Universität Zürich[UZH]

Geographisches Institut

Enhancing Snow Avalanche Forecasting: Developing User-Centered Dashboards for Data Visualization and Decision Support

ESS 511 Master's Thesis

Author: Nils Besson, 20-710-729

Supervised by: Dr. Frank Techel (frank.techel@slf.ch)

Faculty representative: Prof. Dr. Ross Purves

26.08.2025

Acknowledgement

I am deeply grateful to my supervisors, Prof. Dr. Ross Purves of the University of Zurich (UZH) and Dr. Frank Techel of the Swiss Institute for Snow and Avalanche Research (SLF). Their expert guidance, constructive feedback, and unwavering support have been invaluable at every stage of this research.

I also extend my sincere thanks to Thomas Werschlein from the IT department at the Geographical Institute of UZH. His assistance in setting up and maintaining the virtual server environment for the user tests ensured that the technical aspects of this project ran smoothly.

My appreciation goes out to all the participants who volunteered for the survey. Their knowledge, generous time investment, and thoughtful feedback were essential to the validation process and greatly improved the quality of this work.

Finally, I would like to thank my family, my girlfriend, and my friends for their constant encouragement, practical help, and well-timed distractions. Your support sustained me throughout the thesis journey, and I could not have done this without you.

Abstract

Avalanche forecasting remains a cognitively demanding task, requiring experts to integrate heterogeneous data sources under time pressure. While advances in numerical modelling and monitoring networks have expanded the available information, tools that effectively combine and visualise these data sets for operational use are still limited.

This thesis develops and evaluates a prototype dashboard designed to support avalanche forecasters through integrated, interactive visualisation of snowpack, meteorological, and avalanche data. The system brings together interpolated grid fields, point-based measurements, and clustering results from an ensemble HDBSCAN* approach, presented within a multi-tab layout (Cluster, Point, and Grid views). The design follows established geovisualisation principles, emphasising spatio-temporal context, intuitive filtering, and consistency across representations.

To assess the dashboard's effectiveness, I conducted a structured user study including performance tasks, ease-of-use ratings, and a System Usability Scale (SUS) survey. Results show that grid data were interpreted most intuitively, while point data offered detailed accuracy at the cost of higher cognitive load. Cluster-based representations provided a useful overview but were limited by the modest quality of the underlying ensemble HDBSCAN* clustering (Adjusted Rand Index = 0.28). Despite these challenges, participants rated the dashboard as straightforward and coherent, and reported that its interactive design supported efficient pattern recognition.

The main contribution of this work lies in demonstrating how multiple avalanche-relevant data sources can be integrated into a single, user-centred dashboard concept. Beyond its immediate application, the study highlights both the opportunities and limitations of applying ensemble clustering in this context, and establishes a benchmark for future research. The approach and design principles are transferable to other natural-hazard domains where experts must balance complex, uncertain data under time pressure.

Keywords: Snow avalanche forecasting, Data visualisation, User-centred dashboard, Geovisualisation, Cognitive load, IMIS sensor network, grid data, Point-based data, HDBSCAN* clustering, Usability evaluation (SUS)

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Switzerland's terrain consists of three main parts: the Swiss plateau, the Jura, and the Alps, with the alpine region comprising 58% of the area (Schweizerische Eidgenossenschaft, 2023). During the winter, many Swiss and international guests gather in the alpine region for winter activities like hiking, snowboarding, or alpine skiing. This period also corresponds to the highest probability of snow avalanches.

Accurate snow avalanche forecasting is crucial for enhancing safety and mitigating risks in mountainous areas such as the Swiss Alps, where avalanches pose a significant threat to human life, infrastructure, and economic activities (Statham et al., 2018). Avalanche forecasting presents considerable challenges, including processing vast amounts of environmental data, identifying critical parameters, and delivering actionable insights in a user-friendly format. This thesis addresses these challenges by developing an advanced, interactive dashboard tailored to the needs of snow avalanche specialists (Harbola & Coors, 2018).

Avalanche forecasting requires integrating meteorological data, snowpack properties, and terrain variations, making it a complex endeavour (Floyer et al., 2016; Sovilla et al., 2007; Wever et al., 2016). Traditional forecasting methods, while effective, rely heavily on human interpretation of large volumes of raw data and indices, which can be both time-consuming and inefficient. Advancements in computational tools, such as machine learning (ML) algorithms and physically based models like SNOWPACK, have significantly improved prediction accuracy and resolution. However, fully automating these models remains challenging, with final decisions still relying on human expertise (Pérez-Guillén et al., 2022; Techel et al., 2024).

According to the SLF (White Risk, n.d.), the spatial occurrence of snow avalanches is predictable, as forecasting experts know from experience which mountain faces are particularly prone to avalanches. The challenge in avalanche forecasting is determining precisely when an avalanche will occur rather than where. To assess avalanche occurrences in the Jura and Alpine regions, SLF forecasting experts analyse environmental data gathered by the Intercantonal Measurement and Information System stations (IMIS), weather predictions, and incorporate field tests assessing snowpack stability to predict avalanche danger (Mayer et al., 2023; Rapisarda & Pranzo, 2021; Steinkogler et al., 2014).

The avalanche forecasting process involves analysing an enormous amount of data. Experts must produce two bulletins per day, resulting in a high cognitive load and increased potential for errors (Van Herwijnen et al., 2023). This thesis aims to

decrease cognitive load for experts by simplifying current tools and combining all data sets into a central, user-centred dashboard, streamlining analysis and interpretation processes (Van Wijk, 2005).

The proposed dashboard aims to enhance usability and precision, allowing experts to extract critical insights effectively. Precision is essential, as experts evaluate avalanche risks across five main categories (1-5) and three subcategories (+, - and =). Key research objectives include optimising data representation, identifying critical features for hazard analysis, and developing intuitive visualisation methods that simplify complex computations while minimising cognitive load (Young & Kitchin, 2020).

The dashboard facilitates data-driven decisions, offering experts timely insights through a centralised, interactive, and visually engaging interface. This work has broader implications for decision-making in avalanche hazard assessment and could inform similar efforts in other domains of natural hazard management (Hallandvik et al., 2017; Horton et al., 2020; Kunz et al., 2010; Maissen et al., 2024).

These challenges highlight the need for improved approaches to avalanche forecasting, which motivates the research gap discussed in the next section.

1.2 Research Gap

Avalanche forecasting currently requires experts to manually interpret large volumes of observational and model-based data, updated as frequently as every 30 minutes. The need to produce multiple daily bulletins under time pressure imposes a significant cognitive load on forecasters, increasing the risk of fatigue-related errors and reducing overall decision-making efficiency (McClung & Hungr, 1986).

Despite advances in computational tools, including physically based models like SNOWPACK and ML approaches proposed in Pérez-Guillén et al. (2022) or in Fromm and Schönberger (2022), forecasting workflows still depend heavily on manual data inspection and human judgment. These systems often lack user-friendly interfaces that could consolidate, prioritise, and communicate the most critical information (Van Wijk, 2005).

A clear research gap exists in the development of integrated, user-centred dashboards that support snow avalanche forecasting by streamlining data visualisation and interpretation. While some geovisualisation tools exist (e.g., Horton et al. (2020) and Kunz et al. (2010)), there is limited work on operational interfaces designed specifically to reduce cognitive load and enhance expert workflows in this context.

This thesis addresses that gap by designing and evaluating an interactive dashboard

tailored to the needs of avalanche forecasters. The platform aims to reduce mental effort, improve risk communication, and support more consistent, data-driven decisions in complex terrain.

1.2.1 Research Questions

This section outlines the primary aim and research questions guiding this research, focusing on optimising data representation, identifying critical attributes for avalanche forecasting, and proposing data visualisations tailored to expert users. These efforts ultimately aim to enhance decision-making and improve the utilisation and accuracy of snow and avalanche forecasting tools.

This thesis explores the following principal research questions:

1. **Integrating Predictive Analytics with Expert Judgment**: How can machine-learning-driven risk predictions be blended into a dashboard so as to support expert decision-making in avalanche forecasting?
    - Determine the most essential information required for effective snow and avalanche risk assessment, prioritising attributes that provide the greatest predictive and analytical value.
2. **Optimised Data visualisation for Expert Use**: How can critical snow and avalanche data be presented in a clean, intuitive, and user-centred manner, ensuring clarity and ease of interpretation for forecasting experts?
    - Which visualisation and interactive design techniques best enhance clarity, usability, and minimise cognitive load when presenting snow and avalanche data to forecasting experts?
3. **Comparative Analysis of Data Representation**: What are the differences in interpretability between point data, grid data, and cluster data, with a focus on their respective contributions to snow and avalanche forecasting accuracy?
    - What are the advantages and limitations of each data representation method in terms of interpretability for avalanche hazard assessment?

2 State of the Art

This chapter reviews the current state of research relevant to avalanche forecasting and visualisation. It begins with geovisualisation techniques used in natural hazard contexts, before focusing on their application to snow avalanches. Next, it outlines the processes behind avalanche formation and forecasting, highlighting how models, field data, and expert judgment interact in practice. Together, these insights establish the scientific and practical foundation for the dashboard design developed in this thesis.

2.1 Geovisualisation

This section introduces the role of geovisualisation in natural hazard forecasting, with a focus on snow avalanches. It outlines key techniques and data structures used to communicate spatial risk and uncertainty and presents dashboard designs that support decision-making in dynamic environments.

2.1.1 Geovisualisation for Natural Hazards

This subsection provides a general overview of geovisualisation techniques in the context of natural hazard assessment. It describes visual encoding strategies for risk, methods to display uncertainty, and dashboard architectures used to integrate and present this information effectively (Kehrer, 2013; Tory & Möller, 2004).

Visualisation of Patterns and Risks "The ultimate goal of geovisual analytics is to support analytical reasoning about complex spatio-temporal problems by enabling and stimulating human interaction with dynamic maps and visual interfaces" (Andrienko et al., 2016, p. 3). This makes geovisualisation a central tool for interpreting risk patterns in natural hazard forecasting.

Experts must quickly distinguish threat from non-threat scenarios and identify both the intensity and precise location of danger under time pressure (Sweller et al., 1998; Tory & Möller, 2004). Clear, intuitive, and time-sensitive visualisation is therefore essential.

Common techniques include heat maps and kernel density estimation (KDE), which reveal spatial patterns in hazard intensity (Marinova et al., 2018). While KDE smooths discrete data into continuous surfaces, it may oversimplify terrain-specific features (Andrienko et al., 2008; Horton et al., 2020).

Temporal animations, such as precipitation sequences, allow the user to explore change over time using sliders or step controls (Heer & Robertson, 2007; Rapisarda & Pranzo, 2021). Class-based differences like avalanche danger levels are often shown using layered maps and intuitive colour encodings (EAWS - European Avalanche Warning Services, 2018; Horton et al., 2020; MacEachren & Kraak, 2001). Perceptually balanced schemes reduce misinterpretation and cognitive load (Brewer, 2005).

These techniques provide a toolbox for communicating spatial and temporal risk variation, but they always involve trade-offs between clarity, accuracy, and cognitive effort.

Uncertainty Representation Every visualisation of a natural hazard prediction or forecast always contains uncertainty, whether it is visualised or not. Therefore, when displaying the inherent uncertainty of a model on the output visualisation, overconfidence in the map is avoided and also supports decision-makers in their process.

The type of uncertainty and the amount of possible uncertainty varies heavily between models and also natural hazard types. For example, model uncertainty is when there are multiple possible forecasts for the same hazard type, or data uncertainty is when there are missing data values and the models are nevertheless created with less information than they should have, or spatio-temporal uncertainty is when the time and the location of the natural hazard events are not precise (Keim et al., 2008; Kunz et al., 2010).

To visualise uncertainty there are multiple ways to do so:

- Transparency/Fuzziness: Uncertainty is shown with blurry or faint forecast boundaries; the more transparent, the less certain the data.
- Hatching or Pattern Overlays: Striped or patterned areas highlight regions with low confidence in the forecast.
- Ensemble Footprints: Multiple overlapping lines or shaded zones from different forecast runs illustrate variability in predictions.
- Error Bars or Confidence Intervals: Vertical or horizontal lines around data points in a time series indicate the range of possible values.

One problem with the inclusion of the uncertainty visualisation is that the more detailed the uncertainty information is, the more it may confuse the user or undermine the significance of the visualisation. A commonly used approach is to include

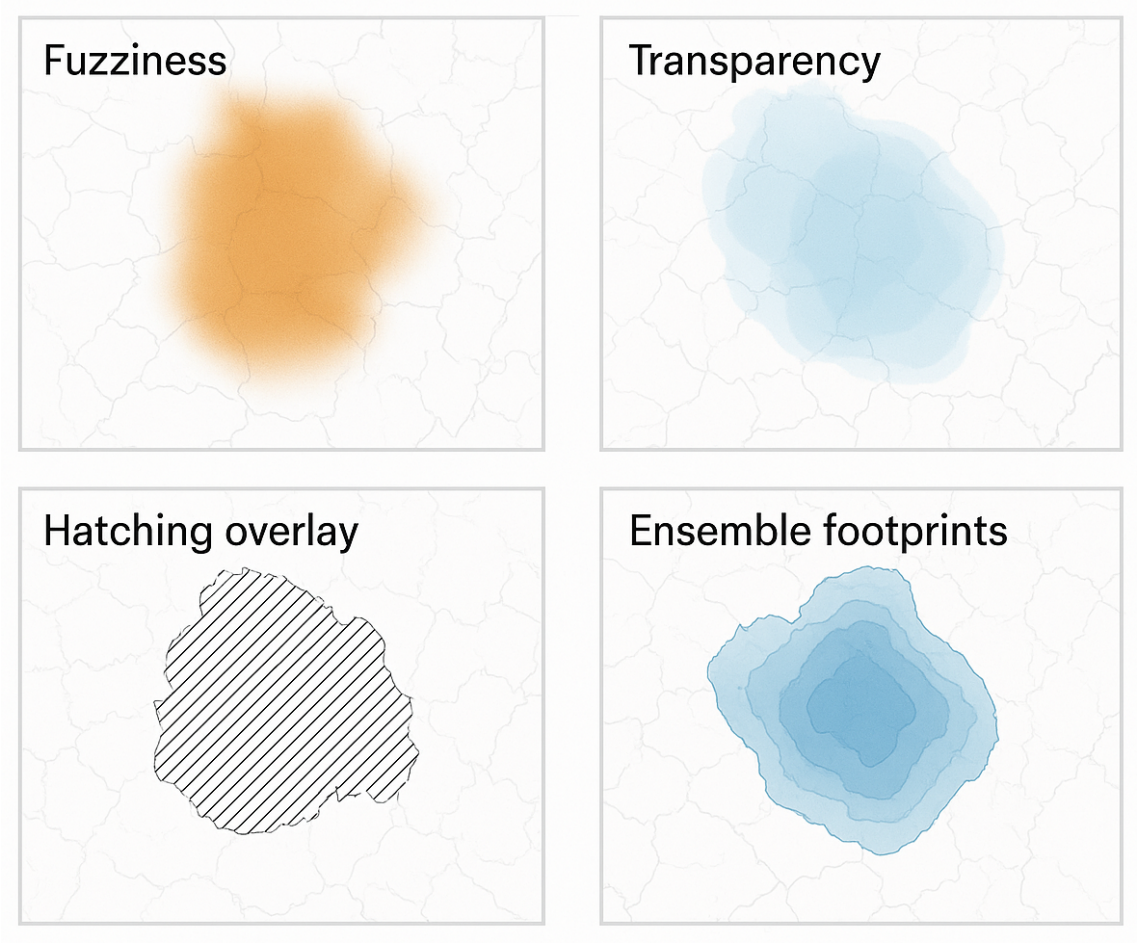


Figure 1: Conceptual comparison of four visual techniques for representing spatial uncertainty in avalanche forecasting. The methods, fuzziness, transparency, hatching overlay, and ensemble footprints, are illustrated using a synthetic hotspot to demonstrate different ways uncertainty can be encoded in map design. Image generated using Sora (AI image generation tool) for illustrative purposes.

uncertainty, when the map is used for experts, but not when it is published to the public (MacEachren & Kraak, 2001; Pohl et al., 2012).

Visualising uncertainty is not only a technical challenge, but a communication one; it requires a thoughtful balance between completeness and clarity.

Dashboard Architectures Dashboard architecture can be summarised as the organisational structure of components and data pipelines that combine into one platform to visualise data in an organised manner.

For a dashboard to be suitable for natural hazard forecasting, it has to be able to combine spatial, temporal and statistical data, as well as support quick decision-making processes in dynamic environments (Saunders et al., 2023). A core component is the modularity, interactivity and real-time responsiveness of the platform so that the user is not hindered in the working process (Badam et al., 2022; Horton et al., 2020; Young & Kitchin, 2020).

Common parts of such platforms are of course the map, which should be situated in the centre of the dashboard. The map should be interactive, that the user can zoom, pan or toggle layers or overlays (Badam et al., 2022; Horton et al., 2020).

There should also be a control panel, where the thresholds, dates and visualised data can be changed, or time-series and animations can be controlled (Badam et al., 2022). Another common addition is to add a panel, where further information, such as ensemble plots, metadata of the visualised data, or time-series graphs, is visualised (Badam et al., 2022).

The design of these dashboards is mostly a compromise. The design has to balance data complexity, user clarity, flexibility, and focus. It should be designed in such a way that the cognitive load of the user is decreased and the insight into the data maximised in comparison to when they would analyse the data without such a tool (Horton et al., 2020; Young & Kitchin, 2020).

2.1.2 Visualising Prognosis Data

This subsection presents techniques for displaying forecast data in intuitive and interactive ways. It highlights how temporal dynamics, ensemble forecasts, and filtering tools enable the user to explore potential hazard scenarios and focus on relevant conditions.

Temporal Interactions The environment is constantly changing, and with it, the potential hazard for a natural hazard event to occur. Thus, it is crucial that

the forecasters have environmental- and forecasting data at their disposal. That the hazard assessment can be completed accurately, the data has to be of good enough temporal resolution (Andrienko et al., 2016; Heer & Robertson, 2007; Horton et al., 2020).

Like previously mentioned, time-series or animations are a good way to display prognosis data. Controls like time sliders let the user scroll through the visualisations to analyse them as exactly as possible. These tools and controls help the user to recognise common patterns and anticipate potential peak risk times. For animations, options like "playback" or a pause button can be very helpful to investigate specific parts of the animations (Heer & Robertson, 2007; Horton et al., 2020; MacEachren & Kraak, 2001).

To compare different forecasting model results, a small multiples approach can be used (see Figure 2). With the side-by-side view of the different results, the differences can be easily identified and the correct conclusions or assessments can be drawn from them (Andrienko et al., 2010; Horton et al., 2020).

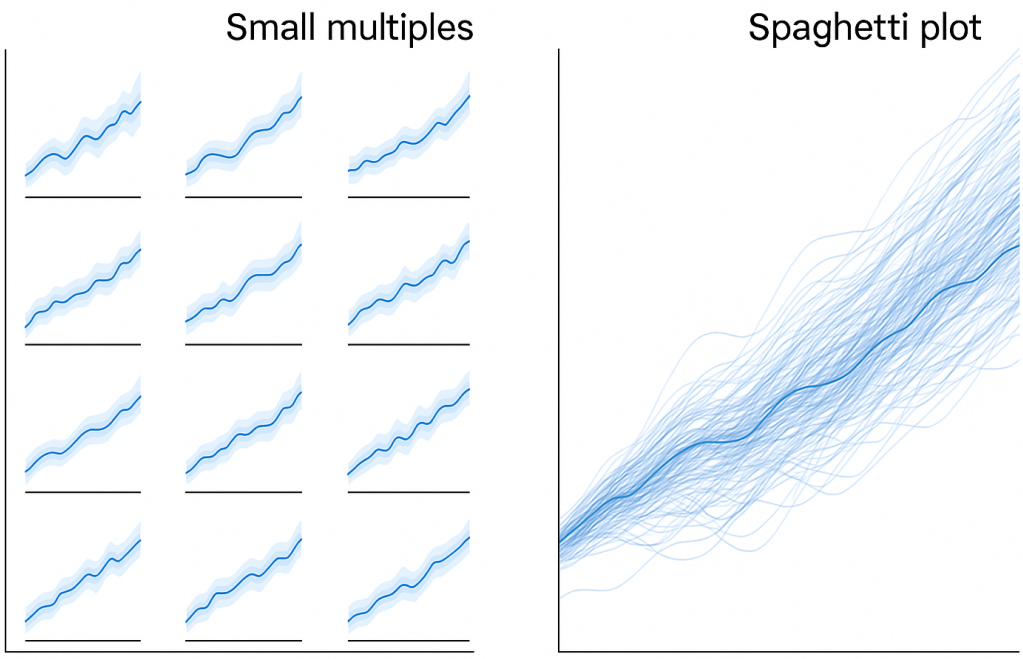


Figure 2: Comparison of small multiples and spaghetti plots for visualising multiple time series. Small multiples offer clarity by separating series, while spaghetti plots overlay them, increasing visual complexity. Image generated using Sora (AI image generation tool).

Ensemble Forecast Displays Often, a single forecasting visualisation cannot capture potential uncertainty. Therefore, ensemble forecasts, where multiple forecasting solutions are overlaid (see Figure 1), are a good way to represent the range of possible forecasts. In avalanche forecasting, this can be especially useful, as small environmental variations can lead to different outcomes (Horton et al., 2020; Kunz et al., 2010).

In ensemble forecasting, multiple semi-transparent forecasts from multiple forecasting model runs are overlaid. Where the colour is most intense, there is the most agreement between the model runs, and the probability for a correct forecasting result is also higher (Horton et al., 2020; Kunz et al., 2010).

When the available variables are line-based (snowline elevation, temperature per location), a "spaghetti plot" can be used (see Figure 2). In this visualisation, each model run is represented by a single line. This format makes outliers easy to spot, but without filtering or smoothing, the plot can be noisy and therefore requires a better understanding of the underlying data to be interpreted reliably (Andrienko et al., 2010; Keim et al., 2008).

Interactive Forecast Filtering All the collected and calculated data are often too much data for the forecasters to analyse at one sight. For that reason, filtering is very important. With filtering enabled, the user can focus on relevant information, reduce the cognitive load for the hazard assessment process and also support scenario-based exploration (Badam et al., 2022; Horton et al., 2020).

There are a number of different applications a filter can have. It could be just a basic filter, where the date, or the attribute can be filtered which is to be visualised. There are also advanced filters, where the user can set upper- or lower-bound thresholds for the visualised values. These filters can either be with a text box (for threshold settings), a slider, or a toggle switch (Badam et al., 2022; Horton et al., 2020).

To enable the user for the filtering process, the options and controls should be visible and intuitive. Further, the response to an action should be immediate and should be reversible and combinable with different actions (Badam et al., 2022; Diehl et al., 2018; Young & Kitchin, 2020).

2.1.3 Field-Based Dashboards

Field-based dashboards are tailored to visualise continuous environmental data over complex terrains. This subsection examines their structure, performance optimisation strategies, and their suitability for operational avalanche forecasting. It concludes with evaluation approaches to ensure usability and reliability.

Field- vs. Object-Based Approaches Dashboards in the field of geovisualisation are generally either field-based or object-based. Field-based approaches have continuous data surfaces, while object-based dashboards have discrete geographic features. These two conceptual approaches differ in how the data is visualised (grids vs. points, rasters vs. lines, interpolated surfaces vs. polygons), what interactions are possible, and how the user can interpret the spatial structure of hazards (Floyer et al., 2016; Harbola & Coors, 2018; Horton et al., 2020).

Table 1: Comparison of field-based versus object-based dashboard designs for avalanche hazard forecasting, highlighting their respective advantages and limitations for operational use.

Dashboard Type	**Data Sources**	**visualisations**	**Description and Application**
Field-Based Dashboards	IMIS station network, SNOWPACK output	Grids, Rasters, Interpolated surfaces	Continuous fields over spatial domains (e.g., snow height, temperature, wind speed, snowpack stability). Ideal for avalanche forecasting due to the relevance of terrain-following gradients. Enables layer overlays and temporal comparisons (Horton et al., 2020; Pérez-Guillén et al., 2022).
Object-Based Dashboards	GIS-based discrete entities (e.g., buildings, roads, flood zones, administrative units)	Points, Lines, Polygons	Distinct, discrete entities typically used in urban or emergency services dashboards. Examples include showing buildings at flood risk. Generally less suited to snowpack forecasting due to dominance of terrain-based processes (Cappabianca et al., 2008; Harbola & Coors, 2018).

In the context of snow-avalanche forecasting, a field-based dashboard remains a highly effective way to explore spatially continuous variables, since avalanche activity is strongly terrain-dependent and can vary dramatically with elevation, slope and aspect (Horton et al., 2020; Pérez-Guillén et al., 2022). However, many critical observations, such as individual avalanche occurrences or snow stability test results, are inherently discrete, point-based measurements that cannot be reliably interpolated across the landscape. For this type of data, object- or point-based visualisations (e.g. map markers, scatter plots, or profile plots tied to specific GPS locations) provide the

most accurate representation. Depending on the nature and density of each data source, a hybrid approach, using a field-based dashboard for continuous layers and complementary point-based widgets for discrete observations, ensures both accuracy and usability.

Components of a Field-Based Dashboard Field-based dashboards for avalanche forecasting are based on a modular layout, which was already presented in Section 2.1.1. The most important components include a terrain-related map view, a control panel for temporal and variable filtering, and a detail window for location-specific diagnostics. In the context of avalanches, these modules are designed to support the interpretation of the snowpack, filtering by altitude, and the integration of real-time models (Floyer et al., 2016; Horton et al., 2020; Wood et al., 2007).

Unlike object-based dashboards, the field-based setup emphasises fluid interaction with continuous spatial data. Interactivity, such as moving the mouse pointer over interpolated snow stability layers or animating forecast steps, helps reduce cognitive load and supports expert workflows.

Performance and Data Management Because the user of these operational dashboards has to use them under time-sensitive conditions, it is essential that the rendering time is low, the interactions are as smooth as possible, and the visualised information is up-to-date (Mohammed et al., 2020). This requires that the management of data volume, rendering methods, and the server be treated as important (Badam et al., 2022; Kalamaras et al., 2018).

To keep the rendering time as low as possible, it is often done that the raster layers (mostly base layers and historic data) are pre-rendered into map tiles for a faster loading time. Thus, with the pre-rendered layers, the user is less flexible for real-time updates of the data (Diehl et al., 2018; Harbola & Coors, 2018).

In cases where near-real-time data (e.g., 30-minute IMIS updates) must be displayed immediately, on-the-fly interpolation becomes necessary. Although this approach is computationally heavier than pre-rendered tiles, its flexibility is essential for integrating the newest available data. Performance can be preserved through optimisations like resolution throttling or viewport-specific computation (Andrienko et al., 2016; Horton et al., 2020).

If, for example, an on-the-fly interpolation is used, the rendering time is much higher than aimed for. To improve the performance of the dashboard, clipping can be used. This feature only renders the part of the visualisation that is currently shown. That decreases the general rendering time per "image", but also implies new

rendering times when the user pans, zooms, or interacts with the map in any other way (Andrienko et al., 2016; Ware, 2013). Another way to improve the performance of the dashboard is via debounced user interface (UI) inputs. This prevents rendering of the visualisation when the user is still interacting with a feature (a slider for example) and only renders when the user has stopped interacting with the feature (Badam et al., 2022; Kalamaras et al., 2018).

To facilitate these performance-enhancing measures, a real-time data pipeline is almost needed. With new data gathered automatically from the IMIS stations, operating the data preparation manually would be inefficient. The application programming interface (API) structure of the dashboard also needs to support efficient data queries, a smart caching behaviour, and potentially a scaling hosting for multi-user access (Young & Kitchin, 2020).

In summary, geovisualisation plays a central role in translating complex environmental data into actionable insights for avalanche forecasting. From foundational data inputs and spatial display methods to the integration of uncertainty and forecasting layers, effective visualisation is crucial for supporting expert judgment under time pressure. Field-based dashboards, in particular, offer a powerful and scalable solution for synthesising spatio-temporal data in mountainous terrain. These tools not only enhance clarity and efficiency in risk assessment but also represent a critical bridge between raw model output and expert decision-making.

2.2 Avalanche formation

To predict the occurrence of avalanches accurately, the forecasting experts have to understand how avalanches form in the terrain, how weak layers within the snowpack form, and what can release an avalanche. This will be discussed in the following section.

2.2.1 Weak-Layer Formation and Properties

Avalanche release begins with the initiation of failure in a weak layer of the snowpack (McClung & Schaerer, 2006; Schweizer et al., 2003). This section involves understanding how stress, snowpack structure, and external triggers interact to create initial fractures. The following subsection outlines the key mechanical conditions that make such failures possible.

The central issue in avalanche formation lies in the stability of the individual snow layers within a snowpack. Metamorphic processes, such as near-surface faceting (for

example, radiation recrystallisation), faceting adjacent to wet layers, temperature-gradient effects, and surface hoar deposition, can transform these layers and, in doing so, significantly alter their stability (Schweizer et al., 2003).

According to Schweizer et al. (2003, 2015) there are two different types of weak layers that can form within a snowpack: non-persistent weak layers and persistent weak layers.

Non-persistent weak layers develop from freshly fallen snow whose grains are loosely packed and poorly bonded. For the first few days after burial, these layers remain unstable and can act as sliding planes within the snowpack. During this brief window, even modest loads or disturbances may cause them to fail. However, as time passes and temperature and pressure allow the grains to sinter and bond more firmly, the layer consolidates and no longer exhibits that early weakness (Schweizer et al., 2003, 2015).

Persistent weak layers, in contrast, are made up of large, angular snow grains, typically originating as surface hoar or depth hoar. These crystals grow under strong temperature gradients within the snowpack, creating a structure that is slow to bond and consolidate. Because the grains remain faceted rather than rounding and sintering together, the layer retains its weakness over an extended period. As a result, these persistent weak layers can pose an avalanche hazard for weeks or even months after their initial formation, only gradually stabilising once environmental conditions allow the grains to metamorphose into more cohesive, rounded shapes (Schweizer et al., 2003, 2015).

The thickness of a weak layer also plays a crucial role in snowpack stability: thicker layers tend to contain larger grain facets and exhibit lower bond densities, which makes them more prone to crack initiation (Schweizer & Jamieson, 2001). Equally important are the stiffness and density of the overlying slab-properties that can be increased by wind packing or freeze-thaw cycles, because a stiffer slab more efficiently transmits stress into the weaker layers below, thereby promoting both fracture initiation and propagation (Sigrist & Schweizer, 2007). Finally, external environmental drivers such as intense solar radiation and steep temperature gradients accelerate the metamorphic transformations that give rise to faceted weak layers, further compromising snowpack integrity (Schweizer et al., 2003).

2.2.2 Failure Initiation Mechanics

Once failure is initiated, the next critical process is the propagation of cracks through the weak layer. Whether a small fracture develops into a full slab avalanche depends on how efficiently stress is redistributed and fractures spread (Schweizer et al., 2003).

This subsection explains the main factors controlling crack propagation and its role in avalanche release.

According to Schweizer et al. (2003) and McClung and Hungr (1986) there are three main failure mechanics:

- shear failure: A weak snow layer collapses when sideways stress exceeds its strength.
- tensile failure: A crack forms when pulling forces exceed the weak layer's tensile strength.
- loss of cohesion: Snow grains lose their grip, start sliding downhill, and pick up more snow.

Shear failure happens when a cohesive slab of snow rests on top of a weaker, less bonded layer. Gravity pulls both layers downhill, and if the gravitational force exceeds the weak layer's shear strength, a fracture develops along a nearly horizontal surface. The likelihood of such shear fractures is closely tied to the underlying snow's grain type and density (McClung & Hungr, 1986; Schweizer et al., 2003).

After a shear failure occurs in the snowpack, the fracture typically propagates upslope through the weak layer towards the crown area of the slab (McClung & Hungr, 1986). As the fracture spreads, stresses redistribute around the slab, particularly near its upper edge. Once the fracture reaches this crown area, the slab must detach from the undisturbed snow uphill. This detachment occurs through tensile failure, which happens perpendicular to the slope (and therefore also perpendicular to the direction of gravitational force). Tensile failure pulls the slab apart across the crown line, allowing the slab to fully release and slide downhill (McClung & Hungr, 1986; Schweizer et al., 2003).

The third described "failure mechanism", as seen in loose snow avalanches, is not a failure in the classical fracture mechanics sense. There is no defined weak layer or cohesive slab involved. Instead, the release occurs when the cohesion within the surface snow becomes insufficient to support its own weight on a steep slope. This loss of cohesion is not triggered by a discrete fracture event, but rather arises gradually due to factors like snow type, temperature, and slope angle. Once a small amount of snow begins to move, it can entrain more snow downslope, resulting in a gravity-driven point release (McClung & Hungr, 1986).

Another key failure mechanism involves initiation by external triggers, most commonly human interaction with the snowpack. When a skier or snowboarder crosses a slope, their weight and dynamic movements concentrate extra stress on a buried weak

layer. If that added load breaks the fragile bonds at one point, a fracture can quickly propagate through the surrounding snow. As the crack spreads across the slab, the overlying mass can detach and rush downhill in an avalanche (Schweizer et al., 2003; Van Herwijnen et al., 2023).

2.2.3 Avalanche Types

Avalanches can be categorised into different types depending on the snowpack structure, release mechanism, and flow characteristics. Distinguishing between these types is essential for both scientific analysis and operational forecasting, since they vary in frequency, predictability, and potential impact. This subsection introduces the main avalanche types relevant to hazard assessment.

The primary distinction is between dry-snow and wet-snow avalanches. Dry-snow avalanches, including slab and loose snow types, occur in cold conditions and are typically more destructive due to their speed and mass. Wet-snow avalanches occur when the snowpack is saturated with liquid water, leading to slower but denser flows (McClung & Hungr, 1986; Schweizer et al., 2003).

A further distinction is made between spontaneous avalanches, triggered without external influence (e.g., due to rapid loading or warming), and human-triggered avalanches, which result from localised disturbances such as backcountry travel. For operational forecasting, identifying the most probable avalanche type is critical for estimating danger levels and communicating terrain-specific risks (EAWS - European Avalanche Warning Services, 2018; Statham et al., 2018).

Overall, avalanche formation results from the interaction of snowpack stratigraphy, the development of weak layers, and external forcing such as snowfall, wind, or warming. These processes jointly determine when and how avalanches are released, and they explain why similar conditions may produce very different outcomes in different places. By outlining the fundamental mechanics behind avalanche release, this section establishes the physical basis on which forecasting approaches and hazard assessments rely.

2.3 Avalanche Forecasting-Process-Chain

This subsection describes how avalanche danger is operationally assessed. It introduces the forecasting process as a sequence of steps, from gathering input data to producing danger levels and bulletins. Each step highlights the balance between model outputs, field observations, and expert interpretation.

The avalanche forecasting workflow at the SLF follows a structured multi-step process, combining real-time data acquisition, model-driven analysis, expert judgment, and public communication. This forecasting chain ensures a reliable and traceable assessment of avalanche risk throughout the Swiss Alps.

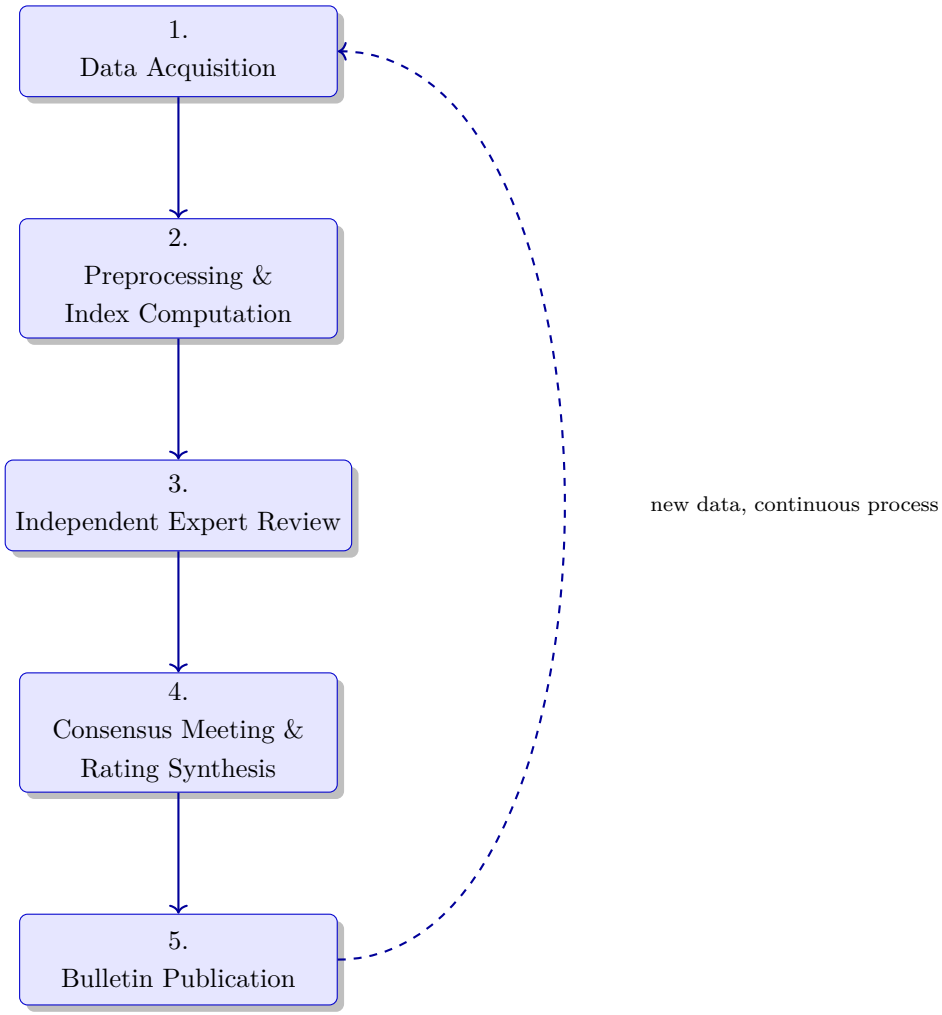


Figure 3: SLF avalanche forecasting process chain. The five-step workflow includes: (1) environmental data collection, (2) automated preprocessing and index computation, (3) expert assessment, (4) consensus forecast synthesis, and (5) public bulletin publication. The process restarts continuously as new data become available.

2.3.1 Forecasting Workflow and Bulletin Structure

This section describes the SLF's end-to-end workflow for producing twice-daily avalanche bulletins. Each stage, from IMIS sensor data collection and automated stability index computation, through expert review and consensus discussion, to the publication of colour-coded danger maps and regional reports, is covered. This overview highlights how data-driven models and human expertise integrate to deliver timely, actionable forecasts.

Data Acquisition The IMIS station network collects environmental measurements automatically twice per hour, including snow depth, air temperature, wind speed and direction, and humidity, and transmits them to a central database for processing (Pérez-Guillén et al., 2022; WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-b).

Each reading is tagged with metadata such as sensor type, elevation, exposure, and timestamp, ensuring proper alignment and validation during downstream processing (Stucki, n.d.; WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-b).

The IMIS network was designed specifically for avalanche monitoring in alpine terrain, with stations distributed across various elevations, aspects, and slope types (Pérez-Guillén et al., 2022). In addition to IMIS, further data inputs include automatic weather stations (AWS), snow profiles, manual field reports and observations, physical snowpack model outputs, and ML model outputs, which help to cross-validate and contextualise automated readings (Schweizer et al., 2021).

Preprocessing & Index Computation Upon arrival, the raw data undergoes automated quality control. Outliers are detected (e.g., extreme jumps in values from neighbouring stations), missing values are imputed, and sensor drift is corrected using spatio-temporal comparisons with nearby stations (Lehning et al., 1999; Pérez-Guillén et al., 2022).

The cleaned data is then passed into statistical and ML modules, which derive indices such as weak layer instability, probability of avalanche release, and the likelihood of falling into a specific avalanche danger level (Lehning et al., 1999; Schweizer et al., 2003). These indices serve as intermediate diagnostic layers, quantifying, for example, the probability of a critical weak layer forming, the likelihood of slab release, or thresholds for new snow accumulation. Many are derived using SNOWPACK outputs, either directly or as features in statistical models. By computing values on a standardised grid, forecasters can more easily assess spatial patterns and identify instability "hotspots." The outputs are interpolated onto a 1×1 km grid to offer spatial consistency across the Alpine domain, aiding in regional comparisons across the Alpine region (Pérez-Guillén et al., 2022).

Independent Expert Review The gridded indices and original IMIS station data are then reviewed independently by two to four SLF forecasters. Each expert conducts a separate analysis, evaluating patterns in the derived indices, sensor readings, and meteorological forecasts (e.g., COSMO-1, European Centre for Medium-Range Weather Forecasts (ECMWF) model outputs) (Pérez-Guillén et al., 2022; Statham et al., 2018; Stucki, n.d.).

While the indices and model predictions provide objective input, expert analysis remains essential for accounting for local context, atypical snowpack evolution, and edge cases. Forecasters apply heuristic knowledge, compare past and present conditions, and interpret subtle patterns that automated systems may miss. This human-in-the-loop approach ensures both consistency and critical oversight (Hartl, 2025).

Based on this assessment, they identify critical areas and assign a preliminary avalanche danger level (1-5). An avalanche danger scale of 1 represents low danger, moderate (danger level of 2), considerable (danger level of 3), high (danger level of 4) and very high (danger level of 5). The danger level 5 is not used very often; that is why it is mostly combined with the danger level 4 (high).

The European Avalanche Warning Services (EAWS) scale is often supplemented by modifiers (+, - and =) to express confidence and nuances within a given level (EAWS - European Avalanche Warning Services, 2018). The "+" indicates a higher than average danger threat for that level and also tending towards the higher danger level, the "-" suggests that the danger is on the lower end of the current level and the conditions are stable on that level. And lastly "=" indicates a stable situation, where the danger is an average for that threat level (EAWS - European Avalanche Warning Services, 2018).

Consensus Meeting & Rating Synthesis Following individual assessments, the forecasters meet to compare results and reconcile any differences. Through discussion and consensus-building, a unified danger map is produced, reflecting shared expert judgment and harmonised spatial classifications (Stucki, n.d.).

The consensus process ensures that the published forecast follows EAWS standards and communicates risk in a unified, regionally coherent way (EAWS - European Avalanche Warning Services, 2018). This step also allows experts to re-evaluate ambiguous zones and decide whether to apply modifiers or adjust boundaries based on collective confidence (Stucki, n.d.).

Bulletin Publication Once consensus is reached, the final forecast is published as the official avalanche bulletin, issued twice daily: the morning edition at 08:00 and the evening update at 17:00. Each bulletin remains valid until the next release and includes a colour-coded danger map, detailed textual descriptions of avalanche problems, and an overview of temporal evolution (see Figure 4) (Pérez-Guillén et al., 2022). For example, a bulletin might highlight new-snow and wind-slab problems above 2 000 m, particularly on north-through-south-east aspects, and depict these hazards in a danger-rose diagram of elevation bands and slope orientations. The

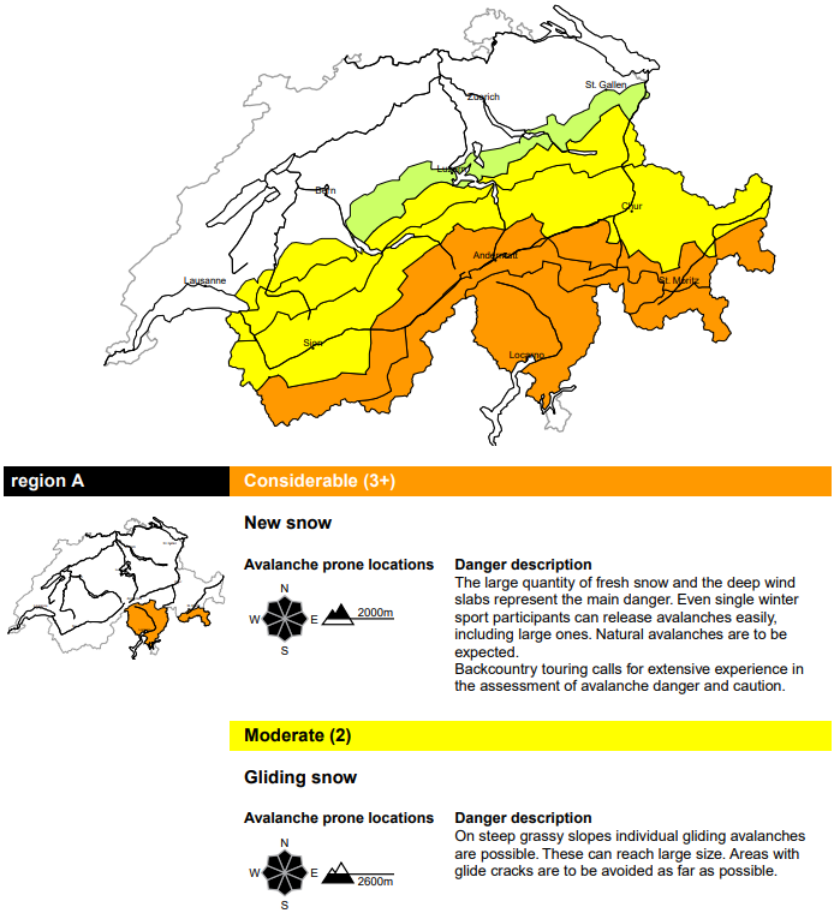


Figure 4: Partial excerpt of the SLF avalanche bulletin issued at 8:00 on 10^{th} February 2024, showing the full regional danger map (top) and detailed breakdowns for two selected regions (bottom). The map uses a standardised colour scale: green indicates low danger (Level 1), yellow moderate danger (Level 2), and orange considerable danger (Level 3). Each region-specific panel includes avalanche type, elevation range, slope aspect, and a textual hazard description.

bulletin describes full-depth release risks and specific avalanche problems (e.g. new snow or gliding snow) in those areas.

Publication occurs via the SLF website, mobile applications, and regional alert systems, ensuring that ski resorts, mountain guides, civil-protection authorities, and the public all receive the forecast. User feedback and post-event validations, such as recorded avalanches or false alarms, are then fed back into operational improvements and model refinements (WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-a). With each bulletin's release, the forecasting cycle immediately restarts, requiring experts to remain continuously engaged with new data and evolving conditions (Stucki, n.d.).

2.3.2 Model-Based Forecasting Inputs and Operational Role

In recent years, model-based forecasting has become a central component of operational avalanche hazard assessment in Switzerland. These models use a combination of physics-based simulations and ML techniques to predict key avalanche-relevant variables, providing objective, spatially distributed inputs to support expert decision-making. This section outlines the primary models used, the forecast variables they produce, and how these outputs relate to bulletin components and the research presented in this thesis.

The SNOWPACK model is the core physical snow cover model used by the Swiss avalanche warning service. Driven by high-resolution numerical weather prediction models (e.g., COSMO-1), SNOWPACK simulates snow stratigraphy, metamorphism, liquid water content, and other key snowpack characteristics at high temporal resolution. These simulations are initialised at over 140 locations using data from the IMIS and AWS networks, covering various elevations and slope aspects representative of avalanche-prone terrain (Pérez-Guillén et al., 2022; Techel et al., 2025).

From the SNOWPACK simulations, several diagnostic variables are derived. These include indicators for weak layer presence, critical layer depth, slab thickness, and indicators of potential slab release. These features form the basis for subsequent statistical and ML models. Among the most prominent models are:

- A random forest classifier trained on historical avalanche danger level ratings to predict the probability of each danger level (1-4) for each grid point (prob_level_3, etc.).
- A second random forest model that predicts the instability probability based on weak layer and slab characteristics (np_pMax, etc.).

- A logistic regression model that estimates the probability of natural avalanche activity (probability_natAval), incorporating snow accumulation and instability predictors (Pérez-Guillén et al., 2022; Techel et al., 2025).

These model outputs are computed at discrete points and subsequently interpolated onto a regular 1×1 km grid using regression kriging techniques (Fromm & Schönberger, 2022). This interpolation supports regional comparison and visualisation, and feeds into both human assessments and automated forecast tools.

Importantly, the outputs of these models are closely aligned with the content of the avalanche bulletin. For example:

- prob_level_1 to prob_level_4 and level_continuous support the assignment of the bulletin danger level.
- size_expected, P_s1 to P_s4 reflect the anticipated avalanche size class.
- Aspect and elevation filters help identify the most relevant terrain bands, consistent with the danger rose diagrams in the bulletin.

While these models offer strong performance, they are known to exhibit limitations in generalisation, especially in complex, real-world domains like snow and terrain modelling (Maxwell et al., 2018). Forecasting uncertainty is further compounded in ensemble approaches, which must account for variability across model members and spatial conditions (Leutbecher & Palmer, 2008).

In this thesis, the model outputs described above serve as the primary data set for clustering and visualisation. Their structure and operational relevance make them well-suited for exploring new methods of data simplification and expert-oriented presentation, as discussed in the following sections.

2.3.3 Structuring and Simplifying Model Data

Input data form the foundation of avalanche forecasting. They include meteorological measurements, snowpack observations, and terrain information, all of which feed into models and expert assessments. This subsection outlines the main categories of input data and their relevance for evaluating avalanche conditions.

Model-based avalanche forecasts provide detailed, high-resolution data products that describe snow stability, avalanche probabilities, and danger levels across complex alpine terrain (Horton et al., 2020; Pérez-Guillén et al., 2022). While this information is operationally valuable, it presents a challenge for both forecasters and end-users:

the data are multivariate, voluminous, and spatially heterogeneous. This complexity can increase cognitive load and reduce interpretability, especially under time pressure (Badam et al., 2022; Horton et al., 2020).

One promising approach to address this challenge is the use of clustering techniques, which aim to reduce complexity by identifying patterns or groups within the data. Clustering methods are commonly used in exploratory data analysis and have seen increasing use in environmental sciences, including meteorology, hydrology, and natural hazard risk assessment (Floyer et al., 2016; Kunz et al., 2010). Their goal is to reveal internal structure by grouping similar data points based on feature similarity, spatial proximity, or shared behaviour over time.

Unlike supervised learning models, clustering methods do not rely on labelled data. This makes them particularly useful in domains like avalanche forecasting, where large volumes of numerical data exist but clear categorical labels (e.g., "danger level") may only be available regionally or retrospectively (Jain et al., 1999; Pan et al., 2013). In this thesis, clustering is used to automatically identify and spatially group avalanche-relevant conditions based on the similarity of model outputs (Horton et al., 2020; Pérez-Guillén et al., 2022).

Several clustering methods exist, each with different strengths and weaknesses:

- K-Means and similar centroid-based methods are computationally efficient but require the number of clusters to be predefined and are sensitive to outliers (Jain et al., 1999; MacQueen, 1967).
- Hierarchical clustering (e.g., agglomerative hierarchical clustering (AGNES)) can reveal nested structures but is often computationally expensive and sensitive to linkage criteria (Hastie et al., 2009; Johnson, 1967).
- Density-based methods like DBSCAN and its variants (e.g., HDBSCAN*) are robust to outliers and do not require specifying the number of clusters. They are particularly well-suited for data with varying density and spatial noise (Stewart & Al-Khassawneh, 2022).

This work builds upon HDBSCAN*, a hierarchical, density-based clustering algorithm that can identify spatially coherent zones with similar snow and avalanche conditions. HDBSCAN* is advantageous for avalanche forecasting because it:

- does not require a pre-set number of clusters
- can handle noisy, sparse, or irregular data
- is able to produce more interpretable and spatially meaningful cluster maps

By applying clustering techniques to model output data, this work explores an alternative to expert-defined forecast regions and enables a data-driven perspective on spatial hazard structure. This supports the overarching goal of making large, multivariate data sets more interpretable and operationally useful.

2.3.4 Visualisation and Interactive Dashboards

Determining the avalanche danger level is the central task of forecasters. This step combines model results with expert judgment to classify current conditions into standardised danger levels. The following section summarises the most common methods used in practice and their respective advantages and limitations.

Visualising avalanche-relevant model data poses a unique challenge: the data is multivariate, spatially distributed, time-dependent, and often contains uncertainty. These characteristics place high cognitive demands on users, especially when rapid decision-making is required under uncertain conditions. Effective visualisations must balance detail and clarity, enabling the user to detect critical patterns without being overwhelmed.

In recent years, interactive dashboards have emerged as a practical tool for visualising spatio-temporal hazard data. Dashboards integrate multiple coordinated views, such as maps, charts, sliders, and filters, into a single interface, allowing the user to explore data from multiple perspectives while retaining spatial and contextual awareness (Badam et al., 2022; Horton et al., 2020). Within avalanche forecasting, such tools can help streamline analysis, reduce error rates, and support faster, more confident decision-making.

Several visualisation strategies have proven effective for operational avalanche settings:

- Layered map displays: Overlaid raster or point data on topographic base maps provide terrain context, allowing the user to connect forecast data to specific slopes and aspects (Kunz et al., 2010).
- Interactive filtering: Sliders, checkboxes, and aspect/elevation filters help reduce visual clutter by isolating relevant subsets of the data (Shneiderman, 1996; Young & Kitchin, 2020).
- Dynamic temporal controls: Time sliders and animations support the assessment of trends, changes in instability, or precipitation over multiple forecast intervals (Horton et al., 2020; Shneiderman, 1996).
- Colour encoding and scale design: Intuitive and standardised colour schemes (e.g., EAWS danger colours) help the user interpret hazard levels quickly and reduce misinterpretation (MacEachren et al., 2005).

- Aggregation and summarisation: Visual strategies such as cluster boundaries or worst-case value previews provide simplified overviews while retaining access to detailed information when needed (Elmqvist & Fekete, 2009; Horton et al., 2020).

While dashboards are powerful, they also pose design challenges. Poor interface layout, ambiguous colour encoding, or overuse of interactive elements can increase cognitive load instead of reducing it. Research on geovisual analytics emphasises the importance of clear task alignment, minimal visual noise, and consistent user feedback loops (Andrienko et al., 2016; Shneiderman, 1996). For avalanche data, this means using terrain-aware base maps, grouping related parameters (e.g., instability and snowfall), and minimising clutter in high-density data zones.

This thesis adopts a user-centred design approach, combining domain-specific insights with general visualisation guidelines. The dashboard developed here integrates cluster-based summaries, terrain filters, and interlinked spatial and temporal views. It aims to reduce expert workload, support situational awareness, and enable interpretable exploration of complex model outputs.

While a range of explainability and interaction design techniques have been proposed in the literature, such as SHapley Additive exPlanations (SHAP) values for local model attribution (Lundberg & Lee, 2017), individual conditional expectation (ICE) plots (Goldstein et al., 2015), or coordinated multiple views (Baldonado et al., 2000; Yi et al., 2007), not all of these were directly integrated into the implemented dashboard. Instead, this thesis focused on visual strategies aligned with operational needs, such as filter-based simplification, dynamic tooltips, and danger-level-based colour encoding (Brewer, 2005). The selection of visualisation elements was guided by usability and clarity, prioritising expert interpretability over full model introspection.

Together, these processes form the foundation of the operational forecasting workflow, which connects directly to the methodological approach of this thesis.

Together, these processes form the foundation of the operational avalanche forecasting workflow, linking the physical drivers of avalanche formation with the information products used in practice. Combined with the geovisualisation approaches discussed earlier, they illustrate both the opportunities and limitations of current forecasting practice. This state of the art establishes the conceptual and practical background on which the methodological approach of this thesis is built.

3 Data

This chapter introduces the data sets used for analysis and dashboard implementation. It describes the IMIS, derived point-based products, interpolated grid fields, and the official avalanche bulletins provided by the SLF. Finally, it presents a correlation analysis to illustrate relationships between attributes. Understanding these inputs is crucial for interpreting later methodological and visualisation choices.

The data sets analysed in this thesis were all provided by the SLF. I obtained two distinct data sets, point measurements and grid data, each offering complementary perspectives on snow and avalanche conditions. All observations were done in the period from February to April 2024.

Throughout this thesis, I have not performed any further calculations or preprocessing on the SLF-provided data. My sole objective remains to present and visualise the data set as delivered, ensuring that the original measurements and derived probabilities are faithfully represented.

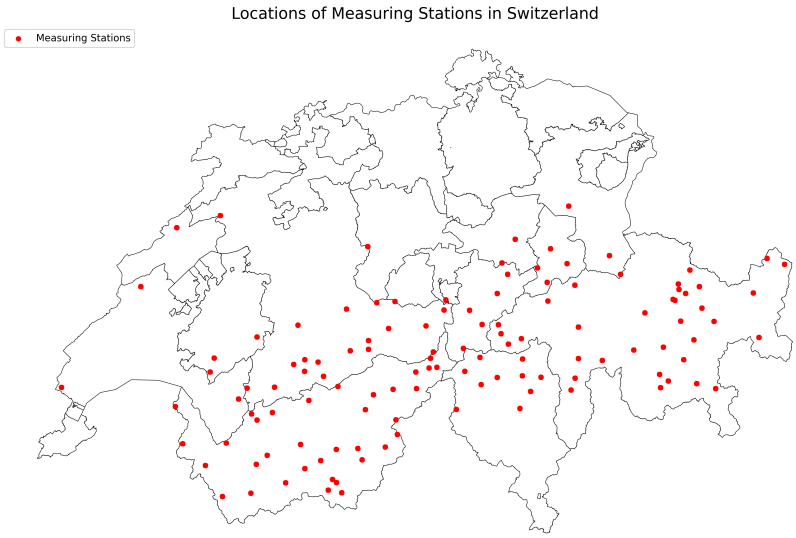


Figure 5: Distribution of the 189 Intercantonal Measurement and Information System (IMIS) stations across Switzerland, which provide automated snowpack and weather measurements used in avalanche forecasting.

3.1 Intercantonal Measurement and Information System

A large part of the data in this thesis was, at the first instance, collected by the automated measuring stations. The IMIS network has stations, which autonomously collect data (Lehning et al., 1999; Pérez-Guillén et al., 2022). There are 189 stations, situated mainly across the alpine region, with some stations in the pre-alpine region and four stations in the Jura region of Switzerland (see Figure 5). According to Pérez-Guillén et al. (2022, p. 3) "15% of the stations are situated at elevations between 1500 and 2000m a.s.l., 61% between 2000 and 2500m a.s.l., and 24% between 2500 and 3000m a.s.l.". This shows that the range of elevations is broad, which is necessary to get complete coverage of measurements.

Figure 6: "IMIS snow station Belalp (canton Valais) at 2556 m: The four poles visible in front of the station carry a fence in the summer to protect the temperature sensors from wild animals and livestock." (WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-b)

The stations are built in flat terrain, which is protected from the wind. This is crucial to get accurate measurements (WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-b). Following are the most central measured variables:

- snow depth
- air- and surface temperature
- wind speed and wind direction
- snow temperature at 25 cm, 50 cm and 100 cm above ground

(For the complete list of measured variables at the IMIS stations, please see (WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-b).)

Every 30 minutes, the stations collect the data on the aforementioned variables and then send them to the SLF and other partners like MeteoSwiss for further usage (WSL – Institut für Schnee- und Lawinenforschung SLF, n.d.-b).

3.2 Point Data

The first data type used in this thesis is point data, which primarily comprises probability values derived from various calculations on the fundamental measurements collected by the IMIS stations (see Table 2 for a complete list of point data attributes). These derived values were generated using random forest models (Mayer et al., 2022; Pérez-Guillén et al., 2022) and a sigmoid shifter (Mayer et al., 2023). The only directly measured values employed in this thesis are the snow depth readings (hn24, hn72, and hs).

Although these probability values are not raw measurements from the IMIS stations, they are still geographically tagged to the exact locations where the basic data were recorded (Mayer et al., 2023; Pérez-Guillén et al., 2022). Figure 7, for instance, illustrates how the point data information is displayed in my dashboard, providing an intuitive spatial context.

The full data set is supplied as a CSV file, with all attributes recorded once per day for each station over the entire three-month period. In contrast to the IMIS stations' half-hourly logging, my data set offers a single daily snapshot per station, uniformly timestamped at 12:00 UTC. This daily resolution simplifies the temporal granularity while still capturing the essential trends. In addition to the attribute values, the "X" and "Y" coordinates in the Swiss LV03 coordinate system, the elevation (in metres above sea level) of the measuring station , as well as the slope aspect are recorded.

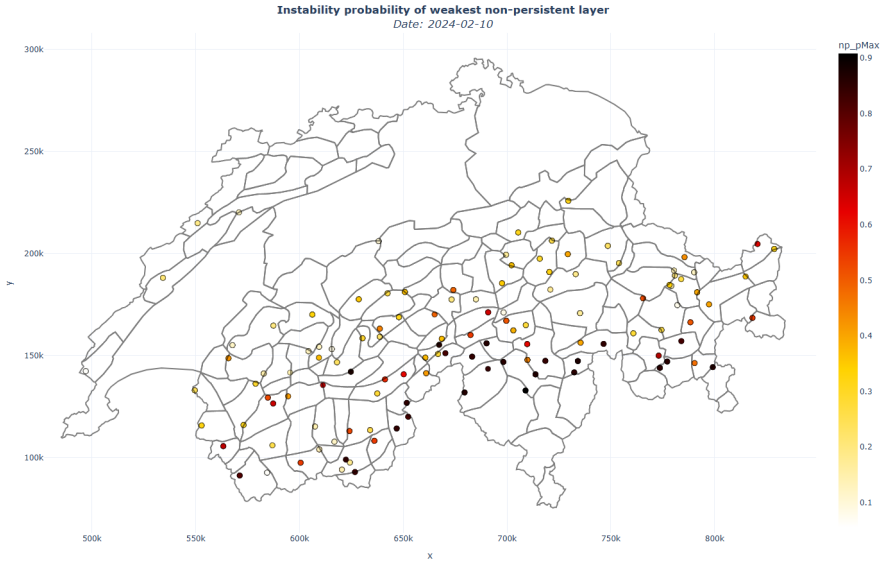


Figure 7: Point-based visualisation of the SLF-provided data set showing the instability probability of the weakest non-persistent layer [np_pMax] across Switzerland on 10^{th} of February 2024.

3.3 Grid Data

The second data set in this thesis is the grid data. Like the probability values in the point data file, these gridded values are derived from the measurements collected at the IMIS stations. However, instead of representing individual station locations, the interpolation produces a continuous surface covering nearly all of Switzerland. Only a few areas on the lower Swiss Plateau are left without values, but since avalanche risk there is negligible, this omission does not affect the overall analysis. Figure 8 provides an example of how these gridded fields appear in the dashboard, and Table 2 lists all available grid data attributes.

Each grid snapshot is generated once per day, with a uniform timestamp of 12:00 UTC. For every grid cell, the data set includes the interpolated probability or model value, the elevation (in metres above sea level), the slope aspect at that location, and the Swiss LV03 X- and Y-coordinates.

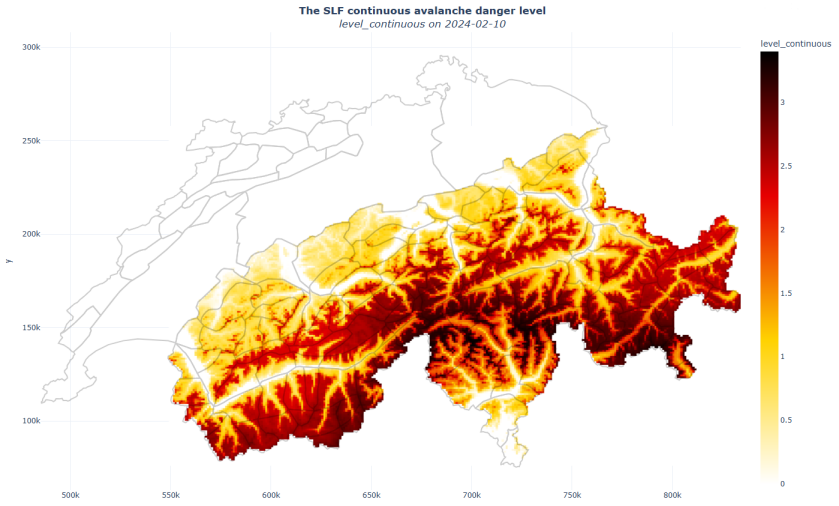


Figure 8: Interpolated grid-based visualisation of the SLF-provided data set of the continuous interpolated danger level [level_continuous] across Switzerland on the 10^{th} of February 2024.

3.4 Avalanche bulletins

The third type of information I received for my thesis comprises the SLF avalanche bulletins for 2023 and 2024, provided in a single CSV file. Each record in this file covers a specific validity period and one or more forecasting regions, or “sectors”, that the SLF uses to enhance the spatial resolution of its danger forecasts. For each bulletin entry, the “drySnow” attribute indicates whether the predicted avalanche would involve dry snow or wet snow. The slope aspect (exposure) is also recorded

to show which mountain faces are at risk. Finally, two key fields convey the danger rating: "level_numeric," which gives the overall danger level on a standardised scale, and "level_detail," which adds sublevel information to refine the forecast (Techel et al., 2022).

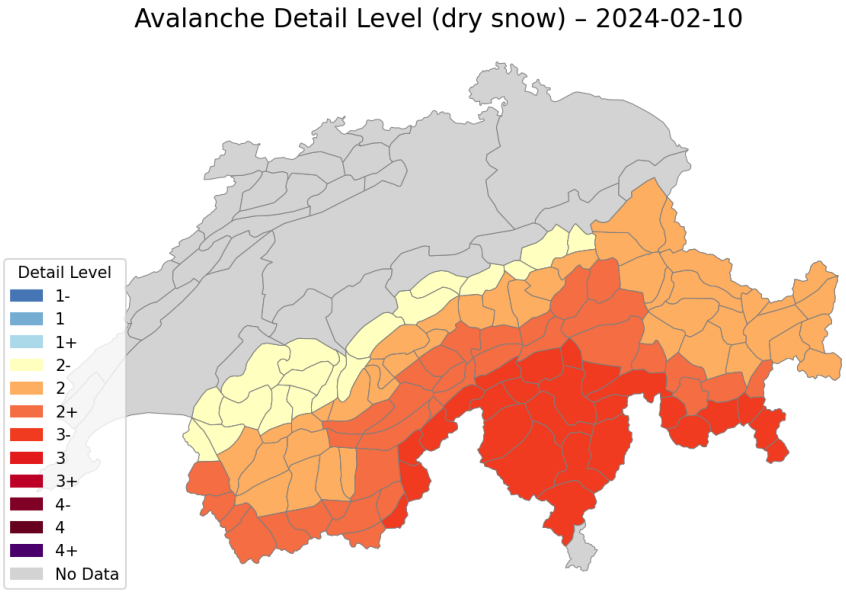


Figure 9: Regional avalanche danger levels for dry snow conditions on 10^{th} February 2024, extracted from the official SLF avalanche bulletin. Colours correspond to the European Avalanche Danger Scale (Level 1 = low to Level 5 = very high)

3.5 Correlation of Data

As a first step , after receiving the data, I decided to analyse the data on correlation. The correlation matrix can be seen in Figure 10.

It is noticeable that the variables P_s2 to P_s5 and size_expected are heavily correlated to each other. This is expected, as the variables P_s2 to P_s5 are all generated from a single determining parameter family (Maissen et al., 2024; Schweizer et al., 2008). The probabilities are then calculated using a sigmoid shifter from left to right for larger or smaller avalanche sizes. Therefore, all of these variables are perfectly correlated with each other (Mayer et al., 2023). Since the size_expected variable contains the same information as the aforementioned variables, but in an ordinal scale, it makes sense that it correlates as well.

Another strongly correlated cluster contains probability_level_3, probability_level_4, and level_expected. The two probability level attributes are derived from the same random-forest "danger score": each tree votes on a discrete avalanche level (1-4), and the overall probability for "level $\geq$ k" is just the fraction of trees

Table 2: SLF-provided attributes used in this thesis, categorised as interpolated grid data, IMIS point data, SNOWPACK model output, or direct measurements, with definitions and descriptions of each variable relevant to avalanche forecasting.

Attribute Name	**Data Type**	**Description**
aspect	Point data (IMIS) / Interpolated grid data	Orientation of slope (°).
elevation	Point data (IMIS) / Interpolated grid data	Elevation (metres above sea-level) of station
date	Point data (IMIS) / Interpolated grid data	Forecast date (YYYY-MM-DD) at 12:00 UTC.
X	Point data (IMIS) / Interpolated grid data	Swiss LV03 X coordinate.
Y	Point data (IMIS) / Interpolated grid data	Swiss LV03 Y coordinate.
dlModelFx_Prob3	Interpolated grid data	Probability that danger level ≥ 3.
instabModelFx	Interpolated grid data	Probability snow cover is unstable.
spontLawModelFx	Interpolated grid data	Probability of spontaneous dry-snow avalanches.
level_continuous	Interpolated grid data	Continuous value of sublevels
sublevel	Interpolated grid data	Ordinal values of sublevels
slfCode	Point data (IMIS)	Unique station identifier.
sector_id	Point data (IMIS)	Warning region ID.
probability_level_1	Point data (IMIS)	Probability of danger level 1 (low).
probability_level_2	Point data (IMIS)	Probability of danger level 2 (moderate).
probability_level_3	Point data (IMIS)	Probability of danger level 3 (large).
probability_level_4	Point data (IMIS)	Probability of danger level 4 (high).
level_expected	Point data (IMIS)	Expected danger level (weighted mean).
P_s2	Point data (IMIS)	Probability of avalanches size ≥ 2.
P_s3	Point data (IMIS)	Probability of avalanches size ≥ 3.
P_s4	Point data (IMIS)	Probability of avalanches size ≥ 4.
P_s5	Point data (IMIS)	Probability of avalanches size ≥ 5.
size_expected	Point data (IMIS)	Expected avalanche size category.
pMax_decisive	Point data (IMIS)	Probability that the weakest layer potentially unstable.
depth_decisive	Point data (IMIS)	Depth (cm) of the weakest layer.
np_pMax	Point data (IMIS)	Probability that the weakest non-persistent layer (np) is potentially unstable.
np_depth	Point data (IMIS)	Depth (cm) of the weakest non-persistent layer (np).
pwl_pMax	Point data (IMIS)	Probability that the weakest persistent layer (pwl) is potentially unstable.
pwl_depth	Point data (IMIS)	Depth (cm) of the weakest persistent layer (pwl).
probability_natAval	Point data (IMIS)	Probability of natural dry-snow avalanche occurence in the vicinity of the station.
hn24	SNOWPACK	Snow height (cm) in last 24 h.
hn72	SNOWPACK	Snow height (cm) in last 72 h (sum of previous three days).
hs	Measured values	Snow depth at the observation site (cm).

voting at or above threshold k. Because levels 3 and 4 correspond to adjacent thresholds on the same score, their probabilities rise and fall almost in lock-step, and the majority-vote "level_expected" sits between them, resulting in near-perfect correlation (Maissen et al., 2024; Pérez-Guillén et al., 2024). By contrast, the cut-off score that defines danger level 1 sits very close to the bottom of the model's risk scale (around the 10th percentile of scores), whereas the cut-off for danger level 2 lies much higher (near the 50th percentile). In other words, probability_level_1 measures the chance of scoring in roughly the lowest 10 % of the risk distribution, while probability_level_2 measures the middle 40 %. Because these two bands do not overlap and span quite different regions of the score range, their values vary independently and consequently show only a weak statistical correlation (Pérez-Guillén et al., 2022). The correlation can be explained that the level_expected score is just the mean of the same probability distribution $\{p_1, p_2, p_3, p_4\}$ that produces probability_level_3 and probability_level_4. Since in the data set the model often assigns most of its mass to levels 3 and 4, those two probabilities dominate the weighted sum. As a result, any shift in the underlying "risk" score that boosts p_3 or p_4 will lift the expected value in almost exactly the same way, hence the near-perfect correlation (Maissen et al., 2024; Pérez-Guillén et al., 2024; Techel et al., 2022, 2024).

Moreover, the four interpolated model outputs (dlModelFx_Prob3, instabModelFx, spontLawModelFx, and level_continuous/sublevel) together with the probability_natAval attribute all follow very similar patterns. This reflects two key points: first, they are derived from the same core SNOWPACK-COSMO-1 framework; and second, and more crucially, each of these metrics encodes the severity of avalanche danger (i.e. higher values mean more severe conditions). Consequently, variables such as the expected danger level (level_expected) and the probability of natural dry-snow avalanches (probability_natAval) tend to rise and fall synchronously, having a strong positive correlation. Such correlations are hardly surprising: all interpolated values stem from the same IMIS station measurements, even though each data set applies a different interpolation method and extracts distinct variables. Because they trace back to identical base data, these attributes naturally move in lock-step (Maissen et al., 2024; Pérez-Guillén et al., 2024). The probability_natAval follows the same principle; it's calculated from the IMIS-derived danger score and then passed through a sigmoid transformation to convert that score into a probability (Maissen et al., 2024; Pérez-Guillén et al., 2024). Understanding these interdependences is crucial, as it highlights how variations in the underlying measurements propagate across multiple derived fields, reinforcing the need for careful interpretation when comparing them (Mayer et al., 2023).

In general, this analysis has shown that the data is strongly correlated. While I keep

the full set of variables for all visualisation tasks, dimension-reduction techniques (as detailed in Section 4) are applied to simplify the feature space for clustering, allowing the analysis to focus on the most informative components. I maintain full visibility of all relevant variables to provide forecasters with a comprehensive picture, while simultaneously applying targeted simplification, such as variable-specific clustering, to make the data more accessible and actionable.

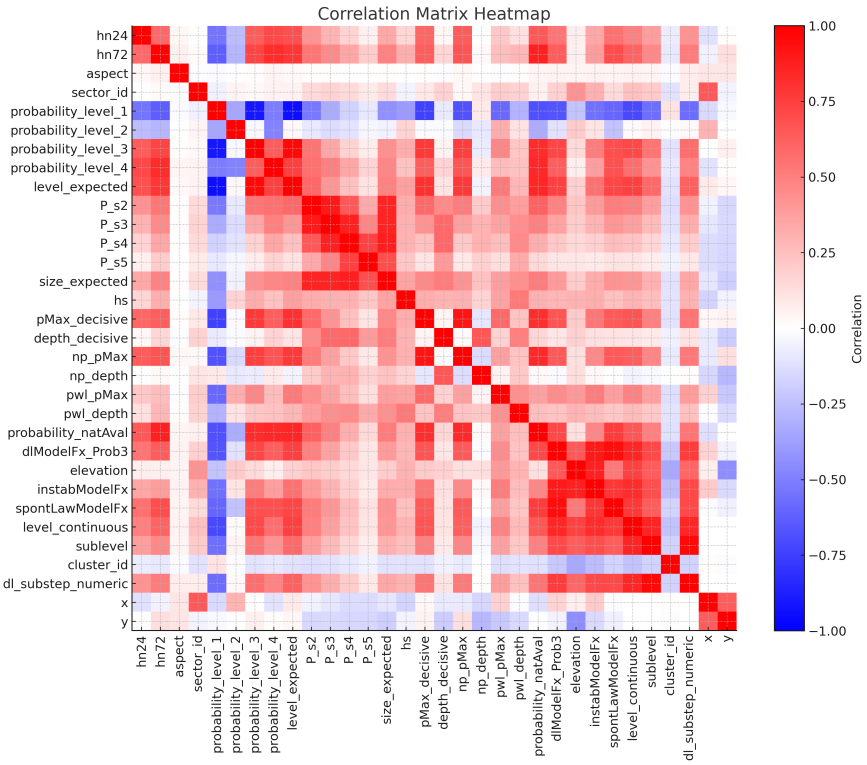


Figure 10: Correlation heat map of avalanche forecasting attributes, showing pairwise relationships between all point- and grid-based variables provided by the SLF. Positive correlations are shown in red, negative in blue.

4 Methods

This chapter presents the methodological framework of the thesis. It describes the data preparation steps and the clustering approach, including the functionality and optimisation of HDBSCAN*. The development of the dashboard is then outlined, with emphasis on layout, design choices, and validation strategies. By combining algorithmic evaluation with user-centred design principles, the chapter demonstrates how the technical and visual components of the system were developed and tested.

4.1 Data Preparation

Before applying clustering or visualisation methods, the raw data sets were pre-processed to ensure consistency and comparability. This step included filtering, transformation, and feature selection to reduce redundancy and highlight relevant attributes. Proper data preparation is essential for both reliable clustering outcomes and meaningful dashboard representations.

The first step to prepare the data for my adapted HDBSCAN* clustering algorithm, I merged and transformed the raw files into a uniform format. Originally, point measurements and gridded values arrived in separate files and featured very different spatial densities, making direct ingestion into the clustering pipeline impractical. To simplify this, I merged all files into a single data set.

Following Archibald et al. (2024), I overlaid Switzerland with a regular 5 km $\times$ 5 km LV03 grid, a compromise between spatial detail and real-time performance. I also tested 10 km $\times$ 10 km and 2.5 km $\times$ 2.5 km alternatives: the coarser 10 km grid excessively smoothed out localised instabilities, while the finer 2.5 km grid increased clustering time by roughly 4$\times$ (to >10 seconds per day) without appreciably improving expert interpretability.

Within each 5 km cell, I computed the mean of every attribute, yielding a uniform "pixel" matrix. Post-aggregation, about one-third of the cells contained no data, roughly half held only interpolated grid values, and the remaining quarter combined both station and grid measurements.

As an alternative, each cell could be assigned the maximum (i.e., worst-case) value per attribute, particularly for strongly elevation-dependent metrics, rather than the mean. This "max-per-cell" approach may better capture critical hotspots, as is often done in operational practice. As shown in Table 3, the overall ARI (across the entire data set) is higher when using the mean value per pixel, whereas in day-by-day evaluations the max-value variant outperforms the mean-value variant.

Table 3: Clustering performance using mean versus maximum aggregation per grid cell. Performance is evaluated with the Adjusted Rand Index (ARI), reported both as overall agreement across the full data set and as the mean ARI across daily clustering runs.

Aggregation Method	**Overall ARI**	**Mean per-day ARI**
Mean value	0.02234	0.15499
Max value	0.01159	0.19411

Finally, I repeated this aggregation for every date in the study period, ensuring that each daily snapshot fed into the clustering algorithm shared the same structured format.

These preparation steps yield a consistent dataset suitable for clustering and dashboard implementation.

4.2 HDBSCAN*

This subsection presents the clustering strategy applied to avalanche data. It first explains the rationale for using clustering in this context and discusses why common algorithms such as K-Means are not suitable. It then introduces HDBSCAN*, the algorithm ultimately adopted, and describes its core functionality, including parametrisation, hierarchy construction, and outlier detection. Finally, the optimisation of clustering parameters and modifications to the ensemble approach are outlined.

Clustering distills high-dimensional snowpack and meteorological data into a few coherent groups, revealing shared terrain features, stability profiles, and risk scores, directly within the dashboard. The goal of this algorithm is not to replace the experts' analysis but to give forecasters a clear visual of how data values cluster spatially, helping them grasp distribution patterns and focus on representative scenarios.

Finding a suitable algorithm was challenging. The most common clustering algorithms like K-Means have a lot of advantages with their easy set-up, easy-to-interpret results, reliability, and most importantly, already widespread usage (Jain, 2010; Tchagna, 2025). But the drawback of K-Means is that the number of clusters the algorithm will calculate has to be known beforehand (Stewart & Al-Khassawneh, 2022; Tchagna, 2025). This limitation alone lets me not use algorithms that function like K-Means. The data, of course, changes per day, which makes it impossible to predict or decide in advance how many clusters there will or should be on any given day.

Ultimately, I decided to use the HDBSCAN* algorithm. The HDBSCAN* is a variable-density-based clustering algorithm. That means that it does not matter to the algorithm how much the data changes between different days; the algorithm can cluster the data. It also has a built-in outlier detection, which increases the reliability of HDBSCAN* (Fromm & Schönberger, 2022; Lewis et al., 2017; Wang et al., 2025).

HDBSCAN* "stands for Hierarchical DBSCAN*. The asterisk suffix indicates an improvement [...] Campello et al. (2013) make to DBSCAN. These authors extend their work in Campello et al. (2015) by introducing a complete framework for cluster analysis and outlier detection which includes an enhanced explanation of the HDBSCAN* algorithm" (Stewart & Al-Khassawneh, 2022, p. 2).

4.2.1 Functionality of HDBSCAN*

This subsection outlines the main steps of the HDBSCAN* algorithm, including its parametrisation, hierarchy construction, and built-in outlier detection.

The input for the algorithm is a set of points in a feature space; these points can, but do not have to, be multi-dimensional. Then a set of parameters has to be set. In the base case, the necessary parameters are min_samples k and min_cluster_size s. These two parameters respectively define the number of neighbours to be estimated in local density and the minimum size for any cluster to be considered as a real cluster (Stewart & Al-Khassawneh, 2022).

Next, the core distance and the mutual reachability are defined. The core distance is the distance for each point x to its k^{th} neighbour (including x itself). By defining the core distance, the algorithm knows the approximate local density. The tighter the neighbourhood, the smaller the core distance. The mutual reachability distance $mreach_k(x,y)$ helps to smooth out the pairwise distance. The mutual reachability distance is defined in a way that it cannot be lower than the core distance of either of the two points (Stewart & Al-Khassawneh, 2022).

After that, the extended minimum spanning tree (MST) is built, which is used as a substrate for the hierarchy construction in the algorithm. First, with the core distance and the mutual reachability distance, the MST, a complete weighted graph, can be calculated. Out of this graph, the extended MST can be extracted by selecting the subset of edges connecting all points with a minimum total weight. After that step, a parameter called self-edge is added to each vertex, which weighs the vertex by its core distance (Stewart & Al-Khassawneh, 2022).

The next step is to construct the hierarchy. This is achieved foremost by edge removal. After initialising the algorithm, a cluster, called root cluster, with all possible points is created. Parallel, a list is maintained with all the active clusters. Then, the weight

of all the edges of the extended MST is sorted in ascending order. The weight of the edges is being interpreted by the algorithm as the density level. Then the first filtering pass is done: the algorithm removes the weight of the smallest edge. If there is more than one edge with the same weight, all of them are filtered out at the same time (Stewart & Al-Khassawneh, 2022).

Whenever an edge is removed from the root cluster the algorithm checks if the number of edges removed is larger or equal to the s (min_cluster_size). If it is, a new cluster is created; if the number is lower than s, the points are labelled as noise. This process is repeated until all points are extracted from the root cluster and the density is at its maximum. At this point, all cluster points are either labelled as noise or are in their own trivial cluster (Stewart & Al-Khassawneh, 2022).

For each one of the newly assigned clusters C, λ_{birth} (density level at which C first appears, when its parents split) and λ_{death} (density level at which it itself splits into children) and λ_{p} (the level at which a point is extracted from the root cluster) are noted. These values serve as the main parameters for the cluster stability calculation (see Formula 1).

$$\text{stability}(C) = \sum_{p \in C} (\lambda_p - \lambda_{\text{birth}}) \tag{1}$$

The higher the persistence of members of a cluster across a wide range of density thresholds, the higher the stability of the cluster (Stewart & Al-Khassawneh, 2022).

In the next step, leaf clusters are identified. Leaf clusters are clusters that never split up into children and therefore, are at the bottom of the hierarchy. Starting at the leaf clusters, a pass to the root of the tree is performed where all the clusters are assigned their own stability (C) and the propagated stability of all the children.

After that, there are two possible cases:

1. the stability (C) $\geq$ the propagated stability of its children
2. or the stability (C) $<$ the propagated stability of its children

If the stability C is higher or equal to the sum of the propagated stabilities, which means that the cluster is more stable than all its subclusters combined, only C is propagated further and C itself is marked as "prominent descendant" for its parents. If the combined stability is higher, the combined stability is propagated and the parents treat all those descendant clusters as their propagated descendants (Stewart & Al-Khassawneh, 2022).

To address the problem of outliers, the HDBSCAN* calculates the Global-Local Outlier Score (GLOSH). Whenever a point x is marked as noise, the edge-weight ε(x) (density-level) and the lowest child-death level $\varepsilon_{\max}$(max) (lowest child-death level propagated to x's last cluster) are noted. With these parameters, the GLOSH can be calculated (see Formula 2) (Stewart & Al-Khassawneh, 2022).

$$\text{GLOSH}(x) = 1 - \frac{\varepsilon_{\max}(x)}{\varepsilon(x)} \tag{2}$$

With the information noted at the point of extraction from the root cluster, the outlier calculation can be done. Values near 1 indicate strong outliers, while values near 0 behave like typical cluster members. The GLOSH score is used to rank and filter the anomaly of the points. Typically, points with a GLOSH score larger than 0.8 are flagged as strong outliers and are removed from the cluster. Once a point is removed, be it because of the GLOSH score or because of other reasons, it will stay noise. In HDBSCAN* there is no noise-reassigning process (Stewart & Al-Khassawneh, 2022).

4.2.2 Optimising Clustering Parameters

To improve clustering quality, I systematically optimised the algorithm's key parameters and compared alternative approaches with labels defined by experts. To find the best combination of parameters, I implemented an automated clustering parameter optimisation script derived from scikit-learn examples. This procedure is described in the following section.

The SLF bulletins serve as the reference labels by providing, for each warning region and date, a numeric danger level ("level_numeric") valid from the bulletin's start to end timestamps. In preprocessing, I normalise each bulletin's valid_from/valid_to interval to a daily "date" field, then merge these labels with the grid-cell data set on both sector_id and date. This assigns every grid snapshot its corresponding expert-defined danger level, which I then treat as ground truth when optimising and evaluating the clustering, quantifying agreement via the Adjusted Rand Index (ARI).

The initial clustering results for the standard HDBSCAN* algorithm had very low ARI scores and no sensible solutions were generated. I therefore decided to try different approaches, to maybe find a better solution than just the plain HDBSCAN* algorithm. I ended up testing four different approaches:

- a normal HDBSCAN* approach
- a baseline HDBSCAN* approach

- an ensemble HDBSCAN* approach
- a seeded K-Means approach

The "normal" HDBSCAN* approach used the regular HDBSCAN* clustering algorithm and checked all combinations of clustering parameters and printed the combination with the highest ARI.

In the baseline approach, I used a fixed set of parameters. In this case, the predefined parameter combination was: min_cluster_size = 10, min_samples = 5 and ε= 0.5. With this first run, the algorithm identifies dense regions and labels all regions not in these clusters as noise. After this initial run, it re-assigns all the noise data to the nearest non-noise cluster and prunes any clusters smaller than the minimum size (Berba, 2020).

The ensemble approach repeats the same HDBSCAN* cluster run five times. Each run, a random seed is used to force the algorithm's stochastic elements to produce slightly different clustering results. From these five labellings of the different runs, an n x n co-association matrix is built. This matrix counts how many times two points ended up in the same cluster across the different runs. This matrix is then converted into a distance matrix, which is fed into average-linkage agglomerative clustering to produce a consensus partition. Because it aggregates information over multiple HDBSCAN* runs, the ensemble consensus is typically more robust and stable than any single run (Pedro, 2022).

The next approach is something completely different and does not even use the same clustering algorithm. I just wanted to test a K-Means based approach and decided on using a seeded K-Means clustering. By averaging all points in each ground-truth class, the true-label centroids are computed. Then the K-Means is initialised on those centroids and runs Lloyd's algorithm to convergence. Because the start points are the ideal "true" centroids, this effectively measures the best-case performance of a centroid-based clustering on the data. There is no randomness in the initialisation, and it shows the upper bound on how well K-Means could match the labels if it started from perfect seeds (Low et al., 2019).

The comparison of all four methods showed that the ensemble HDBSCAN* delivered the strongest performance (see Table 4). Consequently, I adopted an ensemble HDBSCAN* strategy and refined the workflow to further boost results. The details of this enhanced process are laid out in the next subsection.

Table 4: Comparison of four clustering approaches (normal HDBSCAN, baseline HDBSCAN*, ensemble HDBSCAN*, and seeded K-Means), showing Adjusted Rand Index (ARI) values for each method.

Approach	**ARI**
HDBSCAN* (normal)	0.012
HDBSCAN* (baseline)	0.198
HDBSCAN* (ensemble)	0.199
K-Means	0.152

4.2.3 Modified Ensemble HDBSCAN* Approach

This subsection describes the custom enhanced ensemble variant of HDBSCAN*, which combines multiple runs and additional preprocessing steps to improve stability and agreement with expert-defined danger levels.

The first step was to load the bulletin data as ground-truth and the grid data into a common workspace. Following the tidy-data principles of Wickham (2014), the bulletin file was "exploded" so that each date-sector pair carried exactly one danger label. These were then merged into a single data frame. As shown in Table 5, the overall ARI across the full data set is higher when using only the gridded data, whereas including point data only improves the mean per-day ARI, indicating better agreement on individual days.

Table 5: Clustering agreement (ARI) for grid-only features compared to grid and point features. The table reports both the overall ARI across the full data set and the mean per-day ARI when clustering is performed separately for each day.

Feature Set	**Overall ARI**	**Mean per-day ARI**
Grid	0.02251	0.07773
Grid and Point	0.01139	0.15716

To smooth out short-term fluctuations and reduce noise in high-frequency measurements, I computed three-day rolling means and variances for every variable. I also stored each attribute's value from the previous three days, along with day-to-day changes (Δ_1 between today and yesterday, and Δ_2 between yesterday and the day before) (StackExchange, 2024). I tested alternative windows (one-day and five-day) but found the three-day span to have the best balance: a shorter (one-day) window retained too much volatility, obscuring broader trends, while a longer (five-day) window excessively smoothed over sudden large-snow events that can trigger avalanches. Although SNOWPACK already incorporates the full snowpack history, this sum-

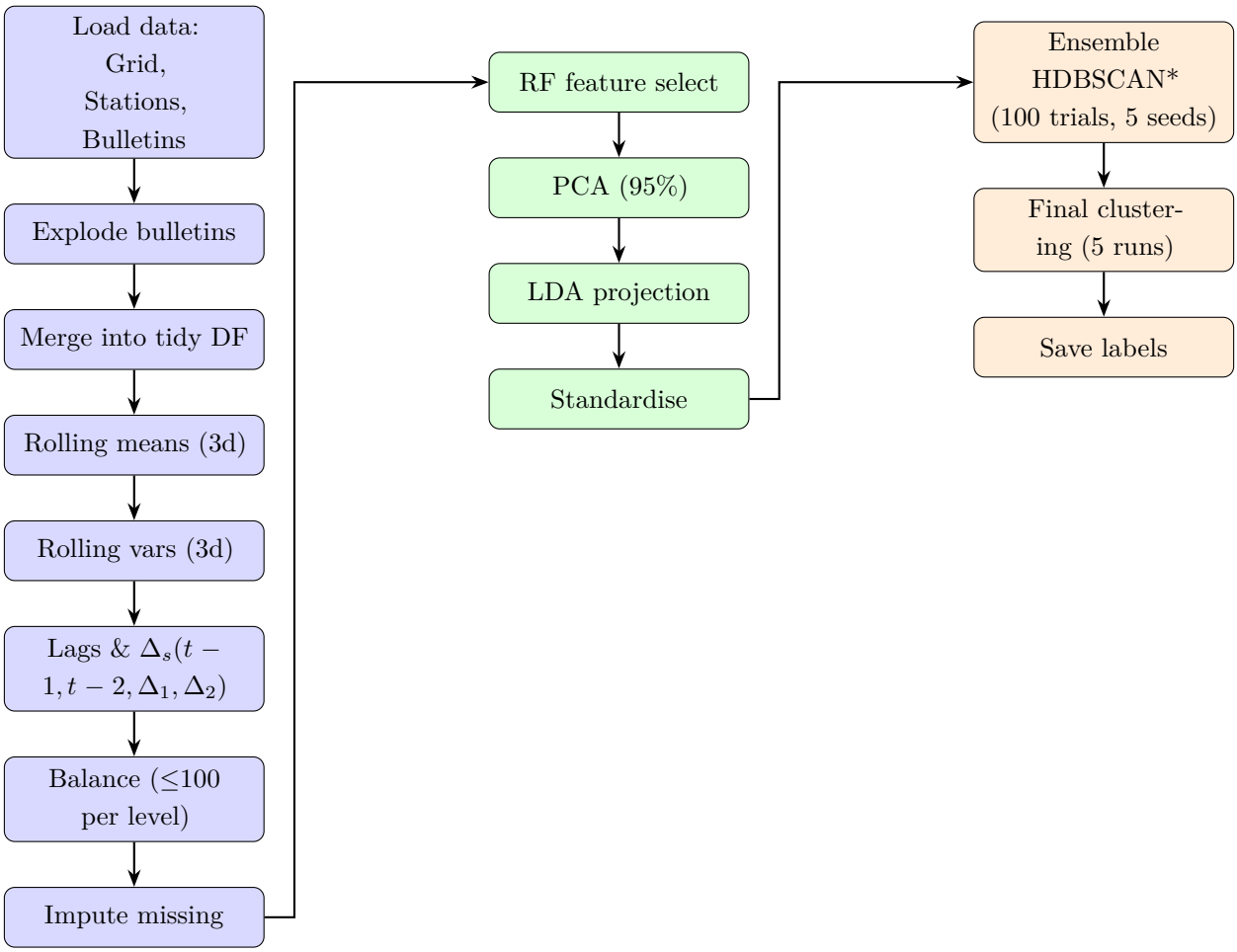


Figure 11: Three-stage processing pipeline for the modified ensemble HDBSCAN* approach. The workflow includes data preparation (blue), dimensionality reduction and feature transformation (green), and final clustering with ensemble HDBSCAN* and label generation (orange)

mary step helps stabilise the clustering input without erasing critical multi-day accumulations (Lehning et al., 1999; Pérez-Guillén et al., 2022).

In the grid aggregation step, I used the mean value of each variable across all elevation bands within a given 5×5 km cell. This choice simplifies the snowpack with multiple layers into a single representative figure per grid cell, making downstream clustering and visualisation more tractable. However, in operational forecasting, it is common to adopt a worst-case approach, for example, the maximum instability probability or the deepest weak layer at each aspect and elevation band, on the premise that a single dangerously unstable zone can trigger an avalanche for the entire cell. However, as I have shown in Table 3, using the mean value per pixel yields a higher ARI across the entire data set and therefore consolidates my choice.

For the clustering parameter search, I balanced the data set following the method proposed by He and Garcia (2009) and drew up to 100 samples per danger level per

day. My original intent was twofold:

1. Preventing dominant classes: Without balancing, the most frequent danger levels (especially Level 3) would vastly outnumber rarer levels, biasing HDBSCAN* toward those dense regions and making it almost impossible to detect meaningful clusters corresponding to low-frequency classes (Campello et al., 2013; Ester et al., 1996; He & Garcia, 2009).
2. Improving runtime: Reducing each class to a fixed maximum of 100 points kept the optimisation runs practical (full-data set trials could take days per parameter set) (Sculley, 2010; Zhang et al., 1996).

However, this balancing step itself alters the true spatial structure of the data: by under-sampling common levels, large areas of moderate risk are effectively "thinned out," being replaced with sparser, artificially uniform samples. This can flatten genuine density gradients, potentially degrading cluster quality (He & Garcia, 2009).

Next, I tackled dimensionality reduction in three stages. First, I filled missing values with the column mean and used a random-forest classifier to identify the most important features (those above the median importance) (Breiman, 2001; Little & Rubin, 2019). I filled all missing values using a single "global" mean for each attribute rather than computing separate daily means. This choice simplifies the workflow and enhances numerical stability, though it may introduce some bias by smoothing over genuine day-to-day fluctuations. Then I ran a principal component analysis (PCA) to retain only the components explaining 95 percent of the variance (Jolliffe, 2011). Finally, I applied linear discriminant analysis (LDA) to maximise separation between the SLF's defined danger levels (Fisher, 1936).

The effects of these transformations on the feature space and clustering performance are summarised in Figure 12. The top-left panel shows the rapid expansion of features after polynomial generation, followed by reductions through feature selection, PCA, and LDA. The top-right panel confirms that 498 PCA components capture 95% of the variance. The bottom-left panel presents the Adjusted Rand Index (ARI) for clustering at each stage with fixed HDBSCAN* parameters, showing a marked increase after LDA. The bottom-right panel compares the number of HDBSCAN* clusters found per day to the number of distinct danger levels in the avalanche bulletin, highlighting both matching and diverging cases.

After feature selection, I standardised all variables by subtracting the mean and dividing by the standard deviation. This keeps each feature on a similar scale, important for methods like PCA, LDA, polynomial feature generation, and random-forest importance, without compressing them into a 0-1 range where outliers get squashed (Scikit-learn, n.d.).

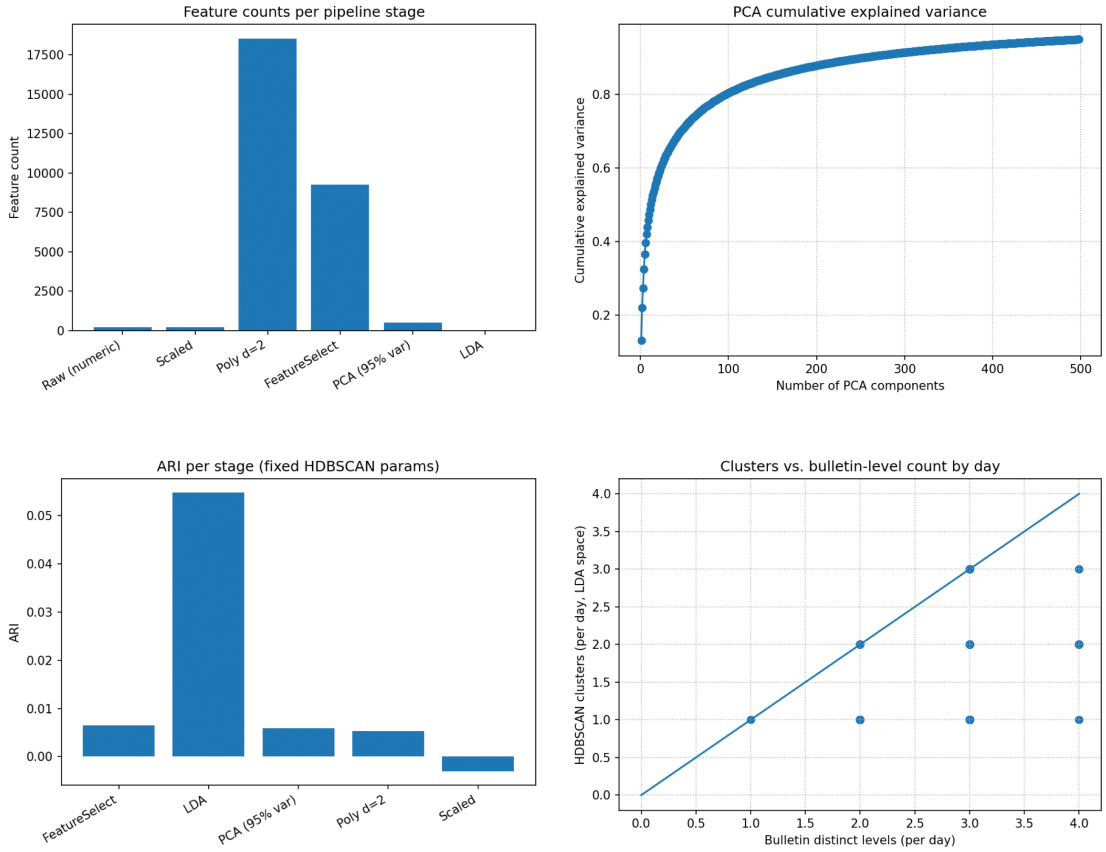


Figure 12: Summary of dimensionality reduction and clustering results for the modified ensemble HDBSCAN* run. Top left: number of features at each pipeline stage. Top right: cumulative explained variance from PCA, showing that 498 components capture 95% of variance. Bottom left: Adjusted Rand Index (ARI) for clustering at each stage with fixed HDBSCAN* parameters, highlighting a notable improvement after LDA. Bottom right: relationship between the number of HDBSCAN clusters found per day and the number of distinct danger levels in the avalanche bulletin.

To reduce spatial noise, I relabelled every data point flagged as “-1” (noise) to match its nearest non-noise neighbour. Any remaining clusters smaller than the minimum cluster size were again marked as noise and reassigned in the same way (Ester et al., 1996).

The final step was an ensemble hyperparameter search using Optuna (StackExchange, 2023). For each trial, I ran HDBSCAN* five times with different random seeds, built a co-occurrence matrix from those five labellings, then converted that into a distance matrix to identify the most stable cluster assignments. I measured cluster quality by computing the ARI against the true SLF levels. Repeating this over 100 trials, I selected the parameter set with the highest average ARI. With those optimal parameters in hand, I performed one last five-seed ensemble on the full data set and saved the resulting cluster labels (Pedro, 2022).

Before clustering, all attributes were standardised to zero mean and unit variance. Because HDBSCAN* uses distance computations in high-dimensional space, any variable with a larger numeric range would otherwise dominate the clustering outcome. Scaling each feature ensures that the chosen correlation-distance metric reflects genuine similarity patterns rather than arbitrary differences in units or scale.

4.3 Implementation

The technical implementation of this thesis was carried out entirely in Python, chosen for its flexibility, mature ecosystem of scientific libraries, and suitability for building interactive, data-driven applications. Compared to prebuilt platforms such as ArcGIS or RStudio, Python provided greater control over both data analysis and visualisation design, enabling a seamless integration of preprocessing, modelling, and user-interface development.

The codebase is organised as a modular pipeline:

1. Data aggregation: integration of grid, station, and bulletin data;
2. Label extraction: derivation of ground-truth classes from SLF bulletins;
3. Feature engineering: rolling statistics, lag variables, and dimensionality reduction;
4. Clustering: ensemble HDBSCAN* models optimised via Optuna;
5. Visualisation layer: preparation of results for interactive display.

The final dashboard was built using the Python libraries "Dash" and "Plotly", presenting coordinated views such as maps, histograms, radar plots, and clustering

summaries. These views are designed to reduce cognitive load while maintaining analytical depth, thereby supporting SLF experts in their daily assessment tasks.

Beyond practical considerations, implementing the system in Python was also a deliberate methodological choice: by scripting each step in one environment, the entire workflow remains transparent, reproducible, and easily extensible for future research.

For transparency and reproducibility, the complete source code is made available in the accompanying GitHub repository (see Appendix I).

4.4 Dashboard Layout

Effective dashboard design goes beyond technical implementation and determines how easily experts can interpret and interact with the underlying data (Andrienko et al., 2016; Few, 2006). In this thesis, design choices were guided by established visualisation principles, with the goal of reducing cognitive load and supporting decision-making in operational forecasting. Key considerations included how to provide an overview while enabling detail-on-demand, how to structure the layout and navigation, and how to use visual encodings such as colours, white space, and dividers. This subsection presents the main design decisions and explains how they contribute to usability and clarity in the final dashboard.

The whole dashboard is based on the information-seeking mantra proposed in the paper of Shneiderman (1996):

> "Overview first, then zoom and filter, and finally details on demand."
> (Shneiderman, 1996, p. 2)

I organised the dashboard into three clearly labelled tabs, Cluster Data, Point Data, and Grid Data, so that the user always knows where they are and what information is displayed. Each tab centres its map display on the screen, with all controls (date picker, attribute selector, elevation and aspect filters, etc.) grouped nearby. This layout gives an immediate overview of the data before any interaction. The user can then zoom and pan the map or apply filters directly, and hovering over any feature reveals exact values without cluttering the view. By keeping design elements consistent, using the same colour scale, typography, and control placement across all tabs, the interface minimises distractions and lets experts focus on interpreting the data quickly and accurately (Kitchin et al., 2016; Prieto et al., 2013; Young & Kitchin, 2020).

The entire dashboard itself functions after the information-seeking mantra principle: the cluster data tab serves as the overview, where all data sets are summarised in a single, central visualisation. When the user wishes to zoom in or filter down to specific attributes or regions, the user can simply click one of the two other tabs; each focuses on a particular data type for deeper exploration. Details-on-demand are built in across every view: whether you're inspecting a grid cell or a station point, the exact values pop up on hover or click, ensuring nothing remains hidden (Shneiderman, 1996).

Another key usability principle comes from Stehle and Kitchin (2020), which highlights the importance of providing the user with immediate feedback after any interaction. Whenever the user clicks on a map feature or adjusts a filter, the interface needs to signal that it is processing the request. To satisfy this requirement, I added a classic "loading" spinner that appears as soon as an action is triggered and disappears once the update is complete. This visual cue reassures the user that the system is responding and keeps them informed throughout their exploration.

4.4.1 Cluster Data Tab

The cluster data tab is unique in that it is the only view for which I generated custom data. The calculation is detailed in Section 4.2. The overall layout of this tab follows the recommendations of Badam et al. (2022), placing the map at the centre of the interface with all controls arranged neatly around it. This arrangement naturally draws the user's eye to the map, arguably the most important element, while keeping tools and settings within easy reach.

Immediately beside the map sits an information panel that presents the point data metrics; I'll return to its specific functions later in this chapter. To minimise visual clutter and support focused exploration, the only control on this tab is a date picker, which allows the user to filter the entire view by day. This streamlined approach ensures that the user spends less time navigating the interface and more time interpreting the data (Few, 2006; Shneiderman, 1996). Above the map is a clear, descriptive title that combines the current date and the currently selected cluster (danger level and substep), ensuring the user always has the essential context they need to interpret what they're seeing.

Overview The central map displays the output of the custom ensemble HDBSCAN* clustering algorithm (described in Section 4.2). The HDBSCAN* uses the grid data and automatically groups grid cells together that share similar environmental conditions. To make these clusters immediately meaningful, each cluster is coloured according to the most frequent danger level and substep among its members. By

summarising complex spatial patterns into a handful of colour-coded regions, the map gives experts a clear, "at-a-glance" view of where different risk levels concentrate right from the start (Galbrun & Miettinen, 2012).

Zoom & Filter When a more in-depth interpretation is needed, the user can click any point within a cluster to select the entire cluster. Within the dashboard, a single click corresponds to the whole cluster, regardless of which point is clicked.

To filter temporally, the date picker lets the user swap between all available days, instantly updating both the map and the accompanying data panel. And, of course, standard map zoom controls are built in, letting the user pan and magnify any area for a closer look.

Details-on-Demand When the user needs to inspect the clusters in even greater detail, they can click on a cluster. If point data is available for the selected cluster, two detailed views appear in the right panel. The information panel displays a series of histograms, one for each point data attribute described in Section 3.2, showing only measurements from that cluster, so the user can see how, for example, snow depth or probability values are distributed across that group. If no point data is available, the panel instead shows "No Station data for this date."

Above the histograms, height-aspect radar plots summarise the spatial prevalence of the cluster's selected danger level. Each ring represents a fixed 500 m elevation band across the full data set range, and each quadrant corresponds to a slope exposure (N, E, S, W). Segments are filled only where stations in that band and aspect report the same main danger level as the cluster. Their opacity reflects how frequently this level occurs relative to all stations in that bin. This visualisation highlights, for example, whether the cluster's danger is concentrated at higher elevations or on south-facing slopes. Together, the histograms and radar plots fulfil the "details-on-demand" stage of the information-seeking mantra proposed in Shneiderman (1996), giving precise, context-rich insights exactly where and when it is needed.

4.4.2 Point Data Tab

The point data tab follows the same basic layout as the cluster data view. The map, the primary display of data, sits front and centre, drawing the eye immediately to its spatial display. To the right of the map, the user can find all of the tab's controls neatly organised, just as recommended by Badam et al. (2022). Above the map is a clear, descriptive title that combines the chosen attribute and the current date,

ensuring the user always has the essential context they need to interpret what they're seeing.

Overview When opening the point data tab, Shneiderman's mantra is applied with the overview stage: the first component the user sees is a static map that visualises the default attribute across all station points. By default, the dashboard shows probability_level_3, because Greene et al. (1999) found that most avalanche accidents occur under a Level 3 ("moderate") danger setting, making it both the most incident-associated and a representative snapshot of typical risk. This immediate, map-based overview gives the user a clear sense of how that probability is distributed spatially before the user can move on to zoom, filter, or drill into specific values.

Zoom & Filter Just like in the cluster data view, the point data tab uses a date picker to let the user scroll through each day's subset. A new feature in this tab is the attribute selector, a drop-down menu that lists every point data variable, so the user can instantly swap the map's focus from e.g. snow depth to probability or any other metric. Naturally, all the usual map controls, panning, zooming, and resetting the view, are built in as well, giving the user full interactive control over both time and attribute without ever leaving the tab.

Details-on-Demand To fulfil the details-on-demand requirement, every data point and grid cell in the visualisation is equipped with hover tooltips. Simply moving the cursor over any map feature instantly reveals the exact value being displayed, whether it is a probability, snow depth, or any other attribute. This immediate, on-hover feedback lets the user explore the data at their own pace, exploring precise numbers without cluttering the interface (Jones & Purves, 2008).

4.4.3 Grid Data Tab

The grid data tab follows the same clear structure as the other views: the map remains the focal point, with all controls aligned to its left, and a concise title bar that states exactly what is being displayed looking at (Badam et al., 2022). This consistent layout reinforces Shneiderman's mantra: overview first, then zoom and filter, and finally details on demand. By giving the user the big-picture map up front before zooming and filtering into specific settings (Shneiderman, 1996).

Overview In the grid data tab, the first view the user sees is the static map view, the "overview" in Shneiderman's mantra. The dashboard applies the same approach

used in the point data tab to the grid view by displaying dlModel_Prob3, which represents the probability that each pixel corresponds to danger level 3. Since Greene et al. (1999) demonstrated that most avalanches occur in areas classified at level 3, this metric offers a logical starting point for analysis.

A key design choice was to represent each pixel by the mean of its underlying values rather than their sum. Using the sum would mix data magnitude with the number of contributing observations, whereas the mean provides a normalised estimate that is directly comparable across pixels. This approach highlights the typical value at each location and avoids artificial inflation of pixel intensities (Cleveland & McGill, 1984; Tufte & Graves-Morris, 1983).

Zoom & Filter The temporal filter in this tab, as in the previous two, is managed by a date selector. To switch between different attributes, I've reused the same drop-down menu, keeping the interface consistent and intuitive.

Two additional controls are added in this tab:

1. Height slider: lets the user set upper and lower elevation bounds for the visualisation, to focus on specific altitude ranges.
2. Aspect selector: lets the user choose which slope exposures to display on the map. Up to three aspects can be chosen simultaneously; if a fourth aspect would be selected a, the selector clears and reverts to showing all aspects.

Details-on-Demand Once again, details-on-demand is handled through hover tooltips. As the user moves the cursor over any point or grid cell, a small pop-up instantly reveals the exact value behind that visual. This lets the user access precise numbers whenever they are needed, without adding extra clutter to the display, and keeps the interaction smooth and consistent across all tabs.

This layout defines the structure of the prototype interface and provides the basis for the subsequent usability evaluation.

4.5 Design Choices

Beyond technical implementation, the effectiveness of the dashboard depends on thoughtful design decisions that influence usability and clarity. These choices were guided by established visualisation principles such as overview first, zoom and filter, and details on demand. This subsection explains the rationale behind layout, interaction mechanisms, and visual encodings, highlighting how design considerations support expert use in an operational forecasting context.

4.5.1 Colour Scale

When visualising the data, one of the key decisions is choosing appropriate colour scales. I opted for a sequential colour scale that comes built into Matplotlib called "Hot_r" (see Figure 13). This particular palette moves from yellow through orange to red, with black marking the highest values. My intention was to create maximum contrast between adjacent data values while keeping the progression, from lower to higher risk, as intuitively obvious as possible (Wells et al., 2021).

Figure 13: Sequential colour scale (“Hot_r” palette) used throughout the dashboard to represent avalanche-related probabilities: yellow indicates low instability, red moderate instability, and black high instability or risk values.

I have decided that I use the same colour scale for all three used data types. This is mostly because of consistency. When the user is switching between different tabs in the dashboard, he or she does not have to look at the colour scale first to be able to understand what information is visualised, but will know from the tab before what values are high and what values are low.

4.5.2 White Space and Dividers

The aim for me was to create the three data type tabs in such a way, that the entire map including the controls of the specific tab fits on the screen entirely (Tufte & Graves-Morris, 1983; Ware, 2013). Of course, I have implemented automatic scaling, depending on the size of the screen.

When the specific tab is loaded, the user has to scroll down to fit the map and the controls on the screen, as the tab selection control does take up some space as well. This could have been avoided when implementing a foldable control panel. I decided against the usage of this, since I did not want to implement unnecessarily complex tools, which do not immensely add to the forecasting process.

4.5.3 Map Controls

For the map controls, I have implemented the basic tools provided by the "Plotly" python-library. I have chosen to use the following map interaction controls:

- Download Plot as PNG

- 2D Zoom
- 2D Pan and 2D Tilt
- Auto Scale
- Reset Axes

The first tool lets the user extract the visualisation as a PDF-file for future usage. The next tool is the basic zooming tool, where the user can zoom in and out using the mouse wheel. This ensures intuitive use, as it is the most common way of zooming used on modern computers. To pan and tilt the image, the user has to click on the visualisation and keep the mouse button clicked and can then move the image with the movement of the mouse. The last tools are "auto scale" and "reset axes". Auto scale lets the user reset the visualised in such a way, that all the information is visible in the image, while "reset axes" resets the image to the way it was, when the page was first loaded.

According to Pohl et al. (2012), simple tools are the most effective when doing cognitively heavy work, such as analysing data. Therefore, I have decided to use these few tools as the only tools available on the dashboard.

4.5.4 Visual Feedback

Like I already mentioned in section 4.4, I have implemented an instant visual feedback after the user interacts with the dashboard. According to Stehle and Kitchin (2020), the interaction feedback is important for the user so they know the interaction with the program has worked, and their actions will have a direct consequence.

4.6 Validation

To assess the reliability and usefulness of the developed system, both the clustering outcome and the dashboard's usability were subjected to validation. Quantitative evaluation of clustering quality was combined with qualitative feedback from domain experts. This subsection explains the validation strategy and the methods used to gather and interpret results.

To validate the dashboard's usability, I applied two complementary approaches that together assess both its user experience and its analytical performance:

1. **System Usability Scale (SUS) survey**: This standardised questionnaire measures usability and intuitiveness from the end user's perspective. By asking

participants to rate statements about ease of use, learnability, and overall satisfaction, the SUS provides a quick yet reliable snapshot of how well the dashboard supports effective interaction (Brooke, 1996).

2. **Adjusted Rand Index (ARI)**: The ARI offers a mathematical measure of clustering accuracy by comparing the HDBSCAN* output against a reference grouping. It accounts for chance agreements and ranges from -1 (no agreement) to 1 (perfect agreement), making it an ideal choice for quantifying how faithfully the algorithm captures known spatial patterns (Hubert & Arabie, 1985).

4.6.1 System Usability Scale

One of the primary objectives of this thesis, and of the dashboard itself, is to reduce the cognitive load on SLF experts and to streamline their forecasting workflow. To evaluate whether the dashboard meets these goals, I conducted a System Usability Scale (SUS) survey in two stages.

First, I designed nine custom questions (see Appendix I.) aimed at specific features of the dashboard. After each question, participants rated the difficulty of responding on a 1–5 Likert scale. This step assessed both the usability of individual features and their intuitiveness in practice.

Second, I administered the standard ten-item SUS questionnaire originally proposed by Brooke (1996), which measures overall ease of use, learnability, and satisfaction.

The SUS score is then calculated using the established procedure: for odd-numbered items, one point is subtracted from the raw response; for even-numbered items, the response is subtracted from five. The adjusted values (each ranging from 0–4) are summed across all ten items and multiplied by 2.5, producing a final score between 0 and 100 (Brooke, 1996).

Because the survey was administered to university students, this approach also provided a form of face validation, checking whether "the model logic and input-output relationships appear reasonable" (Rykiel, 1996, p. 235). At the same time, the SUS results contributed to construct validation by confirming that the dashboard aligns with its theoretical design principles, such as reducing cognitive effort and streamlining decision-making (Rykiel, 1996).

4.6.2 Adjusted Rand Index

To evaluate clustering quality and identify the best-fitting parameter combinations, this thesis applies the ARI. The ARI quantifies the similarity between two clusterings

by pair-counting, that is, by counting how often pairs of points are placed together or separately in both partitions (Zhang et al., 2012). By correcting for chance agreements, it provides a standardised measure of how closely the algorithmic clusters align with the SLF bulletin labels, making it well suited for this application.

First, the rand index (RI) is calculated, where a is the number of point-pairs assigned to the same cluster in both partitions and b is the number of point-pairs assigned to different clusters in both partitions (Zhang et al., 2012).

$$\mathrm{RI} = \frac{a+b}{\binom{n}{2}} \tag{3}$$

The RI then gets adjusted by subtraction of the expected value under a permutation model and normalising by its maximum possible value to get the ARI:

$$\mathrm{ARI} = \frac{\sum_{ij}\binom{n_j}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}} \tag{4}$$

In the equation n_{ji} is the count in cells (i, j) of the contingency table of the two partitions, and a_i, b_j are the corresponding row- and column sums (Zhang et al., 2012).

The ARI is able to quantify the results of the HDBSCAN* clusterings, as it can handle any number of clusters and therefore, does not have to be disclosed beforehand (Hubert & Arabie, 1985; Rand, 1971). This is very practical, as that would make the optimisation process a lot more complex. The ARI has a good way of dealing with noise data and outliers, as it counts all noise data points as their own cluster, which does not influence the results of the cluster points (Campello et al., 2013; Ester et al., 1996). And finally, the ARI can compare HDBSCAN*'s internal stability or co-association matrix directly against any reference or consensus matrix, which is great for building and evaluating ensemble clusterings (Fred & Jain, 2005; Strehl & Ghosh, 2002).

Applying the ARI to compare the clustering results with the avalanche bulletins published by the SLF provides operational validation. As Rykiel (1996, p. 234) puts it, "Operational validation is a test protocol to demonstrate that model output meets the performance standards required for the model's purpose."

This thesis combines data preparation, clustering analysis, and dashboard design into a coherent workflow, ensuring that both evaluation and implementation are grounded in consistent data and sound design principles. Together, these steps form the technical and conceptual basis for the results in the next chapter.

5 Results

This chapter presents the outcomes of the clustering analysis and the dashboard implementation. It reports on the performance of HDBSCAN* and evaluates the effectiveness of the three dashboard tabs (Cluster, Point, Grid). User evaluation results, including interaction questions and usability scores, are summarised to assess interpretability and cognitive load. These results provide the empirical basis for the subsequent discussion.

5.1 Evaluation of HDBSCAN*

This subsection presents the outcomes of applying the clustering strategy to the avalanche data set. It highlights the performance of the ensemble HDBSCAN* approach, compares results across alternative methods, and evaluates agreement with expert-defined danger levels.

In the Optuna-driven search I sampled cluster_selection_epsilon over a logarithmic range from 10^{-3} to 10^{-1}, toggled cluster_selection_method between "leaf" and "eom" and distance_metric between "correlation" and "euclidean", varied min_cluster_size from 5 to 100 (in steps of 5) and min_samples from 1 to 100 (in unit steps), and evaluated both settings of "use_pca" and "use_spatial" (True/False) to identify the combination yielding the highest ARI.

The optimal configuration, $\varepsilon = 0.00387$, leaf selection, correlation distance, min_cluster_size = 145, min_samples = 122, with PCA and spatial features enabled, enforces large, densely packed clusters. Such a setup prioritises stability over granularity, capturing only the broadest regional patterns in avalanche danger.

When compared to the alternative methods tested, the custom ensemble HDBSCAN* approach shows a modest but consistent advantage. It outperforms the normal HDBSCAN* (ARI = 0.012), the baseline HDBSCAN* (ARI = 0.198), and a seeded K-Means upper bound (ARI = 0.152), demonstrating that aggregating multiple stochastic runs enhances overall agreement with the SLF bulletin regions.

Although an ARI of 0.2877 falls into the “very weak agreement” band (0.20-0.35, see table 7), in operational practice this level of correspondence still offers valuable high-level insights (Jain, 2010; von Luxburg et al., 2012). Forecasters can use the resulting clusters as a rough map of danger trends and then investigate further via on-demand histograms and elevation/aspect plots in the dashboard to inform site-specific decisions.

Table 6: Optimised clustering parameters for ensemble HDBSCAN*, showing best-fit values identified through parameter search alongside tested alternatives. Clustering quality is evaluated with the Adjusted Rand Index (ARI).

Parameter	**Best Fit**	**Alternatives Tested**
cluster_selection_epsilon	0.00387	10^{-3}–10^{-1} (log scale)
cluster_selection_method	leaf	eom
distance_metric	correlation	euclidean
min_cluster_size	145	5, 10, 15, . . . , 150
min_samples	122	1, 2, . . . , 150
use_pca	True	False
use_spatial	True	False
Adjusted Rand Index (ARI)	0.2877	

Table 7: Interpretation guide for Adjusted Rand Index (ARI) values, adapted from the thresholds used in the R package proposed in Iacobucci et al. (2020). The scale extends the original four categories into finer levels to provide more nuanced assessment of clustering quality in this thesis.

ARI Range	**Interpretation**
0.90 - 1.00	Excellent agreement
0.80 - 0.90	Very good agreement
0.65 - 0.80	Good agreement
0.50 - 0.65	Moderate agreement
0.35 - 0.50	Weak agreement
0.20 - 0.35	Very weak agreement
0.00 - 0.20	Poor agreement (close to random)
0	Random agreement
< 0	Worse than random

5.2 Dashboard Implementation

This section presents the final dashboard layout, structured into three interactive tabs: cluster data, point data, and grid data. Each tab focuses on one type of data representation and follows the information-seeking mantra proposed by Shneiderman (1996) and uses interactive filters and a consistent layout. The aim was to make the dashboard intuitive to use while supporting fast and accurate decision-making for avalanche experts.

5.2.1 Implementation of the Cluster Data Tab

The cluster data tab provides a broad overview by grouping areas with similar forecast values using the HDBSCAN* algorithm. These clusters are shown on a map using colour-coded risk levels (see Figure 14). Clicking on a cluster reveals more detail: radar plots show how risk varies with slope and elevation, and histograms display distributions of point-based values. This tab gives a quick summary while allowing the user to dig deeper into local risk patterns.

Overview As proposed by Shneiderman (1996), the overview forms the first step in the information-seeking process. When the dashboard loads, the map is centrally positioned and occupies the largest portion of the visual space, reflecting its importance. It presents general trends in clustered values, indicating the danger level per region across the entire Alpine area of Switzerland. Using the forecasting regions as a base map provides immediate spatial context and facilitates understanding of the division into forecasting zones.

Directly above the map, a title is displayed, indicating the currently active tab and selected date. Beneath it, the danger level of the selected cluster is shown in bold, ensuring the user can quickly access this critical information.

Additionally, the colour scale legend is placed next to the map, reinforcing the centrality of key visual elements so that all essential information is visible at a glance.

Zoom & Filter To support zooming and filtering, a date selector is provided at the top of the interface, enabling the user to effortlessly switch between the available dates in the data set. This allows for quick temporal comparisons and enables the user to identify changes in spatial patterns over time. The selector is designed for ease of use, ensuring that the user can navigate through the time series without interrupting their workflow.

In addition to the date selector, several interactive map tools are positioned above the map to allow for more refined navigation. These include options for zooming, panning, and resetting the map view, providing the user with flexible control over the spatial extent of the visualisation. A detailed explanation of these tools and their functionality is provided in section 4.5.3.

Details-on-Demand To the right of the map is the "Details-on-Demand" section of this tab. At the top, an elevation-aspect radar plot displays the distribution of avalanche danger across aspect and elevation bands for the selected cluster. The colour scale matches that of the map, promoting consistent interpretation. Hover interactions reveal the elevation range and aspect represented in each bin.

Below the radar plot, histograms show the distribution of data points within the cluster. Each histogram is labelled with the attribute name and its unit (in brackets). A red line indicates the median value, which is also printed in the top-right corner.

Hover data are also implemented on the map itself: when hovering over a data point, the danger level and corresponding coordinates are displayed.

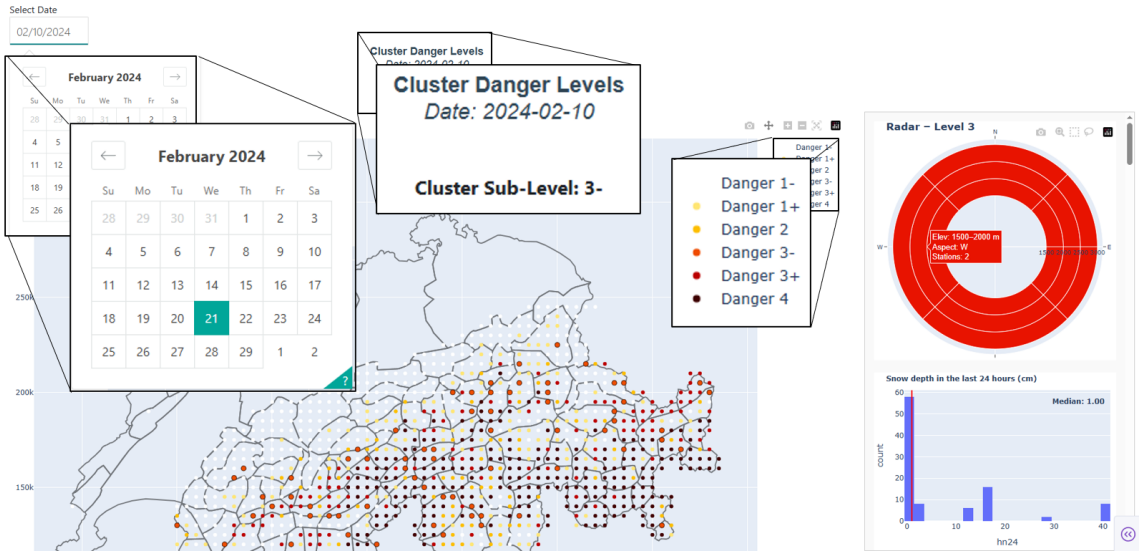


Figure 14: Dashboard implementation of cluster-based avalanche data. The interface displays interactive elements such as a date selector, region-level cluster danger ratings, colour-coded sublevels, and corresponding visualisations (e.g. radar plot and histogram) to support expert assessment and decision-making.

5.2.2 Implementation of the Point Data Tab

The point data tab displays individual IMIS stations and their recorded or derived values, such as snow depth or avalanche probabilities. Each station is shown as

a coloured dot on the map (see Figure 15), with colours representing the selected attribute. The user can switch attributes, change dates, and hover over points for exact values. This view is useful for checking specific locations in detail and comparing station data across time and space.

Overview As in the previously described cluster tab, the map is placed centrally and serves as the main visual component of the tab. It displays the spatial distribution of the point data across the Alpine region of Switzerland. Each point represents a station with associated attribute values, allowing for detailed, location-specific analysis. The visual representation enables the user to quickly assess spatial variations and identify regions with elevated or reduced values.

At the top of the view, a clearly formatted title provides information about the currently selected attribute and the chosen date. This ensures that the user is always aware of the context of the visualised data and can easily compare different attributes or time steps without confusion.

Adjacent to the map, the colour scale legend is displayed. It serves as a key for interpreting the values shown on the map and ensures a quick and intuitive understanding of the visualised patterns.

As in the other tabs, the forecasting regions are used as the base map. This consistent spatial framework enhances geographic orientation and helps the user interpret the values in relation to well-known forecast areas.

Zoom & Filter In this tab, the controls are placed on the right side of the screen, where more space is available compared to the cluster tab. As previously mentioned, a simple control panel allows the user to switch between available dates.

In addition to the map tools and date picker, a new filtering element has been added. The highlighted attribute selector (see Figure 15) enables the user to switch between the available attributes.

Details-on-Demand Details-on-demand in this tab are implemented through interactive hover functionality. When the user moves the cursor over a specific data point on the map, a tooltip appears, displaying both the geographic coordinates and the exact value of the selected attribute for that location.

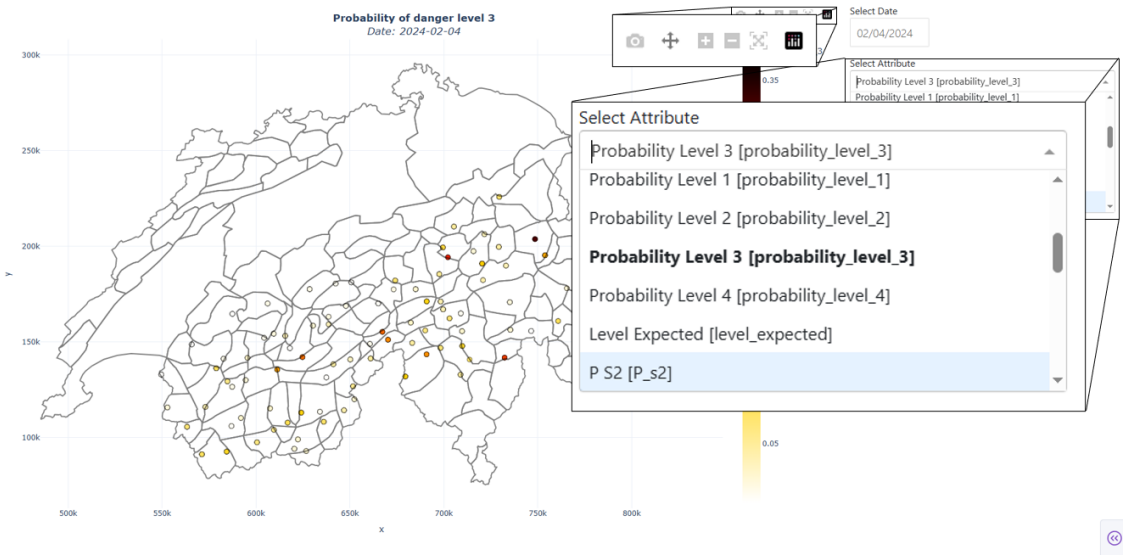


Figure 15: Dashboard implementation of point-based avalanche data. The map displays station-specific values for selected attributes (e.g., probability of danger level 3), with interactive controls for date selection and variable filtering to support targeted expert analysis.

5.2.3 Implementation of the Grid Data Tab

The grid data tab shows interpolated forecast data as a continuous surface across Switzerland (see Figure 16). The user can explore different attributes, filter by elevation and slope aspect, and change the date. Hovering over a cell reveals the exact value. This tab is helpful for spotting regional patterns and comparing terrain conditions across large areas.

Overview As in the other two tabs, the map is the central element of the view. Above the map, a title is displayed, indicating the selected attribute and the corresponding date. Again, the base map consists of the forecasting regions to support spatial orientation.

For the visualisation, 350 × 350 metre grid cells were used. This resolution was chosen as it represents the highest level of detail achievable without negatively impacting dashboard performance or computation time.

Zoom & Filter In addition to the previously mentioned zoom and filter tools, two new controls are introduced in this tab. The elevation range filter allows the user to define an upper and lower elevation threshold for data displayed on the map. The total range corresponds to the minimum and maximum elevations in the data set, divided into equal intervals for selection.

The second new tool is the aspect selector. This enables the user to select one, two, or three slope aspects, filtering the map to show only points corresponding to the chosen orientations. A “clear” button is provided to reset the selection.

Details-on-Demand As in the other tabs, details-on-demand in the grid data tab are provided through hover interactions. When the user hovers over a specific location on the map, a tooltip displays the exact value of the visualised attribute for that grid cell, along with its geographic coordinates in the Swiss LV03 coordinate system.

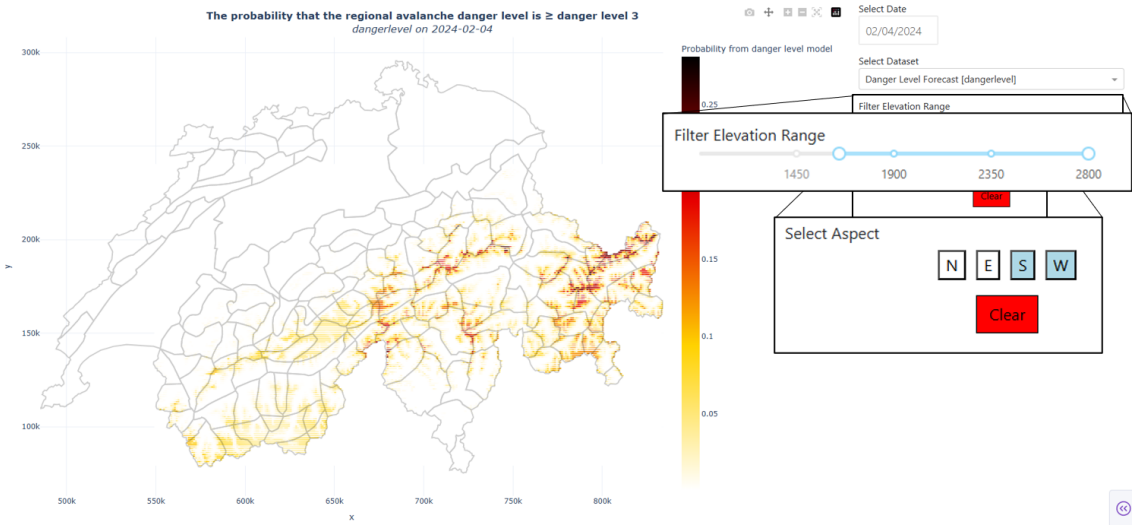


Figure 16: Dashboard implementation of interpolated grid-based avalanche data. The map shows spatial probabilities for regional danger levels with interactive controls for filtering by elevation range and slope aspect, allowing tailored visual analysis based on terrain conditions.

5.3 User Evaluation of Dashboard Interpretability and Usability

This subsection summarises the evaluation of the dashboard, focusing on interpretability, usability, and cognitive load. Through structured tasks and subjective ratings, the survey highlights strengths and weaknesses across the three data representations and provides feedback on the system’s overall effectiveness.

5.3.1 Dashboard Interaction Questions

Across the eight task-focused questions, participants showed solid understanding, though some required more thought (see Figure 17).

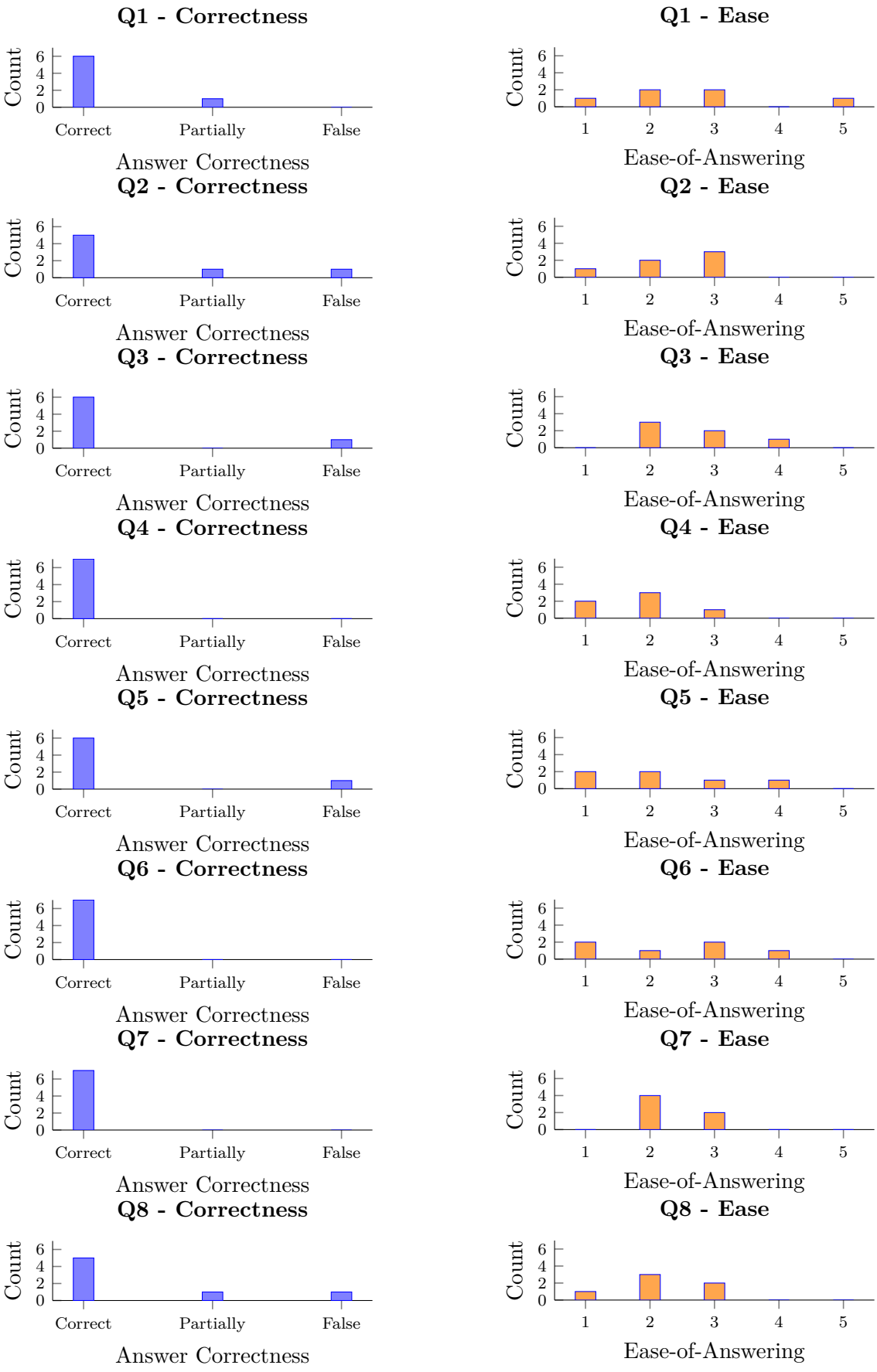
In Question 1, six of the seven testers identified the correct answer, yet their ease-of-answer ratings hovered in the lower-to-middle range, revealing that they needed to pause and consider before selecting the right response. Question 2 echoed that pattern: most responses were accurate, one was only partially correct and another incorrect, and participants again rated the task between “easy” and “medium.” By Question 3, accuracy stayed high, yet ease scores dipped further, underlining that this particular question felt more challenging than its predecessors.

Question 4 showed the best outcome, with every participant answering correctly and most rating it “easy,” except for a single mid-level response. Although Question 5 also achieved near-perfect accuracy, one user flagged it as “difficult,” signalling a notable jump in perceived complexity.

Questions 6 and 7 both recorded flawless correctness, but their ease ratings diverged: users found Question 6 quite demanding, whereas Question 7 landed in that low-to-mid difficulty band. In Question 8, five participants answered correctly, one provided a partially correct response, and one missed the mark entirely; ratings here fell squarely in the moderate-difficulty range.

The ninth custom question, being inherently subjective (it asked which visualisation tab each user preferred), was deliberately excluded from this correctness analysis.

Figure 17: Per-question response analysis for survey items Q1-Q8. Each line shows correctness (left) and ease (right) for that question (1 = "very easy" and 5 = "very hard").



These interaction responses provide context for the formal usability results presented in the next subsection.

5.3.2 System Usability Scale Results

To quantify the dashboard's usability, I followed the standard SUS protocol outlined by Bangor et al. (2008) (see Appendix I. for the full questionnaire). The feedback on Q10 revealed a spread of responses around the midpoint, suggesting users were moderately inclined to adopt the tool in the future (see Figure 18). In Q11, participants generally agreed that the system avoided unnecessary complexity, although a few lower scores hinted at remaining usability hurdles. By contrast, Q12 drew very positive reactions. Most testers found the dashboard easy to use, confirming its intuitive design.

When asked in Q13 whether they would require technical support to operate the dashboard, most respondents disagreed, underscoring the interface's self-explanatory nature and minimal onboarding needs. Q14 echoed that feeling, with users underlining how seamlessly the dashboard's features fit together. Yet Q15 uncovered some inconsistencies in the layout or workflow, as a handful of testers noted areas where the design felt less uniform. Similarly, Q16 exposed slight hesitation around how quickly new users could become proficient, pointing to opportunities for improving the learning curve. The interface's convenience also came under scrutiny in Q17, where several participants still found parts of the navigation somewhat awkward. Finally, Q18 showed that most users felt confident using the system, indicating an overall sense of competence and satisfaction.

Taken together, these responses produced an average SUS score of 68.2. A value of 68.2 is right at the industry benchmark for acceptable usability, corresponding to the adjective rating "OK" (see Table 8). In sum, the dashboard demonstrates solid learnability and well-integrated functionality, though there remains room to simplify complex elements and tighten the user experience for greater consistency.

Overall, the results demonstrate that different data representations influence both analytical accuracy and user perception, with grid data consistently supporting clearer interpretation. The clustering evaluation further showed that meaningful structures could be identified reliably when parameters were tuned to avalanche-relevant attributes. Finally, the usability study indicated that the dashboard achieves acceptable learnability and functionality, though opportunities remain to streamline interactions and reduce complexity. Together, these findings confirm the potential of the developed methods and highlight the key aspects that warrant deeper discussion in the following chapter.

Figure 18: Distribution of System Usability Scale (SUS) responses from university students. Responses are given on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

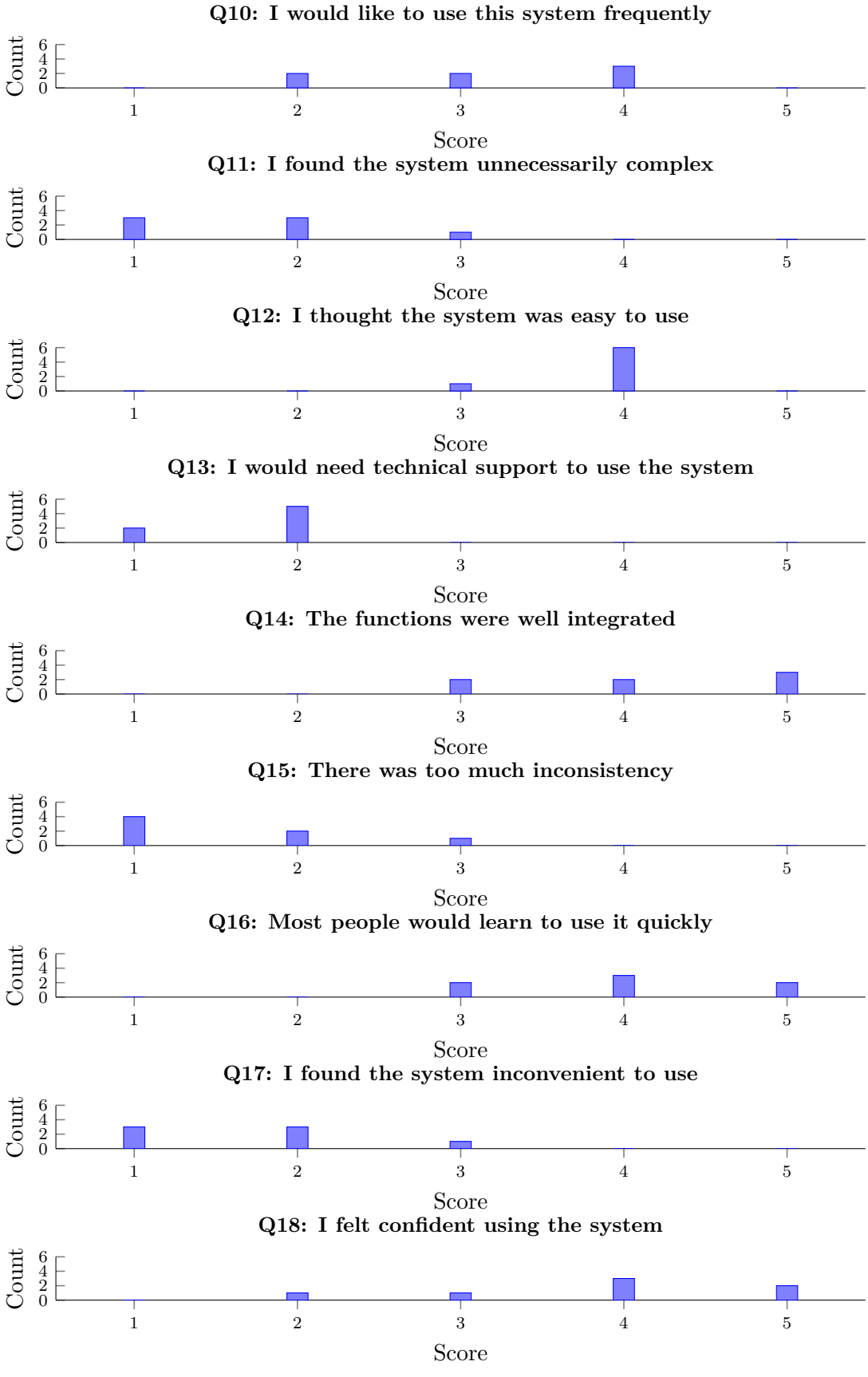


Table 8: Interpretation guide for SUS scores, showing corresponding adjective ratings, percentile ranks, and acceptability ranges, adapted from Bangor et al. (2009) and Sauro (2011).

SUS Score Range	**Adjective Rating**	**Acceptability**
85 – 100	Excellent	Acceptable
70 – 84	Good	Acceptable
68 – 69	OK	Acceptable
51 – 67	Poor	Marginal
0 – 50	Awful	Not acceptable

6 Discussion

This chapter critically reflects on the findings in relation to the research questions. It discusses the prioritisation of attributes, the quality of clustering, and the implications for decision support. It then considers usability aspects and cognitive load, compares the developed dashboard to prior systems, and evaluates the interpretability of different data representations. Limitations of the study are acknowledged, and recommendations for improving both methodology and design are provided.

Besides the custom clustering algorithm (described in section 4.2.3), the central contribution of this thesis is the system's implementation. The multi-tab layout, seamless integration of clustering results, and the design choices grounded in operational use represent substantial advances over existing tools. The dashboard's architecture, described in Sections 4.4 - 4.5, combines real-time data handling with intuitive interaction features such as hover-tooltips, map filters, and visual feedback cues, enabling forecasters to explore complex model outputs without losing situational context.

For better readability, the research questions and the subquestions are repeated and indicated in which subsection they are answered:

1. **Integrating Predictive Analytics with Expert Judgment**: How can machine-learning-driven risk predictions be blended into a dashboard so as to support expert decision-making in avalanche forecasting? → 6.1
    - Determine the most essential information required for effective snow and avalanche risk assessment, prioritising attributes that provide the greatest predictive and analytical value.
2. **Optimised Data Visualisation for Expert Use**: How can critical snow and avalanche data be presented in a clean, intuitive, and user-centred manner, ensuring clarity and ease of interpretation for forecasting experts? → 6.2
    - Which visualisation and interactive design techniques best enhance clarity, usability, and minimise cognitive load when presenting snow and avalanche data to forecasting experts?
3. **Comparative Analysis of Data Representation**: What are the differences in interpretability between point data, grid data, and cluster data, with a focus on their respective contributions to snow and avalanche forecasting accuracy? → 6.3
    - What are the advantages and limitations of each data representation method in terms of interpretability for avalanche hazard assessment?

6.1 Machine-Learning–Driven Risk Prediction and Attribute Prioritisation

This subsection examines which attributes were most influential in the clustering process and how these relate to expert assessments. It highlights which variables consistently contributed to meaningful groupings and considers their implications for avalanche forecasting.

6.1.1 Feature Prioritisation for Risk Prediction

During the programming process, my supervisors and I made the deliberate decision not to exclude any available attributes, even when some showed high correlation or appeared redundant. While this choice prevented me from simplifying the forecasting process by narrowing down the data set, it reflects how avalanche forecasters actually work in practice. Prior research has shown that experts often rely on multiple overlapping indicators to build confidence in their assessments and to cross-validate signals in complex, uncertain environments. Lachapelle (1980) highlighted how forecasters use "redundant" information as part of their mental modelling of snowpack behaviour. In this context, retaining all variables ensures the dashboard supports the nuanced and iterative reasoning processes that professionals depend on.

Consequently, although my first subquestion, identifying the most essential data attributes, cannot be directly answered through exclusion, the design instead reflects a more realistic and operationally valid approach: treating all SLF-provided attributes as potentially important, in line with expert practice.

6.1.2 Clustering Quality & ARI Interpretation

As detailed in Section 5.1, the custom ensemble HDBSCAN* run achieved an ARI of 0.2877, placing it just below the threshold for what might be considered a "weak agreement" for clustering. In other words, the algorithm clearly outperforms a random assignment of points to danger levels, an ARI of zero, but still delivers only modest separation overall (Hubert & Arabie, 1985). This raises the question of why the clustering did not produce more robust results, and several factors likely contribute to its limited effectiveness.

A likely explanation is that the data are simply too noisy for HDBSCAN* to cluster effectively. In creating second-order polynomial (quadratic) features via the pairwise comparisons, described in Section 4.2.1, I used every interpolated grid variable, producing a vast, highly dimensional feature set. While these additional

dimensions can capture complex relationships, they also amplify random variation, making it harder for the algorithm to discern meaningful groupings (Aggarwal et al., 2001). Importantly, these quadratic features are crucial to the ensemble approach of HDBSCAN*, so omitting them is not an option. In practice, reducing the feature set consistently caused the ARI to drop even further, confirming that I must accept the higher noise levels in order to preserve critical information.

One further reason for the low ARI value is the class-balanced subsampling step described in Section 4.2.3. In this process, the script randomly retains at most 100 observations for each date and avalanche danger level, reducing larger groups so that class sizes are equal. Although this prevents the most common values from dominating the clustering, the randomness of selection means it may keep only the highest values, only the lowest, or an unpredictable mix, the latter being ideal but not assured. Consequently, there is a risk that extreme or "peak" observations, which often carry the strongest signals for distinguishing clusters, are systematically discarded (Lohr, 1999). One possible way to mitigate this would have been to apply stratified random sampling, where samples are drawn from defined subgroups (e.g., based on snow depth or instability index bins) within each class (Hastie et al., 2009; Lohr, 1999). This could have preserved a broader range of conditions while still keeping class balance, and may have improved clustering performance (Jain, 2010). Moreover, because the sampling step disrupts the original chronological sequence of measurements within each sector, features such as moving averages, variances, and lagged differences become noisier and less representative of genuine temporal trends (Chatfield, 2004). Combined, these effects tend to flatten key distribution peaks and break temporal continuity, making it more difficult for HDBSCAN*, even after LDA projection, to discover meaningful clusters (Campello et al., 2013). This compromise, though necessary to maintain computational feasibility, inevitably sacrifices some of the data set's richness and temporal coherence (Aggarwal, 2015; Keogh et al., 2001).

I included the class-balanced subsampling step primarily due to practical time constraints: running the parameter optimiser on the full data set could take up to 72 hours per iteration, which was simply untenable for an iterative workflow. Furthermore, because HDBSCAN* operates in an unsupervised manner (as discussed in Section 4.2.1), it inherently lacks any guarantee of recovering the human-assigned bulletin labels. Consequently, even under ideal conditions, the algorithm may struggle to align its clusters perfectly with expert classifications (Hastie et al., 2009).

While the ensemble HDBSCAN* approach yielded a weak ARI of 0.28, this outcome provides valuable insight into the challenges of clustering avalanche-relevant environmental data. In particular, the results demonstrate how noise, high feature correlations, and artificial class balancing can complicate density-based clustering in this domain (Aggarwal, 2015; Jain, 2010). Rather than undermining the value of the

method, the low performance highlights the limitations of current approaches and establishes a benchmark for future studies (von Luxburg et al., 2012). Importantly, documenting this outcome avoids the common problem of "positive results bias" and ensures transparency in reporting what does not work as expected. This knowledge is critical for refining both methodological choices (e.g., alternative algorithms such as OPTICS or agglomerative clustering with spatial constraints (Ankerst et al., 1999; Ester et al., 1996; Müllner, 2011)) and data preprocessing strategies (e.g., stratified sampling (Lohr, 1999)).

6.1.3 Implications for Decision-Support Integration

Given the disappointingly low ARI, including the clustering results in the dashboard risks misleading experts at first glance (Marchionini, 2006; Parasuraman et al., 2000). If the experts try to interpret a weak clustering overview they trust, they may have to spend extra effort overriding or "correcting" it during deeper analysis, ironically increasing their cognitive load rather than reducing it (Parasuraman et al., 2000). Even so, I decided to keep the clusters as an entry point, since they still provide a valuable high-level visualisation of spatial patterns. Coupled with the embedded histograms and elevation-aspect plots, the dashboard maintains strong analytical power, allowing experts to quickly explore key trends before diving into more detailed data inspection (Marchionini, 2006).

Overall, these findings show that while ML techniques can highlight relevant features and capture patterns in avalanche data, their usefulness depends heavily on transparent interpretation and integration with expert judgment.

6.2 Usability and Cognitive Load Assessment of the Dashboard

This subsection discusses the dashboard's usability and its impact on forecasters' cognitive load. Insights from the expert survey are combined with design considerations to assess how well the system supports fast, intuitive decision-making.

These usability assessments are grounded in the dashboard structure and features described in Section 5.2, where the cluster, point, and grid data tabs were implemented according to user-centred design principles.

6.2.1 Visualisation Techniques that Enhance Clarity

My second research subquestion focuses on identifying visualisation techniques that help organise and simplify the data interpretation process.

By following the propositions of Badam et al. (2022) and Shneiderman (1996), the solution to an intuitive, user-centred and easy-to-use dashboard is as follows:

The dashboard design follows the information-seeking mantra of Shneiderman (1996) by guiding the user through a clear and structured exploration of the data. In the cluster data tab, the user first sees an immediate overview of avalanche danger levels across Switzerland via colour-coded regional clusters. This gives a quick sense of spatial risk distribution. From there, the user can zoom and filter by date to focus on specific time frames or areas of interest. Once a cluster is selected, the dashboard provides detailed visual feedback: histograms summarise point-based values within the cluster, and radar plots show how risk varies by elevation and slope aspect. To gather further information about specific attributes or areas of interest, the user can then choose one of the other two available data tabs. Additionally, in all tabs, hovering over a point or grid cell reveals exact values via tooltips, giving the user access to precise data on demand. This layered design supports both quick scanning and in-depth analysis, as intended by the mantra.

According to Badam et al. (2022),the map, as it is serving as the principal medium for data presentation and visualisation, should be positioned centrally on the screen to naturally attract the user's attention. Moreover, design parameters, such as the selected colour scale and the configuration of interface controls, must remain consistent throughout, which can be seen in figures 14-16, showing the consistent layouts of the three different data type tabs.

6.2.2 Cognitive Load Assessment: SUS Findings

As I have already addressed before, the results of the ease-to-answer questions can suggest that the interpretation tasks of my questionnaire were between easy and medium, with some exceptions, where the users found the questions hard to answer.

Based on these findings, I argue that my tool maintains a low cognitive load during data analysis (Tufte & Graves-Morris, 1983; Ware, 2013). Because the participants were students rather than SLF forecasters, the questionnaire tasks had to be simpler than those experts face in practice. Even so, the questions still reflected the essential spatial-analysis steps required to interpret avalanche data, making the results meaningful for evaluating dashboard support. By automating routine data handling, the dashboard enables the user to concentrate on identifying spatial patterns and

formulating avalanche-forecast assessments. In doing so, it reduces mental effort and enhances both the efficiency and clarity of the forecasting process (Parasuraman et al., 2000; Sweller et al., 1998).

To address the second subquestion, how to design a dashboard that minimises cognitive load, I have to focus on three key principles. First, the dashboard should function as a single, integrated tool that automatically loads data for easy comparison, and allows experts to work without jumping between applications or browser tabs (Andrienko et al., 2016; Few, 2006). Second, the interface itself must be deliberately minimalist: every element should serve a clear purpose, and any redundant controls or decorative elements that could distract the user should be removed (Tufte & Graves-Morris, 1983; Ware, 2013). Finally, consistency is important. A uniform layout, using standardised widgets, predictable interactions, and a coherent colour scale, helps the user form reliable mental models, so they can concentrate fully on spatial analysis and forecasting rather than on navigating the interface (Mahajan & Shneiderman, 1997; Nielsen, 2000; Shneiderman, 1996). These design choices, illustrated in Figures 14–16, show how the final implementation puts these principles into practice and supports efficient spatial analysis and forecasting.

6.2.3 Recommendations for Dashboard Refinements

At the end of the SUS-survey, I have implemented a feedback question, where the testers can add their remarks to parts of the dashboard which still need improvement. The feedback was mostly positive remarks, but some parts still needed improvements according to the users.

Originally, only the point data tab included a base map, which made it harder to interpret both the cluster and grid tabs. To improve interpretability and ensure a cohesive visual context, I have added base maps to those two tabs as well, making it easier for the user to understand the data and maintain a uniform representation across all views (Shneiderman, 1996).

Users also flagged the hover tooltip on the cluster tab: it only displays “Cluster-ID” and “Danger Level,” which, while familiar to experts, could confuse anyone not proficient in informatics terminology. Since the target audience already understands these labels, I kept them but added a visual highlight for the selected cluster. This enhancement makes it immediately clear which point is active, boosting usability without diluting the data’s technical precision (Shneiderman, 1996).

Some users suggested adding descriptive labels for each data attribute because they found the distinctions unclear. However, the dashboard was specifically designed for experienced avalanche forecasters who are already familiar with these variables

and their meanings. For that reason, I chose not to add additional explanatory text in order to keep the interface clean and focused, as recommended in Shneiderman (1996) to minimising cognitive load. That said, this feedback highlights a broader challenge in the evaluation process: most of the test participants were not domain experts. As a result, their expectations and needs sometimes differed from those of the intended user group. While this mismatch does not invalidate their feedback, it does suggest that some issues raised, such as a lack of labels, might be less relevant in real expert use but are still important to consider for broader usability or future testing with professionals.

6.2.4 Comparison with Prior Systems

As a point of reference, I compared my dashboard with an earlier prototype, proposed in Purves et al. (2003), designed for expert avalanche forecasting (see Figure 19). This system, which relied on a nearest-neighbour approach, presented key variables such as snow depth, temperature, wind, and historical analogues in a compact matrix-style interface. While it was effective for domain specialists familiar with the data, it lacked spatial visualisations and did not incorporate any form of interactive map or terrain-aware context. In contrast, my dashboard integrates point, grid, and cluster-based spatial data across Switzerland and offers interactive tools for filtering by elevation, aspect, and date.

As aforementioned, Purves et al. (2003) as well as Gassner and Brabec (2002) present nearest-neighbour-based visualisations that display only one attribute value per day, without spatial differentiation and with limited interactivity. In contrast, the dashboard developed in this thesis allows the user to explore multiple attributes simultaneously at a 5×5 km grid resolution, while maintaining both spatial and temporal context. Similarly, many existing avalanche forecasting dashboards show raw model outputs but do not combine clustering approaches or point measurements. Here, these data sources are integrated into a single interface, enabling a richer and more contextual interpretation.

Furthermore, while the earlier system focused primarily on tabular data and supported mostly temporal reasoning, my implementation shifts the emphasis toward spatial reasoning. This allows forecasters to immediately detect patterns, compare regions, and assess terrain effects more effectively. Another important difference is that my dashboard supports modern probabilistic forecast models, such as those derived from grid data, which were not part of earlier designs.

Taken together, these differences illustrate the broader trajectory of avalanche forecasting tools: from static bulletins and coarse regional maps toward highly interactive, high-resolution platforms that merge multiple data sources. The current

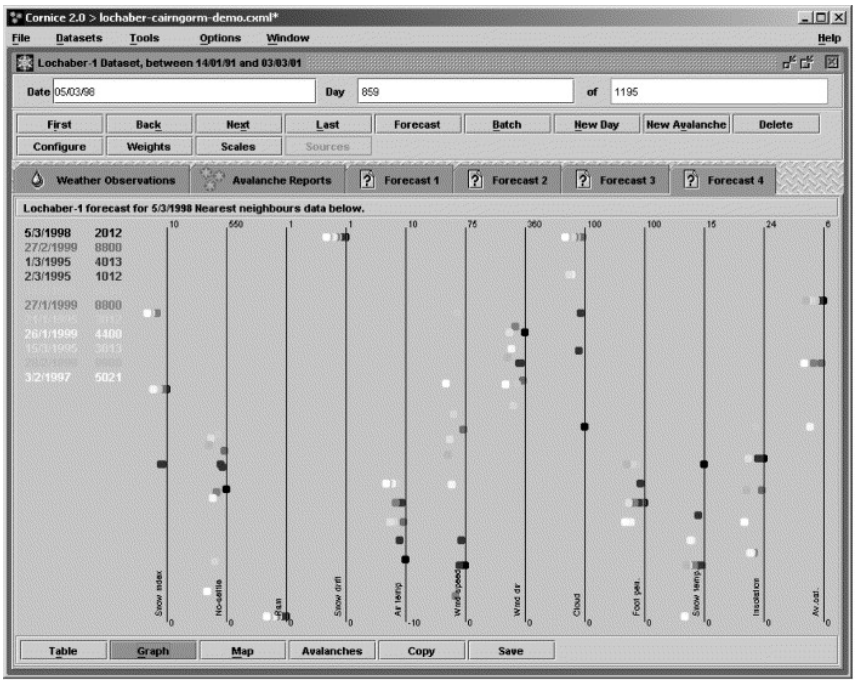


Figure 19: Visualisation of attributes from similar days using the nearest-neighbours method (Purves et al., 2003). The prototype displays historical weather and avalanche forecast data in a compact matrix for expert comparison, but lacks spatial differentiation and interactivity

implementation demonstrates how increased spatial granularity, temporal resolution, and analytical integration can translate into more actionable decision support for operational forecasters.

In sum, the usability evaluation confirmed that the prototype supports expert workflows effectively, though reducing cognitive load further will require simplifying interactions and clarifying complex visual elements.

6.3 Interpretability of Data Representations and Forecasting Accuracy

This subsection evaluates how effectively the three dashboard tabs, cluster, point, and grid, support the interpretation of avalanche-relevant data. It discusses the relative strengths and weaknesses of each representation and considers how well they convey spatial patterns, temporal trends, and critical attributes to expert users.

Each data type was presented to the user in a dedicated tab, following a consistent design structure based on Shneiderman's information-seeking mantra (overview → filter → detail). The cluster view (Section 5.2.1) summarises regional avalanche risk using pre-defined spatial zones. The point view (5.2.2) displays measurements

for individual locations, while the grid view (5.2.3) offers a continuous spatial interpolation of danger levels across the terrain.

In order to assess how intuitively users could interpret these formats, I evaluated their answers to the performance tasks grouped by data type and analysed both the correctness and subjective ease ratings. Figure 20 shows the average scores per visualisation type, offering insight into how well each method supports decision-making.

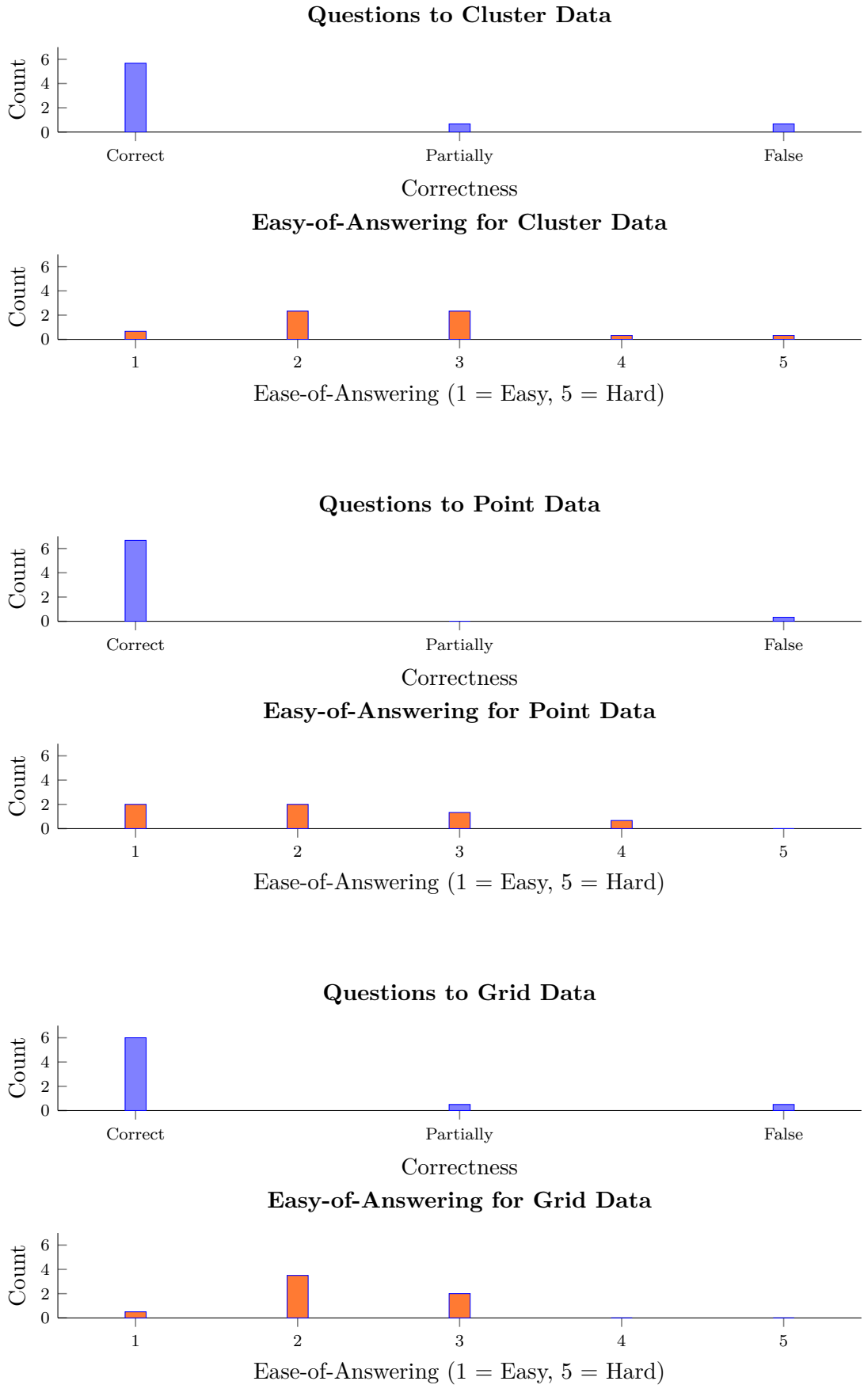
6.3.1 Comparative Performance of the Data Types

When I look only at the correctness of the answers (see Figure 20), the differences between the data types are relatively small. Overall accuracy is high across all three, which limits the insights I can draw about how intuitively each data type is used. For cluster data, 80.9% of responses were correct and 9.6% were partially correct or incorrect. Point data achieved the highest accuracy, with 95.3% correct and 4.7% incorrect, while grid data fell in between at 85.7% correct and 7.1% partially correct or incorrect.

Differences became clearer when considering ease-of-answering (also in Figure 20). Cluster data tasks were most often rated “Easy” (38.9%) or “Medium” (38.9%), with smaller shares marked as “Very Easy” (11.2%) or “Hard/Very Hard” (11.0%). Point data produced a broader spread: one third of responses each were “Very Easy” (33.3%) or “Easy” (33.3%), while “Medium” (22.2%) and “Hard” (11.2%) were less frequent, with no “Very Hard” ratings. Grid data, by contrast, received overwhelmingly positive feedback: “Easy” (58.3%), “Medium” (33.3%), and “Very Easy” (8.3%), with no responses in the “Hard” or “Very Hard” categories.

Taken together, while correctness scores suggest all three representations are usable, ease-of-answering ratings indicate that grid data is perceived as the clearest and most intuitive format, followed by point data and then cluster data (Herla et al., 2024; Horton et al., 2020).

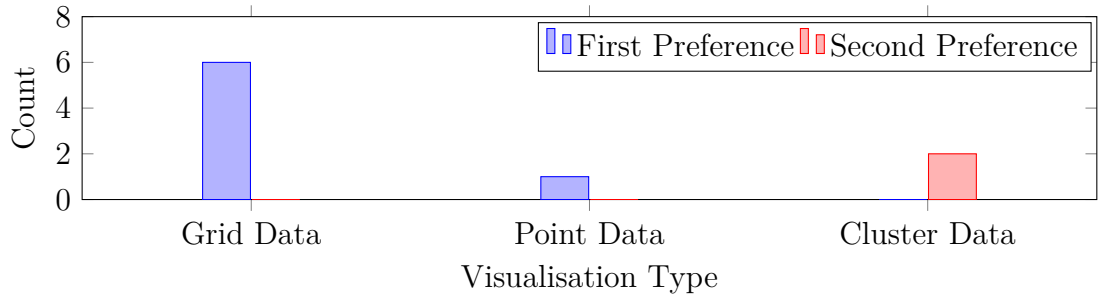
Figure 20: Comparison of participant performance and perceived ease-of-answering questions based on different data types. Each pair of charts shows average correctness (top, in blue) and ease of answering (bottom, in orange).



The question 9 of dashboard interaction questions, which has been left out due to the subjectivity of the question, can now be used. In this question, I ask the users to name the most useful data type to determine the avalanche danger. The results can be seen in Figure 21. I asked only for one data type, but three testers gave two answers, which I nevertheless chose to include in the graphic.

It is visible that the grid data is the clear favourite by the users in regard to the usability to determine avalanche danger. Second is the cluster data and third is the point data type. According to Herla et al. (2024) and Maissen et al. (2024) forecasters using operational dashboards, which have to make decisions under time pressure, gridded (interpolated) data sets, providing seamless coverage at consistent resolution, are preferred.

Figure 21: Participant preferences for data types in location-specific avalanche forecasting, based on responses to Question 9. While users were instructed to select a single preferred data set, Grid, Point, or Cluster, some provided both first and second choices, which are shown in blue and red, respectively.



6.3.2 Insights on Data Readability

Next, I will summarise how each visualisation style compares in terms of intuitive interpretation: cluster data, point data and grid data. I will present which cues helped experts, in this case the testers, quickly grasp conditions and which formats required additional guidance.

Cluster Data The mixed results for this data type stem from several factors. Feedback provided by testers after the survey was carefully reviewed, and adjustments were made where necessary.

When the user hovers over an aggregation point, a “detail-on-demand” tooltip appears. In the cluster tab, this tooltip shows both the Cluster-ID, generated by the HDBSCAN* enumeration, and the cluster’s assigned danger level with the sublevel.

Although those unfamiliar with clustering terminology might find the labels confusing at first, they're essential for distinguishing clusters that share the same danger level colour. By keeping this information on hover rather than displaying it by default, the visualisation remains clean while still offering the precise information when the experts need it (Moore & Benbasat, 1991).

Further, the orientation, when the user has already selected a cluster, was not easy, as it is possible that there are multiple clusters with the danger level, yet are not counted as one. This can happen, as only regions with very similar values are labelled as the same cluster. Therefore, when two clusters have different value distributions in the cluster, but the same danger level scale, they are classified as two different clusters with the same danger level scale (Jain, 2010; von Luxburg et al., 2012). This problem has been addressed by subtly highlighting the chosen cluster with a thin black border around the pixel points. Yi et al. (2007) proposes that highlighting, in this case with the black border, helps the user a lot in differentiating between selected and focusing on complex visualisations.

Finally, the radar plots that show each selected cluster's elevation-aspect distribution were a new feature for the students testing the dashboard, so it's understandable that they initially had some difficulty interpreting them (Kendrick, 2019). I deliberately chose not to explain the radar plots in advance, as I wanted to see whether users would be able to understand them on their own and test the intuitiveness of it.

Point Data The low results seen in figures 17 and 20 of the point data tab are probably due to the nature of the posed questions. I asked the users to determine how many station points of a specific attribute are over a certain value. If the users exactly counted the number of points and had to determine the exact colour hue, this is relatively hard and tedious work. Cleveland and McGill (1984) showed that judging subtle colour hues is among the most difficult visual tasks, and Ware (2013) emphasised that distinguishing such fine hue differences makes counting coloured points both tedious and error-prone.

Nevertheless, the point data carries a big part in the accuracy of the avalanche danger interpretation accuracy.In this tab, the widest range of different data attributes can be visualised, providing a rich basis for analysis that, according to Mayer et al. (2023), is important for understanding snowpack behaviour.

That is why I believe that, even though the exact interpretation of the data might be more difficult compared to the other two visualised data types, the point data tab remains very important. It provides raw data values that may be particularly useful for experts. However, less experienced users, like the testers in my case, tended to prefer data that had already been aggregated in some way.

Grid Data The grid tab was the only one that consistently produced correct answers, and users gave it high ease-to-answer ratings. I attribute this to two main factors. First, like the cluster tab, it provides a continuous spatial display, unlike the point data tab, which is limited to discrete locations. By showing the entire Alpine surface at once, it frees the user from mentally filling gaps between isolated points, making the information immediately clearer and easier to interpret (Horton et al., 2020). The second reason may be familiarity: the testers may be accustomed to interpreting grid-based outputs from their studies, so this representation feels more natural and requires less effort to use effectively.

The other important thing to state is that the tools (the aspect-selector and the elevation-slider) seem to be really straightforward and intuitive to use. Therefore, the user did not have problems using the tools given to them by the dashboard and could extract relevant details easily (Norman, 2013; Shneiderman, 1996).

To summarise, while both the cluster and point data visualisations offer valuable insights, they exhibit limitations in terms of spatial continuity and interpretability. The cluster-based data, despite simplifying regional patterns, occasionally obscures intra-cluster variability. Similarly, point data is precise but can be visually overwhelming and lacks spatial generalisation. In contrast, the grid data provides a continuous and terrain-aware representation of avalanche danger, allowing for immediate recognition of spatial patterns and risk hotspots. Supported by user evaluation results and ease-of-interpretation scores, the interpolated grid emerged as the most intuitive and effective data type for assessing avalanche risk within the dashboard at its current state.

6.4 Limitations

This subsection acknowledges the main limitations of the methods and results. It considers data availability, clustering reliability, and evaluation constraints, outlining how these factors may have influenced the findings.

As discussed in Section 4.2.3, I limited the clustering input to grid data because the point data set not only lacked full alpine coverage but also produced poorer results. In principle, having comprehensive point measurements could enrich the feature set and improve the ARI by introducing valuable local signals (Fromm & Schönberger, 2022; Schweizer et al., 2008). However, adding those point data dramatically increases dimensionality, and with it the level of noise, negating the benefit of any extra information, as shown in Table 5 (Jain, 2010; von Luxburg et al., 2012).

Ideally, the parameter optimiser would have had access to the full daily data set when tuning clustering parameters (Maissen et al., 2024). However, given the

computational limits I faced, it was not practical to process everything at once. As a result, I performed parameter optimisation on a smaller, less-than-ideal subset of the data.

The validation process was constrained by the small, fairly homogeneous test group: only seven participants completed the questionnaire, all of whom were Master's students with substantial GIS experience. According to Nielsen (2000), usability testing can often reveal the majority of interface issues with as few as five participants, which supports the value of this limited sample size. However, Nielsen's guideline primarily applies to general interface usability. Although the dashboard was expressly designed for expert users, including beginners in future evaluations would offer valuable insight into its true intuitiveness and help us understand whether non-specialists can arrive at correct conclusions without prior training.

Another worthwhile enhancement would have been to involve SLF forecasting experts in the formal testing phase. While I did consult closely with an SLF specialist during development, gaining targeted advice on which features to implement and which to omit, a dedicated evaluation by their forecasters would have enriched the results with real-world operational feedback (Kushniruk & Patel, 2004).

At this stage, it is essential to acknowledge that the general approach to natural-hazard assessment inherently embeds a degree of uncertainty. Consequently, even the most precise interpretation of current environmental observations and probabilistic data cannot guarantee flawless forecasts, and the avalanche predictions produced may nonetheless diverge from actual outcomes (Morss et al., 2008).

In summary, the discussion shows that continuous data representations and density-based clustering can improve interpretability, while usability findings underline the need for clarity and simplicity in dashboard design. These insights establish the strengths and limitations of the prototype and point toward refinements that can guide future development.

7 Conclusion

Avalanche forecasting remains a complex and cognitively demanding task that requires synthesising large volumes of spatial and temporal data under time pressure. Despite significant advances in physically based models like SNOWPACK (Schweizer et al., 2003; Wever et al., 2016) and the dense observational coverage provided by networks such as IMIS (Pérez-Guillén et al., 2022), current operational workflows rely heavily on manual data inspection and expert interpretation. This increases the risk of mental fatigue, inconsistency, and delayed decision-making, particularly in high-risk scenarios (Statham et al., 2018).

This thesis addressed this challenge by designing and prototyping an interactive, field-based dashboard tailored specifically to the needs of avalanche forecasters. The system integrates heterogeneous data sources, including model outputs, ensemble forecasts, and real-time sensor data, and visualises them using intuitive, terrain-aware geovisualisation techniques (Horton et al., 2020). Emphasis was placed on design strategies aimed at reducing cognitive load, enhancing temporal interaction, and supporting uncertainty-aware decision-making (Andrienko et al., 2016; Kunz et al., 2010).

The dashboard architecture follows established principles of user-centred design (Young & Kitchin, 2020) and performance optimisation (Badam et al., 2022), enabling fast, interactive exploration of forecasting scenarios. While not yet deployed in operational contexts, the prototype demonstrates how digital tools can help experts shift from manual data scanning to integrated spatial reasoning and scenario assessment. Although this work focuses on snow avalanches, the underlying concepts are transferable to other natural hazards that involve spatio-temporal uncertainty, such as landslides, floods, or wildfires (Cappabianca et al., 2008).

This thesis also demonstrates the scientific value of attempting, documenting, and critically evaluating an ensemble clustering approach for avalanche forecasting even when clustering quality fell short of initial expectations. The modest ARI results observed here reflect not failure but the sensitivity of density-based methods to input representation, feature dimensionality, sampling strategies, and class balance in noisy environmental data (Aggarwal et al., 2001; He & Garcia, 2009; Hubert & Arabie, 1985; Lohr, 1999). Reporting these limitations transparently contributes a valuable benchmark for future research.

The contributions of this thesis lie in two complementary dimensions:

1. Delivering a functional, expert-oriented dashboard that supports exploratory hazard analysis (Horton et al., 2020; Young & Kitchin, 2020).

2. Providing methodological insight by showing that not all ML approaches yield operationally meaningful groupings in avalanche data (Maissen et al., 2024), a result-aware account that future researchers and practitioners can build upon (Jain, 2010).

Taken together, these contributions highlight that the value of this thesis lies not in delivering a definitive operational tool, but in transparently charting both the potential and the limitations of clustering-based approaches for avalanche forecasting. By documenting what worked and what did not, the work provides a clear reference point for researchers and practitioners seeking to develop the next generation of forecasting tools.

7.1 Future Work

Given the underwhelming performance of the modified ensemble HDBSCAN* approach, it is clear that additional investigation is needed. Although I have already detailed the method's limitations in Section 6.4, it seems fruitful to explore two avenues: first, further refining the hierarchical clustering strategy itself; and second, what I believe holds greater promise, evaluating alternative ML algorithms. In particular, supervised techniques such as Random Forests (Breiman, 2001; Mayer et al., 2022), Support Vector Machines (Cortes & Vapnik, 1995), or Neural Networks (Goodfellow et al., 2016) could offer stronger predictive power, and have already shown promise in related avalanche-forecasting studies (Fromm & Schönberger, 2022). However, these methods typically require fully labelled training data and a predefined number of target classes, requirements that lie beyond the scope of this thesis. Nonetheless, pursuing such approaches in future work may ultimately yield more robust avalanche-forecast models.

A promising enhancement not implemented in this thesis would be a "custom view" tab, allowing the user to select any attribute, regardless of data type, for visualisation. By enabling attribute toggles and layering semi-transparent maps, experts could compare two or more variables spatially in one unified view. Similar approaches to modular, user-driven dashboards have proven effective in other domains (Badam et al., 2022; Young & Kitchin, 2020). Integrating the existing analysis tools, such as histograms or elevation-aspect plots, into this tab would be more complex, but I believe it would significantly enrich exploratory workflows. Similarly, adding threshold-based filtering within each data-style tab would offer valuable precision, though I lacked sufficient time to develop these features before project completion.

A logical next step for future research would be to apply this clustering approach to other regions worldwide. By testing it under different climatic conditions, I could

determine whether it performs better elsewhere; perhaps the algorithm struggles specifically with Switzerland's alpine climate and may yield stronger results in more temperate or homogeneous environments. Cross-regional comparisons have already highlighted how operational forecasting chains must be tailored to local conditions, such as the implementation of snowpack models in Norway (Herla et al., 2024), or the adaptation of risk mapping methods in different alpine settings (Cappabianca et al., 2008). Pursuing similar evaluations would help establish whether the limitations observed here are region-specific or more fundamental to the clustering approach itself.

References

Aggarwal, C. C. (2015). *Data mining: The textbook* (Vol. 1). Springer.

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. *Proceedings of the 8th International Conference on Database Theory (ICDT), 1973*, 420–434. https://doi.org/10.1007/3-540-44503-x_27

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I., & Wachowicz, M. (2008). Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research.

Andrienko, G., Andrienko, N., Dykes, J., Kraak, M.-J., Robinson, A., & Schumann, H. (2016). Geovisual analytics: Interactivity, dynamics, and scale.

Andrienko, G., Andrienko, N., & Wrobel, S. (2010). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *12*(2), 38–46. https://doi.org/10.1145/1882471.1882476

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, 49–60. https://doi.org/10.1145/304182.304187

Archibald, C. L., Summers, D. M., Graham, E. M., & Bryan, B. A. (2024). Habitat suitability maps for Australian flora and fauna under CMIP6 climate scenarios. *GigaScience*, *13*, 1–13. https://doi.org/10.1093/gigascience/giae002

Badam, S. K., Chandrasegaran, S. C., & Elmqvist, N. (2022). Integrating annotations into multidimensional visual dashboards. *Information Visualization*, *21*(3), 270–284. https://doi.org/10.1177/14738716221079591

Baldonado, M. Q. W., Woodruff, A., & Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, 110–119. https://doi.org/10.1145/345513.345271

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594.

Bangor, A., Kortum, P. T., & Miller, J. T. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123.

Berba, P. (2020). A gentle introduction to hdbscan and density-based clustering [Accessed on 2025-05-03]. Retrieved May 3, 2025, from https://pberba.github.io/stats/2020/07/08/intro-hdbscan/

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brewer, I. (2005). *Understanding work with geospatial information in emergency management: A cognitive systems engineering approach in giscience* [Doctoral dissertation, Pennsylvania State University]. http://gradworks.umi.com/31/73/3173780.html

Brooke, J. (1996). SUS—a quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science*, *7819*, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

Campello, R. J. G. B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data*, *10*(1), 1–51. https://doi.org/10.1145/2733381

Cappabianca, F., Barbolini, M., & Natale, L. (2008). Snow avalanche risk assessment and mapping: A new method based on a combination of statistical analysis, avalanche dynamics simulation and empirically based vulnerability relations integrated in a GIS platform. *Cold Regions Science and Technology*, *54*(3), 193–205. https://doi.org/10.1016/j.coldregions.2008.06.005

Chatfield, C. (2004). *The analysis of time series: An introduction* (6th ed.). Chapman; Hall/CRC.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554. https://doi.org/10.2307/2288400

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Diehl, A., Abdul-Rahman, A., El-Assady, M., Bach, B., Keim, D. A., & Chen, M. (2018). Visguides: A forum for discussing visualization guidelines. *EuroVis 2018 - Short Papers*, 61–65. https://doi.org/10.2312/eurovisshort.20181079

EAWS - European Avalanche Warning Services. (2018). European avalanche danger scale [Accessed: 2025-03-13].

Elmqvist, N., & Fekete, J.-D. (2009). Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE transactions on visualization and computer graphics*, *16*(3), 439–454.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231.

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179–188.

Floyer, J., Klassen, K., Horton, S., & Haegeli, P. (2016). Looking to the 20's: Computer-assisted avalanche forecasting in canada. *Proceedings of the International Snow Science Workshop (ISSW 2016)*, 1245–1249.

Fred, A. L. N., & Jain, A. K. (2005). Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(6), 835–850.

Fromm, R., & Schönberger, V. (2022). Estimating the danger of snow avalanches with a machine learning approach using a comprehensive snow cover model. *Machine Learning with Applications*, *10*, 100405. https://doi.org/10.1016/j.mlwa.2022.100405

Galbrun, E., & Miettinen, P. (2012). A case of visual and interactive data analysis: Geospatial redescription mining. *ECML PKDD 2012 Workshop on Instant Interactive Data Mining*. http://adrem.ua.ac.be/iid2012/papers/galbrun_miettinen-visual_and_interactive_geospatial_redescription_mining.pdf

Gassner, M., & Brabec, B. (2002). Nearest neighbour models for local and regional avalanche forecasting. *Natural Hazards and Earth System Sciences*, *2*(3-4), 247–253.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org

Greene, E., Wiesinger, T., Birkeland, K., Coléou, C., Jones, A., & Statham, G. (1999). *Fatal avalanche accidents and forecasted danger levels: Patterns in the united states, canada, switzerland and france* (tech. rep.) (Technical Report). Proceedings of the International Snow Science Workshop.

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

Harbola, S., & Coors, V. (2018). Geo-visualisation and visual analytics for smart cities: A survey. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)*, *XLII-4/W11*, 11–18. https://doi.org/10.5194/isprs-archives-XLII-4-W11-11-2018

Hartl, L. (2025). World of science: Frank techel (slf) on people and models [Accessed on 2025-06-11]. Retrieved June 11, 2025, from https://powderguide.com/

en/magazine/bergwissen/world-of-science-frank-techel-slf-on-people-and-models

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Heer, J., & Robertson, G. G. (2007). Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1240–1247. https://doi.org/10.1109/TVCG.2007.70539

Herla, F., Widforss, A., Binder, M., Müller, K., Horton, S., Reisecker, M., & Mitterer, C. (2024). Establishing an operational weather & snowpack model chain in norway to support avalanche forecasting. *Proceedings of the International Snow Science Workshop (ISSW), Tromsø, Norway*, 168–175.

Horton, S., Nowak, S., & Haegeli, P. (2020). Enhancing the operational value of snowpack models with visualization design principles. *Natural Hazards and Earth System Sciences*, *20*(6), 1557–1572. https://doi.org/10.5194/nhess-20-1557-2020

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218. https://doi.org/10.1007/BF01908075

Iacobucci, G., et al. (2020). *Crossclustering: An r package for cluster validation*. Retrieved August 1, 2025, from https://rdrr.io/cran/CrossClustering/

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, *31*(8), 651–666.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, *31*(3), 264–323.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241–254.

Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094–1096). Springer.

Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval.

Kalamaras, I., Zamichos, A., Salamanis, A., Drosou, A., Kehagias, D. D., Margaritis, G., Papadopoulos, S., & Tzovaras, D. (2018). An interactive visual analytics platform for smart intelligent transportation systems management. *IEEE Transactions on Intelligent Transportation Systems*, *19*(2), 487–496. https://doi.org/10.1109/TITS.2017.2727143

Kehrer, J. (2013). Visual analysis of multi-faceted scientific data: Challenges and trends. *IEEE Transactions on Visualization and Computer Graphics*, (3), 495–513.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information visualization: Human-centered issues and perspectives* (pp. 154–175). Springer.

Kendrick, A. (2019). How to measure learnability of a user interface. Retrieved May 23, 2025, from https://www.nngroup.com/articles/measure-learnability/

Keogh, E., Chakrabarti, K., Mehrotra, S., & Pazzani, M. (2001). Locally adaptive dimensionality reduction for indexing large time series databases. *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, 151–162. https://doi.org/10.1145/375663.375680

Kitchin, R., Maalsen, S., & McArdle, G. (2016). The praxis and politics of building urban dashboards. *Geoforum*, *77*, 93–101. https://doi.org/10.1016/j.geoforum.2016.10.006

Kunz, M., Grèt-Regamey, A., & Hurni, L. (2010). Visualizing natural hazard data and uncertainties – customization through a web-based cartographic information system. *Proceedings of the Joint ISPRS Technical Commission IV Symposium and AutoCarto*, 7.

Kushniruk, A. W., & Patel, V. L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, *37*(1), 56–76. https://doi.org/10.1016/j.jbi.2004.01.003

Lachapelle, E. R. (1980). The fundamental processes in conventional avalanche forecasting. *Journal of Glaciology*, *26*(94), 75–84. https://doi.org/10.3189/s0022143000010601

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stöckli, U., & Zimmerli, M. (1999). Snowpack model calculations for avalanche warning based upon a new network of weather and snow stations. *Cold Regions Science and Technology*, *30*(1-3), 145–157. https://doi.org/10.1016/S0165-232X(99)00022-1

Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, *227*(7), 3515–3539. https://doi.org/10.1016/j.jcp.2007.02.014

Lewis, B., Smith, I., Fowler, M., & Licato, J. (2017). The robot mafia: A test environment for deceptive robots. *Proceedings of the 28th Modern Artificial Intelligence and Cognitive Science Conference (MAICS 2017)*, 189–190.

Little, R. J.-A., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Lohr, S. L. (1999). *Sampling: Design and analysis* (1st ed.). Duxbury Press.

Low, J. S., Ghafoori, Z., Bezdek, J. C., & Leckie, C. (2019). Seeding on samples for accelerating k-means clustering. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, 1593–1602. https://doi.org/10.1145/3361758.3361774

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, *30*, 4766–4775. https://arxiv.org/abs/1705.07874

MacEachren, A. M., & Kraak, M.-J. (2001). Research challenges in geovisualization. *Cartography and geographic information science*, *28*(1), 3–12.

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, *32*(3), 139–160.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281–297.

Mahajan, R., & Shneiderman, B. (1997). Visual and textual consistency checking tools for graphical user interfaces. *IEEE Transactions on Software Engineering*, *23*(11), 722–735. https://doi.org/10.1109/32.637386

Maissen, A., Techel, F., & Volpi, M. (2024). A three-stage model pipeline predicting regional avalanche danger in switzerland: A decision-support tool for operational avalanche forecasting [Preprint, under review]. *EGUsphere*, 1–34. https://doi.org/10.5194/egusphere-2024-84

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, *49*(4), 41–46. https://doi.org/10.1145/1121949.1121979

Marinova, S., Petkov, D., & Bandrova, T. (2018). Cartographic modeling and visualization of snow avalanche area. *Proceedings of the 7th International Conference on Cartography and GIS*.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Mayer, S., Techel, F., Schweizer, J., & van Herwijnen, A. (2023). Prediction of natural dry-snow avalanche activity using physics-based snowpack simulations. *Natural Hazards and Earth System Sciences*, *23*(11), 3445–3465. https://doi.org/10.5194/nhess-23-3445-2023

Mayer, S., van Herwijnen, A., Techel, F., & Schweizer, J. (2022). A random forest model to assess snow instability from simulated snow stratigraphy. *The Cryosphere*, *16*(11), 4593–4615. https://doi.org/10.5194/tc-16-4593-2022

McClung, D., & Schaerer, P. (2006). *The avalanche handbook* (3rd). The Mountaineers Books.

McClung, D. M., & Hungr, O. (1986). An equation for calculating snow avalanche run up against barriers. In *Avalanche formation, movement and effects* (pp. 605–612, Vol. 162). IAHS Publication.

Mohammed, H., Wei, Z., Wu, E., & Netravali, R. (2020). Continuous prefetch for interactive data applications. *Proceedings of the VLDB Endowment*, *13*(11), 2297–2311. https://doi.org/10.14778/3407790.3407826

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192–222. https://doi.org/10.1287/isre.2.3.192

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the u.s. public. *Weather and Forecasting*, *23*(5), 974–991. https://doi.org/10.1175/2008WAF2007088.1

Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*.

Nielsen, J. (2000). Why you only need to test with 5 users [Jakob Nielsen's Alertbox, March 19].

Norman, D. A. (2013). *The design of everyday things* (Revised and Expanded). Basic Books.

Pan, W., Shen, X., & Liu, B. (2013). Cluster analysis: Unsupervised learning via supervised learning with a non-convex penalty. *The Journal of Machine Learning Research*, *14*(1), 1865–1889.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Pedro, J. (2022). How to ensemble clustering algorithms [Medium, accessed 2025-04-30].

Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., Van Herwijnen, A., Olevski, T., Obozinski, G., Pérez-Cruz, F., & Schweizer, J. (2022). Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland. *Natural Hazards and Earth System Sciences*, *22*(6), 2031–2056. https://doi.org/10.5194/nhess-22-2031-2022

Pérez-Guillén, C., Techel, F., Volpi, M., & Herwijnen, A. V. (2024). *Assessing the performance and explainability of an avalanche danger forecast model* [Preprint, EGUsphere, under review]. https://doi.org/10.5194/egusphere-2024-2374

Pohl, M., Wiltner, S., Miksch, S., Aigner, W., & Rind, A. (2012). Analysing interactivity in information visualisation. *KI-Künstliche Intelligenz*, *26*(2), 151–159.

Prieto, D., Zeckzer, D., & Hernández, J. T. (2013). UCIV 4 planning: A User-Centered approach for the design of interactive visualizations to support urban and regional planning. *Proceedings of the IADIS International Conference Computer Graphics, Visualization, Computer Vision and Image Processing 2013, CGVCVIP 2013*, *8*(2), 83–90.

Purves, R. S., Morrison, K. W., Moss, G., & Wright, D. S. B. (2003). Nearest neighbours for avalanche forecasting in scotland: Development, verification and optimisation of a model. *Cold Regions Science and Technology*, *37*(3), 343–355. https://doi.org/10.1016/S0165-232X(03)00075-2

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850.

Rapisarda, A., & Pranzo, A. M. R. (2021). Mapping the avalanche risk: From survey to cartographic production. the avalanche bulletin of the meteomont service of the alpine troops command. *Proceedings of the International Cartographic Association*, *4*, 92. https://doi.org/10.5194/ica-proc-4-92-2021

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling*, *90*(2), 229–244. https://doi.org/10.1016/0304-3800(95)00152-2

Saunders, K. R., Forbes, O., Hopf, J. K., Patterson, C. R., Vollert, S. A., Brown, K., Browning, R., Canizares, M. A., Cottrell, R. S., Li, L., Kim, C. J., Stewart, T. P., Susilawati, C., Zhao, X. Y., & Helmstedt, K. J. (2023). Data-driven recommendations for enhancing real-time natural hazard warnings. *One Earth*, *8*(5), 101274. https://doi.org/10.1016/j.oneear.2025.101274

Sauro, J. (2011). Measuring usability with the system usability scale (sus). Retrieved July 17, 2025, from https://measuringu.com/sus/

Schweizer, J., Bartelt, P., & van Herwijnen, A. (2015). Snow avalanches. In J. F. Shroder, W. Haeberli, & C. Whiteman (Eds.), *Snow and ice-related hazards, risks, and disasters* (pp. 395–436). Elsevier. https://doi.org/10.1016/B978-0-12-394849-6.00012-3

Schweizer, J., Bellaire, S., Fierz, C., & Lehning, M. (2008). Evaluation of statistical and rule-based models for snow instability prediction. *Cold Regions Science and Technology*, *51*(2-3), 192–203. https://doi.org/10.1016/j.coldregions.2007.05.008

Schweizer, J., Jamieson, B., & Schneebeli, M. (2003). Snow avalanche formation. *Reviews of Geophysics*, *41*(4). https://doi.org/10.1029/2002RG000123

Schweizer, J., & Jamieson, J. B. (2001). Snow cover properties for skier triggering of avalanches. *Cold Regions Science and Technology*, *33*(2-3), 207–221. https://doi.org/10.1016/S0165-232X(01)00039-8

Schweizer, J., Mitterer, C., Reuter, B., & Techel, F. (2021). Avalanche danger level characteristics from field observations of snow instability. *The Cryosphere*, *15*(7), 3293–3315. https://doi.org/10.5194/tc-15-3293-2021

Schweizerische Eidgenossenschaft. (2023). *Geography – facts and figures*. Retrieved June 11, 2025, from https://www.aboutswitzerland.eda.admin.ch/en/geography

Scikit-learn. (n.d.). Importance of Feature Scaling. Retrieved July 17, 2025, from https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html?utm_source=chatgpt.com

Sculley, D. (2010). Web-scale k-means clustering. *Proceedings of the 19th international conference on World wide web*, 1177–1178.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization* (pp. 364–371). Elsevier.

Sigrist, C., & Schweizer, J. (2007). Critical energy release rates of weak snowpack layers determined in field experiments. *Geophysical Research Letters*, *34*(3). https://doi.org/10.1029/2006GL028576

Sovilla, B., Margreth, S., & Bartelt, P. (2007). On snow entrainment in avalanche dynamics calculations. *Cold Regions Science and Technology*, *47*(1-2), 69–79. https://doi.org/10.1016/j.coldregions.2006.08.012

StackExchange. (2023). *Is repeated hyperparameter tuning can lead to overfitting?* Retrieved April 19, 2025, from https://stats.stackexchange.com/questions/611690/is-repeated-hyperparameter-tuning-can-lead-to-overfitting

StackExchange. (2024). *Effectiveness of using moving averages in timeseries regression analysis with noisy data.* Retrieved April 21, 2025, from https://stats.stackexchange.com/questions/638358/effectiveness-of-using-moving-averages-in-timeseries-regression-analysis-with-no

Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., & Kelly, J. (2018). A conceptual model of avalanche hazard. *Natural Hazards*, *90*(2), 663–691. https://doi.org/10.1007/s11069-017-3070-5

Stehle, S., & Kitchin, R. (2020). Real-time and archival data visualisation techniques in city dashboards. *International Journal of Geographical Information Science*, *34*(2), 344–366. https://doi.org/10.1080/13658816.2019.1594823

Steinkogler, W., Sovilla, B., & Lehning, M. (2014). Influence of snow cover properties on avalanche dynamics. *Cold Regions Science and Technology*, *97*, 121–131. https://doi.org/10.1016/j.coldregions.2013.10.002

Stewart, G., & Al-Khassawneh, M. (2022). An implementation of the hdbscan* clustering algorithm. *Applied Sciences*, *12*(5), 2405. https://doi.org/10.3390/app12052405

Strehl, A., & Ghosh, J. (2002). Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, *3*, 583–617.

Stucki, T. (n.d.). How is the avalanche bulletin produced? Retrieved June 23, 2025, from https://www.slf.ch/de/lawinenbulletin-und-schneesituation/wissen-zum-lawinenbulletin/wie-entsteht-das-lawinenbulletin/

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tchagna, A. (2025). The drawbacks of k-means [Accessed: 2025-03-04].

Techel, F., Mayer, S., Pérez-Guillén, C., Schmudlach, G., & Winkler, K. (2022). On the correlation between a sub-level qualifier refining the danger level with observations and models relating to the contributing factors of avalanche danger. *Natural Hazards and Earth System Sciences*, *22*(6), 1911–1930. https://doi.org/10.5194/nhess-22-1911-2022

Techel, F., Mayer, S., Purves, R. S., Schmudlach, G., & Winkler, K. (2024). Forecasting avalanche danger : human-made forecasts vs . fully automated model-driven predictions. *European Geoscience Union*, *20*(September), 1–31.

Techel, F., Purves, R. S., Mayer, S., Schmudlach, G., & Winkler, K. (2025). Can model-based avalanche forecasts match the discriminatory skill of human danger level forecasts? a comparison from switzerland [Accepted manuscript]. *Natural Hazards and Earth System Sciences*, 1–32.

Tory, M., & Möller, T. (2004). Human factors in visualization research. *IEEE transactions on visualization and computer graphics*, *10*(1), 72–84.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2). Graphics press Cheshire, CT.

Van Herwijnen, A., Mayer, S., Pérez-Guillén, C., Techel, F., Hendrick, M., & Schweizer, J. (2023). Data-driven models used in operational avalanche forecasting in switzerland. *Proceedings of the International Snow Science Workshop (ISSW)*. https://www.researchgate.net/publication/374530389_Data-driven_models_used_in_operational_avalanche_forecasting_in_Switzerland

Van Wijk, J. J. (2005). The value of visualization. *VIS 05. IEEE Visualization, 2005.*, 79–86.

von Luxburg, U., Williamson, R. C., & Guyon, I. (2012). Clustering: Science or art? *Journal of Machine Learning Research: Workshop and Conference Proceedings*, *27*, 65–79.

Wang, S. M., Yang, W. R., Zhuang, Q. Y., Lin, W. H., Tian, M. Y., Su, T. J., & Cheng, J. C. (2025). Application of three-dimensional hierarchical density-based spatial clustering of applications with noise in ship automatic identification system trajectory-cluster analysis. *Applied Sciences*, *15*(5), 2621. https://doi.org/10.3390/app15052621

Ware, C. (2013). *Information visualization: Perception for design* (3rd). Morgan Kaufmann.

Wells, J., Grant, R., Chang, J., & Kayyali, R. (2021). Evaluating the usability and acceptability of a geographical information system (gis) prototype to visualise socio-economic and public health data. *BMC Public Health*, *21*(1), 2151.

Wever, N., Valero, C. V., & Fierz, C. (2016). Assessing wet snow avalanche activity using detailed physics-based snowpack simulations. *Geophysical Research Letters*, *43*(11), 5732–5740. https://doi.org/10.1002/2016GL068428

White Risk. (n.d.). Avalanche terrain maps. Retrieved May 23, 2025, from https://content.whiterisk.ch/en/help/maps/avalanche-terrain-maps

Wickham, H. (2014). Tidy data. *Journal of statistical software*, *59*, 1–23.

Wood, J., Dykes, J., Slingsby, A., & Clarke, K. (2007). Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE transactions on visualization and computer graphics*, *13*(6), 1176–1183.

WSL – Institut für Schnee- und Lawinenforschung SLF. (n.d.-a). Avalanche bulletin and snow situation. Retrieved June 21, 2025, from https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/

WSL – Institut für Schnee- und Lawinenforschung SLF. (n.d.-b). Beschreibung automatischer stationen [Accessed: 2025-05-15].

Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1224–1231. https://doi.org/10.1109/TVCG.2007.70515

Young, G. W., & Kitchin, R. (2020). Creating design guidelines for building city dashboards from a user's perspective. *International Journal of Human-Computer Studies*, *140*, 102429. https://doi.org/10.1016/j.ijhcs.2020.102429

Zhang, S., Wong, H. S., & Shen, Y. (2012). Generalized adjusted rand indices for cluster ensembles. *Pattern Recognition*, *45*(6), 2214–2226. https://doi.org/10.1016/j.patcog.2011.11.017

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). Birch: An efficient data clustering method for very large databases. *ACM sigmod record*, *25*(2), 103–114.

Appendix

I. Dashboard Code

To see the scripts I used for this thesis, please follow the link to my GitHub repository:

https://github.com/streuli7/Master-Thesis

II. Google Forms Survey - Dashboard Evaluation

Usability Test - Avalanche Dashboard

Welcome to the usability study for the
Avalanche Dashboard developed as part of a Master's thesis at the University of Zurich. This dashboard is designed to support avalanche forecasters in analyzing and visualizing model-based avalanche predictions and observations. The aim of this test is to assess the usability of the dashboard—specifically, how intuitive and efficient it is for users to navigate, understand, and interact with its features. Your participation will help identify strengths and areas for improvement, and your feedback will directly contribute to refining the tool for operational and research use. The test consists of three parts:

1. A short explanation and exploration phase.

2. A set of practical feature-based tasks.

3. The System Usability Scale (SUS) questionnaire.

Please take your time with each section. Your honest responses are highly valuable.

When you are ready click on this link to get to the Dashboard.

* Gibt eine erforderliche Frage an

Feature Questions

This section is about how you use the dashboard in practice.
You will be asked to perform specific tasks using different parts of the dashboard. These tasks are designed to help us evaluate how intuitive and functional the system is for real-world use. After each task, you'll rate how easy or difficult it was.

1. Navigate to the Cluster Analysis tab and select the most recent date available. What is the danger level for the cluster with the highest substep label? *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

2. Change the date to the 11th of March. Click on a cluster. In the box on the left side of map scroll down and describe the distribution of the 'Snow depth in the last 72 hours" (hn72) and the median value based on the histogram shown. Remember to put the label of the cluster in your answer. *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

3. Select anohter cluster. On top of the histograms there are four radar plots. Describe the elevation–aspect distribution of the highest available danger level. Remember to put the label of the cluster in your answer. *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

4. In the Point-Data tab, select the 10th of February 2024 and visualize 'pMax_decisive'. What spatial pattern do you see? *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

5. Change the attribute to 'Measured snow depth at station (hs)'. Are there stations with over 300 cm? If yes, how many are there and where are they roughly located?

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

6. In the Grid-Data tab, select the 'SLF Danger Level Sub-steps' dataset for the *

20th of February. Where are the highest values located?

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

7. Change the attribute to 'Instablity Forecast' and the date to the 14th of March. Use the slider to only include areas between 1450–2350 m elevation. How does the visualization change? Can you see a pattern? *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

8. You can now reset the elevation to range to see all the data. For the same date and attribute use the aspect filter to only show South-facing slopes. What do you notice compared to all aspects? (Hint: you maybe need to toggle the selection to see some differences!). *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

9. After exploring and leveraging each of the available data types—point measurements, interpolated grid fields, and the clustered data - which single dataset do you find most effective for producing a precise, location-specific avalanche forecast? *

Meine Antwort

How easy was this question to answer? *

	1	2	3	4	5	
Very Easy	○	○	○	○	○	Very Hard

System Usability Scale Questionnaire

I'd love to hear about your experience using the dashboard.

In this section, please reflect on how it felt to interact with the tool. Your impressions—whether positive or critical—will help us better understand how intuitive, helpful, or frustrating the dashboard may be. Please answer honestly; there are no right or wrong responses.

10. I think that I would like to use this system frequently.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

11. I found the system unnecessarily complex.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

12. I thought the system was easy to use.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

13. I think that I would need the support of a technical person to be able to use this system.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

14. I found the various functions in this system were well integrated.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

15. I thought there was too much inconsistency in this system.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

16. I would imagine that most people would learn to use this system very quickly.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

17. I found the system very inconvenient to use.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

18. I felt very confident using the system.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

19. I needed to learn a lot of things before I could get going with this system.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Getting to know you a bit more

Just a few more questions about you before we finish!

This helps us understand who our users are and how experience might influence dashboard interaction.

How far along are you in your studies? *

- ◯ I am in my Bachelors
- ◯ I am in my Masters

How extensive is your GIS experience?

	1	2	3	4	5	
No experience	◯	◯	◯	◯	◯	Expert (can teach or automate GIS)

Do you have anything you'd like to add?
Whether it's a suggestion, a concern, or simply something that stood out to you—feel free to share anything that's on your mind.

Meine Antwort

Senden Alle Eingaben löschen

Personal Declatation

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

AI Contribution Declaration: The language and style of this report were enhanced using AI language models such as GPT-4 and DeepL. However, all ideas, findings, and conclusions presented are my own.

Winterthur, 26. August 2025 Nils Besson