimensional space is either called space-time cube or space-time aquarium (c). Some data is conceptualised not as vector data, but as raster data. This results in a discrete heterogeneous field space (d). It is possible to divide space into irregular fields (e.g. mobile phone cells) (e). Lastly, some movements are restricted to a network (e.g. transportation). In such cases, we talk about network spaces (f). The conceptual model which underlies the data influences the movement and, therefore, also the tools and methods needed for further analyses. Thus, it is crucial to decide which conceptual model is to be used before conducting thorough movement analysis.

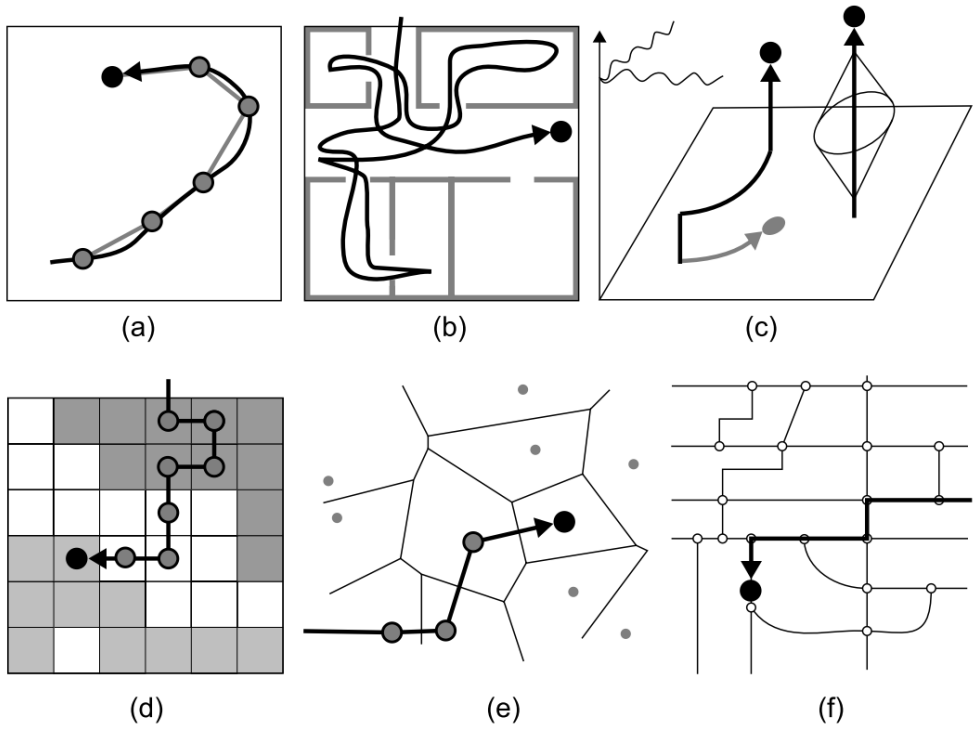


Figure 2.1: *Movement spaces according to Laube (2009). (a) Euclidean homogeneous space, (b) constrained homogeneous space, (c) space-time aquarium, (d) heterogeneous field space, (e) irregular tessellation, (f) network space.*

According to Laube (2014), there are three underlying dimensions which discriminate conceptual spaces. Movement can either be seen from a Lagrangian or Eulerian perspective. The Lagrangian perspective considers the movement of an object's location which results in a sequence of tuples, consisting of location and timestamp. This sequence can then be connected to a polyline. From the Eulerian perspective,

the movement is regarded relative to some fixed points in space (e.g. cell phone towers, RFID tags, etc.). This dimension discerns movement space (e) from all others. The second dimension are constraints on movement. An unconstrained movement allows for simpler conceptual models to be used. However, Laube (2014) notes that most human movement is constrained to some sort of pedestrian or transportation network. Clearly, constraints have an explicit influence on the movement behaviour of objects. In figure 1, spaces (b) and (f) exhibit constraints. The third dimension is the differentiation between continuous and discrete spaces. Spaces (d) and (f) are discrete, whereas the others are continuous. The consideration of all these dimensions before the analysis phase allows selecting the appropriate conceptual model and therefore enables a researcher to select the right analysis methods.

In her PhD thesis, Dodge (2011) introduced an extended conceptual movement framework which is based on the work of Nathan et al. (2008). Four main components make up this conceptual framework, namely, a) internal state, b) movement characteristics, c) movement path, and d) external factors (see figure 2.2). The movement characteristics encompass positional and temporal information on movement as well as movement parameters of the individual. Dodge et al. (2008) were not the first to come up with the notion of movement parameters, however, they introduced the term "movement parameters" which is widely used nowadays. In previous work, movement parameters were also called movement descriptors (Laube et al., 2007) or movement characteristics (Andrienko et al., 2008). The movement parameters are important, as various movement analysis approaches rely on them.

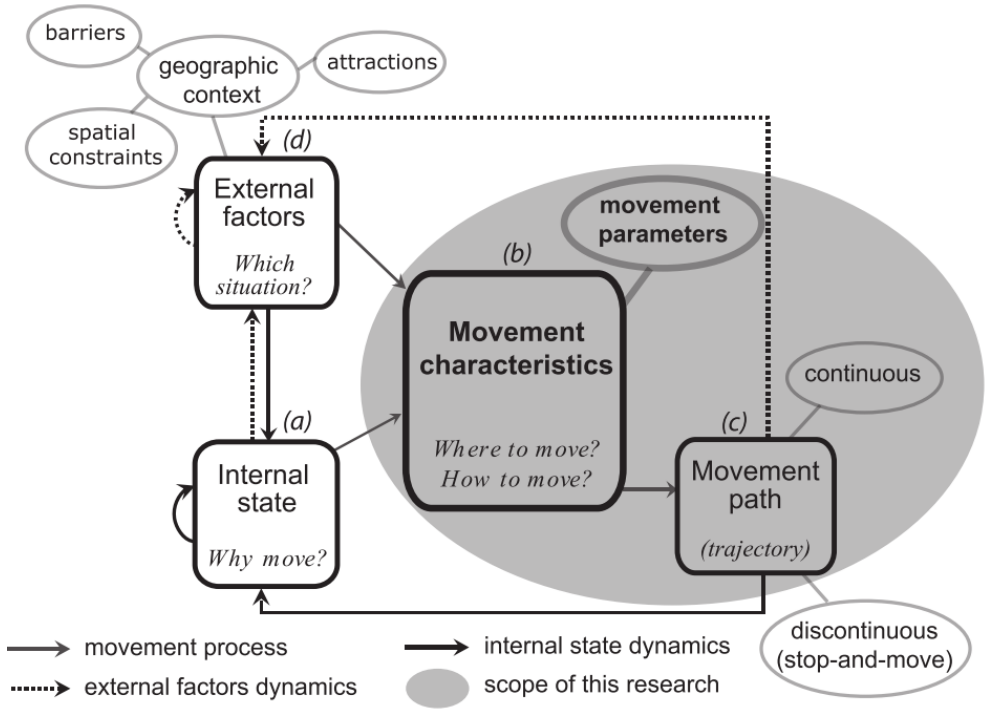


Figure 2.2: *Extended conceptual model of movement from Dodge (2011) based on Nathan et al. (2008).*

Dodge et al. (2008, p. 243) differentiated three different groups of movement parameters: 1) primitive parameters, 2) primary derivatives, and 3) secondary derivatives. These parameters may be further divided into subgroups based on their dimensions: spatial, temporal or spatio-temporal. The primitive parameters simply denote the position (x,y) of an object in space, the time instance at which the position is recorded or the time interval (sampling rate). Primary derivatives encompass dis-

tance, direction and extent in the spatial dimension, duration and travel time in the temporal dimension and speed and velocity in the spatio-temporal dimension. All of them are functions of the primitive parameters ($f(x,y), f(t), f(x,y,t)$). Secondary derivatives are functions of the primary derivatives and include spatial distribution, change of direction, sinuosity in the spatial dimension, temporal distribution and change of duration in the temporal dimension, and acceleration and approaching rate in the spatio-temporal dimension. It needs to be noted that, through advanced sensor technology, many of the derivatives can be gathered from sensors directly nowadays (Dodge et al., 2012, p. 1563).

The motivations for CMA are manifold and depend on the application area. Gudmundsson et al. (2011) present a variety of application areas in which CMA is applied regularly . GPS-tracking of animals allows researchers in behavioural ecology to observe behavioural patterns and gain a better understanding of such. Additionally, mobile phone networks or location-based services can be used to obtain an understanding of human mobility in an urban context as well as give insights into traffic patterns. Furthermore, CMA is used in surveillance and security, marketing and sport scene analysis. Finally, CMA can also be used to analyse human behaviour in avalanche-prone terrain as was done by Hendrikx et al. (2013). The use of volunteered geographic information and CMA in avalanche research will be discussed further in section 2.2.

The subsequent analysis of movement data is known as movement mining. Laube (2014) provides a coherent definition of movement mining:

> *"Movement mining aims for conceptualizing and detecting non-random properties and relationships in movement data that are valid, novel, useful, and ultimately understandable."*
>
> (Laube, 2014, p. 31)

Validity in this context means that the properties and relationships should be general enough to apply to new data. Novelty implies that they are unexpected and nontrivial. The properties and relationships should lead to successful decision-making or new scientific findings, meaning they are useful and, last but not least, they should be understandable and interpretable by human experts (Laube, 2014).
Movement mining is a particular form of data mining which Fayyad et al. (1996, p. 39) define as *"the application of specific algorithms for extracting patterns from data"*. Data mining refers to one specific step in the Knowledge Discovery in Databases (KDD) process. Additional parts of KDD include data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge and proper interpretation of the results (Fayyad et al., 1996, p. 39). The previously discussed selection of an appropriate conceptual model is therefore part of this spanning KDD process. Laube (2014) points out that movement mining, similarly to data mining, is often confronted with data retrieved from multiple, potentially unrelated sources and various data types. Furthermore, movement data is highly spatio-temporally autocorrelated. He states that data mining as a technique is well-suited to finding non-trivial, unexpected relations and patterns in movement data as it can handle the noisy and uncertain nature of movement data.

Laube (2014) roughly categorises movement mining tasks into four groups. 1) Segmentation and data filtering: This task aims at reducing the data complexity and volume. Only the relevant data should be retained for further analysis and this relevant data should be partitioned into smaller parts which are easier to analyse. 2)

Similarity and clustering: There are several similarity measures which allow for a comparison of trajectories. Trajectories might be similar in terms of their shape, location, their inherent properties and other factors. Based on such similarity, trajectories can be grouped together in clusters. 3) Movement patterns: Various patterns can be found in movement data. Laube (2014) emphasises that the definition of these patterns should be grounded in the theory of the according application domain. Nevertheless, Dodge et al. (2008) created a taxonomy of movement patterns which summarises the most frequent patterns. 4) Exploratory analysis and visualisation: The aim of this task is to combine the strengths of computational data processing with the expertise of a human analyst. This analyst might be able to confirm structures that were found by an algorithm or detect things that were retrieved.

The focus of this thesis lies on filtering, similarity and clustering. Therefore, the first two mining tasks will be discussed in more detail in the two subsections (2.1.1 & 2.1.2). The detection of neither movement patterns nor exploratory analysis will be discussed further, as it would exceed the scope of this thesis.

2.1.1 Segmentation & Filtering

The segmentation of trajectories denotes the partitioning of trajectories into multiple subtrajectories called segments. The properties of these segments should be intrinsically uniform (Buchin et al., 2011b). The aim of the segmentation process is to reduce the dimensionality and compress the trajectories to enhance subsequent processing efficiency (Yoon and Shahabi, 2008) or detect underlying semantic meanings such as behavioural states (Edelhoff et al., 2016) or travel modes (Sester et al., 2012). There exist several classifications of segmentation approaches. In this work, the focus will lie on trajectory properties which are used for the segmentation process. Based on these properties, three general ways to segment trajectories can be discriminated: 1) Segment trajectories by only considering geometric properties, 2) segment trajectories by considering movement parameters, and 3) segment trajectories based on underlying semantics.

Geometry-based Segmentation Most work in the field of movement analysis has focused on segmenting spatio-temporal trajectories. However, there exists movement data which does not include a temporal component. Therefore, segmentation approaches only considering geometry are noteworthy. The two most popular segmentation approaches are the Douglas-Peucker algorithm (Douglas and Peucker, 1973) and opening window algorithms. Both of those algorithms are technically line-simplification algorithms that aim at reducing the complexity of a line object. However, with slight modifications, they can be used as segmentation algorithms as well. Cao et al. (2005) applied the Douglas-Peucker algorithm for line segmentation and Gudmundsson et al. (2009) included a temporal component in a variation thereof. Meratnia and de By (2004) outline the mechanics of several segmentation approaches, including Douglas-Peucker and opening window.
Other geometry-based approaches rely on Minimum Bounding Rectangle (MBR) (Anagnostopoulos et al., 2006; Rasetic et al., 2005; Yoon and Shahabi, 2008). These approaches exploit indexing structures in spatial databases, such as R-trees, and build upon those.

Movement-Parameter-based Segmentation Several researchers came up with approaches to segment trajectories based on movement parameters. Buchin et al.

(2010b) introduced a segmentation approach which included various movement parameters or combinations thereof. Their approach ensured that segments are intrinsically homogeneous in terms of these criteria.
Dodge et al. (2009) used profiles of movement parameters for trajectory segmentation and classification. The movement parameter profiles were decomposed based on deviation and sinuosity indices. The decomposition of the profiles can be applied to the trajectories themselves, as they are directly related to them.

Semantics-based Segmentation The previously-discussed approaches use geometry or various attributes of trajectories. These can be considered low-level information, as they have not been interpreted and tied to some semantic meaning. However, Yan et al. (2011) argue that most applications are interested in behavioural aspects of movement rather than simple positional information and movement parameters. They note that semantics may be inferred by spatio-temporal properties of the trajectory, geographic context or other objects related to the trajectory.
Several researchers attempted to enrich trajectories with semantic information. Various approaches segmented trajectories based on episodes of moving and stopping, such as the ones proposed by Alvares et al. (2007) or Yan et al. (2010). Buchin et al. (2013) extended their earlier approach (Buchin et al., 2011b), by including semantics and segmenting trajectories based on movement states. Further noteworthy is the approach proposed by Sester et al. (2012), which segments trajectories at attractive places that are mined from the data.

Filtering The filtering of trajectories aims at reducing noise, removing outliers and compacting data for subsequent processing (Laube, 2014). Commonly-applied filtering techniques include least squares, spline approximation, moving average, kernel-based smoothing and Kalman filtering (Dodge et al., 2009; Eubank, 2005; Jun et al., 2006; Laube and Purves, 2011). All of these approaches aim at smoothing raw GPS data for further processing steps.
Furthermore, some of the data might not be relevant to the analysis and the reduction of data can lead to more efficient processing (Yoon and Shahabi, 2008). This can be done by selection of data using specific queries (Andrienko and Andrienko, 2007).

2.1.2 Similarity & Clustering

The assessment of trajectory similarity is a fundamental challenge in movement analysis and builds the foundation for various further data mining steps, namely clustering trajectories, finding movement patterns and classifying movement data (Dodge, 2011). Ranacher and Tzavella (2014) state that there exists no universal concept of how to assess similarity, even though it is crucial for many further analyses. Lin (1998) provides three intuitions to clarify the concept of similarity: 1) The more commonality there is between two objects, the more similar they are, 2) the more differences exist between two objects, the less similar they are and 3) the maximum similarity between two objects is reached when they are identical. Similarity can either be assessed for whole trajectories or on segments obtained from a preceding segmentation procedure (Ranacher and Tzavella, 2014).
Ranacher and Tzavella (2014) present a thorough review of movement comparison

which is decomposed based on the physical quantities of movement which are of a spatial, temporal and spatio-temporal nature. These correspond to the dimensions on which Dodge et al. (2008) based movement parameters, as was discussed in the previous subsection (2.1.1). Ranacher and Tzavella (2014) distinguished types of similarity measures based on three criteria: 1) applicability to primary or secondary movement parameters, 2) topological or quantitative comparison, and 3) aimed intention of measures. Similar to the previous subsection (2.1.1), the similarity measures presented in this work will be divided into measures solely based on geometric properties (spatial dimension) and measures exploiting advanced movement parameters (temporal and spatio-temporal dimension).

Spatial Similarity There exists a wide variety of similarity measures which are solely based on the spatial dimension of trajectories. Topological comparisons of movement trajectories will not be discussed in this work. The most basic quantitative spatial similarity measures consider the spatial distance between two points. One of the most trivial and also most commonly-used functions to calculate the spatial distance is the Euclidean distance. It describes the length of a straight line between two points in a Euclidean space. Ranacher and Tzavella (2014) note that this function is a special case of the Minkowski distance which is calculated as $d_M(x,y) = \sum_{i=0^m} (|x_i - y_i|^q)^{1/q}$. For $q = 2$ the Minkowski distance corresponds to the Euclidean distance and for $q = 1$ it corresponds to the Manhattan distance. The Euclidean distance can be enhanced with additional dimensions, as long as they can be projected in an Euclidean space. If some other reference system is used, it might not make sense to calculate the length of a straight line between two points. In these cases, distances along curved surfaces or network distances need to be calculated.
Ranacher and Tzavella (2014) further distinguish between measures that evaluate the similarity of whole trajectories (global similarity) and those that only compare some segments of a trajectory (local similarity).
Rinzivillo et al. (2008) introduced a relatively simple measure, for which the distances between the origins and destinations of two trajectories is calculated and averaged. This method can be enhanced by including additional check points between which the distances are measured. If all measured positions in the two trajectories are used to calculate the distance, the measure is referred to as the Euclidean distance between two paths (Zhang et al., 2006).
Andrienko et al. (2007) present a distance function called "common route distance". To calculate this distance, two trajectories are scanned for positions that are within some predefined threshold. Two distance measures are calculated: The Euclidean distance between corresponding positions and a penalty distance for positions that do not match. These two distances are then summed to get the common route distance.
A relatively old, but still widely-used distance measure is the Hausdorff distance. It can be applied to polygons (Alt et al., 1995) and similarly to trajectories (Junejo et al., 2004). The Hausdorff distance corresponds to the maximum distance between two trajectories $d_H(A,B) = max(\delta(A,B), \delta(B,A))$, whereas $\delta(X,Y)$ denotes the maximum distance between two trajectories X and Y.
Another commonly-applied similarity measure is the Fréchet distance. The Fréchet distance is defined as the maximum distance between any two positions of two trajectories that were recorded at the same time (Alt and Godau, 1995). It should be noted that, for this algorithm, the ordering of the locations is critical and not the exact timestamps. Thus, it can be applied to trajectories that lack a temporal com-

ponent. Buchin et al. (2010a) intuitively describe the Fréchet distance as follows: Imagine a person and their dog are walking on two trajectories. Both of them can choose their speed, but they are connected by a leash. The Fréchet distance is the length of the minimal leash that is needed to walk on these trajectories from start to end. Alt and Godau (1995) introduce a variant of the algorithm, which allows the trajectories to move non-monotone. This allows for the comparison of trajectories with an unequal number of vertices.
In order to evaluate local similarity, Lee et al. (2007) combined three distance measures (angular, perpendicular and parallel distance) into one by summing them up. Further noteworthy approaches are One-Way Distance (OWD), and Locality In between Polylines (LIP) (Lin and Su, 2005; Pelekis et al., 2012). OWD is defined by the integral of Euclidean distances between the trajectories, divided by the length of a trajectory. The advantage of this approach is the relatively low computational complexity. LIP uses the area in between two polylines as the distance measure. Pelekis et al. (2012) state that this measure only makes sense if the two polylines follow a similar stable trend.
All previously-described similarity measures only use the spatial positions of trajectories for the similarity assessment. However, there are other purely geometric parameters that can be used for such an assessment. Several works covered the use of travelled distance and range to evaluate similarity (Merrick and Loughlin, 1997; Tottrup et al., 2012). Furthermore, heading can be interpreted as angular measure and, therefore, is independent from a temporal dimension, as was done by Laube et al. (2005) or Melnychuk et al. (2010). Additionally, the shape of trajectories in terms of straightness, sinuosity or fractal dimension can be compared. Benhamou (2004) provides an overview of these measures, which all are a form of tortuosity.

Spatio-temporal Similarity As mentioned in the previous subsection (2.1.1), most movement data contains a temporal component. Hence, trajectories cannot only be spatially similar but also temporally similar. Two trajectories are similar if they move close to one another in space and time (Ranacher and Tzavella, 2014). The inclusion of a temporal component allows for more elaborate similarity measures which are often based on purely geometric methods. The most basic way to include time is to compare positions of trajectories at corresponding times. These methods are called lock-step measures (Ranacher and Tzavella, 2014). Nanni and Pedreschi (2006), for example, calculate the sum of all Euclidean distances between points of two trajectories that have a corresponding timestamp. This value is then divided by the time over which these trajectories exist. Similarly, Buchin et al. (2011a) calculate the average Euclidean distance between trajectories at corresponding times. These two approaches can be seen as extensions of previously-discussed Euclidean distance methods, enhanced by including a temporal component. It needs to be noted that the inclusion of a temporal component may also lead to issues arising from noise and misalignments in time (Ranacher and Tzavella, 2014).
There exist less restrictive approaches, so-called elastic measures, which allow comparisons of positions that do not match in their temporal dimension or only consider some elements of the trajectory. Berndt and Clifford (1994) introduced an elastic method for similarity assessment, which is based on an approach applied in speech recognition, called Dynamic Time Warping (DTW). In order to obtain an optimal match, the temporal dimension of one trajectory is locally compressed and stretched. This allows a minimisation of the distance between the two trajectories, even if there are some small variations in their temporal behaviour.
Vlachos et al. (2002) argued that the DTW approach is relatively sensitive to noise

and presented a more robust approach: Longest Common Subsequence (LCSS). LCSS is a variation of the edit distance which was originally introduced by Levenshtein (1966). The algorithm aims at finding subsequences in two trajectories that are common. Two subsequences are common if the Euclidean distance between them is smaller or equal some predefined distance threshold. The translation of one of the trajectories allows the algorithm to detect similar trajectories that have different sampling rates or speeds, similar motions in different space regions, outliers and different lengths. There exist various other approaches based on the edit distance such as Edit distance with Real Penalty (ERP), Edit Distance on Movement (EDM) or Edit Distance on Real sequence (EDR). These will not be discussed any further in this work (Chen and Ng, 2004; Chen et al., 2004, 2005).
Andrienko et al. (2007) extended their common route distance by including a temporal component. The so-called common route and dynamics distance not only checks trajectories for spatial proximity but also checks whether they were recorded at similar times or not.
Porikli (2004) showed that a Hidden Markov Model (HMM) can be used for trajectory similarity assessment. Based on various attributes, such as coordinates, orientation, speed, aspect ratio, size, etc., a HMM is fitted on each trajectory. The distance is then defined as the cross-fitness of the trajectories to each other's HMM. Even though the basic HMM distance was based on single attributes, Porikli (2004) notes that their mixture would be a perfect candidate for similarity assessment.
Dodge et al. (2012) enhanced the EDM by including additional movement parameters and therewith creating a new distance measure: Normalized Weighted Edit Distance (NWED). The NWED does not consider the raw trajectories for similarity measurement but symbolic representations of the trajectories. The trajectories are segmented based on the profiles of movement parameters and each segment is assigned a class membership. Two measures were used, namely deviation from the mean value and sinuosity of the profiles. The NWED measure is obtained by computing the number of edit operations (i.e. insertion, deletion, substitution) needed to transform one trajectory representation into another. Dodge et al. (2012) only used speed and turning angle to compute the NWED in their experiments. Nevertheless, other movement parameters or even context variables of which profiles can be obtained can be used in the NWED.

Clustering Clustering methods for point data can generally be divided into four major classes according to Han et al. (2009). It is important to understand these clustering methods before looking into clustering methods for whole trajectories as they are based upon these.

Partitioning Methods Partitioning methods classify the whole dataset into a predefined number of classes so that all classes contain at least one element and each object is in a class. In an iterative manner, the objects are reclassified so that the differences between the classes become larger and the differences in-between the objects of those classes become smaller. The most common partitioning methods are k-means (Lloyd, 1982), k-medoids (Kaufman and Rousseeuw, 2005) and the EM algorithm (Dempster et al., 1977).

Hierarchical Methods Hierarchical methods hierarchically decompose a large dataset into classes. This can either be done in an agglomerative (bottom-up) or divisive (top-down) manner. The main idea is to either combine the closest points/clusters in a new cluster or split the largest cluster into two smaller

ones. Some well-known examples are AGNES and DIANA (Kaufman and Rousseeuw, 2005). Han et al. (2009) note that the irreversible nature of the splits and merges is a major drawback of these methods. They discuss two approaches that were developed to avoid this issue: BIRCH (Zhang et al., 1996) and Chameleon (Karypis et al., 1999).

Density-based Methods Density-based methods iterate through all data points and identify clusters based on the point density. Clusters are found by identifying points that have more points in their ε-neighbourhood than some predefined threshold. These points are called core-points. Core-points are summarised in clusters together with their border-points (in ε-neighbourhood of core-point but not enough neighbours). All points that lie outside of the ε-neighbourhood of core-points are outliers. The most prominent representatives of density-based methods are DBSCAN (Ester et al., 1996) and OPTICS (Ankerst et al., 1999).

Grid-based Methods Grid-based methods divide the feature space into a finite number of cells and calculate the density of those cells. The clusters are generated based on the grid cells with high densities and their cell neighbours. The most common grid-based algorithms are STING (Wang et al., 1997), CLIQUE (Agrawal et al., 1998) and DENCLUE (Hinneburg and Gabriel, 2007).

According to Kisilevich et al. (2010), there are several types of approaches that can be applied to trajectories. A major group are clustering approaches that are based on distance/similarity measures. For these approaches, the clustering problem is reduced to a choice of a similarity measure, which have been discussed previously in this subsection, and a choice of a generic clustering algorithm. Both the partitioning and hierarchical methods determine the affiliation of an object to a cluster, based on the similarity of the object to a cluster. In the case of trajectories, the cluster centers are also a trajectory. Therefore, it is possible to assess the similarity of a trajectory and a cluster center using one of the previously-discussed measures. Furthermore, it is also possible to apply density-based methods to trajectories. For every trajectory, an ε-neighbourhood can be computed and it is possible to determine which trajectories lie within this neighbourhood.
Worth mentioning are also visual-aided approaches, which try to combine computational power with human expert knowledge, as was shown by Andrienko et al. (2007, 2009) and Andrienko and Andrienko (2011). Additionally, Kisilevich et al. (2010) note that model-based clustering approaches can be applied to trajectory data. Gaffney and Smyth (1999) used a generative mixture model for clustering trajectories and Alon et al. (2003) presented an approach which is based on a HMM. These approaches will not be discussed in detail in this work.

2.1.3 Geographic Context in CMA

Several researchers recognized the importance of geographic context in movement analysis (Buchin et al., 2012; Gschwend and Laube, 2012; Purves et al., 2014). In Dodge's (2011) conceptual model of movement, external factors are a major element, along with movement parameters. Especially in the absence of a temporal component, the inclusion of a context dimension might provide valuable additional information for analyses. Nevertheless, most CMA approaches focussed solely on the spatial and temporal dimension, disregarding geographic context. Purves et al. (2014) identified this as a research gap in the field of CMA. In this subsection, exist-

ing approaches that include context will be discussed.
Andrienko et al. (2011) introduced an event-based conceptual model for movement analysis including context and provided a general approach thereto. Lautenschütz (2010) conducted a user study investigating the influence of context on trajectory segmentation. However, the focus of her study was on the human identification of breakpoints and not the computational analysis of movement trajectories.
Siła-Nowicka et al. (2016) implemented a semantic segmentation approach with mobility data and identified significant places, similarly as was done in previously-discussed works (subsection 2.1.1). They used point of interest data as a form of contextual information to classify the obtained significant places. Even though geographic context is likely to affect the segmentation process of trajectories, there is a lack of research on computational context-aware segmentation approaches.

Buchin et al. (2012) indicated several forms of geographic context, such as network, land cover, obstacles, terrain, ambient attributes (e.g. meteorological attributes) or other agents. They state that various forms of context can be treated as trajectory attributes. Hence, CMA approaches that consider such trajectory attributes could easily be enhanced with contextual attributes. Buchin et al. (2012) focussed on including context in similarity assessments. They distinguished three approaches to include context in such approaches, based on how the spatio-temporal and the contextual component are handled. These two components can either be treated as 1) equal, creating a multi-dimensional space in which the similarity is computed, 2) independent, whereas similarity for both parts is computed separately, or 3) integrated, where the two components are combined in a new similarity measure. They argue that the first approach is not appropriate, as the mapping of attributes might lead to a loss of information and that the different components (time, space & context) should not be treated equally. Thus, they presented an approach for the implementation of an integrated distance measure that extends existing distance measures (Hausdorff & Fréchet) with context costs.
In a recent study, Sharif and Alesheikh (2017) introduced a context-aware approach which is based on DTW. They distinguished between four different types of context 1) motivation context, describing the reason the individual is moving, 2) movement context, including primary and secondary movement parameters, 3) modality context, characterising the individuals condition, specification and movement capacity, and 4) milieu context, encompassing external factors influencing the movement. These four contexts are added to trajectories as additional dimensions and included in distance measurements which are used for the calculation of the so-called Context-based Dynamic Time Warping (CDTW).

2.2 Volunteered Geographic Information

Quite often, trajectories that are analysed with CMA approaches are produced and provided by users. The emergence of the Web 2.0 in the 2000s, which was described first by O'Reilly (2005), enabled users to share content online and built the groundwork for the general web phenomenon known as User-Generated Content (UGC). Vickery and Wunsch-Vincent (2007) defined UGC, referred to as User-Created Content (UCC), as content made available online that reflects some amount of creative effort and was created by non-professional producers. Goodchild (2007a) introduced the term Volunteered Geographic Information (VGI) which denotes geographic content shared online. Therefore, VGI is a special case of UGC. Other VGI-enabling

technologies besides the Web 2.0 are GPS, georeferencing, geotags and broadband communication. To date, various platforms that allow the sharing of geographic information exist and many platforms that focus on other content (e.g. multi-media) provide the possibility to geotag content. As VGI is created by non-professional users, the quality of such data is uncertain and several issues can arise. These will be discussed in subsection 2.2.1. Nevertheless, VGI is applied in various fields, such as mapping, disaster management, citizen science and avalanche research. The applications of VGI will be discussed in subsection 2.2.2.

2.2.1 Issues with VGI

Nielsen (2006) was the first to describe the participation inequality phenomenon, one of the major issues that come with UGC. He noted that a large share of users in online communities only consume content without producing any. Only a small minority of users frequently add content to those websites and share it with the community. The user-participation mostly follows a 90-9-1 rule, where:

- 90 % of users lurk in the background, read or observe but do not contribute.
- 9 % of users contribute content from time to time but read or observe most of the time.
- 1 % of users regularly contribute content and, therefore, are responsible for most content.

As a result, 90 % of content is produced by only 1 % of the most active users and 10 % of content is produced by 9 % occasional contributors. This leads to an overrepresentation of the 1 %-user group which can bias the understanding of a community and has an impact on quality, coverage and content (Haklay, 2016). Techel et al. (2015) observed that the participation inequality might lead to a geographical bias in VGI, meaning that there is an overrepresentation of some areas. Therefore, it is important to account for participation inequality before the analysis of UGC data. Nielsen (2006) proposes some approaches to address participation inequality, including rewarding contributing users, promoting quality contributors or easing the effort to contribute. Clearly, the 90-9-1 distribution Nielsen (2006) introduced does not hold equally for all communities and may be more or less skewed. These effects were shown for general UGC-platforms, including Wikipedia (Javanmardi et al., 2009), Flickr, and Twitter (Li et al., 2013), but also for specific VGI-platforms, such as Open Street Map (OSM) (Budhathoki, 2010; Neis and Zipf, 2012) and mountaineering websites (Techel et al., 2015).

Closely linked to participation inequality is the digital divide in VGI. Even though VGI are publicly accessible over the internet, two-thirds of the world's population cannot access the world wide web, and, consequently, VGI platforms (Goodchild, 2007a; Sui et al., 2013). With their analysis, (Neis and Zipf, 2012) showed that 72 % of all OSM contributors are European, 12 % American and only 16 % are from the remaining continents. Haklay (2010) found a large disparity in OSM coverage between poor rural and urban areas in the UK. In general, this digital divide leads to an overrepresentation of some user groups, similar to participation inequality, that researchers need to be aware of and account for. Sui (2015) noted that even though there has been progress in overcoming the digital divide in the VGI setting, the world remains divided and further work should address this issue.

Another common issue with VGI data is lack of quality control and gatekeeping. Geographic data from traditional data providers, such as governmental agencies, cartographers and geographers, is usually gathered by trained professionals with a relevant background. Furthermore, the data passes through strict quality control gates. However, VGI is not bound to such processes and people without professional training, so-called neo-geographers (Turner, 2006), can create and share data. Despite such data being relatively cheap, the lack of quality control measures may lead to reduced data quality. As filtering through professional gatekeepers may not be enforced, the data might be poorly organised, out of date, incomplete or inaccurate. Additionally, metadata is not necessarily included and, therefore, source information might be unavailable, masked or entirely missing (Flanagin and Metzger, 2008). Further, this leads to obscurity regarding the producer of the data and its purpose (Purves, 2011).
The motivations to create VGI are manifold and often thought to be altruism and project success. However, Budhathoki (2010) found that the most important motivations are related to monetary, instrumentality of local knowledge and self-view factors. With the increasing popularity of VGI-platforms and services, malicious contributions are to be expected. The user's motivation strongly influences content quality, as users might intentionally bias, reinterpret or manipulate data (Antoniou, 2011; Purves, 2011). The large volumes of VGI that are continuously produced make it difficult to guarantee quality as traditional means are unemployable (Flanagin and Metzger, 2008). While Goodchild noted a lack of mechanisms to ensure quality in 2007, he and a colleague proposed three approaches to assure the quality of VGI data five years later (Goodchild and Li, 2012).

Data quality is closely linked to credibility, which was defined by Hovland et al. (1953) as a two-dimensional concept consisting of expertise and trustworthiness. The authoritative nature initially endowed traditional producers of geographic information with relatively high credibility, which was retained by upholding high quality standards for data production (Flanagin and Metzger, 2008). Budhathoki (2010) showed that only people who are motivated by local knowledge contribute VGI on a regular basis. Hence, VGI contributors might not have benefited from professional training, but they exhibited expertise of their local surroundings. In contrast to authoritative data producers, they are immersed in their environments in various ways and therefore gain access to knowledge which is inaccessible or difficult to obtain for traditional producers. Moreover, Flanagin and Metzger (2008) illustrate how new networking and peer-to-peer credibility assessment approaches help to overcome the idea of authoritative producers being the only credible data sources. However, Purves (2011) warns of circularities in such data sources if they enjoy too much credibility. He notes that further data production might be negatively influenced by previously-gathered erroneous data. Users might associate wrong names with coordinates or assign wrong or inaccurate coordinates which would further lead to ambiguity during geoparsing.

Shortly after the term VGI emerged, several researchers raised concerns about privacy (Elwood, 2008). Obermeyer (2007) pointed out that, despite the benefits of new technologies, risks arise from the misuse of GIS and VGI. In contrast to data collected by the government, which by law needs to be aggregated to protect personal privacy, VGI is not affected by such privacy regulations (Li and Goodchild, 2013). Various sources of information can be combined and potentially expose private information such as significant places, health conditions or other personal data (Mooney et al., 2017). Li and Goodchild (2013) provide an example by showing how home and

work locations of Twitter users can be inferred from geotagged Tweets combined with land use data. Thus, it is important that the user's privacy is protected and the contributed information cannot be linked to them. Mooney et al. (2017) described a number of approaches that work towards privacy protection but noted that there are still several privacy issues that have not been accounted for.

When working with VGI, it is also important to consider ethical issues. Mooney et al. (2017) stated that producers of VGI need to be included in the research process, collecting private data should be avoided, and users should be informed about the usage of their data and be allowed to modify or delete information that they want to keep private. Closely linked to these ethical and privacy issues is the differentiation between Volunteered and Contributed Geographic Information (CGI), as was pointed out by Harvey (2013). He defined CGI as data that was produced automatically and uncontrolled by users. Examples are data that are collected by a navigation system or by using a cell phone. A main difference between the two types is that one is opt-in, meaning that the user has to control the service and willingly shares geographic information, and the other is opt-out, requiring the user to accept all terms and conditions unconditionally, leaving him with no control. The differentiation between these types of data has implications for data quality and potential biases in data.

Lastly, legal issues with VGI need to be acknowledged. According to Mooney et al. (2017), the major legal issues concern liability and data license types. In the case of economic loss or incorrect decisions, which can be linked to erroneous or low quality VGI, it needs to be clarified as to whether and to which degree the contributors and/or platform providers can be hold accountable. Mooney et al. (2017) state that there are currently no clear answers to this question. There are three major open license types which are commonly used with open data: share alike licenses, open licenses and limited open use licenses. Problems occur if platforms merge their datasets or mash-ups from different sources are created and these platforms or sources use different licensing types. Mooney et al. (2017) provide an example from France, where the national mapping agency planned to integrate their address data set, the one from the French Post Office and the OSM data set. All of these datasets were under different licensing and their approach was for users to contribute to two separate data sets that use different licensing that they need to accept. This example shows one possible approach for such a licensing issue but, clearly, future research should tackle this challenge and develop standardised resolutions.

2.2.2 Applications of VGI

The range of applications of VGI is broad. One of the most well-known VGI platforms is Open Street Map (OSM). By the end of January 2018, OSM had attracted almost 4.7 million users of which about 1 % were active contributors (OpenStreetMap, 2018). The provisioning of free current geographic information of the whole world is a major motivation for the project (Haklay and Weber, 2008). Geographic information from authoritative sources is usually rather costly. OSM can provide an alternative for individuals, small businesses or organisations which cannot afford authoritative data. The quality of OSM data is inconsistent and does often not attain the quality of authoritative data. However, the speed at which datasets can be collected is impressive and is one of VGI's main advantages (Haklay, 2010). Especially in usage scenarios where no current authoritative data is available and quick data

collection is necessary, the advantages outweigh the sometimes lacking quality of VGI data. Goodchild and Glennon (2010) argue that, especially during emergencies, communities can effectively assist responders and emergency managers by means of VGI. They illustrated how the community of Santa Barbara contributed to VGI during a series of large wildfires and what lessons could be learned from these incidents. Zook et al. (2010) analysed how VGI was used for disaster relief efforts after the earthquake in Haiti on 12 January 2010. They showed that the available VGI for Haiti greatly increased shortly after the earthquake. Even though this information does not help response teams on a physical level, it still can enhance logistical systems on which the relief efforts are grounded.

One of the oldest areas of application of VGI is citizen science. People share observations and geographic data with researchers for the purpose of knowledge generation and scientific progress. VGI is relevant to citizen science projects, where the location of the collected information is of importance to the research and a major part of the citizens' research activity (Haklay, 2013). Citizen science is disseminated in various fields such as ecology (Wiersma, 2010), environmental monitoring (Connors et al., 2012), remote sensing (Fritz et al., 2012b) and natural hazard monitoring (Longueville et al., 2010), just to name a few.

VGI also plays a role in urban planning. With the help of online mapping interfaces, where design proposals are shared with the public, users can share their ideas and inputs with decision makers. This enables the involvement of citizens, who usually would not be able to contribute to such projects. These tools can help decision makers enhance the information available, get a better idea about public opinion and therefore avoid confrontations, maintain credibility and legitimacy and reduce implementation costs (Seeger, 2008).
Transportation networks are a crucial part of any urban system. In recent years several researchers focused on improved planning of cycling routes, with cycling being a sustainable mode of transport. By analysing GPS tracks of cyclists, it is possible to gain insights into urban travel behaviour and generate route choice models, as was done by Menghini et al. (2010) for the city of Zurich. This enables planners to make more elaborate decisions based on cycling traffic flows. Pánek and Benediktsson (2017) showed how emotional mapping can capture the reaction from cyclists to their environment. Their participatory planning approach aimed to close the gap between urban planners and cyclists. There exist several cycling-specific VGI platforms which allow users to share their trips with the community. This allows users to plan routes based on aspects that are relevant to them, such as motorized traffic volume, safety, required fitness, etc. (Kessler, 2011). Nelson et al. (2015) introduced a cycling safety tool which allows user to report on collisions or near-misses, which not only helps users to choose safer routes, but also allows urban planners to mitigate risks at critical road passages. These works illustrate that VGI can come in various forms, including trajectory data.

There exist many other tracking applications for recreational activities besides cycling, such as running, hiking and backcountry skiing. Hendrikx et al. (2013) were the first to use GPS tracks to gain insight into travel behaviour in avalanche terrain. They handed out GPS devices to experienced backcountry skiers which allowed them to track their tours. The aim was to obtain a better understanding of decision-making processes in avalanche-prone terrain, as these processes play a vital role in avalanche accidents. By aggregating the gathered GPS tracks with a Digital Elevation Model (DEM), they were able to obtain a rich data source for analysing decision-making processes during backcountry travel. Similarly, Haegeli and Atkins (2016)

equipped heliskiing guides with GPS-tracking devices to capture risk management approaches in avalanche terrain. Both argue that an expanded dataset would allow more thorough analysis of avalanche accidents. Consequently, Hendrikx et al. (2014a) created a freely-available smartphone application with tracking capabilities. This allowed them to gather a larger sample set, which was analysed over multiple seasons (Hendrikx and Johnson, 2016).
Techel et al. (2015) took a different approach to gathering VGI than the previously-discussed crowd-sourcing. They obtained existing geotagged tour reports from two social media mountaineering websites. These were used to analyse recreational backcountry usage patterns which were compared with avalanche statistics. They found that weather conditions and avalanche danger level likely are the main criteria during the planning phase of backcountry tours. Plank (2016) also retrieved data from an online platform focusing on outdoor activities. He was able to gather more than 6000 ski tour entries. In combination with other data, such as avalanche danger, weather conditions and snow conditions, the risk of the tours was analysed. He found that many tours are relatively dangerous. Therefore, recreationists should not simply copy tours on such platforms, but consult guide books and other external information. Even though these works were solely first approaches, they show that VGI can also play a crucial role in avalanche research in the future.

2.3 Avalanche Research

Avalanches not only endanger the life of alpine recreationists, but they also threaten infrastructure and transportation networks. Therefore, a thorough understanding of avalanche processes is vital to preventing economic loss and fatalities. An avalanche is defined as "A large mass of snow, mixed with earth and ice, loosened from a mountain side, and descending swiftly into the valley below." (OED Online, 2018). In this work, the focus will lie on the several factors influencing avalanche processes. Specific avalanche situations, detailed snow physics, etc. will be omitted, as they are not relevant to the analysis presented later on. Fredston and Fesler (1994a) summarised the four main factors, weather, snow cover, terrain, and human, in the avalanche triangle (figure 2.3). All of these factors can be divided into subfactors, many of which are interconnected. Subsection 2.3.1 will focus on the geophysical factors, while subsection 2.3.2 will elaborate on the human factor in avalanche processes.

2.3.1 Geophysical Factors

Weather According to Harvey et al. (2012), the various weather parameters need to be analysed together, as their combination influences the snow stratigraphy. The major weather factors are precipitation, wind, temperature and radiation. Precipitation can either come in the form of snow or rain. Fresh snow can influence the snow cover in two different ways. It can form a new layer in the snow cover, which rests on older snow. If these two layers do not bond quickly enough, the new layer might slide on the older layer. Furthermore, fresh snow can act as an additional weight on the existing snow cover. By this weight, existing weak layers might collapse and lead to an avalanche. Temperature and wind influence the density and redistribution of the falling snow (Harvey et al., 2012).
If the precipitation comes in the form of rain, it can introduce heat into the snow cover. However, more influential is the added weight on the snow cover. Similar

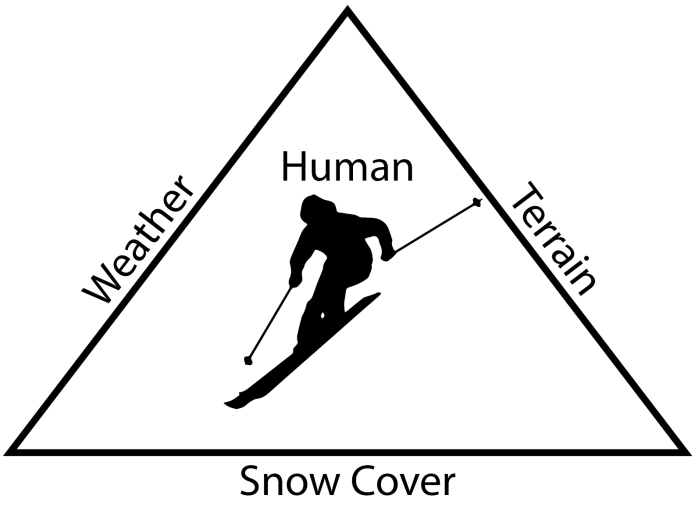


Figure 2.3: *The avalanche triangle summarises the four major factors of avalanches. Own figure after Fredston and Fesler (1994a).*

to the fresh snow, the additional weight might lead to damage to underlying weak layers which can collapse (McClung and Schaerer, 2006).

McClung and Schaerer (2006) differentiates two wind components: horizontal and vertical. The horizontal component is mainly responsible for snow relocation. The relocation is highly dependent on local terrain features. Snow deposition mostly occurs in convex terrain features such as gullies, notches, etc., sharp changes of slope and close to ridges on the lee side. This can lead to local instabilities or cornice formation.
The vertical component determines the amount, rate and distribution of precipitation. Due to vapour condensation, rising air causes rain or snow fall. Therefore, precipitation patterns and intensity are strongly dependent on large-scale topographic features and wind. Furthermore, warm downward winds might cause crust formation on the snow cover, which can later act as a weak layer.

Harvey et al. (2012) state that the snow cover temperature is dependent not only on the air temperature, but also on radiation and wind. Heat exchange between air and snow cover can occur due to turbulent exchange (wind) or condensation of vapour (McClung and Schaerer, 2006). Even though the snow cover temperature rarely is the main factor in avalanche accidents, the influence is not negligible. Harvey et al. (2012) note that the impact of a skier's weight on the snow cover is enhanced if the temperature is relatively high. Furthermore, vaporization of snow to and from the surface of the snow cover can lead to surface hoar, which can act as a weak layer (McClung and Schaerer, 2006).

Radiation can have a heating or cooling effect on snow cover. Two types of radiation influence snow cover: short-wave radiation from the sun and terrestrial long-wave radiation. Whether they have a warming or cooling effect on snow cover depends on the balance of the two, in the absence of additional factors such as wind. The short-wave radiation is strongly influenced by the irradiation angle, and, therefore, by the exposition of a slope. Furthermore, fog and cloud cover can have a greenhouse effect on snow cover (Harvey et al., 2012). McClung and Schaerer (2006) discuss how large temperature gradients in snow cover, resulting from radiation, can lead to the formation of weak layers on the snow cover surface.

Snow Cover The stratigraphy of snow cover is a major factor in avalanche processes. As discussed in the previous paragraph, weather and terrain factors strongly impact the structure of the snow cover. The temperature of the snow cover and the temperature gradient influence snow metamorphism. Snow metamorphism denotes the change of form of snow particles and differentiates between dry and wet snow (McClung and Schaerer, 2006). Snow metamorphism can lead to persistent and non-persistent forms of weak layers in the snow cover. Persistent forms have a low fracture toughness, meaning they can be fractured easily, resulting in avalanches. They can last from months to whole seasons. In contrast, non-persistent forms usually develop with fresh snow and only last for a short period of time. Fresh snow bonds quickly with old snow, strengthening snow cover stability (McClung and Schaerer, 2006).

Terrain Besides direct influences on avalanche processes, the terrain also affects weather and snow cover properties. Slope inclination is a primary factor in avalanche formation, as it influences the gravitational forces on the snow mass (Harvey et al., 2012). McClung and Schaerer (2006) note that there exists no clear inclination boundary, below which slopes can be regarded as safe. Most avalanches, however, occur between 30° and 50° (Harvey et al., 2012). If the slope inclination is lower than 30°, the gravitational forces are usually too small for an avalanche to develop. If the slope inclination is higher than 50°, most fresh snow cannot be deposited and the amount of snow is not sufficient for an avalanche.

According to McClung and Schaerer (2006), the slope exposition influences two weather parameters. Depending on the predominant wind direction in an area, some slopes regularly collect drift snow. This can be observed for slopes on the lee side of ridges. Furthermore, the exposition influences the irradiation. Sun-exposed slopes receive high amounts of sun radiation, which leads to large fluctuation of temperatures in the snow cover between day and night. Thus, new snow can bond to existing layers faster which results in lower avalanche danger. Shady slopes usually receive less sun radiation and, therefore, weak layers can persist over longer periods of time (Harvey et al., 2012).

Certain landforms are more prone to avalanches than others. This is mostly due to the deposition of drift snow which is favoured in gullies and other concave landforms. Additionally, such landforms are often relatively steep which further amplifies the avalanche danger. Areas below ridges and slope drop-offs also favour snow deposition (Harvey et al., 2012). McClung and Schaerer (2006) denote such features as terrain traps. Moreover, convex slopes exhibit increased avalanche danger. This is due to tension in the snow cover (McClung and Schaerer, 2006). Slopes interspersed with boulders also exhibit increased avalanche danger, as stated by Harvey et al. (2012). The boulders not only disrupt snow cover; they often show high temperature gradients in the adjacent snow cover. Nevertheless, slopes with high roughness show smaller avalanche danger than smooth slopes. Rough slopes prevent the formation of continuous weak layers, as is noted by Schweizer (2003).

The altitude only has an indirect effect on avalanche danger. Snowfall, wind and temperatures vary with elevation (McClung and Schaerer, 2006). Harvey et al. (2012) distinguish three main elevation levels: 1) below the tree line, where snow cover stability is relatively good, due to temperature variations and little wind, 2) large homogeneous slopes above the tree line, which are generally smooth and show little wind erosion and consistent snow cover stability and 3) ridge areas, where strong winds

and cold temperatures are prevalent, leading to high instability, cornices, etc.

2.3.2 Human Factor

Even though 95 % of all avalanche accidents are human-triggered, the focus of research has mainly been on geophysical aspects of avalanches (Harvey et al., 2002, p. 449; McClung, 2002, p. 111). However, the realisation that humans play a crucial role in avalanche accidents is nothing new. McCammon (2009) summarised that, as early as 1892, the author of a mountaineering guide wrote that mountaineers should be aware of avalanche conditions and act accordingly. In the following years, several authors developed simple decision-making strategies for trips in avalanche-prone areas. The first forecasting approaches in the 1950s heavily focused on quantitative factors according to McCammon (2009). It was not until the 1980s that the term "human factor" was introduced by Fesler (1981, as cited in McCammon, 2009). Soon after, the term gained wide acceptance and is since commonly used in avalanche education.
There is no clear definition of the human factor and but most authors describe it as a combination of risk propensity, misperceptions, attitude, pressure and other aspects which influence decision-making processes (Fredston and Fesler, 1994b; Harvey et al., 2012; McClung, 2002). As these aspects, being mainly psychological components, are difficult to ascertain, empirical evidence was not available until the 2000s (McCammon, 2009).

Decision-making in avalanche terrain aims at maximising the enjoyment of backcountry skiing, while keeping the risk below a certain level, above which recreationists would be exposed to excessive danger (McClung, 2002). Traditionally, avalanche education has provided recreationists with scientific knowledge about avalanches, which should then be applied in avalanche terrain (Haegeli et al., 2010). However, avalanche processes are composed of a multitude of parameters and, therefore, are of such complexity that a full analytical evaluation in the field is not feasible by non-experts. Hence, various authors proposed heuristics, simplistic rules focusing on key elements of avalanche processes, for simplified decision-making processes or decision aids (Furman et al., 2010; McCammon, 2001, 2009; Munter, 1997). Most studies, analysing decision-making in avalanche terrain, utilised avalanche accident records (Atkins, 2000; McCammon, 2002) or (online) surveys (Atkins and McCammon, 2004; Haegeli et al., 2010). However, recently, the first approaches using recorded GPS tracks were analysed to obtain a better understanding of decision-making processes, as was seen in section 2.2.

Decision-making for backcountry tours starts with the choice of an appropriate destination under given conditions. Recreationists should consider a range of information to obtain an overview of the current avalanche situation, such as avalanche bulletins, weather reports and maps (Harvey et al., 2012). Haegeli et al. (2010) analysed how decision aids can influence the choice of destination. They were able to show that simple decision aids were able to influence human behavior causing a shift towards more avalanche sensitive behaviour.

The analysis in the work at hand will draw upon these findings and try to extend the understanding of decision-making during the planning process of backcountry tours. VGI data will be explored using CMA techniques to determine the influence of avalanche-related information on decision-making and build the groundwork for further analysis of such data.

2.4 Research Gaps

The data used in this work consists of a number of trajectory-like objects. In contrast to GPS-tracks, these assumingly do not include timestamps and therefore lack a temporal component. This has a severe impact on the applicable analysis approaches. As most movement parameters rely on a temporal component, as was shown in section 2.1, the application of algorithms based on movement parameters is not suitable for such data. Therefore, it is inevitable that the focus be on available properties, such as geometric and contextual properties. There exist several approaches relying on such properties, as discussed in section 2.1.2. However, to date, no work on planned route trajectories has been conducted. Therefore, the first research gap identified in this thesis is the lack of appropriate CMA approaches to handle planned route trajectories.

VGI is a rather new data source for avalanche research. Recent work focused on human behaviour in avalanche terrain based on the analysis of recorded GPS-tracks. Analysis of planned movement data has so far been neglected, indicating a further research gap. Therefore, an aim of this work is to introduce planned movement data as a new kind of VGI for avalanche research. Harvey et al. (2012) note that thorough planning of backcountry tours is central to risk reduction. The used data might provide insights into this planning phase, a part of the touring process. This should improve understanding of decision-making processes during the planning phase.

2.5 Research Questions

Based on the state of research and the identified research gaps, four research questions were developed. These research questions should serve as a guideline for the remainder of this thesis. The results obtained through the analysis should help answer those research questions and shed light on the identified research gaps.

Research Question I What are the differences and similarities between planned backcountry routes and real backcountry tour GPS-tracks and how will they affect the further analysis of the planned backcountry routes? How can planned route trajectories be characterised?

Hypothesis I As planned routes are drawn on a map, several properties occurring in actual GPS-tracks, such as timestamps, speed, elevation, etc., are missing, or generated artificially. These missing or artificially generated properties need to be considered for further analysis of the routes. Geometric properties of planned route trajectories should adhere to those from real GPS-tracks.
Contextual information, extracted from secondary sources, will help to characterise planned route trajectories.

Research Question II How can test routes created by users be distinguished from 'real' planned routes? Which measures need to be taken to filter a dataset of planned route trajectories so that most remaining routes depict a realistic planned backcountry tour?

Hypothesis II Most test routes do not adhere to typical route lengths and are generally too short or too long. Furthermore, 'real' backcountry ski tours will be drawn in greater detail, leading to a higher number of vertices and smaller step lengths. Additionally, test routes are expected to show atypical geometries which do not follow any environmental features. This affects attributes extracted from external sources such as digital elevation models (DEMs) and their derivatives. Critical attribute values can be determined, which allow the querying of the dataset. This will enable the removal of unrealistic planned routes from the dataset.

Research Question III What type of intrinsic or extrinsic information can be used to determine the degree of accuracy of planned route trajectories?

Hypothesis III For each realistically-planned backcountry tour, there should exist various similar backcountry tours nearby. Furthermore, realistically-planned route trajectories should exhibit similar geometries as GPS-tracks recorded by users during backcountry ski tours. Some users append additional information to planned route trajectories, such as avalanche danger level and tour date. This information should correspond to information gathered by authoritative sources, if specified correctly by the users.

Research Question IV What can be learned from planned route trajectories about the planning process of backcountry tours in avalanche-prone terrain?

Hypothesis IV Through simple route statistics, typical route characteristics can be identified. By aggregating the routes, an overview of the spatial distribution of planned backcountry activity can be obtained. Furthermore, additional user data can provide insights into avalanche danger or coarse temporal distribution of planned backcountry activity. The combination of avalanche danger and route attributes will allow obtaining an understanding of the influence of avalanche bulletins on the planning process of backcountry tours.

The following two chapters (3 & 4) will cover the data and methods used to answer these research questions. In chapter 6, these research questions will be taken up again and discussed based on the results achieved.

Chapter 3

Data & Study Area

This chapter will provide an overview of the data used in this thesis and the study area that was chosen. The aim of this chapter is to provide insights into data characteristics, how this data was gathered and which geographic areas were considered for the subsequent analyses. Furthermore, a detailed characterisation of the planned route trajectories should allow detecting structural differences with 'real' GPS-tracks. This will provide the initial clues for answering the first research question as well as a solid ground for the selection of appropriate methods, which will be discussed in chapter 4. In order to do so, the data used will be discussed in detail, including its properties and sources. The study areas will be delineated in words and illustrated as maps.
In the first section (3.1) the planned backcountry tours will be introduced and discussed in detail, including the tools with which data were gathered. Then, secondary data that was used will be addressed. These include digital elevation models, avalanche bulletins and recorded GPS-tracks. The study areas will be defined and elucidated in section 3.2. As some computational analyses were computationally-expensive, besides the use of a large study area for large-scale analyses, three small study areas were selected.

3.1 Data

For this thesis, data from different sources was analysed and combined. This section aims at giving an overview of this data, presenting its sources, types and characteristics. The main data is the planned backcountry tours. Other secondary data, such as DEMs, avalanche bulletins, and GPS tracks, was used to enrich the planned tours and put them into context.

3.1.1 Planned Backcountry Tours

The main dataset analysed in this work contains the planned backcountry tours. Data on these tours was gathered through the White Risk platform[1], which was developed by the SLF. White Risk was originally designed as an avalanche learning

[1] www.whiterisk.ch

CD-ROM (Harvey, 2006). It covered various aspects of avalanche theory, including interactive animations, self-assessment tests and multimedia content. The aim was to create an up-to-date supplement for existing analogue information sources that supports autodidactic learning of avalanche theory. Two years later, Harvey et al. (2008) introduced the White Risk Instructor, a presentation tool for avalanche educators. As mobile phones gained popularity, a smartphone application was introduced: White Risk Mobile (Suter and Harvey, 2009). Initially, it included current avalanche and weather information and allowed users in the field to share observations with the SLF. An online GIS-based trip-planning tool followed shortly after (Eckert and Suter, 2010; Eckert, 2011). In 2013, the e-learning, instructor and trip planning tools were consolidated in a single web-based platform: White Risk 2.0 (Harvey et al., 2013). In connection with the White Risk smartphone application, alpine recreationists are provided with a full suite of tools for backcountry trips and other avalanche-related topics.

The prototype of the trip-planning tool was created by Eckert and Suter (2010). It allowed users to draw routes on a digital map. Furthermore, they were able to mark key passages (cruxes), check hill slope, elevation and exposition, as well as current avalanche and weather conditions. The planned tours were stored in a database so users could download their planned tours to their mobile phones. During the trip, georeferenced photos and texts could be shared with the community directly from the mobile application.
With the launch of White Risk 2.0, the trip-planning tool was redesigned and the mobile counterpart was integrated in the White Risk mobile application. The main functionalities remained and users can still draw routes directly on a digital map in their web browser. Besides, recorded tours can be imported from GPX-files and drawn routes can be exported as GPX-files in the main frame. The GPX format is an open GPS exchange format, which can contain locations, timestamps and additional information. White Risk includes maps from several national mapping agencies (Swisstopo, IGN & BEV), as well as from OpenStreetMap. Additional layers including slope angle, wildlife closures, or existing tours from the Swiss Alpine Club (SAC) can be displayed. Besides a main route, alternative routes and cruxes can be planned. A content box showing tour details, such as expected trip time, distance, peak elevation, and elevation differences, is included in the planning interface. All additional information is related to the main route. Alternative routes do not influence the tour details. Figure 3.1 shows a capture of the White Risk trip-planning tool interface.

In the "Assess conditions" tab, users have direct access to the avalanche bulletin, various snow maps and weather forecasts for tours planned in Switzerland. For tours in other European countries, a web-map is displayed that links to the avalanche bulletins of various European avalanche centres. Snow maps and weather forecasts are not available for tours outside of Switzerland. In a side box, users can store current avalanche patterns, danger levels and weather conditions for their planned tour. Figure 3.2 shows a screen capture of the assess conditions tab.

In the "Additional preparation" tab, users can fill out a check list for the tour, add participants of the tour and add remarks. Finally, in the "Share & bring along" tab, users can obtain a link to the tour, create PDFs containing the most important tour information and maps, and export the tour as a GPX-file. The GPX-file can be used with external GPS devices. In White Risk Mobile, users have direct access to their planned tours upon logging in. The tour can be displayed on a map including the user's position and most of the additional stored information.

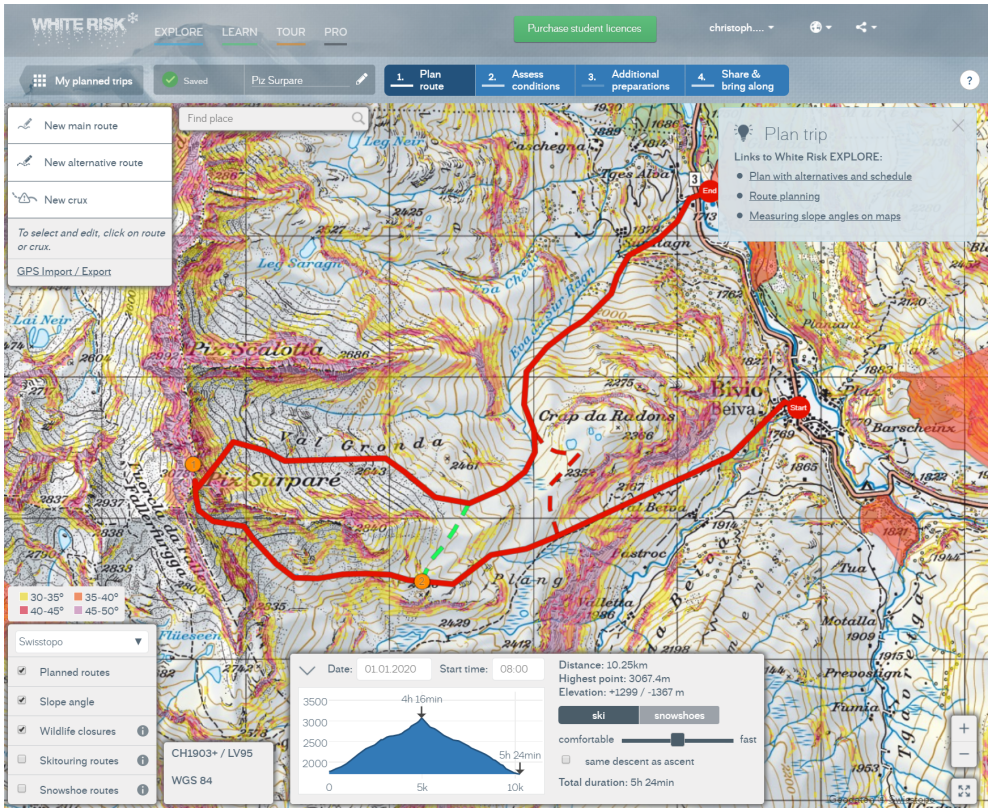


Figure 3.1: *White Risk: Main interface of trip-planning tool with example tour from (`www.whiterisk.ch/en/tour`.)*

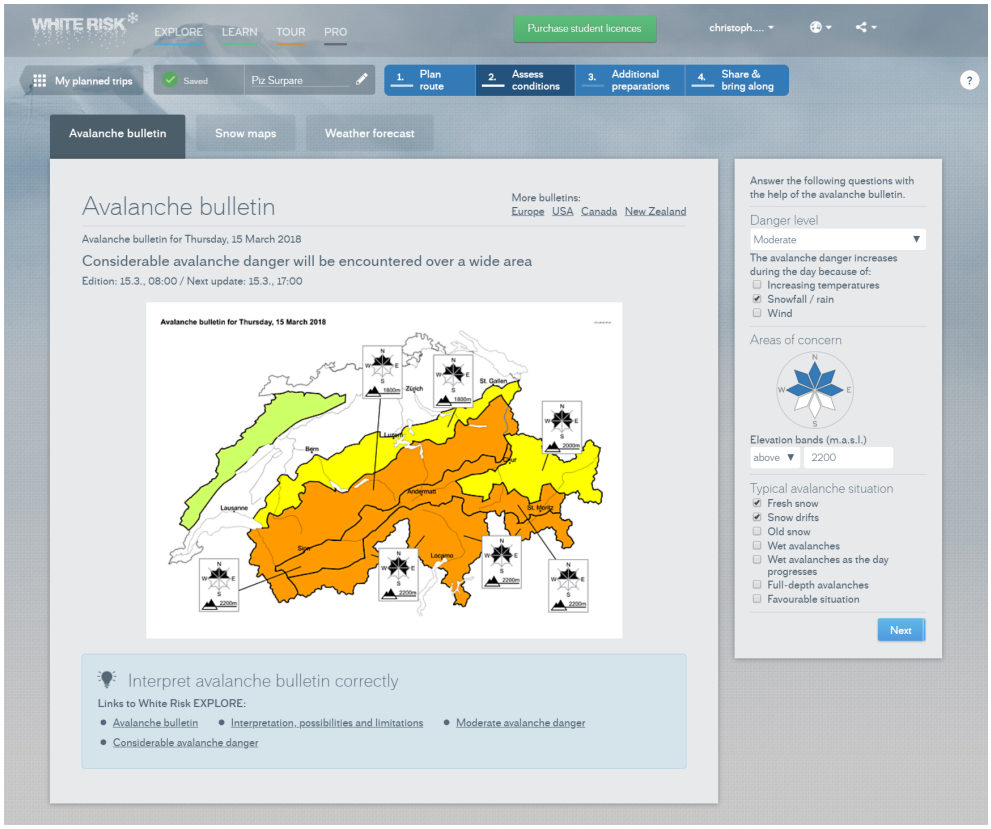


Figure 3.2: *White Risk: Assess conditions tab of trip-planning tool (`www.whiterisk.ch/en/tour`.)*

All data entered in White Risk is stored in a PostgreSQL database at the SLF. For this thesis, an anonymised extract of this database was provided by the SLF. This means that all tables including a reference to the user's identities were stripped from the database. The first extract was created on June 9^{th}, 2017. Hence, it included tours from four winter seasons spanning 2013/14 to 2016/17. On April 16^{th}, 2018, an update of the data was provided, which allowed for an analysis of the winter season 2017/2018 as well. Figure 3.3 shows all database tables that were available for this work, with the tables used marked green. Only attributes that serve as primary or foreign keys, as well as those used for the analysis are included in the diagram. A full database diagram can be found in the appendix (app. A, fig. A.1).

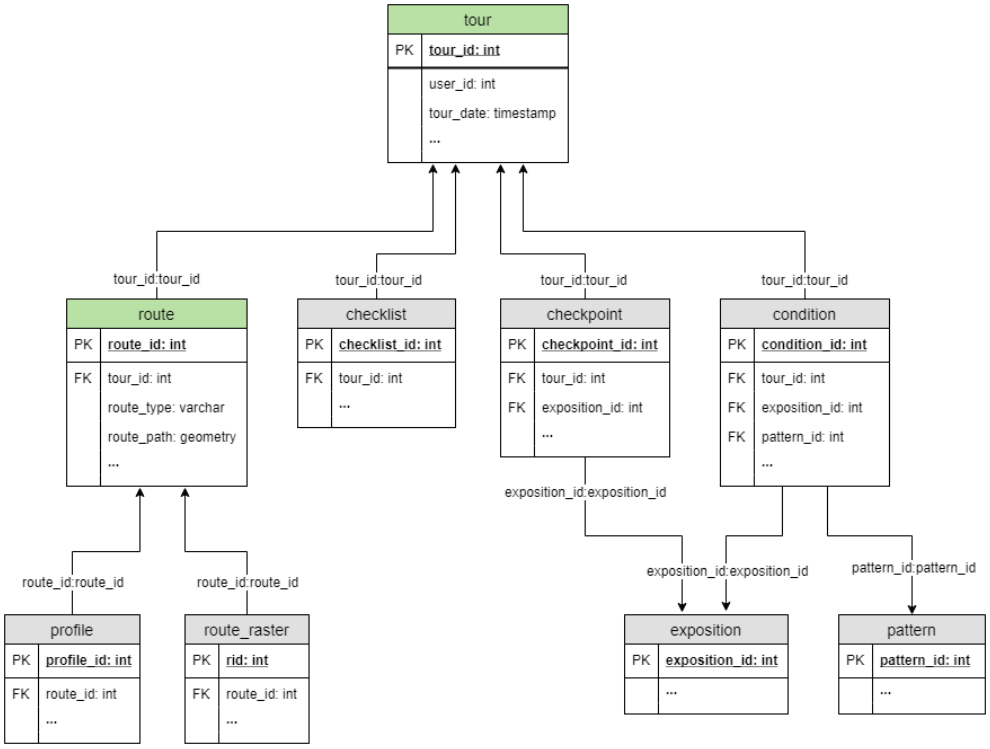


Figure 3.3: *White Risk Database Diagram: Tables used in this work are marked green.*

In total, there were 120'643 tour entries from 11'241 users. Every tour entry can consist of various routes and additional user-stored information about tour dates, checkpoints, conditions, etc. The routes can either be main or alternative routes. Users are limited to one main route per tour, but users can create multiple alternative routes. The recording of neither routes nor additional information is enforced and, therefore, a tour entry in the database can remain empty. This was the case for 32'767 tours. Automatically-generated tour details, such as distance, elevation plus/minus, peak elevation, etc., are only available if a main route is created. For this reason, this information was calculated rather than extracted form the database (see section 4.1). 151'754 routes were drawn in White Risk, of which 87'640 are main routes and 64'114 are alternative routes. These numbers are summarized in table 3.1.

Gudmundsson et al. (2011) defined a movement trajectory as "sequence of time-stamped locations $(x, y)_{T_1}, ..., (x, y)_{T_t}$, where $T_1, ..., T_t$ are t consecutive time steps". Contrasting to this, the planned route trajectories do not include a temporal component. Therefore, they can be defined as an ordered sequence of n locations $(x, y)_1, ...,$ $(x, y)_n$. If these locations are connected in order, they yield a polyline that can self-intersect, similar to movement trajectories (Gudmundsson et al., 2011). This further

Feature	Count
Users	11'241
Tours Total	120'643
Without Routes	32'767
Avg. Tours per User	10.73
Routes Total	151'754
Main	87'640
Alternative	64'114
Avg. Routes per User	13.5
Avg. Routes per Tour	1.73

Table 3.1: *Summary of full route dataset, as obtained from the SLF.*

implies that it is impossible to calculate movement parameters, which are based on the temporal component. As mentioned in chapter 2, the lack of a temporal component severely influences the range of available CMA algorithms. Thus, this inherent characteristic is of major importance to any further processing steps.

In White Risk, users draw routes in a continuous vector space without any restrictions. They are perceived from a Lagrangian perspective. Hence, after Laube (2009), the planned route trajectories are conceptually modelled in a Euclidean homogeneous space (cf. fig. 2.1), similar to GPS-tracks. The used reference system for the routes is the global WGS84.

It should be noted that White Risk is not a social platform intended to share content. Users generally plan routes for themselves and do not intend to share the route with the community. Therefore, in contrast to various existing works, the term 'user' will be used rather than 'contributor' for the remainder of this work. Nevertheless, the data represents a form of VGI data and, thus, several VGI-related issues, as were discussed in subsection 2.2.1, apply. As the full user table was not provided from the SLF, the total number of users is unknown. From the tour table we can obtain an ID list of all users which created at least one tour. There might be several users that have a White Risk account, but never planned a tour. However, as White Risk provides no general platform to share tours with the community, it can be assumed that these users mainly use other parts of White Risk than the trip-planning tool. Thus, users who did not plan any tour are negligible considering the participation inequality. In subsection 4.2.3, it will be shown how the influence of participation inequality was addressed and incorporated.
For planned backcountry tours, there is no clear definition regarding high-quality data. Some users might plan a tour roughly and adjust their route according to given conditions, etc. in the field, whereas others might plan a tour in great detail. Both tours fit the user's need and, therefore, they can be viewed as being of high quality. However, several users created routes for testing purposes. Often, these routes include some random locations and therefore exhibit rather atypical geometries. Hence, these routes can be objectively characterized as low-quality data. As users did not clearly specify the purpose of their tours, there is a need for an elaborate filtering process to dispose test routes. The filtering process will be covered in section 4.1.
In terms of credibility, it is of interest whether the planned tours could veritably represent backcountry ski tours. As mentioned in section 2.2.1, credibility is closely

linked to data quality. Thus, apparent test tours can be easily removed from the data. However, evaluating whether the remaining tours correspond to existing backcountry ski tours is difficult. In section 4.4, how planned tours are compared to recorded backcountry tours to verify credibility will be shown.
As mentioned earlier in this section, only user IDs for each tour were available for this thesis. Thus, no critical information was made available. However, by combining data from other sources with the planned backcountry tours, one might still be able to gather additional information about a user which might lead to a privacy breach. Considerations regarding privacy issues associated the data used will be examined in chapter 6.

3.1.2 Digital Elevation Models

In this thesis, two DEMs were used: DHM25 and SwissALTI3D. Both were provided by the Federal Office of Topography (swisstopo) and will be discussed in the following two paragraphs.

DHM25 DHM25 is a DEM of Switzerland that can be obtained as contour lines or as rasters with various spatial resolutions (25 m, 50 m, 100 m & 200 m cell size). It was derived from digitized contour lines of the "Landeskarte 1:25'000" (LK25), a national map of Switzerland with a 1:25'000 resolution. In this work, the raster with 25 m resolution was used. The mean accuracy ranges between $\pm$ 2.5 m and $\pm$ .5 m. DHM25 is based on the same reference system as the LK25: LV03. Since its creation, no updates were provided for DHM25 and, in 2014, it was superseded by SwissALTI3D (Swisstopo, 2005).
DHM25 was used for all computations that covered the full Swiss Alps area. Due to its comparatively low resolution, processing times could be reduced. The influence of the low resolution on the quality of the results will be discussed in section 5.2.1.

SwissALTI3D SwissALTI3D is a high-resolution DEM of Switzerland. It is distributed as rasters in multiple spatial resolutions (2 m, 5 m, & 10 m cell size). Airborne laser scans serve as a modelling base of the DEM in areas below 2'000 m.a.s.l. For areas above 2'000 m.a.s.l. elevation values are computed through stereo correlation. The accuracy is around $\pm$ 50 cm for areas below 2'000 m.a.s.l. and between $\pm$ 1 m and $\pm$ 3 m for areas above 2'000 m.a.s.l. SwissALTI3D is completely updated every 6 years. It is available in two reference systems: LV03 & LV95 (Swisstopo, 2014). SwissALTI3D with 5 m resolution was used for analyses on smaller study areas. As it is 25 times the size of DHM25, most computations require accordant processing times and power. The available computing resources were not able to handle such quantities of data for the whole Swiss Alps area.

DEM Derivatives Using ArcGIS Pro, four derivatives from DHM25 and SwissALTI3D were generated. All of the outputs are raster files. The following DEM derivatives were calculated:

Slope The slope raster contains a steepness value for each raster cell. Two methods are provided by ArcGIS Pro: planar and geodesic. As the input data is in a planar reference system, the planar method was selected. In ArcGIS Pro, the

slope values are calculated by finding the largest height difference between a cell and its eight direct neighbours [2].

Aspect The aspect raster stores values between 0° and 360°, indicating the compass direction a slope faces. Again, a planar and a geodesic method are available, of which the former was selected. Similar to the slope calculation, ArcGIS Pro uses a 3x3 window for the computation [3].

Profile & Plan Curvature Profile and plan curvatures are a secondary product of the curvature calculation. Curvature denotes the second derivative of the DEM, which is also calculated by using a 3x3 window. The profile curvature is the curvature in the direction of the maximum slope and the plan curvature is perpendicular thereto [4].

For the further processing steps, all rasters were projected to the WGS84 reference system. Furthermore, the rasters were clipped according to the study areas. To ensure that all routes were covered by the rasters, a 100 m buffer around all study areas was generated for the clips.

3.1.3 Avalanche Bulletin

The SLF publishes an avalanche bulletin twice a day during the winter season. It aims to inform and warn a target audience about the current avalanche danger in alpine areas. The target audience includes local authorities, inhabitants of alpine areas, and alpine recreationists, among others. It includes a map of all avalanche-affected regions in the Swiss Alps, Lichtenstein and Jura with their current avalanche danger. The avalanche danger is classified according to the European Avalanche Danger Scale[5]. During typical spring conditions, a double map is published describing the avalanche danger for dry avalanches in the morning and wet avalanches, typically occurring in the afternoon. The avalanche danger maps further include information about specific heights or expositions that are particularly risky. Textual information that accompanies the map provides additional reports on current dangerous avalanche patterns. Additionally, snow pack information, specific weather observations, and weather forecasts are included in the avalanche bulletin. Furthermore, snow maps for all warning regions are available and recommendations for transportation axes, settlements, as well as backcountry travel are provided. The current avalanche bulletin and additional information can be found on the website of the SLF[6].

3.1.4 Recorded GPS Tracks

To put the planned backcountry tours into context and compare them with real backcountry tours, GPS tracks from two mountaineering social media platforms were

[2] https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/how-slope-works.htm

[3] https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/how-aspect-works.htm

[4] https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/how-curvature-works.htm

[5] http://www.avalanches.org/eaws/en/main_layer.php?layer=basics&id=2

[6] https://www.slf.ch/en/avalanche-bulletin-and-snow-situation.html

gathered: Gipfelbuch.ch[7] and camptocamp.org[8]. Data from these two platforms has already served the analysis of backcountry touring activity in Switzerland (Techel et al., 2015). It should be noted that Techel et al. (2015) referred to Gipfelbuch.ch as bergportal.ch, which is the operator of gipfelbuch.ch. Techel et al. (2015) stated that all of Gipfelbuch.ch's users are German-speaking, whereas camptocamp.org is predominantly used by French- and Italian-speaking users. Although these platforms do not capture all backcountry skiing activity, together they may provide a more complete picture of backcountry skiing activity in Switzerland.

Gipfelbuch.ch Gipfelbuch.ch allows users to share conditions for hiking, climbing, backcountry skiing and other alpine activities. To create a condition entry, users first select an existing destination summit and route. If the desired summit or route does not exist, they can add new summits or routes to the list. Users can then specify the date of the tour and overall conditions, add descriptions, photos, videos and GPS-tracks. Most users add information about weather conditions, as well as photos. On May 3rd 2018, 36'838 condition entries for backcountry ski tours had been shared, of which 1'035 included a GPS-track. All reports were written in German.
The GPS-tracks on Gipfelbuch.ch were not stored in a spatial database, which would have allowed a spatial query for their extraction. Thus, all GPS-tracks needed to be extracted manually from the website. To reduce the amount of tours which needed to be checked for GPS-tracks, only tours to summits in the three small study areas, as defined in subsection 3.2.1, were searched. To obtain a list of these summits, the swissTLM3D dataset from swisstopo was used (Swisstopo, 2015). SwissTLM3D is a large-scale topographical landscape model of Switzerland, which includes a point layer with the names of all Points of Interest (POI). Using ArcGIS Pro, this layer was clipped with the small study areas. This resulted in 182 POI, including hills, summits, mountain passes, etc. in these study areas. The names of these POI were then exported as CSV files and used for the manual route search on Gipfelbuch.ch. Gipfelbuch.ch provides a search tool for all stored mountain summits. For each existing summit, a description and a list of tours is available. From this tour list, the tours with GPS-tracks were accessed and the tracks were downloaded as GPX-files. In total, 99 GPS-tracks were retrieved from Gipfelbuch.ch. However, one GPX-file had to be removed, as it was in an unreadable format and could not be processed. The number of tours for each study area is summarized in table 3.2.
As the GPX-file format does not enforce the inclusion of additional data, such as timestamps, elevation, etc., the amount of additional information in the retrieved files is inconsistent. The amount of additional information included in the GPX-file depends on the device used, as well as on the device settings. Furthermore, the GPS-tracks dataset also included several digitized SAC-tours, which did not include any timestamps. Hence, timestamps were completely removed from all files. The digitization was realised by Gipfelbuch.ch and skitourenguru.ch[9], whom were stored as authors in the GPX-files. The GPX-file format is rather lax. This may lead to conflicts when reading the files in R. In three files, tags had to be adjusted from self-closing to normal tags, in order to make them readable.

Camptocamp On Camptocamp users can also share tour reports for alpine activities, similar to Gipfelbuch.ch. Besides general descriptions, information about con-

[7] `https://www.gipfelbuch.ch/`
[8] `https://www.camptocamp.org/`
[9] `http://skitourenguru.ch/`

ditions and photos, users can also upload GPS-tracks or draw their route on a map interface. A filter option by country was not available, and only a spatial filter could be applied using a map extent. As of May 3rd 2018, approximately 80'000 backcountry touring reports were stored for the Central-European Alps. How many of those included a GPS-track is unknown. About 80 % of the reports were written in French, 18 % in Italian and less than 2 % in other languages (German, English, etc.).
The operators of Camptocamp provided a shapefile of all recorded GPS-tracks and drawn routes in Switzerland. A total of 2'357 routes were provided. The shapefile only contained a geometry and a tour ID. Timestamps for the routes were not available. In ArcGIS Pro, all routes which intersected with the small study areas were then extracted. One route had to be modified manually, as it included various line features in the city of Milan. These lines were removed, so that only lines intersecting with the study area remained. Finally, three shapefiles for the lines in each study area were created and processed with R. A summary of the total numbers of routes from Camptocamp can be found in table 3.2.
It is likely that the dataset from Camptocamp included drawn or digitized routes as well. However, no additional information was available to provide any hints. Thus, no differentiation was conducted.

	Number of GPS-Tracks		
Study Area	**Gipfelbuch.ch**	**Camptocamp**	**Total**
Wildstrubel	15	44	59
Urserental	55	71	126
St. Antönien	28	1	29
Total	98	116	214

Table 3.2: *Number of GPS-tracks retrieved from Gipfelbuch.ch and Camptocamp.*

After loading both datasets into R, the files from the two sources were combined. This resulted in three spatial line objects, each containing the GPS-tracks in one small study area. Known digitized routes were marked as such in the spatial object.

3.2 Study Area

As White Risk was developed by a Swiss research institute (SLF), it can be expected that Swiss users make up the largest proportion of users. Most users will mainly plan tours in their surroundings or in mountainous areas they are familiar with. Swiss natural hazard authorities divided Switzerland into 137 warning regions, based on climatological criteria (SLF, 2018b). Additionally, Lichtenstein is included as a warning region. A list of all warning regions including their names can be found in appendix B. For 129 of these 138 regions, the SLF publishes an avalanche bulletin twice a day during the winter months and once a day during the early and late winter season. As backcountry ski tours usually take place in alpine areas, it was decided to focus on areas for which an avalanche bulletin is published. Hence, the Swiss Alps were selected as a study area for large-scale analyses. The Jura Mountains were excluded, even though they are covered by the avalanche bulletin. By excluding the Jura, a contiguous study area could be obtained. 117 warning regions were selected as the study area for the large-scale analyses in this thesis (cf. figure 3.4). This gave a study area of 26'371.4 km^2.

Figure 3.4: *Study area for large-scale analyses, marked in light orange.*

3.2.1 Small Study Areas

To reduce the computational load for further analysis steps, three smaller study areas were selected. For these study areas, the number of routes is drastically reduced, which enhances the speed of computational analyses. Furthermore, the higher resolution DEM, SwissALTI3D, could be used in these areas.
All three small study areas cover popular backcountry skiing areas in the Swiss Alps. Thus, it can be expected that a reasonable number of tours are located within these study areas. To reduce the effect of any geographic correlations, three areas widely distributed across the Swiss Alps were selected. Figure 3.5 shows the locations of the three small study areas.

Wildstrubel Wildstrubel is a massif in the cantons of Valais and Bern, consisting of various mountain peaks and the Wildstrubel glacier. The Wildstrubel study area covers six warning regions with an area of 144 km^2.

Urserental The Urserental lies in the canton of Uri, bordering the cantons of Valais and Ticino. It covers five warning regions and has an area of 144 km^2 as well.

St. Antönien St. Antönien is a small municipality in the canton of Grisons, which is surrounded by several mountains. The study area only covers one warning region and has an area of 100 km^2.

3.3 Conclusion

The definition of planned route trajectories highlighted the major difference to 'real' movement trajectories: the lack of a temporal component. The lack of this temporal component prevents the calculation of various movement parameters and directly influences the CMA methods, which are applicable to the data. Nonetheless, the geometry of planned route trajectories resembles the geometry of 'real' movement

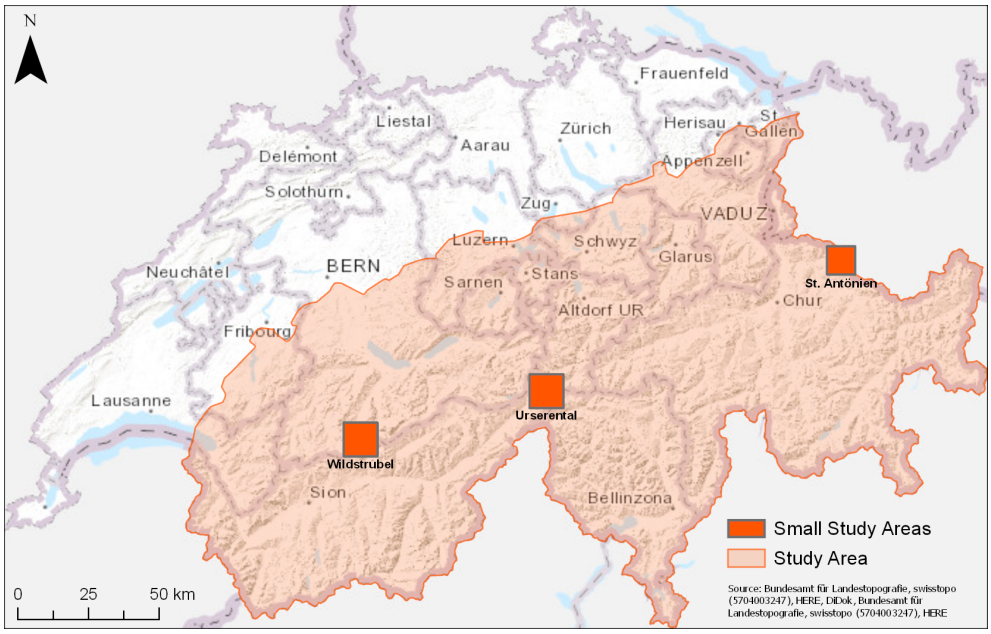


Figure 3.5: *Study areas for small-scale analyses, marked in orange with grey boundaries.*

trajectories to a large degree. Hence, methods relying on geometry should be applied without any issues. Additional contextual information, which is not reliant on a temporal component, should be similar for planned and 'real' movement trajectories. Thus, the inclusion of such contextual information might enable the application of additional methods.

As was discussed, the planned route trajectory dataset is likely affected by typical VGI-issues. These include participation inequality, data quality and credibility. Such issues need to be considered and their influences on the eventual results need to be assessed.

Two rasters with varying resolutions were presented. Different raster resolutions might lead to differences in contextual information, extracted for the planned route trajectories. Thus, it is necessary to examine the influence of the raster resolution on these results.

The avalanche bulletin can be seen as a form of authoritative avalanche information. A comparison of user-stored avalanche danger levels and avalanche danger levels stored by the SLF might provide insights into the accuracy of the planned routes. Similarly, GPS-tracks from other sources might also be used to evaluate the accuracy of planned route trajectories. High similarity between planned routes and GPS-tracks might serve as an indicator of the degree of accuracy of planned routes. However, the number of GPS-tracks was small, which might impair the validity of these results.

Chapter 4

Methods

In this chapter, the methods applied to the data, presented in chapter 3, will be illustrated. This should allow for a full understanding of how the results, as presented in chapter 5, were obtained. Furthermore, maximum reproducibility should be reached. Therefore, for each processing step, the methods applied will be described. Furthermore, parameter settings will be considered where necessary.
This chapter is subdivided into five sections. The first section will cover all steps taken during the preprocessing phase (sec. 4.1), including filtering and data enrichment. Then, a description of how a general overview of the dataset was generated will be provided in terms of temporal & spatial distribution, participation inequality and avalanche conditions (sec. 4.2). Section 4.3 will illustrate how values, extracted from DHM25, were processed and how the influences of raster resolution, participation inequality and avalanche conditions on the route attributes were assessed. Section 4.4 will address which similarity and clustering approaches were applied to the dataset. Finally, the computing environment and software utilized in this work will be described in section 4.5.

4.1 Preprocessing

The raw data, as it was obtained from the SLF, was rather messy. There were duplicate routes, test routes drawn by users and routes for assumingly other purposes, such as cycling or drone routes. Hence, preprocessing of the data prior to conducting the main analyses was needed. This section aims to put forward a possible approach for distinguishing test routes from 'real' planned routes. This builds the groundwork for all subsequent processing steps. The filtering of undesired routes played a crucial role in preprocessing. As users did not mark their test routes as such, all filtering criteria were based on the route geometry or other route characteristics.

To gain additional contextual information, the planned routes were enriched with elevation data from DHM25, derivatives thereof, as well as additional data recorded by users. Figure 4.1 provides an overview of the preprocessing pipeline, through which the data was run. These steps will be discussed in detail in the following subsections. It should be noted that all later analyses were based on the filtered dataset as described in this section.

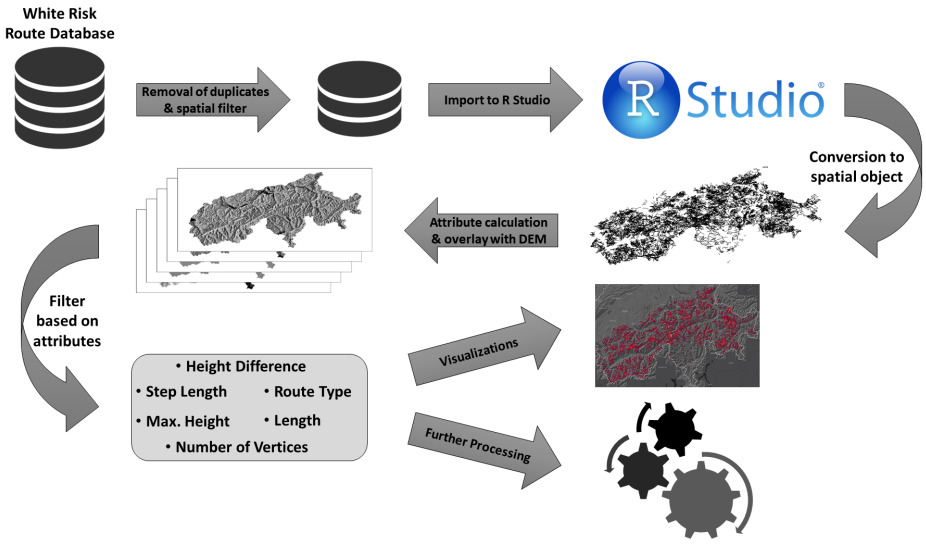


Figure 4.1: *Preprocessing pipeline including various filtering steps and raster overlay that was applied on the planned routes.*

4.1.1 Filtering on Database Level

With the data being stored in a relational database, it was possible to filter the routes before loading them into the analysis environment. On this level, duplicate routes were removed and a spatial filter was applied. Duplicate routes can be introduced to White Risk when users share tours. When a user opens the link of a shared tour, a copy of this tour is created which can be modified by the user without affecting the original version. If the copy is not changed, two identical tours are stored in the database, leading to two entries sharing the same route geometry. Despite this functionality being useful for users, it leads to undesired effects during the analysis of routes. If a user planning a trip shares a tour with all participants, this may lead to a high number of duplicates for a single tour. Even though a group went on a tour together, it may seem like various users planned the same tour independently of each other. This will certainly lead to an overrepresentation of routes from those tours and, therefore, the removal of duplicate routes was necessary.
In the database, a new view was created which only included a single entry for duplicate routes. This was achieved by using the `ROW_NUMBER()` command on the ordered route geometries. Geometries which are duplicates were assigned row numbers greater than 1. Therefore, the selection of entries with a row number of 1 resulted in a set of unique geometries.

The application of a spatial filter allowed extracting routes within the large study area. A large proportion of routes outside the study area seemed to be test routes. It is likely that users draw test routes either in areas they are familiar with or some random location, disregarding its suitability for backcountry skiing. Figure 4.2 shows planned routes in Zurich and Corsica. Clearly, in a city there is no possibility for backcountry skiing. In Corsica, ski touring is possible. However, the map shows a route that looks like a drone route.
For the spatial filtering, a dissolved polygon of the chosen warning regions was loaded into the database. A new view was created based on the previously-generated view, for which duplicates were removed. A spatial query was used to dismiss routes outside of the study area.

Finally, before moving the data to the processing environment, the user IDs were

attached to the routes by joining the tours and routes by the tour ID.

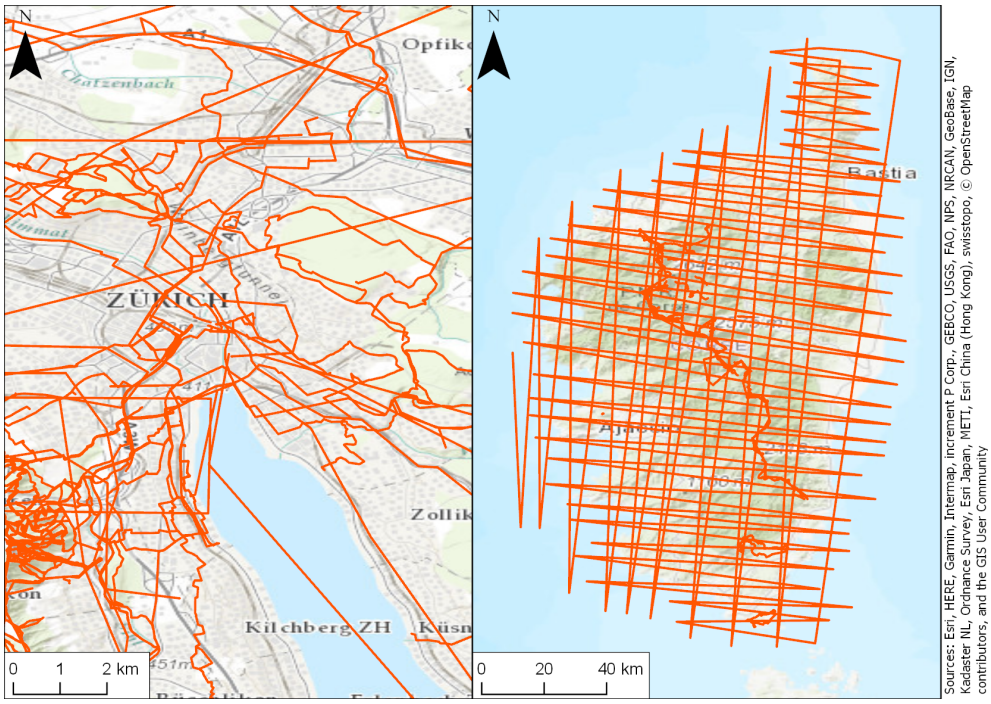


Figure 4.2: *Maps of example test routes, illustrating the need for spatial filtering. Routes in the city of Zurich (left) and Corsica (right).*

4.1.2 Attribute Calculation & Raster Overlay

After the first filtering stage, the data was loaded into R Studio. The routes were converted to spatial objects and several attributes were calculated, such as route length, number of vertices, average step length, minimum & maximum step length and step length deviation. For these calculations, the packages 'sp' and 'geosphere' were used (Bivand et al., 2013; Hijmans, 2017a). A full list of used R packages can be found in subsection 4.5.
The DHM25 raster was loaded into ArcGIS Pro, where the derivatives slope, profile- & plan curvature and aspect were calculated. For slope and aspect calculations, the planar method was selected, as the input dataset had a planar reference system. The Z factor for the curvature calculation was set to 1. These rasters were then projected to WGS84, clipped with the study area and loaded into R Studio. In R, a special raster extraction library called 'velox' was used to extract raster values for all routes (Hunziker, 2017). This library proved computationally much more efficient than the widely-known 'raster' library (Hijmans, 2017b). However, it was not able to return the extracted values ordered along the input line. For the extraction, all rasters were first converted to a velox object. Several subsets of each raster were generated to decrease the memory load during the extraction. Subsequently, for each route, minimum, maximum and mean height, slope, plan and profile curvature were extracted. Additionally, the height difference ($maximum - minimum$) was calculated for each route. All aspect values extracted from the raster were stored as a list for each route. Additionally, the headings of all routes and the route straightness were calculated. The headings were calculated for each segment (two subsequent locations) of the routes and then stored with the routes as a list. The straightness was calculated according to the straightness index formula of Batschelet (1981 in Benhamou, 2004,

pp. 211):

$$straightness = 1 - \frac{distance_{start\ to\ end}}{route\ length} \tag{4.1}$$

This gives a straightness index of 0 for perfectly straight lines ($distance_{start\ to\ end}$ is equal to *route length*) and 1 for a round trip, with start and end points being the same ($distance_{start\ to\ end}$ is 0). Finally, the user-specified date of the tour and the danger level were extracted from the tour table in the database and joined with the routes by the tour ID.

Table 4.1 shows a list of all attributes that were stored with the routes after the enrichment step. Attributes that were used in the ensuing filtering process are highlighted in gray.

Table 4.1: *List of route attributes after enrichment. Attributes used for filtering are highlighted in gray.*

Variable	Description	Unit
route_id	Unique ID of route	
tour_id	Unique ID of tour	
user_id	Unique ID of user who created tour	
route_type	"MAIN" or "ALTERNATIVE" route	
length	Route length	[m]
numberOfVertices	Number of vertices	[m]
stepLength	Average step length	[m]
stepLengthDeviation	Standard deviation of step length	[m]
stepLengthMin	Minimal step length	[m]
stepLengthMax	Maximal step length	[m]
minHeight	Minimum height	[m.a.s.l.]
meanHeight	Mean height	[m.a.s.l.]
maxHeight	Maximum height	[m.a.s.l.]
heightDifference	Maximum height difference	[m]
minSlope	Minimum slope	[°]
meanSlope	Mean slope	[°]
maxSlope	Maximum slope	[°]
aspect	List of aspect values	[°]
minPlanCurv	Minimum plan curvature	[1/100m]
meanPlanCurv	Mean plan curvature	[1/100m]
maxPlanCurv	Maximum plan curvature	[1/100m]
minProfCurv	Minimum profile curvature	[1/100m]
meanProfCurv	Mean profile curvature	[1/100m]
maxProfCurv	Maximum profile curvature	[1/100m]
headings	List of heading values	[°]
straightness	Straightness Index	
tour_date	User-specified date of tour	
danger_level	User-specified avalanche danger level	

4.1.3 Filtering based on Attributes

The filtering of test routes can be categorized as a classification problem. Routes will either be classified as test routes or as assumingly 'real' backcountry ski routes. Thus, the strictness of the filters influences precision and recall of this classification

as defined by Olson and Delen (2008, p. 138). In this case, false positives are routes, classified as ski route, which really are test routes. False negatives are routes, classified as test routes, which are 'real' routes. If the filtering criteria are strict, a large share of test routes can be removed. Hence, most remaining routes are indeed 'real' routes, leading to a higher precision. However, some 'real' routes might be classified as test routes and consequently removed. These false negatives will be lost, leading to a lower recall.

Six attributes were selected to filter test routes (cf. table 4.1). In this subsection, these attributes will be discussed in detail and some examples will be provided that underline their significance. The attributes relating to step length (stepLengthDeviation, stepLengthMin & stepLengthMax) are strongly correlated with the actual route step length. Thus, they did not provide any additional benefit and were not considered for the filtering. The remaining attributes are mostly resulting from the DEM extraction. Thus, they do not represent inherent route characteristics but rather provide additional information for the routes and their context. Their suitability for filtering purposes was evaluated by visualizing routes with extreme attribute values. However, none of them exhibited conspicuous anomalies.

For the classification, a simple heuristic approach was chosen. Thresholds for all selected filtering attributes were generated, based on relatively simplistic route statistics and visual assessments. The selection of the thresholds was discussed with two experts from the SLF (Stephan Harvey & Frank Techel). The combination of several approaches should allow for a sensible threshold selection. Other classification approaches, such as machine learning, would have been conceivable. However, such approaches would have complicated the filtering process and were thus neglected. Nonetheless, other elaborate classification approaches should also give high quality results and should be evaluated in future research concerning this dataset.

Route Type As mentioned in section 3.1, users can draw main and alternative routes. Alternative routes provide users with additional options during a tour in case of deteriorating weather conditions or other incidents. However, users are not forced to plan alternative routes. As the alternative routes often only depict partial routes, it was decided to focus on main routes in this thesis. Hence, all routes of the type 'alternative' were discarded in further analysis steps.

Route Length Filtering based on route length allowed removing routes which were overly short or long. Besides several routes with length 0, many routes had a length of only a few hundred meters. Such short routes could be completed within a few minutes. Therefore, it is assumed that these were not 'real' routes. Figure 4.3.1 & 4.3.2 illustrate two relatively short example routes. The map context, showing small residential areas, underlines the justification of their removal. All routes shorter than 1'683.944 m, which corresponds to the 2.5 % length quantile of all main routes in the study area, were removed.
Furthermore, there were many routes of excessive length. Besides the fact that these routes would suggest extensive touring time, they often showed relatively simple geometries. They often did not follow any environmental features as can be seen in figure 4.3.3. Nevertheless, long routes can also include multi-day backcountry tours. Figure 4.3.4 shows a route exhibiting a reasonable geometry but has a length of 65 km. This length is clearly too long for a day trip. However, the route passes various alpine cabins and, therefore, it can be assumed that it is a multi-day trip.

Combined with alpine cabin data, such routes could have been segmented into day trips. Yet, as only few routes exhibited similar characteristics, this segmentation was omitted. Thus, all routes longer than 21'008.91 m, which corresponds to the 95 % length quantile of all main routes, were eliminated.

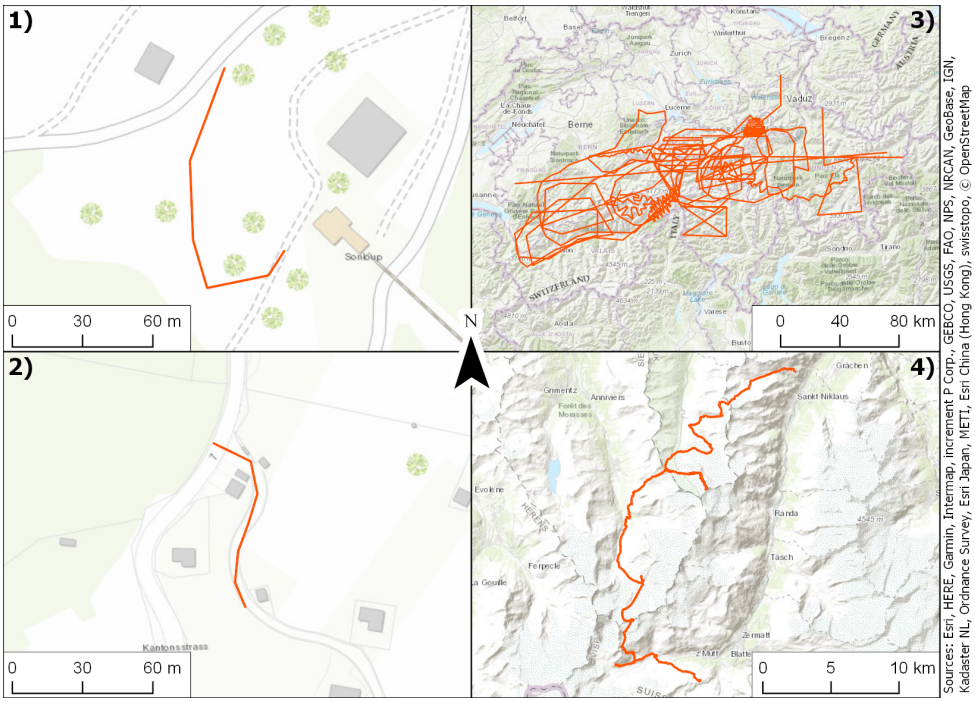


Figure 4.3: *Visualization of routes affected by filtering based on length. 1) & 2) short routes with 136.4 m and 82.3 m length, respectively. 3) Various routes with lengths between 137 km and 266 km. 4) Multi-day backcountry tour with 65 km length.*

Number of Vertices Several routes only included two vertices and, therefore, embodied a simple line. It is clear that these routes were not 'real' routes, as they were too simple. Even with 4 or 5 vertices, most routes are rather inaccurate. Figure 4.4 shows a selection of routes with 3 to 5 vertices in the area of Urserental. It is apparent that many routes do not follow any environmental features but consist of random points. Therefore, these routes can be classified as test routes. Hence, all routes with 5 or less vertices were filtered from the dataset.

Maximum Height As planned backcountry routes are expected to be located in alpine areas, routes were filtered based on their maximum height. This maximum height likely corresponds with the peak elevation of a backcountry route. Routes with a maximum height of less than 1'000 m.a.s.l. are unlikely, as such elevations do not usually have sufficient snow and such areas are not fit for backcountry skiing. Hence, all routes with a maximum height of less than 1'000 m.a.s.l. were removed from the dataset.

Height Difference For backcountry recreationists to enjoy skiing, there should be some height difference during the tour. If there were only small height differences, a tour would mainly consist of walking slopes without considerable rise or fall. To select an appropriate threshold, routes with small height differences were visualized. As can be seen in figure 4.5, there were various problematic routes. 1) shows a route that was drawn around an airport in Saanen. It is clear that this was not a proper

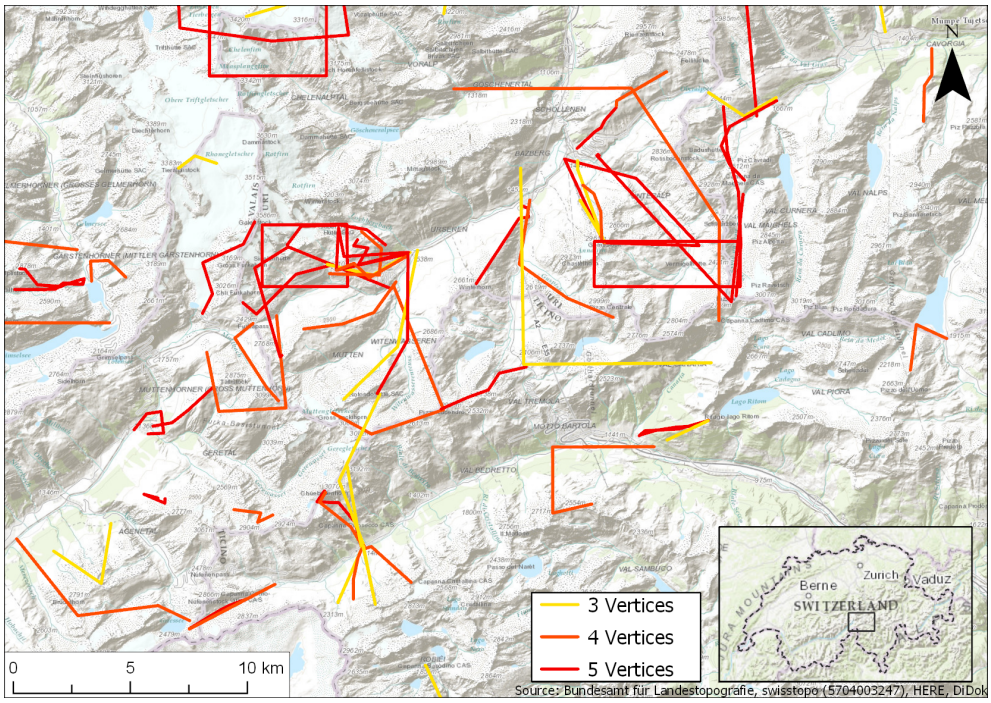


Figure 4.4: *Visualization of routes around Stotzigen Firsten affected by filtering based on number of vertices.*

backcountry ski tour. 2) illustrates a planned route around Lago Ritom. The height difference was only about 220 m and most of the route followed the lakeside. In 3) a route in the form of a fir is displayed, which assumingly is a test route. 4) shows several routes which likely were proper backcountry tours. Most of them were between Gemmipass and Lämmerenhütte, an alpine cabin. As these routes should be retained, the 0.5 % height difference quantile of the remaining routes (183 m) was selected as a threshold. Even though probable test routes, such as the one depicted in figure 4.5 - 3, are still included in the route dataset, this threshold denotes a reasonable trade-off.

Average Step Length The average step length of a route is an indicator of how detailed a user's planned route was. Routes with a large average step length often exhibit major anomalies, which is why this attribute was selected for the filtering process. There might be routes with long monotone ascents, for example on a glacier (cf. fig. 4.6.1), that do not need very detailed planning. However, most routes with long step lengths showed clear signs of test routes, such as artificial forms or unreasonable route behaviour (fig. 4.6.2). The 98 % quantile of the remaining routes (881.82 m) was selected as a threshold for step length filtering.

4.1.4 Route Subsets in Small Study Areas

For the computationally-expensive analyses, subsets for all three small study areas, as described in section 3.2.1, were created. In a first step, the shapefile of the three small study areas was uploaded to the database containing the full route dataset. Then, a view for each study area was created. These views included all routes that were completely confined to a study area. This allowed the extraction of route IDs from routes in the study areas. From the final filtered routes, as described in the

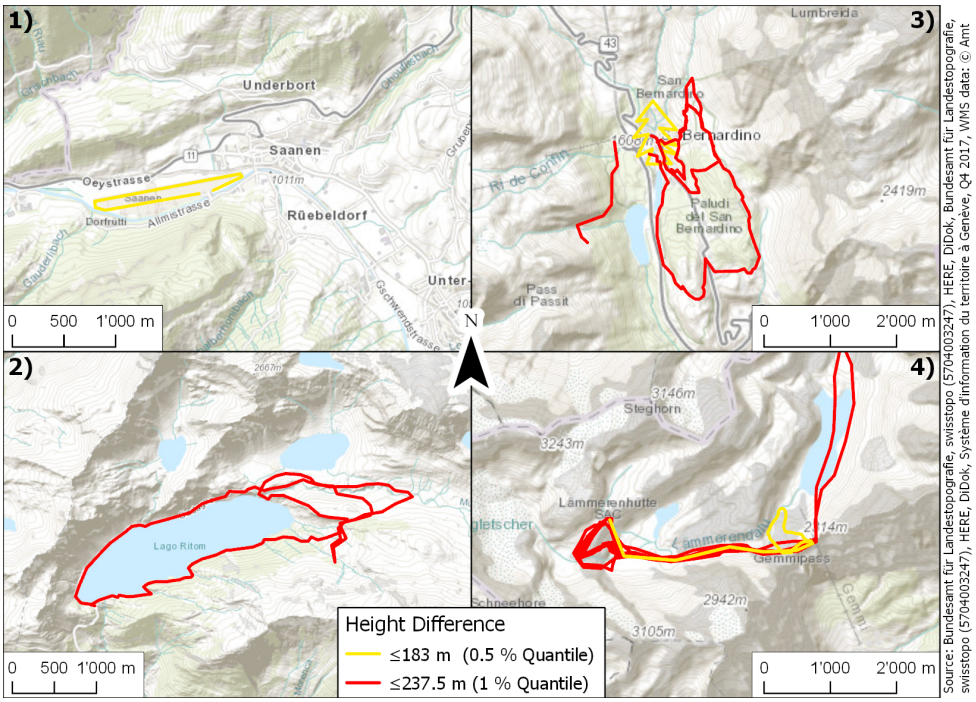


Figure 4.5: *Visualisation of routes with small height difference. 1) Presumable test route around airport in Saanen. 2) Route around lake. 3) Route in the form of a Christmas tree in San Bernardino. 4) Probable 'real' routes from Gemmipass to Lämmerenhütte.*

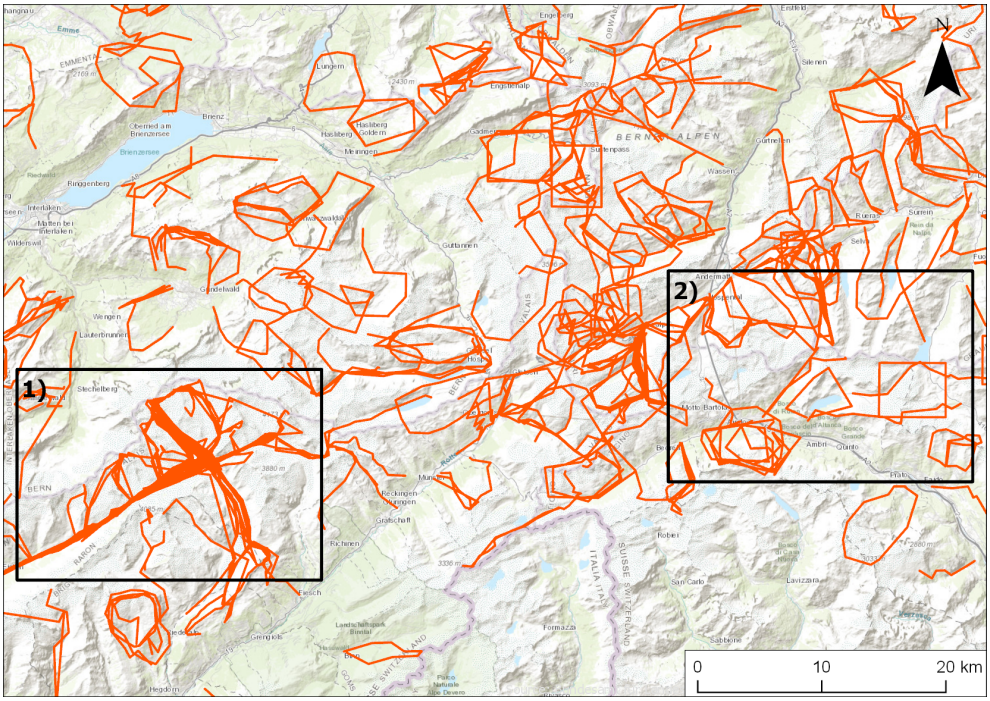


Figure 4.6: *Visualisation of routes with step lengths greater than 881.8233 m. 1) Potential 'real' routes on glacier with monotone ascent. 2) Probable test routes that do not follow any environmental features or show geometric forms, such as squares.*

previous subsections, the routes whose IDs were in this list of route IDs were extracted and stored as new objects. This process could have been achieved relatively easily in R. However, the retrieval of route IDs from the database proved to be more efficient.

4.2 General Overview

As the planned routes from White Risk have never been analysed to date, a general overview of the dataset is necessary. This should provide an idea of the temporal and spatial distribution of the dataset, existing participation inequality and additional data stored with the routes, such as avalanche conditions. Through this general overview, the dataset will be more comparable to existing work on backcountry ski touring activity and thus allow gathering initial indications of the degree of accuracy of planned routes in the dataset. Furthermore, a comparison of user-stored avalanche conditions with authoritative data will show how well user-specified data corresponds to empirically-gathered data.

4.2.1 Temporal Distribution

To analyse the user-stored tour date values, a subset was created. Only routes that included tour dates were considered for the subsequent analysis. Furthermore, the tour date attributes needed some prior data cleaning, as some users stored only the last two digits of the year. For these cases, 2000 years were added to the tour date. Only dates during the winter season should be included in the analysis. There is no fixed date from or to which backcountry skiing is possible. Thus, for all relevant winter seasons, the days of the first and the last avalanche bulletin were extracted from the avalanche bulletin database of the SLF. As the avalanche bulletin is only published if there is enough snow, these dates can be used as an indicator for the beginning and end of the winter season. Consequently, all routes were assigned to a season according to their tour date. Routes with tour dates between seasons were not considered for the consequent visualizations. Several plots were generated to visualize the route distribution by season, month and weekday.

4.2.2 Spatial Distribution

To visualize the spatial distribution of the filtered routes, the line density was calculated using ArcGIS Pro. First, the routes were transformed from WGS84 to LV95 in R and exported as a shapefile. The transformation allowed using metric parameters for the line density calculation. The tool works as follows: For each centroid of the output raster cells, a buffer is created and overlayed with the input lines. The length of all lines within the buffer are then calculated and divided by the area of the buffer. This value is then assigned to the raster cell. More information on ArcGIS Pro's line density tool can be found online[1]. Two buffer radii of 50 m and 250 m were selected. The cell size of the output raster was set to 25 m and the area units to square meters. Two different search radii were chosen as it eases visualization in varying scales.

4.2.3 Participation Inequality

In order to determine the presence of participation inequality, general user statistics of the dataset were extracted and visualized. By doing so, it was easy to obtain a

[1] `https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/line-density.htm`

general estimate of the impact of a few heavy users. To analyse the impact of participation inequality, three subsets of the full dataset were generated. There was a subset of rare users, heavy users and a random sample of the full dataset. All subsets included 10 % of the routes from the full dataset. The rare user sample included 10 % of the routes which were created by users who planned the fewest routes. The heavy user sample included 10 % of the routes which were created by users who planned the most routes. For the random sample, a single sample from the full dataset was drawn in R with a fixed seed of 1. The definition of a seed ensured reproducibility. Similar to the full dataset, general user statistics were extracted and visualized. The three subsets were further used to assess the existence of a geographic bias and for the evaluation of differences in route attributes between rare and heavy user routes.

Techel et al. (2015) noted that there exists a geographic bias in backcountry ski datasets from social media mountaineering websites. Therefore, it is likely that a geographic bias is also present in the similar dataset used in this work. Two χ-maps (Chi-maps) were generated to evaluate a potential geographic bias. In a first step, the three route subsets were converted from WGS84 to LV95 and exported from R as shapefiles. These three shapefiles were loaded into ArcGIS Pro, where two density maps were generated for each shapefile. The search radii for the density maps were 50 m and 250 m, respectively. The cell size for all density maps was set to 25 m. The density map of the random sample was generated first. The density maps for the heavy and rare users were generated in a second step. For these density maps, the extent was matched to the density map of the random sample. Furthermore, they were snapped to the density map of the random sample so that all cells were aligned. This is a prerequisite for the calculation of χ-values. All six generated density maps were then loaded into R. In R, the χ-maps were calculated, using the following formula:

$$\chi = \frac{(obs + 1) - (exp + 1)}{\sqrt{exp + 1}} \tag{4.2}$$

with the density maps of the rare and heavy user samples as 'obs' and the density map of the random sample as 'exp'. In both rasters, several cells had a value of zero. This was the case for areas in which no backcountry touring activity was present. Thus, one was added to all rasters to prevent a division by zero.

To obtain a general idea of spatial differences between the routes of the two subsets, a relatively simple measure was calculated: the mean centers of routes. In order to do so, the routes from both datasets were converted to points, using the 'Feature To Point' tool in ArcGIS Pro. This tool allows extracting centroids from line objects. For the line centroids of the two respective subsets, the mean center was calculated using the according tool 'Mean Center' in ArcGIS. All optional parameters were neglected. Comparing the coordinates of the two mean centers and visualizing them on the respective χ-maps should enhance understanding of an eventual geographic bias.

4.2.4 Avalanche Conditions

For analysing avalanche conditions, only routes which included a user-specified danger level were considered. A histogram was generated to visualize the attribute

distribution and provide a general idea of the prevalent danger levels in the dataset. The user-specified avalanche danger was then compared to authoritative avalanche danger data from the SLF to assess its quality. The avalanche danger extraction from the authoritative data was only possible for routes that included a tour date. Thus, only routes that included a user-specified avalanche danger and tour date were considered for the comparison. The avalanche bulletin published by the SLF, as described in section 3.1.3, was used as authoritative avalanche danger data.
In order to accomplish a comparison, each route had to be matched with the accordant avalanche bulletin from the SLF. For each route, the start and end points were extracted. An overlay of these points with the SLF warning regions allowed to extract one or two warning regions in which each route was located. For each route, the route id, the ids of the start and end warning regions, as well as the user-specified tour date were exported. Based on the warning regions and tour dates, this dataset was then joined with avalanche bulletins, retrieved from the database of the SLF. For each route, an avalanche danger rating for start and end point, as well as expositions and heights of avalanche-prone locations were returned. As users specify the avalanche danger during the planning phase of the tour, the danger levels of the evening bulletin from the day before the specified tour date were extracted. The evening bulletin of the day before the tour is relevant for tour planning. If a double map was published for a tour date, avalanche danger ratings for both conditions, wet and dry, were returned.
The subset was split into two groups: routes with a single authoritative avalanche danger level and routes with authoritative avalanche danger levels for two conditions. Only the routes with a single authoritative danger level were considered for the subsequent tests. As it is unknown for which times the routes were planned, it is not possible to select either the morning or the afternoon bulletin for a comparison. In a first comparison, whether there are differences between danger levels at the start and end point of the full subset was assessed. Then, user-stored danger levels were compared with authoritative danger levels. As attributes of the same routes were compared with each other, a pair-wise test was necessary. The attributes did not follow a normal distribution. Thus, the Wilcoxon signed-rank test was used for comparisons between user-specified data and authoritative data. With growing sample sizes, p-values tend to get smaller (Wasserstein and Lazar, 2016). Thus, it is crucial that effect sizes are calculated in addition to the p-values when large samples are used for statistical tests. These are able to provide a clearer picture of existing effects. As proposed by Fritz et al. (2012a, p. 12), Cohen's R was used as effect size for the Wilcoxon signed-rank test (Cohen, 1988). It can be calculated with the following formula:

$$r = \frac{Z}{\sqrt{N}} \tag{4.3}$$

As the avalanche attributes are on an ordinal scale level, this was the only effect measure available. To visualize the distributions of avalanche danger levels from various sources, as well as differences between the values from those sources, barplots were created.

4.3 Route Attributes

The calculation of route attributes allows obtaining a context for the routes. This context might provide further insights into the planning process of users, extend-

ing the general overview. Furthermore, the evaluation of various influences on the route attributes should allow the identification of factors which influence the user's planning process.

For all calculated attributes, as well as for most attributes extracted from DHM25, median, mean and standard deviations were calculated in R. Furthermore, boxplots were generated to visualize the distribution of the attribute values. For aspect and heading values, it is meaningless to calculate summary statistics and generate boxplots as they are nominal attributes. These attributes were reclassified in eight compass directions (N, NW, W, etc.). Then, the frequency for each direction was calculated for all routes. To obtain an understanding of the full dataset, the frequencies of all routes were summed and normalized by dividing by the maximum frequency sum. The normalized values were then plotted in pie plots. For aspect values, two plots were generated. In the second plot, the aspect values were additionally normalized by the number of pixels of each compass direction in the aspect raster. The reason for this normalization is that, in the Swiss Alps, which show a general southwest to north-east distribution, north-west and south-east values are expected to show higher frequencies. Only raster values which are on an elevation higher than 1'000 m.a.s.l. were considered for the normalization. As valleys and the bordering Swiss Plateau slope values are mostly relatively low, aspect values are not very meaningful. Furthermore, most backcountry ski tours take place in areas higher than 1'000 m.a.s.l. More than 88 % of the filtered routes start above an elevation of 1'000 m.a.s.l. and, after preprocessing, all of them had a maximum height above 1'000 m.a.s.l. Thus, the removal of aspect values below 1'000 m.a.s.l. is justifiable.
Curvature values are relatively hard to interpret as their unit is 1/100 m. Thus, instead of boxplots, the plan and profile curvature was combined. Similar to the aspect visualizations, the curvature values were classified into nine categories (convex, straight, and concave for both directions: plan and profile) and the summed frequency of these classes was obtained. Again, these values were normalized by the maximum frequency sum. To classify curvature values as 'straight', thresholds need to be fixed. It was decided to use the same thresholds as were fixed by Vontobel (2011). This should allow a comparison between route characteristics and typical avalanche release zone characteristics. Thus, all curvature values between -0.2 and 0.2 were classified as 'straight'. Values below -0.2 were classified as convex and values above 0.2 were classified as concave.

4.3.1 Influence of Raster Resolution

Various studies suggest that attributes derived from DEMs vary, depending on the DEM resolution. Enlarging cell sizes leads to a smoothing of the DEM, which causes attribute ranges to narrow (Thompson et al., 2001; Sørensen and Seibert, 2007). Hence, extrema values (minimum / maximum) extracted from rasters are particularly affected by varying cell sizes. It is evident that route attributes extracted from DEMs and derivatives are affected by such varying raster resolutions as well. Thus, the differences between attributes, extracted from the two available rasters, were evaluated. For all routes in the three small study areas, the raster values from DHM25, SwissALTI3D and their slope derivatives were extracted. Then, for each study area, the route attributes from these rasters were compared using a Wilcoxon signed-rank test. As attributes of the same routes were compared, a pairwise test was selected. The sample did not follow a normal distribution, which is why a non-parametric approach was chosen. It is expected that derivatives of the DEMs vary

more than the DEM values. This will be shown by comparing effects on height and slope attributes. The curvature, which corresponds to the second derivation of the elevation model, was not included in the analysis. It is assumed that effects found for slope attributes similarly hold for curvature attributes. Aspect was not included, as no mean and extreme attributes were calculated for this derivative.
As mentioned in section 4.2.4, p-values are rather sensitive to sample sizes. Thus, effect measures are needed to support the statistical tests. Cohen's R was calculated, using the same formula as previously described (eq. 4.4). Due to the central limit theorem, parametric tests are robust against violations of their assumptions (e.g. non-normality). Therefore, if the sample size is large enough, parametric tests show accurate results, despite data not following a normal distribution (Lehmann, 1999). As the samples compared were of a large size, a parametric effect size measure was additionally included to measure the strength of the deviation between the variables. Cohen's D was selected, which allows comparing two means (Fritz et al., 2012a, p. 7). This measure is commonly used in combination with t-tests. The Cohen's D formula is as follows (Cohen, 1988):

$$d = \frac{M_A - M_B}{\sigma} \tag{4.4}$$

4.3.2 Influence of Participation Inequality

To assess the influence of participation inequality on the route attributes, the attributes from rare and heavy user samples were compared. This comparison further allowed evaluating whether heavy users who planned many routes plan different routes than rare users. As the two subsets included different routes each, they were independent. Furthermore, none of the attributes followed a normal distribution, which is why a non-parametric test was chosen: The Mann-Whitney U test. Again, the large sample size would have justified the application of a parametric test (Lehmann, 1999). Cohen's R and Cohen's D were used as effect sizes. They were calculated in the same manner as previously discussed (eq. 4.3 & eq. 4.4).

4.3.3 Influence of Avalanche Conditions

To assess whether the avalanche danger influenced the user's planning process, the influence of the avalanche danger level on route attributes was analysed. For this analysis, only the user-specified avalanche danger was used, as the users planned the routes based on their retrieved information. This was achieved by testing the route attributes for correlation with the user-specified avalanche danger and calculating correlation coefficients. As the avalanche danger levels are on an ordinal scale, rank correlation coefficients were calculated. A well-known correlation coefficient is Spearman's r (r_s). In the absence of ties, r_s can be calculated as follows (Myers and Well, 2003):

$$r_s = 1 - \frac{6\,\Sigma_i\, D_i^2}{N(N^2 - 1)} \tag{4.5}$$

However, as the avalanche danger level can only take one of five values, ties can be expected and this formula is not applicable. Thus, Kendall's Tau (τ) was used for the analysis (Kendall, 1938). The following formula can be used for its calculation

(Myers and Well, 2003):

$$\tau = 1 - \frac{2\,(number\ of\ inversions)}{N(N-1)/2} \tag{4.6}$$

Furthermore, boxplots grouped by avalanche danger and scatterplots including a linear model were plotted to inspect the influence visually.

As all attributes are scattered rather strongly, which affected the τ values, a secondary approach was applied to obtain an understanding of the influence of avalanche danger on route attributes. The routes for which an avalanche danger was present were grouped into two groups based on their avalanche danger level. The first group 'low danger' included all routes with avalanche danger levels 'low' and 'moderate' (levels 1 & 2). The second group, 'high danger', included all routes with avalanche dangers 'considerable', 'high', and 'very high' (levels 3 - 5). For all routes in these two groups, attribute means and Cohen's D were calculated (eq. 4.4). As the presence of a correlation has already been established, the Mann-Whitney U test was not applied to the dataset. Given there exists a correlation between the avalanche danger and route attributes, a difference between the two groups would be the logical consequence. However, Cohen's D might provide some further insights into the strength of the differences between the groups.

4.4 Similarity & Clustering

To identify groups of similar backcountry routes, a clustering approach was applied to the dataset. It was expected that these clusters would correspond to major backcountry routes. Furthermore, high similarities between routes and neighbouring routes are an indicator of the accuracy of a route. If a route deviates strongly from all other routes, it is rather unrealistic. Realistic routes should also correspond rather well to existing GPS-tracks.

As was noted by Kisilevich et al. (2010), the clustering problem can be reduced to a choice of similarity measure and generic clustering algorithm. As a similarity measure will also be needed for the comparison between planned routes and GPS-tracks, such an approach was taken. In any case, the computation of similarities is computationally-intensive. Therefore, only the routes in the three small study areas were used as inputs. The following two subsections will cover the selected similarity measure and clustering algorithm.

4.4.1 Similarity

As mentioned in section 3.1.1, the planned backcountry routes do not include a temporal component. Thus, all spatio-temporal similarity measures are out of the question. As was discussed in section 2.1.2, there exist various purely spatial similarity measures with varying levels of complexity. As the full routes should be compared and no segmentation was applied, only global measures were taken into account. The Fréchet distance was chosen as a similarity measure. Surely, there would have been other measures which would have provided sensible results. However, there exist multiple implementations of the Fréchet distance for trajectories in R and it provided reliable results in previous works. Furthermore, in contrast to other distance

measures, such as the "common route distance", there is no need for setting parameters and it is not as simplistic as for example the simple Euclidean distance.

Alt and Godau (1995) defined the Fréchet distance between two curves A and B, as follows:

$$\delta_{Frechet}(A,B) = \inf_{\substack{\alpha[0,1]\to[a,a'] \\ \beta[0,1]\to[b,b']}} \max_{t\in[0,1]} \left\{ \delta_{Euclidean}\Big(A\big(\alpha(t)\big), B\big(\beta(t)\big)\Big) \right\} \tag{4.7}$$

where $A(\alpha(t))$ and $B(\beta(t))$ each denote a location on the two curves at some point t on these curves. t denotes the index position of a location in the ordered set of points, which represents the curve. $\alpha(t)$ and $\beta(t)$ can only move forward, maximally one index position, and only in a continuous fashion. The Fréchet distance can be implemented using a dynamic programming approach, which has a complexity of $O(mn\ \log(mn))$ for two curves of lengths m and n.
In this work, the implementation of the 'trajectories' package was used, which provides a slightly higher complexity (Pebesma, 2012). However, as the similarity between routes was only calculated for the routes in the small study areas, the process was still sufficiently efficient. To use this implementation, all routes had to be converted to a 'track' object. Additionally, all of them were projected to the LV95 reference system. With a planar reference system, the unit of the resulting Fréchet distance is meters. Besides coordinates, these objects needed a temporal component. As this temporal component has no influence on the Fréchet distance, constant increasing timestamps starting at the same time were generated for all routes. The used implementation needs timestamps to keep track of the point order, making up the line. After converting all routes to 'track' objects, a similarity matrix was generated for all three small study areas, each containing the Fréchet distances between all routes in those areas. This similarity matrix served as input for the following clustering step. Furthermore, k Nearest Neighbours (kNN) distances could be calculated and stored as attributes of the original routes.

The similarities between routes and GPS-tracks were calculated in a similar manner. The GPS-tracks were converted to 'track' objects, which allowed the computation of the Fréchet distance. As only a fraction of the GPS-tracks included timestamps, again, constant increasing timestamps were generated for all of them. As mentioned before, the timestamps have no influence on the Fréchet distance. Then, for each small study area, the similarities between all planned routes and all GPS-tracks were calculated. For all planned routes, "k Nearest GPS-track" (kNG) distances, analogical to kNN, were calculated and stored as attributes.

4.4.2 Clustering

After selecting a similarity measure for the planned route trajectories, the clustering problem is reduced to a choice of a generic clustering algorithm. Partitioning methods require the preliminary specification of the number of clusters. All objects are then assigned to one of the clusters. As it is unknown how many major backcountry routes exist in an area, the choice of an appropriate number of clusters is rather difficult. Furthermore, partitioning methods do not provide the possibility to detect outliers in the dataset, which certainly will exist in the analysed dataset. Hierarchical methods assign all objects to a cluster, similar to partitioning methods. However, for large numbers of clusters, all outliers represent a single class and can therefore

be identified. Nevertheless, the ascertainment of the optimal number of classes is tedious work. Therefore, neither partitioning nor hierarchical methods were considered for clustering the trajectories. There exist no implementations of grid-based clustering approaches in R, which provide the possibility to input a similarity matrix. Thus, they were not applicable to the data at hand.

The two most prominent density-based clustering approaches, DBSCAN and OPTICS, are both implemented in R. The 'dbscan' package provides fast implementations of DBSCAN and OPTICS (Hahsler and Piekenbrock, 2017). DBSCAN requires two preliminary parameters to be fixed: ε and minPts. For every trajectory tr_i, the number of trajectories to which the Fréchet distance is smaller than the specified ε, so-called ε neighbours, is retrieved. All trajectories with equal or more ε neighbours than minPts are classified as core trajectories. All core trajectories between which the Fréchet distance is smaller than the specified ε are grouped in a cluster. Then, all non-core trajectories which are closer to a core trajectory than ε are also appended to the according cluster, while all other trajectories are considered noise.
A careful selection of parameters is critical for the success of the clustering. ε strongly influences the density of found clusters. The smaller ε, the denser and more defined the found clusters are. However, with smaller ε-value, clusters also get smaller and more routes are classified as outliers. This might lead to classifying sparser clusters completely as outliers. If a large ε is chosen, found clusters exhibit lower densities and the trajectories of the same cluster show larger variations. Furthermore, groups of trajectories which follow a similar path but start of at slightly different locations (e.g. a town center and a nearby parking lot) might be included in the same cluster.
Ester et al. (1996) suggest using kNN distance plots to select an appropriate ε-value. ε is set to the first point of a "valley" formed in the plot. However, such valleys are not always easily detected. Figure 4.7 shows the 10th nearest neighbour distances for all three small study areas. At first glance, 1'800 m for St. Antönien and 2'000 m for Urserental and Wildstrubel seem appropriate ε-value. However, considering the strong terrain variations in alpine areas, these values seem rather large. Additional discernible breaks indicating a valley can be found at 800 m for St. Antönien and 1'000 m for Urserental and St. Antönien. These values still seem relatively high, considering alpine terrain variations. Finally, DBSCAN was ran with various ε-values between 400 m and 1'200 m. The resulting clusters were then inspected visually.

Ester et al. (1996) propose always setting minPts to 4 for two-dimensional data. However, this might result in small clusters and more clusters in general. As the aim was to identify clusters that correspond to major backcountry routes, minPts was set to 10. This results in clusters including at least 10 routes.
The clustering was conducted with the OPTICS algorithm as well. OPTICS is based on DBSCAN. However, instead of a cluster structure, it orders the objects, based on which a clustering can be performed. OPTICS was applied to the dataset with various parameters. Yet, the results never reached the quality of those achieved with DBSCAN, which is why it was neglected after all.

To put the resulting clusters into context, two results were visualized together with the GPS-tracks from Gipfelbuch.ch and Camptocamp. This allowed a visual comparison between clusters of planned backcountry routes and recorded GPS-tracks and digitized SAC-tours.

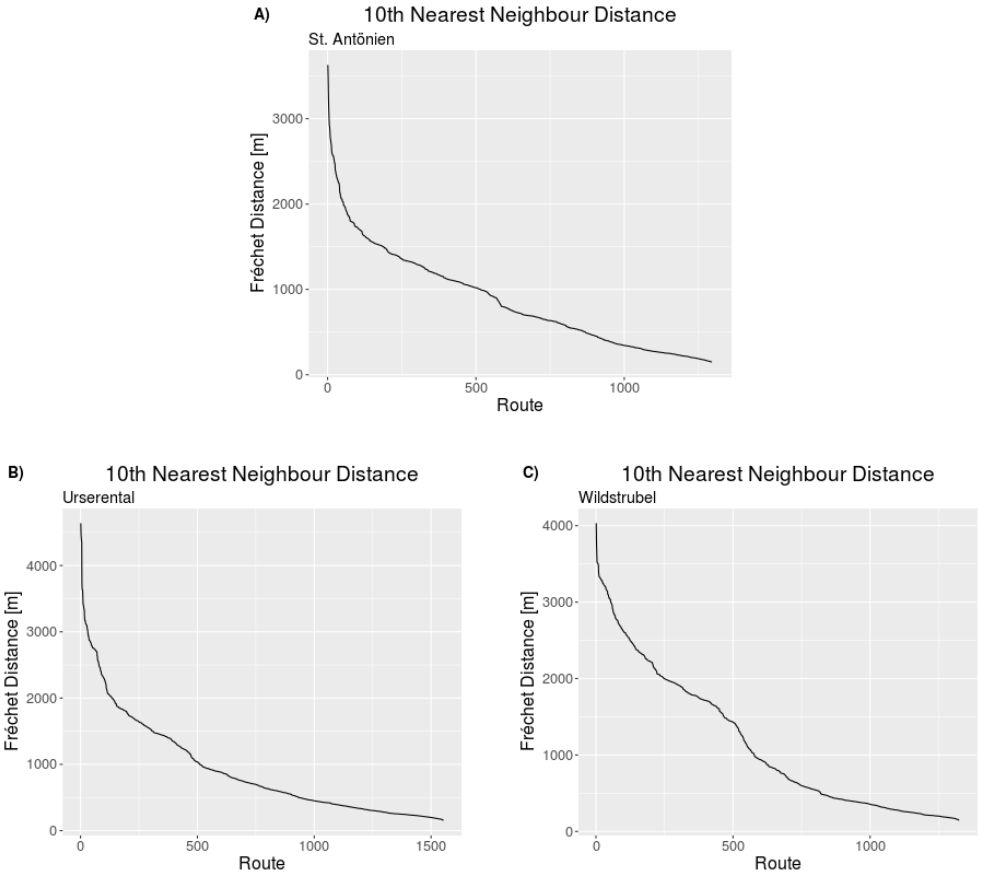


Figure 4.7: *10-Nearest neighbour distance plots for A) St. Antönien, B) Urserental, and C) Wildstrubel*

4.5 Computing Environment & Software

Most computations were conducted on a Linux server running Ubuntu 16.04.4 LTS. The server had a 64 bit architecture and four Intel cores running at 2 GHz. 16 GB of RAM and about 600 GB of disk space was available for the computations. The data was stored in a PostgreSQL 9.6.7[2] database including PostGIS 2.3.3[3], which was installed on that server. Furthermore, R 3.4.3[4] and RStudio Server 1.1.383[5] were installed on the server and used for all computations. Several R packages were installed. A complete list can be found in table 4.2.
To access the data in the database, DataGrip 2017.2.3[6] was installed on a local machine. Google Chrome 65 was used to access RStudio Server. Additionally, ArcGIS Pro 2.0.1[7] served for visualization purposes.

[2] https://www.postgresql.org/
[3] https://postgis.net/
[4] https://www.r-project.org/
[5] https://www.rstudio.com/
[6] https://www.jetbrains.com/datagrip/
[7] https://pro.arcgis.com/en/pro-app/

Table 4.2: *List of used R packages with a short description and reference.*

R Package	Usage	Reference
coin	Statistical analyses	(Hothorn et al., 2008)
cowplot	Visualizations	(Wilke, 2017)
data.table	Data manipulation	(Dowle and Srinivasan, 2017)
dbscan	Density clustering	(Hahsler and Piekenbrock, 2017)
dplyr	Data manipulation	(Wickham et al., 2017)
geosphere	Distance calculations	(Hijmans, 2017a)
ggplot2	Visualizations	(Wickham, 2009)
lsr	Statistical analyses	(Navarro, 2015)
plotKML	Import of GPX-files	(Hengl et al., 2015)
plotrix	Visualizations	(Lemon, 2006)
plyr	Data manipulation	(Wickham, 2011)
raster	Raster manipulation	(Hijmans, 2017b)
reshape2	Data manipulation	(Wickham, 2007)
rgdal	Spatial data manipulation	(Bivand et al., 2017)
rgeos	Spatial data manipulation	(Bivand and Rundel, 2017)
RPostgreSQL	Interface to database	(Conway et al., 2017)
sp	Spatial data manipulation	(Bivand et al., 2013)
spacetime	Creation of trajectories	(Pebesma, 2012)
trajectories	Similarity of trajectories	(Pebesma and Klus, 2015)
velox	Fast raster extraction	(Hunziker, 2017)

Chapter 5

Results

This chapter will outline all results that were acquired during the process of this thesis. These results will be discussed and critically analysed in chapter 6. The chapter is structured as follows: First, a thorough description of the dataset, including general statistics and spatial and temporal distribution, will be given (sec. 5.1). Furthermore, a general overview of participation inequality will be provided and user-stored avalanche conditions will be addressed. Secondly, the route attributes will be examined (sec. 5.2). These include intrinsic attributes and attributes extracted from DEMs. Additionally, influences from raster resolution, participation inequality and avalanche danger on these route attributes will be evaluated. Finally, the results of the similarity assessment and clustering will be presented (sec. 5.3).

5.1 General Dataset Overview

After preprocessing, 56'594 routes created by 7'905 users remained. In general, each tour contained only one main route. Therefore, the number of tours should correspond to the number of routes after the filtering. However, there were five tours with two main routes, which resulted in 56'589 tours. As all other routes corresponded to exactly one tour, all planned route trajectories will be referred to as 'routes' for the remainder of this work. Table 5.1 summarises the obtained results.

Feature	Count
Users	7'905
Tours Total	56'589
Routes Total	56'594
Avg. Routes per User	7.16

Table 5.1: *Summary of filtered route dataset, as described in section 4.1.*

The number of routes in each small study area is described in table 5.2.

Study Area	Number of Routes
Wildstrubel	1'324
Urserental	1'553
St. Antönien	1'295

Table 5.2: *Number of routes in small study areas.*

5.1.1 Temporal Distribution

For 23'765 routes out of the total 56'594 routes, a tour date was specified. However, some dates had unreasonable year values or were during the summers season, when backcountry skiing is not possible. After the removal of those dates, 20'964 tour dates remained. Figure 5.1 illustrates the distribution of routes for the winter seasons 2013/14 to 2017/18. At the time of the data extraction from the database, the winter season 2017/18 had not terminated. Thus, some more routes are to be expected for this season. Figure 5.1 shows that, after the first season (2013/14), more routes were planned the following two seasons (2014/15 & 2015/16). There was a clear decrease for seasons 2016/17 and 2017/18. Possible causes for this decrease will be addressed in the discussion.

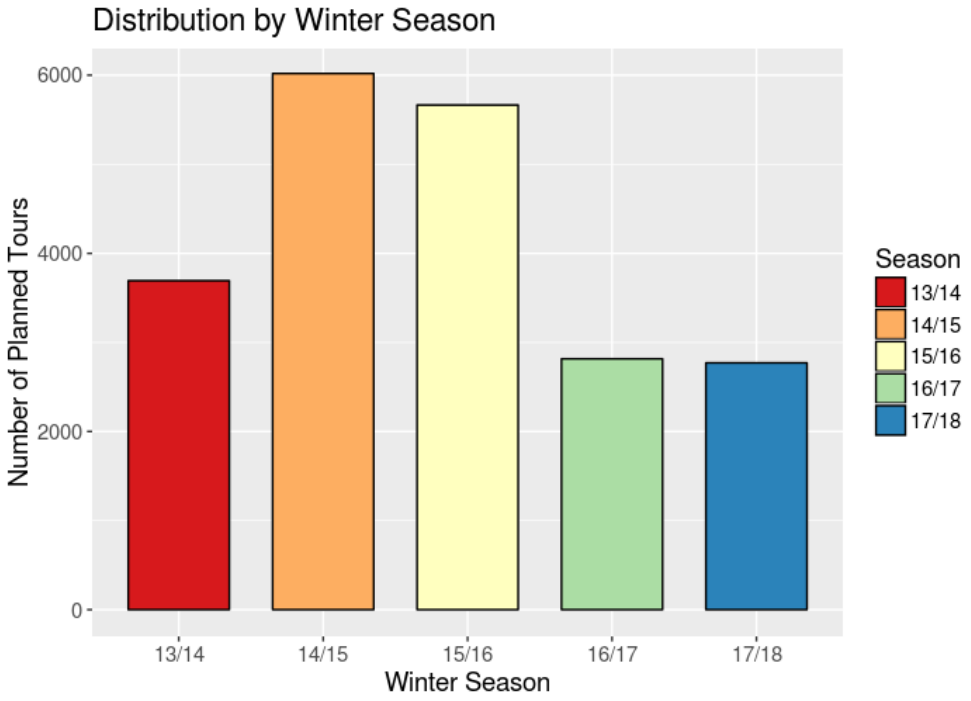


Figure 5.1: *Distribution by winter season of routes which included a tour date for seasons 2013/14 to 2017/18.*

Figure 5.2a provides an overview of the distribution of routes by month. The dates of the routes range from November to May. During all seasons but 2016/17, most routes were planned for dates between January and March, peaking in February for three out of five seasons. On average, the number of routes planned for dates in January to March was about twice as high as for dates in December and April. Routes planned for dates in November and May were rather rare.
Most routes were planned for weekends. The number of routes planned for Fridays was about 1.5 times that for other weekdays. This holds for all seasons covered and can be seen in figure 5.2b. On average, the number of routes planned for weekend days was about four times that for regular weekdays. Saturday was the most popular day for backcountry skiing in all seasons but 2017/18.

5.1.2 Spatial Distribution

Figure 5.3 shows a density map of all routes after the filtering process for the large study area. As mentioned in section 4.2.2, a search radius of 250 m was used for

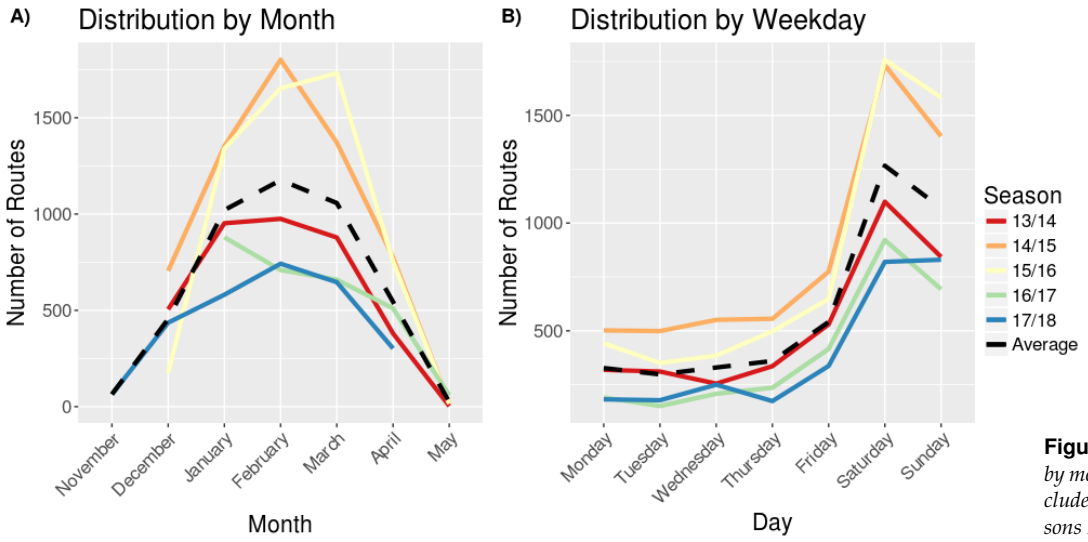


Figure 5.2: *Distribution by month of routes which included a tour date for seasons 2013/14 to 2017/18.*

the overview map. As can be seen in the density maps, most alpine areas showed some planned backcountry touring activity. The relatively low-lying valleys, such as the Rhone Valley, Alpine Rhine Valley or Tessin valleys, showed little to no planned backcountry touring activity. Further, there was generally little planned backcountry touring activity in the Italian-speaking area of Switzerland in the south.

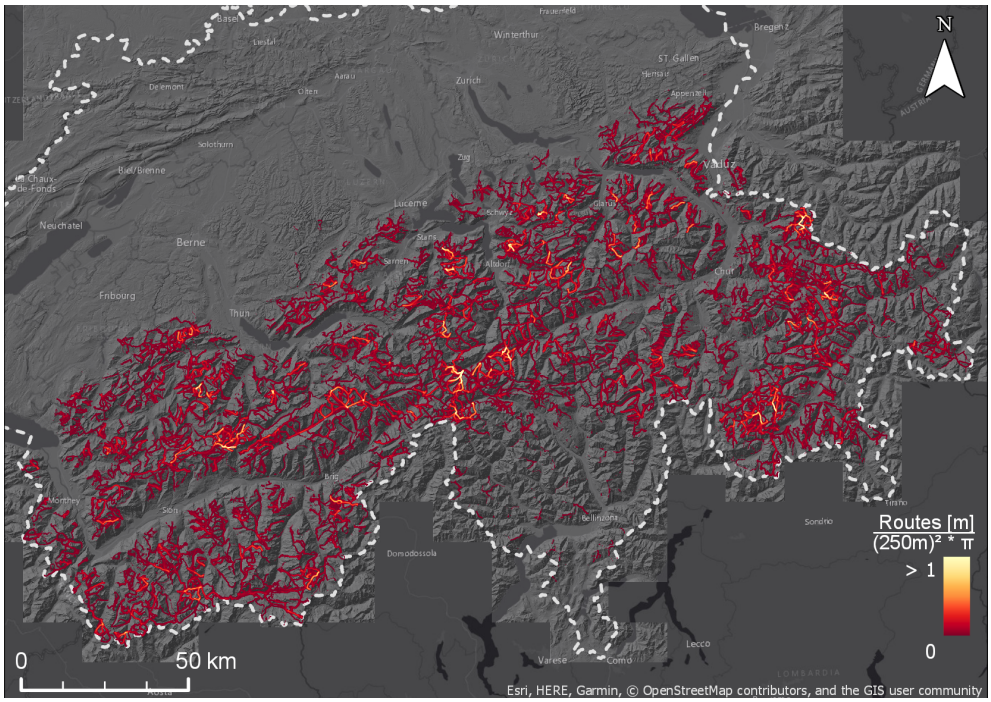


Figure 5.3: *Line density map of filtered routes for large study area. Search radius: 250 m.*

Figure 5.4 shows two higher resolution density maps of the two small study areas Wildstrubel and Urserental. The Wildstrubel area (fig. 5.4.1) shows a wide distribution of routes, with starting points lying on all sides of the Wildstrubel massif. Some popular routes, e.g. from the Gemmi Pass (1c) to the Lämmerenhütte (1b) and further to the Wildstrubel (1a) or from Kandersteg towards the Wildstrubel, are clearly discernible.

In the Urserental study area (fig. 5.4.2), it is clearly visible that most routes start from Realp. Three main directions can be distinguished: 2a) Chli Bielenhorn (north), 2b) Stotzigen Firsten (south, south-west), and 2c) Gross Leckihorn (south).

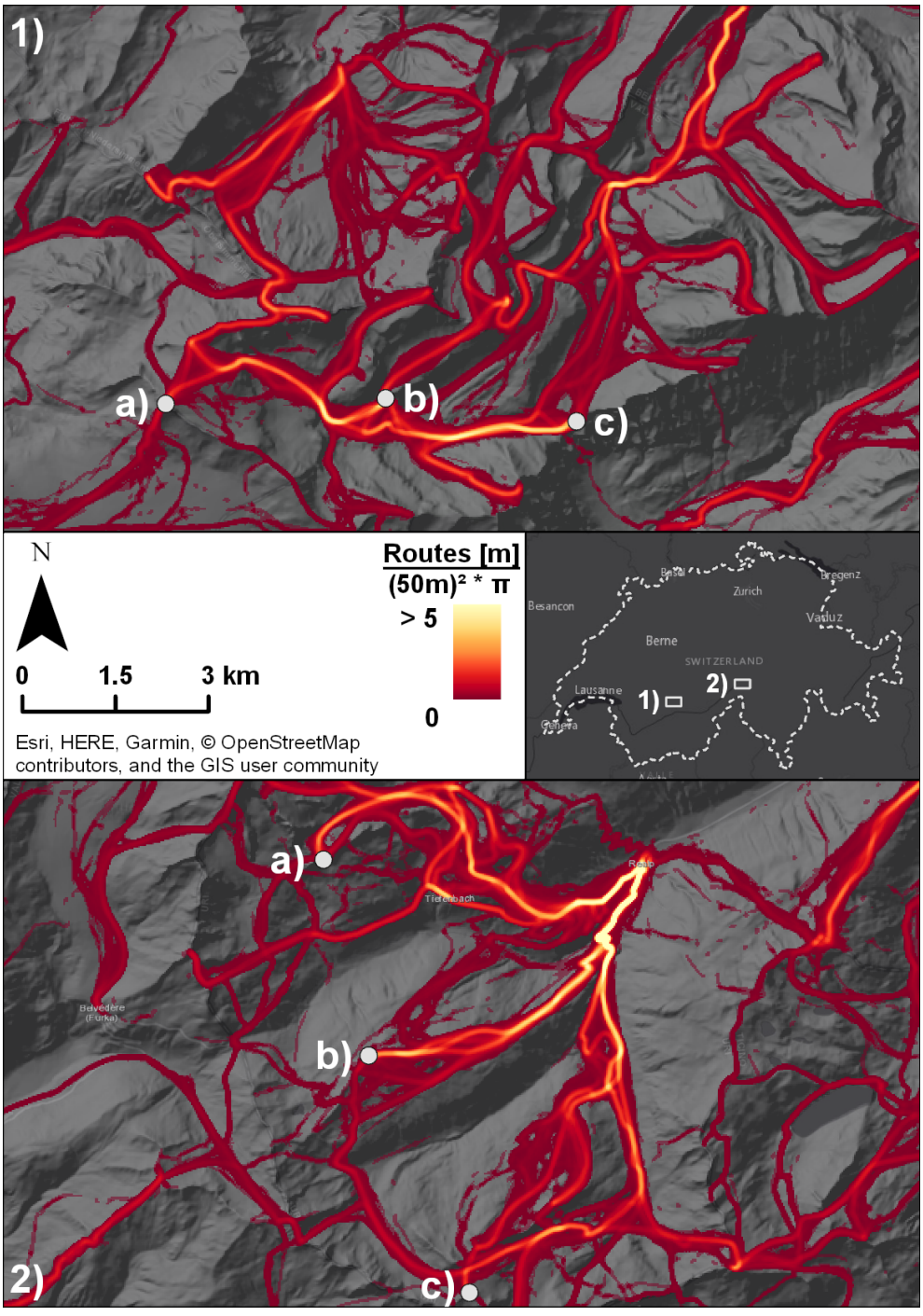


Figure 5.4: *Line density map of: 1) Wildstrubel with a) Wildstrubel, b) Lämmerenhütte, and c) Gemmipass. 2) Urserental with a) Chli Bielenhorn, b) Stotzigen Firsten, and c) Gross Leckihorn.*
Search radius: 50 m.

5.1.3 Participation Inequality

As described in section 5.1, the 56'594 routes were planned by 7'905 distinct users. There was no user ID available for 1'824 routes. These routes had probably been

created by users who deleted their account after the route creation. Figure 5.5a shows a boxplot of the number of routes users planned. It was cropped for clarity. The maximal outlier, thus the person who planned the most routes, recorded 238 routes. Figure 5.5b illustrates the distribution of routes by users and a cumulative frequency after an illustration by Purves (2011).

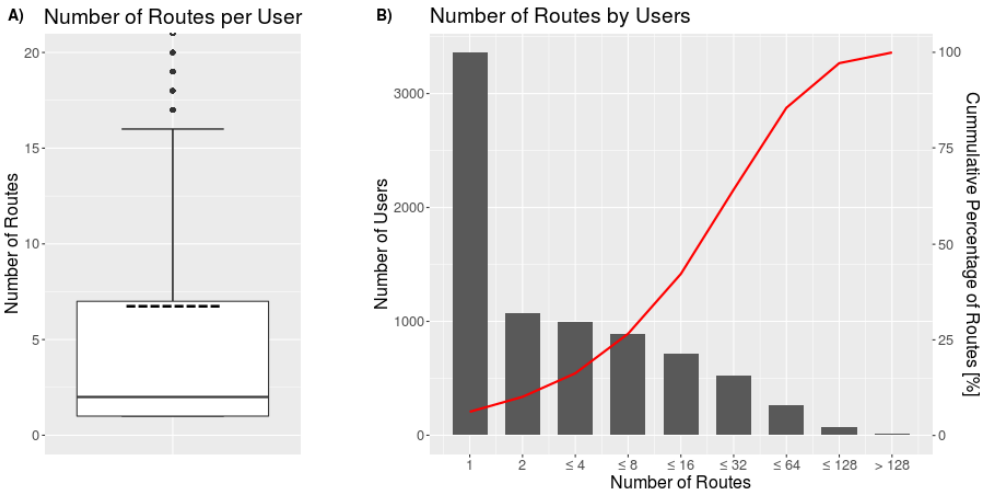


Figure 5.5: *A) Boxplot of number of routes per user. B) Distribution of number of routes by users after Purves (2011).*

It is evident that there exists a participation inequality in the examined dataset. However, the participation inequality is not as strong as the suggested 90-9-1 rule introduced by Nielsen (2006). In this dataset, the top 1 % users account for 13.6 % of the content, the 9 % intermittent users account for 41.7 % of the content and the 90 % rare users, also known as lurkers, account for 44.7 % of the content. Hence, the examined dataset is less skewed than typical social media platforms.

As described in section 4.2.3, the full route dataset was sampled in three groups. This sampling of users in three groups of rare, heavy and random user groups resulted in three sets of 5'659 routes each. The number of users of each group is listed in table 5.3. In the group with random routes, 196 routes did not include a user ID. For both other datasets, routes without a user ID were not considered.

Group	Number of Users
Rare Users	4'512
Random Users	2'558
Heavy Users	56

Table 5.3: *Number of users in each subgroup, created to analyse the effect of participation inequality.*

The distribution of planned routes by users for all groups and the cumulative frequency is shown in figure 5.6. As expected, the distribution of the random users (fig. 5.6a) is similar to the total dataset. The distribution of the rare users (fig. 5.6b) revealed that over 75 % of the users of this group only planned a single route and most of the remaining 25 % users planned two routes. 50 out of the 56 heavy users created between 76 and 147 routes (fig. 5.6c). Only three users created more than 150 routes. Those users who are classified with less than or equal to 75 routes all created 75 routes. However, as only 10 % of the routes were selected, only some of their routes were included in the subset.

These three subsets were used for the generation of χ-maps, which allowed assessing the existence of a geographic bias in the dataset, as will be shown in the next

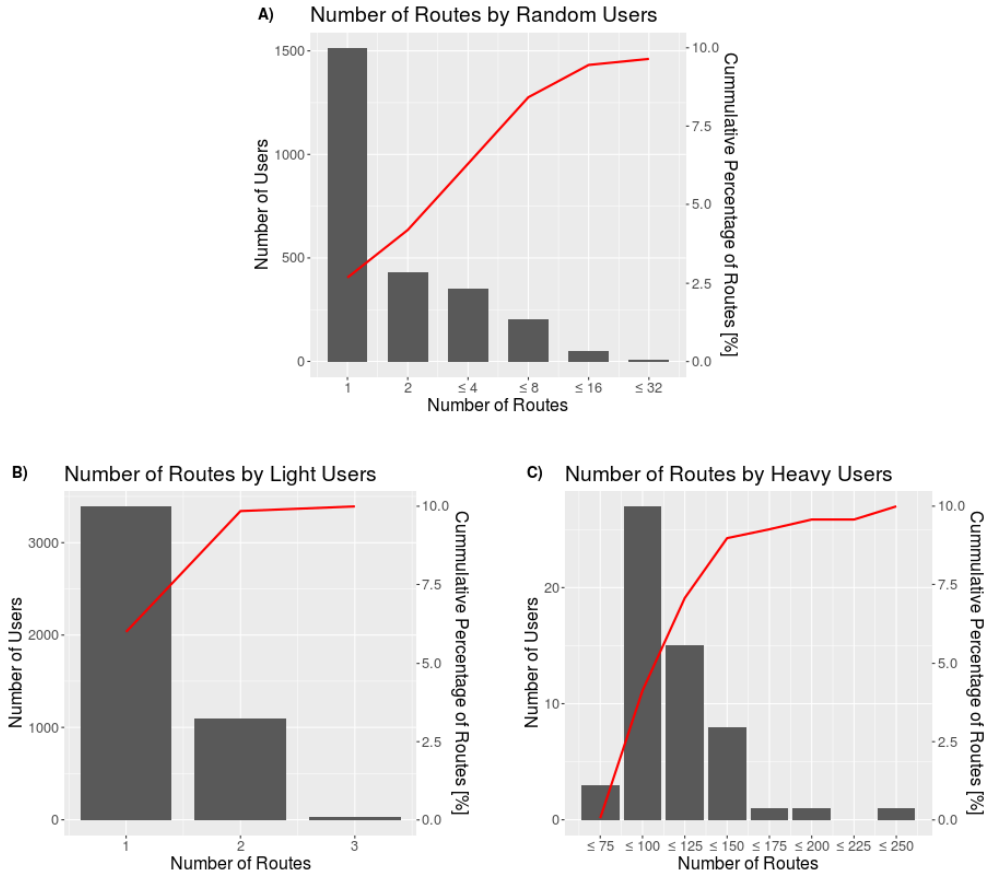


Figure 5.6: *Distribution of number of routes by users for all subgroups: A) Random users, B) rare users, and C) heavy users.*

paragraph. Furthermore, in section 5.2, the differences in route attributes between the user groups rare and heavy will be assessed.

Geographic Bias To evaluate, whether routes from rare and heavy users show a different spatial distribution, χ-maps were generated. The χ-maps can be interpreted as follows: Hot colors (red to yellow) indicate higher route density in the observed raster (density of rare / heavy users) than in the expected raster (density of random sample). Contrarily, cold colors (blue to white) indicate lower route density in the observed raster than in the expected raster.
Observing the χ-map of the rare user group (fig. 5.7), there seems to be a general trend of high χ-values towards the Swiss Plateau. This is supported by the mean center of the line centroids of rare user routes, which lies 14.5 km to the west and 2.2 km to the north of the mean center of line centroids of heavy user routes. High χ-values seem to occur most frequently on the northern side of the Alps. Furthermore, areas with high accessibility, such as the Sion area, Andermatt, and Davos, exhibit relatively high χ-values. Low χ-values mainly occur in inner-alpine areas, such as Wildstrubel, Aletsch, and Bivio.
The χ-map of the heavy user group (fig. 5.8) shows high χ-values for several regions, such as Southern Valais (Verbier, Arolla, & Saas Fee), Aletsch, inner Grisons (especially Engadin), and Glarus Alps. All these areas show exceptionally high χ-values, potentially resulting from few users planning large numbers of routes in these areas. Wide areas in the Bernese and Pennine Alps show low χ-values, except for the previously mentioned regions. Furthermore, the Uri Alps and the whole Davos

area show relatively low χ-values. It is noteworthy that χ-values in the three selected small study areas (Wildstrubel, Urserental, & St. Antönien) are low for both user groups. In general, it is safe to assume that there exists a geographic bias in the dataset, introduced by participation inequality. Rare users plan more routes in areas which are highly accessible from the Swiss Plateau, while heavy users prefer inner-alpine areas.

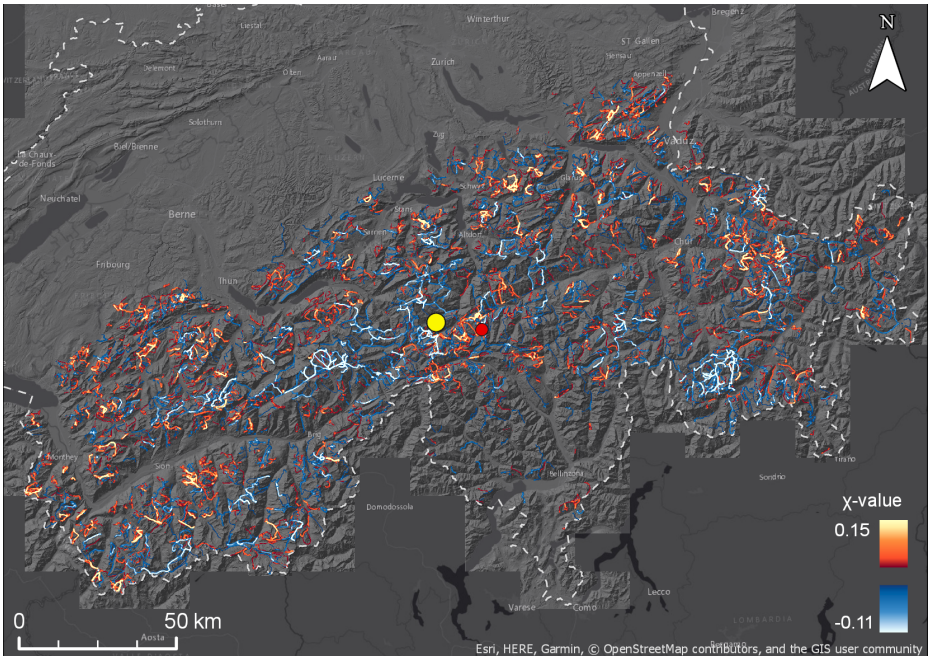


Figure 5.7: *Chi-map for rare users. Density map of routes from rare users as observed and density map of random sample as expected value. Search radius: 250 m. Mean center of route centroids from rare users as yellow circle.*
Mean center of route centroids from heavy users as red circle.

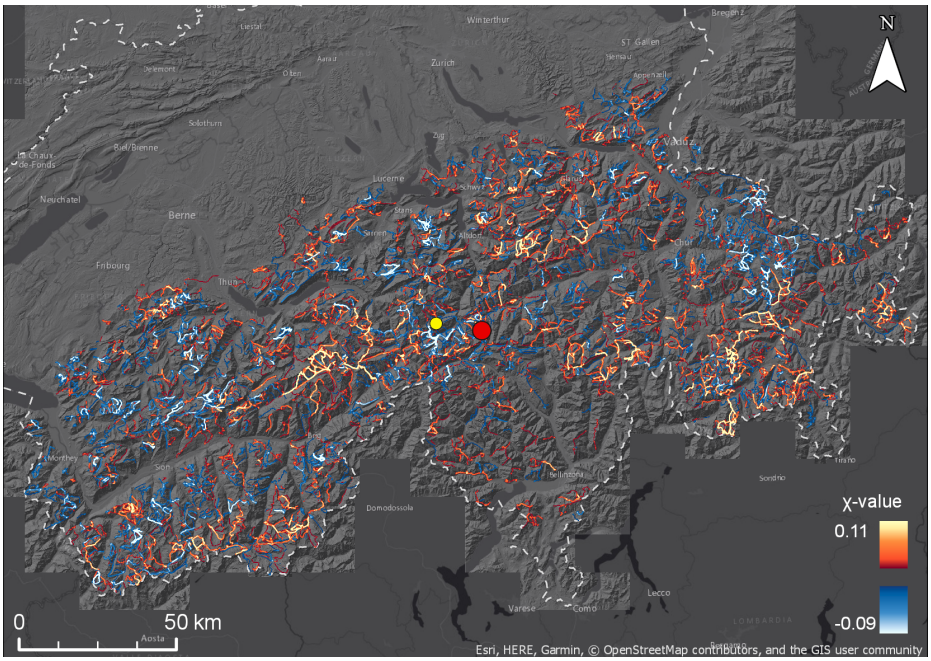


Figure 5.8: *Chi-map for heavy users. Density map of routes from heavy users as observed and density map of random sample as expected value. Search radius: 250 m. Mean center of route centroids from rare users as yellow circle.*
Mean center of route centroids from heavy users as red circle.

5.1.4 Avalanche Conditions

The forecasted avalanche danger synthesises various conditions (weather, snow layer etc.) to provide a general sense of the avalanche danger in alpine areas. These are crucial during the planning process and can give an idea of what routes are possible and sensible, considering the risk of being caught in an avalanche. There were 15'231 routes for which an avalanche danger was specified. For more than 88 % of these routes, the specified avalanche danger was "moderate" (level 2) or "considerable" (level 3). Only for about 10 % of the routes that included a danger level was the specified level "low danger" (level 1). For less than 1 % of these routes, the specified danger level was "high" or "very high" (levels 4 & 5). These numbers are summarised in figure 5.9.

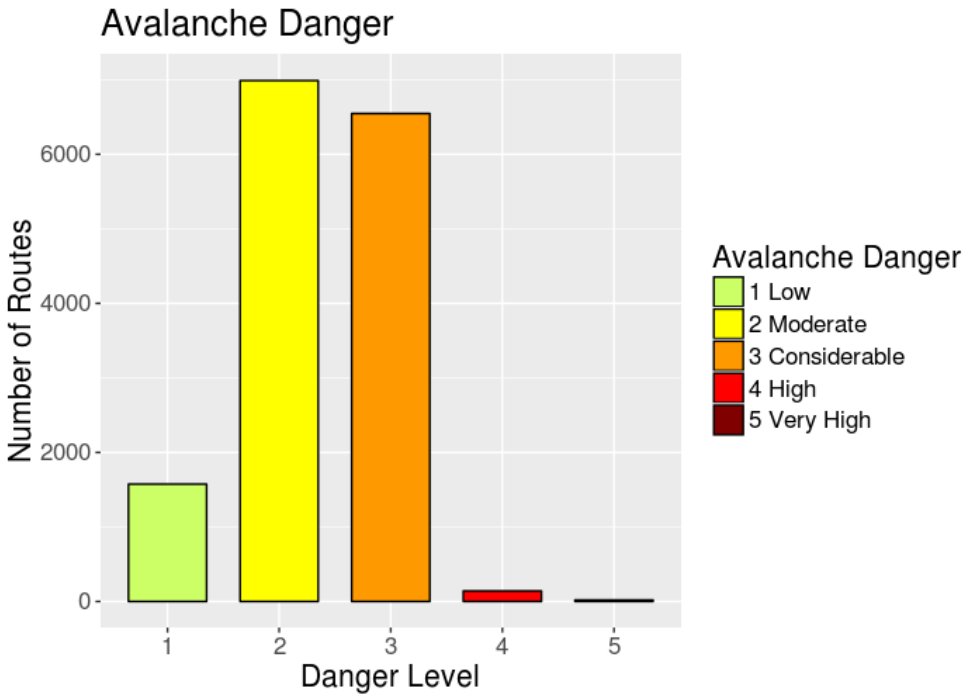


Figure 5.9: *Distribution of user-specified avalanche danger level for filtered routes between seasons 2013/14 and 2017/18.*

Comparison to Authoritative Danger Levels To assess whether users specified the avalanche danger carefully, the user-stored avalanche dangers were compared with the danger level from the avalanche bulletin. The danger level from the avalanche bulletin can be considered authoritative data, with the avalanche bulletin being published by the SLF. From the 56'594 routes that remained after the preprocessing, 23'765 of the routes included a tour date. For 21'322 of these tour dates, the avalanche danger in the evening avalanche bulletin of the day before the tour was extracted. For 2'449 routes, two avalanche bulletins were available (wet and dry conditions). A total of 10'024 of all routes included a user-specified avalanche danger and a tour date, which allowed for a comparison between user-stored and authoritative danger levels. From these 10'024 routes, 9'146 had a single authoritative avalanche danger level and 878 had avalanche danger levels for two conditions. Table 5.4 provides an overview of these numbers.

The danger levels at the start and end points were equal for all routes, including those with danger levels for two conditions. Thus, a distinction of danger levels at

Feature	Count
Routes Total	56'594
Thereof	
including user-spec. danger level	15'231
including tour date	23'765
including authoriative danger level	21'322
including user-spec. & authoritative danger levels	10'024
Thereof	
including user-spec. & single authoritative danger level	9'146
including user-spec. & two authoritative danger levels	878

Table 5.4: *Summary of routes that were used for comparison of user-specified avalanche danger and authoritative avalanche danger.*

start and end points is redundant and was therefore neglected.

In a next step, the user-specified avalanche danger and authoritative avalanche danger were compared. Only the routes that included a single authoritative danger level were used for this analysis. H_0 for the Wilcoxon signed-rank test was: The rank-sums of the differences between the user-specified avalanche danger and the authoritative avalanche danger follow a symmetric distribution around zero. The resulting p-value was 0.01184, which is lower than the significance level of 0.05. This suggests that H_0 can be disproved, meaning that there exist significant differences between the signed ranks of user-specified avalanche danger and the authoritative avalanche danger. However, considering the large sample size, the p-value is still rather large. Furthermore, the effect size is very small with 0.026. This indicates that the shift is not constant in one direction. For 7'156 routes (78.24 %), users specified the same avalanche danger as was stored by the SLF on the day before the specified date. For 1'852 routes (20.25 %), the avalanche danger differed by one level from the authoritative danger level, leaving only 138 routes (1.51 %) with larger deviations. Figure 5.10a illustrates the distribution of the user-specified avalanche danger, side-by-side with the authoritative avalanche danger. Figure 5.10b shows the distribution of the danger level differences. Negative values indicate authoritative danger levels higher than the user-specified avalanche danger level and vice versa.

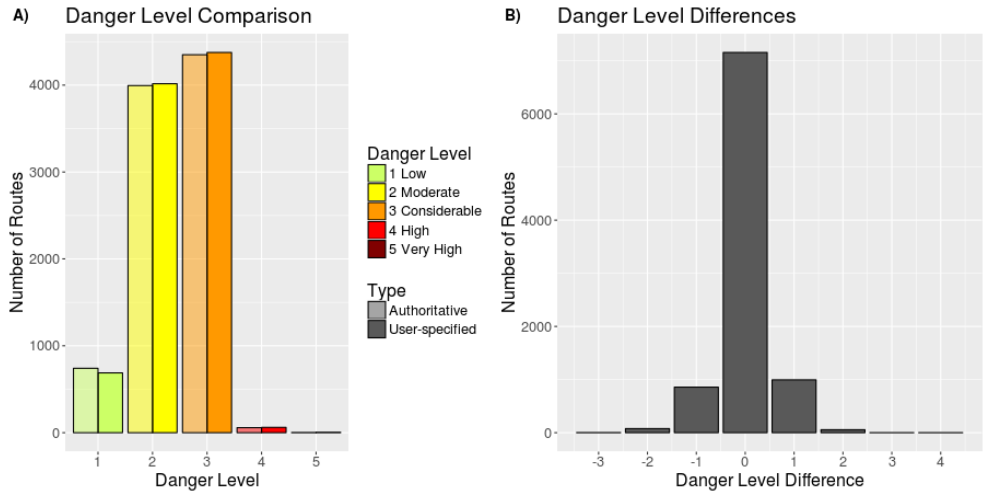


Figure 5.10: *A) Comparison of authoritative and user-specified avalanche level distribution with single condition. B) Distribution of danger level differences.*

5.2 Route Attributes

Intrinsic route attributes, such as length, step length, etc., can provide an idea of the level of detail of routes and of general route characteristics. Terrain attributes extracted from DEM can provide context to the routes. The terrain directly influences the avalanche danger and, thus, the riskiness of routes can be indicated by terrain attributes.
These intrinsic and extrinsic attributes are likely to be affected by certain parameters. The resolution of DEMs is expected to exhibit a direct influence on terrain attributes. Furthermore, whether there are differences between rare and heavy user routes needs to be assessed. To assess whether users adjust their routes based on avalanche danger, the influence of avalanche danger on terrain attributes needs to be evaluated as well.

Table 5.1 shows the summary statistics for all attributes that were either calculated or extracted from the DHM25 and its derivatives, except for aspect and heading. For each attribute, median, mean and standard deviation were calculated. Most attributes exhibit a lower median than mean and are presumably positively skewed. Boxplots for all attributes except aspect, heading, and curvature are visualized in figures 5.11 to 5.15. The mean values are included in the boxplots in the form of dashed lines. Figure 5.11 indicates that most routes range between 5'000 m and 10'000 m with step lengths between 125 m and about 300 m. Most routes are located in heights between 1'250 m.a.s.l. and 3'000 m.a.s.l. with an average mean height of about 2'000 m.a.s.l. (cf. fig. 5.13). Median slope values range between 2°and 41°with a mean slope of 19°. It is noticeable that slope values show less dispersion than other attributes. Straightness values are widely distributed, as illustrated in figure 5.15. The median is 0.38, indicating more routes were direct paths rather than round trips. As discussed in section 4.1.2, round-trips would exhibit straightness values of 1. Plain curvature values are relatively hard to interpret and, thus, a bar plot of the frequencies of the classified curvature values was generated, as described in section 4.3 (fig. 5.16). The normalized curvature frequencies reveal that straight slopes are traversed most, meaning that curvature in both directions is close to zero. This is not surprising as straight slopes are expected to occur most often in the terrain. The most prominent landforms after straight slopes are slopes with concave profile / convex plan curvatures and convex profile / concave plan curvatures. The curvature boxplots can be found in appendix C. Positive curvature values indicate concave surfaces, whereas negative values indicate convex surfaces. The summed counts for all aspect values of the routes reveal that north-west to south-east exposed slopes are traversed more frequently than south to west exposed slopes (fig. 5.17.A). South-eastern slopes are traversed most. However, the counts are distributed relatively evenly between north-west and south-east. As was noted in section 4.3, the Alps show a general south-west to north-east trend. Thus, it is not surprising that the class north-east shows the highest values for the normalized aspect frequencies, followed by north and east (fig. 5.17.B). These values show high frequencies while they do not occur that often in the terrain. The heading denotes the travel direction of the tours. The summed counts for headings show that most routes head west to north with a maximum at north-west (fig. 5.18). The differences for headings vary slightly more than for aspects but are also distributed rather evenly.

Variable		Median	Mean	Std. Dev.
length	[m]	6471.25	7542.85	3998.06
numberOfVertices		34.00	46.92	45.54
stepLength	[m]	194.99	232.54	146.03
minHeight	[m.a.s.l.]	1527.71	1571.57	503.72
meanHeight	[m.a.s.l.]	2051.66	2061.95	486.65
maxHeight	[m.a.s.l.]	2628.84	2627.54	568.69
heightDifference	[m]	1023.50	1055.98	404.97
minSlope	[°]	1.96	2.64	2.30
meanSlope	[°]	18.99	19.11	3.79
maxSlope	[°]	41.01	42.37	9.17
minPlanCurv	[1/100m]	-2.75	-3.08	1.58
meanPlanCurv	[1/100m]	0.02	0.04	0.15
maxPlanCurv	[1/100m]	3.65	4.15	2.25
minProfCurv	[1/100m]	-3.89	-4.66	3.18
meanProfCurv	[1/100m]	0.01	-0.01	0.17
maxProfCurv	[1/100m]	2.46	2.96	1.97
straightness		0.38	0.47	0.30

Table 5.5: *Summary statistics of calculated attributes and attributes from raster extraction.*

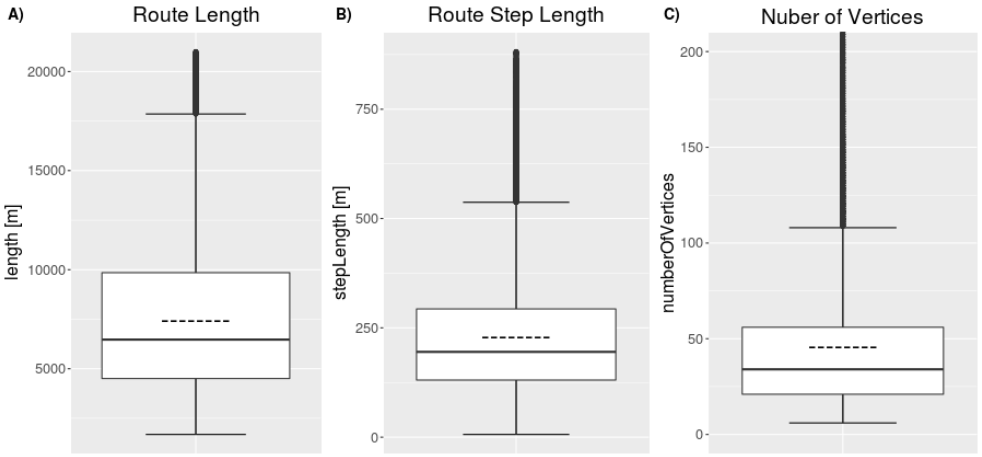


Figure 5.11: *Boxplots of attributes A) length, B) stepLength, and C) numberOfVertices.*

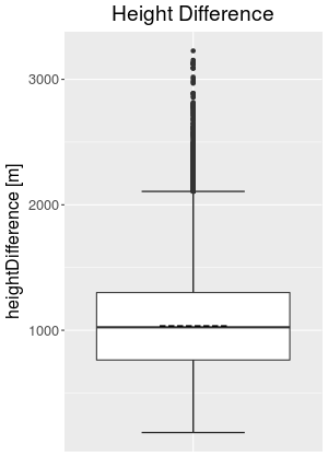


Figure 5.12: *Boxplot of attribute heightDifference.*

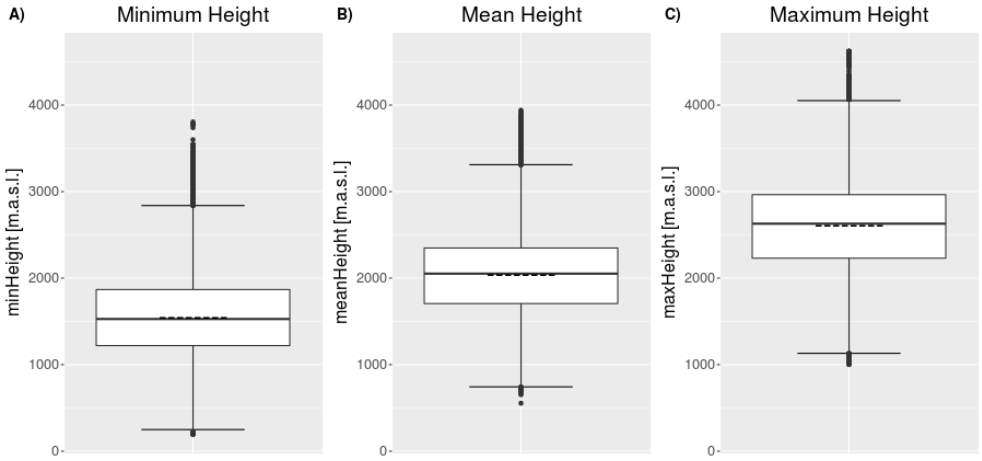


Figure 5.13: *Boxplots of attributes A) minHeight, B) meanHeight, and C) maxHeight.*

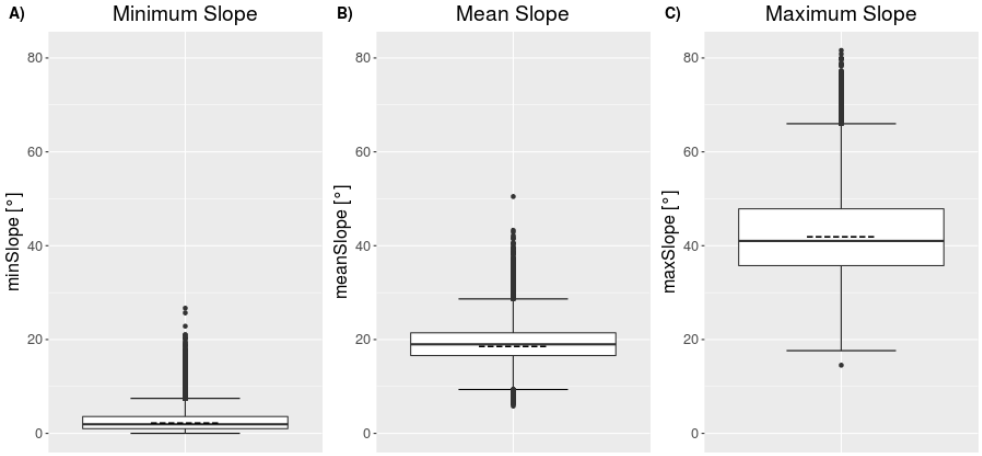


Figure 5.14: *Boxplots of attributes A) minSlope, B) meanSlope, and C) maxSlope slope.*

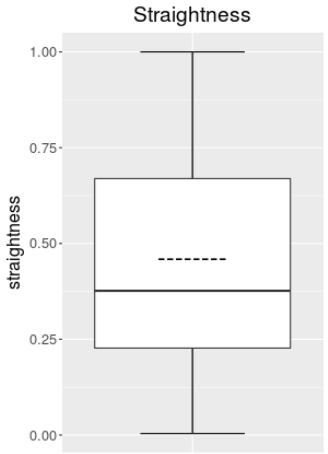


Figure 5.15: *Boxplot of attribute straightness.*

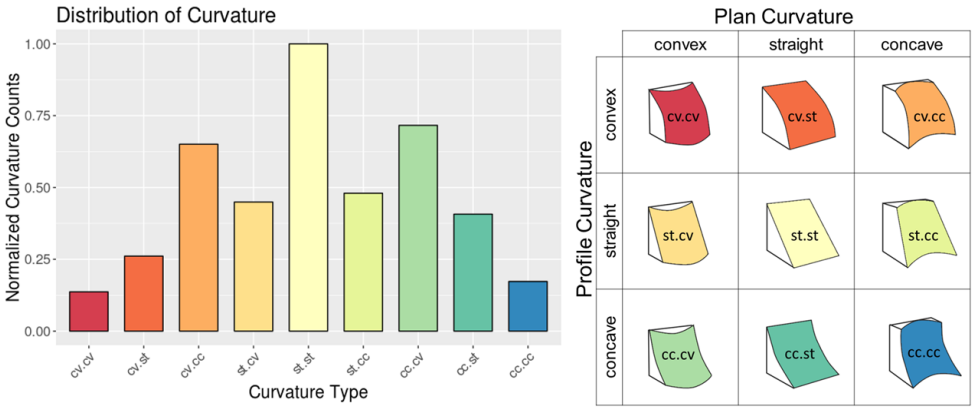


Figure 5.16: *Bar plot of normalised curvature frequencies after the reclassification. Values were reclassified by the maximum frequency. Figure after Dikau (1989) and Vontobel (2011).*

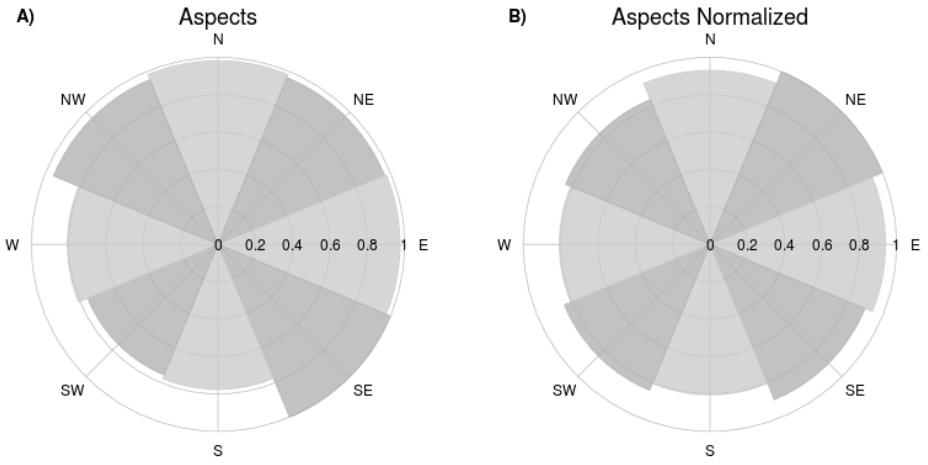


Figure 5.17: *Pie plots of summarised frequencies of A) aspects normalised with maximum frequency and B) aspects normalised with number of pixels of each class and maximum frequency.*

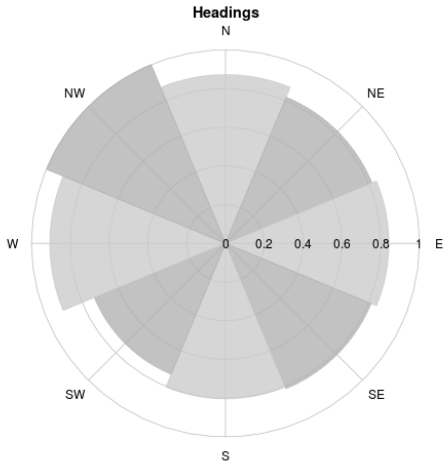


Figure 5.18: *Pie plots of summarised frequencies of headings.*

5.2.1 Influence of Raster Resolution

Table 5.6 provides an overview of the results from the Wilcoxon signed-rank test for the attributes extracted from the two rasters with varying resolution (DHM25 & SwissALTI3D). It includes mean values, p-values and two effect measures, Cohen's R and D. For all study areas, mean minimum values decrease with higher DEM resolution and mean maximum values increase with higher DEM resolution. These differences are introduced by the smoothing when the resolution is reduced. All height differences range between 1 and 4 meters. Differences of maximum slope are especially large in all study areas, with values up to 13°.
To check the statistical significance of the attribute differences between the two rasters, the following H_0 was used for the Wilcoxon signed-rank test: The rank-sums of the differences between attributes extracted from DHM25 and attributes extracted from SwissALTI3D follow a symmetric distribution around zero. The p-value is less than 0.001 for all attributes in all study areas. Thus, the rank-sums of the differences do not follow a symmetric distribution around zero and suggests a rejection of H_0. This means that there exist statistically significant differences in the ranks of the attributes extracted from DEMs of varying resolution.
Both effect sizes are generally lower for height attributes than for slope values. Cohen's R for height attributes ranges from medium to large in the study areas St. Antönien and Wildstrubel. In the study area Urserental, all height attributes show a large Cohen's R. Thus, there is a directional shift in the height attributes extracted from the two rasters. However, Cohen's D for all height attributes is very small and, thus, the shift is small considering the standard deviations. Both effect sizes for most slope attributes are medium to large in all study areas. An exception are minimum and mean slopes in the St. Antönien study area, where Cohen's D is only small to medium.

Even though there is a difference in all attributes, the mean values, especially for height attributes, do not strongly differ considering their range. This is highlighted by the low Cohen's D values. This behaviour is illustrated by the following boxplots of the mean height and slope attributes from both DEMs (fig. 5.19). The shift for height attributes is hardly perceivable but more evident for slope attributes. The extrema values (minimum / maximum) are affected most by varying resolutions.
The boxplots of the differences between the attributes, extracted from the two rasters, further illustrate this behaviour (fig. 5.20). The differences of mean height attributes scatter rather strongly. Means, medians and quartiles of the differences are slightly positive for mean height. Thus, mean heights extracted from DHM25 are slightly larger than those extracted from SwissALTI3D and there is a small but constant shift as suggested by the statistical analysis. The differences of the mean slopes show less scattering. Means, medians and quartiles of the mean slope differences are all negative and only few outliers are positive. Thus, the mean slope values extracted from DHM25 are generally smaller than those extracted from SwissALTI3D. The shift is constant and stronger than for mean height values. This supports the large Cohen's R values, which were acquired during the statistical analysis. The effect can be expected to be stronger for extrema values.
These findings are likely to hold for the full dataset as well. Thus, the DEM resolution has an influence on the extracted values, which is generally low for height attributes but higher for derivatives thereof, as was suggested by the literature.

Attribute	Mean		P-Value	Cohen's	
	DHM25	SwissALTI3D		R	D
St. Antönien					
minHeight	1506.54 m	1505.47 m	< 0.001	0.370	0.003
meanHeight	1862.00 m	1858.21 m	< 0.001	0.336	0.008
maxHeight	2368.98 m	2372.41 m	< 0.001	-0.669	0.008
minSlope	2.74 °	1.33 °	< 0.001	0.752	0.400
meanSlope	17.12 °	18.45 °	< 0.001	-0.859	0.273
maxSlope	36.83 °	44.38 °	< 0.001	-0.850	1.172
Urserental					
minHeight	1722.81 m	1720.24 m	< 0.001	0.648	0.008
meanHeight	2230.24 m	2224.86 m	< 0.001	0.621	0.022
maxHeight	2842.67 m	2845.64 m	< 0.001	-0.536	0.010
minSlope	2.18 °	0.87 °	< 0.001	0.797	0.960
meanSlope	18.09 °	19.93 °	< 0.001	-0.864	0.779
maxSlope	41.10 °	54.21 °	< 0.001	-0.861	1.749
Wildstrubel					
minHeight	1999.18 m	1998.18 m	< 0.001	0.238	0.009
meanHeight	2337.28 m	2335.50 m	< 0.001	0.355	0.030
maxHeight	2857.64 m	2860.40 m	< 0.001	-0.453	0.017
minSlope	0.97 °	0.53 °	< 0.001	0.684	0.807
meanSlope	15.70 °	16.75 °	< 0.001	-0.851	0.523
maxSlope	43.82 °	55.02 °	< 0.001	-0.816	0.992

Table 5.6: *Results of Wilcoxon signed-rank test for route attributes extracted from DHM25 and SwissALTI3D of all small study areas.*

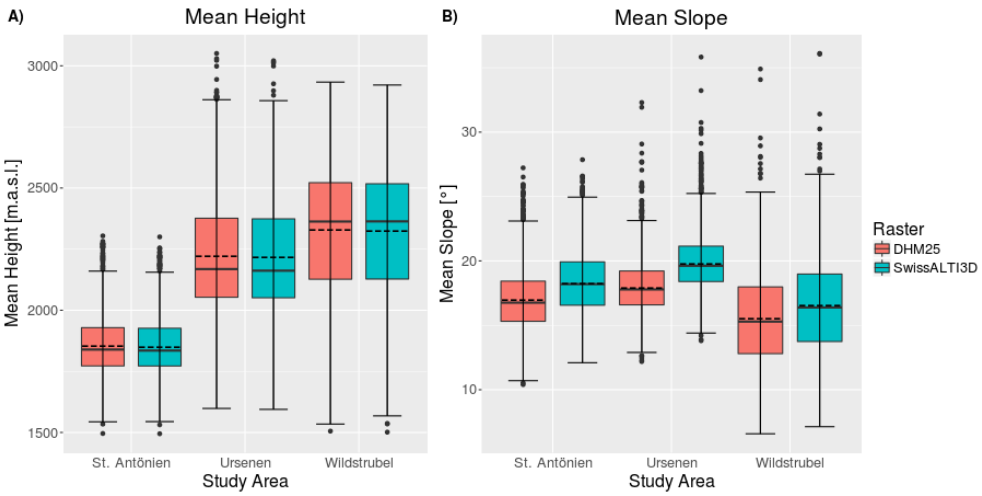


Figure 5.19: *Boxplots of attributes A) meanHeight (DHM25) / meanHeight (SwissALTI3D), B) meanSlope (DHM25) / meanSlope (SwissALTI3D) for all small study areas.*

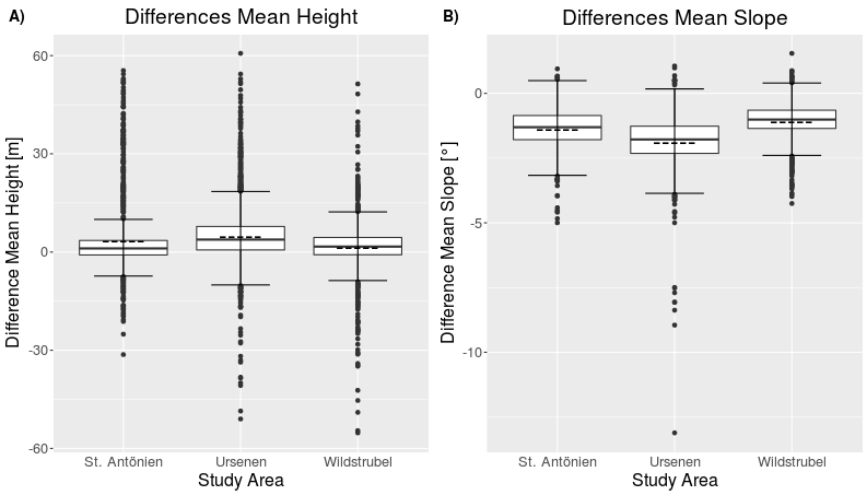


Figure 5.20: *Boxplots of differences between attributes, extracted from the two rasters A) meanHeight (DHM25) - meanHeight (SwissALTI3D), B) meanSlope (DHM25) - meanSlope (SwissALTI3D) for all small study areas.*

5.2.2 Influence of Participation Inequality

The creation of subsets based on user groups (rare & heavy users) allowed checking whether rare and heavy users planned routes with differing route attributes. Table 5.7 provides an overview of mean values for all attributes of both user groups, p-values and Cohen's R and D as effect measures.
It should be noted that, at the mean values, heavy users plan longer routes in higher detail (with smaller step lengths). Furthermore, routes from heavy users are usually higher than those from rare users. Attributes from raster derivatives seem to be less affected than intrinsic and height attributes. To check the statistical significance of these differences, the following H_0 was used for the Mann-Whitney U test: The rank-sums for the attributes do not differ between routes from rare and heavy users. The p-value was less than or equal to 0.001 for all attributes except straightness. For all these attributes, the p-value is thus less than 0.05, which suggests a rejection of H_0. This indicates that there exists a constant one-sided shift for all attributes, except for straightness. However, the first effect measure, Cohen's R, is very small to small for all attributes. The largest Cohen's D values are observable for length, step length, number of vertices and height attributes, ranging between a small and medium size. This supports the previously-made observation of mean differences. This behaviour is further illustrated in the boxplots of those attributes in figure 5.21 and 5.22.
Derived attributes, namely slope and curvature attributes, generally show very small to small values for Cohen's D. Maximum plan curvature and minimum profile curvature are the derived attributes showing the largest values for Cohen's D. Nevertheless, these can still be classified as small. The comparison of the straightness attribute for the two user groups resulted in a p-value of 0.7, which is distinctly larger than 0.05. Hence, H_0 cannot be rejected for the straightness attribute and there seems to be no difference in the rank-sums. This is further underlined by the low effect sizes. Additional boxplots of the remaining attributes can be found in appendix E.

Before comparing the avalanche danger ratings from routes of rare and heavy users, the sample size needs to be considered. As stated in section 5.1.4, not all users specified an avalanche danger level for their route. Thus, only routes including an avalanche danger can be compared, which influences the size of the two samples. In

Attribute	Mean		P-Value	Cohen's	
	Rare Users	Heavy Users		R	D
length	6677.99	7860.10	< 0.001	-0.200	0.301
numberOfVertices	36.83	53.33	< 0.001	-0.360	0.352
stepLength	276.60	210.67	< 0.001	0.274	0.429
minHeight	1510.54	1635.54	< 0.001	-0.153	0.245
meanHeight	1970.76	2131.60	< 0.001	-0.216	0.325
maxHeight	2488.46	2710.06	< 0.001	-0.272	0.387
heightDifference	977.92	1074.52	< 0.001	-0.163	0.241
minSlope	3.00	2.49	< 0.001	0.124	0.212
meanSlope	19.71	19.08	< 0.001	0.133	0.160
maxSlope	42.27	43.04	< 0.001	-0.052	0.082
minPlanCurv	-3.04	-3.12	< 0.001	0.046	0.053
meanPlanCurv	0.04	0.06	< 0.001	-0.060	0.128
maxPlanCurv	3.90	4.48	< 0.001	-0.176	0.255
minProfCurv	-4.29	-5.03	< 0.001	0.192	0.234
meanProfCurv	-0.01	-0.03	< 0.001	0.051	0.127
maxProfCurv	2.85	3.06	< 0.001	-0.086	0.109
straightness	0.46	0.45	0.700	0.005	0.036
danger_level	2.46	2.31	0.026	0.029	-

Table 5.7: *Results of Mann-Whitney U test for route attributes of rare and heavy users. For units, refer to table 5.5.*

the rare user sample, 2'847 routes included an avalanche danger, which corresponds to 50.3 % of all routes in the sample. In contrast, the heavy user sample only included an avalanche danger level for 540 routes, which corresponds to 9.4 % of all routes in the sample. Nevertheless, the difference in sample size should not affect the Mann-Whitney U test, as there are enough samples in both subsets. As the avalanche danger level is on an ordinal scale level, the computation of Cohen's D is meaningless and was therefore omitted. The p-value for the comparison of avalanche danger levels between the rare and the heavy user group was 0.026, suggesting a rejection of H_0. The value of Cohen's R is small and, therefore, the effect is not strong. Nevertheless, the percental distribution of avalanche danger levels for the two user-groups, as can be seen in figure 5.23, suggests a considerable difference in the distributions between rare and heavy users. The avalanche danger levels specified by heavy users seem to follow the distribution of the complete dataset more closely than those specified by rare users.

It is noteworthy that all route attributes from routes of heavy contributors in general are more similar to the attributes of the full dataset compared to attributes from routes of rare contributors (cf. table 5.5). This holds for all attributes, and the possible implications of this behaviour will be addressed in the discussion.

5.2.3 Influence of Avalanche Conditions

It is expected that users adjust their planned routes according to the prevalent avalanche danger. The higher the avalanche danger, the higher the risk of being caught in an avalanche. Thus, users should plan routes in terrain which is less exposed to avalanche risk. As mentioned in the previous section (5.1.4), 15'231 of the filtered routes included an avalanche danger rating. This allowed the assessment of the influence of prevalent avalanche danger on route attributes. All of these

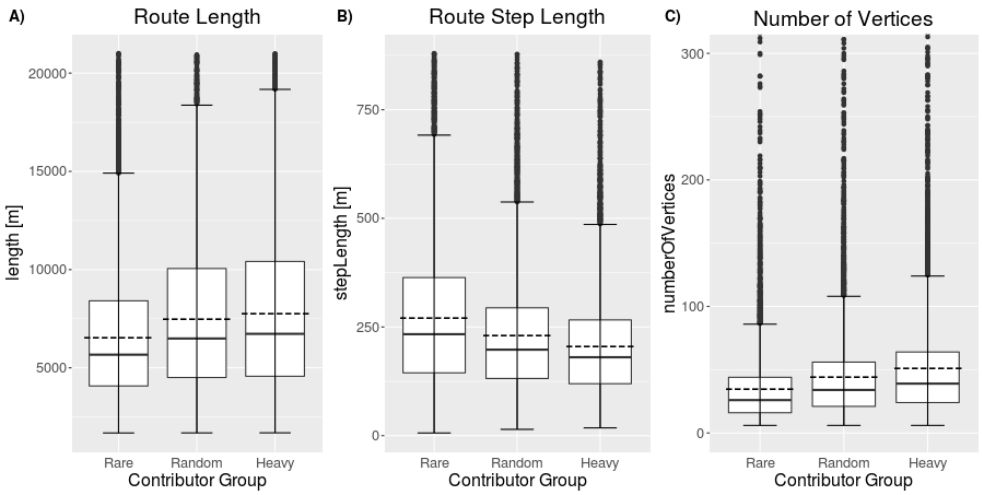


Figure 5.21: *Boxplots of attributes A) length, B) stepLength, and C) numberOfVertices for all three user-groups.*

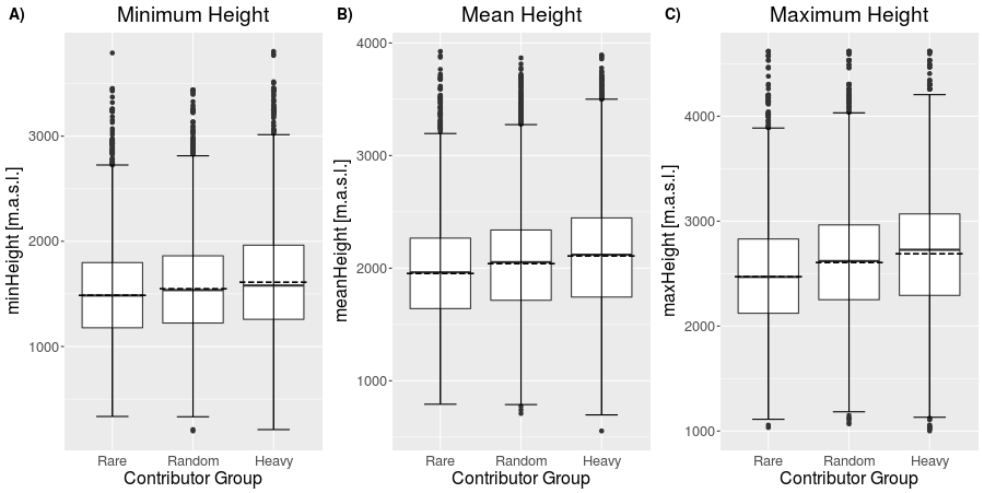


Figure 5.22: *Boxplots of attributes A) minHeight, B) meanHeight, and C) maxHeight for all three user-groups.*

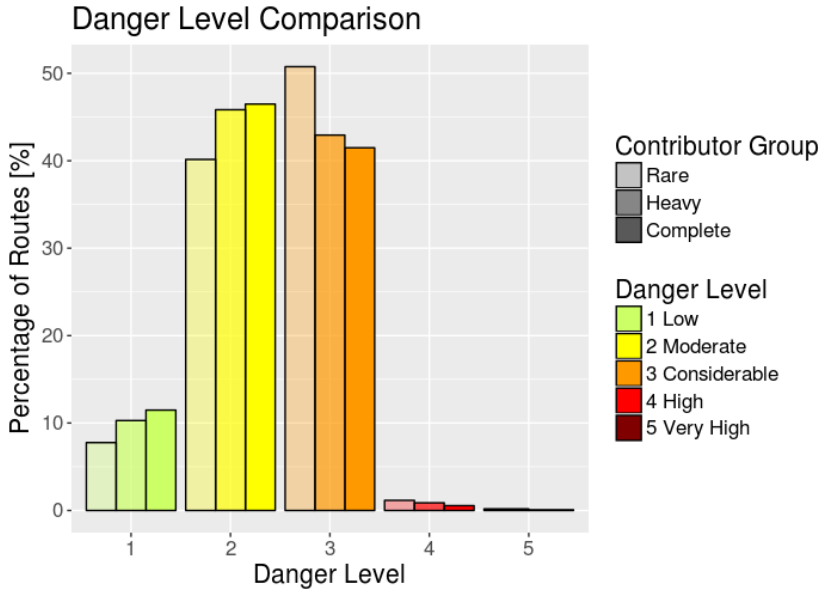


Figure 5.23: *Normalised distribution of user-specified avalanche danger level for rare and heavy users.*

were used to calculate the correlation coefficients between the avalanche danger and various attributes. H_0 for the correlation tests was as follows: The correlation coefficient is close to 0, indicating that there is no correlation between the variable and avalanche danger. As is displayed in table 5.8, the resulting p-values from the correlation tests are below 0.05 for all attributes. This suggests a rejection of H_0 and indicates a significant correlation between avalanche danger and every attribute. Considering the absolute τ-values, all attributes only show a weak correlation. The highest p-values correspond with the lowest τ-values, suggesting that there is the least significant correlation for step length, mean profile curvature and straightness. The strongest correlations would be expected for attributes which are directly related to avalanche danger, such as maximum slope, minimum and maximum curvature attributes and height attributes. All these values show values around $\pm$ 0.1, which is still rather small. However, the low τ-values are likely a result of the strong spread of the attributes, which affects the correlation coefficient. Thus, it can be said that users tend to avoid areas with high elevations, large slopes and curvature values.

Table 5.8: *Results of correlation tests including Kendall's τ. For units, refer to table 5.5.*

Attribute	**P-Value**	τ
length	< 0.001	-0.061
numberOfVertices	< 0.001	-0.057
stepLength	0.006	0.017
minHeight	< 0.001	-0.047
meanHeight	< 0.001	-0.070
maxHeight	< 0.001	-0.096
heightDifference	< 0.001	-0.059
minSlope	< 0.001	0.042
meanSlope	< 0.001	-0.032
maxSlope	< 0.001	-0.116
minPlanCurv	< 0.001	0.083
meanPlanCurv	< 0.001	0.062
maxPlanCurv	< 0.001	-0.098
minProfCurv	< 0.001	0.093
meanProfCurv	0.001	-0.021
maxProfCurv	< 0.001	-0.086
straightness	0.010	-0.016

Mean and maximum attributes of slope and plan curvature are visualized in figures 5.24 and 5.25. Visualisations of other attributes can be found in the appendix D. Figure 5.24 shows that extrema attributes are more affected than mean values. This corresponds to τ-values, with the exception of minimum height. The boxplots (fig. 5.24a & 5.24c) show a clear negative trend of both slope attributes with increasing avalanche danger. Therefore, the higher the danger level, the less steep mean and maximum slope. This is further highlighted by the linear models which are visualised in the scatterplots (fig. 5.24b & fig. 5.24d), whereas the trend is clearer for the maximum slope. Boxplots for the attributes with a danger level of 5 show a contra-linear trend. However, the number of routes planned for very high avalanche danger is small and should not be overrated. Minimum slope values are included in the appendix (fig. D.6).
For curvature attributes, both minimum and maximum values are of special interest. Both their boxplots show a clear trend towards zero (fig. 5.25a & 5.25c). This is supported by the linear models visualised in the scatterplots (fig. 5.25b & 5.25d). Thus, with increasing avalanche danger, routes tend to avoid heavily convex or concave

terrain. Profile curvature behaves accordingly, which can be seen in the visualisations included in the appendix (fig. D.9). Again, curvature attributes of routes with danger level 5 exhibit suspicious behaviour, which is likely an effect of the small size of the sample with those avalanche danger levels.

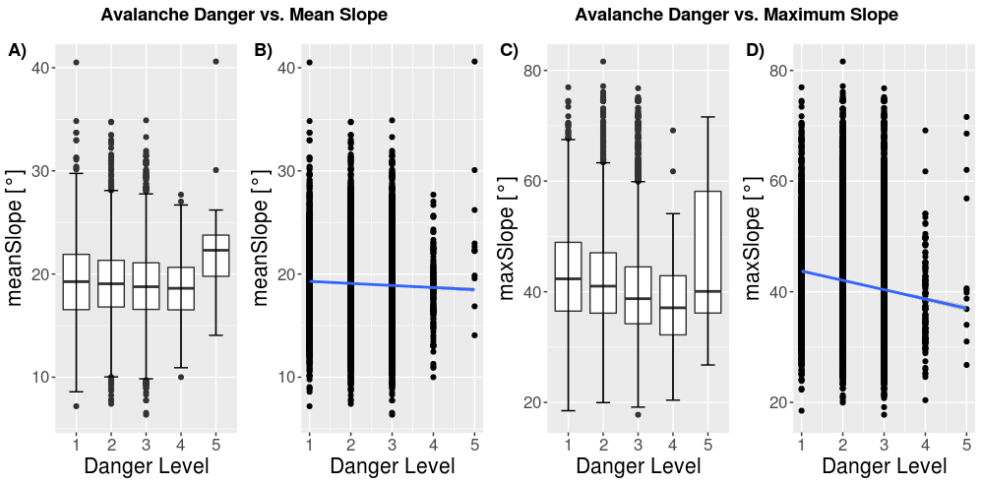


Figure 5.24: *A) Boxplot of mean slope, grouped by avalanche danger level. B) Scatterplot of mean slope and avalanche danger level, including linear model. C) Boxplot of maximum slope, grouped by avalanche danger level. D) Scatterplot of maximum slope and avalanche danger level, including linear model.*

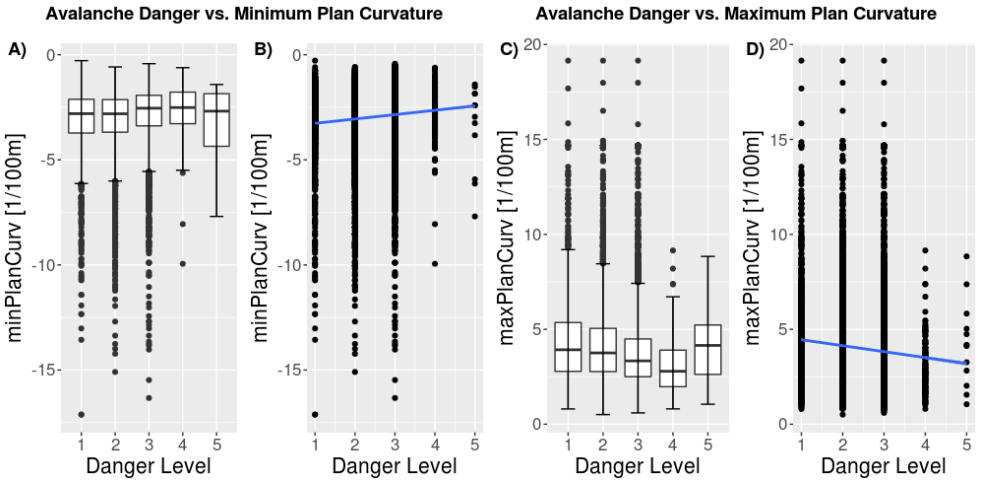


Figure 5.25: *A) Boxplot of minimum plan curvature, grouped by avalanche danger level. B) Scatterplot of minimum plan curvature and avalanche danger level, including linear model. C) Boxplot of maximum plan curvature, grouped by avalanche danger level. D) Scatterplot of maximum plan curvature and avalanche danger level, including linear model.*

As described in section 4.3.3, the routes were grouped based on their danger level in order to assess the strength of the attribute variations. All attribute means, grouped by danger level and Cohen's D, are displayed in table 5.9. The results correspond well to those obtained in table 5.8. Step length, mean profile curvature and straightness show the smallest values for Cohen's D. The largest values for Cohen's D can be observed for maximum height and maximum slope. Nonetheless, these effect measures are still classified as small. Again, extrema values (minimum / maximum) of DEM derivatives show larger values for Cohen's D than the mean values.
Generally, the inspection of mean values grouped by avalanche danger and the Cohen's D values support the findings from the correlation tests. Thus, with higher avalanche danger, users plan routes in lower elevations with less extreme slope and curvature. It is noteworthy that mean attributes for low avalanche danger correspond more closely to the attributes of the full dataset (cf. table 5.5). However, only about 27 % of all routes have been considered for the analysis of the effect of avalanche danger on route attributes. Thus, the remaining 73 % has a rather strong

impact on mean values.

Attribute	Mean		Cohen's D
	Low Danger	High Danger	
length	7451.63	6849.23	0.162
numberOfVertices	46.61	41.95	0.114
stepLength	225.00	228.73	0.027
minHeight	1522.58	1466.44	0.128
meanHeight	2017.42	1934.99	0.195
maxHeight	2595.39	2463.72	0.267
heightDifference	1072.81	997.29	0.202
minSlope	2.61	2.87	0.110
meanSlope	19.15	18.87	0.080
maxSlope	42.46	40.23	0.254
minPlanCurv	-3.11	-2.82	0.200
meanPlanCurv	0.03	0.05	0.147
maxPlanCurv	4.21	3.79	0.198
minProfCurv	-4.76	-4.27	0.158
meanProfCurv	-0.01	-0.02	0.032
maxProfCurv	2.93	2.63	0.167
straightness	0.49	0.48	0.025

Table 5.9: *Attribute means of all attributes grouped by danger level and Cohen's D. Group 'low danger' including routes with danger levels 1 & 2 and group 'high danger' including routes with danger levels 3 – 5. For units, refer to table 5.5.*

5.3 Similarity & Clustering

To compare the inter-object similarities from all planned routes with the similarities between planned routes and GPS-tracks, the Fréchet distance was calculated, as described in section 4.4. This allowed putting the routes into context and also served as input for the subsequent clustering of routes. The GPS-tracks will be denoted as 'tracks' for the remainder of this work.

5.3.1 Similarity

Figure 5.26 provides an overview of the kNN and kNG distances. As discussed in section 4.4, these are the distances to the k nearest routes and tracks, respectively. Naturally, kNN-distances increase with larger k-values. However, in all three study areas, the distances to the 1, 3 and 5 nearest neighbour-routes are less than 1 km for most routes (> 75 %). The median is lower than the mean for all kNN-distances, indicating a positively skewed distribution. Generally, kNN-distances seem to be similar for all three study areas, whereas the interquartile range of kNN-distances in Wildstrubel is larger than in the other two study areas. The number of routes in all three subsets, as shown in section 5.1, is comparable. The Wildstrubel and Urserental study areas share the same size, being almost 1.5 times larger than the St. Antönien study area. Thus, St. Antönien should show the lowest kNN distances as the route density is expected to be higher. Nevertheless, the kNN-distances in St. Antönien are similar to Urserental. The stronger dispersion of kNN-distances in Wildstrubel seems to be inherent.

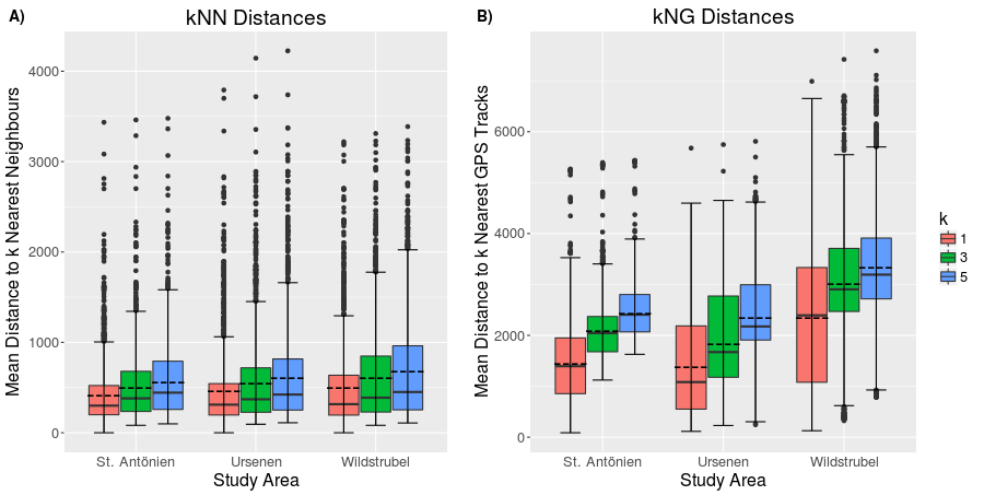


Figure 5.26: *A) Boxplots of distances to 1, 3 and 5 nearest neighbours for all small study areas. B) Boxplots of distances to 1, 3 and 5 nearest GPS-tracks.*

Similar to the kNN-distances, kNG-distances increase with larger k. kNG-distances are generally greater than kNN-distances. Routes in Wildstrubel exhibit larger kNG-distances than routes in Urserental and St. Antönien. For routes in St. Antönien and Wildstrubel, mean and median of the kNG-distances are closer, indicating a less-skewed distribution. This is not the case for routes in Urserental. Furthermore, the interquartile ranges of all kNG-distances tend to be larger than for kNN-distances. This behaviour is not as strong for routes in St. Antönien as it is for routes in Urserental and Wildstrubel. The limited number of tracks is a possible source of the variations in dispersions between kNN-distances and kNG-distances. The variations in dispersion between study areas, however, does not seem to be a result of the varying number of tracks. Hence, other effects are likely to influence this behaviour and will be addressed in the discussion.

5.3.2 Clustering

The application of DBSCAN on the generated similarity matrices resulted in up to 29 clusters. Hence, the visualisation proved rather difficult. For this reason, two peaks were selected as examples to illustrate the clustering results: Stotzigen Firsten in Urserental and Schafberg in St. Antönien. Table 5.10 provides an overview of the number of clusters found using ε-values between 400 m and 1'200 m, as well as the number of outliers and percentage of those. Visualisations of the clusters will be presented in the next two paragraphs.

Stotzigen Firsten On Stotzigen Firsten, there is a total of 330 routes. These are visualised in figure 5.27.1. Some clear outliers are discernible in the unclustered routes, such as the route heading over the Tiefen Glacier north of Stotzigen Firsten. It is apparent that this route is not a typical tour in this area.
Setting ε to 400 m (fig. 5.27.2, DBSCAN is able to identify a single cluster A. This cluster includes 157 routes starting from Realp and finishing on Stotzigen Firsten. The cluster is rather dense for large parts of the route, showing some minor variations near the peak, as can be seen in figure 5.27.2. The single cluster contains almost 50 % of all routes. Hence, it is safe to declare it a major route of Urserental.
By increasing ε to 600 m (fig. 5.27.3), cluster A gains several additional routes which

ε	Clusters	Outliers	Outliers Percentage [%]
Stotzigen Firsten			
Total	-	330	100
400 m	1	173	52.4
600 m	3	98	29.7
800 m	4	70	21.2
1'000 m	3	53	16.1
1'200 m	3	35	10.6
Schafberg			
Total	-	400	100
400 m	4	280	70.0
600 m	5	233	58.3
800 m	7	177	44.3
1'000 m	6	121	30.3
1'200 m	3	57	14.3

Table 5.10: *Overview of clusters and outliers for clustering with DBSCAN for routes to Stotzigen Firsten in Urserental and to Schafberg in St. Antönien.*

more strongly vary and were previously classified as outliers. Furthermore, two additional clusters are found. Cluster B includes routes from Realp to Stotzigen Firsten and back to Realp. These routes follow a similar path in both directions. Additionally, there is a cluster C, including routes from Realp to Stotzigen Firsten and back which descend on the northern slope of Stotzigen Firsten. Overall, there are more than 20 % fewer routes classified as outliers compared to DBSCAN with 400 m as ε.
DBSCAN with 800 m as ε (fig. 5.27.4) is able to identify yet another cluster D, including routes from Realp to Stotzigen Firsten, continuing over the Mutten Glacier to the Rotondo SAC cabin. The clusters A to C remain, each including some additional routes with larger variations. The number of routes classified as outliers is reduced by an additional 8.5 %.
With ε set to 1'000 m (fig. 5.27.5), DBSCAN merges clusters B and C. Thus, routes following the same path up and down and routes taking the northern slope back to Realp were not differentiated. Clusters A and C each gain some additional routes. This leads to a reduction of routes classified as outliers by about 5 %.
By increasing ε to 1'200 m, the number of clusters is unchanged (fig. 5.27.6). However, all clusters include some additional routes. The most apparent change is in cluster D, which now includes routes which ascend to Muttenhorn after Stotzigen Firsten, before finishing at the Rotondo cabin. Again, the number of outliers is reduced by more than 5 %.

In general, DBSCAN is able to detect major backcountry routes on Stotzigen Firsten. A ε-value of 800 m provides the best results, including the most important clusters and removing 21.2 % of the routes as outliers. If ε is set lower, not all clusters can be detected. If ε is set higher, two distinct clusters are merged and some clear outliers are included in the clusters. A reduction of ε to 600 m is conceivable, if minPts is reduced as well. This might allow the detection of additional clusters.

Schafberg There exist 301 routes going to Schafberg. These also include routes to Girenspitz, as a clear distinction between the routes is not possible. Additionally, 99 routes to Sulzfluh were included, as they were clustered together with routes to Schafberg for large ε-values. Figure 5.28.1 illustrates the unclustered routes to Schafberg, Girenspitz and Sulzfluh. Again, there are some clearly distinguishable outliers

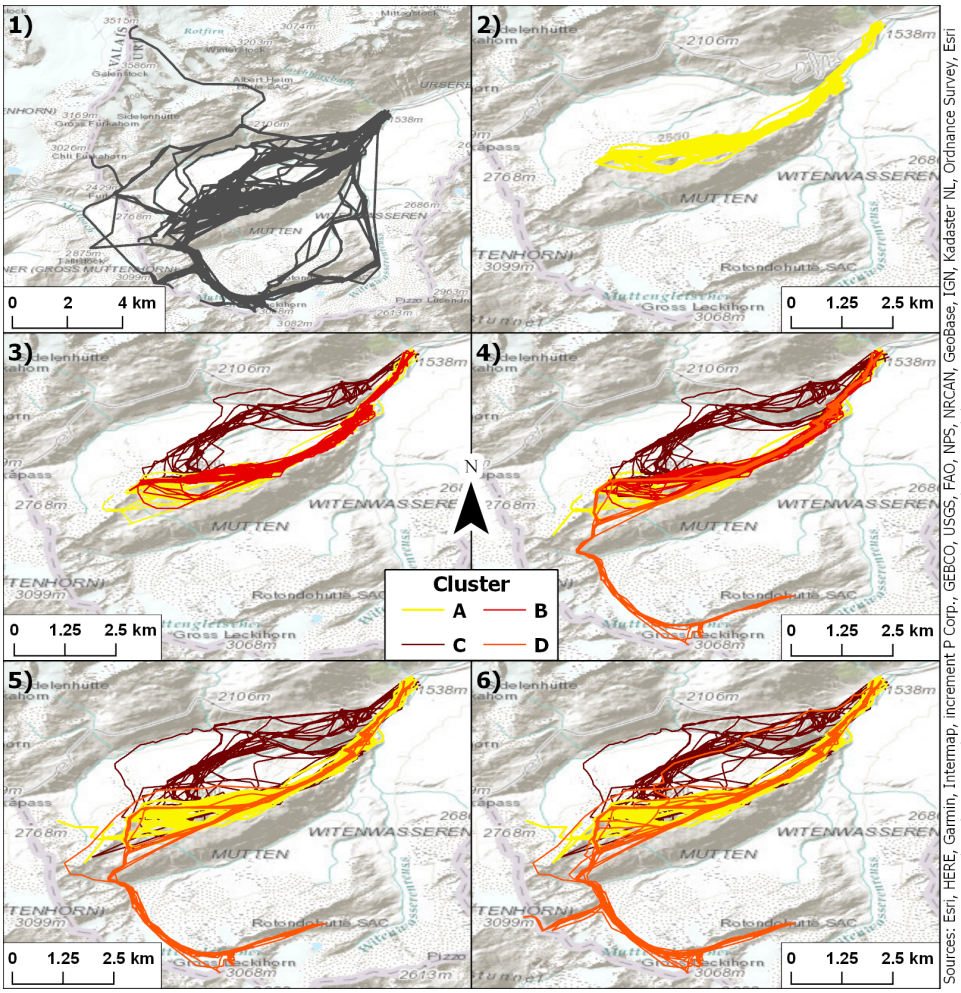


Figure 5.27: *1) All routes on Stotzigen Firsten, visualisation of result of DBSCAN with ε of 2) 400 m, 3) 600 m, 4) 800 m, 5) 1'000 m, and 6) 1'200 m. MinPts was set to 10 for all results.*

to the east and west of Schafberg.

With the minimal ε-value of 400 m, four clusters are found. Two of those clusters, A and B, finish at Girenspitz. Cluster A includes longer routes, starting in St. Antönien, and cluster B includes routes starting in Untersäss. Clusters G and H both end up on Sulzfluh, whereas routes in cluster G start in St. Antönien and routes in cluster H in Partnun. All clusters are relatively dense, showing only minor variations. 70 % of all routes are classified as outliers.

With ε set to 600 m, cluster B is merged with cluster A, leaving only a single cluster of routes to Girenspitz. Two additional clusters, C and D, are found. Both of them include routes to Schafberg, whereas routes in cluster C start in St. Antönien and routes in cluster D go from Partnun. Clusters G and H remain, each including some additional routes with larger variations. The percentage of outliers is reduced to 58.3 %.

By increasing ε by an additional 200 m to 800 m, two additional clusters are found. Cluster B now includes routes from St. Antönien to Girenspitz and back to St. Antönien. Most routes follow a similar path in both directions. Furthermore, a cluster E is found, which includes routes from St. Antönien to Schafberg and back. This cluster shows rather large variations, also including routes that go to Girenspitz and Schafberg. Clusters A, C, D, G, and H, persist, all including some additional routes. The increase of ε leads to a reduction of outliers by 14 %.

When ε is further increased to 1'000 m, cluster C is merged with cluster A. Cluster A then includes various routes to Girenspitz and Schafberg, starting at several locations. Cluster E is merged with cluster B, which now includes various round-trip routes from St. Antönien to Girenspitz and/or Schafberg and back. A new cluster F is found, which includes round-trip routes from Untersäss and Partnun to Schafberg. Clusters G and H are mostly unaffected, only including some additional routes. The outlier percentage is reduced by an additional 14 %.
With 1'200 m as ε, clusters D, G and H are merged with cluster A. Cluster A then includes routes from St. Antönien, Untersäss and Partnun to Girenspitz, Schafberg and Sulzfluh. The inner-cluster variations are large. Clusters B and F gain some additional routes. The percentage of outliers is reduced by an additional 16 %.

Again, DBSCAN performs well with ε-values up to 800 m. If ε is set higher than 800 m, various clusters are merged and variations in the clusters are too large. Furthermore, the merged clusters look rather chaotic. However, only with 400 m as ε, DBSCAN is able to identify two clusters of routes to Girenspitz. Clustering of routes to Schafberg and nearby peaks proved rather difficult, compared to Stotzigen Firsten. Inner-cluster variations seem to be larger and more clusters are merged. Reasons for this will be addressed in the discussion. Nevertheless, these results prove the feasibility of a route clustering based on the Fréchet distance with DBSCAN, but underline the importance of parameter settings.

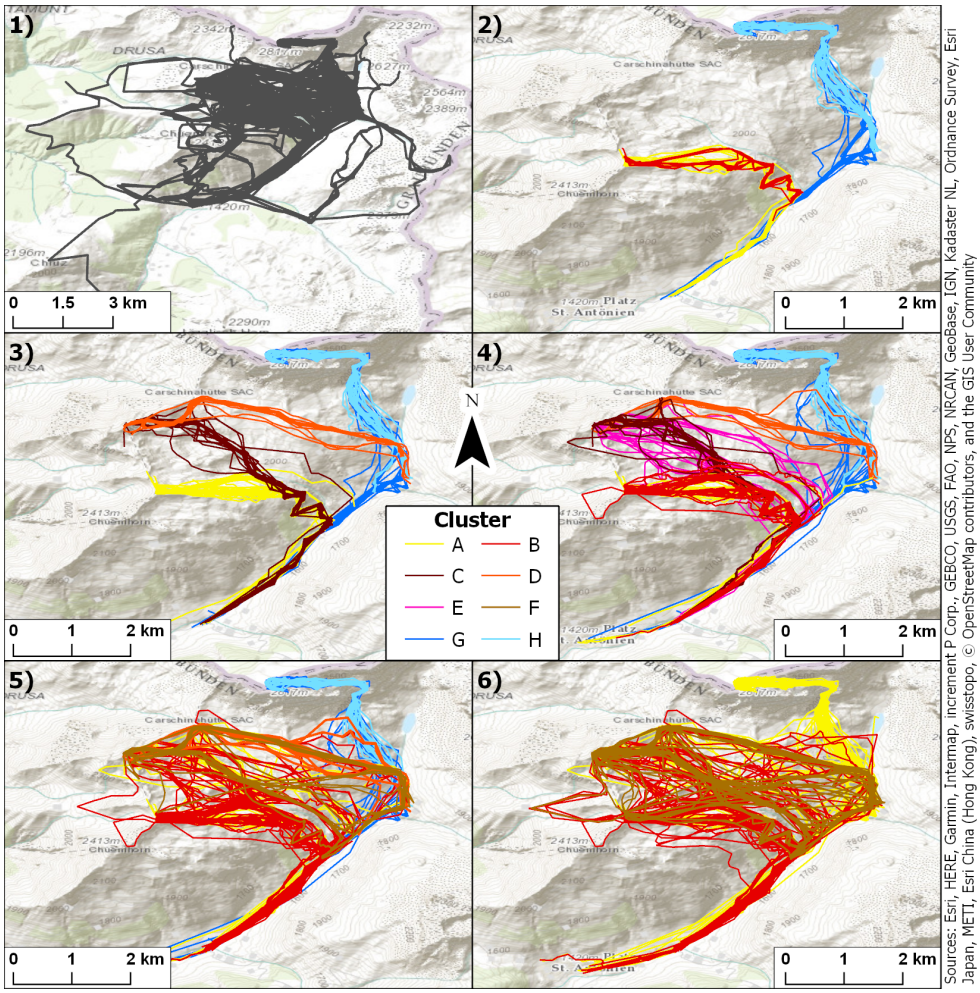
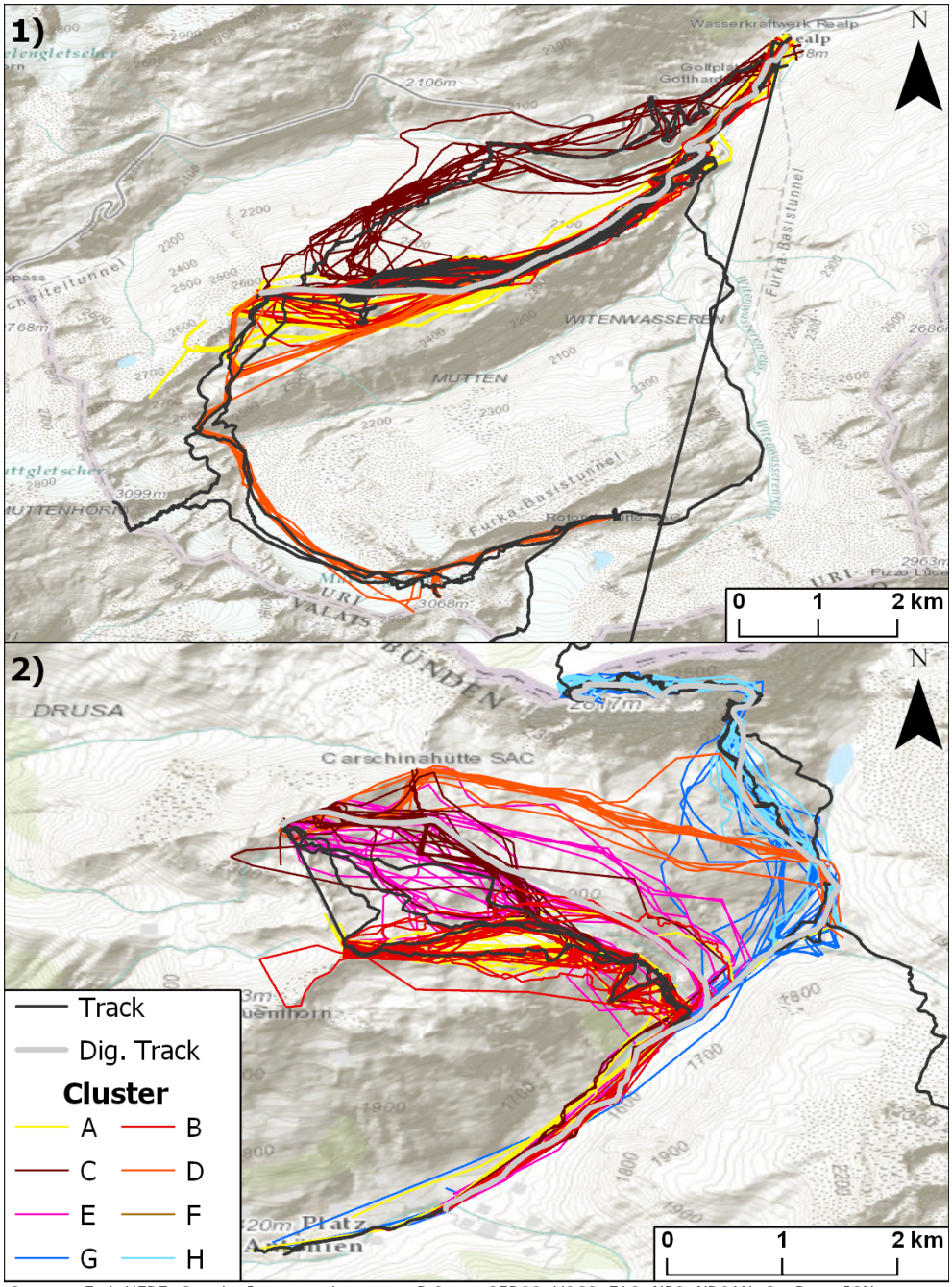


Figure 5.28: *1) All routes on Schafberg, visualisation of result of DBSCAN with ε of 2) 400 m, 3) 600 m, 4) 800 m, 5) 1'000 m, and 6) 1'200 m. MinPts was set to 10 for all results.*

Comparison Clusters & GPS-Tracks For the visualisation with tracks, clustering results with 800 m as ε from Stotzigen Firsten and Schafberg were selected. As mentioned in section 3.1.4, there exist digitised routes in the track dataset. These are visualised in light grey (Dig. Track), whereas assumingly 'real' tracks are visualised in dark grey (Track).
On Stotzigen Firsten (figure 5.29.1), clusters and tracks correspond well. There exist tracks for all identified major backcountry routes. One of the tracks continues to Muttenhorn after Stotzigen Firsten. This track corresponds well to some routes in Cluster D from figure 5.27.6. There exists no cluster following the track which continues from the Rotondo SAC-cabin to Realp. The track has a length of over 25 km. Therefore, similar planned routes were likely filtered in the preprocessing. One track contains an anomaly, which is still included due to the lack of preprocessing of the tracks. The digitised SAC-tour is easily identified by its smooth turns. Routes from cluster A, as identified in figure 5.27.1, seem to follow the only digitised track closely.
On Schafberg, tracks vary more than on Stotzigen First. This corresponds to the findings as presented in the previous paragraph. There exist tracks that correspond with clusters A, B, C, and E to Girenspitz and Schafberg. The tracks to Sulzfluh are similar to routes in clusters G and H. No track follows planned tours in cluster D. Various planned routes follow digitised SAC-tours closely. These form part of clusters E and G / H. There was no digitised SAC-tour available with high similarity to other clusters.
In general, the similarities between planned tours and tracks seem weaker on Schafberg than on Stotzigen Firsten. This contradicts the results presented in section 5.3.1. However, only a subset of the result is analysed and therefore these subsets seem to be special cases. This is underlined by the comparison of kNG-distances in figure 5.30.



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

Figure 5.29: *Comparison between route clusters acquired with ε-value of 800 m and GPS-Tracks on 1) Stotzigen Firsten and 2) Schafberg*

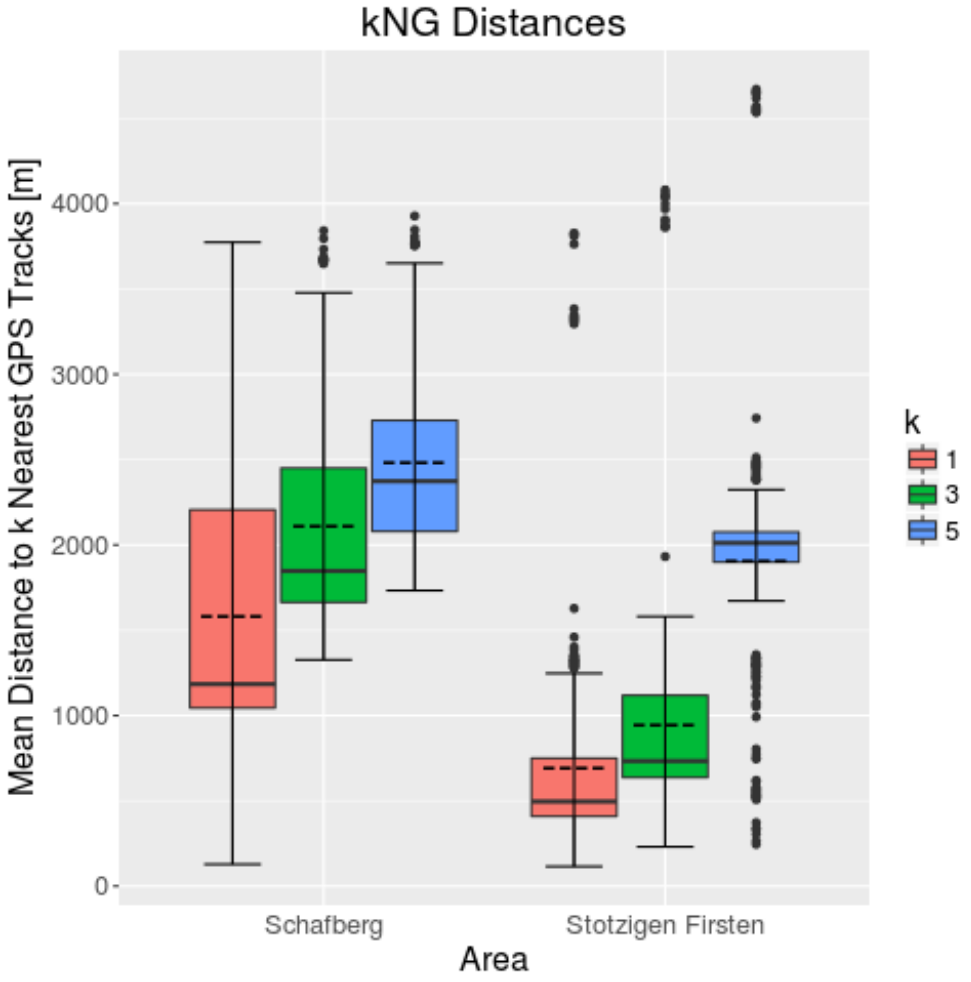


Figure 5.30: *Comparison of nearest GPS-track distances between routes from Stotzigen Firsten and Schafberg*

Chapter 6

Discussion

In the following chapter, the results achieved, as presented in the previous chapter 5, will be put into context. This should show, if the results are sensible and increase their impact. In the first three sections (sec. 6.1 - 6.3), the results from this thesis will be compared to findings in related literature. Comparisons to existing literature should provide further insights into the results and allow to provide in-depth explanations for unexpected findings. This will build the groundwork to answer the research questions of this thesis and lead to the achievement of the overall aim stated in the introduction. Then, the research questions, as formulated in subsection 2.5, will be addressed (sec. 6.4). The research questions should be answered based on the insights gained and results achieved in this thesis. Issues that were encountered and limitations of this thesis will be discussed in the last section (sec. 6.5).

6.1 General Dataset Overview

Temporal Distribution The number of routes planned per year experienced a drastic drop between seasons 2015/16 and 2016/17, as was illustrated in figure 5.1. It stands to reason, that a new tool, as White Risk 2.0 was in 2013, needs some time to gain popularity. Therefore, a rise in number of planned routes in the following season is comprehensible. However, the reason for the decrease in the number of planned routes after season 2015/16 is unknown. A possible explanation for the decrease would be a decline in popularity of White Risk, leading to lower user numbers and thus less planned routes. Unfortunately, no data concerning active users on White Risk was available for this thesis. This would have provided further insight into White Risk's popularity. Even though there exist various trip-planning tools, none of them is specifically designed for backcountry skiing. A product that competes with White Risk to some degree is Skitourenguru. Introduced in 2015, Skitourenguru allows assessing the risk of about 900 popular backcountry tours in Switzerland (Schmudlach and Köhler, 2016; Schmudlach, 2017). Skitourenguru aims at supporting the route choice for backcountry recreationists and providing an indication of the risk of a tour under current conditions. Lacking the possibility to draw routes on a map and load them on mobile devices, it does not cover the full planning process. Thus, it would make sense to use this tool in combination with White Risk. However, users could dispense with the use of White Risk as Skitourenguru provides sufficient information and the time spent on planning can be reduced.

A further explanation for the decrease in the number of planned routes might be unfavourable snow and weather conditions for seasons 2016/17 and 2017/18. According to the SLF's snow and avalanche reports for seasons 2015/16 and 2016/17, the two seasons were distinctly different (Zweifel et al., 2016, 2017). While both seasons exhibited low snow quantities during early winter months, abundant snowfall between January and March 2016 compensated for the season 2015/16. In contrast, season 2016/17 went down in history as one of the shortest winters with the least snow. Hence, the unfavourable snow conditions in 2016/17 provide a sensible explanation for the decrease in the number of planned routes for this season. However, the winter season 2017/18 was characterized by an early start of the winter season and general snow depths were distinctly above average (SLF, 2018d). Even though this led to higher danger levels on average, backcountry skiing conditions in the season 2017/18 were better than in the years before. Despite this improvement, the amount of planned tours remained at a constant low level.
Further reasons for the decrease of planned routes in seasons 2016/17 and 2017/18 might be that less users specified a tour date for their planned routes. This would lead to an ostensible decrease in planned routes, even if in fact the same amount of routes were planned. The date of the last modification of routes might provide further insights into user activity. However, this variable was not considered during the analysis of the dataset.
In conclusion, it is difficult to pinpoint the major cause of the decrease in number of planned routes. It is likely that a combination of factors influenced the number of planned routes. Comparable studies about backcountry skiing activity in Switzerland are not available but might provide further insights into the distribution of such activity over the years.

Considering the planned number of routes by month, the months with a maximum number of routes correspond with the months having the most favourable backcountry skiing conditions, namely January to March (fig. 5.2a). The distribution also indicates late winter starts, as in season 2016/17. During this season, backcountry skiing was not possible in December and, therefore, no routes were planned for this month. The distributions by month for all seasons correspond with the findings from Techel et al. (2015). They noted that backcountry activity was high between mid-December and mid-April. Additionally, they found that backcountry activity is highest during the weekends, which corresponds with the distributions of planned routes by weekday in this work (fig. 5.2b). Aggregating the planned routes to weekdays and weekend-days, the distribution is almost equal to the numbers presented by Techel et al. (2015). Consequently, the comparison of the distribution by day with numbers from Zweifel et al. (2006) exhibited the same differences as were observed by Techel et al. (2015). Techel et al. (2015) claimed that the web-portal contributors often live outside of the Swiss Alps, close to the population centres of Switzerland. Assuming the distribution of users follows the demographic distribution in Switzerland, also White Risk users are likely to live close to population centres. As no user information was available for this work, this assumption cannot be verified. Users living in population centres in the Swiss Plateau need more time to reach areas where backcountry skiing is possible. Thus, they likely prefer weekend-trips. Such effects of demographic differences in user communities have been described by various authors (Antoniou, 2011; Grossenbacher, 2014).

Additional works analysing backcountry skiing data focussed on terrain attributes in relation with avalanche danger and did not include statements about the temporal distribution of backcountry activity (Hendrikx et al., 2013; Hendrikx and Johnson,

2016). Furthermore, some focussed on specific user groups (e.g. heli skiing) (Hendrikx et al., 2014b; Haegeli and Atkins, 2016). Therefore, a meaningful comparison to other works on backcountry activity was not possible.

Spatial Distribution The density map, illustrating the spatial distribution (fig. 5.3), is not directly comparable with the backcountry activity map of Techel et al. (2015, p. 1991). Techel et al. (2015) aggregated the backcountry activity to the level of SLF warning regions. However, hot spots in the density maps correspond rather well with regions of high backcountry activity in the map of Techel et al. (2015). Especially, areas like the Wildstrubel massif, Urserental and Southern Engadin correspond well.
Furthermore, Ticino showed little planned activity and actual backcountry activity in these two works. Techel et al. (2015) noted that the combined proportion of French and Italian posts was almost identical to the proportion of French and Italian speakers in Switzerland. However, they did not differentiate these posts any further. Thus, it is possible that a majority of posts were in French. This would explain the low activity in Italian-speaking parts of Switzerland in their work. In this thesis, routes were not differentiated according to German-, French- and Italian-speaking users. Even though White Risk is available in Italian, Italian-speaking users might favour another tool, leading to bias in the dataset. However, it is also possible that the low activity in Ticino is real and backcountry skiing activity is indeed low in Ticino. An analysis of the user base might provide further insights into the proportion of Italian-speaking users. However, privacy restrictions prevented an in-depth user analysis.
Noticeable differences are detectable for St. Antönien, which served as a study area in this work. The backcountry activity reported by Techel et al. (2015) for this area is low, while figure 5.3 shows high densities of planned backcountry activity. Also, the Prättigau area shows high densities of planned backcountry activity while the backcountry activity reported by Techel et al. (2015) is relatively low. A possible explanation for this might be avalanche prevention courses and backcountry skiing courses which take place in this area. There are various providers of such courses, such as the SAC[1], Mammut[2], Bergführer Davos-Klosters[3] etc. If White Risk is introduced in these courses, it is likely that participants will plan test routes in these areas, leading to seemingly higher planned backcountry activity. These planned tours would not be mirrored in the reported backcountry activity, leading to differences in the two datasets.
Areas in the Mont Blanc massif exhibited lower planned backcountry activity than reported backcountry activity. Possible influences on the spatial distributions of both datasets could be heavy users who planned large numbers of tours in these areas. Such users might not be active on White Risk or platforms examined by Techel et al. (2015), respectively, leading to such differences.
A quantitative comparison between the results in this work and the results from Techel et al. (2015) would have been possible. To do so, planned backcountry activity from White Risk should have been aggregated to warning regions. Due to time constraints, a quantitative comparison could not be included in this thesis. However, the qualitative comparison suggested high similarities between the two datasets.

Similar to the temporal distribution, Techel et al. (2015) were the only ones who

[1] http://www.sac-cas.ch/formation/kurse/courseshop/detail/coursetype/type/lawinen.html
[2] http://alpineschool.mammut.ch/de.html
[3] https://bergfuehrer-davosklosters.ch/

examined the spatial distribution of backcountry activity. This prevented further comparisons to existing works.

Participation Inequality Several authors have noted that participation inequality is an issue with UGC, also with geographic data (VGI) (Grossenbacher, 2014; Neis and Zipf, 2012; Techel et al., 2015). Therefore, it does not come as a surprise that participation inequality was observed in the dataset examined in this thesis, illustrated in figures 5.5 and 5.6. The user distribution of the dataset is not skewed to the extent proposed by Nielsen (2006). However, the user distribution of the data from White Risk highly corresponded to comparable data as analysed by Techel et al. (2015). A major difference between White Risk and typical social media platforms on which users create content is that users of White Risk create the content for themselves and do not share it with a community. Thus, there are no users who only consume content without producing any, which leads to the less-skewed distribution. The fact that Techel et al. (2015) also neglected users who did not produce any content underlines the comparability of the two datasets.

Techel et al. (2015) stated that there exists a geographic bias in the analysed dataset, introduced by participation inequality. They found that heavy users favoured north-western parts of the Swiss Alps close to population centres in Switzerland. The analysis of certain user groups in White Risk confirmed the existence of a geographic bias for the dataset analysed in this thesis. However, the direction of the geographic bias was opposing, as was shown in figures 5.7 and 5.8. Yet, a direct comparison of heavy users from two different datasets is meaningless, as these users are unlikely to be the same people and therefore exhibit different behavioural habits. Nonetheless, the existence of a geographic bias was supported by the findings in this work and its influence should be considered in future research. Other works using VGI in an avalanche research context did not consider participation inequality (Hendrikx et al., 2013; Hendrikx and Johnson, 2016).

Avalanche Conditions The distribution of avalanche danger levels specified for planned routes differs strongly from the findings of Techel et al. (2015). Users in White Risk planned distinctly more routes on days with avalanche danger level 3 ('considerable') than were actually reported. Techel et al. (2015) presented a relative frequency of 26 %, whereas 42.9 % of the planned routes were planned for danger level 3 (cf. fig 5.9). Compared to results from Techel et al. (2015), fewer routes were planned for days with danger levels 1 & 2 ('low' & 'moderate'). These numbers are particularly surprising considering that danger level 2 is forecast more frequently than level 3, as was shown by Harvey (2002, p. 444). The planning of tours usually takes place several days before the actual tour date. However, users only have access to the current avalanche bulletin. They might expect a reduction of the avalanche danger leading up to the day of the tour, but still report the current avalanche danger. Another possible explanation for the observed differences might be imprecise route planning by users or test routes remaining after the preprocessing. Nonetheless, such routes are unlikely to be fully responsible for such differences.
There is a general difference between the dataset examined in this thesis and the dataset analysed by Techel et al. (2015). It is unknown whether the routes planned in White Risk were eventually executed. It might well be that users planned the route and did not execute it due to unfavourable avalanche conditions. Backcountry reports from web-portals are more likely to have taken place and thus exhibit less

favourable avalanche conditions.

A comparison of the user-specified avalanche danger with authoritative data from the SLF showed that 78.24 % of the users specified the avalanche danger correctly (figure 5.10a). Furthermore, for 20.25 % of the routes, the difference between user-specified and the authoritative danger level was 1. This is notable as there exist various possible error sources. Users might have created test routes, thereby specifying the avalanche danger without consulting the avalanche bulletin. Furthermore, users might specify the wrong tour date, leading to a mismatch when combining user and authoritative data. Also, the avalanche danger might change between the date of planning and the date of the tour. Nonetheless, the results suggest that the share of users imprecisely planning a route is rather small. If a user planned routes imprecisely or created test routes, the differences between the datasets would be expected to be larger.
The distribution of the differences between the two datasets shows that the amount of times the authoritative avalanche danger is under- and overestimated is almost equal, as was illustrated in figure 5.10b. There does not seem to be a general trend. Therefore, users from White Risk indeed seem to plan routes with higher avalanche danger as would be expected considering the reported backcountry activity by Techel et al. (2015). There exist no further works outlining the distribution of backcountry activity by avalanche danger.

6.2 Route Attributes

To date, typical characteristics of backcountry tours have not been evaluated. Therefore, a general comparison of the obtained route attributes to existing research is impossible. However, typical terrain attributes can be compared to typical avalanche starting zone characteristics. Ideally, backcountry recreationists avoid terrain which is exposed to higher avalanche risk.

It was learned from the summary statistics that typical routes are between 5'000 m and 10'000 m long (fig 5.11a) with vertices every 125 m to 300 m (fig 5.11b). Over this distance, they usually surmount height differences between 800 m and 1'200 m (fig. 5.12). The typical height of planned routes ranges between 1'250 m.a.s.l. and 3'000 m.a.s.l. (fig. 5.13).
According to Vontobel (2011), accident avalanches show a maximum height of 2'430 m.a.s.l. on average. This value lies within the range of maximum heights of planned routes. Minimum and mean heights are mostly lower (cf. fig 5.13). It is clear that backcountry recreationists favour such elevations due to their snow consistency and generally favourable snow conditions for backcountry skiing. Hence, it is unlikely that backcountry skiers avoid such elevations.
The average slope of accident avalanche starting zones is between 33°and 38°, with an average maximum of 41°(Vontobel, 2011). Median and mean maximum slopes of planned routes are close to the maximum slope of avalanche starting zones (tab. 5.5). Mean slope values of planned routes, however, are distinctly lower with about 19°. Possibly, planned routes exhibit high maximum slope values introduced through imprecise route drawing in the map interface. Routes along ridges might brush steep slopes below the ridge and lead to high maximum slope values. Figure 6.1 illustrates this issue. The routes in blue all show maximum slope values below 40°. The red routes exhibit maximum slope values greater than 50°. These routes can be considered imprecise. They show larger variations and brush raster cells with

high slope values at several critical locations (1-3). These critical locations include a ridge (1), and two extremely steep rock faces (2 & 3). Sometimes, hill slopes with high slope values are still skiable and bold recreationists might be after such challenges. In this case, however, these slopes are almost vertical and, thus, not skiable. Effects of imprecise route planning are likely to be accountable for maximum slope outliers.

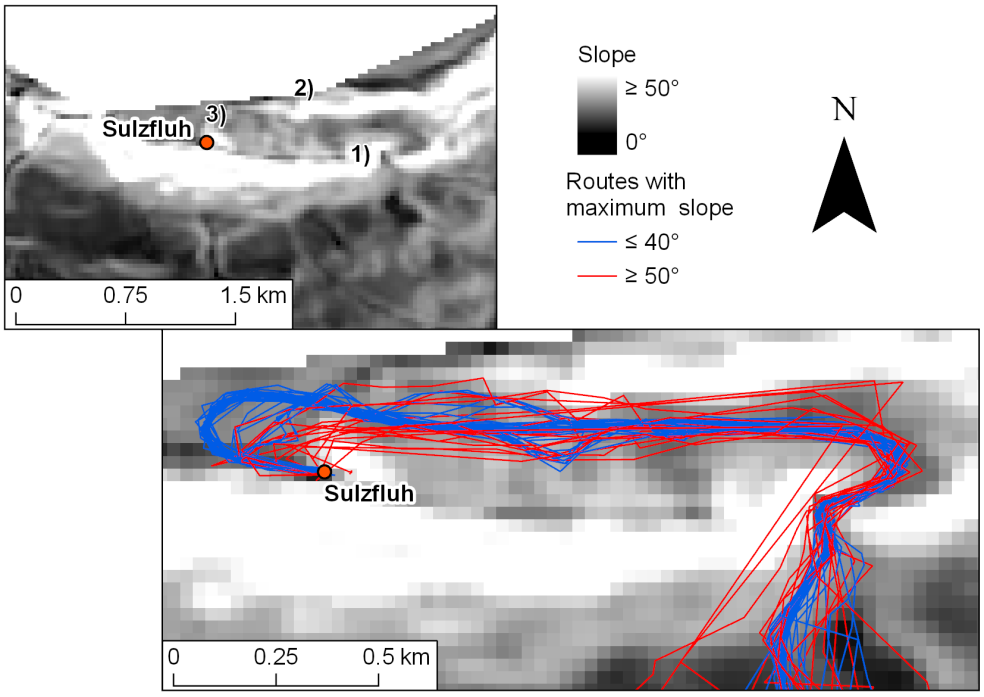


Figure 6.1: *Exemplary routes with high maximum slope values introduced through imprecise route planning on Sulzfluh.*

Vontobel (2011) noted that most avalanche starting zones exhibit convex or straight profile curvature. Straight profile curvature frequencies are above average for planned routes as well, as was shown in figure 5.16. However, the convex curvature frequency was below average and, thus, recreationists seem to avoid slopes with convex profile curvatures. Plan curvature for most starting zones was straight or concave. Planned routes mostly exhibited concave plan curvature as well, but straight curvature frequencies were below average. Generally, backcountry recreationists seem to avoid critical terrain forms. Classes with the highest number of avalanches showed low numbers for the curvature counts, with straight curvature in both directions as an exception. It can be expected that slopes exhibiting straight curvatures in both directions are most frequent in the terrain. Thus, the high numbers obtained for these class for the routes are not surprising. A normalization of curvature counts by class frequencies in the curvature rasters would allow generating further insights on the curvature class distribution. However, due to time constraints, this was not included in the thesis.
The analysis of Vontobel (2011) showed that most avalanche starting zones exhibit aspects between west and north-east. She claims that slopes exposed north are more frequently traversed than others, especially in late winter. According to Munter (1997), north-exposed slopes are generally traversed more frequently as they exhibit more favourable snow conditions. This assumption is supported to some degree by the findings in this thesis. Figure 5.17 showed that north-west- to north-east-exposed slopes are traversed relatively often. However, also east- and south-east-exposed slopes are traversed frequently. The aspect counts normalized

by the class frequency in the raster show that north- and north-east-exposed slopes are traversed more often than would be expected based on the class distribution in the terrain. This confirms the claims of Vontobel (2011) and Munter (1997). It is clear that favourable snow conditions are particularly desirable on the descent, i.e. when skiing. During the ascent, users might accept worse snow conditions, which leads to the fact that other aspects are relatively frequently represented as well. Furthermore, other aspects may also show good snow conditions.
A comparison of heading directions of planned routes with existing literature was not possible due to gaps in the literature. Comparing the heading directions to the slope exposition is rather difficult. If backcountry recreationists ascend to the peak in a direct fashion, the heading direction opposes the exposition, as illustrated in figure 6.2.1. However, to tackle steep hill slopes, backcountry recreationists often rely on the kickturn technique. Instead of walking directly towards the peak, skiers traverse the slope almost perpendicular to the line of the greatest slope (fig. 6.2.2). Nonetheless, most users probably do not plan their tours in such detail that every kickturn is included. During the descent the heading is in the opposite direction. This leads to a relatively even distribution of heading directions, as was seen in figure 5.18. North-west and west are the most frequent heading directions. This could be a result of many skiers choosing these directions during their descent. As mentioned before, northern slopes exhibit the most favourable snow conditions, which might be preferred by many users for the descent. An analysis of selected routes and the segmentation of routes in ascending and descending parts might provide further insights into the relationship between aspect and heading.

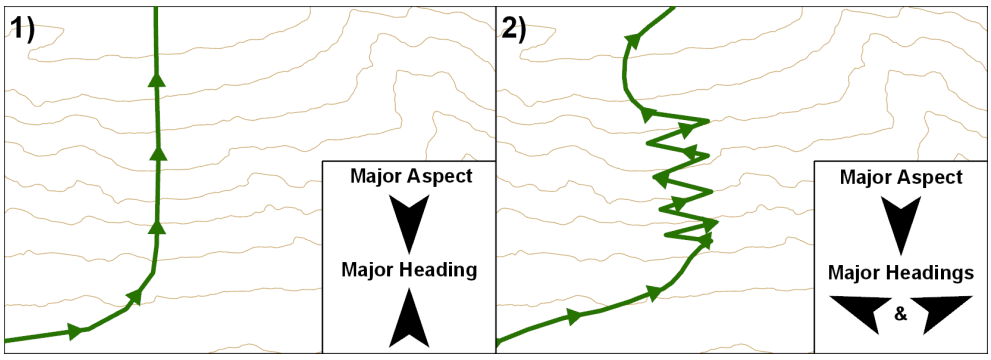


Figure 6.2: *Behaviour of heading in relation to major aspect of slopes during ascent.*

Almost the full range of possible straightness values is represented by the data. There are no straight routes in the filtered dataset, indicated by no straightness value being zero. With 64 %, a majority of the routes are closer to a straight line, however (value of zero), than to a round-trip (value of one), as was displayed in figure 5.15. This might either imply that most users do not return to their starting location or that they only plan the ascent of a full tour. A segmentation of routes based on the order of height values along the routes would have been conceivable. This would have allowed for a distinction in ascending and descending parts of the route. Time constraints and a lack of an appropriate raster extraction tool prevented such a segmentation and a subsequent analysis of differences between ascending and descending routes. As previously mentioned, there was no possibility to compare route attributes, including straightness, with existing literature.

Influence of Raster Resolution As was suggested in various previous works (Thompson et al., 2001; Sørensen and Seibert, 2007), the raster resolution influenced the attributes extracted from DEMs. Generally, the attribute ranges were widened

when the raster resolution was increased, as expected. Naturally, a widening of ranges particularly affects extrema attributes (minimum and maximum). Vontobel (2011) found that slope and curvature values increase with higher raster resolution. However, she noted that the differences are generally small and only the differences for slope were statistically significant. Her findings were based on mean curvature, exposition as well as on mean and maximum slope. This corresponds to the results obtained in this work, which were summarized in table 5.6. Differences were generally small for all height attributes, as well as for mean slope. Minimum and maximum slope attributes showed the largest variations for different DEM resolutions. Both slope and curvature were calculated with finite difference methods. The average maximum technique, as described by Burrough and McDonnell (1998, pp. 190-191), is implemented in ArcGIS Pro to calculate slope. Curvature is calculated using an algorithm introduced by Zevenbergen and Thorne (1987). The 3x3-Kernel, which is utilized in both algorithms, introduces an additional smoothing effect, as neighbouring cells are included in the calculation.
Even though Vontobel (2011) did not find any statistically significant differences for curvature attributes, the route attributes of minimum and maximum curvature were expected to be affected by varying raster resolution. Vontobel (2011) examined mean curvature attributes. Similar to mean height and slope, mean curvature should show less variation due to raster resolution differences than extrema attributes. Strongly related to the attribute range, both minimum and maximum curvature should be more greatly influenced by the raster resolution.
In general, high-resolution DEMs are unnecessary for many analyses in avalanche research. The reason for this is that avalanches are mainly released on weak underlying layers and not directly on the terrain surface (Harvey et al., 2012, pp. 28). Thus, the snow has a smoothing effect on the terrain. Bühler et al. (2011) showed that relatively coarse resolutions around 25 m are sufficient for simulations of avalanche incidents. They noted that influences of micro-topographical factors on avalanche processes are minimal, especially for large avalanches. However, the map interface of White Risk provides a slope layer based on a 5 m resolution DEM. As users base their planned routes on this resolution, future analyses of planned routes should consider the same resolution for DEMs and derivatives, if enough computational power is available.
The route attributes extracted from DHM25 were not used for any classification or further processing, but only for the characterisation of routes. Nonetheless, when working with routes and terrain attributes, one should be aware of the possible influences of raster resolutions on the extracted attributes.

Influence of Participation Inequality To date, there exists no work on VGI in which differences of route attributes between different user groups in terms of usage frequency were considered. To recap, in this thesis it was found that heavy users planned longer routes in more detail. Furthermore, their routes were at higher elevations than routes from rare users. Terrain attributes were less affected than intrinsic route attributes.
Studies that can be related to those findings are those from Hendrikx et al. (2014a) and Hendrikx and Johnson (2016). They examined terrain use as a function of user experience. Clearly, a user who planned large numbers of routes is not necessarily more experienced than a user who planned a single route. It is well possible that an experienced backcountry recreationist had only recently discovered White Risk and had not yet planned many routes on White Risk. Nevertheless, it is unlikely that an inexperienced user planned 70 or more routes. Thus, it is assumed that

heavy users are more experienced than rare users. Hendrikx et al. (2014a) and Hendrikx and Johnson (2016) found that experienced backcountry recreationists used significantly steeper slopes on average than inexperienced ones. The analysis of the planned routes revealed that heavy users plan routes with lower mean slope than rare users. However, the maximum slope of their routes was higher than of those from rare users. Despite both differences being statistically significant, with values between 0.5 and 1.6, the effects for both were small and, thus, the differences were insignificant. Mean values and effect sizes for all attributes are summarized in table 5.7. Generally, the mean attribute values of heavy users are more similar to the full dataset than those of rare users. This indicates that routes from heavy users corresponded more to typical routes (cf. table 5.5). This finding supports the assumption that heavy users are more experienced than rare users. Even though a comparison to the results from Hendrikx et al. (2014a) and Hendrikx and Johnson (2016) was not immensely fruitful, it would be interesting to compare route attributes with larger effects, such as length and height attributes. This might support the assumption that heavy users are more experienced than rare users.

The differences in sample sizes when comparing the user-specified avalanche dangers was surprising. Rare users specified the avalanche danger five times more often than heavy users. It is unclear why this difference could be observed. A possible explanation would be that rare users sought to explore the full range of functionalities when using White Risk for the first time. Contrarily, heavy users might think that the specification of an avalanche danger is not that beneficial, as the planning is usually realised several days before the tour and the avalanche danger is likely to change in the lead up to the tour. Although no obvious reason could be found, this difference could indicate different types of use of the tool across user groups.
Generally, it was shown that rare users plan more routes with higher avalanche danger than heavy users (cf. figure 5.23). It is however unclear whether there exists a real difference between the user groups or this difference is a result of possible different use types of the tool. Assuming that rare users are less experienced than heavy users, they might underestimate the avalanche risk when the forecast danger level is considerable. It would be interesting to see if experienced backcountry skiers undertake less tours during higher avalanche danger (considerable to very high) than less experienced skiers. A self-assessment of White Risk users regarding their experience, as well as their usage of the tool would help to gain further insight into the White Risk user base, their specific planning behaviour and how they use White Risk.

Influence of Avalanche Conditions Hendrikx et al. (2013) were the first to relate terrain use to forecast avalanche danger. They found a weak negative relationship between the forecast avalanche danger and the 99th percentile of used slope angles from GPS-tracks. However, their dataset only included tracks from experienced backcountry skiers and thus they remarked that these results are not generalizable. Three years later, Hendrikx and Johnson (2016, p. 741) used a larger sample of tracks and found that the steepest slope angles used were higher when low avalanche danger was forecast (median ~38°) than for higher danger levels (moderate to high: median ~30°). Both these results correspond to the findings in this work, where a weak negative relationship between mean / maximum slope and avalanche danger was found. However, the differences between the maximum slope medians of various danger levels were not as strong in this thesis, as was shown in figure 5.24 (between ~42°and ~37°). Hendrikx and Johnson (2016) also noted that the range of

slope attributes was very large. These large attribute ranges influence correlation coefficients.
Similar to terrain use as a function of user experience, Hendrikx and Johnson (2016) only used slope attributes as a metric to relate terrain use to avalanche forecast. It would therefore be interesting to compare additional route attributes, such as height or curvature attributes, to gain a better understanding of terrain use and compare it to results obtained in this thesis.

The contra-linear behaviour of route attributes for danger level 'very high' should be addressed at this point. Generally, this danger level is forecast very rarely. In all seasons covered, it was only forecast during the season 2017/2018 over a few days. The SLF advises against all winter sport activities beyond open ski runs, which includes all backcountry activities. When the forecast avalanche danger is 'very high', large and very large avalanches are to be expected which can be released naturally (SLF, 2018a). Harvey (2002, p. 447) noted that only very few people are caught in avalanches while touring during the presence of very high avalanche danger. Techel et al. (2015) did not include danger level 'very high' in their analysis, as it was never forecast during the period of their research. The total number of tours for which the specified avalanche danger was 'very high' was 12. Five of those tours included a tour date and, for two of these dates, the forecast avalanche danger was also 'very high'. Thus, it is likely that many of the planned routes for which danger level 'very high' was stored were either test routes or not accomplished after all. Considering that only a fraction of the planned routes with danger level 'very high' are 'real' planned routes, the contra-linear trend for attributes at this danger level can be ignored. Furthermore, compared to other danger levels, the sample size was small (n = 15'219 vs. n = 12), which underlines the insignificance of this effect.

6.3 Similarity & Clustering

Similarity The Fréchet distance proved to be an adequate similarity measure for movement trajectories in various previous works (Buchin et al., 2008, 2010a; Khoshaein, 2013). In this thesis, it was shown that it is applicable to planned movement data as well. Furthermore, it allowed obtaining similarity between planned movement and 'real' movement data. Figure 5.26a showed that the Fréchet distances between routes (kNN-distance) were similar for the two study areas St. Antönien und Urserental. This is notable as the Urserental study area is almost 1.5 times the size of the St. Antönien study area. The routes in the Wildstrubel study area showed slightly larger values for kNN-distances, as well as larger interquartile ranges. The kNN-distance is influenced by the terrain in the respective study areas. In rugged areas, backcountry routes generally follow relatively clear defined paths. The possibilities to take varying routes are restricted and few outliers can be found in those areas. This leads to lower kNN-distances in Urserental. In Wildstrubel and St. Antönien, the terrain is more open and allows backcountry recreationists to vary routes more freely. The stronger route variations lead to a wider route distribution and thus to larger kNN-distances and larger interquartile ranges. As is shown in figure 6.3, there exist areas with large route variations in all three study areas (marked with black rectangles). In Urserental, most of the routes in this area exhibited 5 Nearest Neighbours (5NN) distances above 1'000 m, which lies outside the interquartile range seen in figure 5.26. In the St. Antönien and Wildstrubel study areas, the 5NN-distances for most routes in these areas were between 250 m and 750 m / 1'000 m.

This results in large kNN-distances for routes in St. Antönien in relation to the size of the study area, and generally larger kNN-Distances for routes in Wildstrubel.

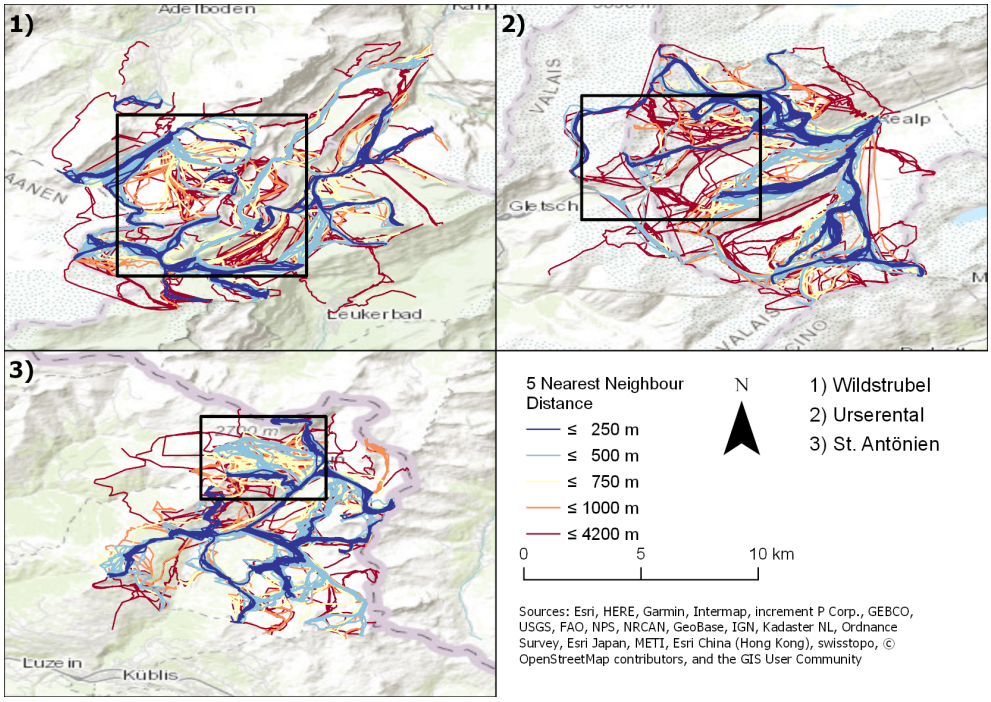


Figure 6.3: *Routes coloured by distance to five nearest neighbours in 1) Wildstrubel, 2) Urserental, and 3) St. Antönien.*

Considering the similarities between routes and GPS-tracks, additional factors come into play. Again, kNG-distances in St. Antönien and Urserental showed generally narrower quartile ranges than Wildstrubel, as was seen in figure 5.26b. While some of these differences are introduced by varying terrain and the smaller study area size of St. Antönien, the number of tracks and their spatial distribution strongly influenced the results. This behaviour is illustrated in figure 6.4. In Urserental, most tracks were available. These tracks include trips to all major backcountry destinations. Many routes showed 5 Nearest GPS-Tracks (5NG) distances below 1'500 m. In St. Antönien, the lowest number of tracks was obtained. This resulted in no route having a 5NG-distance below 1'500 m. However, the tracks were relatively well distributed and covered all major backcountry destinations. Most routes exhibited 5NG-distances between 1'500 m and 3'500 m, resulting in a narrow interquartile range. The large number of tracks in Urserental led to lower median and mean values for kNG-distances compared to St. Antönien, notably for larger k. In Wildstrubel, for multiple areas with high planned activity, none or only a few tracks were available. Combined with the stronger route variations, this led to higher kNG-distances. Even though there exist routes with low 5NG-distances, most of them ranged between 2'500 m and 5'000 m.

The Fréchet distance provides the ideal basis for further analysis on the examined dataset. Buchin et al. (2010a) discussed how additional constraints, such as direction, speed, etc., can be included in the similarity measure. Furthermore, geographic context can be included to create a context-aware similarity measure as was shown by Buchin et al. (2014). This might allow the inclusion of terrain-based avalanche risk factors in a similarity measure in future research. Buchin et al. (2008) presented a trajectory clustering approach which is based on the Fréchet distance. However, the simple usage of the Fréchet distance with a subsequent generic clustering ap-

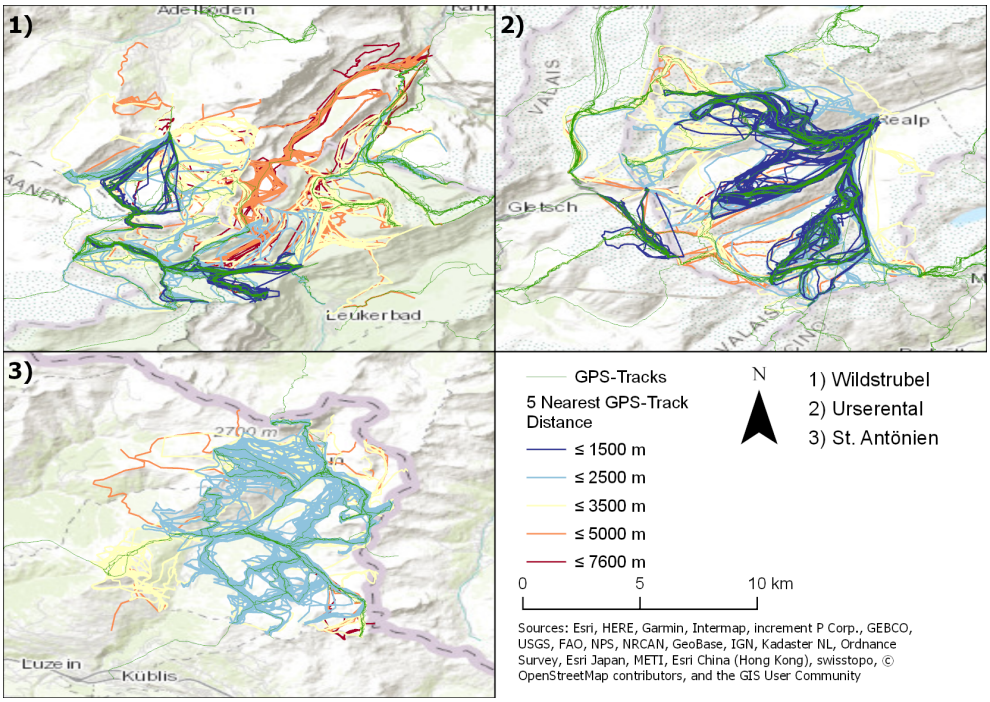


Figure 6.4: *Routes coloured by distance to five nearest GPS-tracks in 1) Wildstrubel, 2) Urserental, and 3) St. Antönien.*

proach is also feasible (Kisilevich et al., 2010). Therefore, the usage of the Fréchet distance enables a wide variety of ensuing clustering approaches.

To date, no work has been done on planned movement trajectories. Therefore, the similarity analysis of planned routes using the Fréchet distance is just one of a number of possible approaches. A general comparison of similarity measures for planned route trajectories would help to obtain a better idea of applicable approaches and to identify advantages and disadvantages for each of them.

It should be noted that most elaborate similarity measures are computationally-expensive. Therefore, concentrating on relatively small subsets is necessary in most cases. In this thesis, a geographic filter was applied to the routes to create smaller subsets. The calculation of the Fréchet distances for these route subsets was unproblematic. Filters other than geographical ones are conceivable, for example based on attributes. Such filters might allow comparing routes with similar attribute characteristics, e.g. high slope values, similar heights, etc. Since these routes may be dispersed over a large area, the coordinates may need to be normalized. The normalization would enable an assessment of geometric similarities between these routes, although they are geographically-scattered. However, such a normalization can lead to a loss of context. The context attributes should thus be calculated prior to the normalization.

Clustering The application of DBSCAN on the route dataset showed that it is possible to obtain a rough route corridor of major backcountry routes, solely based on the route geometry. The clustering was successful, even with routes with relatively low detail. The SAC released several backcountry skiing guide books for Switzerland (Eggenberger, 2015; Maier, 2015). To check the validity of the clusters identified by the DBSCAN as major backcountry routes, they were compared to routes from these guide books. Most clusters found with low ε-values corresponded to sin-

gle or combined routes from the SAC guide books, as is summarized in table 6.1. This underlines that the clusters found actually denote major backcountry ski routes and have an acceptable validity. The lack of cluster C on Stotzigen Firsten in the backcountry skiing guide might indicate the need for a supplement in the guide, as various users have planned this route.

Table 6.1: *Comparison of clusters identified by DBSCAN with routes in SAC guide books from Maier (2015, pp. 447/551) and Eggenberger (2015, pp. 79).*

Cluster	SAC Tour
Stotzigen Firsten (ε: 800 m, fig. 5.27.4)	
A	990
B	990
C	-
D	980a & 980c
Schafberg (ε: 600 m, fig. 5.28.3)	
A	115a
C	114a
D	114b
G	11 & 113a
H	113a

Nevertheless, the results of the clustering revealed that DBSCAN for routes is highly affected by parameter settings. The influence of the parameters is strongly related to terrain. This was shown based on two examples: Stotzigen Firsten and Schafberg. While on Stotzigen Firsten, the terrain predetermines possible route choices relatively clearly, Schafberg is more open. Schafberg is less rugged than Stotzigen Firsten and thus allows for more route variations. Furthermore, Girenspitz is in close proximity to Schafberg and a distinction of routes to these two different peaks is only possible with low ε-values. With higher ε-values, clusters include more core trajectories and additional trajectories are appended to clusters. This leads to larger clusters and lower inter-cluster distances, which facilitates the merging of clusters, as was seen for large ε-values. On Stotzigen Firsten, only two clusters were merged, even for an ε-value of 1'200 m (fig. 5.27). The two merged clusters followed the same path, whereas one depicted round trips and the other one-way trips. On Schafberg, clustering with an ε of 800 m already seemed rather chaotic (fig. 5.28). When ε was further increased, various clusters were merged and making a clear distinction between clusters became increasingly difficult. Thus, for routes in terrain which allows for only small variations, clustering seems to be less affected by parameters than for routes on open slopes which allow for large route variations.
The terrain characteristics also influence the number of routes classified as outliers. Table 5.10 shows distinctly larger proportions of outliers (15 % - 20 %) for Schafberg than for Stotzigen Firsten. This was a result of the greater possibility for route variations on Schafberg.

As a way to reduce the impact of parameter settings, a hierarchical clustering approach based on DBSCAN could be implemented. In a first step, general clusters with high variations could be identified with relatively large ε and $minPts$. This would result in large clusters including various routes to one or few peaks starting at similar locations. In a next step, a clustering with lower ε and $minPts$ would allow finding distinct route corridors to these peaks. The parameters could be adapted according to given terrain characteristics.

An additional approach might be to use contextual features as seeds for the clus-

tering. Conceivable contextual features are mountain peaks, i.e. tour destination, or towns / public transportation stops as starting points. Such an approach might allow identifying routes to specific mountain peaks in a first step and then finding major routes to these peaks in a second step. Future research should evaluate the feasibility and quality of such approaches.

Generally, DBSCAN clustering is relatively difficult for this large route dataset. Even when only small study areas are considered, large numbers of clusters are identified. A sensible visualisation of all clusters is virtually impossible and, therefore, a focus on individual peaks is inevitable. Hierarchical or context-based approaches might assist with the identification of major backcountry areas, for which a consequent detection of major backcountry routes can be conducted.

It is noteworthy that the validity of comparisons with GPS-tracks is limited due to the low number of available tracks. This holds for both comparisons between routes and tracks, as well as between route clusters and tracks. For more comprehensive comparisons, larger track samples would be necessary. However, from the sources considered in this thesis, no additional tracks were retrievable.

6.4 Research Questions

Research Question I What are the differences and similarities between planned backcountry routes and real backcountry tour GPS-tracks and how will they affect the further analysis of the planned backcountry routes? How can planned route trajectories be characterised?

In contrast to GPS-tracks, the planned backcountry routes do not include timestamps, as was noted in section 3.1.1. The lack of a temporal component in planned routes is the major difference between routes and tracks, having a severe influence on applicable analysis methods. As was shown in section 2.1, many CMA approaches exploit the temporal component when gaining additional information about movement, e.g. in the form of movement parameters. These movement parameters support trajectory segmentation, similarity assessment, etc. Therefore, in the absence of a temporal component, only spatial approaches can be applied to the trajectories, which are solely based on the trajectory geometry.
Depending on the GPS tracking device, tracks may include additional information about elevation, speed, air temperature, etc. beside locations and timestamps. Planned routes never include such additional information. If such information is indispensable to further analyses, the routes need to be enriched with secondary data sources, such as DEMs, weather data, or others. Furthermore, it might be possible to estimate features such as speed based on typical movement speeds in a given terrain. Additional information might enable the application of context-aware similarity measures, if they are consistently available.
GPS tracking devices usually record the carrier's location at specific time intervals. This leads to relatively homogeneous step lengths of tracks when the carrier moves at a constant speed. In contrast, the step length of planned routes depends on the level of detail desired by users. This usually varies depending on terrain characteristics and step lengths are generally larger for routes than for tracks. Nevertheless, this difference does not influence further processing. However, it has a definitive influence on the results. An example thereof are high slope values, which result from imprecise route planning, as was previously discussed. Furthermore, the differences

in step lengths may help distinguish routes from tracks.

A definition of planned route trajectories, which implies a general data structure, is given in section 3.1.1. General characteristics of planned route trajectories can be acquired from intrinsic attributes, based on the route geometry, or from extrinsic attributes, obtained through enrichment of routes with secondary data. Characteristics of the planned backcountry routes analysed in this thesis are outlined in sections 5.1 and 5.2. They were discussed in the previous section.

Research Question II How can test routes created by users be distinguished from 'real' planned routes? Which measures need to be taken to filter a dataset of planned route trajectories, so that most remaining routes depict a realistic planned backcountry tour?

Most test routes are characterized by atypical path geometries, which do not follow any environmental features. Often, they do not show sensible route choices, such as routes leading over large cliffs, extremely steep slopes, etc. Furthermore, many test routes are either too short or too long and are planned with little detail. Routes in areas in which no backcountry skiing is possible can also be considered test routes.
The distinction between test routes and 'real' planned routes is a binary classification problem. There exist several machine learning approaches for such problems. However, for the sake of simplicity, a heuristic approach was chosen in this thesis. Several attributes are related to route geometries and can be used to identify atypical behaviour. Thresholds for these attributes were defined based on attribute quantiles and a subsequent removal of assumed test routes. Furthermore, a spatial query allowed the removal of routes outside of areas in which backcountry skiing is possible.

Research Question III What type of intrinsic or extrinsic information can be used to determine the degree of accuracy of planned route trajectories?

After removing test routes during the preprocessing, most remaining routes should exhibit realistic geometries. Test routes are a prime example of unrealistic routes. In section 4.1, it was noted that length, number of vertices, and step length are route attributes which can help to distinguish test routes from 'real' planned routes. These intrinsic attributes were used to filter unrealistic routes. Maximum height was used to identify routes which do not reach necessary elevations for backcountry skiing. As this attribute originated from the enrichment of routes, maximum height is an extrinsic attribute. Furthermore, the application of a spatial filter allowed removing routes which are located in areas where no backcountry skiing is possible and, thus, are unrealistic as well.
A further extrinsic attribute that should be considered to determine the realism of routes is the maximum slope. Slope values of more than 65° to 70° are rather unrealistic, especially if these values were acquired using a low-resolution DEM. Such high slope values do occur in the slope raster, mainly in inner-alpine areas. However, the number of routes exhibiting such high slope values is below 2 %. As was noted in the previous section, the resolution reduction leads to a smoothing of the terrain. With a high-resolution DEMs, small cliffs that are overleapt during a tour can lead to large slope values. With low-resolution DEMs, however, high slope values indicate consistently very steep terrain, which can be considered unrealistic for most users. In any case, high slope values might just be a residue of undetailed route planning as previously discussed.
The comparison between user-stored and authoritative avalanche danger levels

showed that, for over 78 % of the routes, the avalanche danger was stored correctly. Possible reasons for deviations were provided in the previous section. Deviations between user-stored and authoritative avalanche danger of more than one level, as was the case for 1.51 % of the routes, can be interpreted as a suggestion of unrealistic routes.
Similarity to other routes or GPS-tracks can also be used as clue for the degree of accuracy of routes. For a realistic route, it is expected that there exist several other routes with similar geometry. This only holds if sufficient data is available, as was the case for planned routes. For realistic routes, there should also be similar tracks. However, the number of tracks available was insufficient to validate this assumption. The application of clustering on the route dataset allowed easily detecting outliers. These outliers can be considered unrealistic, as no similar routes exist.

Research Question IV What can be learned from planned route trajectories about the planning process of backcountry tours in avalanche-prone terrain?

By analysing an enriched planned route dataset, what typical planned backcountry routes look like was ascertained, in terms of length and terrain attributes. This allowed obtaining a general idea of the characteristics of the movement users plan to make in alpine terrain during backcountry tours. These findings are outlined in section 5.2 and discussed in section 6.2.
Density maps of planned routes provided insights into the spatial distribution of planned backcountry activity in Switzerland. It was shown that this distribution corresponds relatively well with the one reported for actual backcountry activity. The analysis of tour dates allowed generating an overview of the temporal distribution of planned backcountry activity. This distribution followed the distribution of reported backcountry activity closely.
From the analysis of avalanche danger users reported with the routes, it was learned that more routes than expected were planned for days on which the forecast avalanche danger was considerable. Possible reasons for this behaviour were discussed in the previous section. Furthermore, the analysis of the relationship between route attributes and avalanche danger revealed that, for higher avalanche danger, less exposed terrain was selected during route planning. The correlation, however, was only marginal.
The analysis of the spatial route distribution by user group revealed that rare users generally plan routes closer to the Swiss Plateau or well accessible valleys than heavy users. Furthermore, rare users plan shorter routes than heavy users at lower elevations. Additionally, the level of detail of routes from rare users was significantly lower than that of routes from heavy users.
By clustering planned route trajectories, major backcountry routes could be identified. The identified major backcountry routes corresponded with routes described in backcountry skiing guide books from the SAC. Moreover, typical route corridors for these major routes could be found based on the clusters.

6.5 Issues & Limitations

Throughout the thesis, several issues were encountered which will be discussed in this section. These issues are not specific to this thesis and should also be considered in further works on the planned route dataset. Furthermore, the limitations of this thesis will be addressed. The limitations should provide an overview of possible

improvements to the work in this thesis and where further work on the analysed dataset should focus.

6.5.1 Issues

Statistical Significance With a few exceptions, the statistical analyses have consistently shown very low p-values. These low p-values suggest a rejection of the H_0-hypotheses for most tests. As Wasserstein and Lazar (2016) noted, these low p-values do not necessarily imply that there exists a meaningful difference or correlation between attributes. They emphasized that any test can produce low p-values if the sample size is large enough. For all statistical tests in this thesis, the sample size was more than one thousand, which can be considered a large sample size. Wasserstein and Lazar (2016) underline the importance of additional approaches to quantify findings and support scientific reasoning. For this reason, additional effect measures were provided for all statistical analyses in this thesis. These should complement the results of statistical tests and provide a better understanding of the identified effects.

Privacy & Ethical Considerations When working with movement data, location privacy is a delicate topic. Krumm (2009) outlined several threats in relation to location privacy, from physical threats to inferences of additional information such as a user's home or work location, mode of transportation or static properties. In section 3.1.1, it was stated that an anonymised extract of planned movement data was provided for this thesis by the SLF. Clearly, it is not possible to directly infer some users' home location by a simple analysis of planned movement trajectories. However, if this data is combined with additional data, such inferences become possible. Consider the following example: A user plans a tour in White Risk and downloads the tour on his/her mobile phone. During the tour, he/she uses a GPS tracking device to record his/her movement. Upon arriving at home after finishing the tour, he/she uploads a tour report to some social media platform including the stored GPS-track. If the user followed his/her planned route, the recorded track should exhibit high similarity to the planned route. Thus, by combining the planned route with the GPS-track, one might be able to identify a user's social media profile and find a range of information about him/her.

Various researchers noted that geographic information is not truly volunteered if users originally created data for their own benefit or are unaware of how their data is being used (Obermeyer, 2007; Harvey, 2013). When working with user-created data, it is thus important that users are aware of the fact that their data might be used for other purposes, such as research. Generally, platforms that enable the creation of data should inform users about the use of their data.
The data used in this thesis was acquired by the SLF and made available for this thesis. The data will not be distributed to any third party and was solely used for analyses in this thesis.

Data Quality & Credibility Besides participation inequality and privacy issues, the data analysed in this thesis is also affected by other common VGI issues, discussed in section 2.2.1: Missing quality control and credibility. The user's motivation should be addressed as it is closely linked to data quality (Antoniou, 2011). In White Risk, users create content, i.e. plan routes, for their own benefit. As the planned routes

are not shared with a community, common motivations as identified by Budhathoki (2010) are unlikely to be prevalent. Users who plan a route using White Risk with the aim of using it during a tour thus benefit from high-quality planning. The planned routes provide useful information on terrain, avalanche danger, etc. only if they are of a sufficiently high quality. Nevertheless, typical quality control mechanisms cannot be applied. It is the user's decision as to how detailed their planned route should be and how much additional data should be included. Thoroughly-planned routes can be regarded as credible, since they represent "real" planned routes.
When generating low-quality data, the motives are different. It was found that various routes do not follow any environmental features or are located in areas unsuitable for backcountry skiing. It is unlikely that this data was intentionally biased or manipulated, as sometimes occurs on other platforms (Flanagin and Metzger, 2008; Purves, 2011). In contrast to other platforms, no real benefit can be obtained by biasing or manipulating data. It is more likely that new users create routes to test the full range of functionalities of a new tool. The testing of functionalities can be successful even if a unrealistic tour is planned. This, however, leads to data with low credibility. It was outlined in section 4.1 how routes with low credibility were eliminated in the dataset.

6.5.2 Limitations

The classification of routes in test routes and 'real' planned routes was conducted using heuristics, which were based on statistics, visual assessment and expert feedback. This combination should have led to a high-quality result. However, no quality assessment was conducted for the classification process. The analysis of classification errors would have provided further insights into the quality of the classification. Furthermore, this would have allowed for a comparison between the selected classification approach and other approaches.
A supervised learning approach might have provided sensible results for the distinction between test routes and 'real' planned routes. There exists a large range of supervised learning approaches, aiming at classifying objects into predetermined categories. Well-known approaches include decision-trees, neural networks, Bayesian networks or support vector machines, to name but a few. Kotsiantis (2007) provided a review of existing classification techniques, including a comparison thereof as well as their issues. It is unknown whether such an approach would have provided better results than the selected heuristic approach.

The straightness index, which was calculated for the planned routes, is rather simplistic and can only give a general idea of the straightness of a route. However, this index is straightforward, efficient and easy to compute. There exist several more elaborate tortuosity measures, as discussed by Benhamou (2004). Yet, most of those measures are more complex and relatively hard to implement. Furthermore, no existing package was available for the tortuosity calculation, which would have worked on the data without major conversions. It needs to be evaluated as to whether more elaborate tortuosity measures would have truly provided additional insights.

In section 5.1.3, whether the analysed dataset was affected by participation inequality was examined. It was shown that this participation inequality influences the spatial distribution of routes and that there exist differences in the route attributes between different user groups (sec. 5.2.2). However, the impact of participation in-

equality has not been taken into account in any of the results. A stratification approach, similar to that used by Techel et al. (2015), would have allowed extracting a more representative sample for the analyses.

In this thesis, only a single similarity measure and clustering approach was applied to the data. In section 2.1, various similarity measures for movement data were discussed. A qualitative comparison of multiple similarity measures for planned movement trajectories would have allowed a better-informed decision and provided a basis for further work on planned movement trajectories.
Furthermore, context was not included in the similarity measure used in this thesis. The inclusion of context in the similarity measure might have made the measure more accurate. However, the focus of this thesis was not the invention of a new context-aware similarity measure for planned movement data, but only to show the applicability of an existing one to such data.

A normalization of terrain attributes based on the value distribution in DEM and DEM derivatives was only conducted for the aspect attribute. The consideration of value distributions in height, slope and curvature rasters would have allowed obtaining an understanding of the attribute distribution in relation to the basic population. Due to time constraints, this was not included in this thesis.

Users can display a layer showing areas with slope values greater than 30° in White Risk. This layer is based on a DEM with a 5 m resolution. Assuming this layer is used during the planning process of backcountry routes, a layer with the same resolution should be used for all analyses on routes created with White Risk. However, as previously discussed, attribute extractions were not possible for rasters with such high resolutions, due to computational restrictions.

Chapter 7

Conclusion

In this thesis, a novel dataset was analysed to generate insights into the planning phase of backcountry ski tours. The data originated from White Risk, an online-based avalanche prevention platform, which includes a backcountry tour planning tool. Existing avalanche research utilizing movement data has mainly focussed on decision-making in avalanche terrain. Elaborate analysis techniques for movement data, such as CMA approaches, were neglected in recent works. In this thesis, planned movement data was analysed, which exhibits a different data structure than traditional movement data. The lack of a temporal component in movement data prevents the application of various CMA approaches and, thus, appropriate analysis approaches needed to be identified.

After a thorough preprocessing, during which the data was filtered and enriched with contextual information from DEM and DEM derivatives, a general overview of this novel dataset was provided. This allowed obtaining insights into temporal and spatial distribution of planned backcountry activity. Furthermore, the existence of participation inequality, a common issue with VGI data, and its impacts were shown. The distribution of the user-specified avalanche danger was analysed and compared to authoritative data acquired by the SLF.

In a next step, route attributes were summarized. This allowed obtaining an understanding of what typical planned backcountry routes look like. Furthermore, various impacts on these route attributes, such as DEM raster resolution, different user groups in terms of usage frequency and avalanche conditions, were evaluated.

Finally, the Fréchet distance, a well-known similarity measure from CMA, was applied to the data to assess the similarity in-between routes and between routes as well as GPS-tracks acquired from two mountaineering social media platforms. The similarity assessment further provided the foundation for a subsequent clustering of planned routes, which enabled the identification of major backcountry routes and the rough definition of their route corridors.

Together, these analysis steps allowed achieving the overall aim specified in section 1.2 and to answer the research questions specified in section 2.5. The contributions of this thesis will be outlined in the next section.

7.1 Contributions

The overall aim of this thesis was to analyse planned movement trajectories gathered with White Risk and therewith gain better insights into the planning process of backcountry tours. The results of this thesis revealed the days and months most tours are planned and in which areas high numbers of tours are planned. It was shown that these findings correspond well with reported backcountry activity. Furthermore, a large share of the user-specified avalanche danger matched authoritative data from the SLF, indicating adequate dealing with avalanche bulletins by most users during the planning process.
The analysis of route attributes provided a general overview of planned routes. It was shown that the DEM raster resolution directly influences route attributes, extracted from DEMs, and that extrema attributes are most affected. The evaluation of differences between user groups revealed that heavy users generally plan longer routes at higher elevations. Furthermore, routes from heavy users are usually planned in higher detail than routes from rare users. Additionally, the combination of user-specified avalanche danger with terrain attributes indicated that users adjust the planned routes according to the forecasted avalanche danger. For days with higher avalanche danger, routes with less steep routes and less extreme terrain forms are chosen. However, correlation coefficients were only marginal.
The assessment of differences between planned route trajectories and 'real' movement data revealed that the major difference between the two data types is the lack of a temporal component for planned movement data. This drastically restricts the number of applicable CMA approaches. The approaches which were applicable were identified and the Fréchet distance, as well as DBSCAN, based on the Fréchet distance, were applied to the dataset. It was shown that similarity measures are strongly influenced by terrain, in which planned routes are located. The application of DBSCAN on the planned routes allowed the identification of major backcountry routes. A rough route corridor could be defined for these major routes, solely based on route geometry. It was noted that the clustering approach is highly sensitive to parameter settings and, thus, also to the terrain. Comparisons of planned route trajectories to 'real' movement data proved to be rather difficult, due to a lack of sufficient GPS-tracks.
This work provided the first in-depth insights into the route planning behaviour of backcountry recreationists. This important component of backcountry skiing was previously unexplored. The thesis has built the groundwork for further research and should therefore help to advance understanding of the human factor in avalanche processes and, thus, also support the development of future prevention measures for avalanche accidents.
Additionally, a basis was created for further analysis of planned movement data. The approaches applied in this work are not solely applicable to planned backcountry tours but can also be used for different types of planned movement such as running, cycling, etc. However, a thorough comparison of various approaches should be conducted to identify the most effective and efficient approach.

7.2 Outlook & Further Research

This thesis is subject to several limitations, as was discussed in section 6.5. These limitations should be addressed in the future and hopefully be resolved to gain fur-

ther insights into the route planning behaviour of backcountry recreationists. The quality of the route classification during preprocessing should be assessed in order to compare it to other classification approaches. This could be achieved by manually labelling a number of routes from White Risk. These could then be classified using the heuristic approach applied in this thesis and classification errors could be detected. In a further step, additional classification approaches could be applied to the dataset and classification errors could be evaluated. This would allow a quantitative comparison of multiple approaches.

Future research should look at the possibility of creating artificial movement parameters for planned movement trajectories. Such could be generated by analysing typical walking or skiing speeds in a given terrain. Based on these speed attributes, artificial timestamps could be generated. The generation of an artificial timestamp should enable the application of spatio-temporal similarity measures. Such similarity measures can be considered context-aware, as the artificial movement speeds are heavily dependent on the terrain (upward/downward movement, steepness etc.). A comparison with purely spatial similarity measures should reveal whether spatio-temporal similarity measures based on artificial timestamps are superior to spatial similarity measures.

During this thesis, the idea came up to create a median trajectory for the clusters identified by DBSCAN. Such a trajectory would depict a typical route which is taken on a major backcountry route. Unfortunately, there exists no general implementation for median trajectories in R. Etienne et al. (2016) provided a theoretical framework and an implementation of median trajectories. However, no R package could be found that can easily be adopted for various data types.

Schmudlach and Köhler (2016) attempted to rate the avalanche risk of backcountry ski tours. They incorporated terrain data from DEMs, forestation data, as well as the forecast avalanche danger from the SLF. Currently, their tool enables users to rate the risk of several hundred popular backcountry routes which were digitized by Schmudlach and Köhler (2016). However, with a sufficiently efficient algorithm, the avalanche risk rating could be included in White Risk and computed on the fly when users plan backcountry ski routes. This could point users to sensible route choices depending on the current avalanche danger.

A useful extension to White Risk would be the implementation of tracking capabilities. This would allow users to spare an additional tracking device. Furthermore, if users planned a route and tracked it using White Risk, direct comparisons between routes and tracks could be performed. This would enable the detection of points where users deviated from a planned route during their trip. In case of deviations, users could be asked directly for their reasons, which would provide direct insights into decision-making during backcountry skiing.
As was discussed, the number of available tracks for this thesis was limited. The inclusion of tracking capabilities in White Risk might also increase the number of available 'real' movement trajectories. Furthermore, all tracks could be stored in the same format, which would simplify further processing.

Nowadays, there exist various route planning tools, not only for backcountry skiing, but also for other activities, such as cycling, running, etc. Nonetheless, the analysis of planned movement trajectories has been neglected to date. Besides gaining insights into decision-making processes during the planning phase of backcountry tours, the analysis of planned movement trajectories could provide valuable insights in other fields. It is hoped that this thesis sets the starting point for further research

on planned movement trajectories and opens up new avenues for scientific progress in various fields.

Bibliography

Agrawal, R., Gehrke, J., Gunopulos, D., and Raghavan, P. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, pp. 94–105. ACM, Seattle, Washington, USA (1998).

Alon, J., Sclaroff, S., Kollios, G., and Pavlovic, V. Discovering Clusters in Motion Time-Series Data. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pp. I–375–I–381 (2003).

Alt, H., Behrends, B., and Blömer, J. Approximate Matching of Polygonal Shapes. *Annals of Mathematics and Artificial Intelligence*, 13(3-4), pp. 251–265 (1995).

Alt, H. and Godau, M. Computing the Fréchet Distance between Two Polygonal Curves. *International Journal of Computational Geometry & Applications*, 5(1 & 2), pp. 75–91 (1995).

Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., and Vaisman, A. A Model for Enriching Trajectories with Semantic Geographical Information. In *Proceedings of the 15th International Symposium on Advances in Geographic Information Systems - ACM GIS 2007*, pp. 22:1–22:8. ACM, Seattle, Washington (2007).

Anagnostopoulos, A., Vlachos, M., Hadjieleftheriou, M., Keogh, E., and Yu, P. S. Global Distance-Based Segmentation of Trajectories. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 34–43. ACM, Philadelphia, PA, USA (2006).

Andrienko, G., Andrienko, N., and Heurich, M. An Event-Based Conceptual Model for Context-Aware Movement Analysis. *International Journal of Geographical Information Science*, 25(9), pp. 1347–1370 (2011).

Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., and Giannotti, F. Interactive Visual Clustering of Large Collections of Trajectories. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 3–10. Atlantic City, NJ, USA (2009).

Andrienko, G., Andrienko, N., and Wrobel, S. Visual Analytics Tools for Analysis of Movement Data. *SIGKDD Explorations Newsletter*, 9(2), pp. 38–46 (2007).

Andrienko, N. and Andrienko, G. Designing Visual Analytics Methods for Massive Collections of Movement Data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 42(2), pp. 117–138 (2007).

Andrienko, N. and Andrienko, G. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2), pp. 205–219 (2011).

Andrienko, N., Andrienko, G., Pelekis, N., and Spaccapietra, S. Basic Concepts of Movement Data. In *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery* (Edited by F. Giannotti and D. Pedreschi), chapter 1, pp. 15–38. Springer Berlin Heidelberg, Berlin, Heidelberg (2008).

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. OPTICS: Ordering Points to Identify the Clustering Structure. In *ACM Sigmod Record*, pp. 49–60. ACM, New York, NY, USA (1999).

Antoniou, V. *User Generated Spatial Content: An Analysis of the Phenomenon and its Challenges for Mapping Agencies*. Phd thesis, University College of London (2011).

Atkins, D. Human Factors in Avalanche Accidents. In *International Snow Science Workshop*, pp. 46–51. Big Sky, MT, USA (2000).

Atkins, D. and McCammon, I. Differences between Avalanche Experts and Novices. In *International Snow Science Workshop*, pp. 452–461. Jackson Hole, WY, USA (2004).

Batschelet, E. *Circular Statistics in Biology*. Academic Press, New York, NY, USA (1981).

Benhamou, S. How to Reliably Estimate the Tortuosity of an Animal's Path: Straightness, Sinuosity, or Fractal Dimension? *Journal of Theoretical Biology*, 229(2), pp. 209–220 (2004).

Berndt, D. and Clifford, J. Using Dynamic Time Warping to Find Patterns in Time Series. In *Workshop on Knowledge Discovery in Databases*, volume 10, pp. 359–370. AAAI, Seattle, Washington, USA (1994).

Bivand, R., Keitt, T., and Rowlingson, B. rgdal: Bindings for the 'geospatial' data abstraction library (2017). R package version 1.2-16. `https://CRAN.R-project.org/package=rgdal`.

Bivand, R. and Rundel, C. rgeos: Interface to geometry engine - open source ('geos') (2017). R package version 0.3-26. `https://CRAN.R-project.org/package=rgeos`.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. *Applied spatial data analysis with R, Second edition*. Springer, NY (2013). R package version 1.2-6. `https://CRAN.R-project.org/package=sp`.

Buchin, K., Buchin, M., and Gudmundsson, J. Constrained FreeSpace Fiagrams: A Tool for Trajectory Analysis. *International Journal of Geographical Information Science*, 24(7), pp. 1101–1125 (2010a).

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., and Luo, J. Detecting Commuting Patterns by Clustering Subtrajectories. In *International Symposium on Algorithms and Computation* (Edited by S.-H. Hong, H. Nagamochi, and T. Fukunaga), December, pp. 644–655. Springer Berlin Heidelberg (2008).

Buchin, K., Buchin, M., Van Kreveld, M., and Luo, J. Finding Long and Similar Parts of Trajectories. *Computational Geometry: Theory and Applications*, 44(9), pp. 465–476 (2011a).

Buchin, M., Dodge, S., and Speckmann, B. Context-Aware Similarity of Trajectories. In *Geographic Information Science* (Edited by N. Xiao, M.-P. Kwan, M. F. Goodchild, and S. Shekhar), pp. 43–56. Springer Berlin Heidelberg, Berlin, Heidelberg (2012).

Buchin, M., Dodge, S., and Speckmann, B. Similarity of Trajectories Taking Into Account Geographic Context. *Journal of Spatial Information Science*, 9, pp. 101–124 (2014).

Buchin, M., Driemel, A., van Kreveld, M., and Sacristán, V. An Algorithmic Framework for Segmenting Trajectories Based on Spatio-Temporal Criteria. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 202–211. ACM, San Jose, CA, USA (2010b).

Buchin, M., Driemel, A., Van Kreveld, M., and Sacristan, V. Segmenting Trajectories: A Framework and Algorithms using Spatiotemporal Criteria. *Journal of Spatial Information Science*, 3(3), pp. 33–63 (2011b).

Buchin, M., Kruckenberg, H., and Kölzsch, A. Segmenting Trajectories by Movement States. In *Advances in Spatial Data Handling: Geospatial Dynamics, Geosimulation and Exploratory Visualization* (Edited by S. Timpf and P. Laube), pp. 15–25. Springer Berlin Heidelberg, Berlin, Heidelberg (2013).

Budhathoki, N. R. *Participants' Motivations To Contribute Geographic Information in an Online Community*. Phd thesis, University of Illinois (2010).

Bühler, Y., Christen, M., Kowalski, J., and Bartelt, P. Sensitivity of snow avalanche simulations to digital elevation model quality and resolution. *Annals of Glaciology*, 52(58), pp. 72–80 (2011).

Burrough, P. A. and McDonnell, R. A. *Principles of Geographical Information Systems*. Oxford University Press Inc., New York, New York, 11 edition (1998).

Cao, H., Mamoulis, N., and Cheung, D. W. Mining Frequent Spatio-Temporal Sequential Patterns. In *5th IEEE International Conference on Data Mining (ICDM'05)*, pp. 82–89 (2005).

Chen, L. and Ng, R. On The Marriage of Lp-norms and Edit Distance. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*, pp. 792–803. VLDB Endowment, Toronto, Canada (2004).

Chen, L., Özsu, M. T., and Oria, V. Symbolic Representation and Retrieval of Moving Object Trajectories. In *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 227–234. ACM, New York, NY, USA (2004).

Chen, L., Özsu, M. T., and Oria, V. Robust and Fast Similarity Search for Moving Object Trajectories. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pp. 491–502. ACM, Baltimore, MD, USA (2005).

Cohen, J. *Statistical Power Analysis for The Behavioral Sciences*. Lawrence Erlbaum Associates, 2nd edition (1988).

Connors, J. P., Lei, S., and Kelly, M. Citizen Science in the Age of Neogeography: Utilizing Volunteered Geographic Information for Environmental Monitoring. *Annals of the Association of American Geographers*, 102(6), pp. 1267–1289 (2012).

Conway, J., Eddelbuettel, D., Nishiyama, T., Prayaga, S. K., and Tiffin, N. Rpostgresql: R interface to the 'postgresql' database system (2017). R package version 0.6-2. `https://CRAN.R-project.org/package=RPostgreSQL`.

Dempster, A., Laird, N., and Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), pp. 1–38 (1977).

Dikau, R. The Application of a Digital Relief Model to Landform Analysis in Geomorphology. In *Three Dimensional Applications In GIS* (Edited by J. Raper), chapter 5, pp. 51–78. CRC Press, 1st edition (1989).

Dodge, S. *Exploring Movement Using Similarity Analysis*. Phd thesis, University of Zurich (2011).

Dodge, S., Laube, P., and Weibel, R. Movement Similarity Assessment Using Symbolic Representation of Trajectories. *International Journal of Geographical Information Science*, 26(9), pp. 1563–1588 (2012).

Dodge, S., Weibel, R., and Forootan, E. Revealing the Physics of Movement: Comparing the Similarity of Movement Characteristics of Different Types of Moving Objects. *Computers, Environment and Urban Systems*, 33(6), pp. 419–434 (2009).

Dodge, S., Weibel, R., and Lautenschütz, A.-K. Towards a Taxonomy of Movement Patterns. *Information Visualization*, 7(3-4), pp. 240–252 (2008).

Douglas, D. H. and Peucker, T. K. Algorithms for the Reduction of the Number of Points Required To Represent a Digitized Line or Its Caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), pp. 112–122 (1973).

Dowle, M. and Srinivasan, A. data.table: Extension of 'data.frame' (2017). R package version 1.10.4-3. `https://CRAN.R-project.org/package=data.table`.

Eckert, J. SnowSense: A Case Study of User-centered Location-Based Services. *Journal Of Information Architecture*, 3(1), pp. 5–18 (2011).

Eckert, J. and Suter, C. New Mobile and Web-Based Services for Backcountry Skiers. In *International Snow Science Workshop*, pp. 475–480. Squaw Valley, CA, USA (2010).

Edelhoff, H., Signer, J., and Balkenhol, N. Path Segmentation for Beginners: An Overview of Current Methods for Detecting Changes in Animal Movement Patterns. *Movement Ecology*, 4(1), p. 21 (2016).

Eggenberger, V. *Graubünden Nord : Prättigau/Surselva/Rheinwald : Skitouren*. SAC-Verlag, Bern, 4th edition (2015).

Elwood, S. Volunteered Geographic Information: Future Research Directions Motivated by Critical, Participatory, and Feminist GIS. *GeoJournal*, 72(3-4), pp. 173–183 (2008).

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD*, 96(34), pp. 226–231 (1996).

Etienne, L., Devogele, T., Buchin, M., and McArdle, G. Trajectory Box Plot: A New Pattern to Summarize Movements. *International Journal of Geographical Information Science*, 30(5), pp. 835–853 (2016).

Eubank, R. L. *A Kalman Filter Primer*. CRC Press, Boca Raton, FL, USA (2005).

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), pp. 37–54 (1996).

Flanagin, A. J. and Metzger, M. J. The Credibility of Volunteered Geographic Information. *GeoJournal*, 72(3-4), pp. 137–148 (2008).

Fredston, J. and Fesler, D. *Snow Sense: A Guide to Evaluating Snow Avalanche Hazard*. Alaska Mountain Safety Center, Inc., Ancorage (1994a).

Fredston, J. and Fesler, D. The Human Factor - Lessons For Avalanche Education. In *International Snow Science Workshop*, pp. 473–487. Snowbird, UT, USA (1994b).

Fritz, C. O., Morris, P. E., and Richler, J. J. Effect Size Estimates: Current Use, Calculations, and Interpretation. *Journal of Experimental Psychology: General*, 141(1), pp. 2–18 (2012a).

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., and Obersteiner, M. Geo-Wiki: An Online Platform for Improving Global Land Cover. *Environmental Modelling & Software*, 31, pp. 110–123 (2012b).

Furman, N., Shooter, W., and Schumann, S. The Roles of Heuristics, Avalanche Forecast, and Risk Propensity in the Decision Making of Backcountry Skiers. *Leisure Sciences*, 32(5), pp. 453–469 (2010).

Gaffney, S. and Smyth, P. Trajectory Clustering with Mixtures of Regression Models. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 63–72. ACM, San Diego, CA, USA (1999).

Goodchild, M. F. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, 69(4), pp. 211–221 (2007a).

Goodchild, M. F. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2, pp. 24–32 (2007b).

Goodchild, M. F. and Glennon, J. A. Crowdsourcing Geographic Information for Disaster Response: A Research Frontier. *International Journal of Digital Earth*, 3(3), pp. 231–241 (2010).

Goodchild, M. F. and Li, L. Assuring the Quality of Volunteered Geographic Information. *Spatial Statistics*, 1, pp. 110–120 (2012).

Grossenbacher, T. *Studying Human Mobility Through Geotagged Social Media Content*. Master's thesis, University of Zurich (2014).

Gschwend, C. and Laube, P. Challenges of Context-Aware Movement Analysis – Lessons Learned about Crucial Data Requirements and Pre-Processing. In *Proceedings of the GIS Research UK 20th Annual Conference* (Edited by B. Rowlingson and D. Whyatt), volume 1, pp. 241–246 (2012).

Gudmundsson, J., Katajainen, J., Merrick, D., Ong, C., and Wolle, T. Compressing Spatio-Temporal Trajectories. *Computational Geometry: Theory and Applications*, 42(9), pp. 825–841 (2009).

Gudmundsson, J., Laube, P., and Wolle, T. Computational Movement Analysis. In *Springer Handbook of Geographic Information* (Edited by W. Kresse and D. M. Danko), chapter 22, pp. 725–738. Springer, Berlin, Heidelberg, Berlin, Heidelberg (2011).

Haegeli, P. and Atkins, R. Managing the Physical Risk from Avalanches in a Helicopter Skiing Operation - Merging and Contrasting Terrain Use Data with the Operational Guiding Perspective. In *International Snow Science Workshop*, pp. 104–111. Beckenridge, CO, USA (2016).

Haegeli, P., Haider, W., Longland, M., and Beardmore, B. Amateur Decision-Making in Avalanche Terrain With and Without a Decision Aid: A Stated Choice Survey. *Natural Hazards*, 52(1), pp. 185–209 (2010).

Hahsler, M. and Piekenbrock, M. dbscan: Density based clustering of applications with noise (dbscan) and related algorithms (2017). R package version 1.1-1. `https://CRAN.R-project.org/package=dbscan`.

Haklay, M. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, 37(4), pp. 682–703 (2010).

Haklay, M. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice* (Edited by D. Sui, S. Elwood, and M. Goodchild), chapter 7, pp. 105–122. Springer, Dordrecht (2013).

Haklay, M. Why is Participation Inequality Important? In *European Handbook of Crowdsourced Geographic Information* (Edited by C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, and R. Purves), chapter 3, pp. 35–44. Ubiquity Press, London (2016).

Haklay, M. and Weber, P. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7(4), pp. 12–18 (2008).

Han, J., Lee, J.-g., and Kamber, M. Overview of Clustering Methods in Geographic Data Analysis. In *Geographic Data Mining and Knowledge Discovery* (Edited by H. Miller and J. Han), chapter 7, pp. 149–187. CRC Press, 2nd edition (2009).

Harvey, F. To Volunteer or to Contribute Locational Information? Towards Truth in Labeling for Crowdsourced Geographic Information. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice* (Edited by D. Sui, S. Elwood, and M. Goodchild), chapter 3, pp. 31–42. Springer, Dordrecht (2013).

Harvey, S. Avalanche Incidents in Switzerland in Relation to the Predicted Danger Degree. In *International Snow and Science Workshop*, pp. 443–448. Penticton, BC, Canada (2002).

Harvey, S. White Risk - Interactive Avalanche Learning CD. In *International Snow Science Workshop*, pp. 274–281. Telluride, CO, USA (2006).

Harvey, S., Aegerter, S., and Landolt, D. White Risk 2 . 0 – A New Web-Based Platform for Avalanche Education. In *International Snow Science Workshop*, pp. 507–510. Grenoble, France (2013).

Harvey, S., Eisenegger, P., and Landolt, D. White Risk Instructor – A Presentation Tool for Avalanche Educators. In *International Snow Science Workshop*, pp. 591–593. Whistler, BC, Canada (2008).

Harvey, S., Rhyner, H., and Schweizer, J. *Lawinenkunde - Praxiswisse für Einsteiger und Profis zu Gefahren, Risiken und Strategien*. Bruckmann Verlag GmbH, München (2012).

Harvey, S., Signorell, C., and Genswein, M. Avalanche Accidents in Back Country Terrain of the Swiss Alps: New Investigations of a 30 Year Database. In *International Snow Science Workshop*. Penticton, BC, Canada (2002).

Hendrikx, J., Johnson, J., Hardy, G., and Weinberg, W. Using Global Crowd-Sourced Data To Understand Travel Behaviour in Avalanche Terrain. In *International Snow Science Workshop*. Banff, AB, Canada (2014a).

Hendrikx, J., Johnson, J., and Southworth, E. Understanding Travel Behaviour in Avalanche Terrain: A New Approach. In *International Snow Science Workshop*, 2004, pp. 511–515. Grenoble, France (2013).

Hendrikx, J. and Johnson, J. D. Understanding Global Crowd Sourcing Data to Examine Travel Behavior in Avalanche Terrain. In *International Snow Science Workshop*, pp. 737–743. Breckenridge, CO, USA (2016).

Hendrikx, J., Shelly, C., and Johnson, J. Tracking Heli-Ski Guides to Understand Decision Making in Avalanche Terrain. In *International Snow Science Workshop*, pp. 1021–1027. Banff, AB, Canada (2014b).

Hengl, T., Roudier, P., Beaudette, D., and Pebesma, E. plotKML: Scientific visualization of spatio-temporal data. *Journal of Statistical Software*, 63(5), pp. 1–25 (2015). R package version 0.5-8. `https://CRAN.R-project.org/package=plotKML`.

Hijmans, R. J. geosphere: Spherical trigonometry (2017a). R package version 1.5-7. `https://CRAN.R-project.org/package=geosphere`.

Hijmans, R. J. raster: Geographic data analysis and modeling (2017b). R package version 2.6-7. `https://CRAN.R-project.org/package=raster`.

Hinneburg, A. and Gabriel, H. DENCLUE 2.0: Fast Clustering based on Kernel Density Estimation. In *International Symposium on Intelligent Data Analysis*, pp. 70–80. Springer Berlin Heidelberg (2007).

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. Implementing a class of permutation tests: The coin package. *Journal of Statistical Software*, 28(8), pp. 1–23 (2008). R package version 1.2-2. `https://CRAN.R-project.org/package=coin`.

Hovland, C. I., Janis, I. L., and Kelley, H. H. Communication and Persuasion. Psychological Studies of Opinion Change. p. 315. Yale University Press, New Haven, CT, USA (1953).

Hunziker, P. velox: Fast raster manipulation and extraction (2017). R package version 0.2.0. `https://CRAN.R-project.org/package=velox`.

Javanmardi, S., Ganjisaffar, Y., Lopes, C., and Baldi, P. User Contribution and Trust in Wikipedia. In *5th International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pp. 1–6 (2009).

Jun, J., Guensler, R., and Ogle, J. Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance, Speed, and Acceleration Profile Estimates. *Transportation Research Record: Journal of the Transportation Research Board*, 1972, pp. 141–150 (2006).

Junejo, I. N., Javed, O., and Shah, M. Multi Feature Path Modeling for Video Surveillance. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pp. 716–719 (2004).

Karypis, G., Han, E.-H., and Kumar, V. Chameleon: Hierarchical Clustering Using Dynamic Modeling. *Computer*, 32(8), pp. 68–75 (1999).

Kaufman, L. and Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, 1st edition (2005).

Kendall, M. G. A New Measure of Rank Correlation. *Biometrika*, 30(1/2), pp. 81–93 (1938).

Kessler, F. Volunteered Geographic Information: A Bicycling Enthusiast Perspective. *Cartography and Geographic Information Science*, 38(3), pp. 258–268 (2011).

Khoshaein, V. *Trajectory Clustering using a Variation of Fréchet Distance*. Master thesis, University of Ottawa (2013).

Kisilevich, S., Mansmann, F., Nanni, M., and Rinzivillo, S. Spatio-Temporal Clustering. In *Data Mining and Knowledge Discovery Handbook* (Edited by O. Maimon and L. Rokach), chapter 44, pp. 855–874. Springer, 2nd edition (2010).

Kotsiantis, S. Supervised Machine Learning: A Review of Classification Techniques. In *Emerging Artificial Intelligence Applications in Computer Engineering* (Edited by I. G. Maglogiannis, K. Karpouzis, M. Wallace, and J. Soldatos), pp. 3–24. IOS Press (2007).

Krumm, J. A Survey of Computational Location Privacy. *Personal and Ubiquitous Computing*, 13(6), pp. 391–399 (2009).

Laube, P. Progress in Movement Pattern Analysis. In *Behaviour Monitoring and Interpretation - BMI - Smart Environments* (Edited by B. Gottfried and H. Aghajan), pp. 43–71. IOS Press (2009).

Laube, P. *Computational Movement Analysis*. Springer, Heidelberg (2014).

Laube, P., Dennis, T., Forer, P., and Walker, M. Movement Beyond the Snapshot - Dynamic Analysis of Geospatial Lifelines. *Computers, Environment and Urban Systems*, 31(5), pp. 481–501 (2007).

Laube, P., Imfeld, S., and Weibel, R. Discovering Relative Motion Patterns in Groups of Moving Point Objects. *International Journal of Geographical Information Science*, 19(6), pp. 639–668 (2005).

Laube, P. and Purves, R. S. How Fast Is a Cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS*, 15(3), pp. 401–418 (2011).

Lautenschütz, A.-K. How Context Influences the Segmentation of Movement Trajectories – An Experimental Approach for Environmental and Behavioral Context. In *GIScience 2010: Sixth International Conference on Geographic Information Science*, September. Zurich, Switzerland (2010).

Lee, J.-G., Han, J., and Whang, K.-Y. Trajectory Clustering: A Partition-And-Group Framework. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pp. 593–604. ACM, Bejing, China (2007).

Lehmann, E. L. *Elements of Large-Sample Theory*. Springer (1999).

Lemon, J. Plotrix: A package in the red light district of r (2006). R package version 3.6-6. `https://CRAN.R-project.org/package=plotrix`.

Levenshtein, V. I. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics - Doklady*, 10(8), pp. 707–710 (1966).

Li, L. and Goodchild, M. F. Is Privacy Still an Issue in the Era of Big Data? - Location Disclosure in Spatial Footprints. In *21st International Conference on Geoinformatics*. IEEE, Kaifeng, China (2013).

Li, L., Goodchild, M. F., and Xu, B. Spatial, Temporal, and Socioeconomic Patterns in the Use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40(2), pp. 61–77 (2013).

Lin, B. and Su, J. Shapes Based Trajectory Queries for Moving Objects. In *Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems*, pp. 21–30. ACM, Bremen, Germany (2005).

Lin, D. An Information-Theoretic Definition of Similarity. In *Proceedings of the Fifteenth International Conference on Machine Learning*, volume 98, pp. 296–304. Morgan Kaufmann Publishers Inc. (1998).

Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), pp. 129–137 (1982).

Longueville, B. D., Luraschi, G., Smits, P., Peedell, S., Groeve, T. D., and Commission, E. Citizens As Sensors for Natural Hazards. *Geomatica*, 64(1), pp. 41–59 (2010).

Maier, M. *Zentralschweizer Voralpen und Alpen : Einsiedeln bis Gotthard : Skitouren*. SAC-Verlag, Bern, 2nd edition (2015).

McCammon, I. Decision Making for Wilderness Leaders: Strategies, Traps and Teaching Methods. In *Wilderness Risk Manager's Conference*, pp. 16–29. Lake Geneva, WI, USA (2001).

McCammon, I. Evidence of Heuristic Traps in Recreational Avalanche Accidents. In *International Snow Science Workshop*. Penticton, BC, Canada (2002).

McCammon, I. Human Factors in Avalanche Accidents: Evolution and Interventions. In *International Snow Science Workshop*, pp. 644–648. Davos, Switzerland (2009).

McClung, D. M. The Elements of Applied Avalanche Forecasting Part I: The Human Issues. *Natural Hazards*, 26(2), pp. 111–129 (2002).

McClung, D. M. and Schaerer, P. *The Avalanche Handbook*. Mountaineers Books, 3rd edition (2006).

Melnychuk, M. C., Welch, D. W., and Walters, C. J. Spatio-Temporal Migration Patterns of Pacific Salmon Smolts in Rivers and Coastal Marine Waters. *PloS one*, 5(9) (2010).

Menghini, G., Carrasco, N., Schüssler, N., and Axhausen, K. W. Route Choice of Cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, 44(9), pp. 754–765 (2010).

Meratnia, N. and de By, R. A. Spatiotemporal Compression Techniques for Moving Point Objects. In *9th International Conference on Extending Database Technology*, pp. 765–782. Springer, Heraklion, Greece (2004).

Merrick, R. L. and Loughlin, T. R. Foraging Behavior of Adult Female and Young-Of-The-Year Steller Sea Lions in Alaskan Waters. *Canadian Journal of Zoology*, 75(5), pp. 776–786 (1997).

Mooney, P., Olteanu-Raimond, A.-M., Touya, G., Juul, N., Alvanides, S., and Kerle, N. Considerations of Privacy , Ethics and Legal Issues in Volunteered Geographic

Information. In *Mapping and the Citizen Sensor* (Edited by G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, and V. Antoniou), chapter 6, pp. 119–135. Ubiquity Press, London (2017).

Munter, W. *3 x 3 Lawinen: entscheiden in kritischen Situationen*. Agentur Pohlmann & Schellhammer (1997).

Myers, J. L. and Well, A. D. *Research Design and Statistical Analysis*. Lawrence Erlbaum Associates, 2nd edition (2003).

Nanni, M. and Pedreschi, D. Time-Focused Clustering of Trajectories of Moving Objects. *Journal of Intelligent Information Systems*, 27(3), pp. 267–289 (2006).

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., and Smouse, P. E. A Movement Ecology Paradigm for Unifying Organismal Movement Research. *Proceedings of the National Academy of Sciences*, 105(49), pp. 19052–19059 (2008).

Navarro, D. Learning statistics with r: A tutorial for psychology students and other beginners. (version 0.5) (2015). R package version 0.5. `https://CRAN.R-project.org/package=lsr`.

Neis, P. and Zipf, A. Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(3), pp. 146–165 (2012).

Nelson, T. A., Denouden, T., Jestico, B., Laberee, K., and Winters, M. BikeMaps.org: A Global Tool for Collision and Near Miss Mapping. *Frontiers in Public Health*, 3(March), pp. 1–8 (2015).

Nielsen, J. The 90-9-1 rule for participation inequality in social media and online communities (2006, October 9). Retrieved 20 May 2017, from `https://www.nngroup.com/articles/participation-inequality/`.

Obermeyer, N. Thoughts on Volunteered (Geo)Slavery. In *Workshop on Volunteered Geographic Information*, pp. 1–3. Santa Barbara, CA, USA (2007).

OED Online. avalanche, n. (2018). Retrieved 27 Feburary 2018, from `http://www.oed.com/view/Entry/13596`.

Olson, D. and Delen, D. *Advanced Data Mining Techniques*. Springer Berlin Heidelberg (2008).

OpenStreetMap. OpenStreetMap Stats (2018). Retrieved 25 January 2018, from `https://wiki.openstreetmap.org/wiki/Stats`.

O'Reilly, T. What Is Web 2.0 (2005, September 30). Retrieved 17 January 2018, from `http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html`.

Pánek, J. and Benediktsson, K. Emotional Mapping and Its Participatory Potential: Opinions About Cycling Conditions in Reykjavík, Iceland. *Cities*, 61, pp. 65–73 (2017).

Pebesma, E. spacetime: Spatio-temporal data in R. *Journal of Statistical Software*, 51(7), pp. 1–30 (2012). R package version 1.2-1. `https://CRAN.R-project.org/package=spacetime`.

Pebesma, E. and Klus, B. trajectories: Classes and methods for trajectory data (2015). R package version 0.1-4. `https://CRAN.R-project.org/package=trajectories`.

Pelekis, N., Andrienko, G., Andrienko, N., Kopanakis, I., Marketos, G., and Theodoridis, Y. Visually Exploring Movement Data via Similarity-Based Analysis. *Journal of Intelligent Information Systems*, 38(2), pp. 343–391 (2012).

Plank, A. The Hidden Risk in User-Generated Content: An Investigation of Ski Tourers' Revealed Risk-Taking Behavior on an Online Outdoor Sports Platform. *Tourism Management*, 55, pp. 289–296 (2016).

Porikli, F. Trajectory Distance Metric Using Hidden Markov Model based Representation. In *IEEE European Conference on Computer Vision, PETS Workshop*, volume 3. IEEE (2004).

Purves, R., Laube, P., Buchin, M., and Speckmann, B. Moving Beyond the Point: An Agenda for Research in Movement Analysis with Real Data. *Computers, Environment and Urban Systems*, 47, pp. 1–4 (2014).

Purves, R. S. Methods, Examples and Pitfalls in the Exploitation of the Geospatial Web. In *The Handbook of Emergent Technologies in Social Research* (Edited by S. Nagy Hesse-Biber), chapter 26, pp. 592 – 624. Oxford University Press (2011).

Ranacher, P. and Tzavella, K. How to Compare Movement? A Review of Physical Movement Similarity Measures in Geographic Information Science and Beyond. *Cartography and Geographic Information Science*, 41(3), pp. 286–307 (2014).

Rasetic, S., Sander, J., Elding, J., and Nascimento, M. A Trajectory Splitting Model for Efficient Spatio-Temporal Indexing. In *Proceedings of the Thirty-first International Conference on Very Large Data Bases - Volume 31*, pp. 934–945. VLDB Endowment, Trondheim, Norway (2005).

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., and Andrienko, G. Visually Driven Analysis of Movement Data by Progressive Clustering. *Information Visualization*, 7(3-4), pp. 225–239 (2008).

Schmudlach, G. Einführung Skitourenguru. Technical report, www.skitourenguru.ch (2017).

Schmudlach, G. and Köhler, J. Automated Avalanche Risk Rating of Backcountry Ski Routes. In *International Snow Science Workshop*, pp. 450–456. Beckenridge, CO, USA (2016).

Schweizer, J. Snow Avalanche Formation. *Reviews of Geophysics*, 41(4) (2003).

Seeger, C. J. The Role of Facilitated Volunteered Geographic Information in the Landscape Planning and Site Design Process. *GeoJournal*, 72(3-4), pp. 199–213 (2008).

Sester, M., Feuerhake, U., Kuntzsch, C., and Zhang, L. Revealing Underlying Structure and Behaviour from Movement Data. *KI - Künstliche Intelligenz*, 26(3), pp. 223–231 (2012).

Sharif, M. and Alesheikh, A. A. Context-Awareness in Similarity Measures and Pattern Discoveries of Trajectories: A Context-Based Dynamic Time Warping Method. *GIScience and Remote Sensing*, 54(3), pp. 426–452 (2017).

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., and Fotheringham, A. S. Analysis of Human Mobility Patterns from GPS Trajectories and Contextual

Information. *International Journal of Geographical Information Science*, 30(5), pp. 881–906 (2016).

SLF. Danger levels (2018a). Retrieved 4 June 2018, from `https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin/danger-levels.html`.

SLF. Division into regions (2018b). Retrieved 26 March 2018, from `https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/about-the-avalanche-bulletin/division-into-regions.html`.

SLF. Lawinenkunde und prävention (2018c). Retrieved 23 May 2018, from `https://www.slf.ch/de/lawinen/lawinenkunde-und-praevention.html`.

SLF. Winterbericht 2017/18 (2018d). Retrieved 31 May 2018, from `https://www.slf.ch/de/lawinenbulletin-und-schneesituation/wochen-und-winterberichte/201718/winterbericht.html`.

Sørensen, R. and Seibert, J. Effects of DEM Resolution on the Calculation of Topographical Indices: TWI and Its Components. *Journal of Hydrology*, 347(1-2), pp. 79–89 (2007).

Sui, D. Alternative GIS (alt.gis) and the Six Sensesof the New Mind: Is alt.gis TransformingGIS into a Liberation Technology? In *Advances in Spatial Data Handling and Analysis* (Edited by F. Harvey and Y. Leung), chapter 1, pp. 1–11. Springer, Cham (2015).

Sui, D., Elwood, S., and Goodchild, M. Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice* (Edited by D. Sui, S. Elwood, and M. Goodchild), chapter 1, pp. 1–12. Springer, Dordrecht (2013).

Suter, C. and Harvey, S. Mobile Information Systems for Avalanche Topics. In *International Snow Science Workshop*, pp. 391–394. Davos, Switzerland (2009).

Swisstopo. DHM25 - Das digitale Höhenmodell der Schweiz. Technical report, Bundesamt für Landestopographie, Wabern, Switzerland (2005).

Swisstopo. swissALTI3D - Das hoch aufgelöste Terrainmodell der Schweiz. Technical report, Bundesamt für Landestopographie, Wabern, Switzerland (2014).

Swisstopo. swissTLM3D Version 1.3. Technical report, Bundesamt für Landestopographie, Wabern, Switzerland (2015).

Techel, F. and Zweifel, B. Recreational Avalanche Accidents in Switzerland: Trends and Patterns with an Emphasis on Burial, Rescue Methods and Avalanche Danger. In *International Snow Science Workshop*. Grenoble, France (2013).

Techel, F., Zweifel, B., and Winkler, K. 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), pp. 1985–1997 (2015).

Thompson, J. A., Bell, J. C., and Butler, C. A. Digital Elevation Model Resolution: Effects on Terrain Attribute Calculation and Quantitative Soil-Landscape Modeling. *Geoderma*, 100(1-2), pp. 67–89 (2001).

Tottrup, A. P., Klaassen, R. H. G., Strandberg, R., Thorup, K., Kristensen, M. W., Jorgensen, P. S., Fox, J., Afanasyev, V., Rahbek, C., and Alerstam, T. The Annual

Cycle of a Trans-Equatorial Eurasian-African Passerine Migrant: Different Spatio-Temporal Strategies for Autumn and Spring Migration. *Proceedings of the Royal Society B: Biological Sciences*, 279(1730), pp. 1008–1016 (2012).

Turner, A. J. *Introduction to Neogeography*. O'Reilly Media (2006).

Vickery, G. and Wunsch-Vincent, S. Participative Web And User-Created Content: Web 2.0 Wikis and Social Networking. Technical report, Organization for Economic Cooperation and Development (OECD), Paris (2007).

Vlachos, M., Kollios, G., and Gunopulos, D. Discovering Similar Multidimensional Trajectories. In *Proceedings 18th International Conference on Data Engineering*, pp. 673–684. IEEE (2002).

Vontobel, I. *Geländeanalysen von Unfalllawinen*. Msc thesis, University of Zurich (2011).

Wang, W., Yang, J., and Muntz, R. STING : A Statistical Information Grid Approach to Spatial Data Mining. In *Proceedings of the Twenty-third International Conference on Very Large Data Bases - Volume 23*, pp. 186–195. VLDB Endowment, Athens, Greece (1997).

Wasserstein, R. L. and Lazar, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), pp. 129–133 (2016).

Wickham, H. Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), pp. 1–20 (2007). R package version 1.4.3. `https://CRAN.R-project.org/package=reshape2`.

Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York (2009). R package version 2.2.1. `https://CRAN.R-project.org/package=ggplot2`.

Wickham, H. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), pp. 1–29 (2011). R package version 1.8.4. `https://CRAN.R-project.org/package=plyr`.

Wickham, H., Francois, R., Henry, L., and Müller, K. dplyr: A grammar of data manipulation (2017). R package version 0.7.4. `https://CRAN.R-project.org/package=dplyr`.

Wiersma, Y. Birding 2.0: Citizen Science and Effective Monitoring in the Web 2.0 World. *Avian Conservation and Ecology*, 5(2) (2010).

Wilke, C. O. cowplot: Streamlined plot theme and plot annotations for 'ggplot2' (2017). R package version 0.9.2. `https://CRAN.R-project.org/package=cowplot`.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., and Aberer, K. SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories. In *Proceedings of the 14th International Conference on Extending Database Technology*, pp. 259–270. ACM, Uppsala, Sweden (2011).

Yan, Z., Parent, C., Spaccapietra, S., and Chakraborty, D. A Hybrid Model and Computing Platform for Spatio-Semantic Trajectories. In *Proceedings of the 7th International Conference on The Semantic Web: Research and Applications - Volume Part I*, pp. 60–75. Springer, Heraklion, Greece (2010).

Yoon, H. and Shahabi, C. Robust Time-Referenced Segmentation of Moving Object Trajectories. In *2008 Eighth IEEE International Conference on Data Mining*, pp. 1121–1126. IEEE, Pisa, Italy (2008).

Zevenbergen, L. W. and Thorne, C. R. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, 12, pp. 47–56 (1987).

Zhang, T., Ramakrishnan, R., and Livny, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pp. 103–114. ACM, Montreal, Canada (1996).

Zhang, Z., Huang, K., and Tan, T. Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes. In *18th International Conference on Pattern Recognition*, volume 3, pp. 1135–1138. IEEE, Hong Kong, China (2006).

Zook, M., Graham, M., Shelton, T., and Gorman, S. Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2), pp. 7–33 (2010).

Zweifel, B., Pielmeier, C., Marty, C., and Techel, F. Schnee und Lawinen in den Schweizer Alpen - Hydrologisches Jahr 2016/2017. Technical report, WSL-Institut für Schnee und Lawinenforschung SLF, Davos, Switzerland (2017).

Zweifel, B., Pielmeier, C., Marty, C., Techel, F., Stoffel, L., Ruesch, M., and Bühler, Y. Schnee und Lawinen in den Schweizer Alpen - Hydrologisches Jahr 2015/2016. Technical report, WSL-Institut für Schnee und Lawinenforschung SLF, Davos, Switzerland (2016).

Zweifel, B., Raez, A., and Stucki, T. Avalanche Risk for Recreationists in Backcountry and in Off-Piste Area: Surveying Methods and Pilot Study at Davos, Switzerland. In *International Snow Science Workshop*, pp. 733–741. Telluride, CO, USA (2006).

Appendices

Appendix A

Database Diagram

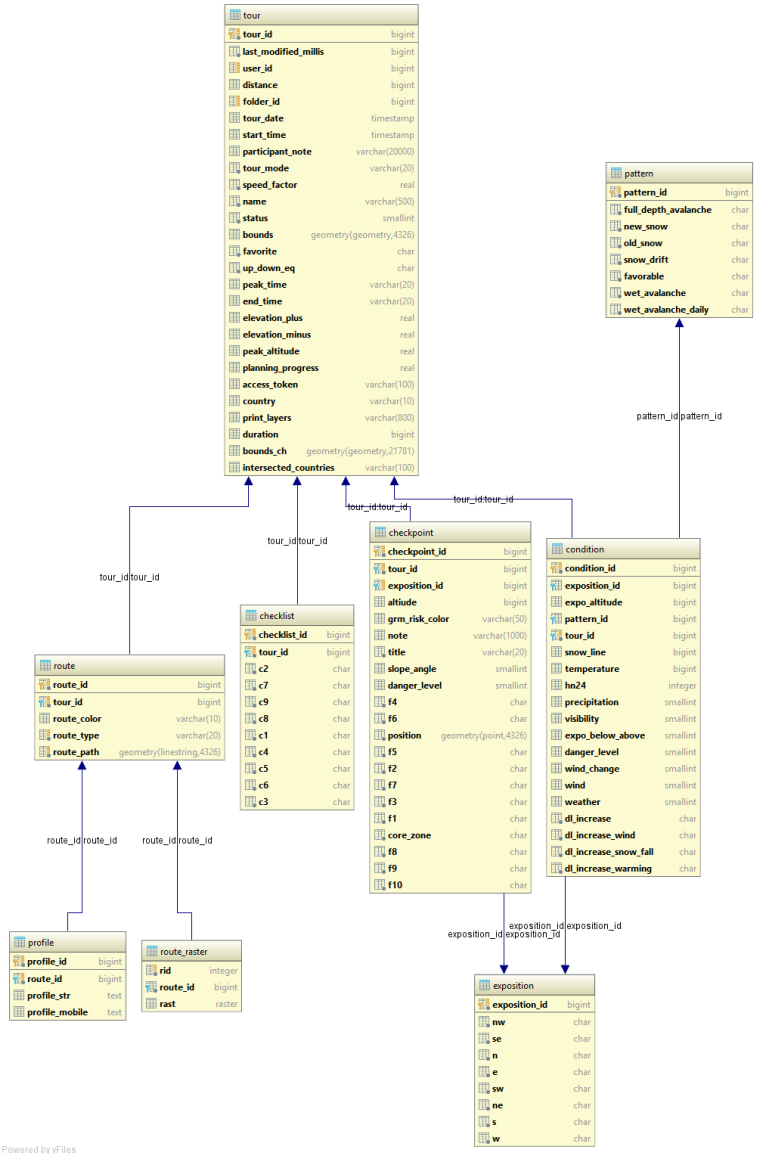


Figure A.1: *White Risk database diagram.*

Appendix B

Warning Regions

Western Part of the Northern flank of the Alps

1111 Vaud Prealps
1112 Pays d'Enhaut
1113 Aigle-Leysin
1114 Bex-Villars
1121 Fribourg Alps
1211 Western Bernese Prealps
1212 Eastern Bernese Prealps
1213 Hohgant
1221 Niedersimmental
1222 Gstaad
1223 Wildhorn
1224 Lenk
1225 Iffigen
1226 Adelboden
1227 Engstligen
1228 Obersimmental
1231 Kandersteg
1232 Blüemlisalp
1233 Lauterbrunnen
1234 Jungfrau – Schilthorn
1241 Brienz-Interlaken
1242 Grindelwald
1243 Schreckhorn
1244 Hasliberg – Rosenlaui
1245 Guttannen
1246 Gadmertal
1247 Grimsel Pass
1311 Vouvry
1312 Monthey-Val d'Illiez

Central Part of the Northern flank of the Alps

2111 Entlebuch
2121 Glaubenberg
2122 Engelberg
2123 Melchtal
2131 Schwyz Prealps
2132 Muotatal
2211 Schächental
2212 Uri Rotstock
2221 Meiental
2222 Maderanertal
2223 Northern Urseren
2224 Southern Urseren

Eastern Part of the Northern flank of the Alps

3111 Northern and central Glarus
3112 Southern Glarus-Grosstal
3113 Southern Glarus-Sernftal
3211 Appenzell Alps
3221 Toggenburg
3222 Alpstein – Alvier
3223 St Gallen Oberland
3311 Liechtenstein

Valais

4111 Emosson
4112 Génépi
4113 Val d'Entremont-Val Ferret
4114 Conthey-Fully
4115 Martigny-Verbier
4116 Haut Val de Bagnes
4121 Montana
4122 Val d'Hérens
4123 Arolla
4124 Val d'Anniviers
4125 Mountet
4211 Lötschental
4212 Turtmanntal
4213 Aletsch region
4221 Lower Valleys of Visp
4222 Upper Mattertal
4223 Upper Saastal
4231 Northern Simplon region
4232 Southern Simplon region
4241 Reckingen
4242 Binntal
4243 Northern Obergoms
4244 Southern Obergoms

Northern and Central Grisons

5111 Northern Prättigau
5112 Southern Prättigau
5113 Western Silvretta
5121 Calanda
5122 Schanfigg
5123 Davos
5124 Flims
5211 Northern Tujetsch
5212 Southern Tujetsch
5214 Obersaxen – Safien Valley
5215 Val Sumvitg
5216 Zervreila
5221 Domleschg – Lenzerheide
5222 Schams
5223 Rheinwald
5231 Albulatal
5232 Savognin
5233 Avers
5234 Bivio

Central Part of the Southern flank of the Alps

6111 Val Bedretto
6112 Upper Valle Leventina
6113 Val Blenio
6114 Upper Valle Maggia
6115 Lower Valle Leventina
6121 Lower Valle Maggia
6122 Riviera
6131 Lugano area
6132 Mendrisio area
6211 alto Moesano
6212 basso Moesano

Engadine / eastern Part of the Southern flank of the Alps

7111 Corvatsch
7112 Bernina region
7113 Zuoz
7114 St Moritz
7115 Val Chamuera
7121 Samnaun
7122 Eastern Silvretta
7123 Sur Tasna
7124 Val Suot
7125 Val dal Spöl
7126 Val S-charl
7211 Val Bregaglia
7221 Upper Val Poschiavo
7222 Lower Val Poschiavo
7231 Val Müstair

Jura

8111 Saint-Cergue
8112 Vallée de Joux
8113 Yverdon - Bevaix
8114 Val de Travers
8211 Val de Ruz - Colombier
8212 Bienne - Neuchâtel
8213 Vallon de Saint-Imier
8214 Moutier - Tavannes
8215 Thal
8216 Olten-Gösgen
8221 La Chaux-de-Fonds - Le Locle
8222 Franches-Montagnes
8223 Ajoie
8224 Delémont - Bellelay
8225 Laufental
8226 Basel
8227 Oberes Baselbiet
8228 Rheinfelden

Swiss plateau

9111 western Swiss Plateau
9211 central Swiss Plateau
9311 eastern Swiss Plateau

Figure B.1: *SLF warning regions including names.*

Appendix C

Curvature Boxplot

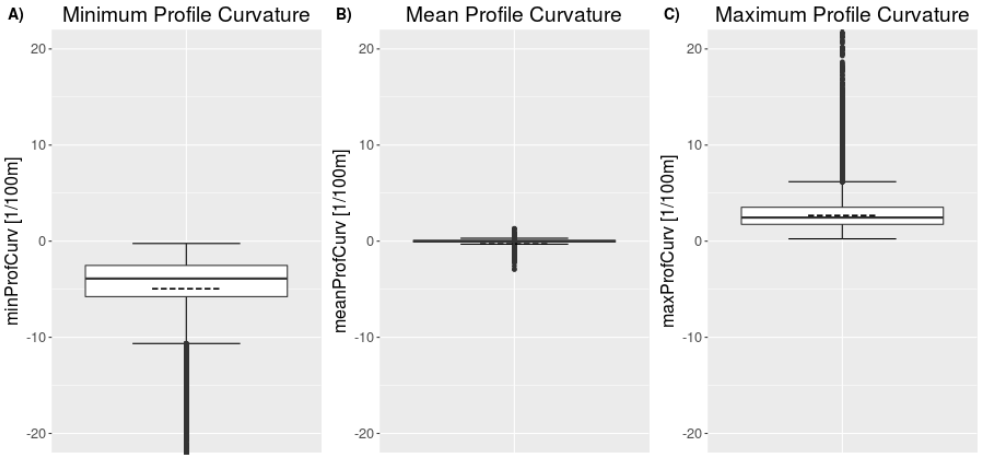


Figure C.1: *Boxplots of attributes minProfCurv, meanProfCurv and maxProfCurv.*

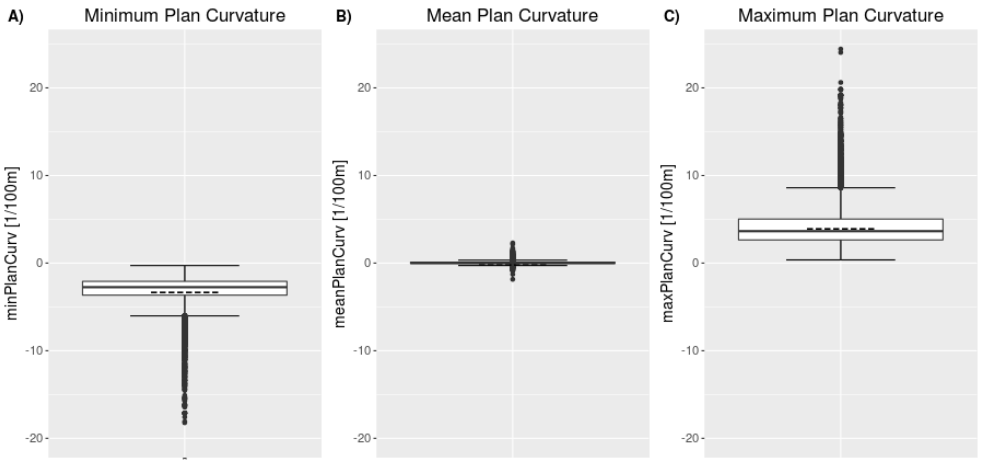


Figure C.2: *Boxplots of attributes minPlanCurv, meanPlanCurv and maxPlanCurv.*

Appendix D

Attribute Correlation with Avalanche Danger

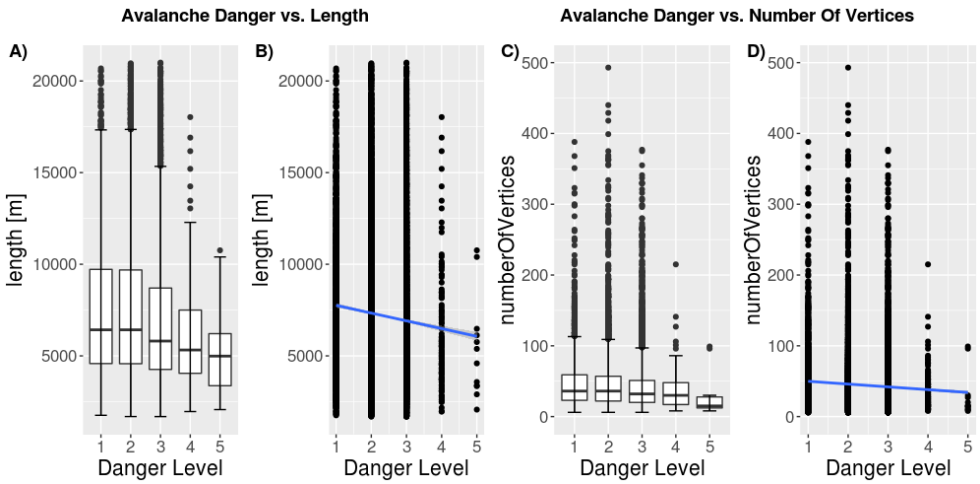


Figure D.1: *A) Boxplot of length, grouped by avalanche danger level. B) Scatterplot of length and avalanche danger level, including linear model. C) Boxplot of number of vertices, grouped by avalanche danger level. D) Scatterplot of number of vertices and avalanche danger level, including linear model.*

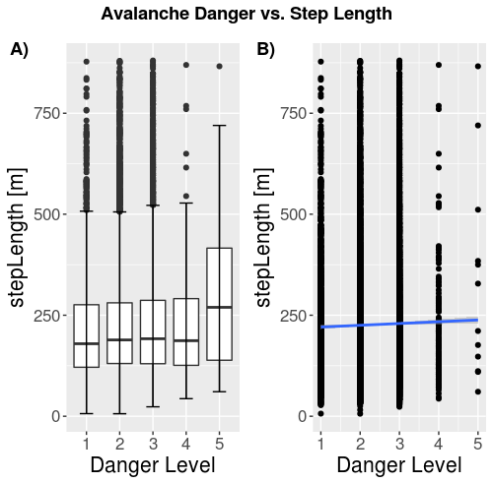


Figure D.2: *A) Boxplot of step length, grouped by avalanche danger level. B) Scatterplot of step length and avalanche danger level, including linear model.*

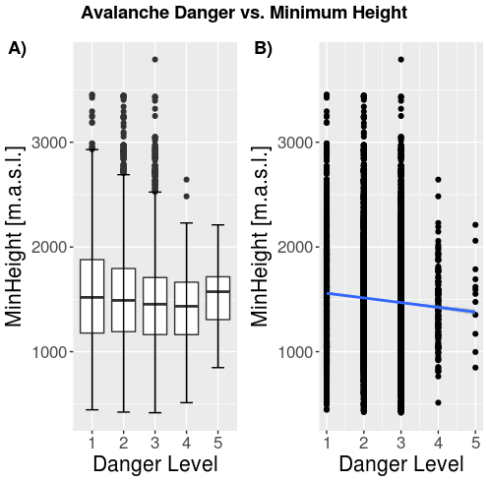


Figure D.3: *A) Boxplot of minimum height, grouped by avalanche danger level. B) Scatterplot of minimum height and avalanche danger level, including linear model.*

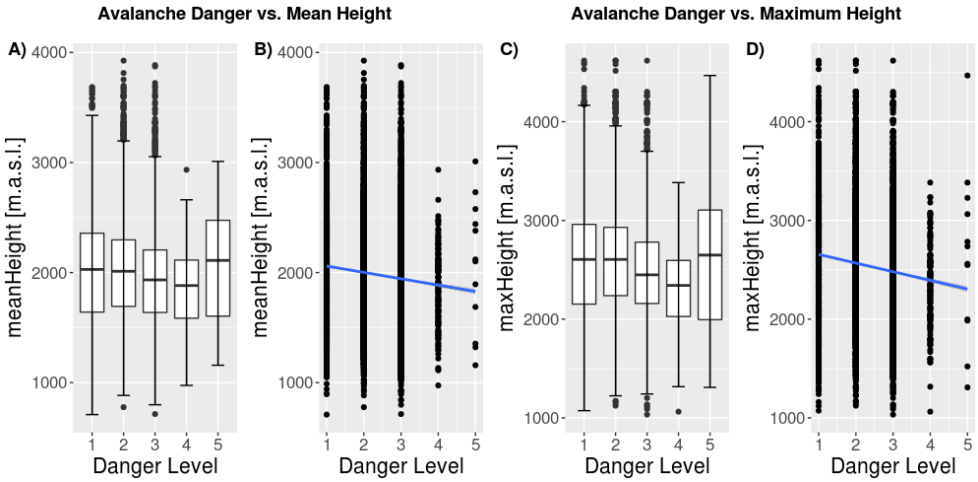


Figure D.4: *A) Boxplot of mean height, grouped by avalanche danger level. B) Scatterplot of mean height and avalanche danger level, including linear model. C) Boxplot of maximum height, grouped by avalanche danger level. D) Scatterplot of maximum height and avalanche danger level, including linear model.*

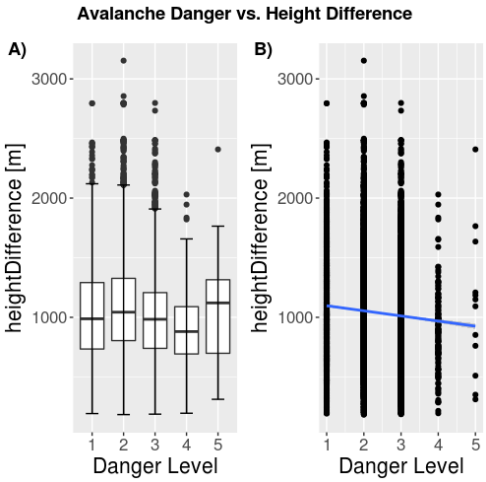


Figure D.5: *A) Boxplot of height difference, grouped by avalanche danger level. B) Scatterplot of height difference and avalanche danger level, including linear model.*

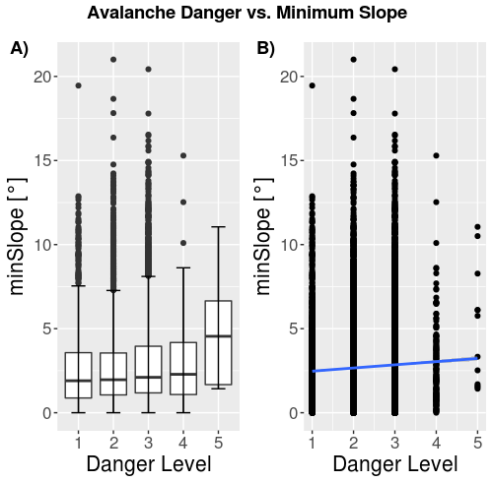


Figure D.6: *A) Boxplot of minimum slope, grouped by avalanche danger level. B) Scatterplot of minimum slope and avalanche danger level, including linear model.*

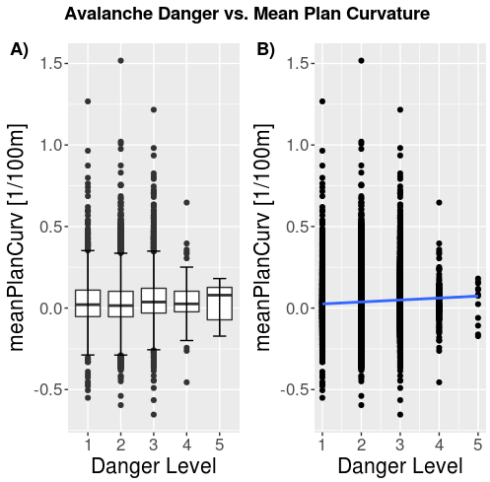


Figure D.7: *A) Boxplot of mean plan curvature, grouped by avalanche danger level. B) Scatterplot of mean plan curvature and avalanche danger level, including linear model.*

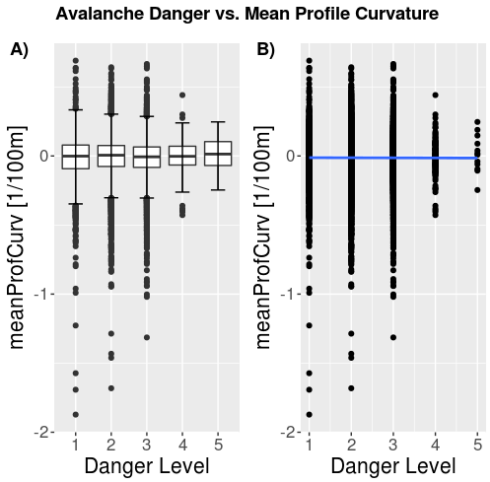


Figure D.8: *A) Boxplot of mean profile curvature, grouped by avalanche danger level. B) Scatterplot of mean profile curvature and avalanche danger level, including linear model.*

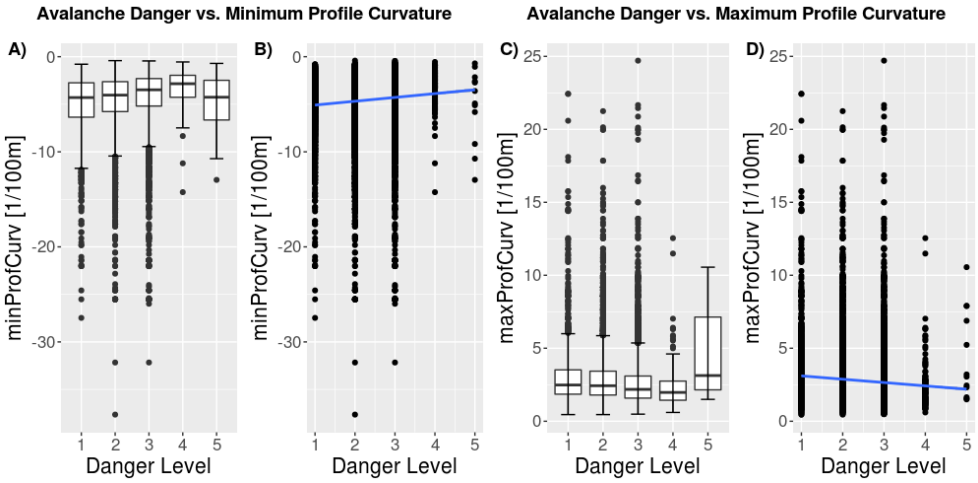


Figure D.9: *A) Boxplot of minimum profile curvature, grouped by avalanche danger level. B) Scatterplot of minimum profile curvature and avalanche danger level, including linear model. C) Boxplot of maximum profile curvature, grouped by avalanche danger level. D) Scatterplot of maximum profile curvature and avalanche danger level, including linear model.*

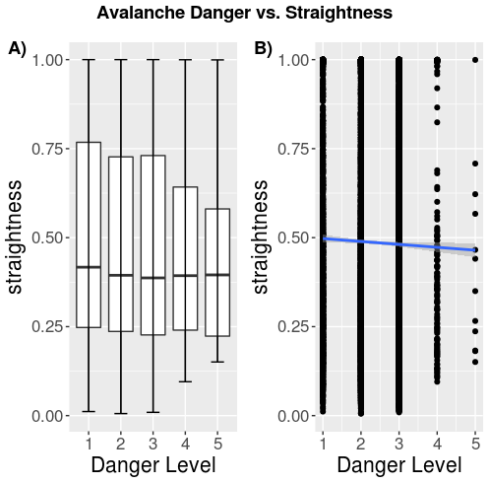


Figure D.10: *A) Boxplot of straightness, grouped by avalanche danger level. B) Scatterplot of straightness and avalanche danger level, including linear model.*

Appendix E

Boxplots Participation Inequality

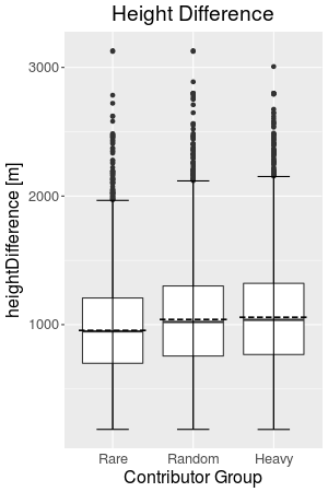


Figure E.1: *Boxplot of attribute heightDifference.*

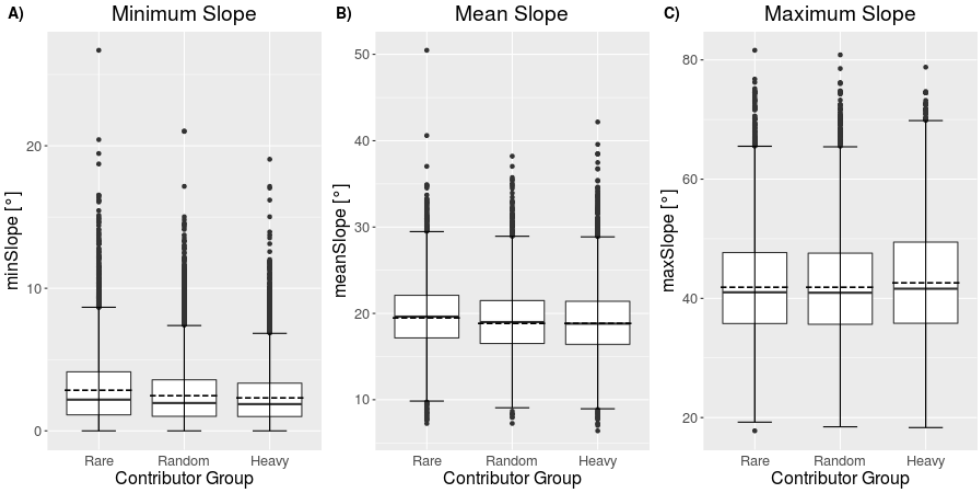


Figure E.2: *Boxplots of attributes A) minSlope, B) meanSlope, and C) maxSlope slope.*

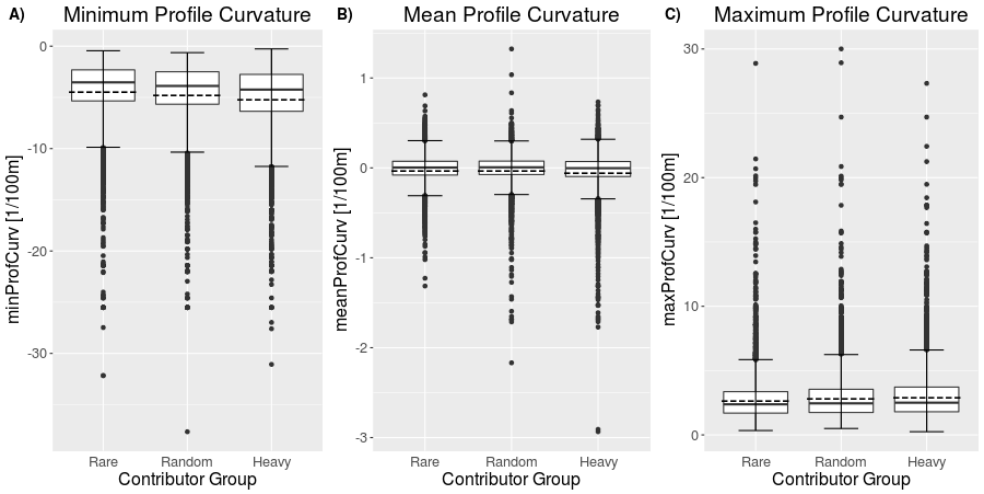


Figure E.3: *Boxplots of attributes minProfCurv, meanProfCurv and maxProfCurv.*

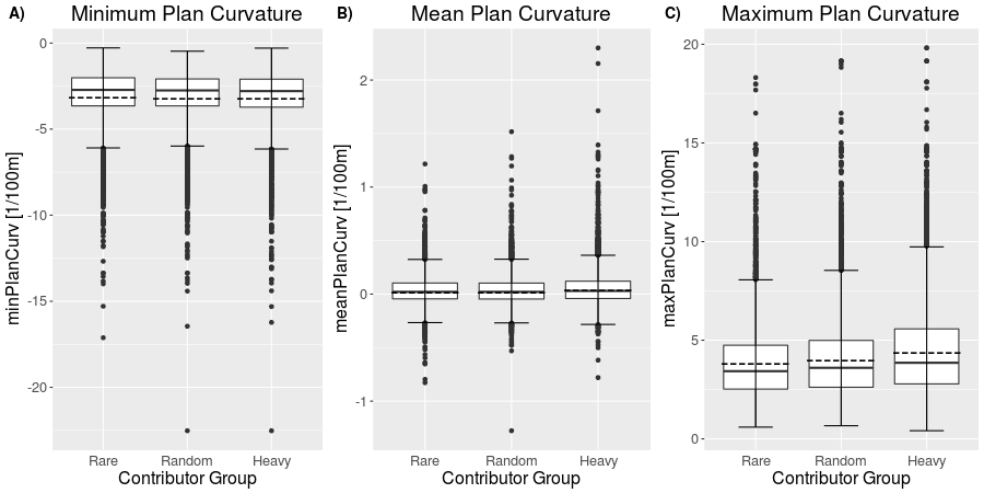


Figure E.4: *Boxplots of attributes minPlanCurv, meanPlanCurv and maxPlanCurv.*

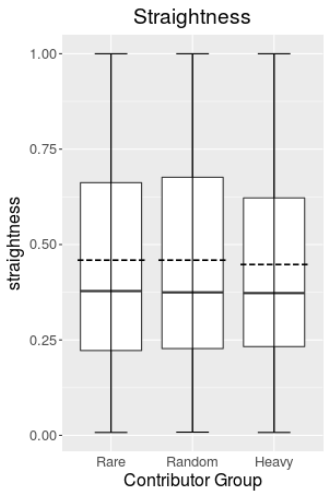


Figure E.5: *Boxplot of attribute straightness.*

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work.
All external sources are explicitly acknowledged in the thesis.

27.07.2018

Date

Christoph Schönenberger