Geographical Information Management and Applications

Exposure to avalanche terrain in crowd-sourced ski touring route plans

Final thesis report

Student: Maarten Hogeweij
Supervisors:
Dr. F.O. Ostermann (University of Twente)
Prof.dr. R. Purves (Zürich University)
S. Harvey (WSL Institute for Snow and Avalanche Research SLF)
Responsible professor:
Prof.dr. M.J. Kraak (University of Twente)

Universiteit Utrecht

TU Delft

UNIVERSITY OF TWENTE.

WAGENINGEN UR

Abstract

Recreational snow avalanche accidents claim about 100 lives yearly in the Alps. Since those avalanches are often triggered by the recreationist themselves, growing research attention is given to the routes travelled by backcountry recreationists. This thesis aims analyze the relation between avalanche terrain and planned ski tours. This is done using a dataset of over 50.000 planned ski tours and 800 GPS tracks from Switzerland, combined with new data on avalanche terrain in Switzerland. This data allows for new ways to classify and compare routes based on the terrain they intersect.

First, the mean and 95th percentile values for terrain hazard in routes were calculated. Then, the routes were clustered using k-means based on the avalanche terrain they intersect. They were also clustered spatially using DBScan. Then, the relation between avalanche terrain and route location was investigated. Lastly, the GPS tracks and planned routes were compared to place the planned routes into a broader context.

The new terrain data proved useful to differentiate between routes based on avalanche terrain. Five avalanche terrain clusters were found, with 17% of routes being part of a steep terrain cluster. 34% of routes were assigned to 655 spatial clusters. The spatial clusters were generally quite homogeneous in terms of avalanche terrain. Heterogeneity within a cluster was often caused by routes following different planning strategies. In some cases, official ski touring routes from the Swiss Alpine Club were more hazardous in terms of terrain than routes drawn by individuals. The GPS tracks intersected less hazardous terrain in the entire dataset, but within individual spatial clusters they often intersected more hazardous terrain. In the future, the results from this could be used to suggest safe routes to recreationists between given start and end points.

Acknowledgements

This thesis would not have been possible without the efforts from a number of people. First, I would like to thank Stephan Harvey from the Swiss avalanche institute SLF. He provided me with the collected data of planned ski tours, and also let me use the data from his avalanche terrain model. Besides this, he also gave a lot of useful inputs on the content of this thesis, taking time to meet me both on Skype and in person. In the beginning there were a lot of different analysis directions possible, and Stephan's input helped me focus on the important parts.

Ross Purves from the University of Zürich was very helpful and provided me with many useful ideas regarding the content and methods of the research. I'm especially thankful that he and Stephan took time to let me visit them in Switzerland, which was very fun and informative. The creativity of Ross regarding possible methods to analyze the data really inspired me.

Frank Ostermann, my supervisor, gave me a lot of help in structuring my thoughts and my writing. It was at times a struggle to keep the overarching goal of the research in mind, and Frank helped me keep track of the logic of the argumentation used.

Numerous people I had conversations with at the International Snow Science Workshop in Innsbruck inspired me and provided interesting ideas, which I am thankful for. Lastly, I'd like to thank my parents for introducing me to the mountains and Lotte for her love and support.

I hope you enjoy reading this thesis, I enjoyed writing it.

Maarten Hogeweij
Amsterdam, 24-02-2019

Table of contents

1. Introduction

1.1. Avalanche hazard and recreational behavior

Avalanches are one of the main hazards faced by backcountry winter recreationists in the Alps and mountainous regions elsewhere. Around 100 deaths are recorded in the Alps every year as a consequence of avalanches (Techel et al., 2016). Most of those are recreationists travelling in the mountains, for example backcountry skiers and mountaineers. Because of this, much research has been devoted to improving the knowledge needed to avoid recreational avalanche accidents. An important focus of this research is on the geophysical properties of avalanches, such as the terrain types where they are likely to occur, the snowpack properties, and the influence of weather (Zweifel & Haegeli, 2014). This has led to a body of knowledge on the properties of recreational avalanche accidents. For example, it is known that in recreational avalanche accidents, the avalanche is most often triggered in slopes between 35 and 40 degrees, with a north-facing aspect, and that fresh snow and heavy wind are factors that make avalanches more likely (Tremper, 2008). This knowledge has helped improve avalanche education courses, which now aim to teach backcountry recreationists to detect and avoid areas and situations where triggering and getting caught in an avalanche is more likely. However, despite this knowledge, still many people choose to venture in this terrain, and take high risks (Hendrikx, 2018). Analysis of past avalanche accidents shows that in many cases the people who got caught or triggered the avalanche were educated, and knew the hazard, yet still chose to ski a hazardous slope (McCammon, 2002).

This gap between knowledge of risks among recreationists, and acting based on this knowledge, has led to another research theme in recreational avalanche prevention, the human factor (Hallandvik, Andresen, & Aadland, 2017). Instead of focusing on geophysical properties of avalanches, the human factor focuses its research on the role of humans in triggering recreational avalanches. The reason for this is that in the vast majority of recreational avalanche accidents, the avalanche was triggered by the skier themselves, or somebody from their party (Techel & Zweifel, 2013). This makes recreational avalanche accidents different from many other natural hazards: because humans cause their own avalanches, humans are also able to avoid causing those avalanches. Still, as noted before, in many cases recreationists still fail to avoid causing avalanches and getting caught in them.

Recently, GPS devices have become more commonly used by recreationists to track their routes. This emergence of high-resolution route data makes it possible to analyse routes commonly travelled in the backcountry more accurately than before. When this route data is combined with data on avalanche terrain, it can be used to assess how hazardous a route is. There are two reasons why this is useful. First, it helps assess the risks people take related to avalanche terrain in their ski tours, which adds to insight into the human factor of avalanches. The importance of this knowledge is exemplified by various recent publications that focus on studying the routes people travel in the backcountry (Haegeli & Atkins, 2016; Hendrikx & Johnson, 2016).

Secondly, it is useful to evaluate the hazards in common ski touring routes in a fine granularity. In most descriptions of ski touring routes found in route books or online, a broad description of the route and possible hazards is provided[1]. However, when combining crowd-sourced route data with avalanche terrain data, it becomes possible to assess this hazard on a more detailed level. Also, it becomes possible to identify higher and lower hazard variants within a general route corridor, and to compare the hazard between different routes.

Another recent development has been on spatial modelling of potential avalanche terrain. As noted before, terrain factors (e.g. slope) play a crucial role in the probability of an avalanche being triggered, all other factors being equal. Therefore, using terrain data to model and classify avalanche terrain is

[1] For an example see:
https://www.gipfelbuch.ch/tourenfuehrer/routen/id/12310/Skitour_Snowboardtour/Chueenihorn

another focus in the avalanche research community. Many studies use a simple discrete classification such as between simple, challenging, or complex terrain (Campbell & Gould, 2013). Also, many studies up to this point have focused on the points where an avalanche can be triggered. However, a recent model by Harvey et al. (2018) uses both potential triggering points and the areas an avalanche can run out into. Also, the model takes into account to what extent serious injury is to be expected as a consequence of an avalanche. Therefore, it enables a more detailed modelling of the hazard in avalanche terrain.

In this thesis, the goal is to add to the body of knowledge of backcountry route trajectories in relation to potential avalanche terrain. For this, the output from the model from Harvey et al. (2018) is used. This dataset is combined with a dataset of 53553 planned ski tours, crowdsourced from the route-planning platform Whiterisk.ch. The main novelty is that this thesis is the first to use this newly modelled avalanche terrain data to analyse the hazard in routes. This is important as it allows for a classification of routes based on their hazard attributes. This has not been attempted before at such a high level of detail, since the avalanche terrain data used in this thesis is new. A secondary novelty is the fact that this thesis used planned route data. This is different from other studies which mostly use GPS tracks of actually travelled tours. The interesting thing about planned route data is that it gives an insight into the planning phase of backcountry ski touring, which is one of the distinct phases in which recreationists should try to avoid risky terrain. The insights gained from this work could therefore describe common mistakes made in route planning, which could help to prevent skiers planning hazardous routes.

1.2. Research objectives

1.2.1 Research questions

This research focuses on avalanche terrain and the relation this has to planned backcountry recreation routes. As such, the main research question is:

"What is the relation between planned backcountry routes and the surrounding avalanche terrain?"

The expectation is that people use the Whiterisk platform to plan routes that avoid avalanche terrain as much as possible. However, to a degree the type of avalanche terrain faced by routes is can't be avoided with careful planning, as there is no way around it with the given start and end points of a route. Therefore, the relation between route planning and avalanche terrain is complex and varies in different locations. The sub-questions go deeper into the specifics of this relationship. The first sub-question is related to the different ways in which avalanche terrain can be treated as a route attribute. This question is:

1. "Which different methods can be used to relate avalanche terrain to planned routes, and what are their (dis)advantages?"

The answer to this question will be a description of the methods used to establish avalanche terrain as a route attribute and its visualization in maps. Also, the advantages and disadvantages of the methods will be discussed. The second sub-question attempts to classify routes into similar groups based partly on the output from sub question 1:

2. "What clusters exist in the route data based on their avalanche terrain characteristics?"

Such clusters are interesting to find because they can tell us something about what types of routes are being planned. An example of such a category could be: a route that is in general safe, but has one very dangerous section. This is a more meaningful description of a route than just the route length or the mean slope encountered within a route. The expected outcome of this question is a list of the route categories and a description of their characteristics. These categories will be called avalanche terrain clusters from this point onwards. Given the large number of routes, it is impossible to discern patterns in the route data by eye, therefore a clustering algorithm will be used to categorize the routes. Besides patterns in the avalanche terrain data in routes, there are also spatial patterns in the routes. The next question is related to this:

3. "What spatial clusters exist in the route data?"

For this, a clustering algorithm will be used. The aim is to filter out route corridors. One glance at the route data shows that there are specific areas of high density where many routes are very similar. This is mostly related to typical start and end points, e.g. towns and mountain peaks. Also, many trajectories follow known routes, for example as defined by the Schweizer Alpen-Club (SAC). Clustering routes spatially helps when comparing them, since it enables comparing routes within same spatial cluster directly to each other, which helps highlight route-planning decisions. Also, it allows to analyse to what extent the avalanche terrain in routes (from question 1) is related to their spatial location. It will be interesting to see whether it is possible to plan different routes in terms of avalanche terrain within the same spatial corridor. The next question goes deeper into this:

4. “To what extent is avalanche terrain similar for routes within a spatial cluster?”

The reason why this is interesting is that it shows how detailed route planning can be used to avoid avalanche terrain. If some routes within a spatial cluster are safer than others, this means that detailed planning is useful in that cluster. To answer this question, the variety of terrain clusters in each spatial cluster will be calculated, as well as the variety of the terrain hazard within routes. Also, visual inspection will be used to investigate locations where routes are either very homogeneous or heterogeneous in terms of avalanche terrain.

The last question concerns the GPS tracks of routes, downloaded from Wikiloc. It will be interesting to see if the common attributes of planned routes, defined in the prior questions, are also to be found in GPS tracks of routes. While the planned routes and GPS tracks cannot be directly compared, since they were made by a different group of people, there could be common characteristics. This question is:

5. "Is there a difference in avalanche terrain of the planned routes and the GPS tracks?"

To answer this question, the methods from question 1 to relate avalanche terrain and routes will be used. Also, the terrain clusters from question 2 will be computed for the GPS tracks. This will allow for a quantitative analysis of avalanche terrain in both datasets. Then, popular corridors are taken where many GPS tracks and planned routes are, and visual inspection is used to compare them. This will serve to put the planned routes into a broader context.

1.2.2. Planned routes in relation to an avalanche hazard map

The main problem that this thesis addresses is the question how planned routes in the backcountry are related to exposure to avalanche terrain. This is interesting and relevant because little research has been done on how recreationists deal with avalanche hazard in the route planning phase. This is important to know as it can be used to improve avalanche education, as well as provide more insight than before in the relative hazard of different ski tours. The scientific relevance is in the fact that this thesis adds to the body of knowledge of terrain preferences of backcountry recreationists. The societal relevance is in the fact that knowledge from this thesis can be used to improve avalanche education. Also, the evaluation of hazard in the routes studied in this thesis can be used by backcountry skiers who want to compare safe touring options. Because of the new avalanche terrain data used this can be assessed more accurately than before. Finally, this thesis is relevant because it serves as an exploration of the modelled avalanche terrain data from Harvey et al (2018). This data is a new approach of modelling avalanche terrain, and it will be interesting to see how it relates to typical backcountry routes, as well as its usability for analysing routes.

1.2.2. New approaches in this thesis

The data that is used comprises ski tours planned on Whiterisk.ch, an application that allows users to plan their ski routes in a map where some of the avalanche hazard terrain factors are visualised. This is a new approach for two reasons. First, as noted before, it uses planned routes as opposed to recorded GPS tracks. Second, it uses information that was contributed for personal goals as opposed to research purposes. Most other studies in this domain use volunteers who deliberately track their routes with the idea that it is then used in scientific research. The people contributing data in this thesis drew their routes to use for themselves, unknowing that it would be used in a research. In other words, it can be seen as contributed instead of volunteered data (Harvey, 2013). A third way in which this thesis uses a new approach, is by making use of a new dataset in which avalanche terrain hazard is mapped continuously. This is different from previous studies into this domain, which either used separate terrain metrics (such as slope steepness), or used a simple discrete terrain classification, for example ATES which maps avalanche terrain into three classes (Campbell & Gould, 2013).

The Whiterisk.ch dataset was already used in the MSc thesis of Christoph Schönenberger (2018). This thesis builds on his in the following ways. First, the new avalanche terrain model from Harvey et al. (2018) is used. Secondly, the main focus of Schönenberger was to process the raw route data so that they could be used in research. In this thesis the focus will be more in-depth on the avalanche terrain in routes.

1.2.3. Analysis schema

The prior two pages have given an overview of the research objectives. In the schema on the next page, the structure and reasoning in the research objectives, as well as the analytical steps needed to achieve them, is outlined. The schema is tilted 90 degrees to improve ease of printing. The methods are briefly presented here, and will be further explained in chapter 3.

To summarize, the proposed outcomes of the research questions are as follows. Q1 will present a description of different ways in which avalanche terrain can be used as a route attribute, and a discussion of the advantages and disadvantages of these methods. Q2 will give a description of the clusters, and descriptive statistics showing the characteristics of the terrain clusters in routes. Also, a map will show the geographical of the terrain clusters. Q3 will show the number of clusters existing in the route data, as well as the percentage of routes that is appointed to a cluster. Statistics will be used to describe the characteristics of spatial clusters (e.g. the average number of routes per cluster). Q4 will have qualitative and quantitative outcomes. The qualitative outcome will be a description of a few selected spatial clusters, and show with maps how the avalanche terrain is similar or variable within those spatial clusters. The quantitative outcome will be a statistical measure of the variability of avalanche terrain within spatial clusters. Q5 will have similar methods. Again, maps are used to show how GPS tracks and planned routes are different in a few selected locations. The quantitative outcome here will be descriptive statistics, and will also describe how terrain cluster membership is different for GPS tracks and planned routes.

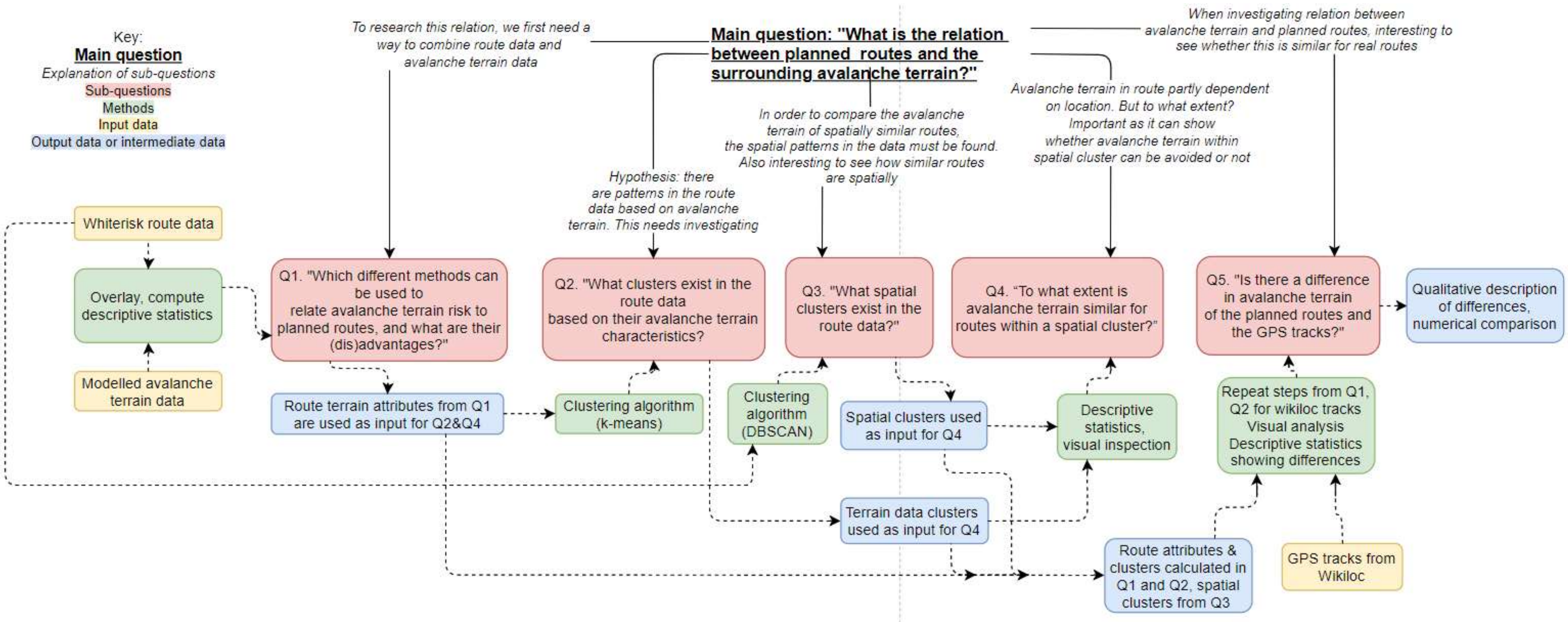

Key:
Main question
Explanation of sub-questions
Sub-questions
Methods
Input data
Output data or intermediate data
Main question: "What is the relation between planned routes and the surrounding avalanche terrain?"
To research this relation, we first need a way to combine route data and avalanche terrain data
Hypothesis: there are patterns in the route data based on avalanche terrain. This needs investigating
In order to compare the avalanche terrain of spatially similar routes, the spatial patterns in the data must be found. Also interesting to see how similar routes are spatially
Avalanche terrain in route partly dependent on location. But to what extent? Important as it can show whether avalanche terrain within spatial cluster can be avoided or not
When investigating relation between avalanche terrain and planned routes, interesting to see whether this is similar for real routes
Whiterisk route data
Overlay, compute descriptive statistics
Modelled avalanche terrain data
Q1. "Which different methods can be used to relate avalanche terrain risk to planned routes, and what are their (dis)advantages?"
Route terrain attributes from Q1 are used as input for Q2&Q4
Q2. "What clusters exist in the route data based on their avalanche terrain characteristics?
Clustering algorithm (k-means)
Q3. "What spatial clusters exist in the route data?"
Clustering algorithm (DBSCAN)
Spatial clusters used as input for Q4
Terrain data clusters used as input for Q4
Q4. "To what extent is avalanche terrain similar for routes within a spatial cluster?"
Descriptive statistics, visual inspection
Route attributes & clusters calculated in Q1 and Q2, spatial clusters from Q3
Q5. "Is there a difference in avalanche terrain of the planned routes and the GPS tracks?"
Repeat steps from Q1, Q2 for wikiloc tracks
Visual analysis
Descriptive statistics showing differences
GPS tracks from Wikiloc
Qualitative description of differences, numerical comparison

Figure 1. *Analytical schema (previous page)*

1.2.4. Thesis outline

The rest of this thesis is set-up as follows. In the next chapter, the scientific theories relevant to the subject are discussed. In chapter 3, an overview is given of the general methods and data used to answer the research questions. In chapter 4, the results are presented, and the research questions are answered. The more specific analytical steps needed to achieve specific results are also discussed in this chapter as opposed to chapter 3, in order to keep the methodological chapter orderly. Then, in chapter 5, a discussion of the results follows as well as the conclusions. The meaning of the results will be interpreted here, as well as the research limitations and recommendations for future research.

1.3. Definitions

Some of the terms used often in this thesis have ambiguous meanings in daily use. Therefore, the definitions used of those terms are now clarified.

1.3.1. Risk

Risk is a generally accepted concept, but there are slight differences between different risk definitions. For this thesis, the risk definition from disaster management literature is used. The term risk is defined as “a measure of the probability and severity of loss to the elements at risk, usually expressed for a unit area, object, or activity, over a specified period of time” (Bründl, Bartelt, Schweizer, Keiler, & Glade, 2011, p.54). Three elements are present in many scientific definitions of risk: hazard, exposure, and vulnerability. When those three elements are known, risk can be defined in an equation as:

$R = H \times E \times V$

Where R is risk, H is hazard, E is exposure, and V vulnerability. Hazard can be defined as:

"A potentially damaging physical event, phenomenon and/or human activity, which may cause loss of life or injury, property damage, social and economic disruption or environmental degradation." (Schneiderbauer & Ehrlich, 2004, p.10). Exposure in this context is the number of people at risk, and vulnerability is determined by how vulnerable those people are in the event of a disaster. It is important to note that in order for a situation to be risky, all three elements need to be present. For example, if there is a very steep slope with lots of fresh snow, H (hazard) will be very high. However, if there are no people present (E), or the people present are not vulnerable (V), there is still no risk. In avalanche situations, people present when an avalanche occurs are generally vulnerable, since it is usually impossible for humans to free themselves when they are caught in an avalanche. The exposure and hazard at a specific location do vary. In this thesis the focus is on avalanche terrain, and there is little knowledge of temporal aspects of the avalanche risk situation. Therefore, for the hazard part of the equation, the terrain hazard is used. The routes form the exposure part of the equation: if we assume people follow the routes they plan, then having routes intersect hazardous terrain implies exposure to this terrain. Therefore, sometimes the word "risk" is used in this thesis to describe routes that go through particularly hazardous terrain.

1.3.2. Terrain hazard

One of the datasets used in this thesis is of modelled avalanche terrain hazard. This is the amount of avalanche hazard the terrain carries. For this, both the potential of triggering an avalanche and the consequences of an avalanche happening are used as inputs. The "terrain-" part specifies that this is not the temporal avalanche, e.g. snowfall or wind influences on the snowpack. Those factors are communicated in the avalanche bulletin, but are not used in this thesis. To diversify, when discussing the temporal factors from the avalanche bulletin, the word "danger" is usually used instead of "hazard". This is in line with most English-language avalanche bulletins, that often use "danger rating" or "danger scale" to describe the temporal avalanche situation.

1.3.3. Triggering potential

In this thesis, triggering potential is used as a term to described the potential for slopes for triggering an avalanche. This is preferred above likelihood or probability, as those terms imply a measurable chance of an avalanche being triggered. However, to be able to calculate that, far more factors need to be considered, and even then it remains a broad guess. A distinction is made between triggering zones and remote triggering zones. In the former, an avalanche can start. In the latter, an avalanche is not likely to start, but they are connected to triggering zones above, where an avalanche can start by loading the remote triggering zone with extra weight (e.g. a skier). Lastly, there are runout zones, where no triggering or remote triggering potential is present, but where an avalanche triggered from above can runout into.

2. Literature review

In the following section, an overview is provided of the relevant theories surrounding this topic. This thesis topic combines three themes of research: avalanche research, VGI, and movement analysis. Based on this, the theoretical framework is divided into three sections.

2.1 Avalanche research

2.1.1. Introduction

The Cambridge dictionary defines an avalanche as "a large amount of ice, snow, and rock falling quickly down the side of a mountain." (Cambridge dictionary online, 2018). There are three types of snow avalanches: slab, wet snow, and loose snow avalanches. Slab avalanches occur when a homogeneous layer of snow starts sliding as a whole, as it loses contact with other layers beneath it as a consequence of increasing pressure. Wet snow avalanches happen when snow loses strength due to moisturizing as a consequence of rising temperatures or rainfall. Loose snow avalanches happen in very steep terrain (over 40 degrees) with snow that has little cohesion between snow crystals, usually due to persistently low temperatures.

Slab avalanches are by far the most dangerous, as the slabs are able of completely burying a person and suffocating them or causing deadly impact trauma, more so than loose-snow avalanches. Also, slab avalanches are able to reach very high speeds (up to 130km/h in 5 seconds, dependant on slope angle), more so than wet snow avalanches (Tremper, 2008). Because of this, most avalanche research focuses on slab avalanches (SLF, 2018). For backcountry recreationists, such as tour skiers, off-piste skiers, or snow-shoe hikers, slab avalanches are one of the major hazards to navigate. Around 100 deaths are recorded in the Alps every year as a consequence of avalanches, with the majority of those being mountain recreationists who triggered the avalanche themselves (Techel et al., 2016). Besides this, avalanches also endanger the infrastructure of mountainous regions, as well as being a potential threat to the lives of people living in mountainous regions.

Because of this, much academic attention is being devoted to understanding avalanche mechanics, the causes, and the effects of avalanches. Understanding these factors should help in decreasing the risks faced by mountain recreationists and inhabitants of mountainous regions alike. The aim of this research is not to list all the different directions that avalanche research has taken, but the main research fields will be outlined to serve as theoretical framework. To guide this theoretical review, avalanche research is divided into two main themes: research into the geophysical aspects of avalanches, and research into the human factor in avalanches. Other research fields include the technicalities of avalanche rescue (Ayuso, Cuchí, Lera, & Villarroel, 2015) and the sensing of avalanche occurrences (Eckerstorfer, Bühler, Frauenfelder, & Malnes, 2016). However, as the goal of this thesis is to analyse route skier choices related to avalanche terrain, these research fields are not relevant and will not be delved into further. The division between geophysical and human factors in avalanche accidents is often used in avalanche research (see for example Hendrikx, Johnson, & Southworth, 2013; Zweifel & Haegeli, 2014). Nonetheless, it is important to note that both human and geophysical factors are at play in recreational avalanche accidents, and as such, they are often interrelated and not to be seen as completely separate. As a simple example, the terrain characteristics (geophysical) influence decision making as recreationists try to avoid dangerous terrain (human). In this thesis, a combination of human and geophysical variables will be researched, and as such, both theoretical themes will be reviewed here.

Another division can be made between recreational avalanche accidents and large-scale avalanche disasters. The former are relatively small and are often triggered by a recreationist (e.g. a skier), as well as having recreationists as casualties. The latter are very large and happen rarely, but when they do happen they threaten infrastructure and residential areas. An example of such a disaster is the Rigopiano disaster in Italy in 2017 which killed 29 people (Smith-Spark & Messia, 2017). Although large-scale disasters are tragic, the vast majority of avalanche deaths occur in the first type, recreational avalanches

(Tremper, 2008). Since the scales and implications of both types of avalanches are so different, they are also often researched separately. This thesis focuses strictly on recreational avalanche accidents. Therefore, the theory presented below also is mostly relevant for recreational avalanches, although many concepts overlap for both types of avalanches.

2.1.2. Avalanche education

Much of the theory provided in the following section is synthesised in so-called avalanche education programs. These are theoretical and practical courses in which recreationists learn about the danger of avalanches and how to cope with this danger (e.g. https://www.avalanche.ca/training/courses) . Different levels of avalanche education exist, ranging from introduction courses to professional education, for example mountain guide education. Besides learning to recognize the danger signs, individuals also learn rescue techniques and group tactics. Another source of avalanche education, besides courses, are many books about avalanche risk written for recreationists. Since this thesis is about behaviour of individuals in avalanche terrain, it is important to note that many recreationists do not randomly choose their routes, but often are able to use avalanche hazard factors as a variable influencing their behaviour. Also, knowing that avalanche education exists, helps illustrate the relevance of this thesis and many other avalanche research projects. Many of the findings of avalanche research eventually get distilled into avalanche education.

2.1.3. Geophysical hazard factors

The geophysical aspects of avalanches can be summarized as weather, snowpack, and terrain. Together they combine to form the so-called avalanche data triangle.

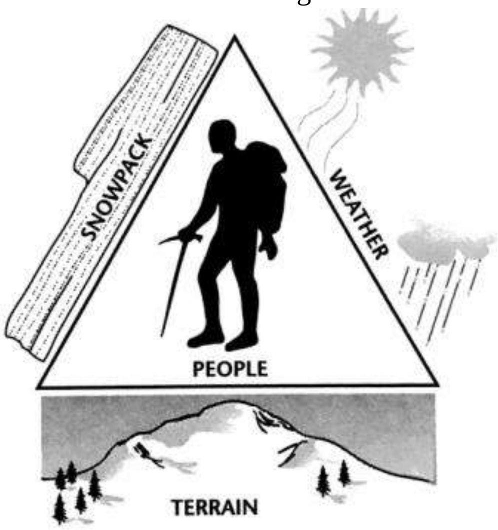


Figure 2. *Avalanche Data Triangle (Fredston & Fesler, 2011).*

In the middle of this are people, forming the human factor. The three factors combine together to form the expected avalanche hazard in a specific place at a specific point in time. Also, they are correlated to each other, e.g. certain snowpack qualities are often found in certain terrain, and so on. In avalanche safety courses, much attention is devoted to reading these three factors and knowing the risks related to them. It is important to note at this stage that although knowing the factors will greatly help decrease the risk of getting caught in an avalanche, they cannot exactly predict avalanches. Rather, they serve to predict where the probability of an avalanche with severe consequences occurring is higher. Individual avalanches never happen at a completely predictable moment and place.

Terrain

The type of terrain at a specific location is of great influence on the avalanche hazard. One of the main terrain features that influences the likelihood of avalanches occurring, is the slope angle. Logically, the steeper the slope, the more easily snow will slide off the mountain, making avalanches more likely. However, this mechanism only works to an extent. The avalanche risk increases with steeper slopes, but this effect decreases at about 45 degrees. Above 50 degrees, slab avalanches become increasingly rare, as these slopes are so steep that much of the snow rapidly slides off after falling, before being able to form a coherent slab. On those extremely steep slopes, loose snow avalanches are common, but those only cause a very minor part of avalanche victims (Tremper, 2008). In general, a slope angle of over 30 degrees is needed for a slab avalanche to form (Schweizer, Bartelt, & van Herwijnen, 2014). The sweet spot for avalanche accidents lies at around 39 degrees: here, the highest percentage of accidents occur (Harvey, 2002). One research shows that about half of human-triggered avalanche accidents occur at slopes between 37 and 42 degrees (Schweizer & Jamieson, 2000).

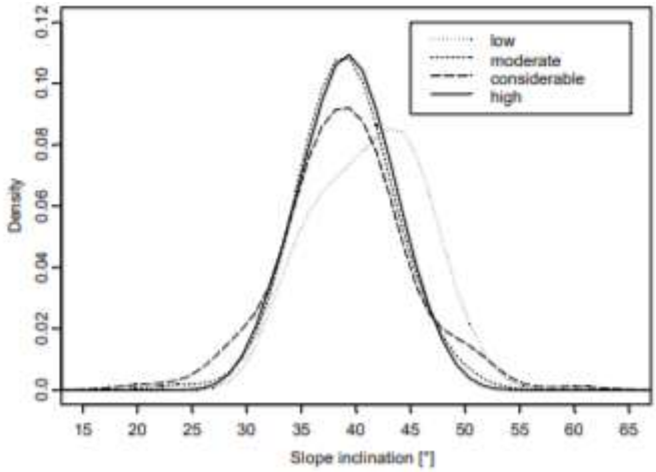


Figure 3. *Rate of avalanches as a function of slope steepness and reported avalanche danger rating as stated in the avalanche bulletin, between 1988 and 1999 (Harvey, 2002)*

Another important feature of terrain in relation to avalanche hazard is the slope aspect. Sun radiation limits the degree to which snow is able to form weak layers and fastens the bonding process of different snowpack layers (Tremper, 2008). Therefore, as north- and east facing slopes get the least sunshine, most avalanche accidents occur there. This is a general rule, there are some specific situations in which south- and west facing slopes are actually more dangerous. For example, in wet snow conditions, which mostly occur in spring, south- and west facing slopes produce more wet snow avalanches due to increased sun exposure (Tremper, 2008). Besides sun exposure, slope aspect is also important in relation to prevailing winds. In the weather section on the next page this will be elaborated on.

A final important part of terrain is the shape. Narrow ridges are dangerous because they often are subject to a high degree of wind loading, which means snow gets blown over the edge and forms a weak layer. Specific terrain shapes such as gullies are dangerous because snow gets blown into it from the side. Quick changes in steepness, e.g. a gully that is steep on the side but flat in the middle, form so-called terrain traps. Because of the specific shape, even a small slide can bury a victim very deeply, decreasing chances of survival (Tremper, 2008). Lastly, convex terrain shapes are more hazardous than concave shapes, as the snow cover has a higher degree of tension (Engel, 2000).

Weather

The second factor influencing avalanche risk, is weather. The most relevant weather variables for avalanches are precipitation, wind, air temperature, and sun radiation (Tremper, 2008). Precipitation can come in the form of snow or rain. Fresh snow is risky as it can add weight on top of a layer that is not bonded well to the layer underneath it. This can lead to a critical loading level, causing the layer underneath the freshly fallen snow to slide. Another possibility is snow falling on top of a layer with bad bonding capacities, a so-called weak layer. If this snow then forms a coherent layer (slab) and a person adds weight load to it by skiing or walking on it, it may be triggered and form a slab avalanche.

The intensity and amount of snowfall are very important to avalanche risk. Intense snowfall within a short time-frame is much more hazardous than gradual snowfall (Conlan, Tracz, & Jamieson, 2014). Rain can change the temperature within the snowpack, resulting in layers being less well bonded. Secondly, rain can increase the weight of the snow on top, causing it to slide (Tremper, 2008).

Wind is a factor in avalanche hazard as it transports loose snow. As most mountains have a prevailing wind direction, in which the wind is usually blowing, snow often gets transported from one side of the mountain to the other. This effect is mostly noticeable close to mountain peaks and ridges, where wind blows snow from windward slopes and drifts it onto lee slopes, forming cornices along the ridge and wind pillows beneath it (Tremper, 2008). Also, here convex shapes are more dangerous as snow is transported and concentrated more easily by wind there. Transported snow often forms weak layers in the snowpack, increasing avalanche hazard. Wind is mostly dangerous combined with specific terrain types, for example gullies, ridges, or mountain peaks, which allow for a high concentration of wind-blown snow. Also, wind combined with snowfall is extra risky, as the fresh snow is immediately blown towards terrain-specific spots.

Changes in temperature influence the snow cover, and thus the avalanche risk. Warmer snow is generally heavier than lighter snow, leading to higher avalanche risk in some specific cases. Numerous studies have found that rapidly increasing temperatures often precede avalanche events, as they decrease the homogeneity of the snowpack (Jamieson, Geldsetzer, & Stethem, 2001; Tracz, 2012). However, the relation between temperature and avalanche risk is complicated, and dependant on the specific situation. In some situations, increasing temperatures decrease avalanche risk, as they speed up the setting process of new snow. The effect of sun radiation on snow temperature is much larger than the effect of air temperature (Tremper, 2008). As could be read in the terrain section, sun radiation usually decreases avalanche risk.

Snowpack

As slab avalanches are a consequence of a homogeneous layer of snow breaking loose from the layers underneath it, the internal structure of the snow cover is a major factor. To form an avalanche, two things are needed within the snow pack: a slab, and a weak layer. Many different specific types of snow exist, all leading to their own avalanche problem. For the scope of this thesis, it is not feasible to go into them all, but a good overview is given in Tremper (2008) chapter 5.

Important to know is that changes in weather, especially snowfall, temperature, and wind, create layers of snow that combine to form hazardous situations. For example, a rain crust forms when rain falls on top of snow at low elevations, creating a slippery surface. When new snow falls on top of this layer, it does not bond well, and creates a slab laying on a weak layer, which is a prime avalanche risk situation.

Avalanche bulletin

Together, the risk factors are assessed every day during winter in alpine regions by the local avalanche authorities. They are elaborated on in the avalanche bulletin, which gets published online daily in most mountain regions during winter. The avalanche bulletin describes the particular hazard factors, for example windblown snow or weak layers within the snow pack. Also, a general warning level is given, which ranges from one (low) to five (extreme). The information in the bulletin can be used by backcountry recreationists to plan their route (relatively) safely. In general, the higher the danger rating in the bulletin, the more recreationists should try to avoid hazardous terrain (e.g. steep slopes). However, levels four and five are also relevant to officials, as they could mean more general safety measures need to be taken, for example closing roads or skiing resorts. During danger levels four and especially five, large-scale avalanche disasters become more likely (Tremper, 2008).

2.1.4. Human risk factors

Besides attention to geophysical factors, many researchers focus on the human factor (Zweifel & Haegeli, 2014). This is a logical step, as in the majority of incidents, avalanches are triggered by humans (Techel et al., 2016). One can assume that many people who travel in avalanche-prone areas are educated and know the risks, but still take them, which can be fatal. Many factors are at play in the human factor, including peer pressure, risk appetite, heuristics, and personal features. What many lines of research within the human factor have in common is that they analyse human behaviour in situations of avalanche risk (McClung, 2002). This information is useful as it can be used in education, for example by warning people who travel in avalanche-prone terrain for route-planning mistakes that others have made.

Numerous studies focus on the psychological aspects of risk-taking in avalanche terrain. For example Mannberg et al. (2018) use surveys combined with hypothetical route choices in avalanche terrain to find out the acceptability to risk among backcountry skiers. Sensation-seeking attitudes in other aspects in life turn out to also predict willingness to ski risky slopes. Also, many respondents agree to ski slopes that are steeper than they prefer, highlighting the risk of peer pressure and group dynamics. This does not fully compute with a qualitative study by Frühauf et al. (2017). Here, expert backcountry skiers are interviewed about their attitudes to risk. According to this paper, these skiers are not sensation seekers, and instead try to minimize risk at all times, while accepting that natural risks are part of their sport. The difference between the two studies may be due to the study population: one was a general group of backcountry skiers, while the other were only professional skiers who spend a lot of time in the mountains. Also, the methods used were different, one was quantitative and the other qualitative.

A study by Fitzgerald et al. (2016) assesses the differences in risk appetite among skiers in backcountry and so-called sidecountry terrain. Backcountry and sidecountry share the same general hazard in terms of avalanches, but the difference is that sidecountry can be accessed by ski lifts, whereas backcountry is completely remote. The main finding of their research is that skiers in terrain that is classified as sidecountry are willing to take more risks, and often have less avalanche training. Thus, sidecountry is perceived as more 'safe', while in reality the same hazards apply as in backcountry terrain.

Other studies focus on route behaviour. As avalanches are much more likely to occur on specific type of slopes (in terms of aspect, steepness, and shape), it is important to assess which slopes will be traversed in a route and avoid slopes that are too prone to avalanches. A question that is central to a number of research papers is whether backcountry recreationist take care to avoid these dangerous slopes. For example, Plank (2016) looks at routes shared by users on an online platform aimed at mountain recreationists such as tour skiers. The main finding is that many of the ski touring routes submitted contain hazardous slopes. According to the author this can lead to others copying this risky behaviour. However, the study only looks at general characteristics of tours (e.g. maximum slope, risk rating in the avalanche bulletin), and not on their specific routes. A research project called White Heat Tracks (http://www.montana.edu/snowscience/tracks.html) aims at gathering GPS tracks of chosen routes from backcountry skiers and snowboarders. Here, the routes are studied on a finer scale. In this project, behavioural economists and avalanche experts work together to create better understanding of the route choices of backcountry skiers. The GPS tracks are combined with data gathered from surveys. One paper focused specifically on Alaskan heli skiing guides shows that on the level of individual skiing tracks, they chose less risky tracks on days where the avalanche risk was higher (Hendrikx et al., 2015). However, on the entire data set of skiing tracks, there weren't significant differences between routes chosen on different days. However, this could be related to the low resolution of the terrain data used.

Another study focusing on the personal features of the skiers who delivered the GPS tracks, has some interesting findings. Mainly, people with more expert avalanche training and better skiing skills, in general choose more hazardous slopes (Hendrikx & Johnson, 2016). Also, people who are familiar with an area, because they have been there before, will take more risks. A third paper from this project asked respondents whether the avalanche risk rating (1 to 5) of the day made them alter their goal (e.g. a specific peak or slope) for the day. It turns out that the higher the risk rating of the day, the more likely the avalanche hazard influences whether the goal was reached (Hendrikx et al., 2013). Many of these findings completely logical and not very surprising, but the important thing is that quantitative analysis is made to prove them.

Whereas the White Heat Tracks project focuses on deliberate data gathering from recreationists, others look to scrape data from the web. An example of this is a paper by Techel, Zweifel, & Winkler (2015), which uses forums where mountain recreationists share trip reports in Switzerland, combined with locations of severe avalanche incidents. Being able to quantify the frequency of backcountry usage on different days (e.g. days with fine or poor weather, days with high or low avalanche danger ratings), enables a calculation of relative risk, as many avalanches are caused by the skiers themselves. An important finding of this method is that the risk of being involved in an avalanche accident increases with a factor five when the avalanche rating goes from 1 to 2, and a factor two when it increases from 2 to 3 (Techel & Zweifel, 2013). Schmudlach, Winkler, & Köhler (2018) use a similar approach to come up with a quantitative risk reduction method, using GPS routes submitted to a ski touring forum. They calculate the amount of avalanches relative to the proportion of travel in specific terrain and avalanche risk situations. As such, they have both the exposure and hazard of terrain and temporal avalanche factors .With this, they can compute the relative risk of specific terrain in specific avalanche situations.

Another interesting analysis is provided in (McCammon, 2002). Here, 715 accidents are reviewed to see whether errors in judgement were made by those involved. Many of the victims are drawn into so-called heuristic traps, where instead of using the information presented to them by the avalanche conditions, they use social cues to make their decisions. For example, mixed groups of men and women often take higher risks than all-male groups, presumably as men don't want to be seen as overtly careful by women.

These studies all have in common that they focus on human decisions related to avalanche hazard.

2.1.5. Avalanche terrain modelling

This thesis uses a fairly new dataset of quantitatively modelled avalanche terrain. Throughout the thesis, this dataset will be used to classify and analyse the planned routes. In order to understand this data and its novelty, it is required to introduce the topic of avalanche terrain modelling here. An important development in avalanche terrain modelling has been the Avalanche Terrain Exposure Scale (ATES) (Statham, McMahon, & Tomm, 2006). Here, a set of rules is used to classify avalanche terrain as either simple, complex, or challenging. Important inputs for this are slope angle, terrain shape, and the presence of terrain traps such as gullies or steep cliffs. The classification is static and insensitive to temporal avalanche risk factors. ATES mapping can be performed manually by experts (Statham et al. 2006) or in a GIS algorithm (Delparte, 2008). The advantage of this method for mapping avalanche terrain is that the three classes are easily understood and communicated, making it suitable for informing backcountry recreationists. The disadvantage is that detailed differences in avalanche terrain are not included, and that there is always manual input required to model the terrain. Bühler et al. (2013), among others, developed a method to automatically detect potential avalanche release areas from digital elevation models (DEM). The model is validated using historical Swiss avalanche data. The advantage is that it can be used for remote areas where little historical data is known, as the algorithm works universally. However, it only identifies potential release areas, which is not as useful for recreationists. Schmudlach & Kohler (2016) developed an algorithm to automatically derive ATES classes from a DEM. This algorithm takes into account the slope shape at each cell, as well as that of surrounding cells. The outcome is a map of continuous ATES values ranging from simple (0) to challenging (100). The approach from Harvey et al (2018) continues on this idea. An important novelty of their method is that they use the numerical RAMMS avalanche dynamics model (Christen, Kowalski, & Bartelt, 2010). This model was designed for predicting large-scale avalanches, but can also be used for medium-sized avalanches (Dreier et al. 2014), which are most relevant for recreationists. Using this model, it was possible to calculate possible burial depth for different avalanche release areas. Also, potential fall zones were identified. This together with triggering risk of avalanches, leads to a continuous value depicting the avalanche hazard. This value includes the potential of triggering an avalanche, combined with the severity of consequences when an avalanche is triggered. A second novelty of the approach from Harvey et al (2018), was that they differentiate between triggering areas and remote triggering areas. The difference between the two is that in the former, there is a potential avalanche start point at that location. In the latter, an avalanche is not likely to be triggered at that location, but it is possible to remotely trigger an avalanche on a nearby slope by loading that location. This is possible because snow slabs are often large and connected across space. Both maps can be used for route planning in relation to avalanche terrain. The maps are designed for optimal use with avalanche danger ratings 2 and 3. The model used to create the two maps is shown in figure 4.

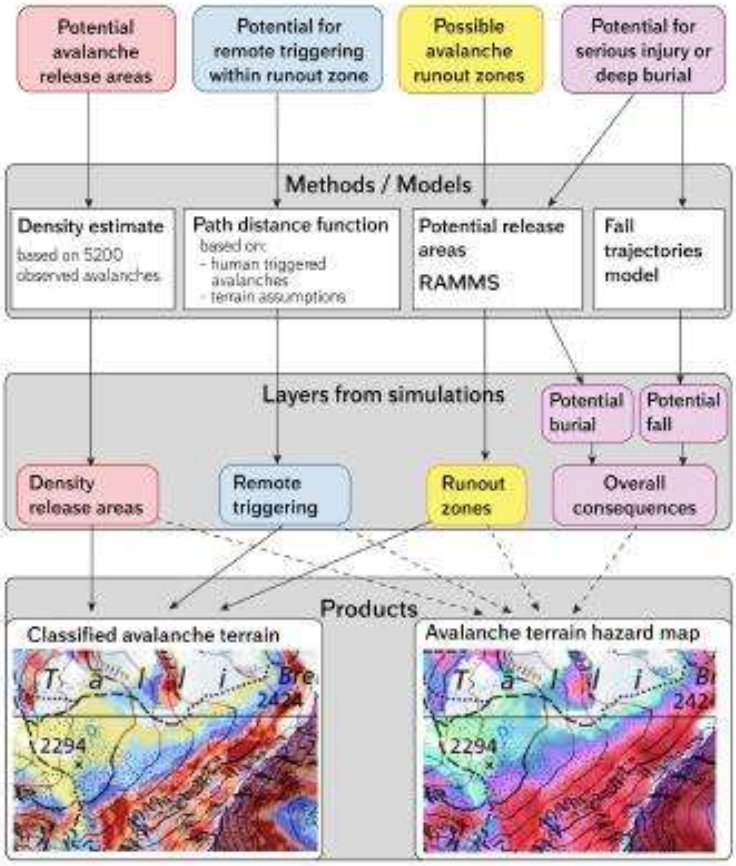


Figure 4. *Method used to compute 2 avalanche terrain maps by Harvet et al (2018), p.*

2.2. Volunteered Geographic Information

2.2.1. Introduction

Volunteered Geographic Information or VGI is the collection, assembling, and dissemination of geographic information by private citizens. The first well-known use of the term is in an article by Goodchild (2007). The main difference between VGI and regular geographic information is that VGI is gathered by private citizens, as opposed to national mapping agencies or other official mapping institutions. VGI is related to the concept of crowdsourcing, with the difference being that in VGI there could be an individual gathering the information, whereas in crowdsourcing there's always a mass of people gathering information (Goodchild, 2007). VGI is a type of user generated content (UGC). UGC is any content on the web that is made by customers or end-users of a website and is publicly available to other end-users. This is a central feature of the so-called web 2.0. The web 2.0 has emerged in the 2000s, and differs from the internet before that in the way that users uploading their own data and content play a central role (Antoniou, 2011). Prior to the web 2.0, the internet was used primarily as a dissemination, from producer to user. Today, users are at the same time also producers of content, and thus a new user type has emerged, the produser (Budhathoki, Bruce, & Nedovic-Budic, 2008).

2.2.2. The Volunteered Geographic Information concept

Although any geographic data that was sourced from private persons can be called VGI, there are often large differences between different VGI datasets. Obviously, VGI data can be about a broad range of topics. Besides this, there are also more fundamental, conceptual differences between datasets. These mostly relate to the type of data, the data collection process, and the purpose of the data. Those differences have implications for the possible uses of the data.

Within the data collection process a distinction can be made between active and passive collection. In active collection, volunteers deliberately contribute data for a specific project. An example of this is OpenStreetMap (OSM), in which volunteers work together to map areas of the world. In passive collection, volunteers collect data for any other purpose, for example to track their athletic progress or share their activities. The data can then be used by people with another purpose, for example academic research or mapping. Since the data was not collected for this purpose, it is in fact re-used. This difference has implications for the quality of data and for privacy. The quality is affected since in a purposeful data-collection it is easier to assign quality thresholds a priori, whereas in re-used data there are no such thresholds, since individuals don't know that their data will be re-used. The privacy is affected since in purposeful data-collection, individuals deliberately choose to collect their data, whereas in passive collection their data is often re-used without their knowing. This difference in collection method marks the difference between volunteered, and contributed geographic information (CGI) (Harvey, 2013).

Related to this is the type of data that is collected. Since active VGI is collected with a specific goal in mind, it is often possible to collect detailed and specific information. In the example of OSM, semantic information is collected about shops and services in an area, as opposed to just geographic locations. In the example of the Whiteheat avalanche research project, personal information such as age, gender, and education level is collected in combination with GPS tracks. This personal information is part of the dataset. As such, many active VGI data have semantic value in their own right. Passively collected information, on the other hand, often needs added context before it becomes valuable. An example of this are activity trajectories from fitness apps such as Strava. These trajectories are used in various research domains, such as urban planning. However, a single trajectory is not of much semantic value. Instead, the patterns recorded from a large set of Strava trajectories need to be combined with other data such as land use, density, or road types, in order to be useful as research data (Griffin & Jiao, 2015). In reality, this difference is not absolute, as almost all applications of VGI data fall somewhere in between. For example in the Whiteheat project, the GPS tracks with the personal data could be analysed by themselves, but they only become interesting when they are overlaid with terrain data such as slope and aspect.

Another distinction is made in the type of data by Stefanidis, Crooks, and Radzikowski (2013). They differentiate between data that is explicitly spatial, such as OSM maps, and data that is not spatial per se but has a spatial footprint. A prime example of this second type are geo-tagged tweets. Tweets themselves are not spatial in nature, they are a piece of text which is not necessarily related to location. However, when many geo-tagged tweets are collected and analysed as a set, they can provide useful spatial information. For example, they can convey information about spatiotemporal trends and preferences. A major difference between this so-called ambient geospatial information and "typical" VGI is that in the latter, the volunteers act as sensors, recording the world around them. In the former, however, the world around is not the main subject of research, and instead the "volunteers" are the observable phenomena (See et al., 2016). In a broad way this is similar to Harveys notion of contributed geographic information (Harvey, 2013), and indeed in some instances both concepts overlap. Yet, the definitions of both are different, as both authors focus on a different component of VGI. Whereas Harvey focuses on the collection process, Stefanidis et al. (2013) focus on the content of the data. As such, geospatial data that is actively collected by volunteers but at the same time has the volunteers as main research subject, is not CGI but is ambient geospatial information. An example of this is the white heat project, where volunteers agree to collect data, making it VGI and not CGI. At the same time, the volunteers are the ones being researched, not the world they sense per se, making it ambient geospatial information. For the rest of this chapter, VGI is usually used to avoid confusion, as most articles about VGI-related concepts use that term.

2.2.3. Volunteered Geographic Information Challenges

VGI can offer interesting, novel forms of data, which official mapping agencies cannot provide. For example, VGI offers insight into the behaviour of people submitting the data. It is therefore better suited than traditional sources of geographical information to map citizen behaviour in time and space. Also, VGI allows for a participation of ordinary citizens, which enables collaboration between them to tackle common problems, democratizing the process of mapping the world (Antoniou, 2011). However, there are several challenges related to the collection of VGI. These will be addressed in the next section.

Quality

One of those challenges relates to the quality of the data. As many contributors of VGI do not have a background in GIS or cartography, their contributions may be inaccurate. Also, as the data in VGI comes from many different contributors, it is hard to hold the contributor accountable for any errors (Turner, 2006). Expert editing of the collected data can be a measure to counteract this (Antoniou, 2011).

Quality is not a straightforward concept, as it has evolved over time. Whereas in the beginning period of geographical information science, quality was merely judged on whether the positions on a map were accurate with the real world. Later, this evolved to a wider concept, including whether data is fit for its intended use (Van Oort, 2006). The responsibility of judging whether data is fit for its intended purpose has shifted from the data provider to the user (Antoniou, 2011).
A good overview of spatial data quality measures is given in (Antoniou, 2011). These are as follows.

- Completeness: whether or not there is data missing that should have be included as it falls within the scope of the product specifications or user requirements. Conversely, also whether data is included that falls out of this scope.
- Logical consistency: whether the volunteered data adheres to basic rules of logic. For example, when a GPS track is uploaded where somebody is skiing at 1000 km/h, this does not follow logical consistency.
- Positional accuracy: how accurate the positions in the recorded dataset are with the real world.
- Temporal accuracy: how accurate the temporal features of the data are.

- Thematic accuracy: all other accuracy measures not included in the previous two. For example, whether an attribute stored in the spatial data is accurate (e.g. temperature in a weather map).
- Purpose and usage: whether the recorded data fits the intended purpose. This can be viewed from the perspective of the data contributor or from the perspective of the data collector.
- Lineage: the recording history of the data. For example: the data and time of recording, who recorded it, which software was used.

Testing a dataset against these measures is a way to ensure its quality is sufficient. However, whereas some of the measures are objective (i.e. positional accuracy), others depend on the person reviewing the data. Whether data is fit for the purpose depends entirely on the purpose. Also, how much importance is given to specific aspects of quality depends on the purpose. Since the data used in this thesis has already been processed in a prior thesis project, some of the quality concerns will have been taken away or lessened. Some challenges still persist. Most notably, whether the data mapped fits the intended purpose remains an interesting question, as the data originally was mapped for personal use and not for research purposes. The positional accuracy is another challenge that remains. Although the trajectories can be thought of as positionally accurate in the strict sense, they may not be accurate in the sense that people actually travelled them. These and other challenges will be further addressed in chapter 4.

Privacy

Another major challenge in VGI is ensuring personal privacy in collected data. To determine how to ensure privacy we first need to define what exactly privacy is. A famous definition (and the one followed here) of privacy by Alan Westin is that privacy is the right of an individual or group to decide for themselves which information about them is shared with others (Westin, 1967). A question that rises from this definition, is: what exactly can be defined as "information about a person"?

An important document on privacy related to data processing is the general data protection regulation of the EU (European Parliament and Council, 1995). In this document, the following definition is given of personal data: personal data is any information relating to an identified or identifiable person. What is important here, is that the information has to be related to an individual. For example, somebodies occupation alone is not personal information, as many more other people will have this occupation. A unique identifier such as a phone number, on the other hand, is personal data. Thus, the more specific a piece of information is about a person, the more likely it is to be considered personal data. Also, somebodies occupation in combination with their full name is considered personal data, as they can be used to uniquely identify a person. Multiple pieces of information that can be combined to form personal data, are to be considered personal data (van Loenen, Kulk, & Ploeger, 2016). In this sense, geospatial data takes a special role. Many geospatial pieces of data are anonymous or easy to anonymize. For example in the case of backcountry skiers' GPS tracks, it is easy to remove the names and other identifiers in the processing phase. However, when combining geospatial data points with each other, or combining geospatial data with other types of data, it is often possible to identify a person. An example of this is the app Strava, which is used by cyclists and runners to track their sports activities. Anonymous data published on the Strava global heat map, which contains all public Strava tracks, can be combined to find a person's address or name (Moody, 2018; Brewster, 2018).

Although privacy is not the biggest subject among researchers that make use of VGI (Granell & Ostermann, 2016), some articles have been written about it. The following general guideline for academic use of VGI is given in (Mooney et al., 2017):

"During the dissemination of research outputs, care must be taken not to expose the identities of or other private information related to the citizens who contributed to the VGI project. Patterns and inferences made about the contributors of the data must be carefully considered so as not to breach the privacy of those citizens." (Mooney et al, 2017, p. 124)

The first part of this quote is not hard, simply leaving out names or personal data will suffice for this thesis research. However, the second part requires some consideration. When using crowd-sourced GPS data, it is key to make sure the GPS tracks cannot be used to find out personal facts about the persons who contribute them. Oksanen, Bergman, Sainio, and Westerholm (2015) deal with this problem. Their solution is quite simple: they count the number of tracks at a given location and filter out tracks that are too unique, as they are most easily used to track personal details (e.g. addresses).

However, their article uses bike tracks, which are more inherently problematic than ski tracks as people often bike from their home, making it possible to track a person's address. Also, people often bike in a city and use their bike to commute to places of interest such as their work place, which can also be seen as personal information. Ski tracks often start somewhere at a parking lot or on a mountain top, leaving out this problem. Therefore, the expectation is that simply anonymizing ski tracks will suffice to make sure they do not qualify as personal data. Still, it is important to keep in mind that the data, when combined with other public datasets, does not become personal data. This, however, is only an issue when this combination is relatively easy. For example, when it takes days of computer processing to turn a dataset into personal data, it is not to be considered an issue (van Loenen et al., 2016).

Inequality

A third set of issues in VGI is related to inequality. in this context, inequality refers to the fact that VGI is often collected by a small portion of the general population, which can make it problematic in terms of representativeness. Different types of inequality exist within VGI participation due to the different reasons for inequality. The first inequality in VGI is a consequence of the so-called digital divide (Ferster, Nelson, Robertson, & Feick, 2018). This divide separates people without internet and computer access from those with it. In many cases, the digital divide follows other divisions based on privilege, for example social class, gender, ethnicity, and other social and economic factors (Foster & Dunham, 2015). As a consequence, some groups are better able to participate in VGI, and as a result are also better represented by VGI (Foster & Dunham, 2015).

Aside from inequality stemming from the digital divide, there is usually also a general participation inequality, even within user populations that all have access to internet. The reason for this is that in many cases, a very small group of contributors is responsible for a very large proportion of the work being done. Both in physical and digital volunteering projects, this is the case (Haklay, 2016). Participation inequality follows the 90-9-1 rule: 90 percent of users don't contribute, and instead only consume information. 9 percent contribute occasionally, and 1 percent contributes by far the most content. Of course, in reality these percentages vary, but the principle is the same. An extreme case such as Wikipedia sees only 0.003% of users contribute two-thirds of the content (Haklay, 2016).

The problem that arises from this inequality is that there is also an inequality in the type of contributions made by volunteers. In other words, topics that are of interest to the 1 percent of high-level contributors, are also more reflected in the contributions. For example, OSM contributors are predominantly male, young, and educated. As a result, points of interest related to activities that are traditionally seen as feminine, such as child care, are underrepresented on OSM (Stephens, 2013). Also, geographic inequality may be a result of participation inequality. This happens when the main contributors are mostly contributing about their surroundings, leading to an overrepresentation of that area within the data (Haklay, 2016). When using VGI as a main data source, these consequences of participation inequality are important as they may lead to a biased dataset.

Most studies that use VGI as a data source, recognize the fact that participation inequality exists (e.g. Foster & Dunham, 2015; Ogie, Clarke, Forehead, & Perez, 2018). However, what is less obvious is how to deal with it. In general, three approaches can be distilled. Firstly, there are techniques to encourage more users to contribute, for example by making contributing easier or by rewarding quality contributions (Nielsen 2006). As I will be working with data that has already been contributed, this is not possible for my research. The second approach is to assess the participation inequality in a dataset, and explicitly adjust any conclusions to this. The third approach means editing the dataset so that the effects of participation inequality are decreased. An example of this is given in Oksanen et al. (2015). This article goes into the possibilities of creating a heatmap of cycling GPS tracks where participation inequality has a decreased role. Instead of counting how many times a GPS track is created along a certain line, they look at the number of unique users creating a track along that line. Another option they use is to compute a density map that uses the diversity of users in an area as a secondary input, thus combining the number of different users with the use intensity of a line. However, this article focuses on cycling tracks, which are inherently different from backcountry skiing tracks as they follow roads and are thus set in specific locations. Also, cycling is often used as a means of commuting, which means that a single user making one track multiple times is more likely. Another approach to deal with participation inequality is given in Techel et al. (2015): they divide the group of data contributors based on the proportion they contributed, and analyse the differences between the groups.

2.2.4. Route Volunteered Geographic Information

As this thesis is about using GPS tracks and planned routes as passive VGI, some similar studies are now outlined. This serves to guide the spatial analysis of routes in this thesis. Many articles focus on GPS tracks of cycling routes, this is interesting in the realms of urban planning and health (e.g. Griffin & Jiao, 2015; Jestico, Nelson, & Winters, 2016; Menghini, Carrasco, Schüssler, & Axhausen, 2010). A major difference between cyclists' GPS tracks and backcountry skiers' tracks, is that the former are generally tied to a road network, whereas the latter aren't. This influences the type of analyses that are possible. For example, Griffin and Jiao (2015) analyse which types of roads (e.g. bike-specific roads or streets with broad shoulders) are most popular based on Strava activity. Also, many cycling studies focus on urban features, such as building densities and land use diversity (Griffin & Jiao, 2015), which is also irrelevant skier tracks. However, some areas of analysis can be applied to both skier and cyclist tracks. For example, many studies assess the role of slope gradient on cycling routes. Also, although ski tours do not follow official roads, they often do follow existing routes through the terrain, for example routes published by the Schweizer Alpen-Club. What most studies into VGI trajectories have in common, is that they focus on the route behaviour of humans in relation to their physical environment. As such, they work from an assumption of a sufficiently modelled or realistic physical environment, and analyse the trajectories in relation to this environment. This is also the goal of this thesis.

2.3. Route and movement analysis

2.3.1. Introduction

Since this thesis deals with the spatial analysis of recorded movement paths and planned route trajectories, it is important to place it within the existing body of work in this field. The analysis of planned routes differs from the analysis of recorded routes in two major aspects. First, in the former, one cannot be sure that the routes were actually travelled, in other words that the person who planned the route went through with it. In recorded GPS tracks of routes, one can be sure that the route was travelled, since it cannot be recorded otherwise. Secondly, in the former, there is no time recorded, whereas in the latter there is. Most existing studies focus on spatiotemporal analysis of recorded routes with time stamps, and not on planned routes. Therefore, these will be outlined here. It is important to keep in mind that any techniques that require time stamps in routes cannot be performed here, or at least not to their full extent.

2.3.2. Basics of movement analysis and spatiotemporal analysis

One of the earliest systems used to analyse people's behaviour within a spatiotemporal context was proposed by Hägerstrand (1970). His paper highlights the movement of people in time and space when faced with different types of constraints. Although these constraints are mostly related to daily movement patterns within cities, the idea that an individuals' movement must be analysed within constraints is relevant for this research.

Hägerstrand's work can be seen as the beginning of the time-geography perspective. In time-geography, human behaviour (e.g. movement) is analysed with time and physical space as basic dimensions (Lenntorp, 1999). This behaviour can be captured in spatiotemporal data. This is data with both spatial and temporal attributes. Spatial objects have a location variable. Temporal objects have a variable time and are not existent indefinitely (Andrienko, Andrienko, Bak, Keim, & Wrobel, 2013). This spatiotemporal data broadly allows three categories of queries, based on three aspects of the data, namely objects ("what?"), space ("where?"), and time ("when?"). One can also combine these queries, for example asking: "where and when was this object moving?" (Peuquet, 1994). This is the most basic definition of spatiotemporal analysis. Movement analysis is a form of spatiotemporal analysis.

Central within movement analysis are so-called movement observations, which are spatiotemporal data points (Dodge, Weibel, Ahearn, Buchin, & Miller, 2016). The central questions that movement analysis tries to answer are: why, how, and by which forces does an object move? The movement ecology paradigm by (Nathan, Getz, Revilla, & Holyoak, 2009) distinguishes between the intrinsic motivations of an object, its internal capacity to move and navigate, and the surrounding context (e.g. external constraints), which together influence the movement parameters. Dodge, Weibel, and Lautenschütz (2008) elaborate on this, noting four categories of factors that influence movement:

1. intrinsic properties of agents
2. spatial constraints
3. environment
4. the influence of other agents.

Of particular interest here is the distinction between spatial constraints and environment. Whereas spatial constraints are 'hard' rules on where an object can and can't move (e.g. a steep cliff or a body of water), different environments can have varying degrees of attractiveness for objects to move into. For example, depending on skiing level, a person will find progressively steeper terrain increasingly attractive to ski, until a certain steepness where it gets too scary and the attractiveness starts going down again. Since the data in this thesis does not contain intrinsic properties of the agents (beyond some basic metrics), the focus is on the latter three categories of factors. Most important here are spatial constraints and environment. The direct influence of other agents can't be tracked in the dataset, as it is impossible to see which agents where at a route at the same time. The routes can be compared to each other, which forms an interesting line of analysis, but this is not part of the direct influence of other agents.

2.3.3. Computational movement analysis and trajectory analysis: techniques

Several techniques exist to analyse the trajectory of a moving object. These will be outlined below. Laube (2014) has written a comprehensive book on recent issues and techniques in computational movement analysis. Computational Movement Analysis (CMA) is *"the interdisciplinary research field studying the development and application of computational techniques for capturing, processing, managing, structuring, and analysing data describing movement phenomena, both in geographic and abstract spaces, aiming for a better understanding of the processes governing that movement"* (Laube, 2014, p. 4-5). Although the dataset of planned routes is not technically movement data (since there is no timestamp), many of the techniques outlined in his book do apply. Another point to stress is that many articles cited here do not explicitly mention the concept of CMA, the methods or concept described do fit within the definition above.

General considerations

Movement data is usually recorded as a collection of spatial points with time stored as an attribute. However, when these points are recorded at a high enough granularity, they allow an almost continuous trajectory to be constructed from them (Long & Nelson, 2013). Laube (2014) states the importance of deciding on a conceptual model in which movement data is to be analysed. This and other decisions influence the outcome of analysis, and are important to fit to the type of data and type of analysis. He distinguishes six conceptual model spaces for trajectories, depicted in figure 5:

1. A space in which objects move completely freely between positions without any restrictions is called homogeneous (unconstrained) Euclidean space (a)
2. A space in which the possible movement space is constrained, for example by terrain features such as water or steep slopes. This is constrained Euclidean space (b)
3. Some visualisation tasks ask for a space-time cube as per Hägerstrand (1970) (c)
4. Movement can also be captured between discrete steps in space, for example a grid. This is called heterogenous field space (d)
5. A specific conceptual type is based on cell phone reception areas, which create an irregular grid of visited tiles, named irregular tessellation (e)
6. The final type is based on a transport network of nodes and edges: network space (f)

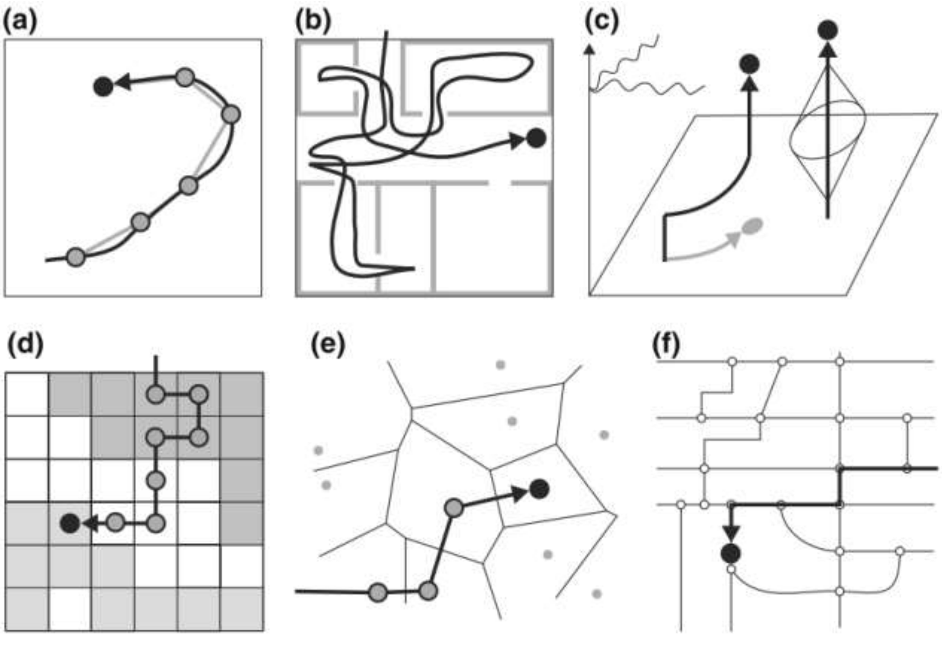


Figure 5. *(previous page) Conceptual movement spaces: (a) Euclidean homogeneous space, (b) constrained Euclidean space, (c) space-time aquarium, (d) heterogeneous field space, (e) irregular tessellation, (f) network space Source: Laube, 2014, p.12*

These six conceptual spaces are based on three dimensions. Firstly there is a distinction to be made between a Lagrangian and Eulerian movement perspective. In the Lagrangian perspective, positions are captured at a time interval, and a line is drawn between those positions to annotate the movement trajectory. In the Eulerian perspective, a position is recorded every time the object passes a checkpoint, for example a toll booth at a road. As the data in this thesis is GPS data or unconstrained routes, they are to be considered Lagrangian trajectories. However, they could also be considered in an Eulerian perspective, for example when counting how many routes pass a specific point of interest (e.g. a mountain peak).

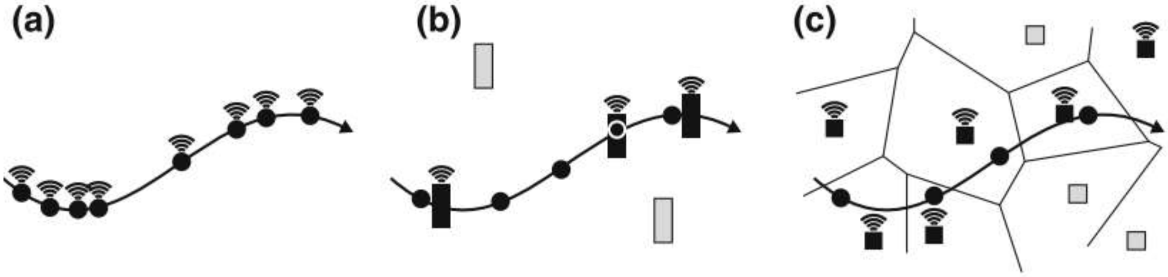


Figure 6. *Two perspectives of movement: Lagrangian, for example GPS tracking (a) versus Eulerian, for example traffic check points (b) or cell phone reception areas (c). Source: Laube, 2014, p.13*

A second dimension to consider are constrained or unconstrained movement spaces. In this context, constrains determine where objects can or cannot move. Although it may seem tempting to model movement in an unconstrained space, this is often not realistic. Most objects experience some sort of constraint, as originally mentioned by Hägerstrand (1970). According to Laube (2014), human movement is typically constrained as well, for example by road or path networks. Ski-touring and freeriding are quite special in that respect, as they do not follow paths. However, a constraint is that there has to be snow to ski on, and a slope to ski off. Also, skiing routes generally cannot access extremely steep slopes. Furthermore, most skinning too steeply uphill is experienced as uncomfortable by many tour skiers. As such, many ski touring routes follow the same terrain features. Choosing whether to model movement in a constrained space is important, as it changes the options. In the example of human movement, modelling in a constrained space allows for map-matching with a road network. This is generally not applicable to the free route choices in ski touring. However, an interesting study is provided in Taczanowska et al. (2017), who model ski touring GPS tracks in a node- and edge network based on a trails system. This network is then enhanced based on the intensity of use, focusing on the most important nodes. This shows that even if ski touring does not officially follow roads, methods taken from road and network modelling could apply to ski touring as well.

The third dimension Laube mentions in his book is whether data is recorded continuously (top row in figure 5) or discretely (bottom row in figure 5). As mentioned before, GPS data with a fine granularity could be seen as continuous when a smooth line is depicted of it. However, Laube (2014) mentions that GPS data can also deliberately be treated as discrete if the research goal is helped by this. An example is an urban network, where the entire segment in which a person is on a specific edge (e.g. a tram line) is treated as one discrete step. A similar thing is done in the aforementioned study by Taczanowska et al. (2017).

Movement parameters

To analyse movement, one needs to address its characteristics, or movement parameters. Movement characteristics or movement parameters both mean the same thing: they are variables that describe the movement. Although there is some variance, most studies use the same basic set of parameters (Laube, 2014). Andrienko, Andrienko, Pelekis, and Spaccapietra (2008) distinguish between moment-related and overall characteristics. Whereas moment-related characteristics are at a specific moment or timestep (e.g. the speed at a specific place) , overall characteristics are about the whole trajectory (e.g. the overall average speed). A different distinction is made by Dodge et al. (2008). They differentiate between three groups of parameters: primitive parameters, primary derivatives and secondary derivatives. A second dimension of distinction is between spatial, temporal and spatiotemporal parameters. The primitive parameters are the coordinate positions at time instances and intervals. The primary derivatives are derived from this, for example the distance or the direction. The secondary derivatives are derived from the primary derivatives, for example the acceleration, which is derived from the speed by calculating the change in speed between two moments or two positions. Also, the derivatives are computed by combining different parameters. For example, to compute direction, one needs multiple sets of coordinates as well as the times at which these coordinates were travelled. The classification from Dodge et al. (2008) is depicted in figure 7.

Parameters/Dimension	*Primitive*	*Primary derivatives*	*Secondary derivatives*
Spatial	Position *(x,y)*	Distance *f(posn)* Direction *f(posn)* Spatial extent *f(posn)*	Spatial distribution *f(distance)* Change of direction *f(direction)* Sinuosity *f(distance)*
Temporal	Instance *(t)* Interval *(t)*	Duration *f(t)* Travel time *f(t)*	Temporal distribution Change of duration *f(duration)*
Spatio-temporal *(x, y,t)*	–	Speed *f(x,y,t)* Velocity *f(x,y,t)*	Acceleration *f(speed)* Approaching rate

Figure 7. *A classification of movement parameters. Source: Dodge et al., 2008, p. 243*

Although many of the derivatives are also stored directly in advanced GPS sensors, from a transparency and control perspective it is advised to (re)compute them as well (Laube, 2014).

Movement patterns

Movement parameters offer a useful framework to investigate an individual movement trajectory. However, even more interesting is to analyse a set of trajectories and their interrelations. To do this, one has to look at the movement patterns. A useful overview of possible movement patterns is given in Dodge et al (2008), although explaining all of them would be beyond the scope of this chapter. What's important, is that different movement patterns can be classified based on the variables in figure 7. Pattern analysis is often applied to data where different objects move in congruence with each other, for example a flock of birds or football players on a field. However, it can also be interesting to find out about places where trajectories concentrate at different moments in time. Since the former is impossible with the dataset used here, the focus will be on the latter.

2.3.4. Challenges in movement analysis

Several challenges can be identified in computational movement analysis. From those challenges, future research directions are formulated in articles that deal with CMA.

Context

Several authors stress the need to include context in movement analysis (e.g. Buchin, Dodge, & Speckmann, 2014; Dodge, 2016; Purves, Laube, Buchin, & Speckmann, 2014; Laube, 2014). Context can be defined as "the locational circumstances of a moving agent" (Buchin et al., 2014, p.2). Dodge et al. (2016) argue that within movement analysis many studies focus on the characteristics of trajectories in isolation, not paying attention to context. Context can be improved by recording additional data with the trajectories, such as demographics. Another way to add context is to add extra geographic information to the study, such as a road network within which trajectories can be placed. This is an example of the importance of constraints that limit movement possibilities as per Hägerstrand (1970). There are quite some examples of recent trajectory studies that include context in their analysis. Buchin et al. (2014) study hurricane and albatross flight trajectories and model their similarity taking into account the context. They differentiate between contexts that enable and that limit movement, and between partially or fully enabling or limiting. For example, a road fully limits a car's movement, whereas a tractor is partially limited as it can go off-road but will be slower. Their article offers an interesting technique for quantitatively taking into account context in similarity analysis. Basically, they increase the distance between two trajectories when their contexts differ, for example one on land and one is one water. This corrects trajectories that are close in absolute sense but far away contextually, and conversely trajectories that are further away in distance but contextually closer. This is something that also needs to be thought of when analysing skier trajectories: two tracks may be very close in space, but contextually far apart, because a steep ridge separates them, creating a different spatial context.

Siła-Nowicka et al. (2016) use a combination of GPS tracks of human movement with contextual information such as points of interest to analyse urban movement patterns. Conversely, they also draw conclusions about the context based on the movement patterns. For example, they evaluate which places are frequently visited by humans in their daily life. In this way, human movement and geographic place are studied together, instead of separately, improving the semantic value. An example of a contextual point of interest in ski touring is a mountain top or pass, which is often used as a target of a ski tour.

Including context in the analysis in this thesis is of particular interest as it could make up for the lack of temporal information in the planned routes.

Uncertainty

Although GPS data is often treated as reliable, there are many possible inaccuracies in the capture. Three categories of uncertainty surround movement data:

- Uncertain specifications: what exactly is meant by a description of a trajectory. For example, exactly which trajectories do we include when we report on "daily trips?"
- Uncertain measurements: this can be due to inaccuracy in the measuring tool, but also because GPS is recorded at discrete time steps, and what happens in between is uncertain.
- Uncertain transformations: stems from the transformation of raw measurements into processed information.

(Laube, 2014).

How important measurement inaccuracies are, depends on the goal of study. When the goal is aggregating a large number of routes, inaccuracies aren't that important. However, when point-in-polygon analysis has to be applied, inaccuracies can be important, as they may lead to incorrect classifications (Laube, 2014).

3. Methodology and data

In the following section the methods used in this thesis are outlined. Prior to this, the study area and data used are described.

3.1. Study area

This thesis is limited to the Swiss alps. The reason for this is that this is the only area for which all used datasets have coverage. The original datasets have coverage for the entirety in Switzerland. However, the planned ski tours were already processed in the thesis of Schönenberger (2018), and he decided to focus on the Swiss alps and exclude the Jura mountains and the flat areas of Switzerland which have no avalanche bulletin, so that he could have a contiguous study area. This are also contains the majority of planned routes. The area of study is presented in the figure below.

Figure 8: *Study area of this thesis. After Schönenberger (2018), p. 36*

A small test area was chosen to develop the methods needed and to provide examples of some of the results. This area comprises the routes leading up to Pigne d'Arolla (3796 m.a.s.l). The reason this peak was chosen is that it has multiple possible routes leading to the top with different terrain attributes, and that there were many GPS tracks on Wikiloc leading to this peak. Within this area, there are 173 planned routes and 31 GPS tracks.

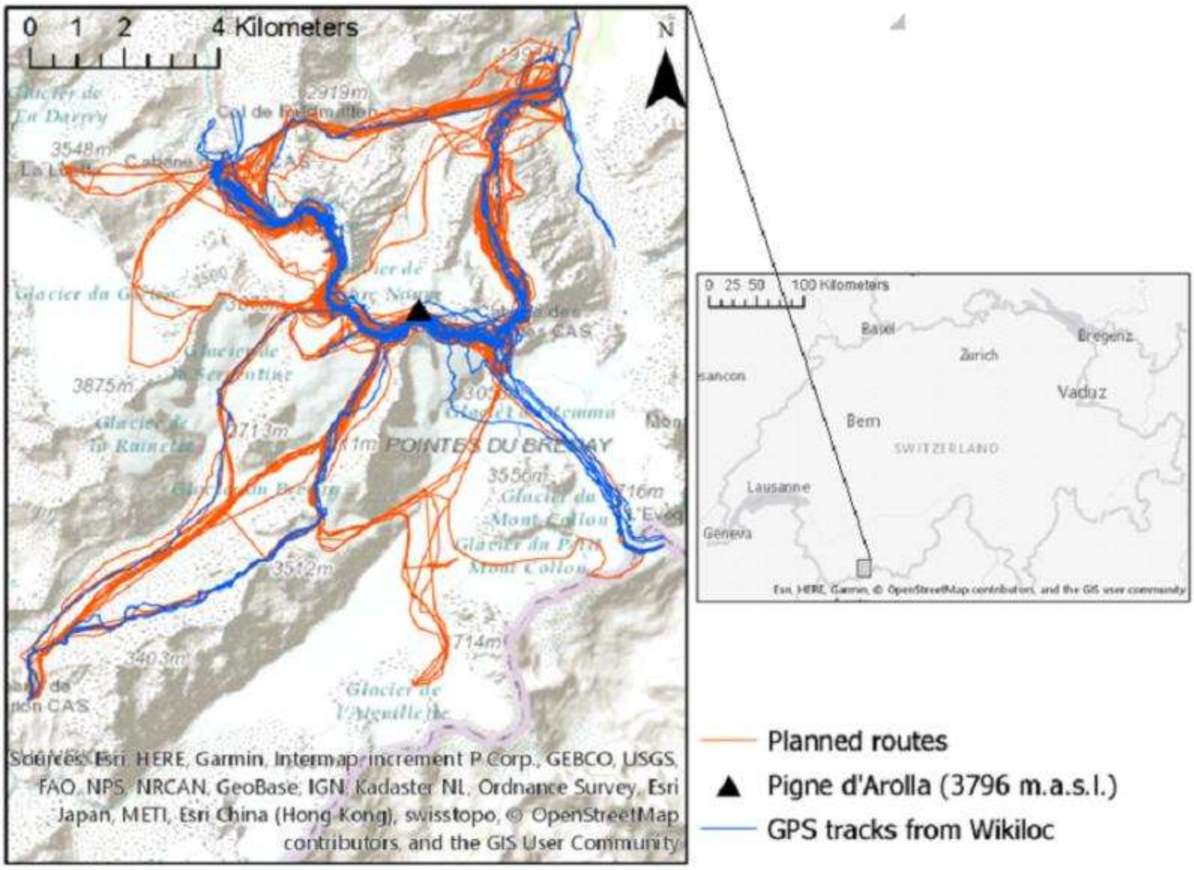


Figure 9. *Smaller study area around Pigne d'Arolla*

3.2. Data

Several datasets are used in this thesis. They are described in the following section.

3.2.1. Planned ski tours

The main dataset that is subject to analysis in this thesis are the planned ski tours from Whiterisk.ch. In total, they are 53553 lines in a vector dataset in .shp format. As noted before, they have already been used in a prior MSc thesis last year (Schönenberger, 2018). A large part of the workload in that thesis was processing the routes and filtering out routes that were not suitable. In that thesis, routes were filtered out because they were unrealistic, too short, because they were not ski touring routes, or because they were not within the study area. Because of this, the dataset is already quite reliable, although some tours will still contain errors. Also, some key terrain characteristics of the planned tours were already included, such as the average and maximum altitude and slope of each slope. In total, 26 of such characteristics are included. Table 1 shows the attributes stored in the processed routes.

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	

Table 1. *Attributes in Whiterisk routes. Source: Schönenberger, 2018, p. 42*

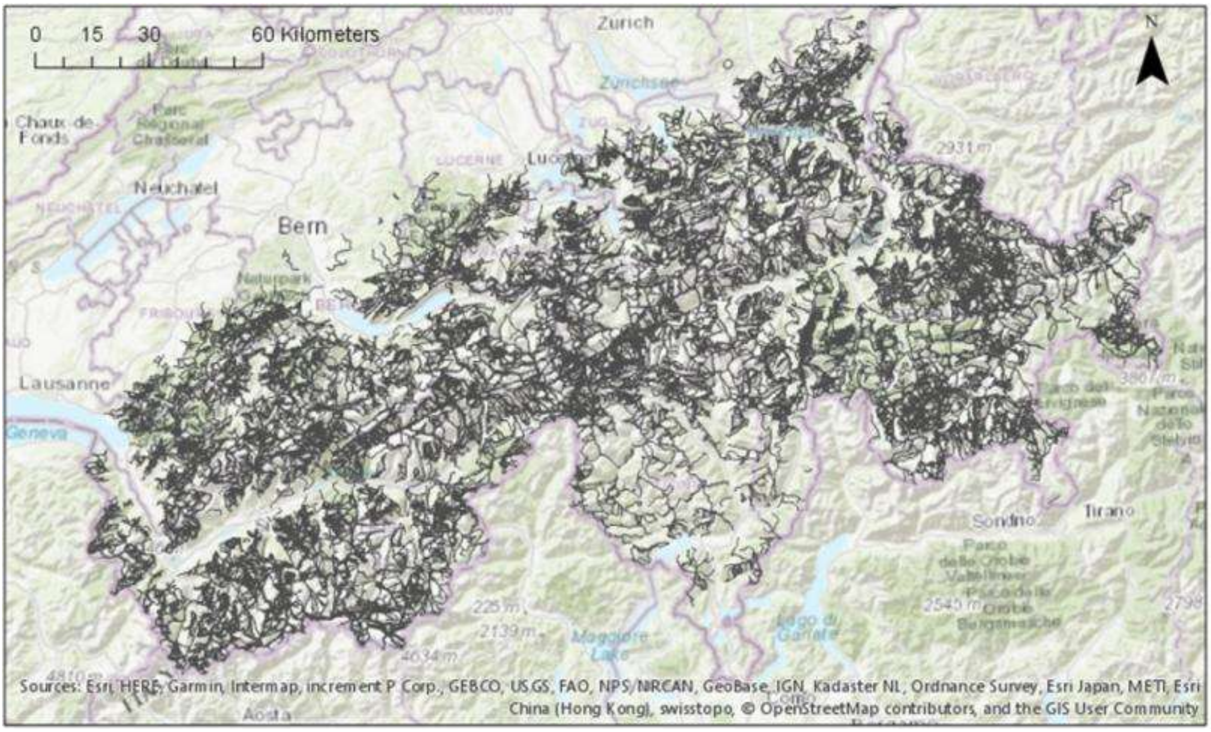


Figure 10. *all the ski touring routes from Whiterisk.ch (source: Whiterisk.ch, processed by C. Schönenberger, edited M. Hogeweij)*

The entire dataset has been visualized in figure 10. Some of the major valleys can be identified since they have no routes. Also, it is clear that the Italian-speaking part has less activity. This is in line with other studies into ski touring activity (Techel et al., 2015). The mean route length is 7628 meters, and the mean altitude difference within a route is 1067 meters.

3.2.2. GPS tracks from Wikiloc

A secondary dataset comprises GPS tracks of ski tours. Those were downloaded manually by the author from Wikiloc.com, a social media platform on which users can share tracks of their outdoor activities. The main difference between this dataset and the one from Whiterisk is that these tracks were all recorded with GPS devices, whereas the ones from Whiterisk were all drawn on a digital map.

When querying the website, thresholds were put in to leave out trips which were probably not ski tours (less than 200m altitude difference or less than 3km distance). Also, to make the data comparable to the Whiterisk data, probable multiday trips were filtered out. The threshold for this was 21008 meters in distance, the same as used by Schönenberger (2018) when filtering the planned route data. Then, some tracks where it was obvious a part was recorded in a car or where ski lifts were used, were manually filtered out. In the end this resulted in a dataset of 777 ski (and splitboard) touring tracks. As can be seen on figure 11 below, there are far less tracks. Also, the data is more concentrated in some specific areas, compared to the Whiterisk data which has coverage throughout the Swiss alps. To compare the coverage of the two datasets, the number of routes in each municipality with the size of the municipality they are in to make up their density. Then, the relative differences between municipalities in track densities were compared. Given the low number of Wikiloc tracks, this gave a better view of the route density than a normal line density measure. The density is quite evenly spread throughout the Whiterisk dataset. In the Wikiloc dataset however, there is an obvious concentration of tracks in some regions, and zero tracks in other regions.

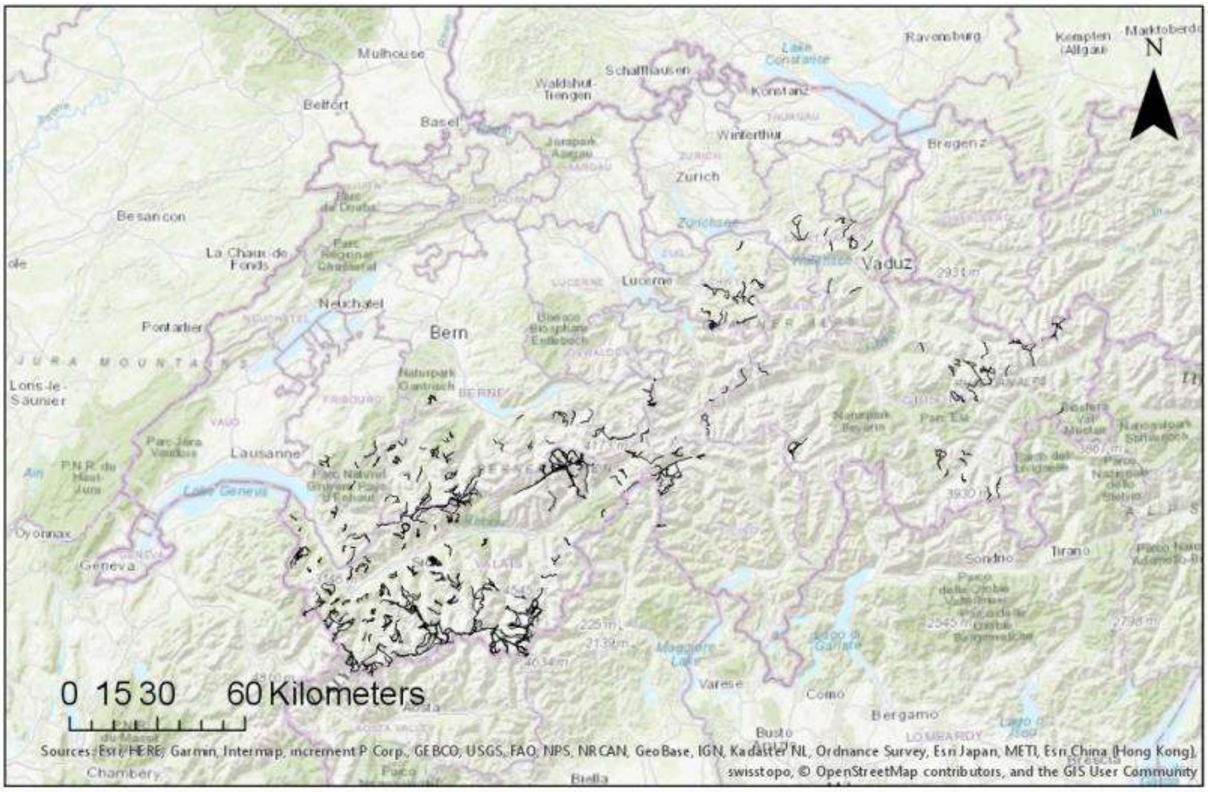


Figure 11: *All the used ski touring GPS tracks from Wikiloc.com (source: Wikiloc.com, edited M. Hogeweij)*

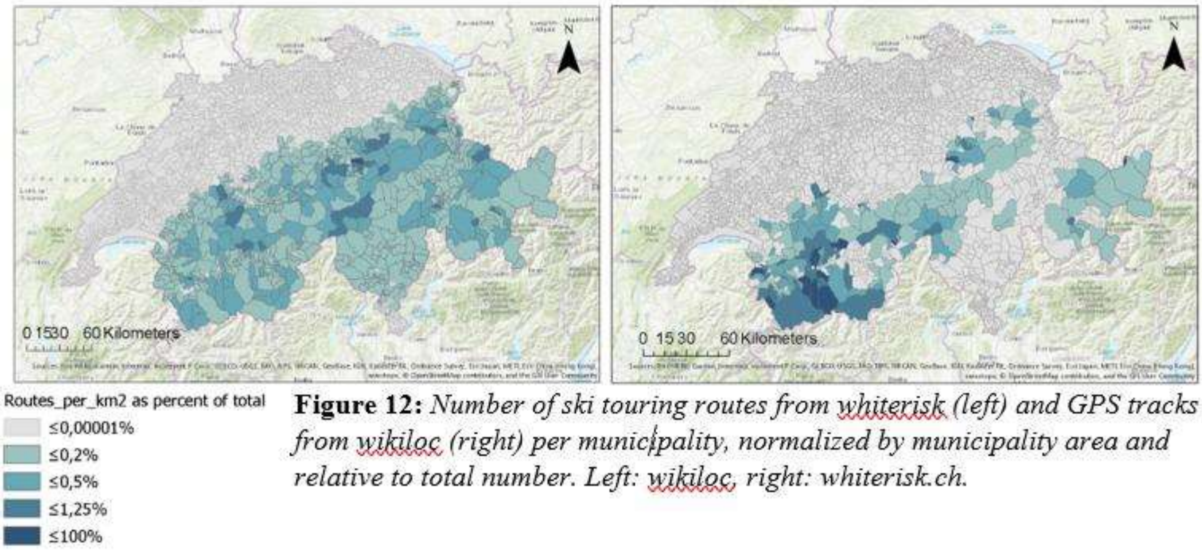


Figure 12: *Number of ski touring routes from whiterisk (left) and GPS tracks from wikiloc (right) per municipality, normalized by municipality area and relative to total number. Left: wikiloc, right: whiterisk.ch.*

There are various possible reasons for this uneven spread in the Wikiloc data, but it is not very important for the scope of this thesis. Since there are less Wikiloc routes, there are also more areas for the Wikiloc dataset with no routes at all. The high density regions are almost exclusively located in the southwest corner of the swiss Alps, in the French-speaking part of Valais/Wallis canton. This may tell us something about the set of users of Wikiloc, maybe it is more popular among people from the French-speaking part of Switzerland.

3.2.3. Avalanche terrain hazard model

Another dataset that is used in this thesis is the modelled avalanche terrain from Harvey et al. (2018). They modelled two different raster maps of avalanche terrain. One depicts the potential of triggering a small or medium-sized avalanche (which form the majority of recreational avalanche accidents). The other depicts the hazards (consequences) of triggering within the avalanche terrain.

Classified avalanche terrain

In the classified avalanche terrain map, each raster cell (resolution of 5 by 5 meters) is assigned to a nominal terrain class. The values that are not avalanche terrain are given a value of 0. The rest is split into areas from which avalanche releases are possible (trigger points), and areas where remotely triggered avalanches can runout into (so-called runout zones). These areas were modelled based on the terrain characteristics of 5200 mapped avalanche starting zones. The runout zones were computed with the avalanche simulation model RAMMS:EXTENDED (Bartelt, Buser, Valero, & Bühler, 2016). The triggering areas are divided into four classes, as are the remote triggering/runout zone areas. Although the different classes cannot be compared numerically, there is a decrease in triggering potential from "high triggering potential" down to "maximum runout zone". To give an impression of this dataset, a zoomed in view is provided in figure 13 of the area directly around Saas-Fee in the Wallis canton. Note that areas with no avalanche terrain are made transparent so that the underlying terrain is visible.

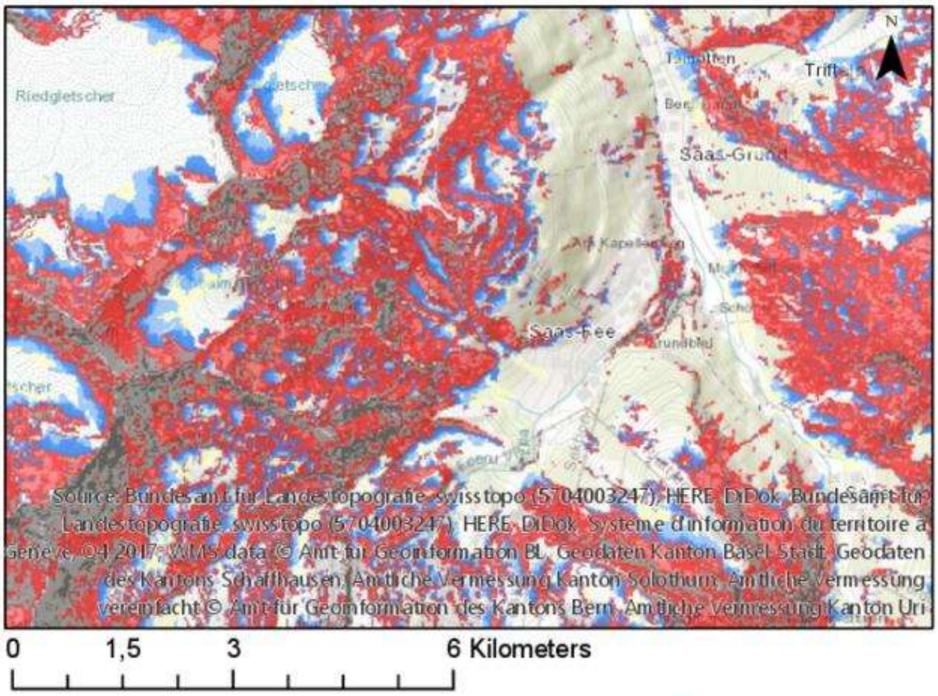


Figure 13: *Classified avalanche terrain for small area around Saas-Fee (Source: Harvey et al, 2018 edited by author)*

Avalanche terrain hazard map

In this map, each cell (5 by 5 meters) is assigned a value between 0 and 1, where 1 is most dangerous. The inputs for this are the map of modelled avalanche triggering potential and runout zones. This is combined with the potential and severity of burial in case of an avalanche, and with the potential of injury by falling when caught in an avalanche. The potential of falling was modelled by using a 10m DEM. Since the severity of potential consequences is combined with the avalanche release potential, it gives an overview of the hazard of the terrain, instead of just the potential for triggering an avalanche. For example, a place where triggering an avalanche is not extremely likely since it is not very steep, may be assigned a low value for triggering potential. However, if it is right above a steep drop, it will have a higher value for terrain hazard, since the potential consequences are severe. The downside of this approach is that the thematic difference between different sources of hazard cannot be seen by the viewer. In other words, it can't be seen from just looking at the values whether a place is dangerous because of falling risk, because of triggering risk, or because it is in a run-out zone. The advantage of this dataset is that cells can be compared numerically as they are ratio type. An example of this data is provided in figure 14, showing the same area around Saas-Fee.

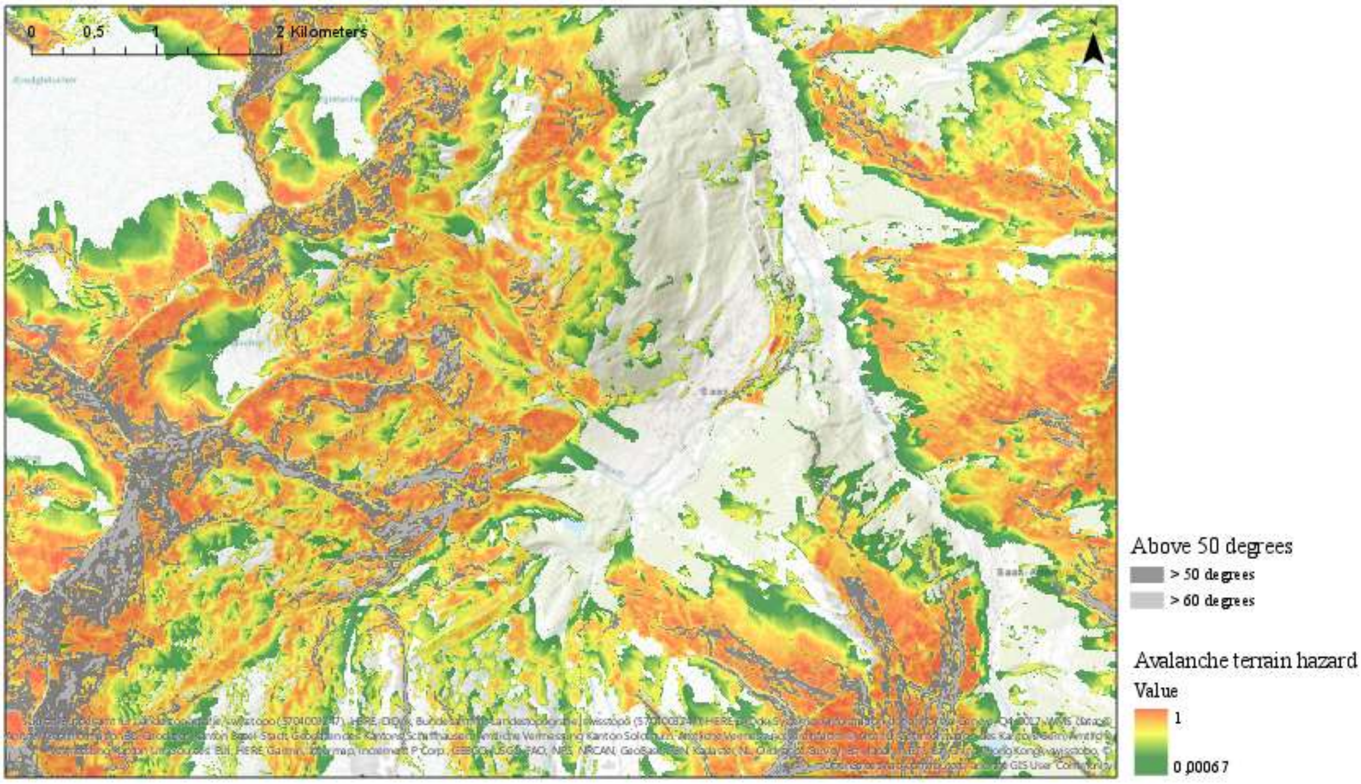


Figure 14. *avalanche terrain hazard for small area around Saas-Fee (source: Harvey et al, 2018, edited by author)*

3.3. Methods

The following subsections present the methods in the order of the research questions that they relate to.

3.3.1. Processing terrain rasters

In this thesis the avalanche terrain data is used to compare and classify routes. Before this can be done, the avalanche terrain data has to be processed. The data was clipped with the Swiss country boundaries as well as with the outer bounds of the area in which all the planned routes are located.

In order to interpret the results of the raster overlay with the routes, some knowledge is needed on the distribution of values in the original classified avalanche terrain raster. For this purpose, the data was clipped with the extent of the area in which the routes were planned, so that the non-mountainous regions were excluded. For this, the minimum bounding box around the routes was used. The goal of this is to exclude terrain that is out of the area where routes are planned. Then, the total surface of each terrain class within that polygon was calculated. This gives an overview of the spread of values in the avalanche terrain raster. However, within the bounding polygon of all planned routes, many areas where no skiing activity takes place are included. There are some very wide valleys within these boundaries. Therefore, to make a more accurate assumption of the values in the terrain rasters, they were clipped again, but this time with each individual bounding box of each route, and the surface of each raster class was again calculated. The difference between both methods is shown in figure 15.

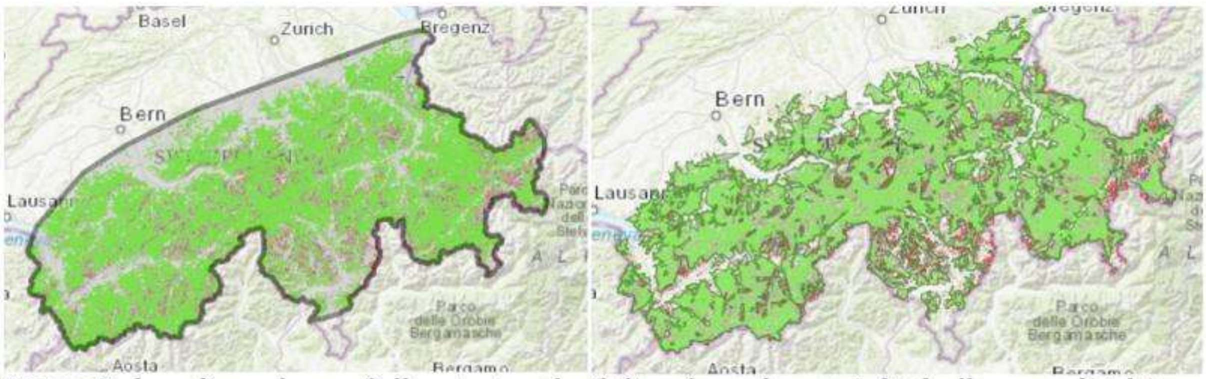


Figure 15. *bounding polygon of all routes together(left) or for each route individually, merged(right)*

The reason why this was done is to be able to compare the terrain in planned routes to the general terrain. The results of this are discussed in section 4.1.

3.3.2. Defining the avalanche terrain characteristics of planned routes

A goal of this thesis is to analyse to what extent planned routes are traversing avalanche terrain. Therefore, the information contained in the terrain rasters needs to be added to the vector datasets of the routes. This is done differently for both raster datasets, since the data contained in them is of a different nature. In the terrain hazard dataset, the data is at a ratio measurement scale. Because of this, descriptive statistics can be calculated based on the intersection of each route with the hazard data, and these can be used to compare routes. The classified avalanche terrain data on the other hand is of ordinal nature. Cells in the class "high triggering potential" have a higher potential for triggering avalanches than those in the class "medium triggering potential". However, this difference cannot be quantified.

Mean terrain hazard values

The terrain hazard data was used to calculate the mean "hazard rating" across all routes. This is shown for the routes in the Pigne d'Arolla study area in figure 16.

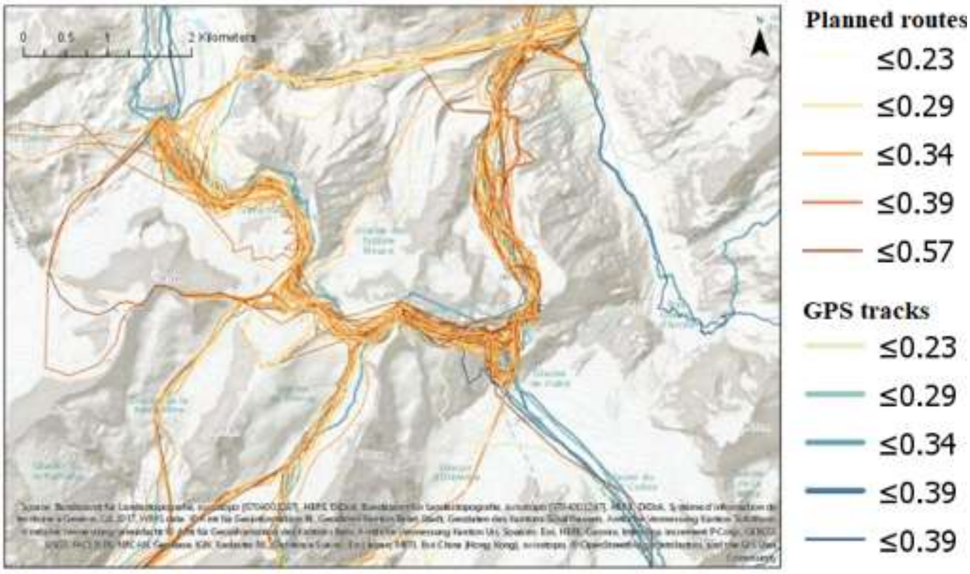


Figure 16: *GPS tracks and planned routes leading up to Pigne d'Arolla, coloured by mean terrain hazard rating.Source: Whiterisk.ch, Wikiloc.com, edited by author*

To calculate this mean terrain hazard value, the cumulative value of all terrain hazard cells intersected by a route is divided by the total number of cells in that route. The results of this operation are discussed in more detail in section 4.1.

95th percentile values

When one wants to know how dangerous a route is, the mean hazard rating has some flaws. This is further discussed in section 4.1. As is often the case with mean values, longer routes tend to have lower values, even when they traverse the same dangerous sections as short routes. Therefore, to assess how dangerous a route is, the maximum values for terrain hazard may be more useful. However, Schönenberger (2018) shows in his thesis that using maximum values is also problematic due to minor mistakes which lead to large differences in route planning when a route intersects a high value cell for a short amount of time. Hendrikx et al. (2015) show that using top percentile values is a useful measure when analysing route trajectories. Here, instead of using the value of the highest valued terrain cell, one uses the value of the 95th percentile highest terrain cell. The advantage of this is that a single cell with a very high value, which may be there due to an inaccuracy in the data, does not give a route an artificially high value as it would when using maximum values. In other words, the 95th percentile values are deemed more robust to minor inaccuracies than the maximum values.

However, computing top percentile values is not straightforward here, as the route data used are polylines instead of points, which makes discrete sampling impossible. Also, the avalanche terrain data is continuous. To avoid this issue, the terrain hazard is rounded and divided into 1000 discrete classes (from hazard 0.001 to 1, with intervals of 0.001). This makes it possible to see the percentage of each terrain hazard value that is traversed by a route. From this, the 95th percentile values are calculated. First, the cells intersected by a route are ordered by value. Then, the 95th percentile point of this list is found. Then, the value of the cell at this point is returned. This is illustrated in figure 17. As an example, the 95th percentile terrain hazard values for the Pigne d'Arolla sample are shown in figure 18. The pattern in the 95th percentile values shows some similarities to that in the mean hazard values. The difference between both of the measures is discussed in more detail in section 4.1 in the results chapter.

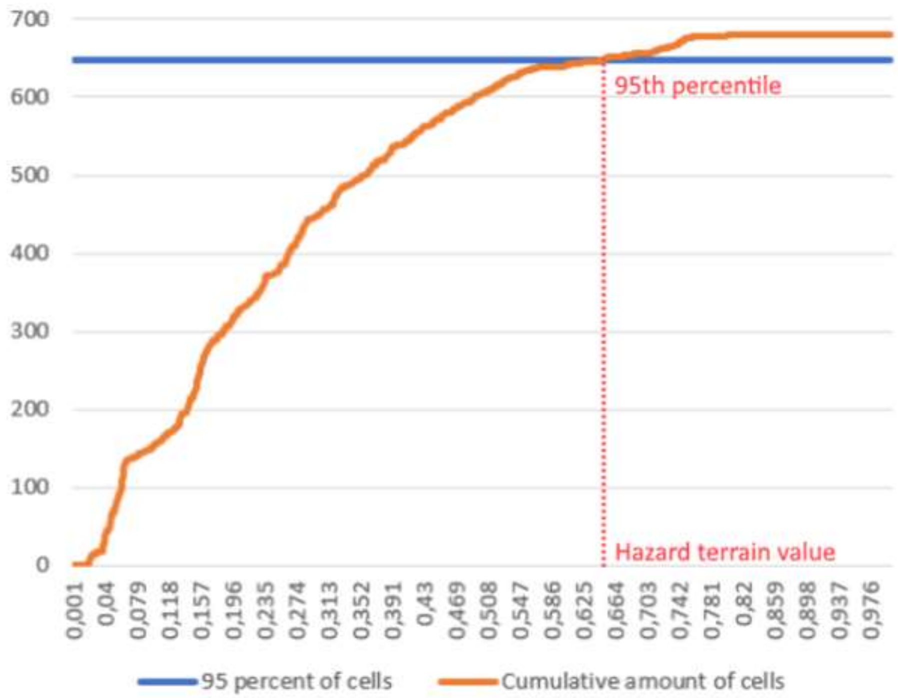


Figure 17. *Graph showing the 95th percentile of terrain hazard in a route*

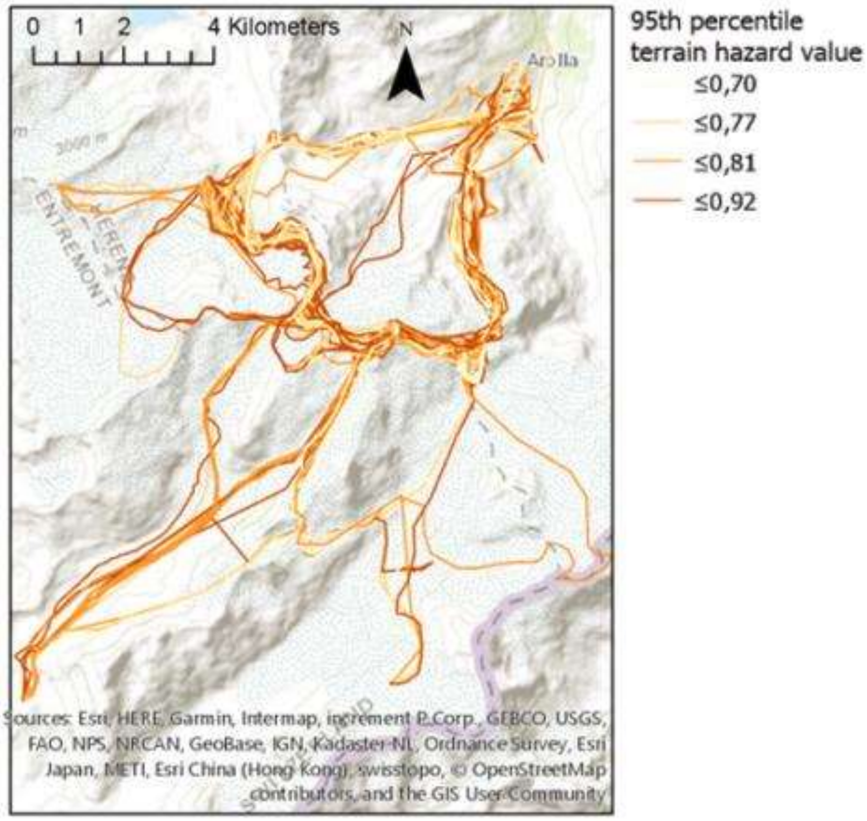


Figure 18. *95th percentile terrain hazard around Pigne d'Arolla*

3.3.3. Clustering based on avalanche terrain characteristics

A next step is clustering routes based on avalanche terrain. For this, the discrete terrain classes were used instead of terrain hazard. The reason for this is that they hold more semantic meaning. Two routes may have similar values for 95th percentile or mean hazard, but still be highly different routes. For example, one could be very steep in some sections while very flat in the rest, whereas the other is moderately steep throughout. To do this, the percentage of each route in each of the 12 terrain classes was calculated. Then, these values are normalized using:

$$z = \frac{(x-u)}{s}$$

where x is the percentage of a route in that terrain class, u is the mean value for that class amongst the whole population, and s the standard deviation. As an example, the chart in figure 19 shows those scores for a single route.

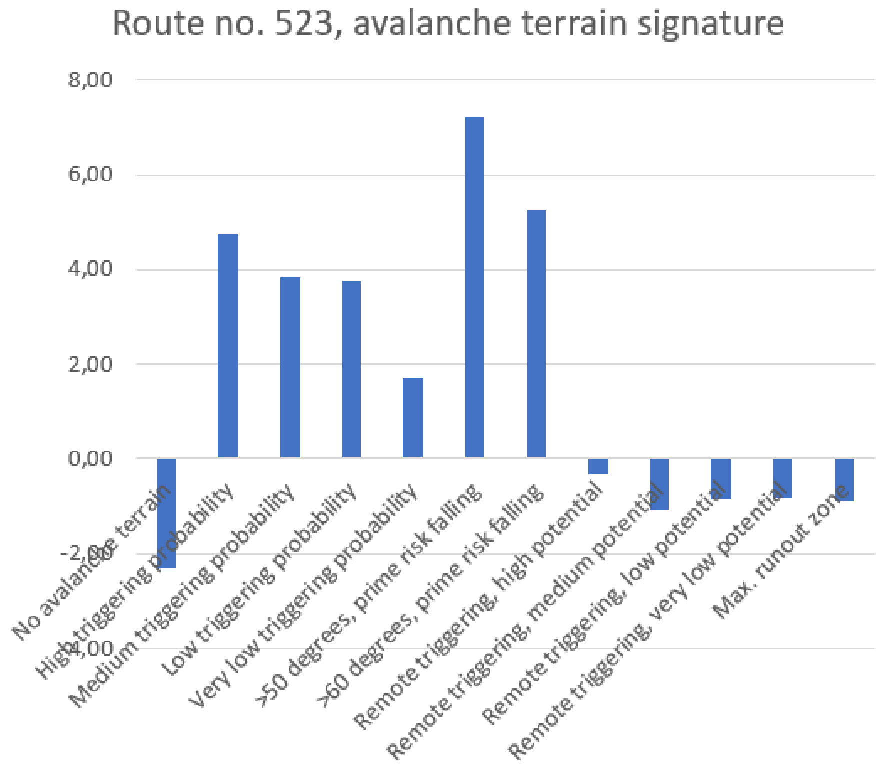


Figure 19. *Avalanche terrain "signature" of single route showing normalized proportional area in each discrete avalanche terrain class*

For each terrain class, it is shown how much that route traverses it, relative to the mean of all routes. These scores can be seen as the "terrain signature" of a route. As can be seen, this route has below average scores for the terrain classes "no avalanche terrain" and all the remote triggering terrain classes. It has above average scores for high to very low triggering terrain, and very high scores for above 50 and above 60-degree terrain. As such, this route traverses relatively much of that terrain. Therefore, this route probably is planned in a steep area, for example a high alpine peak.

A hypothesis at this point is that there exist patterns in the data related to this avalanche terrain signature. For example, upon manual inspection it was found that routes leading to high alpine peaks (for example the Weisshorn at 4505 m.a.s.l.) show very high scores for terrain above 50 and 60 degrees, since to reach those peaks they traverse very steep ridges. This makes them very different from almost all routes, since most skiers tend to avoid terrain over 50 degrees.

Because of this hypothesis, it is useful to try to cluster the routes based on their terrain signatures. First because this can give insight into the characteristics of planned routes in general. For example it is interesting if there exists a class of routes with extreme terrain, and how many routes are part of that class. Second because these terrain signature clusters can be compared to spatial clusters of routes. Whether or not terrain clusters and spatial clusters are related is interesting to know for route planners.

For example, heterogeneous areas in terms of terrain clusters are more interesting for avalanche safety training courses, because when planning a route there, more meaningful planning decisions can be made.

For the clustering based on avalanche terrain, different methods were considered. At first, the DBScan algorithm was tried. The reason for this is that it wasn't known beforehand how many clusters exist in the data, and in DBScan the algorithm decides the amount of clusters suitable for the data. This algorithm uses 2 parameters, epsilon and minPts. For each point, it is assessed whether it has minPts around it within a sphere of diameter epsilon. If this is the case, the point is labelled as a core point. Then, for each core point, the points within that points epsilon are added to the cluster. Then, the cluster is grown by adding every point that is within epsilon of any point in the cluster. However, the results here were not satisfactory. The algorithm ended up either choosing a large number of small clusters, or one very large one. However, the goal here is to categorize the routes in a way that they can be compared across the dataset, which is impossible with numerous clusters of around 30 members. Different parameters were tried, but the results did not end up being useful.

Another method tried was principal component analysis (PCA). This is useful when there are multiple dimensions, as is the case here, since it can help reduce the dimensionality and thus increase the clarity of the clusters. In this technique, the variables are converted to principal components, with each principal component correlating negatively or positively with each of the variables. Then, the correlation between each point (in this case each route) with each component is shown. However, too much information was lost when doing this. Also, the correlation between the principal component and each of the variables was generally quite low (between -0.3 and 0.3) which makes it hard to use the components to say something useful about the variables. Besides, the number of dimensions was not so high (12) that dimensionality reduction was seen as absolutely necessary.

The next approach which was tried was k-means. Here, the only user-defined parameter is k. In the first iteration, k centroids are computed for the data. Then, each object is assigned to the cluster belonging to the nearest k. Then, the centroids of each cluster are recomputed as the centroid of all points assigned to that cluster. Steps two and three are repeated until the centroids no longer move. The value given to k is of great influence on the outcome of clustering. Different methods exist for defining k. The one chosen here is the elbow method. This method comprises calculating the sum of squared distances of each point to its cluster centroid. This is done repeatedly for an increasing k value. The higher k, the smaller the inertia. Ideally, both inertia and k should be as low as possible. Therefore, when plotting k and inertia, the ideal value for k is at the elbow of the graph, which is the point after the largest relative decrease of inertia.

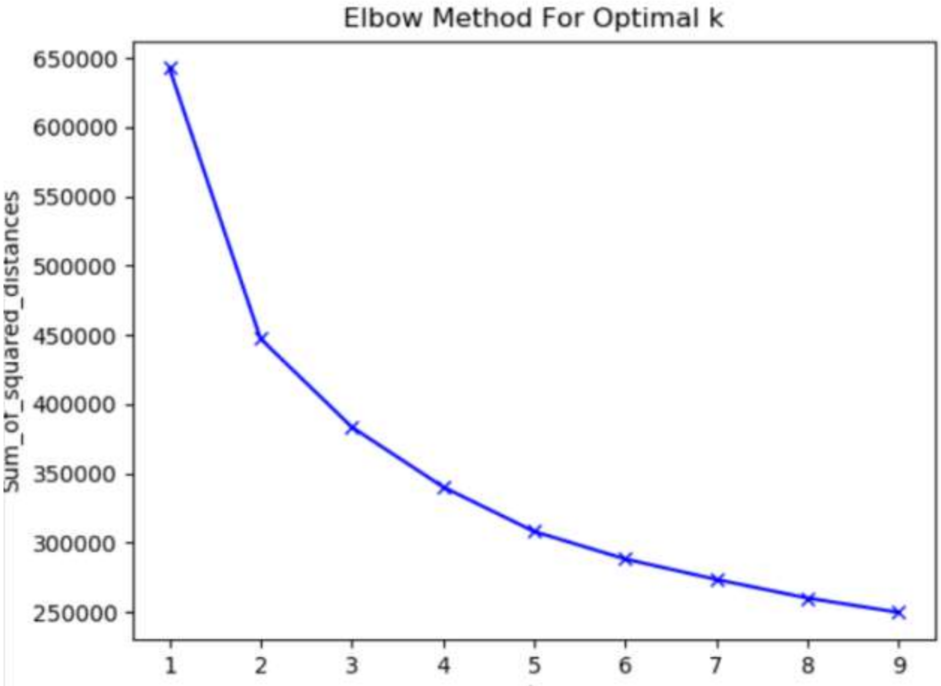


Figure 20. *K means and inertia for the attribute terrain classes in routes*

The elbow in figure 20 is found at k=5 (although there is also a rapid increase in slope at k=1, but this would make clustering useless). Therefore, this value was chosen for k. The elbow method does not always produce the optimal amount of clusters for the data and goal of analysis provided. Also, the elbow in figure 20 is not extremely clear, one could also argue that there is an elbow at k=4 or k=6. Therefore, other values for k were also tried to see whether any more meaningful clusters showed up, but this was not the case. Therefore, 5 was chosen as the best value for k. The results of this are provided in chapter 4.2.

3.3.4. Spatial clustering

Visual inspection shows that there exist spatial corridors in the routes. In the Pigne d'Arolla example, some clear patterns can be defined, as there are a few standard routes to the peak that are more or less followed by most routes. Therefore, it makes sense to cluster the routes spatially. The reason why this is useful is that it allows for comparison of individual routes. Routes with the same general trajectory can be compared to see whether they are different in terms of avalanche terrain, and what planning strategies are the cause of those differences.

Different methods exist for computing the similarity of two lines. The most simple is to compute the centroid of each line and calculate the distance between the centroids. This however is often inaccurate, as routes can be planned in any direction, and this method also does not take into account route length. A slightly more complex method is Hausdorff distance. This calculates the minimum distance from any point in a line to any other point in the other line. However, this does not take into account the sequence of the points which make up the lines, and is also inaccurate when lines have some points closely together but a majority of points far away from each other (Alt & Scharf, 2012).

The measure chosen here is discrete Fréchet distance (Eiter, Mannila, & Eiter, 1994). This is a variety of Fréchet distance. Fréchet distance is the maximum distance between two ordered trajectories. This is often imagined as a man walking a dog along a leash, from a starting point towards an end point. They both cannot backtrack. The Fréchet distance is the minimal length of the leash between the man and the dog needed to traverse the two trajectories. The reason this measure is chosen is that it is suitable for ordered trajectories, and takes into account subtle differences between trajectories, as well as using the whole of the trajectory instead of just the closest points.

The discrete Fréchet distance is an approximation of this distance which only takes into account the nodes of both lines. This method uses the sequence of points within each line, but does not need a time stamp. Instead, the points are seen as an ordered list. This is ideal for the dataset at hand, as the points in this dataset have no timestamp. Further, a python module (trajectory_distance) was available for download which already included an implementation of discrete Fréchet distance which takes variable record lengths into account, which is required for this dataset. The disadvantage of the method used here is that when there is a large difference between number of points, the method loses accuracy, as the line with less points will be assigned a larger distance to other lines then it has in reality. However, this is accepted as lines with less points are generally less accurate, whichever method is used.

A problem faced at this point was the size of the dataset. Since in the most simple implementation for each line, the Fréchet distance to every other line has to be checked, giving the algorithm a complexity of $O(n^2)$. To speed this up, the Euclidian distance between route centroids was first computed, and only the 200 closest lines in Euclidian distance were assessed when calculating the Fréchet distance. Although computing the Euclidian distance of each line to each other line also has $O(n^2)$ complexity, it is much simpler since it only needs to process a single metric for each line, as opposed to each point of each line.

To cluster the routes, an algorithm has to be picked. Several algorithms exist for this purpose. They can be grouped into hierarchical, partitioning, grid-based, and density-based algorithms (Ram, Jalal, Jalal, & Kumar, 2010). Partitioning methods are not suited here, as they need a priori definition of the number of clusters. Given the continuous nature of this dataset, it is impossible to establish the number of clusters needed beforehand. Hierarchical methods suffer from the same problem. Grid methods were considered, but no python implementation could be found, and developing an algorithm seemed tedious work. Therefore, a density-based approach was chosen. The chosen approach was DBSCAN, which has an implementation in the scikit python module. How this algorithm works was already explained in section 3.3.3. The implementation of DBSCAN used here allows the user to input a custom distance matrix, instead of calculating the distances between lines as part of the clustering. The Fréchet distance from each route to the nearest 200 routes was input for this. An added advantage of density-based clustering is that the problem of differences in number of points, identified above, becomes less serious. The reason for this is that to join a cluster a line need not be similar to each line in that cluster, it only has to be within the threshold of one other line.

Using the right parameters is important for DBScan, as it greatly influences whether clusters are accurately classified as such. In this case this was problematic, as the spatial similarity of routes is influenced by the terrain. For example, at alpine peaks all routes follow a narrow ridge, and therefore they are very similar. On large glaciers however routes are quite far apart while still being part of the same cluster. Therefore, it is impossible to assign the perfect parameters for the entire dataset. Experimenting with different parameters showed that the best parameters for the whole dataset were epsilon = 600 and minPts = 10. This was on the strict side, to prevent false positives, with the drawback of not including some routes that were broadly part of a cluster. The results of this are shown in chapter 4.3.

3.3.5. Assessment of similarity of avalanche terrain for routes within a spatial cluster

To find out the influence of the surrounding terrain on a route, the spatial clusters are used. These are analysed in combination with the avalanche terrain attributes in the routes, and the avalanche terrain clusters. As such, the outputs of research questions 1 to 3 are used as input for research question 4. A combination of qualitative and quantitative methods will be used for this. The quantitative part will mean calculating the intra-cluster variance in terms of cluster membership and 95th percentile values.

For the 95th percentile value, Pearson's r will be calculated for each spatial cluster to assess whether the values are normally distributed. If they are, standard deviation is used as a measure of variance. To measure the variance in terms of terrain cluster membership, the ratio between the count of the most occurring terrain cluster and the total number of routes within each cluster is calculated. The higher this number, the more homogeneous a spatial cluster is in terms of terrain cluster membership.
The qualitative part of this question is be comprised of visual inspection. Spatial clusters that are either very heterogeneous of homogeneous are projected on a map. Then, they are inspected in order to find out why they are like that. Also, the area around the routes is viewed in the Whiterisk website to find out whether there are any pointers there as to why routes are planned in the way they are.

3.3.6. Comparison of GPS tracks and planned routes

To answer this question, the GPS tracks from Wikiloc are compared to planned routes. First, the values calculated in question 1 are compared for the entire dataset. This gives an overview of the avalanche terrain in both datasets. After this, the terrain clustering steps are repeated for the GPS tracks. For this the percentage of each GPS track that goes through each discrete avalanche terrain class is calculated. Then, this data is added to the signatures of the planned routes and the clustering is repeated. The clustering is not performed for the GPS tracks separately as that would lead to different clusters which makes comparing them to the planned routes impossible. The third comparison will be on a more detailed level. A number of spatial clusters will be chosen with a sufficient number of planned routes and GPS tracks. Then, it will be reviewed whether there are any differences between the two in dealing with the avalanche terrain in their routes.

4. Results

In this chapter, the results of the analysis are presented in the order of the related research questions.

4.1 Relating modelled avalanche terrain to planned routes

4.1.1. Bounding polygons

The first step in this thesis was establishing a method to treat modelled avalanche terrain as a route attribute. For this, several methods were used. First, the bounding polygon of the terrain in which routes are planned were used. In this bounding polygon, relatively much terrain is in the category "not typical avalanche terrain or no data". This is because this area includes some major valleys without steep slopes. Then, the bounding polygon of each route individually was used to clip the avalanche terrain and calculate the surface of each terrain class. As can be seen in table 3, there is now less terrain in the "no avalanche terrain" class. That is because some of the terrain where no skiing takes place because it is too flat, is now excluded.

Figure 21. *Bounding polygon of all routes, inner areas without routes filled*

Value	Area in km²	% of area
Not typical av. terrain or no data	15175.94	58.08
High triggering potential	1153.12	4.41
Medium triggering potential	2645.20	10.12
Low triggering potential	1529.14	5.85
Very low triggering potential	820.97	3.14
> 50 degrees slope, primary risk falling	958.74	3.67
> 60 degrees slope, primary risk falling	399.66	1.53
High potential for remote triggering	337.06	1.29
Medium potential for remote triggering	884.75	3.39
Low potential for remote triggering	473.23	1.81
Very low potential for remote triggering	971.11	3.72
Max. runout zone	779.55	2.98
Total	26128.48	100.00

Table 2. *area of each class of classified avalanche terrain within bounding polygon*

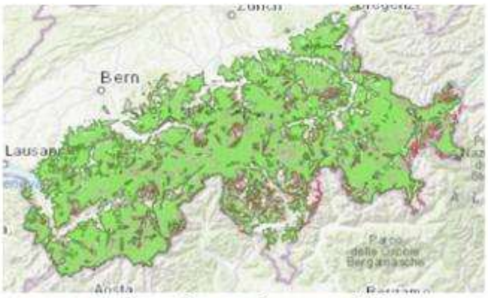


Figure 22. *bounding polygons of all routes, inner areas without routes not filled*

Value	Area in km	% of area
No avalanche terrain or no data	8477.30	49.95
High triggering potential	858.01	5.06
Medium triggering potential	2024.27	11.93
Low triggering potential	1184.59	6.98
Very low triggering potential	641.41	3.78
> 50 degrees slope, primary risk falling	666.87	3.93
>60 degrees slope, primary risk falling	265.17	1.56
High potential for remote triggering	276.15	1.63
Medium potential for remote triggering	724.93	4.27
Low potential for remote triggering	389.23	2.29
Very low potential for remote triggering	803.25	4.73
Max. runout zone	659.60	3.89
Total	16970.76	100.00

Table 3. *Area of each class of classified avalanche terrain, within minimum bounding polygons of planned routes*

As a next step, for each route the intersecting cells of the discrete terrain classes were counted in an overlay operation. The result is a table like tables 2 and 3, but for each route individually. When counting up the amount of cells in each route and calculating the terrain, the resulting table gives an overview of the relative amount of terrain of each class intersected by an average route.

	GPS tracks	Planned routes
Value	**% of routes**	**% of routes**
No avalanche terrain or no data	59,29	60,81
High triggering potential	2,60	3,19
Medium triggering potential	4,87	5,67
Low triggering potential	2,39	2,68
Very low triggering potential	1,28	1,32
>50 degrees slope, primary risk falling	0,35	0,45
>60 degrees slope, primary risk fallig	0,12	0,11
High potential for remote triggering	1,78	1,52
Medium potential for remote triggering	5,79	5,15
Low potential for remote triggering	3,77	3,34
Very low potential for remote triggering	8,75	7,85
Max. runout zone	9,01	7,9
Total	100,00	100

Table 4. *Mean percentage of terrain classes intersected by routes (route length taken into account)*

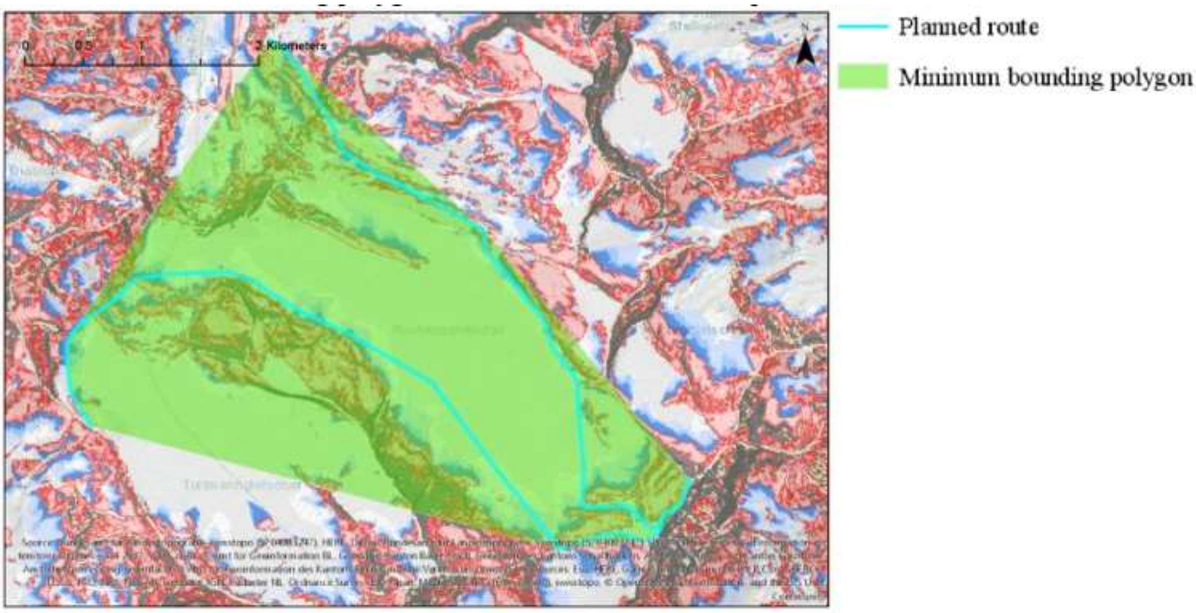


Figure 23. *route with bounding polygon*

Interestingly, the routes on average traverse more terrain defined as no avalanche terrain, and less potential triggering terrain than is included in their minimum bounding polygons. Perhaps one would expect the routes to traverse steeper terrain on average then is included in their bounding polygons, since steeper terrain is often seen as more fun to ski. This is, however, not reflected in the data. When visually inspecting some bounding polygons and the routes they are based on, it becomes clear why this is the case. The routes often follow terrain corridors that are easy to follow, and usually go around extreme terrain. The bounding polygons however include this steep terrain. This is clearly visible in the above example: whereas the route mostly traverses two rather flat glaciers, the bounding polygon includes the very steep ridge in between the glaciers, as well as some other steep terrain. Because of this, minimum bounding polygons should only be used to give an overview of the modelled avalanche terrain, and are not suited to represent the terrain traversed by actual routes.

4.1.2. Mean terrain hazard

When trying to classify routes based on the avalanche terrain they traverse, the modelled terrain hazard is a suitable dataset. One way of relating the routes and the terrain data, is by calculating the mean value for intersecting cells of the terrain hazard dataset for each route. This is done by simple counting the values of all raster cells intersected by a route and dividing this number by the total number of cells intersected. Since the terrain hazard is modelled as a ratio value between 0 and 1, the mean value for terrain hazard can be compared between routes. For the entire population of routes both planned routes and GPS tracks, the mean of this value is 0,307, and the standard deviation is 0,088. The histogram of the mean terrain values is shown in figure 24. The goal of using this data as a route attribute, is ranking or classifying routes based on the avalanche terrain they traverse. As such, the mean terrain hazard value should be a good representation of how "good" the route planning is at avoiding avalanche terrain. To what extent this is the case is illustrated on the next pages by using the Pigne d'Arolla area as an example.

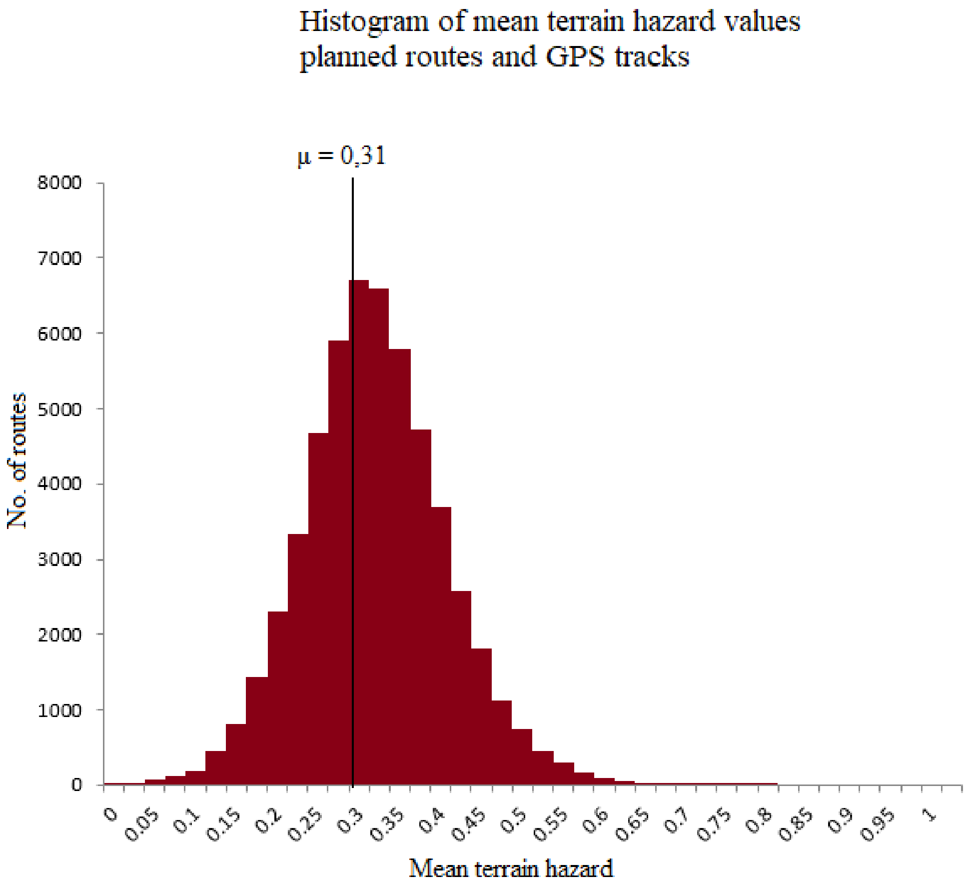


Figure 24. *mean terrain hazard value of all routes*

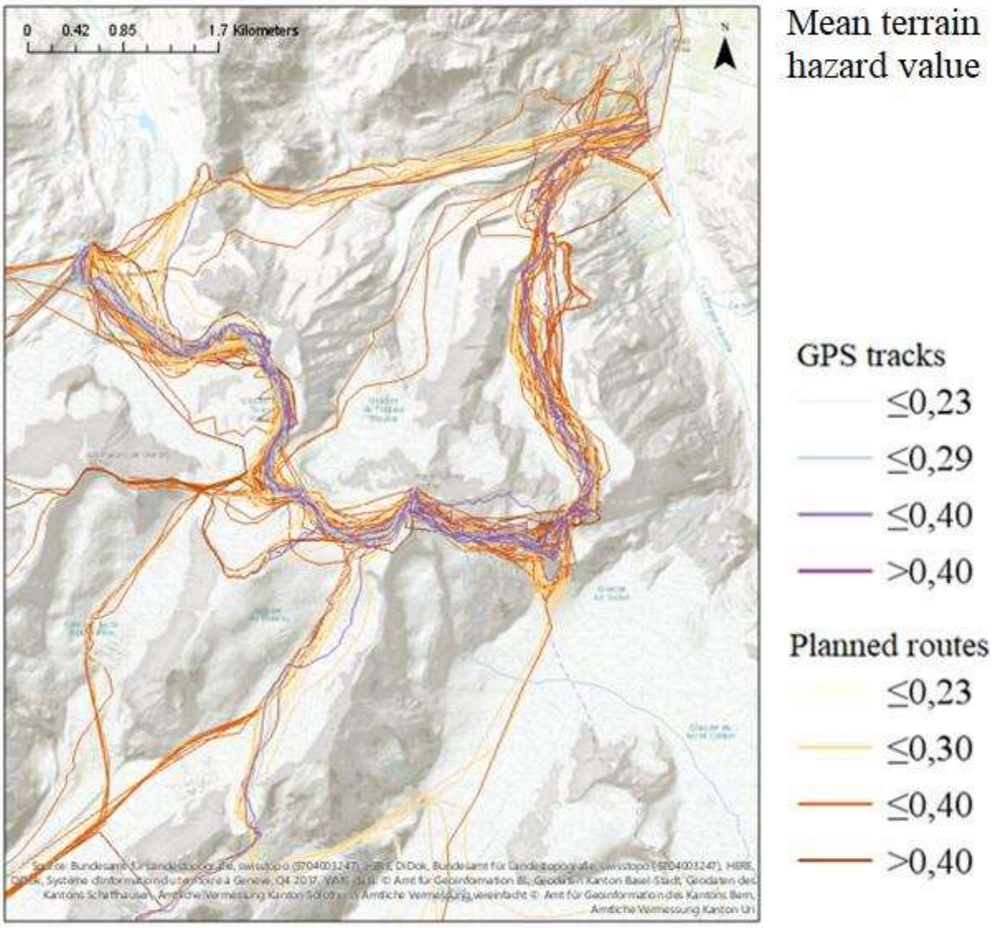


Figure 25. *routes leading to the peak of Pigne d'Arolla, coloured by mean terrain hazard*

The above figure shows the mean hazard value for planned routes and GPS tracks around Pigne d'Arolla. There is not a clear pattern to be identified in the mean terrain hazard values in those routes. On first eye it seems that the routes that approaches the peak from the east have slightly higher values than the ones coming from the west. Also, there is not a clear difference between the GPS tracks and the planned routes as far as the mean terrain hazard values go. What is interesting is the fact that the GPS tracks generally follow the same terrain corridors as the planned routes. Mostly though, there seems to be a high variance of mean hazard values between routes that follow a similar trajectory. This may indicate that mean hazard values are not suitable for measuring the amount of risky terrain a route traverses: spatially similar routes should have similar values. However, for a large part this is due to the fact that many routes have similar trajectories, but different start and end points. For example, some routes finish at the peak, while other traverse it and go down on the other side, which of course influences the terrain they meet on the way. When only viewing routes with similar start and end points, there is a clear relation between spatially similar routes in terms of mean hazard values.

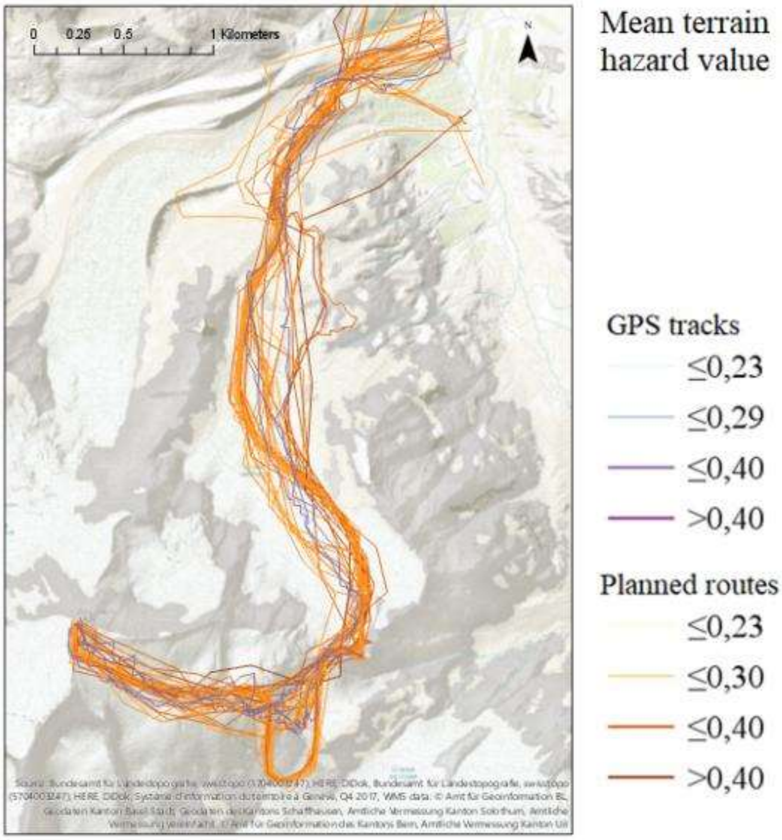


Figure 26. *routes within one corridor to Pigne d'Arolla, coloured by mean avalanche terrain hazard*

Using the same colouring scheme as before, most routes now seem quite similar in terms of mean hazard values. The mean value is now 0,32, with a standard deviation of 0,04. Still, there are differences between the routes. Two passages seem important in determining the mean hazard value assigned to routes. This is shown in figure 26. Here, only routes that are one standard deviation above or below the mean terrain hazard value are shown. There are two passages that seem to determine the value. First, the routes pass a through a broad glacier. The eastern way through clearly passes more terrain with high hazard values. Then, just before the top, the low-value routes all take a detour south, and by doing this they go around an area with high values.

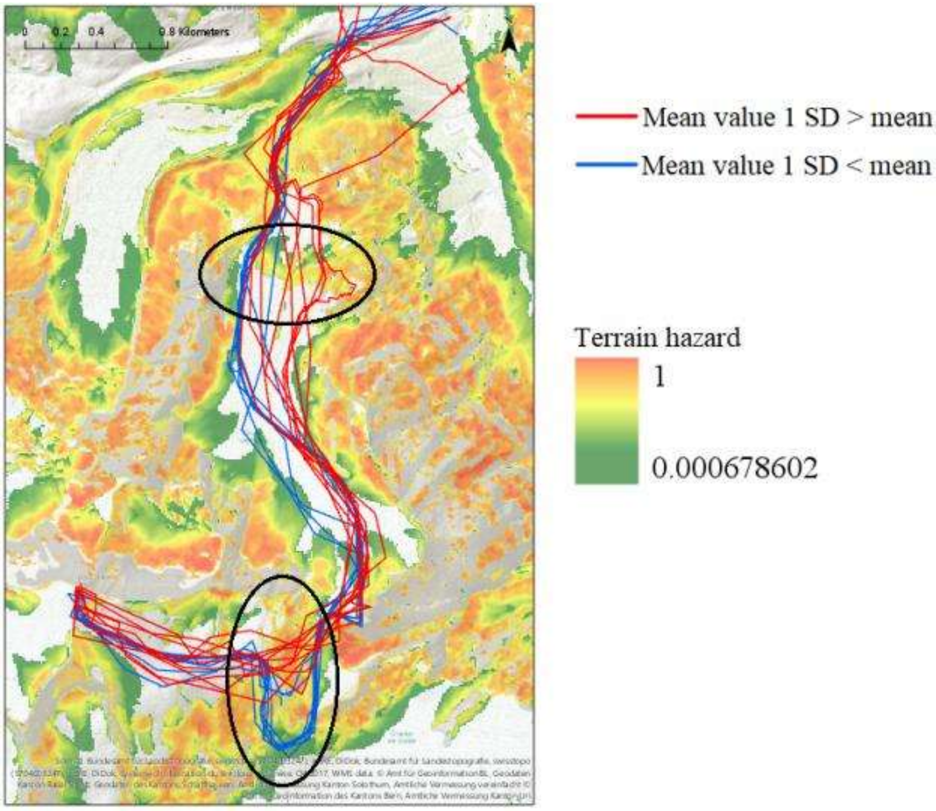


Figure 27. *routes in one corridor to Pigne d'Arolla, 1 sd above or below mean terrain hazard value*

This example shows that mean terrain hazard values do well at showing whether a route passes through terrain with high hazard values. However, there are also cases when it does not work so well. Longer routes tend to have lower values than shorter ones that pass the same high-risk sections. This is of course logical. However, it is not an "accurate", since these routes can be seen as carrying the same risk, they just have more low-risk sections bringing the value down. This effect is enhanced when a long route also passes through a lot of cells with values near zero. These cells may have almost the same terrain as adjacent no-data cells, but no-data cells do not count towards the mean value. An example of this is shown in figure 27 below. This route has a mean hazard rating of 0,24, well below the mean. However, it passes through two sections with very high terrain hazard, encircled in black. As such, it is not a "safe" route. However, because it passes through many cells with near-zero values (encircled in blue), it gets assigned a low mean hazard value.

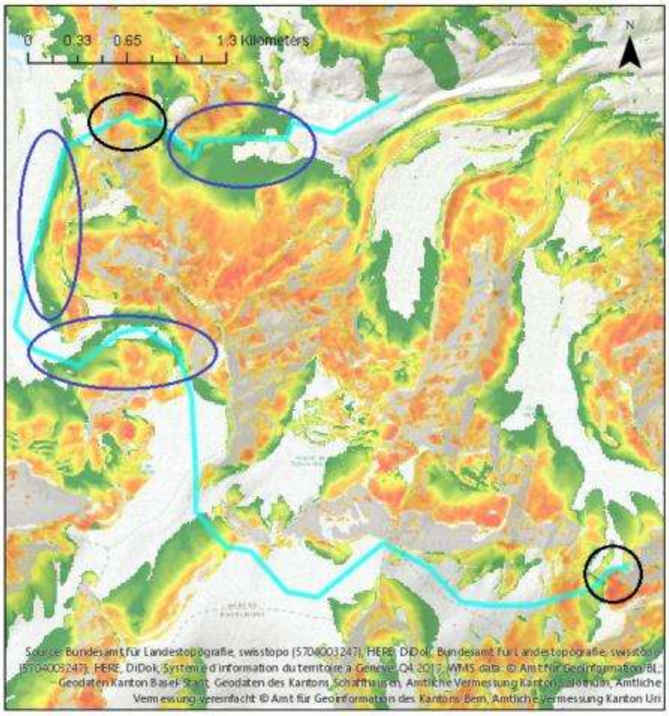


Figure 28. *a long route near Pigne d’Arolla with a low mean hazard value*

For the majority of routes, mean hazard values do give a good view of the risk in a route. As such, they will still be used throughout this thesis as an indication of the terrain hazard in a route.

However, the above example has shown that it does not work well in all cases. Passing dangerous sections should give a route a high value for avalanche danger, regardless of the other terrain it passes through. As such, a next step is to find a way to represent the amount of risky terrain passed by a route, ignoring the flat sections.

4.1.3. 95th percentile hazard values

The previous section has shown that in some cases, it is necessary to show the top values of avalanche terrain hazard in a route as opposed to the mean value. One could simply calculate the maximum value intersected by a route. However, Schöenberger (2018) has already shown that calculating max. values for terrain intersected by planned routes is problematic. The reason for this is that minor differences in planning can make a very large difference when a single high-risk cell is intersected. Therefore, instead of maximum values, the 95th percentile values of terrain hazard in routes was calculated. How this was done was already explained in the methods chapter. The mean 95th percentile hazard value is 0,64.

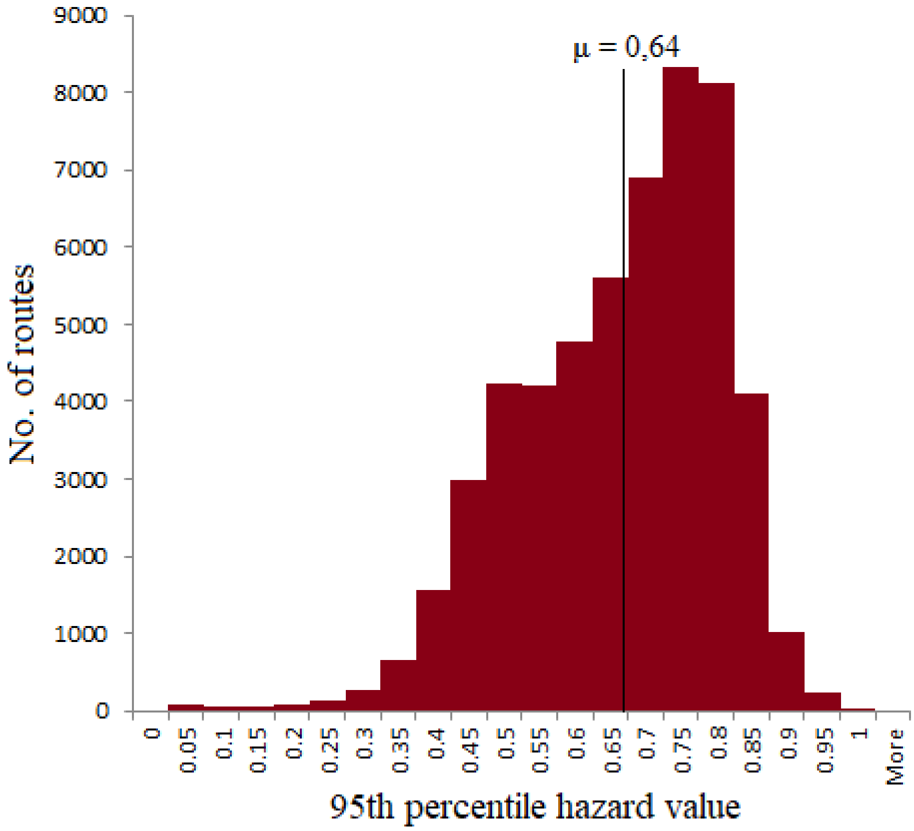


Figure 29. *histogram of 95th percentile hazard values*

The histogram above shows that almost no routes have a 95th percentile value of over 0,9 or under 0,35. Also, the distribution is skewed negatively, with the mean being less than the most occurring values. As such, there are more extreme values on the low end of the distribution than on the high end.
The mean hazard values and 95th percentile values are somewhat related, with a Pearson's r correlation coefficient of 0,21 ($p<0.01$). Again, the example of Pigne d'Arolla was used to inspect the advantages and disadvantages of this measure.

The pattern identified in figures 26 and 27 is also there for the 95th percentile values. The same two key sections seem to be important in determining whether a route has a high value. A difference is that the 95th percentile hazard values show larger differences within the same corridor. The standard deviation in this corridor for the 95th percentile values is 0,8, against 0,4 for the mean hazard values. The reason for this is that the routes are now judged based on the upper 5% of cells instead of for all the intersecting cells. As such, a small difference in route drawing has more effect. This is useful as it shows the true difference between routes in terms of how risky they are.

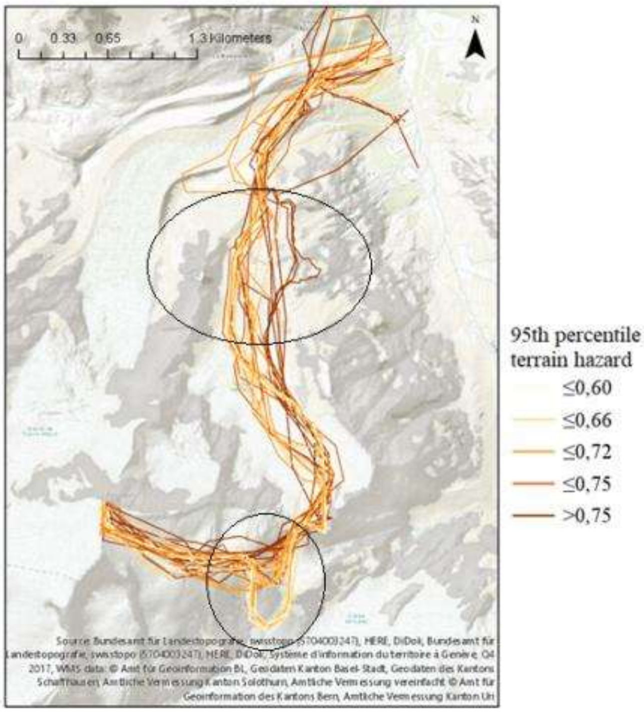


Figure 30. *single corridor of routes to Pigne d'Arolla, coloured by 95th percentile value*

In the example above, the routes that take the detour south around the steep section (lower circle) get assigned vastly different values. This is a good representation of the reality, as they manage to avoid a very steep section with high potential for avalanche triggering. Also, the long route from figure 28 now has a high value (0,80) assigned which accurately represents the fact that it intersects two areas with high terrain hazard. These are clear advantages of using 95th percentile hazard over mean terrain hazard. There is one situation though in which mean hazard values should be better, namely when analysing routes on the aggregated level instead of on individual level, treating the above routes as a single corridor. This is the case when one does not want to plan routes in detail but instead wants to know about the characteristics of a general route corridor. In that case, the high variability of 95th percentile values may make them less accurate, if one does not exactly follow individual routes.

The mean terrain hazard and 95th percentile hazard values of individual routes are useful for pairwise comparisons between individual routes. However, as was discussed in section 3.3.3, they do not give a lot of information about the characteristics of a route. In the next section, the discrete terrain classes are used to cluster the routes based on avalanche terrain. For this, the percentage of each terrain class intersected by each route is used.

4.2 Clustering based on avalanche terrain

4.2.1. Cluster attributes

Clustering based on avalanche terrain resulted in five classes of routes. The mean percentages of each terrain class in each cluster is presented in the table below.

Cluster	HT	MT	LT	VLT	50°	60°	RTH	RTM	RTL	RTVL	MR	No_av	No. of routes	% of routes
0	**8,4**	**12,8**	6,1	2,9	**6,5**	**3,0**	1,9	5,3	3,1	7,0	6,2	36,7	**860**	1,6
1	1,5	2,2	0,8	0,3	0,1	0,0	0,5	1,9	1,3	2,5	2,6	**86,3**	**16098**	30,1
2	**5,6**	**11,5**	6,5	3,8	0,9	0,1	3,0	8,1	4,4	9,3	8,1	38,7	**8223**	15,4
3	3,6	6,1	2,6	1,1	0,3	0,1	2,0	8,1	6,0	**15,5**	**15,2**	39,5	**10080**	18,8
4	3,2	5,5	2,4	1,1	0,3	0,1	1,4	5,1	3,1	7,2	7,8	62,8	**18292**	34,2
All, μ	3,2	5,7	2,7	1,3	0,4	0,1	1,5	5,1	3,3	7,7	7,6	61,4	53553	100,0
All, σ	2,4	4,0	2,5	1,8	1,2	0,5	1,4	3,5	2,3	5,8	6,3	21,2	53553	100,0

Table 5. *Mean values for the four clusters. Route length is not taken into account*

HT	High triggering potential
MT	Medium triggering potential
LT	Low triggering potential
VLT	Very low triggering potential
50°	50-60 degrees, prime risk falling
60°	>60 degrees, prime risk falling
RTH	Remote triggering, high potential
RTM	Remote triggering, medium potential
RTL	Remote triggering, low potential
RTVL	Remote triggering, very low potential
MR	Maximum runout zone
No_av	No avalanche terrain
All, μ	Mean of all routes
All, σ	Standard deviation of all routes

Table 6. *Terrain class abbreviations*

Cluster 0 is small in terms of number of routes, and contains dangerous routes. It has the highest value for high triggering potential terrain. Also, it has very high values for both of the steep terrain classes with above 50 and 60 degree slopes. This last bit is interesting as this terrain is very rarely encountered in the other routes, with mean values of 0.4 percent and 0.1 percent respectively. Cluster 1 is large, containing 34% of all routes. The main distinguishing feature of this route is the high value for non-avalanche risk terrain, while all other values are lower than the dataset mean. Therefore, these routes are generally the safest, traversing the least typical avalanche terrain of all clusters. Cluster 2 contains 15% of all routes and consists of routes that traverse quite a lot of high triggering potential terrain. Also, the routes in this cluster traverse more medium triggering potential than average. What distinguishes this cluster from cluster 0 is the fact that far less extremely steep (<50 degrees) terrain is traversed. Cluster 3 has about average values for triggering terrain. However, the remote triggering terrain is much more prevalent in this cluster than in the rest of the data. Also, the no avalanche terrain values are lower. Routes in this cluster are probably safer than the routes in cluster 0 and 2, but not as safe as cluster 1. The final cluster, cluster 4, has very similar values to the means of the whole dataset. Routes in this cluster can be considered typical or average. Based on these characteristics, the following names are given to the clusters.

Cluster no.	Cluster name	No. of routes	% of routes
0	Steep routes with extreme terrain	860	1,6
1	Very flat routes	16098	30,1
2	Steep routes	8223	15,4
3	Flat routes, many remote triggering zones	10080	18,8
4	Typical routes	18292	34,2

Table 7. *Cluster names*

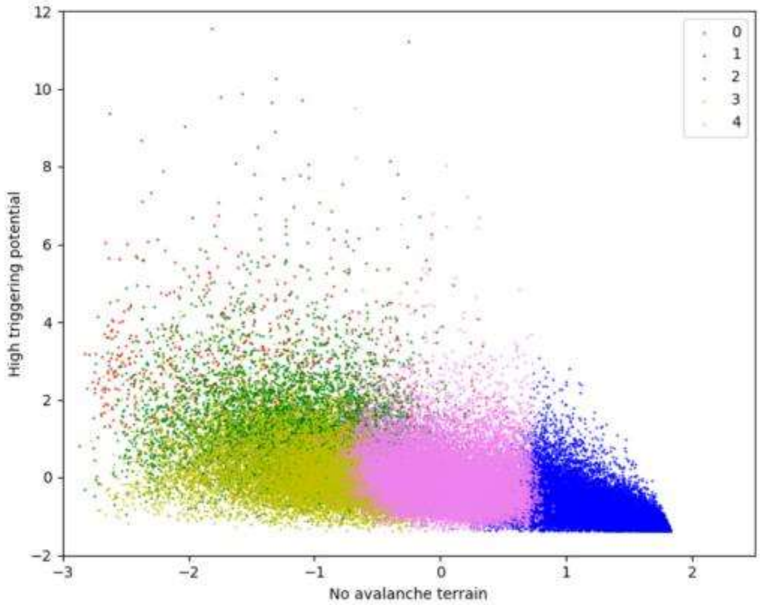


Figure 31. *The five k-means clusters plotted based on two attributes: the percentage of the route not traversing avalanche terrain and the percentage traversing the most 'high triggering potential' terrain class.*

To give an idea of the way the routes within the clusters are positioned in terms of avalanche terrain, they have been plotted in the figure above. For this, the two most telling terrain classes, namely "high triggering potential" and "no avalanche terrain", have been used. The normalized scores of each route on both these terrain classes is shown. Because of the large number of points, it is hard to see the differences in density between cluster centres and cluster edges. To show the density of points within each cluster, the points are coloured by cluster density in figure 32. Also, the cluster means have been added into the plot. This highlights the concentration of the points. It has to be noted that this is a plot of just two attributes. Some clusters may seem to overlap based on these two attributes, but are different in other attributes.

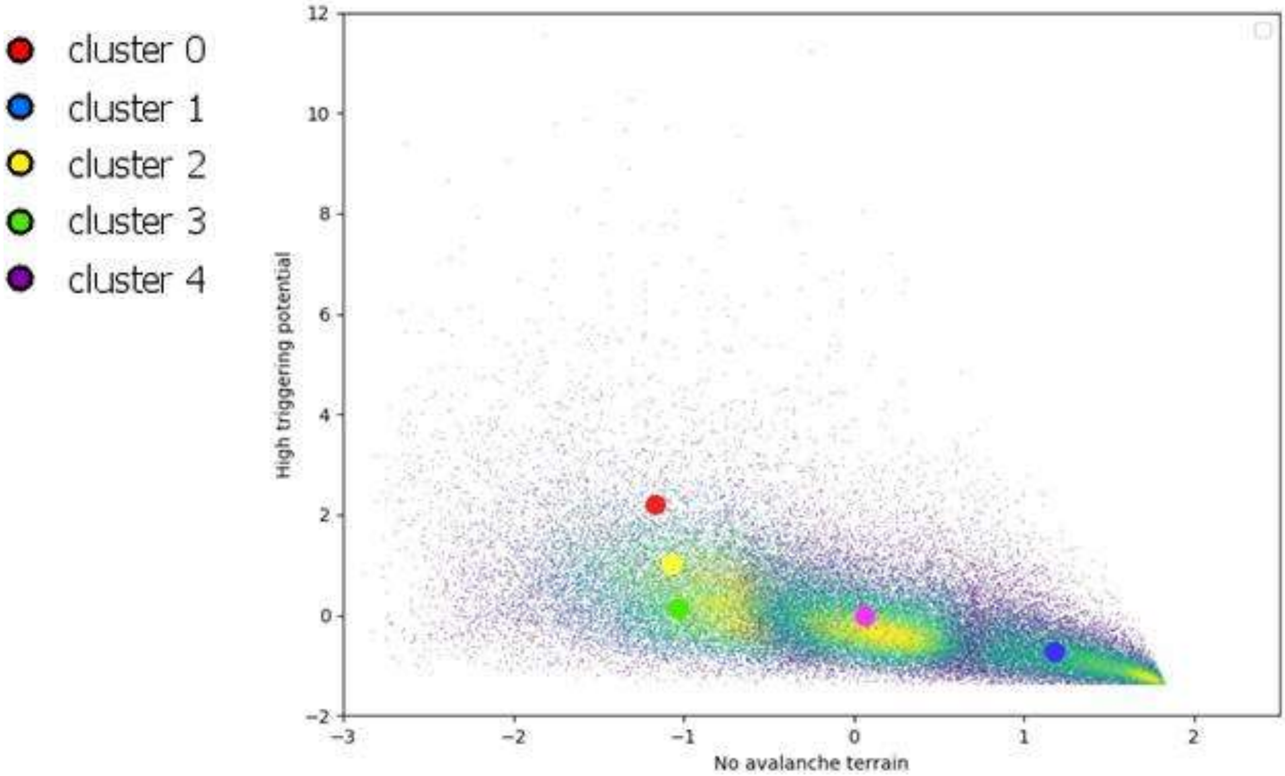


Figure 32. *Kernel density plot of terrain clusters, including cluster centres.*

The difference in density around the cluster centres of cluster 4, 2 and 1 highlight the different characteristics of those clusters. This is less obvious between cluster 0 and 3, As a next step, the routes were grouped based on cluster membership and the mean values for other route characteristics (e.g. length) were calculated to assess whether there is any pattern in this. Then, t-tests were performed to check whether these values were significantly different. For this, Welch's t-test was used, which is more accurate than student's t-test for variable sample sizes (Fagerland & Sandvik, 2009). The results of this are shown in figure 32. All bolded values are significantly different from the mean with $p < 0.05$. The values with a green background are higher than the mean, and the ones with red backgrounds are lower. As expected, the flat routes have lower mean terrain hazard values and lower mean 95th percentile values, and very steep routes and relatively steep routes have higher values then average here. What is interesting besides this, is that the two steep and relatively dangerous clusters (numbers 0 and 2) have below-average danger ratings as defined in the avalanche bulletin. This means that these routes are more often planned when the bulletin gives off a lower warning number. This is interesting as it means that people plan more dangerous routes when the conditions are more favourable. Also, the very flat routes are often planned when the warning number from the bulletin is higher than average. This is again an expected outcome, as it is expected that people take the warnings from the avalanche bulletin into account when planning routes, and adjust the terrain traversed in their routes accordingly. However, it is still interesting to see this reflected in the data. A note that has to be made here is that only about a quarter of the routes have the avalanche warning rating defined, and for the others it is not possible to assess in what conditions they were planned. Furthermore, the very steep routes (cluster 0) relatively often have no danger rating defined by the user.

Cluster no.	0	1	2	3	4	
Cluster name	Steep routes with extreme terrain	Very flat routes	Steep routes	Flat routes, many remote triggering zones	Typical routes	All routes
Length	**7008.9**	**7084.0**	**7388.6**	**7908.9**	**8089.6**	7628.3
No. of vertices	**57.0**	48.3	48.0	49.3	48.5	48.7
Step length	210.4	**209.3**	**222.5**	217.2	**227.2**	219.0
Min. height	**1814.9**	**1407.4**	**1633.1**	**1715.5**	**1575.3**	1563.9
Max. height	**3001.1**	**2301.4**	**2816.8**	**2827.4**	**2710.8**	2630.6
Mean height	**2366.2**	**1829.7**	**2185.5**	**2215.7**	**2103.4**	2059.0
Height difference	**1186.2**	**894.0**	**1183.7**	**1111.9**	**1135.5**	1066.7
95th percentile terrain hazard	**0.686**	**0.377**	**0.676**	**0.571**	**0.561**	0.527
Mean terrain hazard	**0.347**	**0.296**	**0.337**	**0.301**	**0.306**	0.308
Danger rating	**2.12**	**2.49**	**2.19**	**2.23**	2.33	2.34
No. with danger rating specified	119	4758	2013	2538	5073	14502
% with danger rating specified	13.8	29.6	24.5	25.2	27.7	27.1

Table 8. *Mean values for route attributes per cluster (green = significantly below population mean, red = significantly above population mean)*

What's interesting besides this, is that the very flat routes (cluster 1) have far lower values for route height than the mean. This suggests that the more mellow terrain traversed by these routes is more often found on lower slopes. Also, the very steep routes have by far the highest values for height, with an average max. height of 3001 meters. This suggests that these routes are often planned in high alpine terrain, where many of the extremely steep slopes that are typical for this cluster are found. Finally, the step length of routes gives an interesting picture. The very flat routes have the shortest step length, suggesting these routes are planned most carefully. This fits with the idea that these routes are the safest category. However, cluster 4, the typical routes, have the biggest step length. This is not expected, since these routes are not the most unsafe. Apparently, the relation between careful route-drawing and avoidance of avalanche terrain is not straightforward.

4.2.2. Cluster locations

Of the routes in cluster 0 and 2, relatively many are located in the south-western corner of Switzerland. This is where many of the high alpine routes can be found. In contrast, the routes in cluster 1, the very flat routes, are more often located in the Northern part of the Swiss alps. Here, terrain with less altitude and mellower slopes can be found. However, there is not a strong spatial trend in the terrain clusters. Most areas have some routes in each of the terrain clusters, which is why the route centroids are not very far apart.

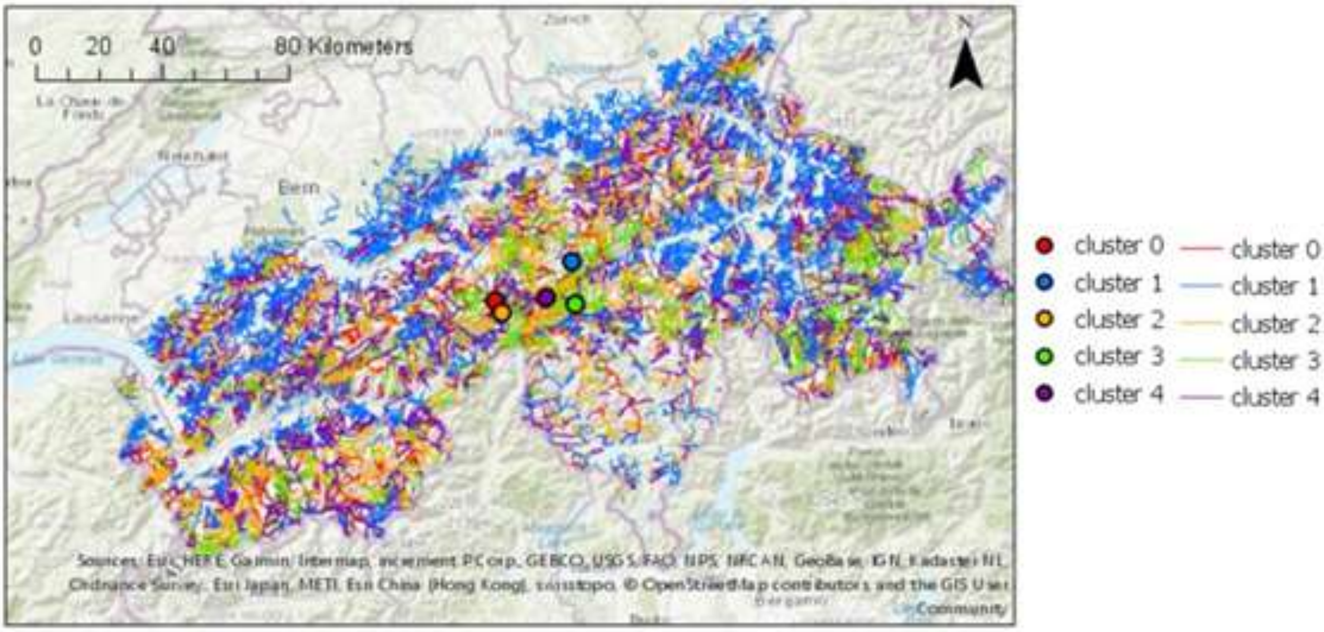


Figure 33. *Cluster centroids and all routes coloured by their terrain cluster label*

4.3 Spatial clusters

Using the DBSCAN clustering algorithm with the parameters defined in chapter 3.3.4, 655 spatial clusters were found. 18225 (34%) of the 53553 routes were assigned to a cluster, with the rest being labelled as noise. The mean number of routes per spatial cluster is 28,9, the median is 19, and the minimum and maximum are 10 and 181 respectively. All route clusters are shown in figure 34. They have been assigned random colours.

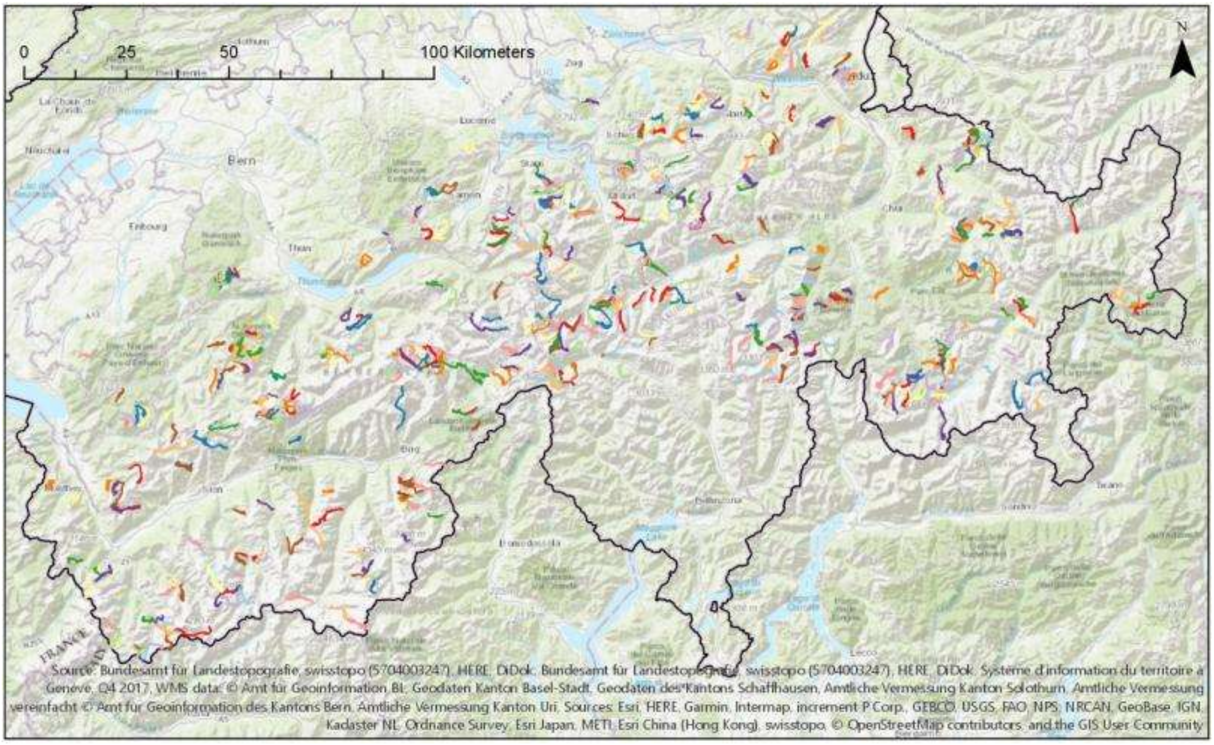


Figure 34. *All spatial clusters in the planned route data*

Clusters usually start either in towns or at mountain refuges, and usually lead to mountain peaks or other mountain refuges. There are areas where many routes follow the same terrain corridor, but still not cluster is found. This happens when routes follow a corridor but then disperse after, for example when some of the routes end at a peak while others traverse it.

As was mentioned in chapter 3, it is hard to find a clustering algorithm and parameters that perfectly clusters all route data without false positives or false negatives. As such, there are some cases where routes are divided into separate clusters while in fact they should have been one, and vice versa. This is shown on the figure below. The example on the left shows five route clusters that have been correctly assigned to separate clusters. They all have the same start points, but are leading to different end points. The yellow and brown clusters lead to the same end point, but take a different route to get there. One of the brown routes has been incorrectly assigned as such, and should have been yellow instead. This is a downside of the DBScan clustering algorithm: when a point can be included in two different clusters, the cluster it is assigned to is dependent on the order the data is traversed. This particular route probably can be included in both clusters, but was traversed first by the brown cluster and as such joined that one.

The right figure shows a set of routes where the clustering algorithm didn't work that well. When inspecting these routes by eye, it is clear that the brown routes on the west form a separate cluster from the rest, as they go over a ridge whereas the rest go through an open field. The algorithm however does not take this context into account, and as such assigns them to the same cluster. The green cluster has more routes on the eastern side, but both clusters have a number of routes going through the middle.

Ideally, all routes here would have been assigned either to one cluster or to two clearly defined clusters. With a smaller epsilon, the clusters would be more clearly defined, and the routes in the middle would perhaps not be joined to the brown cluster. With a larger epsilon, the two clusters would have been joined. Whether or not routes are “correctly” clustered seems partly to depend on the terrain. In cases where there is a uniform start and end point, the clustering algorithm performs better. In the case such as the right figure, the strand and end points are dispersed, which makes clustering messier. The way to solve this would be to change clustering parameters, but this would lead to poor results in locations with more spatially homogeneous routes. As a test the epsilon was increased, and this did result in the two clusters merging. However, it also ended up assigning to them merging with another cluster nearby, which makes them even worse as a representation of planned routes. Cases such as the right figure are however quite rare in the results. There are situations where seemingly uniform routes get assigned to different clusters, but usually this has a reason and they are actually different routes in reality, for example having different start points. Still, it is important to keep in mind when further analysing the clusters that they are partly a result of the parameters chosen.

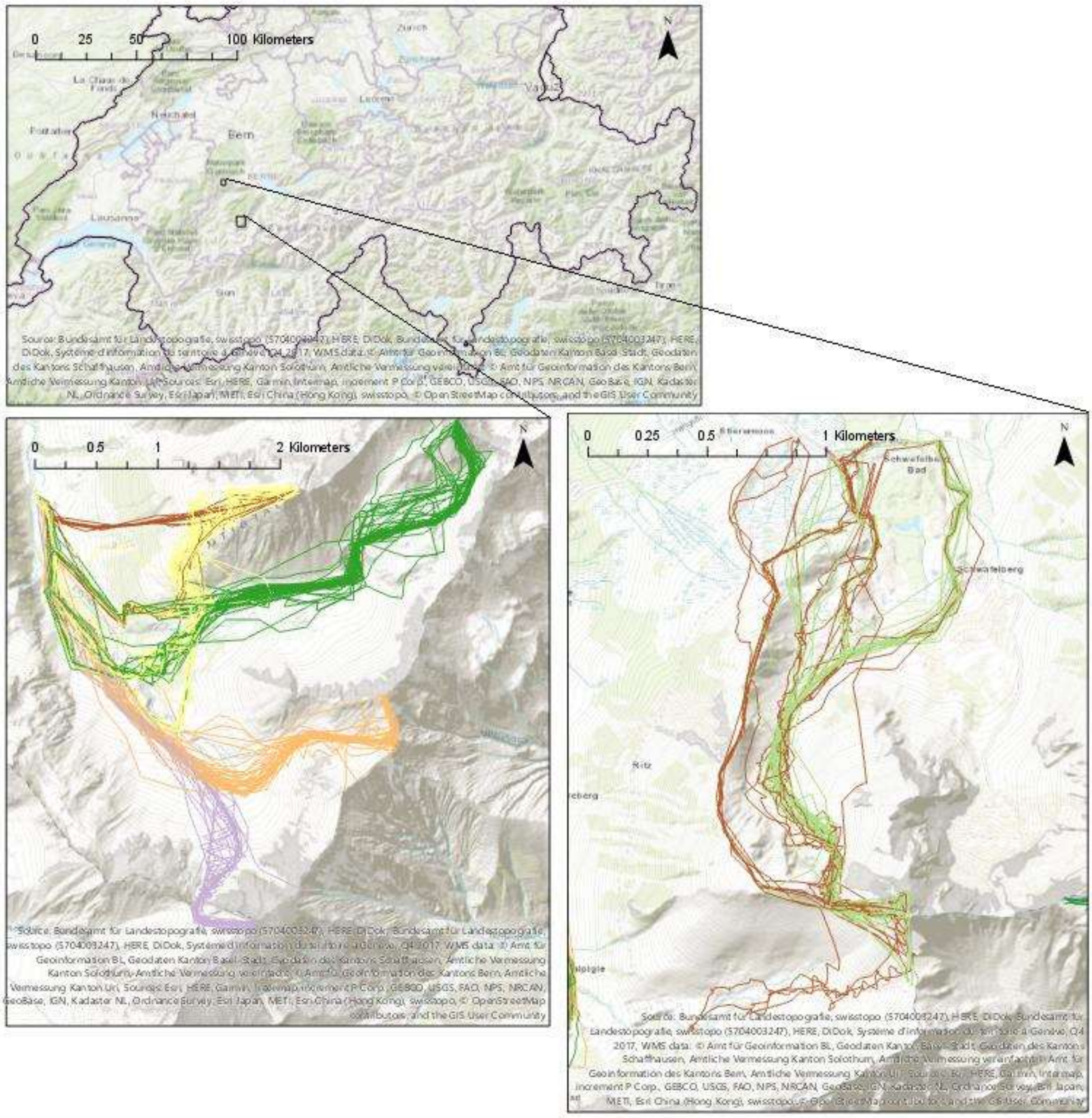


Figure 35. *two examples where clustering worked relatively well (left) and poorly (right)*

4.4 Relation between spatial clusters and avalanche terrain

The next step was to investigate how much the avalanche terrain attributes in routes are coherent within spatial clusters. To do this, the 95th percentile values from Q1 and the terrain clusters from Q2 are used.

4.4.1. Terrain cluster membership

To see how varied spatial clusters are in terms of terrain cluster membership, the count of the most occurring terrain cluster was divided by the total number of routes in for each terrain cluster. This results in a value between 0.2 and 1, where 0.2 is a completely heterogeneous spatial cluster (with 20% of routes belonging to each of the five clusters respectively). 1 means all routes in a spatial cluster are part of the same terrain cluster. The mean for this value is 0.80, the median 0.83, the minimum 0.333 and the maximum 1. The division of this value is shown in figure 36.

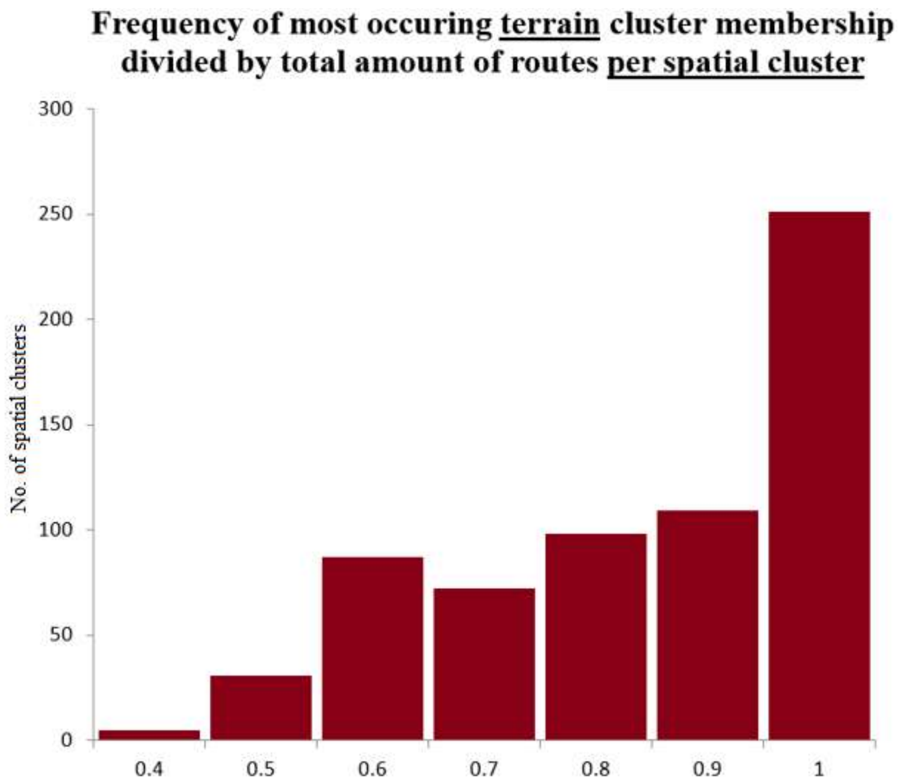


Figure 36. *Variability of terrain clusters within spatial clusters (closer to 1 means more uniform)*

As can be seen, many spatial clusters (252) have a value of over 0.9, meaning 90% of the routes in that spatial cluster have been assigned to the same terrain cluster. This supports the hypothesis that spatial cluster membership and terrain cluster membership are related. There is no trend between the number of routes in a cluster and the variability of terrain cluster membership within that cluster.

4.4.2. 95th percentile values

A next step is to evaluate the variability in terms of 95th percentile values within spatial clusters. As noted, an assumption is that routes that are similar spatially have similar 95th percentile values. To evaluate this, the standard deviation of 95th percentile values within spatial clusters was calculated. However, using standard deviation as a degree of variability is mostly meaningful when the data is normally distributed. Therefore, for each spatial cluster the D’Agostino k-squared measure was calculated for the 95th percentile values (D’Agostino, Belanger, D’Agostino jr., 1990). For 560 of the 652 spatial clusters, the 95th percentile values were normally distributed. Of those 560 clusters, the standard deviation was calculated. The mean 95th percentile terrain hazard value for those clusters was 62,73, and the standard deviation was 6,97. To compare: for the whole dataset the mean is 63,68 and the standard deviation 14,04, twice as high as within spatial clusters. This supports the assumption that routes within a spatial cluster are similar in terms of avalanche terrain. However, between the spatial clusters, the degree of similarity varies. In figure 37, a histogram is shown of the standard deviation in terms of 95th percentile terrain hazard within clusters. Only the clusters where this value is normally distributed are shown.

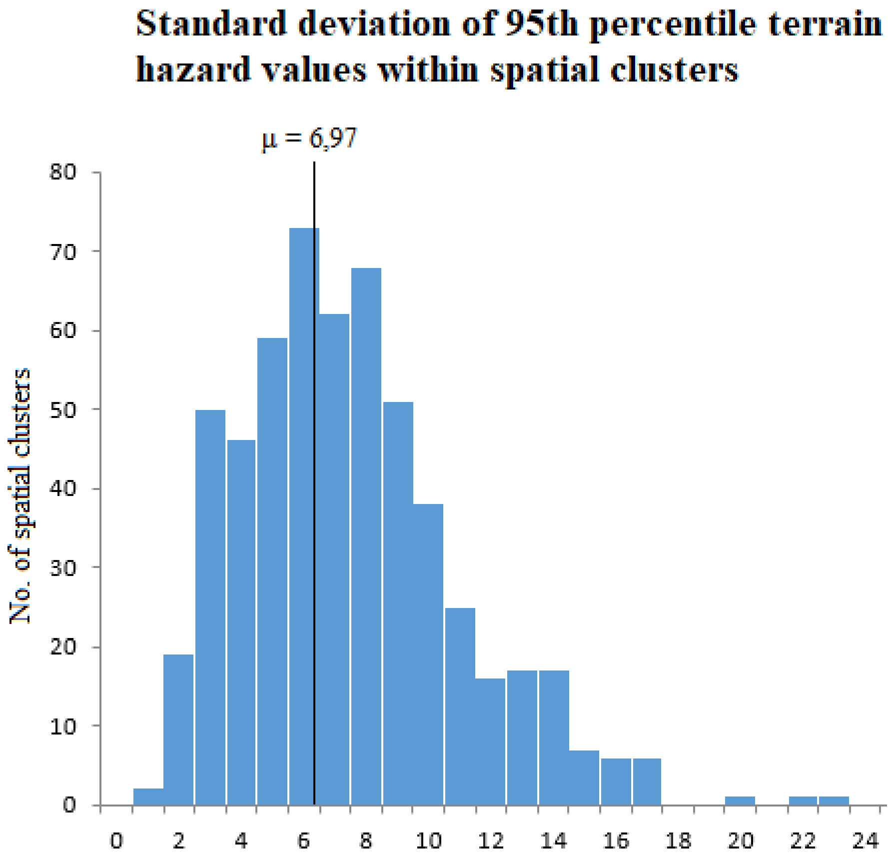


Figure 37. *histogram of SD for 95th percentile values within spatial clusters*

Perhaps an expectation would be at this point that within a spatial cluster, routes that with shorter step length possess lower 95th percentile values, as they are drawn in more detail and thus better able to avoid dangerous sections. However, this was only to be true found in a limited part of the data. Using Pearson's r, only in 97 of the spatial clusters a significant relationship ($p<0,05$) was found between step length and 95th percentile hazard. In those 97, the correlation coefficient was 0,36, which does support the idea that longer step length leads to higher 95th percentile values.

4.4.3. Relation between variability and route attributes

The degree of variability within a route cluster can be seen as a characteristic of the routes within that cluster. Therefore, it is interesting to see whether this variability has any relation to the other attributes within routes. To do this, Pearson's r was calculated as a measure of correlation between the cluster variability of routes and other route attributes. The results of this are shown in table 9. Only the results that are significant with $p<0,05$ are included.

Route length	-0.27
No. of vertices	-0.28
Mean step length	0.08
Min. step length	0.24
Max. step length	-0.10
Max. height	-0.22
Min. height	-0.11
Height difference	-0.33
Max. slope	-0.34
Straightness	-0.17

Table 9. *Pearson's r correlation coefficients for correlation between standard deviation for 95th percentile values in spatial clusters and mean route attributes within clusters*

There are no very strong relations between the route attributes and variability within a cluster. It seems that shorter route cluster with less height difference tend to show more variance in terms of terrain hazard. Possibly, because of the short length, small variations in route planning have relatively more influence. One would perhaps expect the step length in a route cluster to be of influence on the variability. Routes with longer step lengths tend to be planned more coarsely and thus may more often be different from the other routes within their spatial cluster. However, this assumption is not really reflected in the data, only for the minimum step length, and there the relation is not very strong.

No. of vertices	0.08
Max. height	-0.13
Mean height	-0.13
Height difference	-0.12
Straightness	-0.15

Table 10. *Pearson's r correlation coefficients for correlation between terrain cluster membership variety in spatial clusters and mean route attributes within clusters*

In table 10, the same is done for the terrain cluster membership variety. In general, the trends here are similar, although the relationships are even weaker and there are less significant relationships.

4.4.4. Visual inspection – 95th percentile values

The previous section has shown that avalanche terrain is often similar for routes within a spatial cluster. However, figures 36 and 37 show that there are clusters where this is not the case, or to a lesser degree. In the following section, some of these clusters will be discussed. This serves as an exploration into the patterns in the route planning. Given the large number of routes it is impossible to visually inspect them all. Therefore, care should be taken when generalizing the findings to all the routes.

Below is the histogram of 95th percentile values in spatial cluster 9512, where the standard deviation is far higher than average. The mean value for 95th percentile hazard in this cluster is 54,33, and the standard deviation 13,84, and 40 routes are part of this cluster.

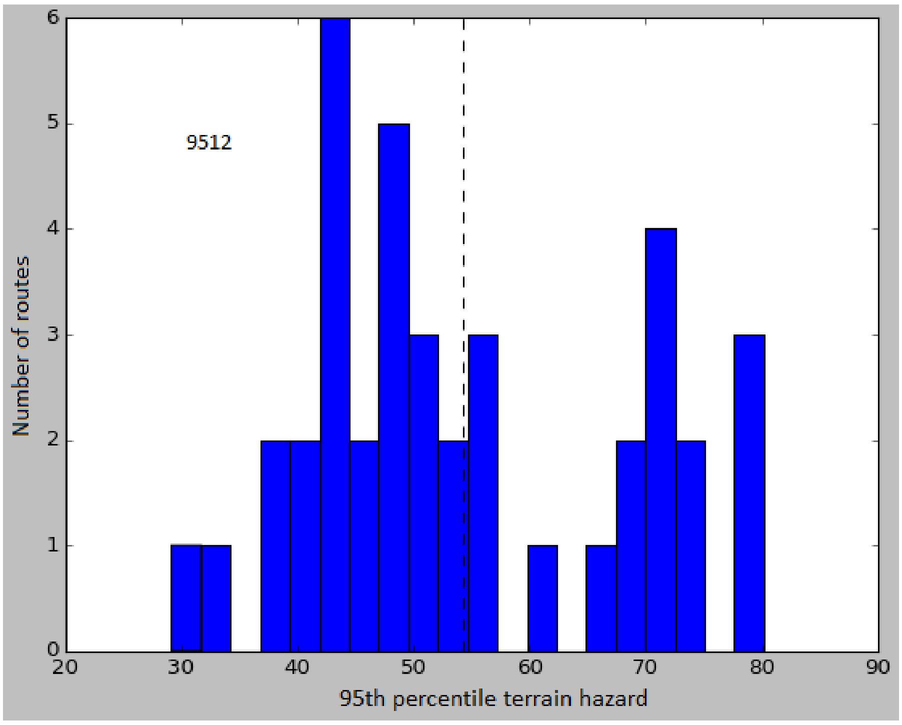


Figure 38. *Histogram of 95th percentile hazard values within spatial cluster 9512*

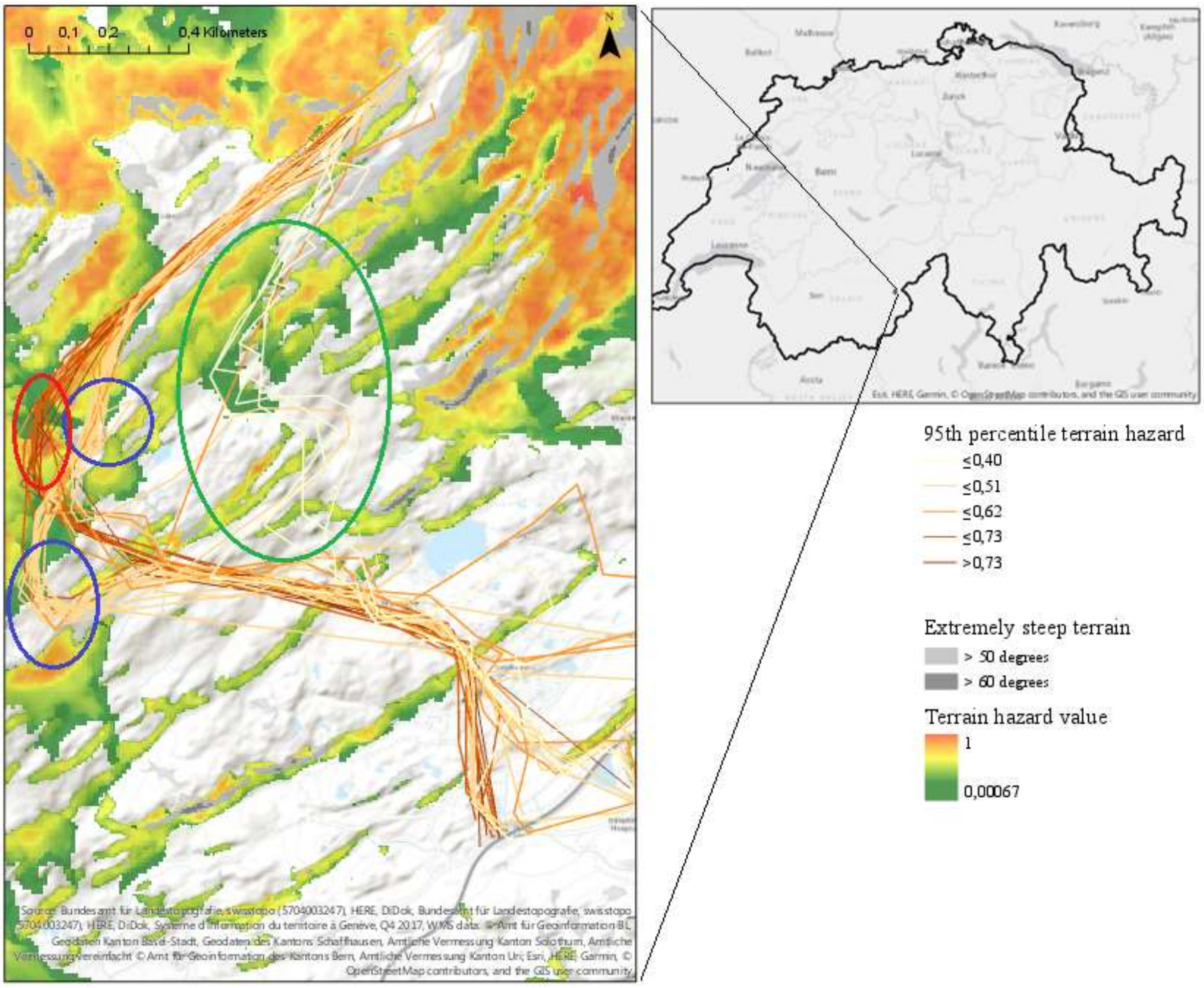


Figure 39. *Spatial cluster near Brig, coloured by 95th percentile hazard value*

When visualizing this route cluster, the reason for the high variability manifests itself. Three variants within this spatial cluster can be identified. The first one, shown with a green circle, encompasses only a few routes and avoids some of the hazardous terrain by taking a shortcut. The red-encircled route takes a longer way but actually end up going straight through an area with highly hazardous terrain, shown in red in the background. The blue-encircled routes follow a similar trajectory but avoid this dangerous section and as such have lower 95th percentile values. Why are these three variants visible in the data? It seems they follow three approaches when drawing routes. The red circle routes follow the SAC ski-touring route. This is visible as a map layer on Whiterisk. The blue circle follows a regular hiking trail, which is also visible on the Whiterisk map. The green circle does not follow any existing trail, but instead seems to take the safest route by avoiding the steepest slope angles on Whiterisk. The three approaches are shown in figure 40. This example shows how different ways of route-planning lead to different amounts of hazardous terrain encountered within a route.

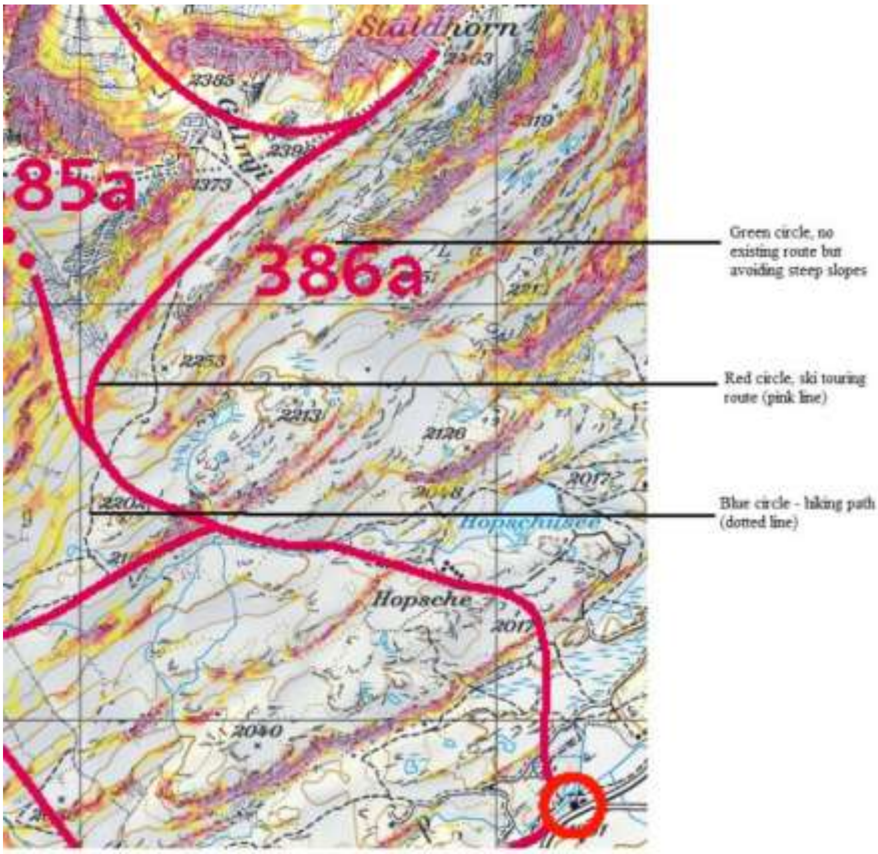


Figure 40. *The three ways routes are planned in this case based on the Whiterisk visualization*

4.4.5. Visual inspection – terrain clusters

The spatial clusters also vary in terms of terrain cluster membership. In the following section, four sets of spatial clusters are inspected to highlight cases where this variety is very high or very low.

Two uniform spatial clusters near Pigne d'Arolla

The following two spatial clusters are uniform in terms of terrain cluster membership. They are respectively majority cluster 3 and cluster 4. These terrain clusters are quite similar, with both being not extremely dangerous but also not extremely safe. The main difference is that cluster 3 traverses far more remote triggering areas, whereas cluster 4 traverses more non-avalanche terrain. The two spatial clusters are clearly visible in figure 41, one leading to Pigne d'Arolla (west) and the other to Cabane de Bertol (east). The variety measure for the clusters in terms of terrain cluster membership is 0,73 and 0,89 respectively. When looking at the terrain surrounding both route clusters, it becomes clear why the routes in the respective spatial clusters are assigned to the different terrain clusters.

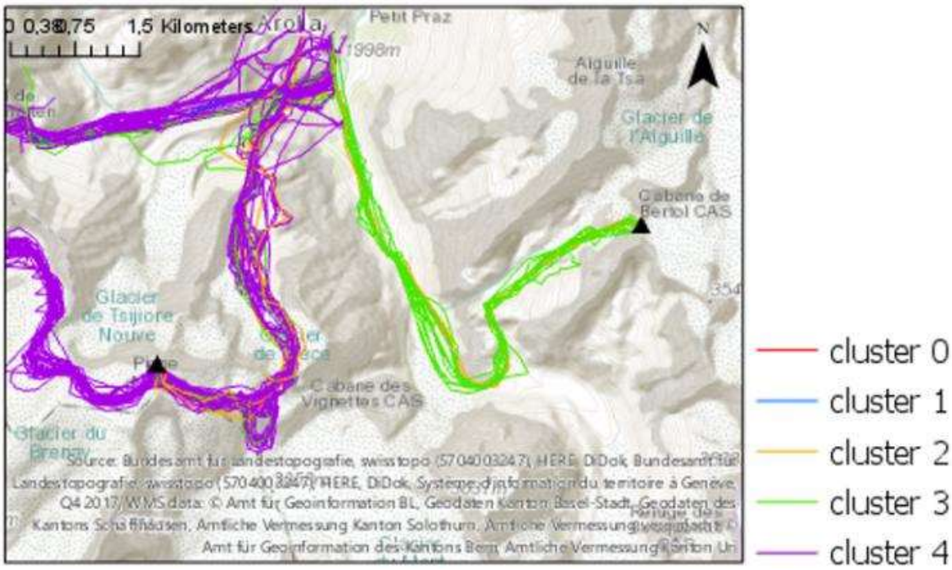


Figure 41. *two spatial route clusters with majority in semantic cluster 3 or 4*

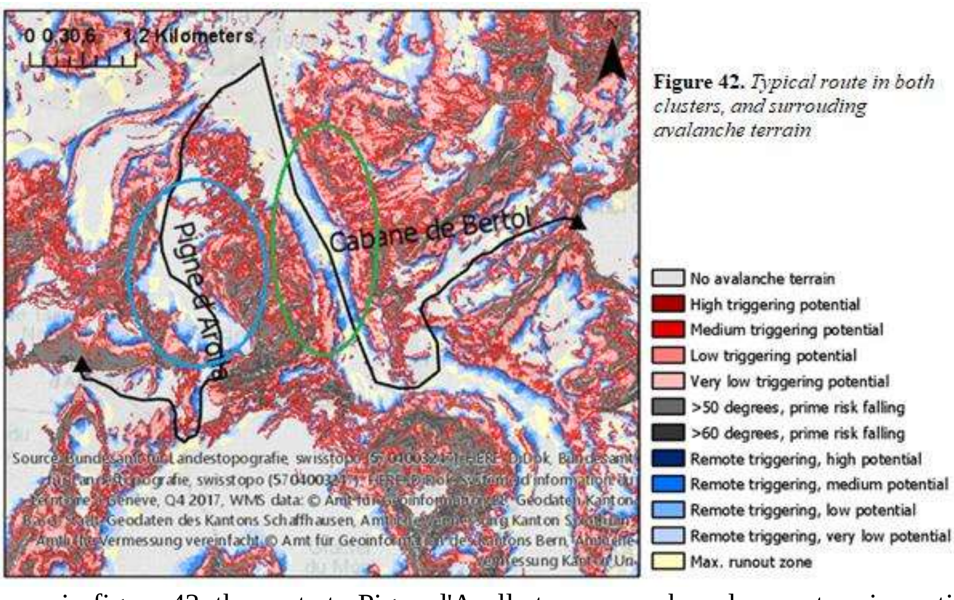


Figure 42. *Typical route in both clusters, and surrouding avalanche terrain*

As can be seen in figure 43, the route to Pigne d'Arolla traverses a broad, open terrain section, marked with a blue circle. The Cabane de Bertol route, on the other hand, goes through a far narrower section of terrain, marked by the green circle. Then, both routes make a turn and go up to their destination. Here, the same pattern can be seen: the Pigne d'Arolla route traverses mostly grey, relatively flat terrain. The

Cabane de Bertol route does so as well, but again the corridor it follows is narrower, and therefore more subject to avalanche runouts or remote triggering. When viewing the terrain of both routes in google earth, this difference is clearly visible. The green routes have steep slopes on both sides for the first section, before going into a more open terrain. The purple routes on the other hand are consistently on open terrain, without any steep slopes above them. In this case, it is clear that the terrain faced on the way to a destination is determent to the cluster a route belongs to. In other words, when planning a route between those two points, one always faces the same terrain. Therefore, the majority in both of the route corridors is assigned to a single terrain cluster. The standard deviation for 95th percentile values in the respective clusters is 7,98 and 6,50.

Figure 44. *Both spatial clusters, viewed in google earth imagery. (source: google earth pro)*

Two uniform spatial clusters in Graubünden

Another example of this are the two spatial clusters in figure 46. These are two sets of routes from Cresta in Graubünden. The south-directed set of routes are almost all in the blue cluster (cluster 1: very flat routes), with a few in cluster 4: typical routes. The set of routes in the northern direction on the other hand are mostly in clusters 2 and 3: steep routes and relatively flat routes with many runout zones. Again, this is a consequence of the terrain faced. As can be seen, the blue routes are mostly traversing a broad, grey slope, meaning an area with mostly no avalanche terrain. The orange routes on the other hand have to cross many sections with high triggering potential terrain. Where possible, they manage to choose the grey terrain, but this is not always possible. For example, the right zoomed in view shows the terrain they have to pass before reaching the peak which is their destination (the Piz Platta, 3392 m.a.s.l.). The blue routes, on the other hand, are also faced by some steep terrain sections, but always have safe options nearby, allowing them to go around the dangerous sections.

When visualizing both routes in 3d, this difference is more clear. The blue routes have wide and open area on their way to the top, which allows them to avoid steep sections. The orange routes, especially near the top, have no options of avoiding very steep terrain. Again, the terrain that is faced by routes between their start and end points is crucial in deciding to which terrain cluster they are assigned.

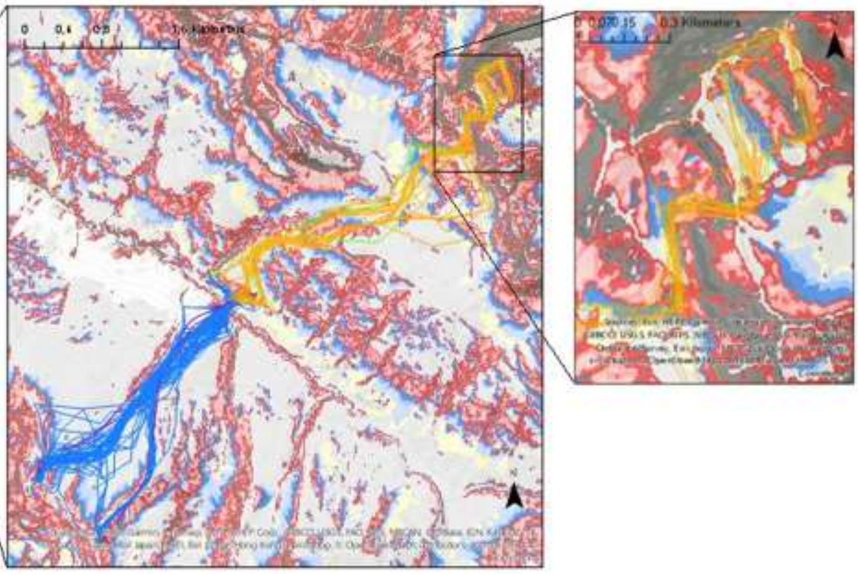


Figure 45. *Two spatial clusters belonging to different terrain clusters*

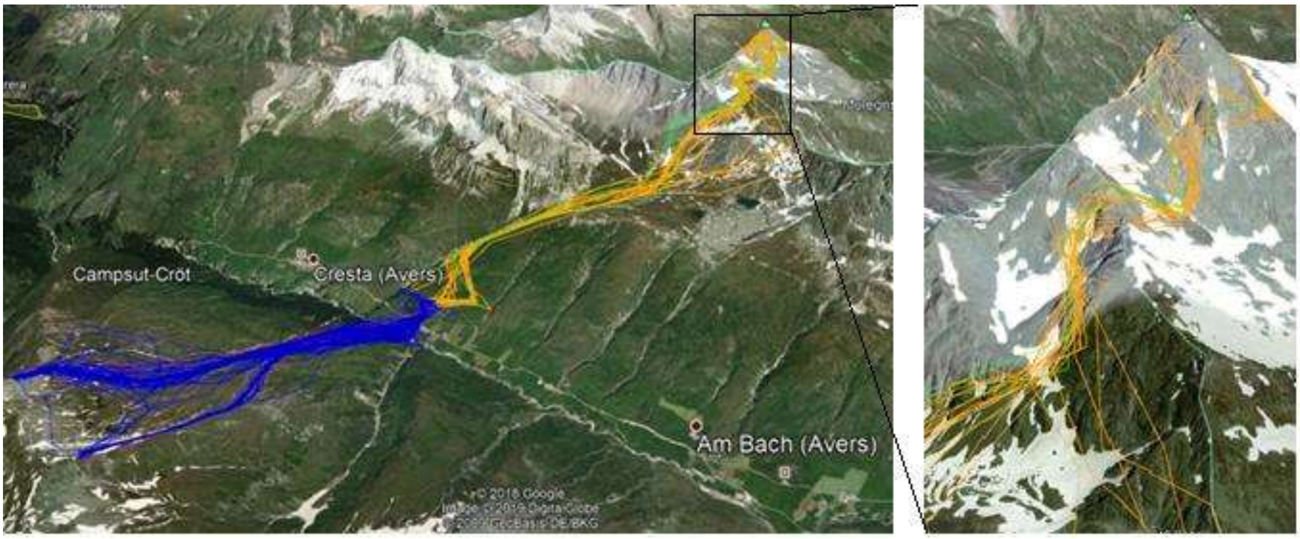


Figure 46. *Both sets of routes seen in google earth*

A spatial cluster with three possible variants, all belonging to different terrain clusters
However, in other cases, this dependency on terrain is not so clear. Many spatial clusters have a mix of different terrain clusters. An example of this is given below. This is a spatial cluster of 64 routes that all start in the town of Schwenden im Diemtigtal and go up south to slightly different destinations.
They are highly varied in terms of cluster membership, with a variety measure of 0.38 For the first half, the routes follow the same trajectory, but then they divert into three branches. These three branches get assigned to different terrain clusters. Why this happens becomes clear when viewing the routes in google earth imagery. From the left image it is clear that the routes follow the same trajectory for the majority of the distance. However, when they are near the top, they branch out into three different areas, leading to slightly different peaks. The distance between each of the goals is only 400 meters. Through these different route choices, the routes get assigned to different terrain clusters. The routes on the right take the steepest path, which is steep going up, and also has steep slopes to the sides. Therefore, the routes here get assigned to the clusters "relatively steep" and "relatively flat, many remote runout zones". The routes on the left aim for a lower, more mellow goal. Therefore, they manage to avoid most of the steep terrain, and consequently are assigned to the cluster "very flat routes". The routes in the middle take something of a middle ground and are thus assigned to "typical routes".

This shows how routes that are quite similar spatially, can still be different in terms of the terrain they encounter, by planning differently the key passages. An implication of this could be that somebody who wants to do a ski tour here looks at the avalanche bulletin of the day and decides which variant he chooses consequently. This is also clear from the figure below, where the routes have been coloured by their 95th percentile terrain hazard values. There is a clear pattern to be seen here, with the routes that lead to the high peak being more dangerous than the ones leading to the lower destination on the left. This example shows that in some spatial clusters there is a high variety of terrain options. It has to be noted however that these routes don't lead to the exact same end destination. Thus, although being defined as a single spatial cluster, the routes are not exactly similar.

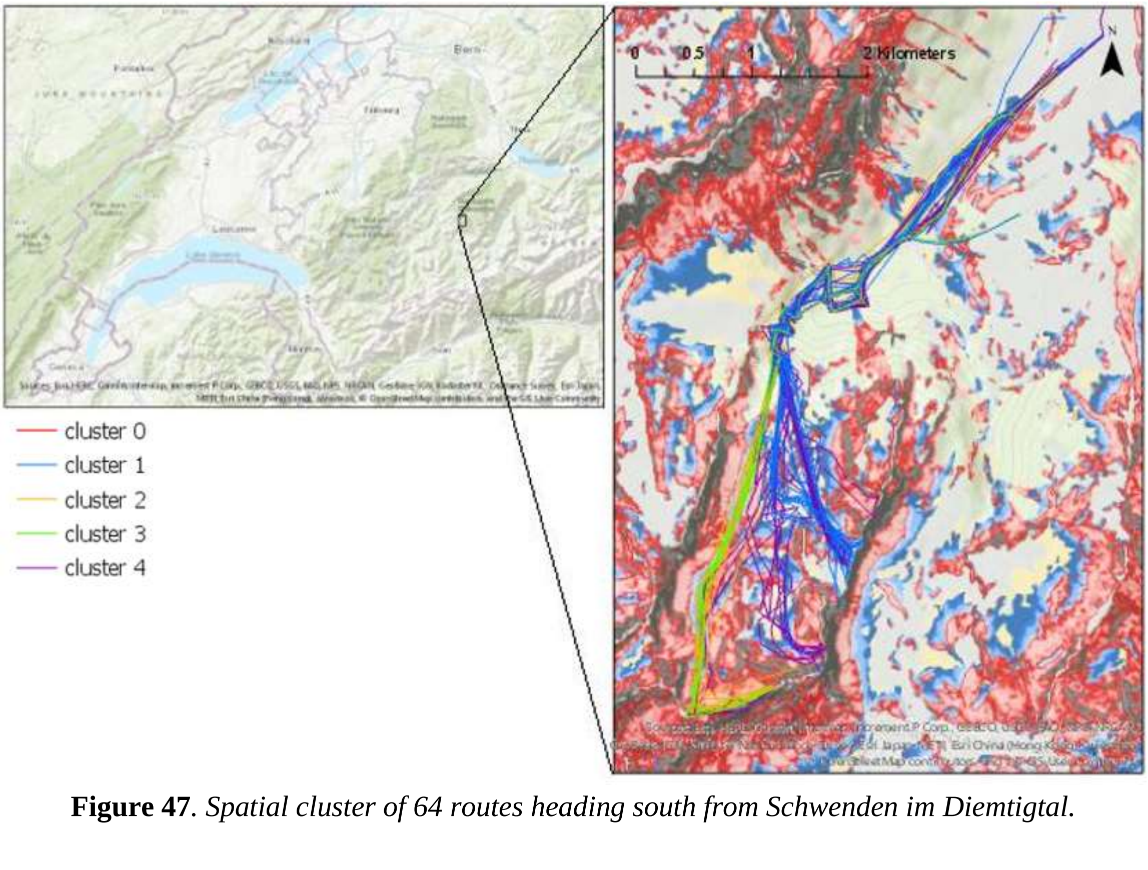
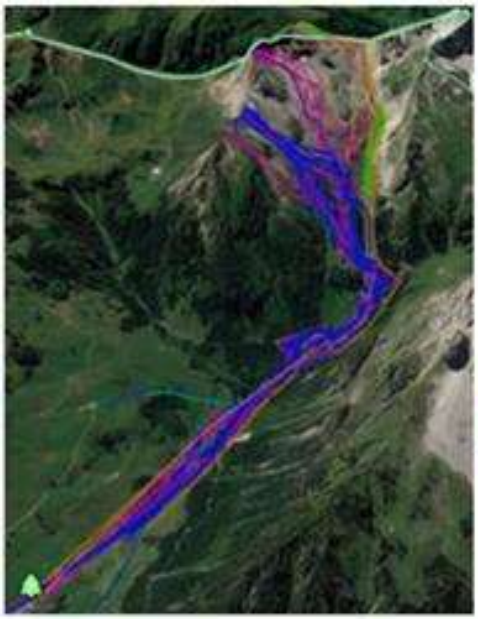


Figure 47. *Spatial cluster of 64 routes heading south from Schwenden im Diemtigtal.*



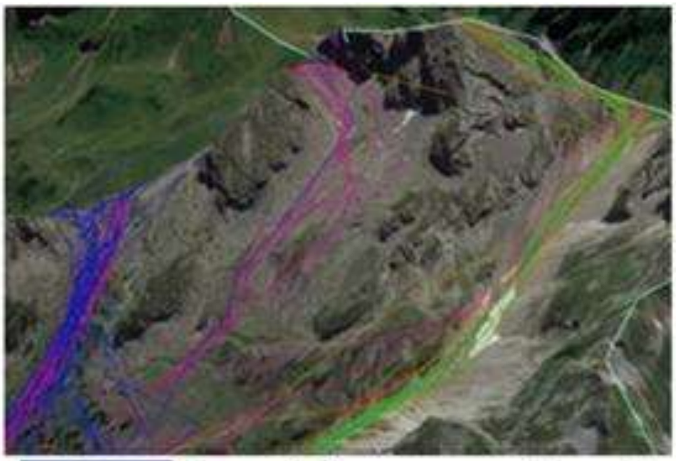


Figure 48. *Imagery view of spatial cluster near Schwenden im Diemtigtal (viewer facing South)*

Figure 49. *Imagery view of spatial cluster (viewer facing South) coloured by 95th percentile hazard*

Although the three variants are not exactly similar spatially, this example does show that minor alterations in route destination can have large effects of the type of avalanche terrain a route meets. The example below shows how routes leading to the same destination can have differing avalanche terrain characteristics by small variations in the path followed between start and end point.

Spatial cluster near Alt. St. Johann

This route cluster is highly varied in terms of cluster membership, with a variety measure of 0,41. These routes start at a lift station in the Toggenburg ski area in north-eastern Switzerland. In general, three varieties of routes can be identified. The eastern routes are in cluster 2 and 3: slightly steep or many runout zones. This is because they are planned underneath a very steep ridge, with a few sections where they traverse steep terrain. The western routes are part of cluster 4: typical routes. These routes traverse less triggering potential terrain than the western routes, but are still not traversing enough “no avalanche terrain” to be part of cluster 1. They are still crossing steep terrain in the last part of the route. The blue routes in the middle go up through an open field, with mellow slopes. There are also no steep slopes around them. Therefore, they are part of the safest cluster, cluster 1. The routes in the center also have the lowest 95th percentile terrain hazard value within this spatial cluster. When looking at the terrain here, it seems clear that the middle corridor is the gentlest slope going up. The reason why the majority of routes follow the eastern corridor, is probably the fact that this is where the official SAC route runs. Again, exactly following this line does not lead to the safest route.

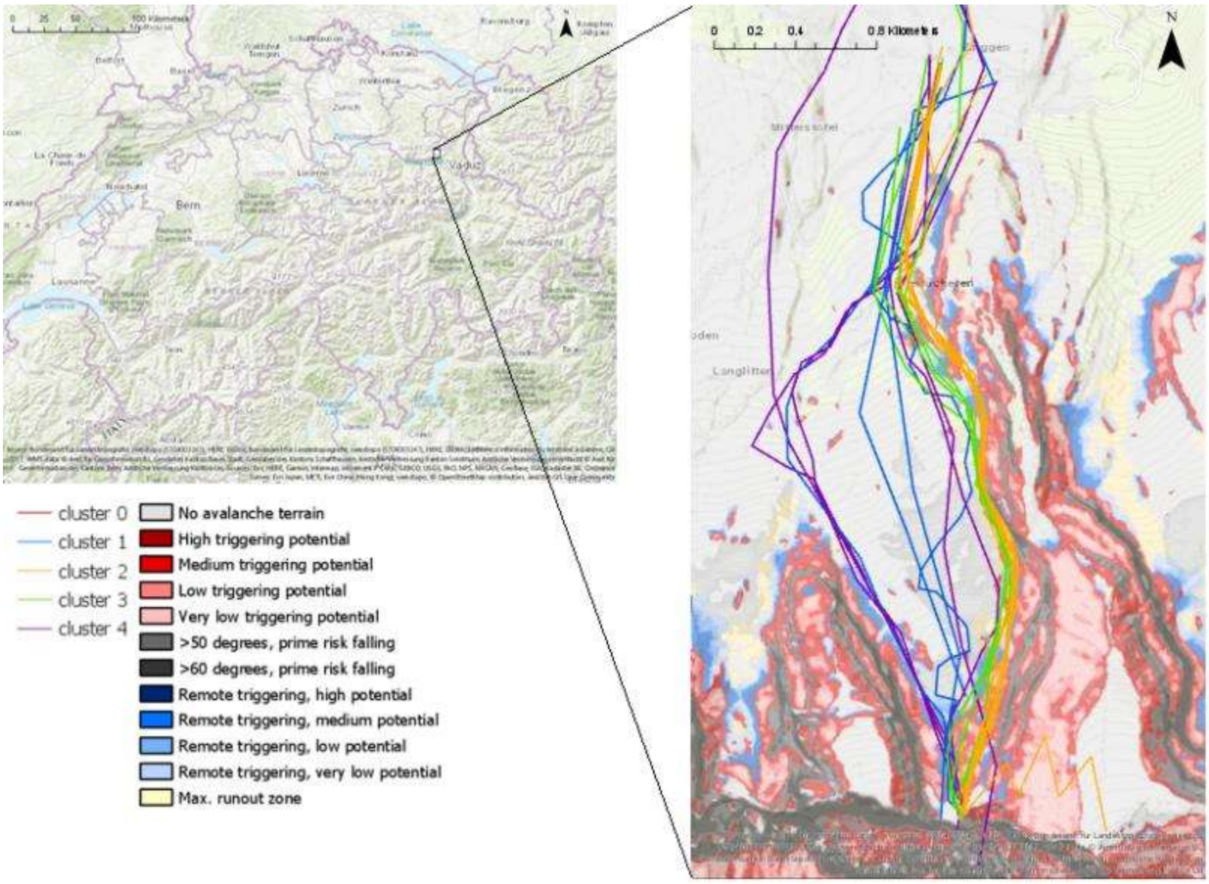


Figure 50. *spatial cluster near Alt St. Johann.*

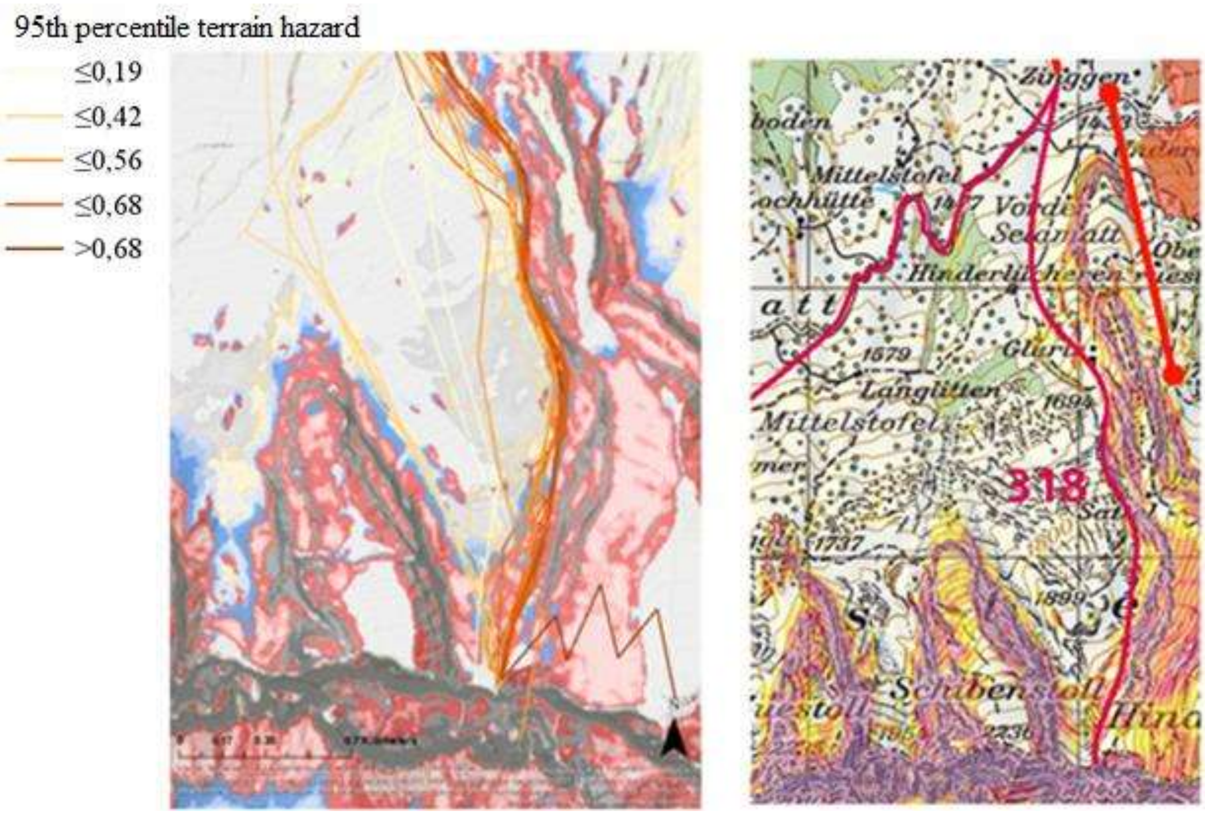


Figure 51. *Zoomed in view of spatial cluster near Alt St. Johann, coloured by 95th percentile terrain hazard values (left), view of route in Whiterisk website (right)*

The zoomed-in view of the upper section of these routes shows clearly how the decision to strictly follow the line of the SAC routes leads to higher 95th percentile hazard values. Namely, this line goes directly parallel to a steep slope with remote triggering and triggering zones. It is quite interesting that, like in the example in figure 40, directly following the SAC route actually leads to less safe routes than planning an alternative route. The Whiterisk website mentions that SAC routes should only be used as general guidelines, and aren't meant to be followed precisely. However, in reality the data reflects that they are followed precisely by a large proportion of users.

4.4.5. Discussion

The previous four examples have shown how avalanche terrain in routes can be spread out or similar within a spatial cluster. In the first two cases, the terrain faced by a route is determent. The routes are constrained in the degree in which they can avoid avalanche terrain classes, since they cannot reach the intended goal without passing through the same terrain types. In the last two cases, on the other hand, there are possibilities to change the avalanche terrain signature of the route through careful planning. This degree of variability within a spatial cluster has implications for planning. Route clusters with high variability can be more suitable for avalanche training courses, as planners can make more meaningful decisions. Also, if there is a safe variant within those clusters, they can be suited for experienced route planners if the conditions are unstable. On the other hand, less experienced route planners could be better off sticking to safe route clusters without many planning options, as those will not allow them to make route-finding mistakes easily. Unsafe corridors with little variety should only be traversed in stable conditions, because here it is impossible to avoid the dangerous sections.

Although these results are interesting, there are some limitations. First, the routes are mere plans, and it is unknown to what degree they are exactly followed by users. In the last case for example, many people followed the relatively unsafe SAC line, but it is well possible that once they arrived at the location they saw that the more westerly route had more mellow terrain and followed that. In any case, it is important to remember that many people choose to follow skin tracks, and thus may stray from their planned route. Also, the degree of variability is subject to the parameters used in spatial clustering. In the third case for example, there are three variants to the route that have slightly different end destination. These destinations are about 400 to 500 meters apart. Given the fact that the maximum distance used in the spatial clustering was 600 meters, the routes were assigned to the same spatial cluster. With stricter parameters, they would be three different clusters, all with a low degree of variability. Converesly, the same goes for case one and two, where more lenient clustering parameters would have included more routes and perhaps a higher degree of variability.

Aside from the specific visual inspection cases, it is interesting to see that routes are often relatively homogeneous in terms of avalanche terrain within a spatial cluster. A large number (252 of 655) of spatial clusters had over 90% in the same terrain cluster. Also, the standard deviation for 95th percentile values within spatial clusters on average was half that of the entire dataset. More research is needed to decide what makes some spatial clusters more heterogeneous than others. The hypothesis here is that terrain plays a role in it. Open, unconstrained terrain with different route options leads to more heterogeneous clusters. This is supported by the fact that route height is negatively correlated with variability within a cluster, since low terrain is often more open than high alpine terrain, although this trend is not very strong.

4.5. Comparison of GPS tracks and planned routes

In the next section, the GPS tracks and planned routes are compared. This is done in two ways. First, some descriptive statistics are presented to compare the two datasets. Then, a number of locations is inspected visually to see how GPS tracks and planned routes deal with avalanche terrain differently.

4.5.1. Quantitative differences

Regarding the avalanche terrain traversed by both sets of routes, they are quite similar. On average, the GPS tracks seem to traverse a bit less hazardous terrain. The main difference though is in the route length. The GPS tracks are far longer on average. Visual inspection shows that many GPS tracks include multiple peaks and are often more complex than the planned routes. One reason why this may be the case is that people plan their routes in shorter segments, whereas GPS tracks include the entire route. Also, some GPS tracks include long traverses through flat valleys for example to get back to the car park. This is not really relative terrain for planning related to avalanche terrain, which might be the reason these sections are often left out in Whiterisk route plans. Thirdly, GPS tracks include all the cut-backs made on steep terrain sections, whereas most people on Whiterisk draw those sections as a straight line.

	Planned routes	GPS tracks	Planned routes	GPS tracks	Planned routes	GPS tracks
	Mean terrain hazard	Mean terrain hazard	95th percentile terrain hazard	95th percentile terrain hazard	Length in meters	Length in meters
Mean	0.31	0.27	0.63	0.59	7629	11547
Median	0.30	0.28	0.66	0.63	6550	11198
SD	0.09	0.11	0.14	0.20	3931	4408
n	53553	777	53553	777	53553	777

Table 11. *mean, median, and standard deviation for terrain hazard values and route length for planned routes and GPS tracks*

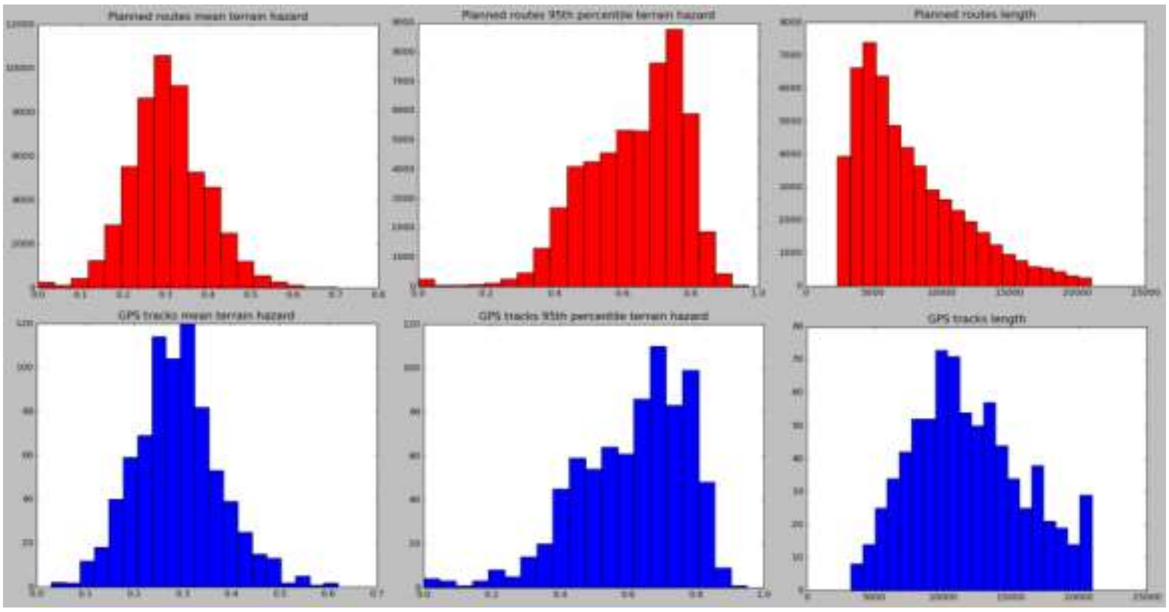


Figure 52. *Histograms of terrain hazard values and route length for planned routes and GPS tracks*

The histogram of the route length is interesting. Not only is the mean length of the GPS tracks higher than that of the planned routes, there is also a peak just before the cut-off point of 21 kilometres. Among the planned routes, there are hardly any routes near this point. This is the threshold used to filter out routes that are probably multi-day trips. This indicates that there are probably more multi-day trips among the GPS tracks. This further supports the idea that people may plan their routes in segments. Many of the planned routes are leading to mountain huts, and are thus part of a multi-day tour. Yet, hardly any routes are planned as such. One reason why this may be the case is that the Whiterisk app is used for route-finding. This is simpler when they can load a new route for each day.

Another result is the fact that the routes are quite similar in terms of terrain hazard values. The GPS tracks are a little bit safer in general. However, this does not say as much for the entire dataset, since the geographical spread of the routes is different (the GPS tracks are concentrated in the south-west whereas the planned routes are evenly spread over the Swiss alps). Therefore, in a next step the planned routes and GPS tracks are compared at a spatial cluster level. First, however, the terrain clusters in the GPS tracks are analysed in order to see if there is a different pattern there.

4.5.2. Terrain clusters

The GPS tracks had the percentage of each route in each of the discrete terrain classes calculated. On average, these percentages are as follows:

	GPS tracks, μ	GPS tracks, SD	Planned routes, μ	Planned routes, SD
High triggering potential	**2,7**	3,3	**3,3**	3,2
Medium triggering potential	**4,7**	4,2	**5,7**	4,0
Low triggering potential	**2,3**	3,0	**2,7**	2,5
Very low triggering potential	**1,2**	1,9	**1,3**	1,8
> 50°, prime risk falling	**0,6**	3,3	**0,4**	1,2
> 60°, prime risk falling	**0,1**	0,7	**0,1**	0,5
Remote triggering, high potential	**1,7**	1,9	**1,5**	1,4
Remote triggering, medium potential	**5,5**	4,0	**5,1**	3,5
Remote triggering, low potential	**3,5**	2,5	**3,3**	2,3
Remote triggering, very low potential	**8,0**	5,8	**7,7**	5,8
Maximum runout zone	**8,2**	6,1	**7,6**	6,3
No avalanche terrain	**61,3**	22,1	**61,4**	21,2

Table 12. *discrete terrain classes for entire GPS track dataset and entire planned route datset*

Although the percentage of each terrain class traversed by both datasets is quite similar, there is one interesting difference. The planned routes have higher values for all triggering terrain, while the GPS tracks have higher values for all remote triggering terrain. Since remote triggering terrain usually borders triggering terrain, this means GPS tracks tend to go around triggering terrain more often. This is the reason why GPS tracks have slightly lower values for terrain hazard. Based on these values, the GPS tracks were assigned to terrain clusters. This allows for a comparison between the cluster membership of GPS tracks and planned routes, shown in table 13.

Cluster no.	Cluster name	% GPS tracks	% Planned routes
0	Steep routes with extreme terrain	1,5	1,6
1	Very flat routes	31,6	30,1
2	Steep routes	12,9	15,4
3	Flat routes, many remote triggering zones	21,9	18,8
4	Typical routes	32,0	34,2

Table 13. *terrain cluster membership for GPS tracks and planned routes*

The fact that GPS tracks traverse more remote triggering terrain is reflected in the higher percentage of routes being assigned to the cluster “flat routes, many remote triggering zones”. Also, the lower terrain hazard values and lower percentage of route segments traversing triggering terrain is reflected in the lower percentage of routes being classified as “steep routes”.

4.5.3. Visual inspection

Cabane Prafleuri

Two locations were chosen to inspect the difference in planned routes and GPS tracks in detail. The first runs between Cabane Prafleuri and the Nendaz ski area in Valais. What is interesting here is that the GPS tracks and planned routes follow the same spatial pattern. However, some of planned routes traverse less dangerous terrain, and have lower 95th percentile values. Two locations are key in this, one at the beginning, the other at the start of the routes. These are shown in figure 54.

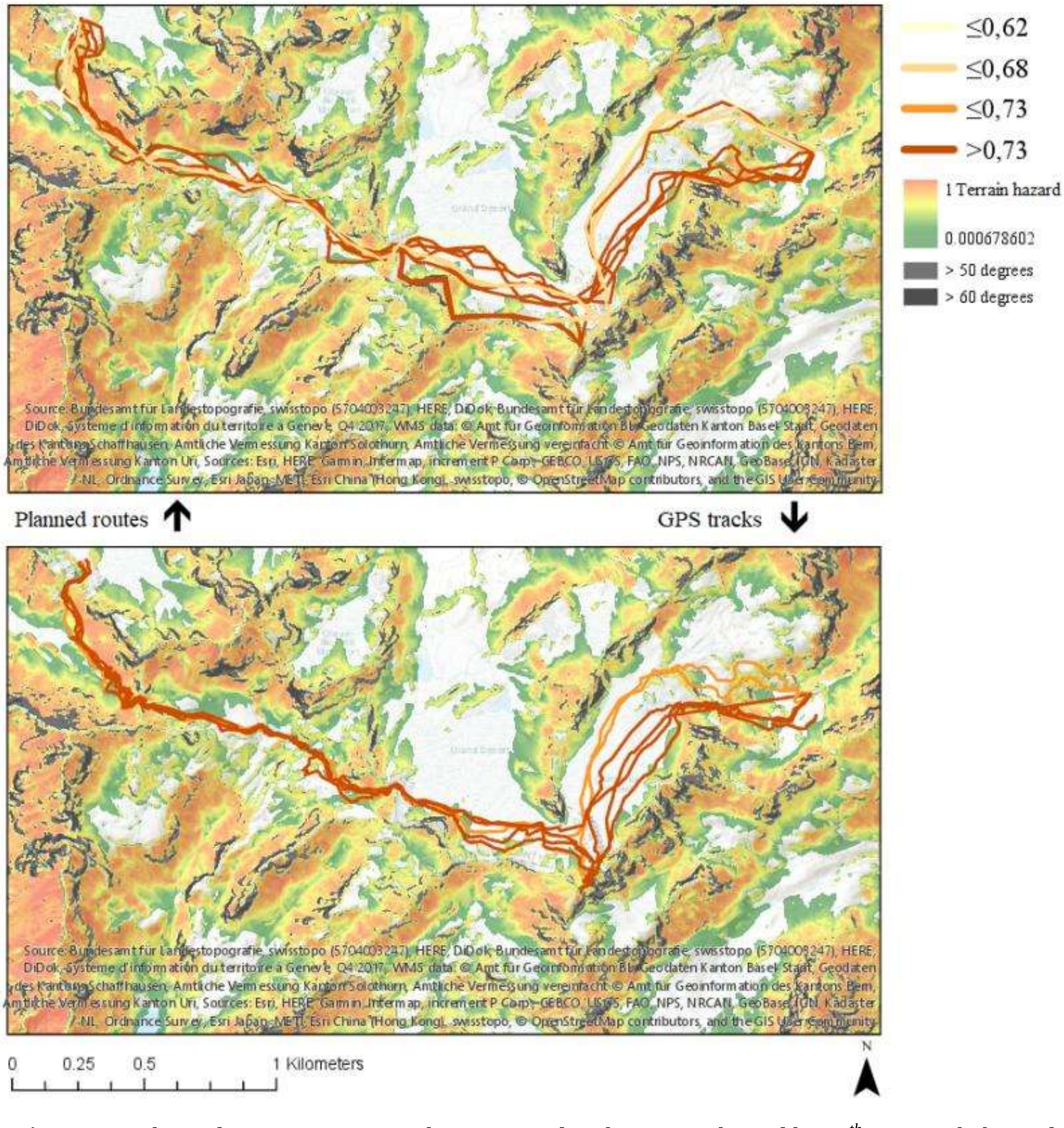


Figure 53. *Planned routes & GPS tracks near Nendaz ski area, coloured by 95th percentile hazard*

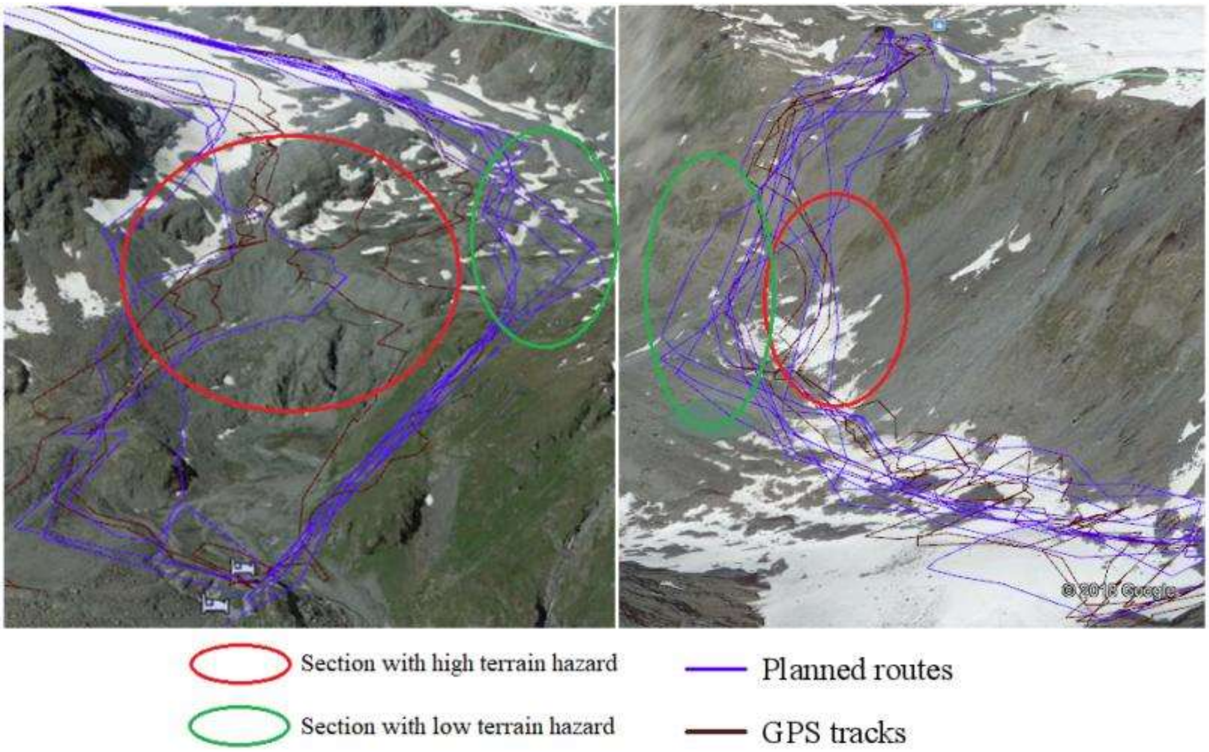


Figure 54. *two key sections in the planned routes and GPS tracks near Nendaz ski area*

These two sections are important because they have high values for terrain hazard. Some of the planned routes and GPS tracks overlap, but there are also a number of planned routes that take a considerably safer trajectory. In both cases this means taking a longer but flatter route to go around steep and dangerous sections. This is not done in the GPS tracks. Perhaps the reason for this is that it takes longer. In the right image, the safer option also involves a traverse that goes up and down in quick succession. This is especially tiring, as involves either walking up without climbing skins or putting on climbing skins for a short period of time. Although the planned routes have lower values for 95th hazard, the mean hazard values are actually higher.

Pigne d'Arolla – Western side

A similar pattern can be detected in the following set of routes traversing Pigne d’Arolla from the western side. Again, the planned routes have lower 95th percentile values than the GPS tracks. This is mostly because of the encircled section. Here, some of the planned routes take a longer way in order to avoid some steep terrain. Only one of the GPS tracks does this, which leads to high 95th percentile values. This section is encircled in figure 55, and a zoomed-in view of it is provided in figure 56.

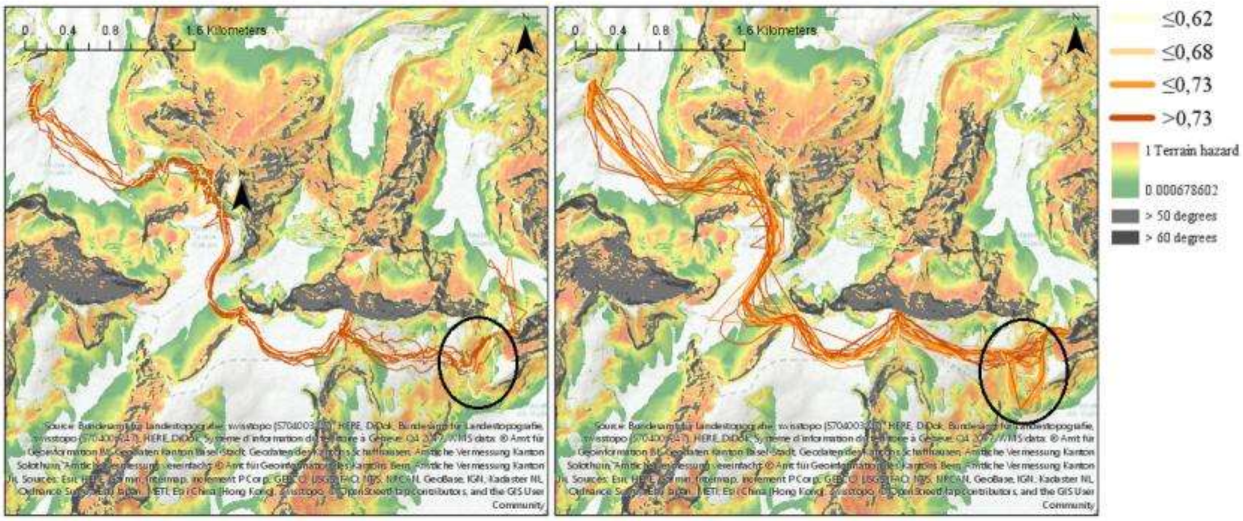


Figure 55. *Planned routes & GPS tracks traversing Pigne d'Arolla, coloured by 95th percentile hazard*

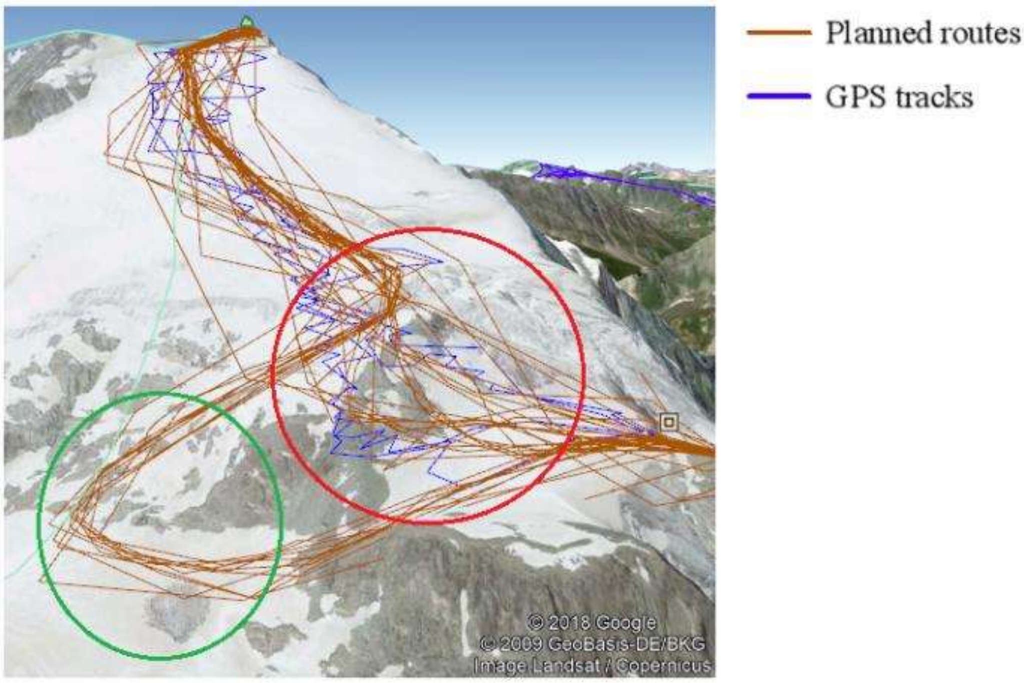


Figure 56. *key section in routes traversing Pigne d'Arolla from west side*

As can be seen by the many cut backs in the GPS tracks, they were going up at this section. The safer way would have been to follow the wide bend around the steep section. However, then they would have to spend a period of time going slightly downhill, and then uphill again. This is more tiring, and perhaps that is why this shortcut was taken.

Although the 95th percentile values in the GPS tracks were higher, the mean values were lower again. This could be due to the greater level of detail in the GPS tracks. As seen by the many cut backs in figure 56, tour skiers often try to follow a way through the terrain that takes as little effort as possible. In doing so they may be avoiding some slightly steeper terrain cells subconsciously. The planned routes draw a straight line at those sections, ignoring minor variations.

In the above two cases, the GPS tracks had higher 95th percentile values but lower mean values. Given the low density of GPS tracks, it was not possible to reliably check whether this pattern persists for other spatially similar routes. A hypothesis at this point is that in general people plan routes able to avoid risky sections successfully, compared to how well people avoid those sections in real life. However, more GPS data needs to be collected to prove this, as there are far fewer GPS tracks than planned routes, which makes drawing conclusions problematic.

This comparison has shown that on dataset level, the differences between GPS tracks and planned routes are minor. The GPS traverse a bit more remote triggering terrain and a bit less triggering zones, which leads to lower hazard values for the GPS tracks. However, on individual spatial cluster level, the GPS tracks are quite a bit more dangerous, and this seems to come from the GPS tracks seemingly prioritizing short and efficient routes over safety. However, the small number of GPS tracks makes it hard to generalize this idea. Also, participation inequality probably plays a role in the different patterns in both datasets. Lastly, temporal data would be needed to test whether this pattern persists with different temporal avalanche danger ratings.

5. Discussion & conclusion

The findings from the results chapter are now discussed and interpreted. First, a discussion of the results and the considered and used methods is presented based on the five sub-questions. Then, concrete conclusions are drawn per sub-question. Limitations and further research directions are identified as well.

5.1. Discussion

5.1.1. Relating modelled avalanche risk terrain to planned routes

Several methods were assessed for relating avalanche risk terrain factors to the route data. In the end, the 95th percentile hazard and mean hazard were chosen for comparisons between individual routes. Other measures were also considered but not included in the thesis. One of those was the maximum terrain hazard. Visual inspection shows this correlates with 95th percentile values. However, it was deemed that the 95th percentile values were more robust to small errors in route planning and inaccuracies due to the digital representation of routes. Another possibility was to choose one of the discrete terrain classes and compare how much routes intersect it. For this, the class "high triggering potential" seemed useful. However, since this is a discrete class, it was decided to be less suited than the continuous terrain hazard data. Small variations between routes can lead to high differences when they are near the border of a class. With the continuous data, this problem is not present.

Minimum bounding polygons are useful to describe the general terrain in which the routes are planned. However, visual inspection shows that they are not accurate on an individual route level. Another possibility is the proportion of a route intersecting each of the discrete terrain classes. This is useful for clustering the routes and to get an idea of the characteristics of a route. However, for differentiating between pairwise routes, the terrain hazard measures are better since they summarize the terrain in a route in a single number.

There are limitations to the methods described above. The planned routes are a digital representation of reality. In reality, many tour skiers choose to follow existing skinning tracks as this takes less effort than making their own track. As such, they may not follow their planned route in the amount of detail that is assumed in the methods described here. This especially holds true for the 95th percentile values. A drawn route may do well to go around a small dangerous section and because of that it will have a lower 95th percentile value than a route that goes through it. In reality, they may have followed a track that goes through the steep section. However, for this thesis the planned routes were treated as real routes, because the goal was to analyse the planning phase of touring, not the travelling phase. As such, the planned routes should be compared with each other, since it is impossible to know whether the people who drew them followed them in reality.

Another limitation is regarding no data in the calculations. Cells with no data are not used in the calculations of mean and 95th percentile terrain hazard values. As such, a route that traverses many routes with values near zero will have lower values than a route that follows a similar trajectory but has more no data cells, while in reality the former is actually more dangerous. The obvious solution to this would be to treat no data cells as zero. However, this results in long routes having even lower values, and the terrain hazard values in routes becoming too much related to route length. This effect is reversed for the 95th percentile values. When including no data cells, there are more cells included with values lower than the 95th percentile value, which makes the value higher.

The calculation of 95th percentile values is based on a discrete representation of the continuous terrain hazard data. To do this, the values in the terrain hazard data were multiplied by 1000, and a discrete raster was built containing those values. As such, the precision is brought down to three decimal points. Some accuracy is lost in this process. However, the values computed in this way are not compared on a very fine scale anyway, and the processing speed won in this process is deemed worth the loss in precision.

Using visual inspection, the 95th percentile was judged as a useful measure to compare individual routes. However, field observations and expert input would be needed to further prove this if it were to be used in the future to select safe routes.

5.1.2. Clustering based on avalanche terrain

The next step was to cluster route based on the avalanche terrain they intersect. For this, the discrete terrain classes were used. The goal of this was to detect patterns in the avalanche terrain classes traversed by routes. Five clusters were identified using the k-means algorithm with k=5. The differences between the clusters are clear when viewing them on a map, and routes that follow similar trajectories were usually assigned to the same cluster. The other route attributes for the majority differed significantly between clusters. The very flat routes, for example, are usually in lower terrain, whereas the very steep routes are significantly shorter than average. Also, for routes that have the avalanche bulletin danger rating specified, there was a logical pattern. The very flat cluster had higher mean danger ratings than the steep and extreme terrain clusters.

Some decisions had to be made to come to this clustering. The first was the attributes to perform clustering on. Other route attributes such as length or height could have been included. This would perhaps have led to more clearly defined clusters (with bigger inter-cluster differences). However by not doing this it was possible to compare how those attributes differed between clusters. If they would have been included, the algorithm would have actively used them to divide the data, which would perhaps have made the inter-cluster differences in those attributes less interesting. It would also have been possible to include mean and 95th percentile hazard values in the clustering. However, since those values are strongly correlated to the discrete terrain classes, it would not have made a large difference. Also, the goal was to diversify based on the type of terrain, instead of just on terrain hazard. For example, clusters 3 and 4 on average intersect different terrain. Cluster 4 has almost twice as much "no avalanche" terrain, and only half in most remote triggering terrain classes. As such, those clusters are found in different locations: cluster 4 is more often in broad, mellow terrain, such as glaciers, whereas cluster 3 is more often found in gradually inclining terrain with very steep slopes to the sides, such as narrow valley floors. However, their mean values for 95th percentile and mean terrain hazard are almost similar. Therefore, including those values in the clustering algorithm would have perhaps decreased the inertia for routes in those clusters and placed the optimal value for k at 4, effectively merging clusters 3 and 4. Therefore, not including those values has in this case made the clusters better defined.

Another consideration that is relevant in nearly all clustering tasks is the chosen algorithm and parameters. Since the number of clusters was hard to dissect beforehand from the data, DBScan was first deemed a good algorithm. Here, the user does not need to choose the number of clusters beforehand. However, the results here were not useful. The algorithm either found a large number of very small clusters or one very large cluster. This is due to the nature of the data. DBScan works best when there are clear, discrete boundaries to the clusters. This is the case in the spatial distribution of routes. However, the route attributes have a smoother distribution. Therefore, DBScan will keep adding routes to each cluster, as most routes have at least one route that is somewhat similar.

This problem is not there when using k-means. However, since the data is so smooth, clusters from k-means will be somewhat arbitrary. Quite a large proportion of routes will be close to their cluster boundary. Therefore, the cluster membership of a route should be analysed with some care. This is also why in a section after this, the clusters were visually inspected. This showed that despite smooth nature of the data, the clustering algorithm has done quite well at assigning different routes to different clusters. However, it also became clear in the visual inspection that homogeneous routes weren't always assigned to the same cluster. Often, a majority of routes following the same trajectory was assigned to the same cluster, with one or two routes being assigned to a different one. In many cases this was because this route was actually significantly different. In some, however, those routes were just over the boundary of another cluster.

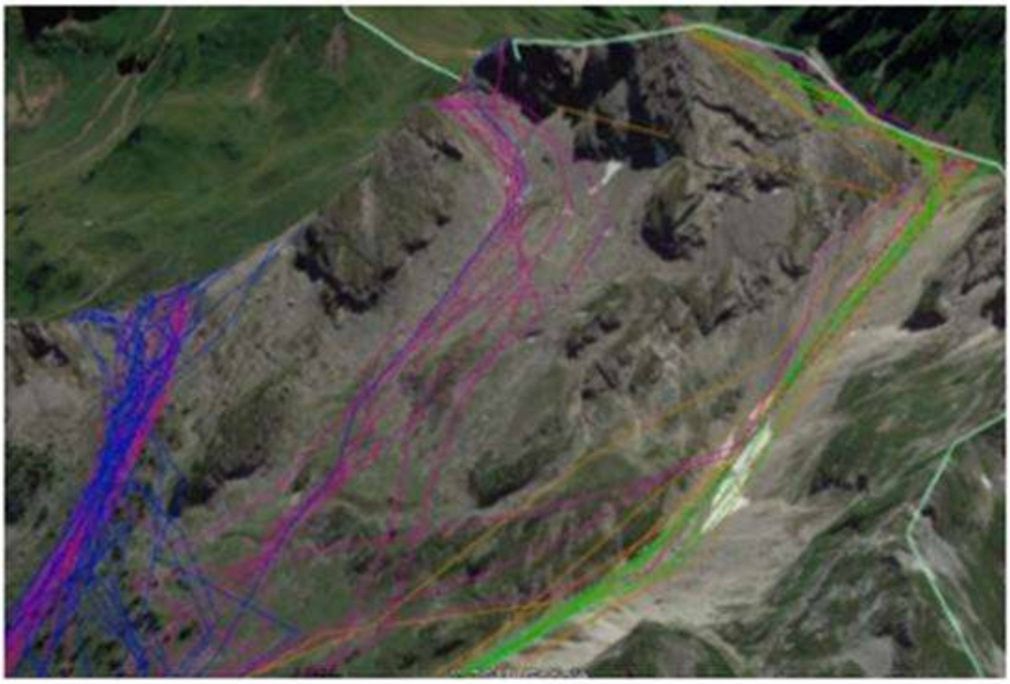


Figure 53. *three different variations within one spatial cluster*

This is for example the case in the above figure from chapter 4.4. There are three clearly distinguishable trajectories among the routes. They have rightly been assigned to three different terrain clusters. However, among each of the trajectories there are also numerous routes in another cluster than the majority. In the green one, there are several orange and purple. In the purple one, there is a blue one and vice versa. Those are probably routes near the boundary of their terrain cluster. Cases like this have been ignored in the visual inspection section. However it is important to acknowledge that in a better fitting clustering approach they should have probably been assigned to a different terrain cluster. Given more time, a systematic approach should have been used to differentiate between a larger number of clustering algorithms, parameters, and input data. Internal validation measures such as silhouette analysis or the Davies-Bouldin method (Baarsch & Celebi, 2016) could be used for this. However, it was decided after inspection that the method used was good enough for the purpose and scope of this thesis.

5.1.3. Spatial clusters

In a next step the routes were clustered based on spatial locations. For this the Frechet distance was used to compute the similarity, and then input to the DBScan clustering algorithm. This proved a useful measure for the purpose of clustering, which was to compare routes that are spatially similar. In the results chapter it was already discussed that it is impossible to find perfect parameters for this dataset. The reason for this is that with varying terrain, the similarity between routes becomes different. Areas with narrow valleys should have stricter parameters than broad terrain. One way to solve this issue is to include context in the clustering. A good example of this is provided in Buchin et al. (2014). This could make clustering here more accurate. For example, a steep slope separating two routes should increase the measured similarity, while a homogeneous area between two routes should decrease it.

5.1.4. Relation between spatial clusters and avalanche terrain

The relation between spatial clusters and avalanche terrain was then researched. It was found that in general, there is a strong homogeneity among routes within spatial clusters in term of terrain cluster membership. The same goes for 95th percentile values. However, for both there are also heterogeneous spatial clusters. Those were visually inspected. It seems that terrain is of influence on the homogeneity within a spatial cluster. Areas that are homogeneous in terrain also produce more homogeneous routes. For the heterogeneous routes, two things seem to cause the variability. First, some routes are treated as a spatial cluster as they have the majority of their trajectory follow the same line. However, they disperse at some point and end up at different destinations. If these destinations are different in terms of avalanche terrain, it causes a large degree of heterogeneity within the spatial cluster. This is only possible if the terrain is open enough that different routes can be planned. Some spatial clusters are limited by the terrain, for clusters that traverse narrow ridges or narrow valleys. Clusters in more open terrain seem to be more heterogeneous. However, this hypothesis would have to be tested quantitatively. One way this could be done is to assign a score to the terrain in a spatial cluster, with more limiting terrain having a lower score. Then, correlation between this score and the variability in a spatial cluster could be calculated.

The second cause is when routes are planned using different “strategies”. It seems many routes follow SAC ski touring routes and regular hiking trails, which are visible as a map layer in Whiterisk. However, those are not always the optimal routes in terms of avalanche safety. In a future research, it would be interesting to digitize the SAC routes, and the compute terrain hazard measures for them. They could then be compared per spatial cluster to the planned routes. In this thesis this was done for two examples, both of which had safer alternatives drawn by users. However, to test whether the alternatives drawn by users are structurally safer, they would have to be compared quantitatively for each spatial cluster. In any case, the SAC routes probably weren’t meant to follow precisely. However, when people draw routes that exactly follow them, and then used those routes for navigation in the Whiterisk app, this is what they are in fact being used like.

Other methods were considered to investigate the relation between surrounding terrain and avalanche risk in a route. Analysing the surrounding terrain within a defined zone per route, instead of the terrain in other surrounding routes, was tried. The bounding polygons provided an exploration into this, but produced inaccurate results. Other methods to do this could be to use a watershed model to make sure only spatially similar terrain is included, or a simple buffer around each route. The terrain in this buffer could then be compared to the terrain within the route. However, the spatial cluster was chosen as it was considered more realistic, since it makes use of planned data from other people, and as such is more sensitive to spatial context.

5.1.5. Comparison between GPS tracks and planned routes

The planned routes and GPS tracks were compared for two reasons. First, to establish whether there are any notable differences between route planning and actual travel. Second, to put the planned routes into the broader context of backcountry routes. Across the datasets, there were no major differences in terms of the types of terrain traversed by both sets of routes. The GPS tracks on average were slightly safer. This was due to them intersecting more remote triggering terrain and less direct triggering terrain. However, as we don’t know when and by whom the planned routes and GPS tracks were undertaken, it is hard to compare them directly. Further, the GPS tracks are more often multi day trips, since they have a far higher mean length. This also makes comparisons across the datasets difficult, since the GPS tracks often include long traverses trough valleys, e.g. to get back to the start point. Most route planners exclude such traverses.

Because of this, two sites were chosen to compare both datasets within the same spatial cluster. Here an interesting finding was that within a spatial cluster, GPS tracks are actually more often intersecting avalanche terrain. The reason for this seems to be that GPS tracks tend to avoid longer, safer routes if this means a lot of extra effort. The planned routes more often do take those detours. However, given the low number of GPS tracks, it is hard to prove this hypothesis. Also, again, it is not known when the routes were travelled. Perhaps the GPS tracks that avoided the safe routes were all travelling in very stable situations. Also, participation inequality probably plays a role in the different patterns. For future research, it would be interesting to let people first plan a route and then equip them with trackers to see to what extent they really follow those routes.

The GPS tracks are also more precise than the planned routes. In steep sections, the GPS tracks are using cut-backs to find an efficient way up, whereas many planned routes draw a straight line. Because of this, it can be expected that GPS tracks have artificially high values for terrain hazard, as they spend more of the route proportionally in the steep hazardous terrain. This is artificial because in reality the people who drew the planned routes will not take a straight line up, and also spend more time there. However, in the visual inspection it was shown that GPS tracks also take more dangerous routes when ignoring this.

5.1.6. Route suggestions

Given the fact that many people follow SAC routes, and that in some examples those routes turned out to be relatively unsafe, it seems like a logical next step to offer an alternative route layer that users can base their plan on in Whiterisk. In the figure below, an example of this is shown for the Pigne d'Arolla. Here, a spatial cluster is first identified. Then, the route in this cluster with the lowest 95th percentile value is selected. This is shown in the figure below.

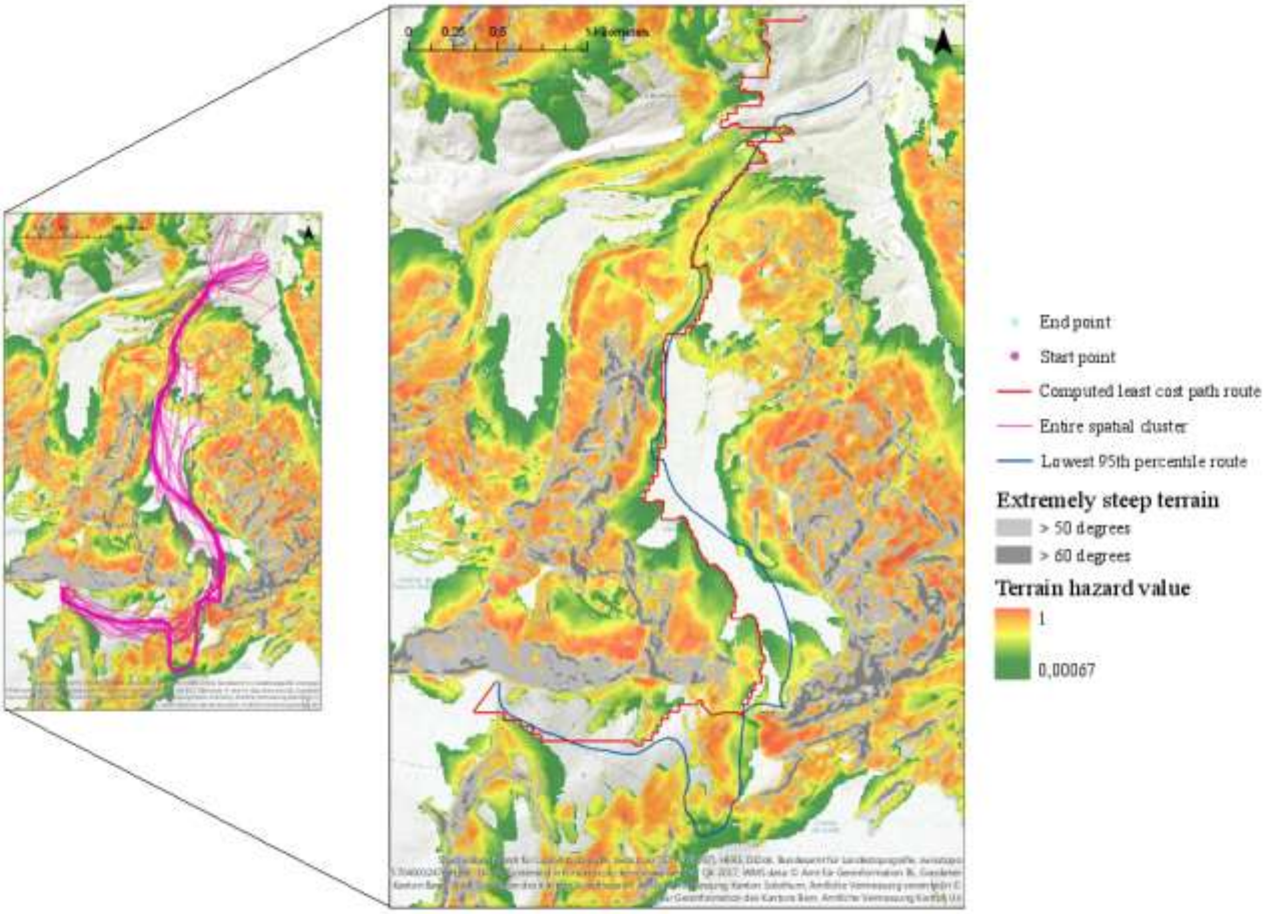


Figure 54. *Lowest 95th percentile planned route within spatial cluster near Pigne d'Arolla, and least cost path computed route using similar start and end points*

This route can then be shown on Whiterisk as a guideline. This is useful for people who know the start and end point and want a safe route in between. The way this would work is users specify a start and end point. Then, it is checked whether there are already routes in the data with similar start and end points (within a search radius). Then, the safest of those is shown and users can base their route on that. However, it should probably only be shown if a. there is at least a specified number of routes with a similar trajectory (a spatial cluster) and b. one or more of those routes is significantly (above a specified threshold) safer than the rest. If those requirements are not met, the suggested trajectory may not be that safe, relative to the specified start and end point, and the user may be better off drawing their own route from scratch. In the figure above, another approach is also shown. This makes use of least cost path analysis (de Smith, Michael, & Goodchild, 2016) to calculate the shorterst path from start to end point, including terrain hazard as a cost factor. This path also does well at avoiding hazardous terrain. However, there are several reasons why in my opinion the optimal route based on the planned route data is better than the computed least cost path. First, the computed optimal route takes an alternative path that has low values for terrain hazard but is still avoided by most routes. Near the top, it follows a good path in terms of avalanche terrain, but this goes through an area with many crevasses (Ski libre, 2011). This is probably the reason why the planned routes avoid this section. Secondly, as the computed least cost path is looking to get the shortest possible route, it stays very close to the boundary of the avalanche terrain hazard where this is possible. As such, if someone follows this route but goes a bit off the track, chances are they will end up in hazardous terrain. Thirdly, the planned routes are not just safer, but also take other preferences from users into account, such as taking an aesthetically pleasing track or passing by points of interest. Those things are hard to model as inputs for the least cost path. As such, my suggestion would be to use the data to give users a suggested route where this is possible given the two demands stated above. If this is not possible, a least cost path could be shown. In both cases, users will have to be made clear that those routes are mere suggestions and they should still look for themselves if they agree with them.

Another method that uses terrain data to automatically compute the safest (and also most efficient) route between a given start and end point, is made by Eisenhut (2011). He used a multi-criteria assessment that gives a penalty for steep slopes, and also takes land cover into account. Instead of drawing a single line, he shows a buffer within which planning is relatively safe, called a corridor. To compare, the lowest 95th percentile route, the computed route, and the corridor from Eisenhut are shown below.

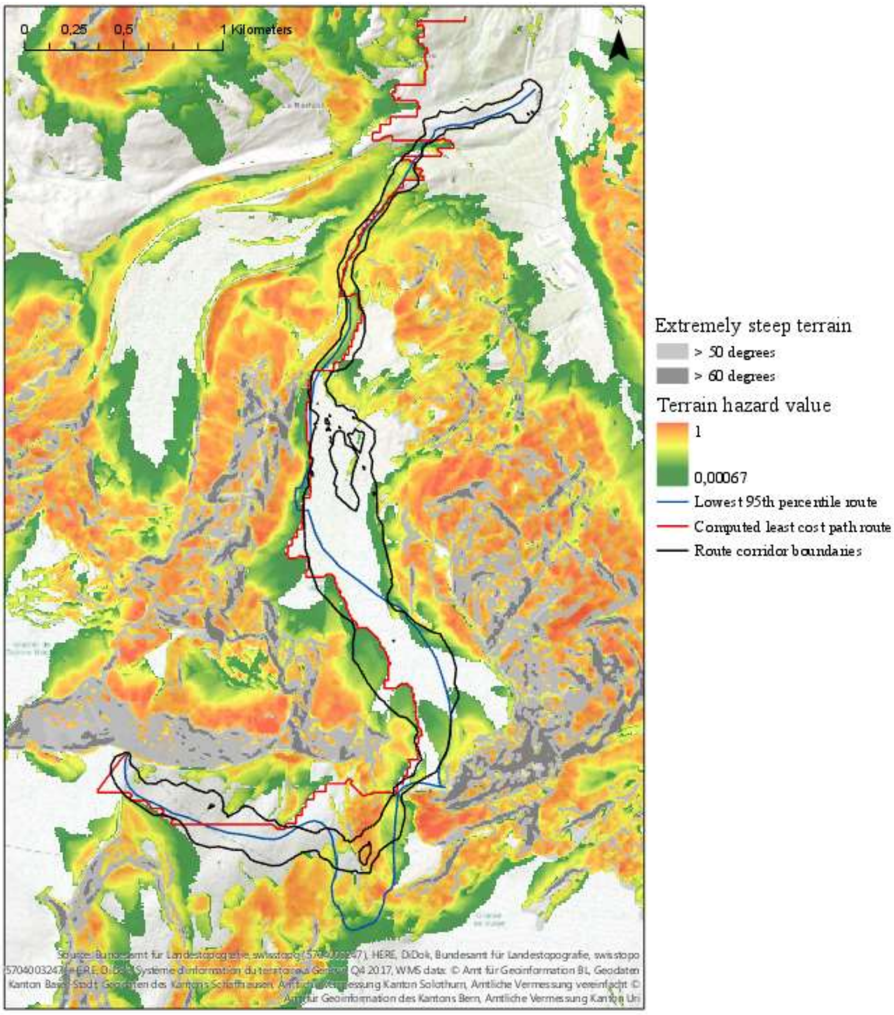


Figure 55. *Lowest 95th percentile route, computed least cost path route, and Eisenhut's corridor*

The main difference between the safest planned route and the corridor is the way they approach the key section in the southern end of the route. In previous sections, it was shown that this is an important factor in the 95th percentile values of route. Of the planned routes, the ones that take a detour around this section generally have lower 95th percentile values. Therefore, it is interesting that the corridor does not take a detour. This is due to the way terrain is interpreted in the algorithm from Eisenhut. In the terrain hazard, possible consequences are also integrated, which leads to different results. Also, in Eisenhut's algorithm, the route with the least effort is preferred. The detour is probably seen as more effort than the short route up. The data give a good new way to give safe route suggestions, as they take into account preferences by users and at the same time prioritize avoiding steep terrain. Also, the safest 95th percentile route has the lowest 95th percentile of the three, 0,51. The least cost path has 0,57, and a typical route drawn in the middle of the Eisenhut corridor has 0,54.

5.2. Limitations

Some specific limitations regarding the research questions were already identified. There are also more general limitations to this research. The data used in this research has biases which are common in most user-generated content, such as demographic and spatial biases. This was already identified by Schönenberger (2018). Some of the hypotheses in this thesis were based on specific study sites. For example, the fact that routes following SAC trajectories are more unsafe than routes in the same cluster that don't. This should be researched quantitatively. Another idea that came forward in this thesis is the fact that many people plan routes that are unsafe but shorter than routes that take a detour around dangerous terrain. This would mean they regard a little less safety acceptable if it means an easier way up. However, it is not known to what extent the routes are exactly followed. It is possible that people still deviate from their route in the field when it turns out that another alternative is safer.

Another important limitation is on the visual inspection used. Here, cases were selected where a sufficient number of routes were drawn, and some other attribute of the drawn routes made the case especially useful for inspection. This should be seen as an exploration. More quantitative analysis would be needed to generalize the findings of the visual inspection to the whole population. Lastly, temporal data on the routes would really make the research more robust. It would allow for an analysis of how well people plan to avoid specific risk factors, such as instable snow on one slope aspect.

5.3. Conclusion

The goal of this research was to identify the relationship between planned routes and avalanche terrain. In the introduction it was already stated that this relation is complex and dependant on location. However, some interesting results were produced. Those are presented here per sub-question.

1. "Which different methods can be used to relate avalanche terrain risk to planned routes, and what are their (dis)advantages?

Of the two modelled terrain datasets from Harvey et al. (2018), the terrain hazard data was deemed most suitable for calculating route terrain attributes, since this data is scaled from 0 to 1 and thus allows for numerical comparisons. Two measures were computed from this data: the mean terrain hazard rating and the 95th percentile hazard rating. Of those, the mean hazard rating is less sensitive to minor variations between spatially similar routes. As such, it is a useful measure to describe the general character of a spatially similar set of routes in terms of the terrain hazard. The 95th percentile is more sensitive to small planning differences. Therefore, it is more useful to identify the safer individual trajectory within a spatial cluster of routes.

2. "What clusters exist in the route data based on their avalanche terrain characteristics?"

To cluster the routes, the proportion of each route intersecting the discrete terrain classes was used. Five clusters were identified. Visual inspection showed that the terrain clusters relate quite well with the real differences between routes. The clusters reflect the terrain they are planned in. For example, the routes with extreme terrain were often planned leading to high, alpine peaks with steep ridges. The class of routes with many remote triggering zones were often in narrow valleys with steep slopes going up on either side of the route.

3. "What spatial clusters exist in the route data?"

Using density based clustering, 34% of the routes was assigned to a spatial cluster. 655 clusters were found. The route clusters usually start in towns or mountain huts, and end on peaks or mountain huts.

4. “To what extent is avalanche terrain similar for routes within a spatial cluster?”

The standard deviation within spatial clusters for 95th percentile values is half of that for the entire population. Also, in 252 of 655 spatial clusters, over 90% of the routes was assigned to the same terrain cluster. As such, in general there is quite a strong relationship between spatial cluster membership and avalanche terrain in a route. However, there are also numerous spatial clusters where this pattern does not hold up. Visual inspection shows that the variability of avalanche terrain within a spatial cluster is dependent on the terrain. Terrain where there are not many options to deviate, such as narrow valleys, have more homogeneous routes than more open terrain where different route planning options exist. Secondly, route swithin a spatial cluster are often planned using different strategies. Some routes strictly follow the suggested SAC trajectory, while others follow hiking trails, and again others seem to mostly pay attention to the slope angles. These different strategies lead to variety within spatial clusters.

5. "Is there a difference in avalanche terrain of the planned routes and the GPS tracks?"

On dataset level, the differences between GPS tracks and planned routes are quite small, with the GPS tracks being on average somewhat safer and traversing more remote triggering terrain. However, on spatial cluster level, the GPS tracks are more dangerous. GPS tracks often seem to take the most economical route in terms of effort, avoiding long traverses even if it leads to a safer route. However, given the small amount of GPS tracks it is hard to generalize this finding.

5.4. Future research

5.4.1. SAC routes

Several future research directions were identified during this thesis. The first one is related to the SAC routes and their avalanche characteristics. Since it turns out that they are often followed by Whiterisk users, it makes sense to evaluate how well they deal with avalanche terrain. The way to do this is to filter out the routes that are part of a spatial cluster, and then compare them to the routes in that cluster. They will have similar avalanche terrain values to many routes in their spatial cluster, as these routes follow the SAC routes. However, in some of the more heterogeneous clusters there will be alternative routes that have different avalanche terrain ratings. Those can be compared to the SAC routes in their spatial cluster. Researching the avalanche terrain in the SAC routes in this regard is interesting as they are used by so many people to base their Whiterisk routes on. Furthermore, they are also used outside of Whiterisk when people use more traditional ways of route planning, e.g. paper maps or tour books. The modelled avalanche terrain makes it possible to quantify the avalanche risk in SAC routes.

5.4.2. Analysing route decisions

In this thesis the focus was on assessing how much avalanche terrain planned routes traverse and how they deal with avoiding this terrain. In a few examples visual inspection was used to differentiate between routes that carefully plan to avoid avalanche terrain and routes that don't. It is perhaps possible to quantatively analyse the process of drawing routes There was no strong relation found between the average step length and avalanche danger in this thesis. One way this could be researched further is by counting at which types of locations people draw vertices. In theory they should draw more vertices near key sections. Converting the lines back to vertices allows for counting points within each discrete terrain class, or within a range of cells above a specified terrain hazard value. This can then be compared to the proportional amount of a route intersecting each terrain class. The outcome of then could then be correlated to mean or 95th percentile hazard values. It would be an interesting finding if there is a pattern in this within spatial clusters. Perhaps routes that plan proportionally more vertices near steep sections are safer than spatially similar routes that plan proportionally more vertices in flat areas.

	% of vertices	% of route
No avalanche terrain	56,25	58,96
High triggering potential	12,50	5,91
Medium triggering potential	9,38	8,58
Low triggering potential	3,13	4,26
Very low triggering potential	0	2,22
> 50 degrees	3,13	1,02
> 60 degrees	0	0
Remote triggering, high potential	0	0,70
Remote triggering, medium potential	3,13	4,83
Remote triggering, low potential	0,00	2,67
Remote triggering, very low potential	6,25	5,78
Max runout zone	6,25	5,08

Table 14. *proportion of route vertices per terrain class vs. proportion of total route in each terrain class*

The route in the example above has drawn far more vertices in high triggering potential terrain. It would be interesting to see whether this holds for the entire dataset, and whether this has influence on the relative hazard values of routes.

References

Alt, H., & Scharf, L. (2012). Computation of the Hausdorff distance between sets of line segments in parallel, 1–13. Retrieved from http://arxiv.org/abs/1207.3962

Andrienko, G., Andrienko, N., Bak, P., Keim, D., & Wrobel, S. (2013). Visual Analytics of Movement. Berlin/Heidelberg: Springer-Verlag.

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).

Antoniou, V. (2011). User Generated Spatial Content: An Analysis of the Phenomenon and its Challenges for Mapping Agencies. Doctoral Thesis, 324. Retrieved from http://discovery.ucl.ac.uk/1318053/

Ayuso, N., Cuchí, J. A., Lera, F., & Villarroel, J. L. (2015). A deep insight into avalanche transceivers for optimizing rescue. Cold Regions Science and Technology, 111, 80–94. https://doi.org/10.1016/j.coldregions.2014.12.005

Bartelt, P., Buser, O., Vera Valero, C., & Bühler, Y. (2016). Configurational energy and the formation of mixed flowing/powder snow and ice avalanches. Annals of Glaciology, 57(71), 179–188. https://doi.org/10.3189/2016AoG71A464

Brewster, T., (2018, 29 Jan). Why Strava's fitness tracking should really worry you. Forbes magazine. https://www.forbes.com/sites/thomasbrewster/2018/01/29/strava-fitness-data-location-privacyscare/#3d54745655c3 retrieved on 25-09-2018

Bründl, M., Bartelt, P., Schweizer, J., Keiler, M., & Glade, T. (2010). Review and future challenges in snow avalanche risk analysis. In I. Alcantara-Ayala & A. S. Goudie (Eds.), Geomorphological Hazards and Disaster Prevention (pp. 49-61). Cambridge: Cambridge University Press.

Buchin, M., Dodge, S., & Speckmann, B. (2014). Similarity of trajectories taking into account geographic context. Journal of Spatial Information Science, 9(9), 101–124. https://doi.org/10.5311/JOSIS.2014.9.179

Budhathoki, N., Bruce, B., & Nedovic-Budic, Zorica (2008). Reconceptualizing the role of the user of spatial data infrastructure. GeoJournal. 72. 149-160. 10.1007/s10708-008-9189-x.

Bühler, Y., Kumar, S., Veitinger, J., Christen, M., Stoffel, A., Snehmani. (2013) Automated identification of potential snow avalancherelease areas based on digital elevation models, Nat. HazardsEarth Syst. Sci., 13, 1321–1335, doi:10.5194/nhess-13-1321-2013, 2013

Cambridge dictionary online, 2018. avalanche. https://dictionary.cambridge.org/dictionary/english/avalanche retrieved on 07-09-2018

Campbell, C., & Gould, B. (2013). A proposed practical model for zoning with the Avalanche Terrain Exposure Scale, 385–391.

Christen, M., Kowalski, J., & Bartelt, P. (2010) RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain, Cold Regions Science and Technology, 63, 1-14

Conlan, M. J. W., Tracz, D. R., & Jamieson, B. (2014). Measurements and weather observations at persistent deep slab avalanches. Cold Regions Science and Technology, 97, 104–112. https://doi.org/10.1016/j.coldregions.2013.06.011

de Smith, M.J., Goodchild, M., & Longley, P. (2016) "Cost Distance", Geospatial Analysis, 3rd Edition

Delparte, D., 2008. Avalanche Terrain Modeling in Glacier National Park, Canada. PhD Thesis, University of Calgary, Calgary, AB, Canada

Dodge, S. (2016). From Observation to Prediction: The Trajectory of Movement Research in GIScience. Advancing Geographic Information Science: The Past and Next Twenty Years, 123–136.

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., & Miller, J. A. (2016). Analysis of movement data. International Journal of Geographical Information Science, 30(5), 825–834. https://doi.org/10.1080/13658816.2015.1132424

Dodge, S., Weibel, R., & Lautenschütz, A. K. (2008). Towards a taxonomy of movement patterns. Information Visualization, 7(3–4), 240–252. https://doi.org/10.1057/palgrave.ivs.9500182

Dreier, L., Bühler, Y., Steinkogler, W., Feistl, T., & Bartelt, P. (2014). Moddeling small and frequent avalnches. In: Proceedings, International Snow Science Workshop, Banff, Alberta, Canada, 2014

Eckerstorfer, M., Bühler, Y., Frauenfelder, R., & Malnes, E. (2016). Remote sensing of snow avalanches: Recent advances, potential, and limitations. Cold Regions Science and Technology, 121, 126–140. https://doi.org/10.1016/j.coldregions.2015.11.001

Eiter, T,; Mannila, H., (1994). Computing discrete Fréchet distance. Tech. Report CD-TR 94/64, Christian Doppler Laboratory for Expert Systems, TU Vienna, Austria.

Eisenhut, A. (2011) Skitourenplanung auf Knopfdruck? Die optimale Aufstiegsroute bezüglich Anstrengung du Lawinensicherheit, modelliert mit einem Geographischen Informationssystem (GIS). MSc thesis for the UNIGIS-program, University of Salzburg, supervised by Prof. Dr. Josef Strobl

Engel, Z. (2000). Skier-triggered avalanches. Acta Universitatis Carolinae, Geographica, 35(SUPPL.), 247–263.

European Parliament and Council. (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data [1995] OJ L 281/31

Fagerland, M. W., & Sandvik, L. (2017). Performance of five two-sample location tests for skewed distributions with unequal variances Performance of fi ve two-sample location tests for skewed distributions with unequal variances. *Contemporary Clinical Trials*, *30*(5), 490–496. https://doi.org/10.1016/j.cct.2009.06.007

Ferster, C. J., Nelson, T., Robertson, C., & Feick, R. (2018). Current Themes in Volunteered Geographic Information. Comprehensive Geographic Information Systems, 26–41. https://doi.org/10.1016/B978-0-12-409548-9.09620-2

Fitzgerald, K., Kay, J., Hendrikx, J., & Johnson, J. (2016). Backcountry and Sidecountry Users: an Intercept Survey Approach. International Snow Science Workshop 2016 Proceedings, Breckenridge, CO, USA, (October), 1057–1063.

Foster, A., & Dunham, I. M. (2015). Volunteered geographic information, urban forests, & environmental justice. Computers, Environment and Urban Systems, 53, 65–75. https://doi.org/10.1016/j.compenvurbsys.2014.08.001

Fredston, J., Fesler, D., 2011. Snow Sense: A Guide to Evaluating Snow Avalanche Hazard (5th edition). Anchorage, USA: Alaska Mountain Safety Center, Inc

Frühauf, A., Hardy, W. A. S., Pfoestl, D., Hoellen, F. G., & Kopp, M. (2017). A qualitative approach on motives and aspects of risks in freeriding. Frontiers in Psychology, 8(NOV), 1–13. https://doi.org/10.3389/fpsyg.2017.01998

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69(4): 211-221

Granell, C., & Ostermann, F. O. (2016). Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management. Computers, Environment and Urban Systems, 59, 231–243. https://doi.org/10.1016/j.compenvurbsys.2016.01.006

Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. Journal of Transport and Health, 2(2), 238–247. https://doi.org/10.1016/j.jth.2014.12.001

Haegeli, P., & Atkins, R. (2016). Managing the Physical Risk from Avalanches in a Helicopter Skiing Operation - Merging and Contrasting Terrain Use Data with the Operational Guiding Perspective. Proceedings of the International Snow Science Workshop, Breckenridge, CO, 104–111. Retrieved from http://arc.lib.montana.edu/snow-science/objects/ISSW16_O5.02.pdf

Hägerstrand, T. (1970). What About People In Regional Science?, 'Papers in Regional Science', 24(1), 6-21

Haklay, M. (2016). Why is participation inequality important? European Handbook of Crowdsourced Geographic Information, 35–44. https://doi.org/10.5334/bax

Hallandvik, L., Andresen, M. S., & Aadland, E. (2017). Decision-making in avalanche terrain–How does assessment of terrain, reading of avalanche forecast and environmental observations differ by skiers' skill level? Journal of Outdoor Recreation and Tourism, 20(December 2016), 45–51. https://doi.org/10.1016/j.jort.2017.09.004

Harvey, S. (2002). Avalanche Incidents in Switzerland in Relation to the Predicted Danger Degree. International Snow and Science Workshop, 2(Figure 1), 443–448.

Harvey, S., Schmudlach, G., Bühler, Y., Dürr, L., Stoffel, A., & Christen, M. (2018). Avalanche terrain maps for backcountry skiing in Switzerland. In International snow science workshop proceedings 2018 (pp. 1625-1631). Retrieved from http://arc.lib.montana.edu/snow-science/item/2833

Harvey, F. To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. In Crowdsourcing Geographic Knowledge; Sui, D., Elwood, S., Goodchild, M., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2013; pp. 31–42.

Hendrikx, J., & Johnson, J. (2016). Understanding global crowd sourcing data to examine travel behavior in avalanche terrain. Proceedings, International Snow Science Workshop, (October), 737–743. Retrieved from www.montana.edu/snowscience/tracks

Hendrikx, J., Johnson, J., Hardy, G., & Weinberg, W. (2014). Using Global Crowd-Sourced Data To Understand Travel Behaviour in Avalanche Terrain. International Snow Science Workshop Banff, 2, 2–5.

Hendrikx, J., Johnson, J., & Shelly, C. (2015). Using GPS tracking to explore terrain preferences of heli-ski guides. Journal of Outdoor Recreation and Tourism, 13, 34–43. https://doi.org/10.1016/j.jort.2015.11.004

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

Hendrikx, J. (2018) Understanding avalanche risk: A new paradigm | Dr. Jordy Hendrikx | TEDxBozeman". https://www.youtube.com/watch?v=kWOgyyDUD2w retrieved on 05-11-2018

Jamieson, B., Geldsetzer, T., & Stethem, C. (2001). Forecasting for deep slab avalanches. Cold Regions Science and Technology, 33(2–3), 275–290. https://doi.org/10.1016/S0165-232X(01)00056-8

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. Journal of Transport Geography, 52, 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006

Laube, P. (2014) Computational Movement Analysis. Berlin: Springer. Springer Briefs in Computer Science. ISBN 978-3-319-10267-2. https://doi.org/10.1007/978-3-319-10268-9

Lenntorp, B. (1999). Time-geography - At the end of its beginning. GeoJournal, 48(3), 155–158. https://doi.org/10.1023/A:1007067322523

Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. International Journal of Geographical Information Science, 27(2), 292–318. https://doi.org/10.1080/13658816.2012.682578

Mannberg, A., Hendrikx, J., Landrø, M., & Ahrland Stefan, M. (2018). Who's at risk in the backcountry? Effects of individual characteristics on hypothetical terrain choices. Journal of Environmental Psychology, 59(August), 46–53. https://doi.org/10.1016/j.jenvp.2018.08.004

Moody, G., (2018, 5 feb). The real Strava heatmap story is not threats to national security, but about privacy and de-anonymization. Privacy news online https://www.privateinternetaccess.com/blog/2018/02/real-strava-heatmap-story-not-threatsnational-security-privacy-de-anonymization/ retrieved on 25-09-2018

Marengo, D., Monaci, M. G., & Miceli, R. (2017). Winter recreationists' self-reported likelihood of skiing backcountry slopes: Investigating the role of situational factors, personal experiences with avalanches and sensation-seeking. Journal of Environmental Psychology, 49, 78–85. https://doi.org/10.1016/j.jenvp.2016.12.005

McCammon, I. (2002) Evidence of heuristic traps in recreational avalanche accidents. In: International snow science workshop, Penticton, BC, 30 Sept –4 Oct

McClung, D. M. (2002). The elements of applied avalanche forecasting part I: The human issues. Natural Hazards, 26(2), 111–129. https://doi.org/10.1023/A:1015665432221

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. Transportation Research Part A: Policy and Practice, 44(9), 754–765. https://doi.org/10.1016/j.tra.2010.07.008

Mooney, P., Olteanu-Raimond, A.-M., Touya, G., Juul, N., Alvanides, S., & Kerle, N. (2017). Considerations of Privacy , Ethics and Legal Issues in Volunteered Geographic Information. Mapping and the Citizen Sensor, 119–135. https://doi.org/https://doi.org/10.5334/bbf

Nathan, R., Getz, W., Revilla, E., & Holyoak, M. (2009). A movement ecology paradigm for unifying organismal movement research, (January). https://doi.org/10.1073/pnas.0800375105

Ogie, R. I., Clarke, R. J., Forehead, H., & Perez, P. (2018). Computers , Environment and Urban Systems Crowdsourced social media data for disaster management : Lessons from the PetaJakarta . org project. Computers, Environment and Urban Systems, (September), 0–1. https://doi.org/10.1016/j.compenvurbsys.2018.09.002

Oksanen, J., Bergman, C., Sainio, J., & Westerholm, J. (2015). Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. Journal of Transport Geography, 48, 135–144. https://doi.org/10.1016/j.jtrangeo.2015.09.001

Plank, A. (2016). 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, 289–296. https://doi.org/10.1016/j.tourman.2016.02.013

Peuquet, D. J. (1994). It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. Annals of the Association of American Geographers, 84(3), 441–461.

Purves, R. S., Laube, P., Buchin, M., & Speckmann, B. (2014). Moving beyond the point: an agenda for research in movement analysis with real data. Computers, Environment and Urban Systems, 47, 1–4. Retrieved from http://www.zora.uzh.ch/104614/1/2014_PurvesR_Moving_beyond_the_point_final.pdf

Ram, A., Jalal, S., Jalal, A. S., & Kumar, M. (2010). A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. International Journal of Computer Applications, 3(6), 1–4. https://doi.org/10.5120/739-1038

Schmudlach, G., Kohler, J. (2016). Method for an Automatized Avalanche Terrain Classification. In: Proceedings, International Snow Science Workshop, Breckenridge, Colorado, 2016

Schmudlach, G., Winkler, K., & Köhler, J. (2018). Quantitative risk reduction method (QRM), a data-driven avalanche risk estimator. In: Proceedings, International Snow Science Workshop, Innsbruck, Austria, 2018

Schneiderbauer, S., Ehrlich, D. (2004). Risk, Hazard and People's Vulnerability to Natural Hazards: a Review of Definitions, Concepts and Data .A review of definitions, concepts and data. Technical Report ·for the European Comission Joint Research Centre. EUR 21410 EN

Schönenberger, C. (2018). Analysis of Planned Route Trajectories to Gain Insights into Route Planning Behaviour for Backcountry Ski Tours. MSc Thesis. Department of Geography, Geocomputation. University of Zurich.

Schweizer, J., Bartelt, P., & van Herwijnen, A. (2014). Snow Avalanches. Snow and Ice-Related Hazards, Risks, and Disasters. https://doi.org/10.1016/B978-0-12-394849-6.00012-3

SLF, 2018. Avalanche formation. https://www.slf.ch/en/avalanches/avalanche-formation.html retrieved on 07-09-2018

Schweizer, J., & Jamieson, J. B. (2000). Field observations of skier-triggered avalanches. Proceedings International Snow Science Workshop, Big Sky, Montana, U.S.A., (October), 2–6.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., … Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. ISPRS International Journal of Geo-Information, 5(5), 55. https://doi.org/10.3390/ijgi5050055

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. International Journal of Geographical Information Science, 30(5), 881–906. https://doi.org/10.1080/13658816.2015.1100731

Ski Libre, description of Pigne d'Arolla ski tour via Col des Vignettes. https://www.ski-libre.com/topos/topos-ski/pigne-darolla-par-les-vignettes-3796m/ retrieved on 22-02-2019

Smith-Spark, L. & Messia, H. (2017). Italy avalanche: Hotel search ends with 29 dead, 11 rescued. CNN online: January 26, 2017. https://edition.cnn.com/2017/01/26/europe/italy-avalanche-hotel/index.html retrieved on 01-12-2018

Statham, G., McMahon, B., and Tomm, I.: The avalanche terrain exposure scale, in: Proceedings of the International Snow Science Workshop, Telluride, CO, USA, 491–497, 2006.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. GeoJournal, 78(2), 319–338. https://doi.org/10.1007/s10708-011-9438-2

Stephens, M. (2013). Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. GeoJournal, 78(6), 981–996. https://doi.org/10.1007/s10708-013-9492-z

Taczanowska, K., Bielański, M., González, L. M., Garcia-Massó, X., & Toca-Herrera, J. L. (2017). Analyzing spatial behavior of backcountry skiers in mountain protected areas combining GPS tracking and graph theory. Symmetry, 9(12), 1–15. https://doi.org/10.3390/sym9120317

Techel, F., Jarry, F., Kronthaler, G., Mitterer, S., Nairz, P., Pavšek, M., … Darms, G. (2016). Avalanche fatalities in the European Alps: Long-term trends and statistics. Geographica Helvetica, 71(2), 147–159.

https://doi.org/10.5194/gh-71-147-2016

Techel, F., & Zweifel, B. (2013). Recreational avalanche accidents in Switzerland: Trends and patterns with an emphasis on burial, rescue methods and avalanche danger. International Snow Science Workshop, (January), 1106–1112.

Techel, F., Zweifel, B., & Winkler, K. (2015). Analysis of avalanche risk factors in backcountry terrain based on usage frequency and accident data in Switzerland. Natural Hazards and Earth System Sciences, 15(9), 1985–1997. https://doi.org/10.5194/nhess-15-1985-2015

Tracz, D. (2012). Deep Snow Slab Avalanches. MSc thesis, Department of Engineering, University of Calgary.

Tremper, B. (2008), Staying Alive in Avalanche Terrain. Seattle: The Mountaineers Books

van Loenen, B., Kulk, S., & Ploeger, H. (2016). Data protection legislation: A very hungry caterpillar. The case of mapping data in the European Union. Government Information Quarterly, 33(2), 338–345. https://doi.org/10.1016/j.giq.2016.04.002

Turner, A.J. (2006). Introduction to Neogeography. O'Reilly Media, Inc: Sebastopol, CA

Van Oort, P. (2006). Spatial data quality: from description to application. Publications on Geodesy 60. Retrieved from http://library.wur.nl/WebQuery/wdab/1788022

Westin, A. (1967). Privacy and Freedom. New York, Atheneum

Zweifel, B., & Haegeli, P. (2014). A qualitative analysis of group formation, leadership and decision making in recreation groups traveling in avalanche terrain. Journal of Outdoor Recreation and Tourism, 5–6, 17–26. https://doi.org/10.1016/j.jort.2014.03.001

Appendix A. python modules used

The following python modules were used in this thesis:

Name	Purpose	Link
Matplotlib	Data visualization	https://github.com/matplotlib/matplotlib
Numpy	Processing large arrays	https://github.com/numpy/numpy
Scikit-learn	Clustering, scaling data, PCA	https://github.com/scikit-learn/scikit-learn
Trajectory-distance	Computing Fréchet distance	https://github.com/maikol-solis/trajectory_distance/tree/master/traj_dist/pydist
csv	Reading csv files	https://docs.python.org/2/library/csv.html
codecs	Reading csv files	https://docs.python.org/3/library/codecs.html
pandas	Reading data frames	https://github.com/pandas-dev/pandas
scipy	Computing descriptive statistics	https://github.com/scipy/scipy
random	Creating random spatial cluster numbers	https://docs.python.org/3/library/random.html
pyshp	Reading route .shp data	https://github.com/GeospatialPython/pyshp

Table 14. *python modules used*