Modelling Avalanche Danger and Understanding Snow Depth Variability

Von der Fakultät für Georessourcen und Materialtechnik der Rheinisch-Westfälischen Technischen Hochschule Aachen

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigte Dissertation

vorgelegt von

Michael Schirmer

aus Freiburg

Berichter: Prof. Dr. Christoph Schneider
Dr. Michael Lehning

Tag der mündlichen Prüfung: 12.07.2010

Contents

Summary

This thesis addresses the causes of avalanche danger at a regional scale. Modelled snow stratigraphy variables were linked to [1] forecasted avalanche danger and [2] observed snowpack stability. Spatial variability of snowpack parameters in a region is an additional important factor that influences the avalanche danger. Snow depth and its change during individual snow fall periods are snowpack parameters which can be measured at a high spatial resolution. Hence, the spatial distribution of snow depth and snow depth change due to individual snow storms were observed [3]. Furthermore, this spatial dataset was characterised with a fractal analysis and results were related to deposition processes [4]. In the following, each subject is described in more detail:

[1] In the past, numerical prediction of regional avalanche danger using statistical methods with meteorological input variables has shown insufficiently accurate results, possibly due to the lack of snow stratigraphy data. Detailed snow-cover data were rarely used because they were not readily available (manual observations). With the development and increasing use of snow-cover models this deficiency can now be rectified and model output can be used as input for forecasting models. We used the output of the physically based snow cover model SNOWPACK combined with meteorological variables to investigate and establish a link to regional avalanche danger. Snow stratigraphy was simulated for the location of an automatic weather station near Davos (Switzerland) over nine winters. Only dry-snow situations were considered. A variety of selection algorithms was used to identify the most important simulated snow variables. Data mining and statistical methods, including classification trees, artificial neural networks, support vector machines, hidden Markov Models and nearest-neighbour methods were trained on the forecasted regional avalanche danger (European avalanche danger scale). The best results were achieved with a nearest neighbour method which used the avalanche danger level of the previous day as additional input. A cross-validated accuracy (hit rate) of 73% was obtained. This study suggests that modelled snow-stratigraphy variables, as provided by SNOWPACK, are able to improve numerical avalanche forecasting.

[2] Snow stability, or the probability of avalanche release, is one of the key factors defining avalanche danger. Most snow stability evaluations are based on field observations, which are time-consuming and sometimes dangerous. Through numerical modelling of the snow cover stratigraphy, the problem of having sparsely measured regional stability information can be overcome. In this study we compared numerical model output with observed stability. Overall, 775 snow profiles combined with Rutschblock scores and release types for the area surrounding five weather stations were rated into three stability classes. Snow stratigraphy data were then produced for the locations of these five weather stations using the snow cover model SNOWPACK. We observed that (i) an existing physically based stability interpretation implemented in SNOWPACK was applicable for regional stability evaluation; (ii) modelled variables equivalent to those manually observed variables found

to be significantly discriminatory with regard to stability, did not demonstrated equal strength of classification; (iii) additional modelled variables that cannot be measured in the field discriminated well between stability categories. Finally, with objective feature selection, a set of variables was chosen to establish an optimal link between the modelled snow stratigraphy data and the stability rating through the use of classification trees. Cross-validation was then used to assess the quality of the classification trees. A true skill statistic of 0.5 and 0.4 was achieved by two models that detected "rather stable" or "rather unstable" conditions, respectively. The interpretation derived could be further developed into a support tool for avalanche warning services for the prediction of regional avalanche danger.

[3] Terrestrial and Airborne Laser Scanning (TLS and ALS) techniques have only recently developed to the point where they allow wide-area measurements of snow distribution in varying terrain. Multiple TLS measurements are presented showing the snow depth development for a series of precipitation events. We observe that the pattern of maximum accumulation is similar for the two years presented here (correlation up to $r = 0.97$). Storms arriving from the Northwest show persistent snow depth distributions and contribute most to the final accumulation pattern. Snow depth patterns of maximum accumulation for the two years is more similar than the distribution created by any two pairs of individual storms. A decrease in variance of snow depth change with time was observed, while variance of snow depth was increasing. Based on the strong link between accumulation patterns and terrain, we investigated the ability of a model based on terrain and wind direction to predict accumulation patterns. This approach of Winstral, which describes wind exposure and shelter, was able to predict the general accumulation pattern over scales of slopes but failed to match observed variance. Furthermore, a high sensitivity to the local wind direction was demonstrated. We suggest that Winstral's model could form a useful tool for application from hydrology and avalanche risk assessment to glaciology.

[4] We present analysis of high resolution laser scanning data of snow depths in the Wannengrat catchment (introduced in [3]) using omni-directional and directional variograms for three specific terrain features; cross-loaded slopes, lee slopes and windward slopes. A break in scaling behavior was observed in all sub-areas, which can be seen as the roughness scale of summer terrain which is modified by the snow cover. In the wind-protected lee slope a different scaling behavior was observed, compared to the two wind-exposed areas. The wind-exposed areas have a smaller ordinal intercept γ, a smaller short range fractal dimension D and a larger scale break distance L than the wind-protected lee slope. Snow depth structure inherits characteristics of dominant NW storms, which results e.g. in a trend towards larger break distances in the course of the accumulation season. This can be interpreted as a result of surface smoothing at increasing scales. Similar scaling characteristics were obtained for two different years at the end of the accumulation season. Since snow depth structure is altered strongly by NW storms, this inter-annual consistency may strongly depend on their frequency in an accumulation period. The analysis of directional variograms suggests that existing anisotropies can be explained by the orientation of terrain features with respect to the predominant wind direction.

Zusammenfassung

Diese Doktorarbeit befasst sich mit dem Thema “Regionale Lawinengefahr”. In den ersten beiden Kapiteln wurde untersucht, ob modellierte Schneedeckeneigenschaften [1] vorhergesagte Lawinengefahr oder [2] beobachtete Schneedeckenstabilität erklären können. Die räumliche Variabilität von Schneedeckeneigenschaften in einer Region hat Einfluss auf die Lawinengefahr. Die Schneehöhe ist zurzeit die einzige Schneedeckeneigenschaft, die mit einer hohen räumlichen Auflösung gemessen werden kann. In dieser Studie wurde sich daher auf diesen Parameter beschränkt, indem die Schneehöhenverteilung und deren Veränderung nach einzelnen Stürmen untersucht wurden [3]. Die räumliche Struktur der Schneehöhe wurde genauer mit dem Konzept der fraktalen Geometrie analysiert. Die Ergebnisse ergaben Hinweise auf unterschiedliche Ablagerungs- und Verfrachtungsprozesse [4]. Im Folgenden werden die vier Punkte [1]-[4] genauer beschrieben:

[1] Bisher konnten statistische Modelle nur unzureichende Ergebnisse für die Vorhersage der Lawinengefahr erzielen. Ein Grund hierfür war, dass diese Modelle meist auf meteorologischen Eingangsparametern basierten und die Informationen aus der Schneedecke fehlten. Diese wichtigen Informationen erhält man durch Beobachtungen im Feld und sind daher nicht immer verfügbar. Durch die Entwicklung von Schneedeckenmodellen konnte dem Problem von fehlenden Schneedeckeninformationen begegnet werden. Für diese Arbeit wurden die Ergebnisse des Schneedeckemodells SNOWPACK sowie meteorologische Messdaten zur Vorhersage der regionalen Lawinengefahr genutzt. Die Schneedecke wurde mit den Messungen einer Wetterstation in der Region Davos (Schweiz) simuliert. Der erhaltene Datensatz umfasst einen Zeitraum von neun Wintern. Wegen der unterschiedlichen Ursachen, die zu Nassschneelawinen führen, wurde der Zeitraum auf Situationen mit trockenem Schnee eingeschränkt. Eine Reihe von Algorithmen kam zum Einsatz, welche aus einer großen Anzahl modellierter Schneeeigenschaften die Wichtigsten bestimmen sollten. Diese Auswahl wurde als Input für statistische Modelle wie Klassifikationsbäume, künstliche neuronale Netze, Support Vector Maschinen, Hidden Markov Modelle und Nächste-Nachbarn-Klassifikation benutzt. Trainiert wurden diese Modelle mit der vorhergesagten Lawinengefahrenstufe der entsprechenden Region. Die besten Ergebnisse wurden mit der Nächsten-Nachbarn-Klassifikation erzielt, die die Lawinengefahr des Vortages als zusätzliche Information erhielt. Eine kreuz-validierte Genauigkeit (hit rate) von 73% wurde erreicht. Die erzielten Ergebnisse legen nahe, dass modellierte Schneedeckeneigenschaften die numerische Lawinenvorhersage verbessern.

[2] Die Schneedeckenstabilität ist definiert als die Wahrscheinlichkeit einer Lawinenauslösung und ist eine der wichtigsten Größen zur Bestimmung der Lawinengefahr. Die meisten Einschätzungen der Schneedeckenstabilität basieren auf zeitaufwendigen und manchmal gefährlichen Feldmessungen. Aktuelle Informationen stehen daher nur selten zur Verfügung. Um diese Lücke zu schließen, wurden Schneedeckenmodelle entwickelt. In der vorliegenden Studie wurden modellierte Schneeeigenschaften

genutzt, um eine Einschätzungen über die Stabilität zu erhalten. Hierfür wurden 775 Schneeprofile und Rutschblock-Tests in der Nähe von fünf Wetterstationen genutzt. Diese Beobachtungen wurden drei Stabilitätsklassen zugeteilt. SNOWPACK lieferte mit den Daten dieser Wetterstationen die zugehörigen modellierten Schneeeigenschaften. Dabei konnte erstens festgestellt werden, dass für eine regionale Stabilitätseinschätzung eine bereits in SNOWPACK implementierte Interpretation geeignet war; zweitens konnte gezeigt werden, dass wichtige Variablen zur Einschätzung der Stabilität im Feld keine entsprechende Erklärung im Modell bieten; jedoch konnten drittens andere modellierte Variablen gefunden werden, die gut zwischen den Stabilitätskategorien unterscheiden können. Darunter befinden sich einige, die im Feld nicht messbar sind. Zuletzt konnten mit einem objektiven Auswahlverfahren die wichtigsten Variablen bestimmt werden. Diese wurden als Eingangsparameter für Klassifikationsbäume benutzt, um einen optimalen Bezug zwischen modellierter Schneedecke und beobachteter Schneedeckenstabilität zu erzielen. Zwei Modelle, eines zur Bestimmung von "eher stabilen" und eines zur Bestimmung von "eher instabilen" Verhältnissen, erreichten Werte bei der so genannten True Skill Statistik von 0.5 bzw. 0.4. Diese Modelle können in weiterentwickelter Form als Hilfsmittel für die Lawinenwarnung verwendet werden.

[3] Terrestrisches und Airborne Laser Scanning (TLS and ALS) erlauben eine flächendeckende Vermessung der Schneehöhenverteilung. In dieser Studie präsentieren wir mit einer Reihe von TLS Messungen, wie sich die Schneehöhenverteilung im Verlauf eines Winters entwickelt. Ebenso wurde die Veränderung der Schneehöhe durch einzelne Schneefallereignisse untersucht. Es zeigte sich, dass am Ende der Akkumulationsphase zweier Winter eine große Ähnlichkeit bestand (Korrelation bis zu $r = 0.97$). Auch einzelne Stürme mit vorherrschender Windrichtung Nordwest wiesen über den ganzen Winter starke Ähnlichkeiten auf. Diese Stürme prägten die Struktur der Schneehöhe am Ende der Akkumulationsphase maßgebend. Allerdings ist die Ähnlichkeit am Ende von zwei verschiedenen Akkumulationsphasen deutlich ausgeprägter, als die Ähnlichkeit zweier einzelner Stürme. Im zeitlichen Verlauf ging die Varianz der Schneehöhenveränderungen zurück, während die Varianz der Schneehöhe zunahm. Da eine starke Abhängigkeit zwischen Ablagerungsmustern und Terrain beobachtet wurde, zeigt diese Studie die Möglichkeiten und Grenzen eines Modells zur Beschreibung der Schneehöhenverteilung basierend auf einer einfachen Geländeinterpretation auf. Diese Interpretation, die in dieser Form von Winstral entwickelt wurde, beschreibt ein Maß wie stark ein Geländepunkt in Bezug auf eine vorherrschende Windrichtung ausgesetzt oder geschützt ist. Mit diesem Modellansatz konnten Muster der Schneeverteilung gut reproduziert werden, nicht so allerdings die gemessene Varianz. Außerdem wurde gezeigt, wie sensibel dieses Modell auf die Wahl der vorherrschenden Windrichtung reagiert. Die präsentierten Ergebnisse zeigen, dass dieser Modellansatz für Anwendungen in der Hydrologie, Glaziologie und für die Einschätzung der Lawinengefahr nützlich ist.

[4] Die in [3] vorgestellten Schneehöhen-Messungen mit TLS wurden in drei Gebieten – einem Luv- und einem Leehang, sowie einem Hang, der in Bezug zur Hauptwindrichtung seitlich umströmt wird – mit omni-direktionalen und direktionalen Variogrammen analysiert. In allen Gebieten wurde ein Skalenbruch festgestellt. Dieser kann als die Rauhigkeitslänge interpretiert werden, bis zu der das Sommergelände von der Schneebedeckung stark verändert wird. Der windgeschützte Leehang unterschied sich von den beiden windexponierten Gebieten durch einen Skalenbruch bei einer kürzeren

Distanz, einem größeren y-Achsenabschnitt γ und einer größeren Fraktalen Dimension D vor dem Skalenbruch. Im zeitlichen Verlauf konnte während der Akkumulationsphase festgestellt werden, dass die Struktur der Schneedecke vermehrt die Eigenschaften von dominanten NW Stürmen annahm, was unter anderem eine Zunahme an Distanz, bei welcher der Skalenbruch festgestellt wurde, bedeutet. Dies spricht für eine Glättung des Sommergeländes durch Schnee auf zunehmenden Rauhigkeitslängen. Die Struktur der Schneehöhe zeigte sehr ähnliche Eigenschaften am Ende der Akkumulationsphase von zwei unterschiedlichen Wintern. Die Beobachtung, dass diese Struktur stark von dominanten NW Stürmen verändert werden kann, spricht allerdings dafür, dass diese Ähnlichkeit zwischen zwei Wintern stark von der Häufigkeit solcher NW Stürme abhängig ist. Die Analyse mit direktionalen Variogrammen zeigt, dass beobachtete Anisotropien der Schneehöhenstruktur mit der Anordnung von Geländeformen zur dominanten Windrichtung erklärt werden können.

1 Introduction

1.1 Motivation

Avalanche forecasting is regularly based on data delivered by weather stations, manual snow cover observations and numerical weather forecasts (Meister, 1995). Many weather stations are available especially in the Swiss Alps, which transmit data continuously. However, the more important snowpack information is rare, both in time and in space. Numerical models describing snow cover stratigraphy or stability offer a solution to this problem.
The original motivation to start this study was based on the observation that avalanche danger is regularly forecasted by experts mainly without the help of numerical models of snow cover development. Models describing the relevant processes have been developed over the last decades. Although considerable attention has been paid to verify those models in regard to its associated sub-process (i.e. is snow-layering adequately represented?), less research has been devoted to show if modelled results can be used for avalanche danger prediction. This research gap should have been addressed with the present study. In order to assess this question, objectives were formulated at the beginning of this study:

- Measure spatial variability of snowpack properties at a regional scale in order to determine possible causes of spatial variability with the help of high resolution distributed surface process modelling
- Verify modelled spatial variability with field measurements
- Relate modelled spatial variability to regional avalanche danger.

In the next section, it will be shown how these objectives are related to avalanche danger prediction.

1.2 Background

For avalanche danger prediction a definite set of rules does not exist. This is in part because the process chain beginning with the weather forcing and ending with an avalanche release is highly complex and not fully understood. In the following, a short introduction will be given, reaching from formation of single avalanches to time and space integrated regional avalanche danger.

1.2.1 Avalanche formation

In a review paper, Schweizer et al. (2003a) defined avalanche formation as the "complex interaction between terrain, snowpack, and meteorological conditions leading to avalanche release". A

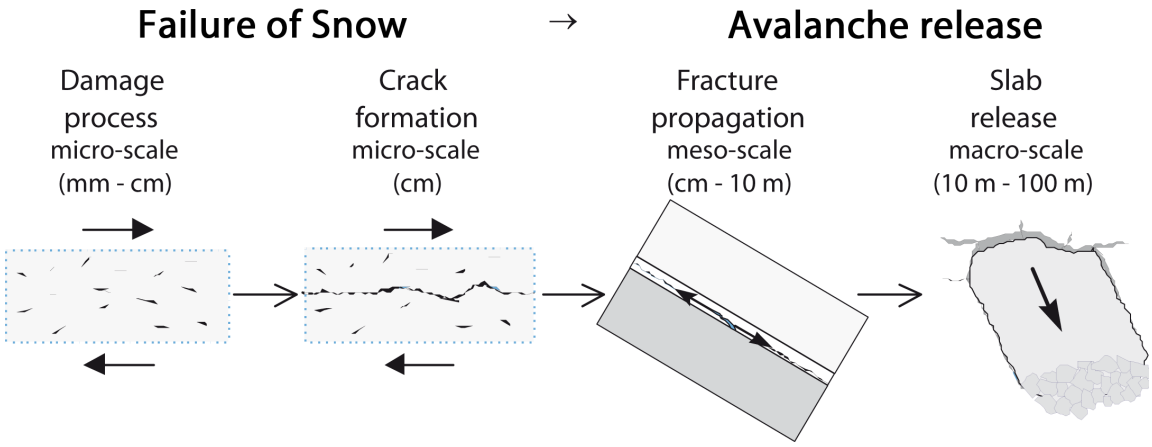

Figure 1.1: Conceptual model of snow slab avalanche release (Schweizer et al., 2003a).

prerequisite for a slab avalanche is a continuous weak layer or weak interface below an overlying cohesive slab. In a conceptual model Schweizer et al. (2003a) described the sequence of processes from failure at a micro scale, over fracture propagation and slab release (Fig. 1.1). In Figure 1.2 a large avalanche is shown demonstrating the extent of fracture propagation over hundreds of meters. Contributory factors for avalanche formation are known and the physical meaning is understood, at least on a conceptual level. Contributory factors are terrain (inclination), new snow precipitation, wind, temperature, radiation, and snowpack stratigraphy (Schweizer et al., 2003a). Spatial variability of those contributory factors at a slope-scale will influence avalanche release probability.

1.2.2 Spatial variability at a regional scale

For regional avalanche danger prediction it is common to focus on areas that are particularly dangerous (Meister, 1995). Those areas are classified using critical altitudes, aspects and sections of terrain. Hence, knowledge on the spatial variability at the regional scale is important. Causes of spatial variations are known in a qualitative sense resulting mainly from wind and radiation influences. In the last decades the focus of research has been on quantifying spatial variations of snowpack properties including stability. For a comprehensive review of those studies, see Schweizer et al. (2008a).

1.2.3 Avalanche danger

In Europe and North America, a consistent international danger level definition exists ranging from "Low" to "Very High" in five levels. It is based on snowpack stability (i.e. the probability of avalanche release), the frequency of trigger points (spatial distribution of instability) and the size and type of anticipated avalanches (Meister, 1995). Since much of the avalanche release processes remain unknown, the "where" and "when" of a single avalanche event cannot accurately be predicted

Figure 1.2: Large avalanche observed during field work on the 25.02.2009.

(Schweizer et al., 2003a). Thus, avalanche forecasting is restricted to a danger description at a larger scale (in Switzerland: >10 km length).

1.3 Problematic issues

1.3.1 Stability interpretations

The assessment of a regional stability distribution is problematic for several reasons. First, snowpack stability cannot objectively be measured. Stability tests combined with snow profiles are the most reliable approximation to a stability measurement. However, different stability tests exist, which produce different results (Winkler and Schweizer, 2009). Moreover, different stability interpretations based on the same observations also produce different results (Schirmer et al., 2009b,c). Hence, considerable uncertainty is associated with snow stability interpretation.
Second, spatial variability exists at smaller scales, e.g. at the slope-scale. A stability test with a measurement support of 3 m^2 has only a limited predictive value when assign slope stability. Several studies have attempted to quantify stability variations at the slope scale (e.g. Jamieson and Johnston, 1993; Campbell and Jamieson, 2007). These studies show that the influence of the slope scale variability cannot be neglected for interpretations of stability test results at a regional scale. Hence, many observations are needed to assess regional variations of stability.
Schirmer et al. (2009b) quantified the uncertainty of a regional stability interpretation with 25 observations. In most situations, this number of observations is not enough to draw reliable conclusions. Since stability observations are time-consuming and sometimes dangerous, a large number of well educated field workers are needed to reach a number of observations significantly larger than 25. Such resources are not practical and were not available for the present study.

1.3.2 Modelling

For numerical avalanche prediction a reasonable first step is to test the informative value of a one-dimensional model which describes the temporal development of a representative snow cover only. In a next step, this analysis should be extended to a spatially distributed model approach. A recent study (Lehning et al., 2008) used such a model setup considering terrain induced wind flow and its interaction with snow cover properties to describe the spatial varying snow depth development of a single storm. They showed that the model was able to capture the ridge-scale snow distribution patterns. The model suggests that these large scale patterns are mainly formed by preferential deposition, which is the spatially varying precipitation due to a terrain induced wind field. For the study discussed here, it was planned to use a similar model setup and extend the investigation to a more complex area. In an area with varying aspects, it was supposed to study the quantitative effect of wind on snowpack stability. For avalanche danger prediction the aim was to analyse modelled patterns not only for one single storm, but for a sequence of storms or an entire snow season. However, both extensions, i.e. a more complex area and longer time scales, turned out to be to be problematic in the course of this study. Furthermore, validation data for the distributed model results were missing. Also for snow depth, which is a simple measurement in comparison to snow stability observations, the spatial characteristics were not known at the beginning of this study. Thus it was not known which spatial characteristics a model needs to fulfill.

A refinement of the objectives was therefore necessary. On the one hand, objectives should focus on reliable measurements describing spatial characteristics of distributed snow data, which can be used in future for model validation. On the other hand, modelling with regard to avalanche prediction should concentrate firstly on a one-dimensional approach, which includes, however, important shortwave radiation variations due to aspect and inclination. Both objectives are described, motivated and formulated in more detail in the following section.

1.4 Objectives

Since high resolution snow stability data is difficult to obtain, if not impossible, field measurements were restricted to snow depth and snow depth change due to individual storms. Wind and the resulting snow distribution is one of the most relevant factors of avalanche formation (Schweizer et al., 2003a). Hence quantifying snow depth variability in a mountain catchment and observing how individual storms contribute to the variance of the snow depth distribution at the end of an accumulation season is of primary importance. However, a reliable high resolution measurement system of snow depth, which can be applied directly before and after snowfall events, needs to be found. Hence we tested the quality of a terrestrial laser scanner under harsh mountain conditions and published results in (Prokop et al., 2008). It was shown that snow depth can be measured with a sufficient accuracy at a high spatial resolution (1 m). Thus far, there are no studies describing snow depth development throughout a whole accumulation season with such a high spatial resolution.

In order to directly verify simple model approaches in regard to avalanche danger prediction, a one dimensional snow cover model was used to explain the time-development of regional avalanche danger. Spatial variability was considered in this model approach with two weather stations and with varying radiation input to slope simulations of four different aspects. Modelled snow stratigraphy was linked to predicted avalanche danger levels. Since this approach includes mistakes of the human decision process, stability observations were used additionally. In both case, the aim was not only to establish a link between modelled snow stratigraphy and avalanche danger in the sense of a black box. The aim was also to determine which of the huge amount of modelled properties show strength in explanation. Hence, a guideline was searched how a snow cover model can be integrated in operational avalanche danger prediction.
These solutions resulted in the final objectives which are presented in the following chapters of the present study.

1. Determine most important modelled snow stratigraphy variables relating to predicted regional avalanche danger. Provide a link between modelled variables and avalanche danger and quantify its predictive potential. Detect if the consideration of limited modelled spatial variability (i.e. two stations, slope simulations of four aspects) adds new strength of explanation to the system.

2. Due to reliability deficits of predicted avalanche danger levels, perform a similar analysis as in 1, but with observed stability as target variable.

3. Use a terrestrial laser scanner to determine the spatial variability of snow depth and snow depth change due to individual storms in an Alpine catchment throughout an accumulation season.

4. Characterise the spatial structure of snow depth and snow depth change with a fractal analysis. Relate the scaling behaviour found with the fractal analysis to different storm types and to spatially varying deposition and erosion processes.

2 Statistical forecasting of regional avalanche danger using simulated snow-cover data

2.1 Introduction

Regional avalanche forecasting attempts to predict current and future snow stability, relative to a given triggering level on the scale of a mountain range or a considerable fraction thereof (e.g. McClung and Schaerer, 2006). Forecasts are issued on a daily basis to warn the public about the level of avalanche danger. These public bulletins play a key role in the prevention of avalanche fatalities. Adequate avalanche warnings, combined with avalanche education and efficient rescue have probably prevented an increase of avalanche fatalities in parallel with the increased recreational use of avalanche terrain at least in some countries (Harvey and Zweifel, 2008). Reliable and consistent avalanche forecasts are, therefore, very much needed. To assess the avalanche danger level, most avalanche warning services rely on a combination of manual observations, automatic weather stations, weather forecasts (including model output) and snow profiles (Meister, 1995). For the locations of the automatic weather stations in the Swiss Alps, the amounts of new snow and drifting snow are additionally derived from the numerical snow cover model SNOWPACK (Lehning et al., 1999; Lehning and Fierz, 2008). Based on all these data, the forecaster uses experience, intuition and local knowledge of the mountain range to estimate and describe the avalanche danger in the public bulletin.

Over the past decades there have been many attempts to create an objective process of danger evaluation, which may also work as a support tool for the avalanche warning service. The French model chain SAFRAN/Crocus/MÉPRA (SCM) provides an automated avalanche danger prediction for virtual slopes (Durand et al., 1999) and is the only real operational derivation of a risk level on the basis of physical snow modelling. Durand et al. (1999) published a contingency table between modelled risk and avalanche activity with a hit rate of 75%. Since Murphy (1991) described the difficulties of comparison between two forecasting systems under different conditions (e.g. different datasets), the published accuracy measures of the studies described below are not further reviewed here. Several studies have used observations of avalanche activity as an indicator of the avalanche danger (e.g. Buser, 1983; Heierli et al., 2004; Pozdnoukhov et al., 2008). As Schweizer et al. (2003b) pointed out, the problem with this target parameter is that it does not distinguish between lower danger levels and that observations may be inconsistent, mainly due to limited visibility during times of avalanche activity. In comparison, Schweizer and Föhn (1996) forecasted the avalanche danger level instead of the avalanche activity using commercial decision-making software. Input variables included snow cover information observed manually. They trained their models using a

verified danger level instead of the forecasted one. The verification was based on additional data and observations, including data not yet available at the time of the forecast. The cross-validated hit rate was 63%, but adding knowledge in the form of expert rules to the system, the performance improved to a hit rate of ∼70%. However, their models did not run fully automatically but required manual input of the snow cover information. Therefore, Brabec and Meister (2001) used a nearest neighbour method with only meteorological variables and snow information, such as penetration depth or surface characteristics, so that the model could be run for the whole area of the Swiss Alps. The accuracy was only ∼52%. The absence of snow cover information was given as a reason for the poor results.
Model output from snow cover models such as SNOWPACK can provide the snow cover information with the required resolution in space and time. This study explores whether the performance of data-based forecasting models can be improved with modelled snow cover data as additional inputs.

2.2 Methods

2.2.1 Data

In order to establish a link between regional avalanche danger and modelled snow cover variables, a test region was chosen where both modelled snow cover data and an estimate of the regional avalanche danger was available. We selected the region of Davos in the Eastern Swiss Alps (225 km^2), which covers typical avalanche-release zones with an elevation range of 1600 to 2800 m a.s.l. The study plot in this test region indicates that the automated weather station Weissfluhjoch, at 2540 m a.s.l., shows an average maximum snow depth of approx. 2.2 m for the nine winters (1999-2007; 1229 days) when the required data were available.
The forecast of the regional avalanche danger level (Fig. 2.1), which is issued every day at 0800h, was used as a proxy target parameter as an accurate measure of the avalanche danger is not available (Meister, 1995). The frequency of the danger levels in the chosen time period (black columns) can be seen in Figure 2.1 (1:“Low”, 2:“Moderate”, 3:“Considerable”, 4:“High” and 5:“Very High”; ‘Very High” did not occur in this time period). Most frequent were the danger levels “Moderate” and “Considerable”. The frequency in the chosen time period was substantially different from the one in the periods used in the studies of Brabec and Meister (2001) (dark grey) and Schweizer and Föhn (1996) (light grey). For the danger levels “Low” to “Considerable” the avalanche danger was characterised by a very high persistence. The probability that the avalanche danger tomorrow will be the same as today was around 80%.
As input variables, modelled snow cover data were generated for the location of Weissfluhjoch. The relevant processes influencing the regional avalanche danger (e.g. new snow, wind or weak layer formation) are assumed to be represented in the chosen study plot. The snow cover model SNOWPACK was used to model settling and layering of the snow cover as well as its energy and mass balance (Bartelt and Lehning, 2002; Lehning et al., 2002a,b). This model requires meteorological data as input. We focused on dry-snow situations since dry-snow avalanches are the main threat over most of the winter. For forecasting wet-snow avalanches models need to be trained separately.

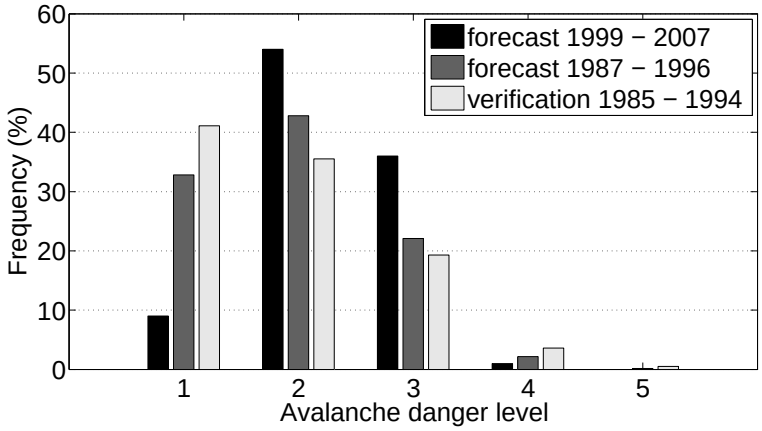

Figure 2.1: Relative frequency of the avalanche danger levels for the region of Davos. Black columns show the forecasted levels during the time period considered in this study. Dark grey columns show the forecasted levels for the study of Brabec and Meister (2001) and light grey show the verified levels used in the study of Schweizer and Föhn (1996).

For each winter a date was determined for the beginning of wet-snow conditions (i.e. two weeks before the snow cover modelled for the Weissfluhjoch study plot became isothermal), which is usually around the beginning of April. We furthermore restricted the dataset to days with snow depth >75 cm, because we think that the stability part of SNOWPACK produces more reliable results with this restriction. This restriction excluded only a limited number of days because the avalanche danger forecast for our region started at about the time when that snow depth was reached – for most winters by the end of November.

A stability index (Schweizer et al., 2006) defined the potential weak layer interface in the modelled snow cover. Motivated by a study which evaluated stability from observed snow stratigraphy (Schweizer and Jamieson, 2003), characteristics of the weak and adjacent layer and of the slab were considered. These model variables were completed with measured and calculated meteorological and snow-surface variables (e.g. wind velocity, outgoing longwave radiation or surface albedo).

All statistical methods described here were tested on a dataset in order to assess their classification capability. This test-dataset (Fig. 2.1) was used for the study of Schweizer and Föhn (1996) and contained a verification instead of a forecast of the regional avalanche danger level and observed, rather than modelled, snow cover data as input variables. This dataset was more reliable than the dataset covering the years 1999-2007 and therefore adequate to test the statistical methods for this specific problem. The test-dataset contained 10 winters between 1985 and 1994. Unfortunately no modelled SNOWPACK data can be produced for that time period, because no automated weather station existed before the 1990s. This meant modelled and measured snow cover data and their explaining power could not be directly compared.

2.2.2 Performance measures

The quality of a method was assessed by a cross-validated hit rate (HR) for all danger levels and by the true skill score (TSS) for each of the four danger levels separately (Wilks, 1995; Doswell et al., 1990).
Since both input and target variables were auto-correlated, e.g. a weak layer might have an influence over a long time period, random cross-validation turned out not to be a useful method, giving unrealistically high hit rates. Therefore, each winter was forecasted using a model and a variable selection based on the remaining eight winters, i.e. both model parameters and input variables could change in this cross-validation scheme.
It seems useful to draw special attention to days on which the danger level changed: first, these days might be the most important to predict correctly; second, because of the high persistence of the target parameter, an obviously useless forecast which always predicts the level of the day before, would show good hit rates for all days. Introducing HR and TSS for only those days on which the danger level changed, helped to identify methods which tended to a persistent forecast. Similarly, the HR and TSS were also considered for the days on which the modelled danger level increased or decreased.

2.2.3 Variable selection

Since the snow cover model delivers a large number of variables at high temporal resolution, variable selection is useful, firstly for data reduction, which increases the speed of the final statistical forecasting model, secondly to achieve an overall improved performance and thirdly to understand which variables are important (Guyon and Elisseeff, 2006).
For many variables, it may make sense to also consider their sum, mean, extreme values or rate, for different time intervals or time lags. For example, the 24 hour change of the air temperature may be more correlated with the regional avalanche danger than the midday value, or the previous day's snowfall more correlated than that of the current day. This leads to a rapid increase in the number of possible variables. A simple univariate variable selection, a Fisher's discriminant analysis (e.g. Bishop, 2006), was performed to determine, for each variable, the two most important derived variables. After this step, a large number (300) of variables still remained. Subsequently a rating of this variable set was carried out with Fisher's discriminant analysis. An alternative ranking was obtained with univariate classification trees (Breiman et al., 1998). To avoid overfitting, the best tree was determined by a pruning algorithm based on cross-validation, and the number of terminal nodes was limited to 10. The cross-validated HR of these univariate pruned trees delivered a ranking between the variables. Only variables that were not pairwise linearly correlated ($r^2 < 0.6$) were considered. Since correlated variables are not redundant per se, the redundancy of excluded variables was visually examined with scattergraphs (Guyon and Elisseeff, 2006). For support vector machines (SVM) a variable selection algorithm was used, based on the Fisher's discriminant analysis in combination with SVM (Chen and Lin, 2006).
Additionally, for the nearest neighbour method (KNN) a combination with a genetic algorithm (GA) was used to rank variable relevancy (Li et al., 2001). Several variable subsets (each of 10

variables) were tested with the KNN method. "Good variable subsets" (see below for the selection criteria) were stored in a final pool. Once the final pool was filled up to a certain threshold, the cardinality of each single variable was interpreted as a relevancy ranking: the more often a single variable was selected as a member of a good variable subset, the more relevant the single variable was assumed to be. Since testing all possible ten-member variable subsets of 300 different variables would cost too much computing time, a genetic algorithm (GA) was used as a search tool to create "good parameter subsets": the GA maximised a fitness function which depended on the selected variable subsets. The fitness function was a combination of the three different hit rates described above (i.e. for all days, for the days on which the avalanche danger level changed, and for days on which the modelled danger level changed its value). In more detail, the GA was initiated with randomly selected 100 ten-member variable subsets. The GA then determined a maximum by continuous mutation and stored it in the final pool. This procedure was reiterated until the final pool contained 100 variable subsets, which was large enough to achieve reproducible results.
For all methods variables were scaled linearly to [-1,1], except for categorical variables which were translated to Booleans for each category. As input for the statistical methods subsets of the 5, 10 and 30 most important variables of each algorithm were tested.

2.2.4 Statistical methods

In order to find an optimal link between input variables and predicted regional avalanche danger, we used the following statistical methods:

- classification trees (TREE),
- artificial neural networks (ANN),
- nearest-neighbour methods (KNN),
- support vector machines (SVM),
- hidden Markov models (HMM).

In the following each method is described in more detail:
A simple classification tree with only one measured variable, the 3 day sum of the new snow (*HN3d_meas*), performed very well and was therefore used as a benchmark for more complex models with more (especially modelled) input variables. For this tree and more complex trees using more input variables, generalisation was achieved with a pruning algorithm using 10-fold cross-validation on the training sets.
Artificial neural networks (ANN), both feedforward and recurrent (e.g. Bishop, 2006; Elman, 1990), with different setups (hidden layer size, training algorithms, early stopping or Bayesian regularisation) were trained for each winter. Since results depend on the initial weights, the networks were initialised five times and the mean of the results was considered. The results of ANNs discussed in the next section were gained with a recurrent network which used adaptive learning, 100 hidden neurons and 100 passes through the sequence.

The same nearest neighbour approach as used by Brabec and Meister (2001) was applied for a direct comparison to previous work. In addition, this method was modified in two ways. (1) The avalanche danger level of the previous day was used as an additional input to predict the current day. With this information the training dataset was reduced: only those days that showed the same avalanche danger level on the previous day were considered as possible nearest neighbours for the current day. This implicated an even more unbalanced classification problem due to the persistence of the target parameter. (2) The classification was modified. A decision boundary was used to determine if the danger level should change to a new value on a particular day. Using 10 nearest neighbours, three of these days must show a higher (or lower) danger level. This decision boundary was obtained by optimising the TSS (Heierli et al., 2004). Subsequently the danger level was determined with a majority vote between the neighbours showing higher (or lower) danger levels. For breaking a tie the nearest neighbour among them was used. Weights in KNN methods allow certain variables to have more influence while calculating the distance between days. Optimal weights were determined with a genetic algorithm (Purves et al., 2003).
Support vector machines (Schölkopf and Smola, 2002) were applied to the problem using the LIBSVM software package of Chang and Lin (software package available at `http://www.csie.ntu.edu.tw/cjlin/libsvm`). Gaussian radial basis functions with radius γ were used as kernel functions. Cross-validation on the training sets was performed to obtain γ and the penalty variable C (Chen and Lin, 2006).
In addition, hidden Markov models were used since they are able to predict time series (Rabiner, 1989; Bishop, 2006). To obtain a discrete input for the HMM, the continuous input vectors were mapped into a discrete codebook index with K-Means vector quantisation.

2.3 Results

2.3.1 Variable selection

Since the variables were selected for each forecasted winter separately, we present in this section a summary of the results obtained for each winter. In Table 2.1 the most important variables are listed, which were selected by the three methods Fisher's discriminant analysis (Fisher); univariate classification trees (TREE) and the combination of genetic algorithm and nearest neighbour (GA/KNN) described above. In the last column the sign of correlation with the avalanche danger level is given to allow a plausibility check. Except for the variables strain rate of the weak layer, 3h rate of crust thickness and net longwave radiation a physical interpretation can easily be given. Important variables were the new snow depth (*HN*, *HN3d*) and the other new-snow variables) with larger values at higher avalanche danger levels. Surface albedo, 3h rate of the slab thickness and relative humidity can also be related to new snow situations, although they are not pairwise linearly correlated. The snow transport index (Lehning and Fierz, 2008) was also selected by all three algorithms with larger index values (more drifting and blowing snow) at higher levels. Higher maximum wind speeds in the last 24 hours were also correlated with higher avalanche danger levels. The deformation index (Lehning et al., 2004) was also found to be important, despite the fact that

a bug has recently been discovered in the description of the temperature dependence, although the index should still describe a relation between critical and actual stress in the bonds. As the definition suggests, small index values are correlated with high avalanche danger. The variable crust is positively correlated to the number of elapsed days without snowfall and therefore negatively correlated to avalanche danger. The modelled profile type "Four" (Schweizer and Lütschg, 2001) was related to the danger level "Considerable". The profile type "Four" is the most frequent in the dataset which describes a weak base. Although a weak base is not itself conclusive (Schweizer and Wiesinger, 2001), it makes sense that a weak base is related to the danger level "Considerable" as a sign of structural instability. Small bond size and low density of the weak layer might also be considered as signs of structural instability.

A result of a variable selection with the combination of GA/KNN methods can be seen in Figure 2.2, which shows every single variable on the x axis and their frequency in the final pool on the y axis. Variables with a high pick-frequency were interpreted as important. This method selected, amongst others, the three-hour rate of the outgoing longwave radiation, which was highly correlated with the rate of the snow surface temperature. Strong warming of the snow-surface before midday was related to lower avalanche danger. It is worth noting that this method chose especially meteorological or snow-surface variables. This may be due the fact that the avalanche danger level of the previous day was introduced as an additional input. This already provides a certain level of stability to the system so that snow stratigraphy variables become less important. Accordingly, only weather or surface properties were important because they describe the change in danger level. To confirm this hypothesis, the GA/KNN method was performed without that additional information. However, the classification power of the KNN method without the additional information of the danger level on the previous day was too poor to reach a conclusion. However, better methods were too computationally demanding for the combination with the genetic algorithm.

Table 2.1: Overview of the variable selection results

Selected variables	measured/modelled	Selected by algorithm Fisher	TREE	GA/KNN	Sign of correlation with target
24 hour new-snow sum (*HN*)	modelled	×	×	×	+
72 hour new-snow sum (*HN3d*)	modelled	×	×	×	+
24 hour rate of *HN3d*	modelled	×			+
3 hour new-snow sum on day-1	modelled	×		×	+
Deformation index	modelled	×	×	×	-
Strain rate of weak layer	modelled	×	×		-
Thickness of surface crust on south slopes (crust)	modelled	×			-
3 hour rate of crust thickness	modelled	×			+
Snow transport index	modelled	×	×	×	+
Profile type "Four"	modelled	×			3[a]
Bond size of the weak layer	modelled		×		-
Density of the weak layer	modelled		×		-
3 hour rate of the slab thickness on day-1	modelled		×		+
3 hour rate of outgoing longwave radiation	modelled			×	-
Net longwave radiation	modelled			×	-
Surface albedo	modelled			×	+
24 hour maximum wind speed	measured			×	+
Relative humidity	measured			×	+

[a]To distinguish between "Considerable" and the other danger levels

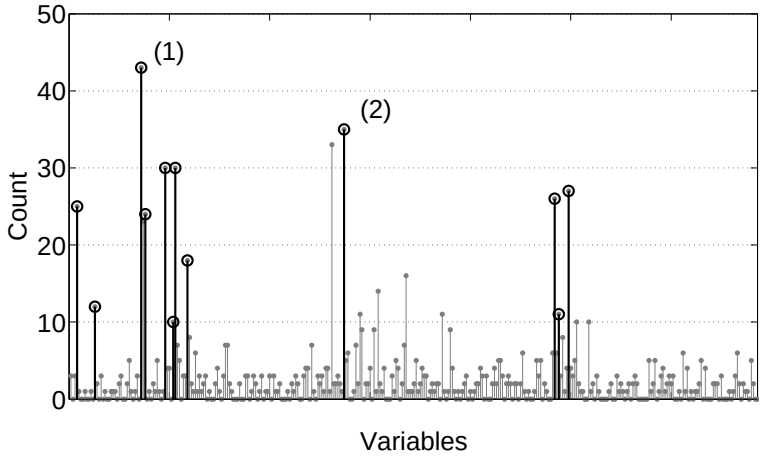

Figure 2.2: Variable selection with the GA/KNN method for winter 1999/2000. Frequency (y-axis) for each single variable (x-axis) in the final pool. The 12 variables selected are marked, e.g. the two most important variables were (1) *HN* and (2) the 3 hour rate of outgoing longwave radiation.

2.3.2 Statistical methods

The a priori capability of the statistical methods was compared with the cross-validated HR of the test-dataset (verified danger levels). Classification trees had the lowest performance, while SVM, ANN, HMM and the KNN method using the avalanche danger level of the previous day as an additional input achieved HRs of around 60%. These results were comparable to the DAVOS4 model (63%) used by Schweizer and Föhn (1996) or the Kohonen neural network (61%) used by Schweizer et al. (1994).
Since the distribution of the forecasted avalanche danger level of the data set covering the years 1999-2007 (Fig. 2.1) was substantially different to previous studies (Schweizer and Föhn, 1996; Brabec and Meister, 2001), results were not directly comparable. Therefore the method used by Brabec and Meister (2001) with measured meteorological data as input (BRABEC) was applied to the time period of this study.
In the following, results will be given for a selection of models, which represent the range of model quality. No results for Support Vector Machines (SVM) are presented since they delivered no additional improvements. The selected models and their input variables are:

- BRABEC (Brabec and Meister, 2001), input variables and their weights are presented in Table 2.2.
- Classification tree with only the measured variable *HN3d_meas* as input (TREE).
- Hidden-markov model (HMM) with the five best variables selected by the Fisher's discriminant analysis
- Recurrent Artificial Neural Network (ANN) with the five best variables selected by the Fisher's discriminant analysis.

Table 2.2: Input variables used in the model BRABEC (Brabec and Meister, 2001)

Description	Weight
New-snow depth (cm)	5
Snow depth (cm)	1
Weather and intensity (code)	1
East component of wind (kt)	3
North component of wind (kt)	3
Air temperature (°C)	2
Snow temperature (°C)	2
Snow surface (code)	1
Penetration depth (cm)	2
Density of new snow (kg m^{-3})	2
72 hour new snow sum (cm) (*HN3d_meas*)	2
72 hour rate of air temperature (°C)	3

- Nearest neighbour method with the 12 best variables selected by the GA/KNN combination and the avalanche danger level of the previous day as an additional input (KNN).
- Same KNN method as above, but with the same input variables as used in BRABEC and the avalanche danger level of the previous day as an additional input (KNN_vg).

Figure 2.3 shows the performance of the selected methods considering all days. All new methods reached a higher HR than the BRABEC method (55%). Even the simple TREE method achieved a remarkable HR of 65%. However, the TREE method produced only the most frequently used danger levels "Moderate" and "Considerable". Thus the TSS values of the other danger levels were 0%. For most of the nine winters the split value for 3 day sum of new snow (*HN3d_meas*) was 12 cm, which discriminated between the danger levels "Moderate" and "Considerable". The variable *HN3d_meas* was also amongst the input variables for the BRABEC method (Brabec and Meister, 2001). Using prior probabilities, classification trees are, in general, also able to handle unbalanced target parameters as exist in our case (Fig. 2.1). Whereas this improved the TSS value for the danger levels "Low" and "High", it produced insufficient overall hit rates. Also more variables did not improve the results.
The other methods achieved better hit rates and better true skill scores for each single danger level. With an HR of 73%, the KNN method performed best.
Considering all days, also the KNN_vg method (as KNN, but with measured, not modelled, input variables) achieved similar results. However, Figure 2.4 indicates the weakness of that method; it shows the same methods and performance measures as in Figure 2.3, but only for those days on which the target parameter changed. The low values in HR and TSS indicate that the KNN_vg

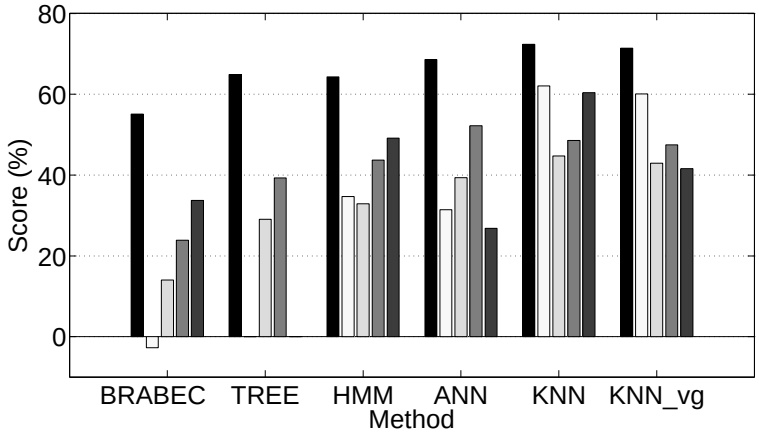

Figure 2.3: Comparison of the various model performances for all days. Abbreviations are explained in the text. The black column is the cross-validated HR of all danger levels, the grey columns the cross-validated TSS value for each of the four danger levels (lightest to darkest indicate "Low" to "High").

method was not able to predict these important days satisfactorily. All other methods showed similar hit rates. For the KNN method it is interesting that the TSS values were greater for the higher avalanche danger levels. It seems to be easier for this model to predict an increase in avalanche danger than a decrease. The HMM and ANN methods best predicted a decrease in the avalanche danger, which can be seen by the more balanced TSS values between each danger level. Also the simple TREE method was reasonably able to predict the days on which the target parameter changed. Using the variable *HN3d_meas* as input it may not be surprising that the TREE method was able to predict an increase in danger level, but a decrease was also predicted correctly in 50% of all cases.

Since the detection of days on which the avalanche danger decreased or increased is important information, models were trained for the binary criteria "Increase or not" and "Decrease or not". However, no improvement was achieved for these, at the first glance, simpler classification problems compared to the methods which forecast the four avalanche danger levels. In the case of decreasing danger one possible explanation is that it might be very difficult for human forecasters to decide consistently whether the avalanche danger level should decrease on a certain day and not on the day(s) before or after. Therefore, it might also be difficult to find reasons in the parameter space of the input variables presented to the statistical methods. In the case of increasing danger the reasons for the change in danger level might differ between danger levels, especially between an increase from "Low" to "Moderate" and from "Considerable" to "High". This information is lost if the models are trained only on these binary problems.

As we saw an improvement when adding the avalanche danger level of the previous day as additional information to the KNN method, this procedure was also implemented for the other methods. Due to the persistence of the target value, this implicated a highly unbalanced classification problem.

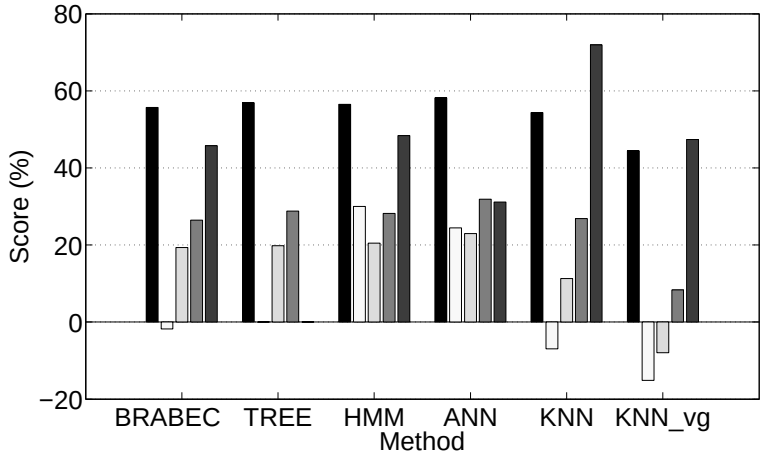

Figure 2.4: Comparison of the various model performances for days on which the danger level changed. Abbreviations are explained in the text. The black column is the cross-validated HR of all danger levels, the grey columns the cross-validated TSS value for each of the four danger levels (lightest to darkest indicate "Low" to "High").

Although most methods provided a formal method to include this additional information, for example prior probabilities (Breiman et al., 1998) or different penalties in the training error function, none of the methods showed an improvement.
A final test was whether the spatial variability of the input variables had additional classification power. Simulations of virtual slopes of the four main aspects and the flat field at two automated weather station (Weissfluhjoch, 2540 m a.s.l. and Hahnengretji, 2490 m a.s.l.) in the test region were used as input for the statistical methods. The variability was implemented either (1) as the maximum difference between the 10 simulations of each variable or (2) with additional categorical data describing the station and the aspect of the simulations. Neither option improved the performance of the statistical methods.
In Figure 2.5 shows a comparison between the KNN method and the forecasted avalanche danger for a representative winter.

2.4 Discussion and conclusion

The object of this paper was to analyse whether modelled variables of the snow cover model SNOWPACK improve the performance of forecasting models that use statistical methods, in comparison to models that are based only on measured weather data. The simple classification tree which uses only one measured (not modelled) variable (*HN3d_meas*) distinguished well between the two most frequent danger levels of "Moderate" and "Considerable". With measured meteorological input variables, as used by Brabec and Meister (2001), a more balanced model performance for all danger levels resulted, but with a strong decrease in the hit rate (HR), independent of the statistical method used. Using additional SNOWPACK variables increased the overall hit rate and produced

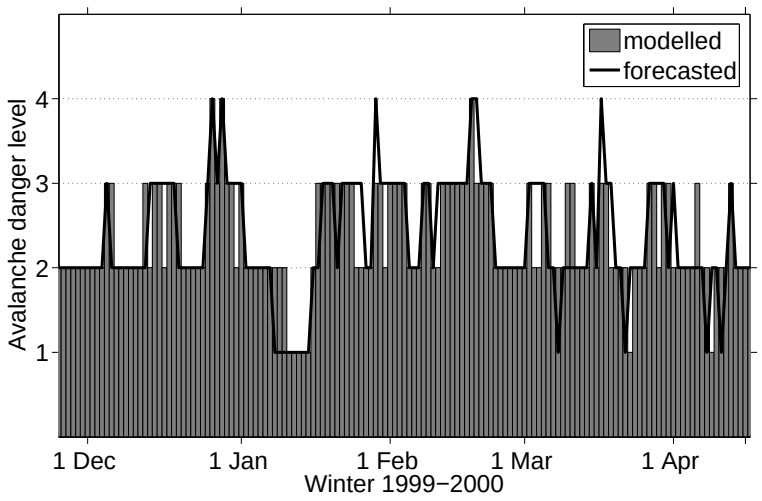

Figure 2.5: The avalanche danger forecasted and modelled with the KNN-method for the winter 1999/2000.

a balanced performance for all danger levels at the same time. The best results were achieved with a nearest neighbour method which used the avalanche danger level of the previous day as additional input. A cross-validated HR for all days of 73% was obtained. Evaluating the accuracy of the models for the days only on which the target parameter changed (e.g. the avalanche forecast changed from "Considerable" to "High"), showed that it is only possible to obtain reasonable results with additional SNOWPACK variables. These findings suggest that for a balanced performance between all danger levels and for good overall accuracy, especially for days when the danger level changes, modelled snow stratigraphy data as provided by SNOWPACK are needed.

Model performance did not improve using SNOWPACK simulations of two automated weather stations and/or virtual slopes simulations. This may lead to the conclusion, that the modelled snow cover variation (though limited) was not related to regional avalanche danger. Similarly, Schweizer et al. (2003b), who found typical measured point stability distributions for the regional danger levels, did not find a relation between mean stability and stability variation. However, we do think that considering spatial variability is important for assessing local to regional avalanche danger. Our conclusion only concerns the situation in which we tried to include more than one meteorological station of the region in the analysis. This approach does not deal with the important local variability, which is given by variations in local energy balance (Helbig et al., 2009), or preferential deposition and drifting snow (Lehning et al., 2008).

Nonetheless, a hit rate of around 70% reveals a remarkable discrepancy between the operational avalanche danger forecast by human experts and a statistical model. Since we have tested a wealth of methods with a huge range of complexity, we feel that the main problem is that it seems impossible to reproduce the human decision on the avalanche danger with the input used in our analysis. This missing link was recognised as a problem in the study of Schweizer and Föhn (1996), despite the fact that they used observed snow cover variables. A study by Schweizer et al. (2008b) showed

that a level of uncertainty exists in the detection of unstable slopes with Rutschblock tests and snow profiles (probability of detection of 70%). Such snow profile interpretations are important arguments for the human prediction or verification of the regional avalanche danger level. Two possible conclusions remain: (1) additional information, which is not formalised at present enters the decision process such as the experience and intuition of the individual; (2) the forecasted danger level is not a good target variable, since it might be erroneous due to incorrect data at the time of the forecast or due to variations in human perception (McClung and Schaerer, 2006).

An operational prediction of the avalanche danger level with a statistical model augmented with SNOWPACK variables as additional input variables would include the following steps: First the present snow cover is simulated with measured data, second the development of the snow cover is predicted with forecasted meteorological data for the next day. This predicted snow cover provides the additional input variables for the statistical methods presented. This would be a fully automated process which could be applied to the whole area of the Swiss Alps.

In comparison to the study of Durand et al. (1999), who compared the output of the French SCM chain to avalanche activity (which can be forecasted reliably without modelled snow cover parameters; (e.g. Pozdnoukhov et al., 2008)), our study presented models trained and tested on the regional avalanche danger as forecasted in the public avalanche bulletin. Using this target parameter the models with modelled snow cover input variables performed better than models that used mainly meteorological input variables.

Our study also showed that the uncertainty in the prediction of the avalanche danger level needs to be quantified. For example, the shift in the danger level distribution to the intermediate danger levels that obviously occurred during previous years should be clarified.

3 Statistical evaluation of local to regional snowpack stability using simulated snow-cover data

3.1 Introduction

The European avalanche danger scale is defined based upon snowpack stability (i.e. the probability of avalanche release), the frequency of trigger points (spatial distribution of instability) and the size and type of the anticipated avalanches (Meister, 1995). Stability is the only property that can be estimated through observations and interpretations of snow profiles and stability tests (Schweizer and Wiesinger, 2001; Schweizer et al., 2008b). Since observations are time-consuming and sometimes dangerous, avalanche warning services sparsely receive information about snowpack stability. Numerical modelling of snow cover stratigraphy and stability has been proposed as a solution to this problem.

The question arises how reliable are evaluations provided by numerical models. Durand et al. (1999) compared their modelled stability estimate of the SAFRAN/Crocus/MÉPRA (SCM) chain to observed avalanche activity. Despite the fact that two different datasets could not be strictly compared, the forecast quality was similar to what is typically achieved with weather information alone (Heierli et al., 2004; Pozdnoukhov et al., 2008). The link between avalanche activity and corresponding danger level has previously been shown to be inconsistent. Schweizer et al. (2003b) reported this discrepancy was mainly due to limited visibility during periods of high activity. Furthermore, said study concluded that avalanche occurrence data are not suitable to verify lower danger levels ("Low", "Moderate" and "Considerable"). Schirmer et al. (2009a) linked statistically simulated snow cover data to forecasted avalanche danger. It was shown that simulated snow cover information was useful for statistical danger level prediction and provided additional benefit in comparison to weather information alone. Lehning et al. (2004) summarised the stability evaluations implemented in the snow cover model SNOWPACK and assessed their quality in comparison to the forecasted avalanche danger level. Schweizer et al. (2006) developed a new stability evaluation based on SNOWPACK simulations obtaining critical thresholds between three stability classes. However, those results were not cross-validated because of a limited dataset of $N = 33$. Hence the question of how well this evaluation might perform on an independent dataset could not be answered.

The first aim of this study was to assess the quality of existing stability estimates implemented in SNOWPACK using a large dataset of observed stability. The second aim was to test modelled variables equivalent to manually observed variables Schweizer and Jamieson (2003) found to be sig-

nificantly discriminating between stable and unstable observed profiles. The third aim was to detect additional modelled variables which discriminate well between stability categories. These variables were then used to link modelled stratigraphy to observed stability through the use of classification trees.
In addition to other automatic methods potentially supporting avalanche danger forecasting (for detecting avalanche days with weather data (e.g. Buser, 1983; Heierli et al., 2004; Pozdnoukhov et al., 2008), for predicting the avalanche danger level itself using measured (Schweizer and Föhn, 1996) or simulated snow stratigraphy (Schirmer et al., 2009a)), the approach developed in this study for estimating stability, covers a supplementary facet of the avalanche danger prediction process, which could be used as a support tool for avalanche warning services.

3.2 Methods

3.2.1 Overview

The main purpose of this study was to relate modelled snow cover data to measured stability information. A summary of the methodology follows:

1. 775 stability observations from one region were used as target variable, which were interpreted with two (one subjective and one objective) rating schemes.
2. Modelled snow cover information evaluated at five automatic weather stations in that region was used to explain the observed stability.
3. Due to the large amount of modelled snow cover information a reduction of data was applied:
   a) Only modelled properties of the slab, the weak layer and the surface were considered together with measured and modelled meteorological variables.
   b) The large amount of remaining variables (∼300) were rated with the Fisher criterion to select the 20 best uncorrelated variables.
4. Classification trees were trained with the 20 best variables as input to explain observed stability. Similarly, univariate trees were built in order to test stability estimates already implemented in the snow cover model.
5. The results achieved from the model were validated through comparison to expert knowledge and agreement with observations.
6. Additionally, a discussion on whether the classification trees could be used for a probability forecast is presented.

3.2.2 Data

In order to relate stability observations with simulated snow cover data using automatic weather stations (AWS) as input, a test region was chosen, where many snow profiles with stability tests in

the surrounding of weather stations were available. An analysis of the SLF snow profile database showed that only the region of Davos, in the Eastern Swiss Alps, had a sufficient number of stability observations for a statistical analysis. Five AWS are located in the region. We selected observations within a 5 km radius and an elevation band of ±300 m of the stations. These thresholds were chosen to consider the two following conflicting aspects: (i) increasing the dataset would make the statistical analysis more reliable, while (ii) observations at larger distances to the AWS might have less relation to the simulated snow cover data. We obtained 775 cases where both the stability observation and the simulated snow cover were available.

The observations included a Rutschblock test and a snow profile. Since the Rutschblock score is dependent on the inclination, we considered only observations from slopes of an inclination >20°. Observations with snow depth less than 50 cm were not included, as it was assumed that with this restriction the stability interpretation would be more reliable. Since we were mainly interested in dry snow situations, only observations between November and April were considered.

The observations were rated into the three stability classes "poor", "fair" and "good". This was achieved through the application of two existing stability interpretations. The first of which is a subjective interpretation scheme developed with expert knowledge (Schweizer and Wiesinger, 2001). Expert knowledge is also needed to apply this rating. For the dataset used in this study profiles were rated by different people. We must therefore assume that the rating is not fully consistent, on the other hand, the expert is able to include a broad spectrum of information in the rating. The second rating is an objective, rule-based method. This method was statistically developed in trying to find differences between observations recorded on slopes that were adjacent to skier-triggered avalanches ("unstable") and those that were skied but not triggered ("stable") (Schweizer et al., 2008b). Applying both interpretations, a relatively similar distribution of the three categories was obtained (30% "poor", 30% "fair" and 40% "good"). However, in only 60% of the cases did the ratings agree. Slight differences in the distribution indicated that the objective interpretation is more conservative (tends to produce more unstable ratings).

The snow cover model SNOWPACK was used to generate the corresponding snow stratigraphy (Bartelt and Lehning, 2002; Lehning et al., 2002a,b). The model provides a huge amount of data in high temporal resolution. Therefore we reduced the data by considering mainly failure layer and slab properties. The stability index (SSI) developed by Schweizer et al. (2006) defined the potential weak layer interface in the modelled snow cover. Similar to a study of observed stratigraphy, the softer layer was chosen as failure layer and the harder as adjacent layer (Schweizer and Jamieson, 2003). These modelled stratigraphy variables were completed with meteorological and snow-surface variables (e.g. measured wind velocity or modelled surface albedo). In addition to modelled values at noon, we considered also sum, mean, extreme values or rate, for different time intervals. However, this leads to an increase in the number of possible variables making a reduction of the variables even more necessary.

We evaluated existent stability evaluations implemented in SNOWPACK, which are mainly stability indices relating parameterised shear strength with shear stress. Those are Sk_{38} (Jamieson and Johnston, 1998) and SSI delivering continuous values, while a combination of both, which is abbreviated as "SC" (Schweizer et al., 2006), delivers the three stability categories "poor", "fair"

and "good".

3.2.3 Rating of variables

A simple univariate rating was performed using the Fisher criterion, which is defined as the ratio of the between-class variance to the within-class variance (e.g. Bishop, 2006) and for a two class problem is given by

$$J = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}, \tag{3.1}$$

where m_i is the mean and s_i the standard deviation of class i, i = 1,2. As input for the classification trees described below the 20 best non-pairwise linearly correlated variables were considered (r^2 <0.6).

3.2.4 Classification

Classification trees (Breiman et al., 1998) were used to discriminate between the stability categories. The number of stability categories were reduced to minimise the dimensionality of the forecast verification problem. Murphy (1991) defined the dimensionality of a forecast with the number of quantities needed to reconstruct the joint distribution of forecast and observation. Since the absolute forecast verification problem of a three-category forecast is already eight dimensional (square of number of categories - 1), the (interrelated, but) multi-facetted nature of forecast quality will be easier to obtain when the number of categories is reduced (Murphy, 1991, 1993). Therefore we trained and verified the trees not on the three categories of the stability observation ("poor", "fair" and "good"). Instead, trees were built for the detection of "poor" observations, and other trees for the detection of "good" observations, while the remaining two categories were pooled.
The classification trees were obtained by optimising the misclassification costs and the complexity (size) of a tree (Breiman et al., 1998). Furthermore, altered prior probabilities were used to adjust the individual class misclassification. Choosing a prior probability larger than the observed relative frequency will tend to decrease the misclassification of that class. This will also introduce a bias between modelled and observed class relations in form of a higher modelled relative frequency (Breiman et al., 1998). We decided to adjust priors to achieve a forecast with larger values for another – for this study more important – quality characteristic (skill, definition in section 3.2.5 and in Table 3.3), which outweighed the disadvantage of the bias in class relations. Through varying the prior probabilities we optimised the true skill statistic (Doswell et al., 1990). Due to the previously mentioned difference in characteristics between modelled and observed classes, from here on "good" and "poor" conditions will be referred to as "rather stable" and "rather unstable", respectively. Both model parameters, tree size and priors, were determined through cross-validation.

Table 3.1: Types of goodness of a forecast (Murphy, 1993)

Name	Description
Consistency	Type 1 goodness, consistency with existing (physical) process understanding.
Quality	Type 2 goodness, agreement between forecasts and corresponding observations, which can be expressed with measures listed in Table 3.3.

Table 3.2: Contingency table for a binary forecast ("1": event, "0": non-event). Total number of cases: $N = a + b + c + d$

		Observation x	
		1	0
Forecast f	1	a	b
	0	c	d

3.2.5 Validation

Two distinct types of goodness for a forecast system were considered as identified by Murphy (1993) (Table 3.1): The first is the correspondence between forecasts and judgements. This is the so-called type 1 goodness or consistency. We evaluated whether the objective data analysis led to a model which was mechanistically consistent with existing (physical) process understanding as expressed by the current expert knowledge. A discussion will be presented later as to whether the tree branching can be reasonably explained and whether the rating of the input variables with the Fisher criterion is comparable to the experts' rating. Consequently, we asked five experts to take a survey to select up to ten different modelled variables which they believed to discriminate between stable and unstable conditions.

The second type of goodness is the straightforward agreement between the forecasts and corresponding observations (i.e. the type 2 goodness or quality). This was assessed through cross-validation (CV). The quality of a forecast can be assessed with the joint distribution of forecast f and observation x, which can be displayed in terms of a 2 x 2 contingency table in the case of a binary categorical forecast (Table 3.2.5). Although the table is comprehensive, the information is more accessible when factorised in conditional distributions (Murphy and Winkler, 1987). In order to determine how well the forecast discriminates between observation classes, the probability of detection (POD) and the probability of false detection (POFD) were chosen (Table 3.3). The frequency of correct null forecasts (FOCN) and the false alarm ratio (FAR) deliver additional insight into how reliable forecasts are. These measures are sample estimates of the conditional distributions. To inspect further aspects of quality, we chose to highlight the accuracy expressed with the proportion correct (PC) and the skill of a forecast (Wilks, 1995). The skill of a forecast is defined as the relative accuracy with respect to a standard reference forecast. This reference forecast is random

Table 3.3: Quality measures (Doswell et al., 1990; Wilks, 1995)

Measure	Description
POD (Probability of detection)	Probability that event "1" was forecasted when it was observed, $p(f = 1\|x = 1)$. Estimated with $a/(a + c)$.
POFD (Probability of false detection)	Probability that event "1" was forecasted when it was not observed, $p(f = 1\|x = 0)$. Estimated with $b/(b + d)$.
FOCN (Frequency of correct null forecast)	Probability that the non-event "0" was observed when it was forecasted, $p(x = 0\|f = 0)$. Estimated with $d/(c + d)$.
FAR (False alarm ratio)	Probability that the event "1" was not observed when it was forecasted, $p(x = 0\|f = 1)$. Estimated with $b/(a + b)$.
PC (Proportion correct)	Probability that the observed event was correctly forecasted. Estimated with $(a + d)/N$.
TSS (True skill statistic)	Measure of skill. Skill is the relative accuracy with respect to a reference forecast. Estimated with POD-POFD = $(ad - bc)/(a + c)(b + d)$.

and unbiased for the true skill statistic (TSS), which is obtained by subtracting POFD from POD (Wilks, 1995). An overview of the evaluation quantities used is given in Table 3.3.

3.2.6 Probability forecast

Classification trees can be applied to create a probabilistic forecast using the class relations at a terminal node. Each terminal node i is now denoted as a separate forecast f_i. The corresponding forecast probability $p(f_i)$ can be obtained from the class relations by applying a large dataset. Since prior probabilities were used, these class relations need to be adjusted to calculate the forecast probability for each node:

$$p(f_i, j) = \frac{\pi_j N_{ij}/N_j}{\sum_k \pi_k N_{ik}/N_k}, \tag{3.2}$$

where $p(f_i, j)$ is the forecast probability for the forecast f_i and class j, π_j the prior probability, N_j the initial class frequency and N_{ij} the class frequency at node i for class j, while k = 1,2.
The quality of a probabilistic forecast can be assessed with an attribute diagram (Wilks, 1995). An attribute diagram relates the forecast probability $p(f_i, j = 1)$, obtained from the training parts of the CV blocks, to the observed relative frequency $p(x = 1|f_i)$ obtained from the test parts of the CV blocks. These quantities were calculated for the same trees and with the same cross-validation blocks, which were used for the verification of the categorical forecast (classification, see section 3.2.4). The attribute diagram delivers insight into quality aspects as reliability and resolution. Reliability of a probability forecast is the correspondence between forecast probability and the observed relative frequency, while resolution is the ability of the forecasts to sort the observed events into groups that are different from each other (Wilks, 1995). Results in terms of reliability and resolution are discussed in section 3.3.2.

3.2.7 Suitability of proposed methods

One single stability observation has only limited strength of explanation for a regional evaluation. Schweizer et al. (2003b) presented characteristic stability distributions for different danger levels. They concluded that only a sufficient number of observations allows a reliable regional stability estimate. However, the dataset in this study would not have contained enough examples if only days that had such a sufficient number of observations were considered. Therefore, days with only one stability observation were also included. The question then was whether a learning system is able to adapt rules correctly, when the stability observations, in some cases, are of limited value for the true prediction parameter "regional stability"? We assumed that there are some limited combinations of factors explaining regional stability and that these patterns repeat in time. An example being snow fall in combination with large wind speeds, or a weak layer of surface hoar crystals together with certain slab properties. Both of these may be such repeating patterns, which have an influence on regional stability. Consequently, even though this assumption was only partly fulfilled, it made sense to include days to the dataset with only one stability observation.
In the dataset used for this study, which contains 775 stability observations, 314 observations were unique per station and day. On average, 1.8 stability observations were available per day and per station. Multiple observations associated with one station on a single day were sometimes inconsistent due to the in-region variability or the uncertainty of the observation. Since simulated snow cover data were only available once per station, multiple observations were applied in the learning phase, although sometimes inconsistent, to the same modelled input variables of each day. This was intended to filter out the in-region effects and help find the main factors of repeating patterns influencing the regional stability. In the evaluation phase, median values of observations made on the same day were used, which where weighted with the number of observations per day, since the classification trees were not able to reproduce the in-region variance of observed stability. The rare cases with median values between two classes were neglected, which lead to the varying number of cases in Tables 3.5 and 3.6 compared to the original dataset ($N = 775$).
A model validation is more reliable if a better estimate of the regional stability can be used. Over several periods of time a verified regional avalanche danger level, based on a sufficient number of observations, was available for the region of this study (Schweizer et al., 2003b; Schweizer and Kronholm, 2007). Therefore, the performance of the trees for these reduced time periods were independently tested.
Modelled snow cover variables are auto-correlated over long periods of time (Schirmer et al., 2009a), which prohibited random cross-validation (Elsner and Schmertmann, 1994). Modelled values at nearby days (both past and future) introduce a bias to the estimation of forecast skill, as these values contain noise correlated to noise for the omitted day. Furthermore, nearby future days were likely to be especially informative about omitted target variables. However, they would be unavailable in real forecast situations, which would bias the estimate of the forecast skill towards higher values. Therefore, blocks of data that were uncorrelated in time were removed. In the case of this study, it was necessary to select blocks for entire winters. We applied such a winter-by-winter CV to select input variables, model parameters (tree size and prior probabilities) and to evaluate quality

aspects. One has to keep in mind that the cross-validated quality of a classification tree through CV is assessed by dissimilar trees, which are built with the training parts of the cross-validations blocks. These trees differ in input variables, nodes and tree size, both with respect to each other and to the classification tree for which the quality is validated.

3.3 Results

3.3.1 Rating of variables

The rating performed with the Fisher criterion (Eq. 3.1) was applied twice, first for the detection of the category "poor" and second for "good". For the detection of the category "good" higher values of the Fisher criterion were achieved. Higher values were also obtained when the subjective stability interpretation was used for classifying the manual observations in comparison to the objective interpretation. Furthermore, quality characteristics of classification trees trained on stability observations rated with the subjective interpretation were better in comparison to the objective interpretation. Subsequently, we will only show results obtained from the subjective stability interpretation.

The two stability indices Sk_{38} and SSI implemented in SNOWPACK showed different strength of discrimination. For both the detection of rather stable or rather unstable conditions, the values of the Fisher criterion for Sk_{38} were larger than for SSI. Similar results were achieved with a Kruskal-Wallis test for the three stability categories: The Sk_{38} was able to discriminate between the three categories ($p<0.001$), while the SSI showed no significant strength of discrimination ($p>0.05$). The combination of the indices ("SC") was also not significant.

For some of the variables which discriminated well between stable and unstable profiles for observed stratigraphy, i.e. failure layer grain size, hardness, and their differences to the adjacent layer (Schweizer and Jamieson, 2003) the sign of correlation was different for modelled profiles. For example, while in unstable measured profiles the difference in grain size across the failure interface was typically large, it was small in modelled profiles. This indicates a low consistency between model and observations.

For other significant variables in the observed stratigraphy, the signs of correlation agreed. Rather unstable profiles had shallower snow depth and lower failure layer shear strength than rather stable profiles. They were typically classified as profile types 7 and 4 (while rather stable profiles were classified as profile type 6) (Schweizer and Lütschg, 2001). Depth hoar was more often found in failure layers of rather unstable profiles (in modelled stratigraphy also faceted crystals), while rounded grains were more often found in stable profiles.

In contrast to the observations found by Schweizer and Jamieson (2003), the slab was significantly thicker in rather unstable modelled profiles. Slab density, one of the most important modelled variables according to the Fisher criterion (lower densities corresponded to more unstable profiles), was not significant in observed stratigraphy. However, there are some indications that soft slabs are more dangerous (Schweizer and Lütschg, 2001; Schweizer and Jamieson, 2003).

Table 3.4: The 20 best pair wise uncorrelated variables selected with the Fisher criterion for the detection of rather stable conditions and rather unstable conditions. The abbreviation "diffminmax" stands for absolute difference between maximum and minimum value of the mentioned time interval

Variable	Selected for rather stable	Selected for rather unstable	Selected by experts
Mean slab density	×	×	×
Mean slab bond size	×	×	×
Difference in hand hardness between slab and failure layer	×	×	×
Ratio of mean slab bond size to grain size	×	×	×
Failure layer bond size times grain size	×	×	×
Ratio of failure layer bond size to grain size	×	×	×
Failure layer 3D coordination number (Lehning et al., 2002a)	×	×	×
Depth hoar within one meter beneath the penetration depth	×	×	×
3 day new snow sum	×	×	×
3 day new snow sum (24 hour diffminmax)	×	×	
24 hour new snow sum (24 hour diffminmax)	×	×	
24 hour maximum of 3 hour new snow sum	×	×	
Ski penetration depth	×	×	
24 hour mean wind speed	×	×	×
Sensible heat flux (24 hour diffminmax)	×	×	
Surface temperature (24 hour diffminmax)	×	×	
Snow temperature at 10 cm below the surface	×	×	
12 hour rate of snow temperature at 10 cm below the surface	×		
Strain rate of adjacent layer (Lehning et al., 2002a)	×		×
Failure layer contains surface hoar	×		×
Ski penetration depth (24 hour diffminmax)		×	
24 hour mean of energy fluxes at surface		×	
Adjacent layer consists of faceted crystals		×	

Variables representing a change in time were rarely rated as important. This suggests that manual snow profiles contain the most relevant information or that snowpack changes are slow.
As input for classification trees, only the 20 best not pair wise linearly correlated variables were considered (r^2 <0.6) and are listed in Table 3.4. The majority were selected both for the detection of rather stable and rather unstable conditions. Most selected variables are modelled. Exceptions are wind speed and surface temperature, which were measured at the weather station.
As mentioned in section 3.2.7, experts were independently asked to select modelled variables which should discriminate between rather stable and rather unstable conditions. If selected variables were also chosen by experts, they are marked in Table 3.4. For most other variables selected by experts high values of the Fisher criterion were obtained. Exceptions were, for example, the difference in hardness or density between failure and adjacent layer, an increase in air temperature in the last 24 hours or the existence of a crust in the slab, which were chosen by experts but achieved low values of the Fisher criterion.

3.3.2 Classification

The trees using as input the best variables shown in Table 3.4 are presented in Fig. 3.1a for the detection of the rather stable conditions and in Fig. 3.1b, for the detection of rather unstable conditions (model "Best_20"). To evaluate the first type of goodness of a forecast as introduced above, the consistency, we discuss the physical or logical meaning of the presented classification trees ("Best_20").

For the detection of rather stable conditions (Fig. 3.1a) the most important variable was the 3-day new snow sum. On the one hand, it seems logical for many situations that no new snow is related to rather stable conditions as suggested by the tree. On the other hand, in the subjective interpretation scheme (which was applied here to rate the manual observations) no reference to the new snow amount is given. However, the new snow amount will affect other measured quantities such as hardness or the Rutschblock score, which are used by this stability interpretation. The next node used the ratio of failure layer bond size to grain size, where large values are related to rather stable conditions as large bond size in comparison to grain size suggests large shear strength. At the next node the tree classified the conditions as rather stable if depth hoar was not present within a depth of one meter below ski penetration. This is appropriate since failures in depth hoar layers deeper than one meter below the ski penetration is unlikely (Schweizer and Camponovo, 2001). Weak layers up to this depth may potentially be released by skiers (cp. Schweizer and Jamieson, 2003) and depth hoar is related to low shear strength (Jamieson and Johnston, 2001).

For the detection of rather unstable conditions (Fig. 3.1b) the most important variable selected by the tree was the ski penetration depth. Larger values were attributed to rather unstable conditions, which is consistent with the observations mentioned previously that soft slabs are related to unstable conditions. In many situations large penetration depth corresponded to new snow conditions, which points to rather unstable conditions. Additionally, a large penetration depth facilitates triggering of deep weak layers. The next nodes (on the right hand side of Fig. 3.1b) have previously been discussed. On the left hand side the next node is defined by the absolute difference between the minimum and the maximum of the snow surface temperature T_{ss} in the last 24 hours. Low values are related to cloudy conditions, which can explain the relation to unstable conditions. The tree suggests from the last node on the left hand side rather unstable conditions if the failure layer bond size is small. Small bond size should be related to low values of shear strength, hence rather unstable conditions.

In interpreting of the sub nodes in the tree hierarchy, an important consideration is that the detected dependencies are, in fact, only valid for the subgroup of the dataset reaching this node: the conditions of the preceding nodes have to be applied first. However, all dependencies detected for these subgroups were validated through testing against the whole dataset. In conclusion, the nodes and threshold values are plausible and the trees displayed a high degree of consistency when compared to the judgement of the experts.

Table 3.5 and Table 3.6 present the cross-validated contingency tables for the model "Best_20". In both cases, i.e. the detection of rather stable and unstable conditions, event "1" ("good" for the detection of rather stable, "poor" for the detection of rather unstable) was forecasted more frequently

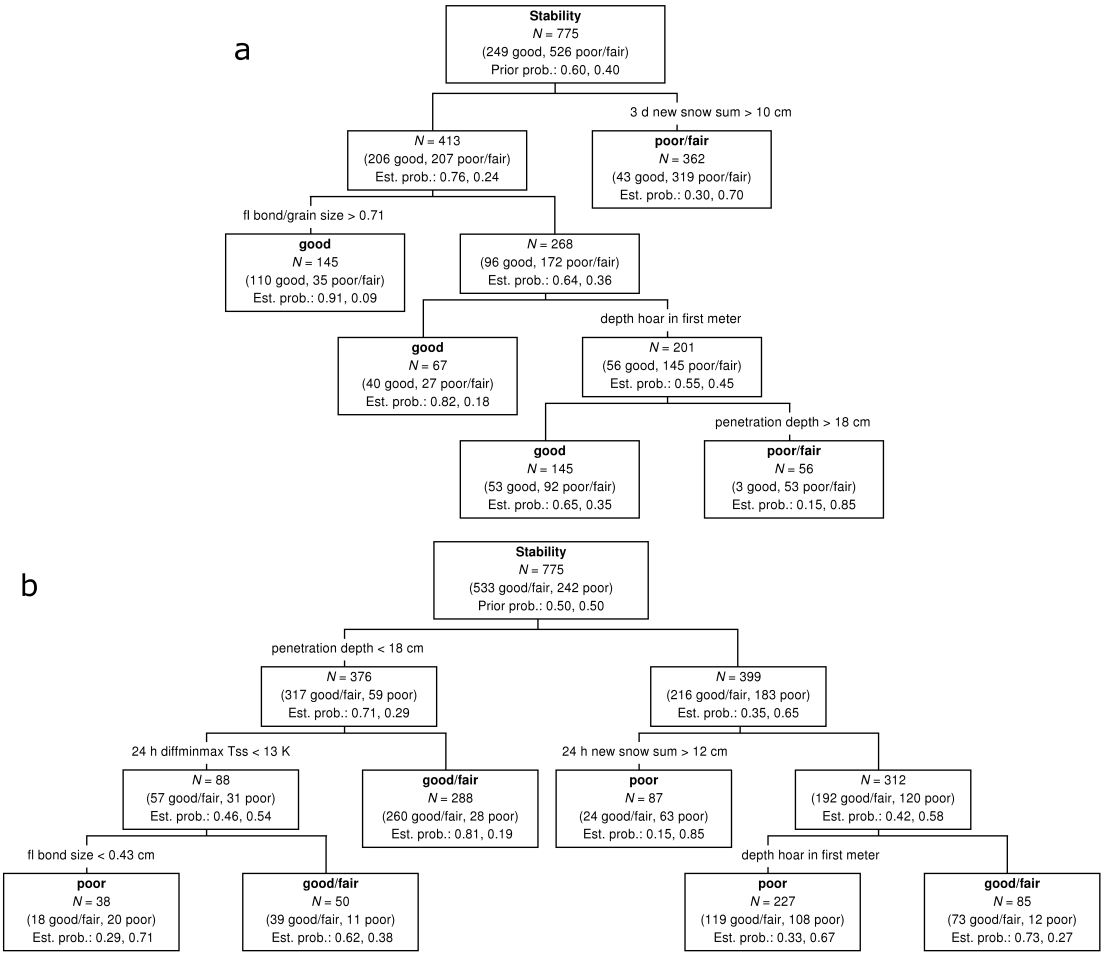

Figure 3.1: Classification tree for the detection of (a) rather stable and (b) unstable conditions using the 20 best uncorrelated variables defined with the Fisher criterion ("Best_20"). For each node the count N of examples reaching that node, the class relations and the estimated forecast probabilities ("Est. prob.") are recorded. In each first node, the values of the altered prior probabilities ("Prior prob.") are noted. Failure layer is abbreviated with "fl", temperature of the snow surface with "Tss" and the absolute difference between minimum and maximum with "diffminmax".

Table 3.5: Contingency table of the model "Best_20" for the detection of rather stable conditions. The base rate (fraction of observations of class "1") was 0.34

		Observation x		
		1: good	0: poor/fair	Total
Forecast f	1: good	190	128	318
	0: poor/fair	61	364	425
	Total	251	492	743

Table 3.6: Contingency table of the model "Best_20" for the detection of rather unstable conditions. The base rate (fraction of observations of class "1") was 0.25

		Observation x		
		1: poor	0: fair/good	Total
Forecast f	1: poor	139	201	340
	0: fair/good	41	336	377
	Total	180	537	717

than it was observed (bias larger than one (Wilks, 1995)). This was artificially produced with the altered prior probabilities, as was mentioned in section 3.2.4. Recall, the prior probabilities were optimised so as not to obtain an unbiased forecast, rather a forecast with the largest skill; with the consequence that the event "1" is more frequently predicted than observed.
The corresponding performance measures are visualised in Fig. 3.2, together with the models "Sk_{38}" and "SC". Results for trees using only the input variable SSI are not shown, since their performance measures were not as good as trees using the Sk_{38}. In the detection of rather stable conditions (Fig. 3.2a), a proportion correct (PC) of 0.75 was obtained for the model "Best_20". This high value has only limited implication, as a constant forecast of "poor/fair" would achieve a PC of $\sim$0.7 since the dataset is unbalanced with a base rate of 0.34. Such a constant forecast would have no skill, whereas the model "Best_20" achieved a true skill statistic (TSS) of 0.5. This model was able to discriminate between the observations, with a probability of detection (POD) of 0.76 and a probability of false detection (POFD) of just 0.26. The model was also quite reliable as when "poor/fair" was forecasted, it was correct with a probability of 0.86 (FOCN, frequency of correct null forecasts). The relatively high value (0.4) of the false alarm ratio (FAR) indicates some deficits in reliability. In all aspects of quality discussed, the model "Best_20" performed better than the models "Sk_{38}" and "SC". However, the differences to the "Sk_{38}" were small. The "Sk_{38}" model achieved a lower PC (0.70), a lower POD (0.72), a lower TSS (0.42), a lower FOCN (0.83) and

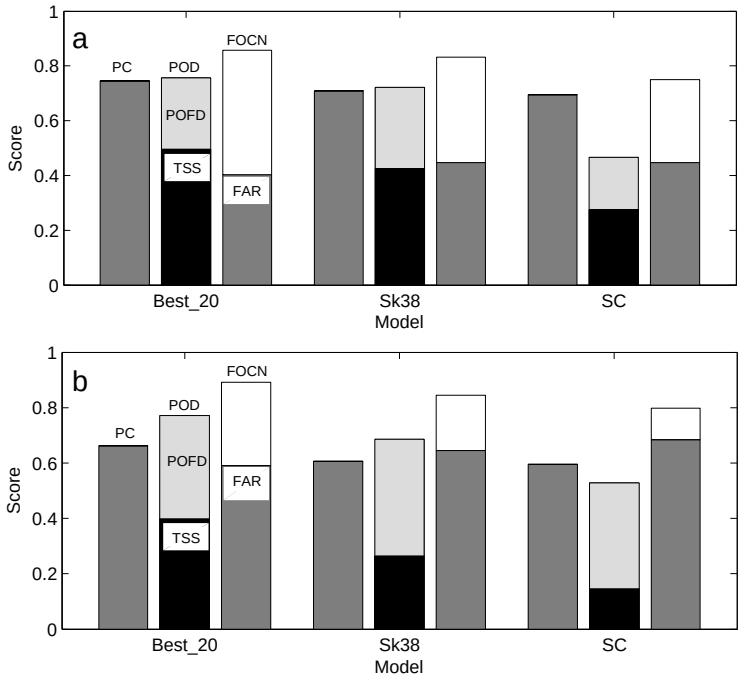

Figure 3.2: Overview of the models' quality measures for the detection of (a) rather stable and (b) rather unstable conditions. Presented are the proportion correct (PC), the probability of detection (POD), the probability of false detection (POFD) and the true skill statistic (TSS). Furthermore, the frequency of correct null forecasts (FOCN) and the false alarm ratio (FAR) are also shown.

higher FAR (0.44).

The quality of the detection of rather unstable conditions can be assessed with Fig. 3.2b. While the POD for the "Best_20" model was again quite large (0.77), the relatively large value for POFD (0.37) resulted in a lower TSS (0.4) than for the detection of rather stable conditions. While the model was very reliable when "fair/good" was forecasted (FOCN = 0.9), the same level of reliability was not present when "poor" was forecasted (FAR = 0.6). Again, for all aspects of quality discussed, the model "Best_20" performed better than the other two models.

In summary, the detection of the rather stable conditions were easier to detect, which is consistent with larger values of the Fisher criterion. Reasons for the low performance measures for the "SC" model can be found in the small dataset used in Schweizer et al. (2006) ($N = 33$). The published non cross-validated accuracy could not be achieved when cross-validation was applied using our large dataset.

The univariate tree for Sk_{38} had a single split, which suggests rather unstable conditions for values smaller than 0.58 and rather stable conditions for larger values.

As discussed in section 3.2.7, a model validation would benefit from a verified regional stability estimate. There are several days for which a reliable verified avalanche danger level could be used for validation, which is shown in Table 3.7 for the model "Best_20". This was done for the categor-

Table 3.7: Results of the cross-validated classification trees using the best 20 variables as input ("Best_20") applied to periods with verified regional danger level (Schweizer et al., 2003b; Schweizer and Kronholm, 2007). Cases for which the models did not correspond to the verified danger level were marked with an asterisk

Date	Forecasted danger level	Verified danger level	Tree to detect "poor"	Tree to detect "good"
11-12 Dec 2002	2	1	poor*	good
21-23 Jan 2002	1	2	fair/good	poor/fair
12-13 Feb 2002	3	3	poor	poor/fair
26-27 Feb 2002	3	3	poor	poor/fair
18-19 Mar 2002	2	1-2	fair/good	poor/fair*
20 Mar 2002	3	3	poor	poor/fair
7 Jan 2003	2	3	poor	good*
13 and 15-17 Jan 2003	2	1	fair/good	good
7 Feb 2003	4	3-4	poor	poor/fair
17-20 Feb 2003	2	1	fair/good	good

ical (non-probabilistic) forecast. However, what can be expected when modelled stability estimates are compared to avalanche danger levels? Referring to the typical stability distributions found by Schweizer et al. (2003b), the tree which detects the rather stable conditions must detect the danger level "1: Low": the binary distribution "good" vs. "poor/fair" is 90% to 10% for level "Low" and 25% to 75% for "2: Moderate". The situation for the tree which detects rather unstable conditions is not as clear: at danger level "3: Considerable", the typical distribution of the binary variable "poor" vs. "fair/good" is mostly balanced. However, since this tree produced more "poor" situations than are present in the observations (bias of 1.5), it is expected that the detection of rather unstable conditions might correspond to the detection of danger levels greater than or equal to "Considerable".

Considering these arguments, on 11-12 December 2002 the tree to detect rather unstable conditions failed. The verified danger level was "Low", but the tree output was "poor", which corresponds to danger levels greater than or equal to "Considerable". On the other hand, the tree to detect rather stable conditions delivered the output "good", which is consistent with the verified danger level. The inconsistent output of the two trees ("poor" vs. "good") can be interpreted as an indication of uncertainty. In the next period the verified danger level was "Moderate", hence the tree to detect rather unstable conditions should not show "poor", since the avalanche danger level is lower than "Considerable". This was correctly reproduced. Similarly, the tree for the detection of rather stable conditions should not show "good", since the danger level was not "Low", which is also correctly reproduced. All cases for which the models did not correspond to the verified danger level are marked

with an asterisk in Table 3.7 (three cases in twenty). There were no instances where both models are wrong. By comparison, for five in ten cases the forecasted danger level was one level different from the verified danger level. However, it is necessary to recall that the classification trees were used as a nowcast, while the forecasted danger level was issued on the afternoon of the preceding day.

3.3.3 Probability forecast

It was tested whether the classification trees of the "Best_20" could be used for a probabilistic forecast. Figure 3.3 shows the attribute diagram (Wilks, 1995), which relates the forecast probability to the observed relative frequency. A circle for each separate forecast determined with the terminal nodes of the trees for each cross-validation block is drawn. Some terminal nodes were more frequently used when the test dataset was applied. This frequency is expressed with the size of the circles. The reference to the categorical use of the trees (classification) is given as follows: the examples of the test dataset with a forecast probability larger than the prior probability would have been classified as event "1". In the detection of rather stable conditions (Fig. 3.1a) a dependency between forecast probability and observed relative frequency is recognisable. Frequently used nodes with larger forecast probability do show relatively more observations of the category "good". This is visualised with larger circles close to the perfect reliability line (iii). It can be seen that examples classified as "poor/fair" also more often contributed positively to the forecast skill; they are more often in the grey zone, while many examples, which would be classified as "good", are in the white zone thus contributing negatively to the forecast skill. This is consistent with the verification of the categorical forecast showing large FOCN values and intermediate FAR values. Only one very frequently used node showed perfect reliability.
Consistent with the observations for the categorical forecast, a lower quality was observed for the detection of rather unstable conditions (Fig. 3.3b); many circles lie outside the grey zone, hence contributing negatively to the forecast skill. Again, examples with low forecast probabilities contributed positively to the skill more frequently.

3.4 Discussion and conclusion

The first objective of this study was to determine if existing physically based stability estimates implemented in SNOWPACK were applicable for regional stability assessment. Based on the Fisher criterion, this is not the case for the stability index SSI, while for the Sk_{38} this was confirmed. Such a result seems plausible since the SSI uses the variables difference in hardness and grain size to adjust the Sk_{38}. These variables were implemented under the hypothesis that modelled variables equivalent to those identified as significant for observed stratigraphy would show similar strength of classification. However, this hypothesis now seems incorrect, since for the most important variables in observed stratigraphy an opposite sign of correlation to stability was found for modelled variables. Those variables were difference in hardness and grain size across the failure interface, failure

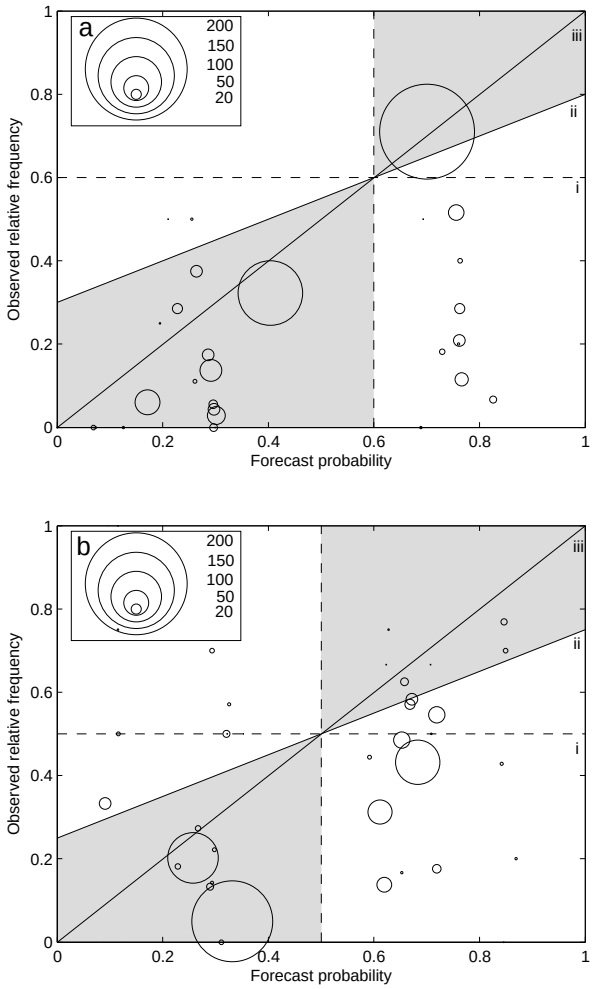

Figure 3.3: Attribute diagram for the verification of the probabilistic forecast of model "Best_20" for the detection of (a) rather stable and (b) rather unstable conditions. It relates the forecast probability of each node of the trees obtained from the training parts of the CV blocks and the observed relative frequency obtained from the test parts. The size of the circles shows how often the nodes were used by the test parts. Circles on line (i) have no resolution, which is plotted at the level of the prior probability for class "1". Circles on line (ii) indicates no skill, while line (iii) implies perfect reliability and skill. Circles in the grey zone contribute positively to forecast skill.

layer grain size and hardness. Nevertheless, for some variables an agreement between modelled and observed snow stratigraphy variables and their relation to stability was found , e.g. grain type and weak layer shear strength. Of particular importance is that other variables, which cannot be measured or were not considered so far (e.g. mean properties of the slab and differences of these to weak layer properties), did show a good strength of classification.

The majority of important variables were from the modelled snow stratigraphy, rather than the measured meteorological data (Table 3.4). This suggests that snowpack modelling delivers an additional benefit to the measured input variables when snow stability needs to be estimated, which is consistent with the findings of Schirmer et al. (2009a).

As the SSI was used for failure layer detection, it was based upon modelled differences in hardness and grain size. In contrast, the SSI was not able to distinguish between stable and unstable conditions, though it was constructed to both find a relevant weak layer and give a stability estimate (Schweizer et al., 2006). In fact, convincing failure layers and interfaces were detected, e.g. buried surface hoar, depth hoar and faceted crystals as well as crusts. Furthermore, modelled properties of these layers were important to distinguish between stable and unstable conditions, e.g. the Sk_{38} but not the SSI itself. Thus it seems that detecting a failure layer and estimating stability are different processes requiring different variables. In the case of the observed stratigraphy, this is confirmed by a study performed by Schweizer and Jamieson (2007), which recognised that a method developed for stability estimation can only be partially adapted for detecting potential failure layers.

Several forecast qualities, i.e. discrimination, reliability and skill (TSS = 0.5 and 0.4) were computed for the classification trees presented in section 3.3.2, as overall performance cannot be expressed by a single quality measure. The classification trees using the most important variables as selected by the Fisher criterion performed better than stability estimates already implemented in SNOWPACK. When periods with a verified regional danger level were tested the trees performed convincingly as only three in 20 cases the models failed. The probabilistic forecast provided by the classification trees can be used with limitations, at least for the detection of rather stable conditions a reliable forecast quality was observed.

Since separate classification trees were used to detect rather stable and rather unstable conditions, potentially conflicting tree results can occur (Table 3.7). Classification trees applied on the original three categories achieved lower quality characteristics, which was to be expected with the higher dimensionality of the classification problem. Furthermore, the two class problem simplified the validation task (as discussed in section 3.2.4). It is believed that the disadvantage of sometimes conflicting tree results are outweighed by the advantages of (i) a higher performance of the separate trees and (ii) the simplified validation. Additionally, conflicting results can be used as an indication of uncertainty in addition to the estimated probability delivered by the classification trees.

Consistency with expert knowledge was found: the variables independently selected for the model by experts were often rated as important with the Fisher criterion. Physical and logical explanations of the rules delivered by the classification trees were found.

Our models for stability evaluation could be used by avalanche warning services as nowcast or forecast, though only in regions with similar climatic characteristics as in the region of Davos. It is characterised by a transitional snow climate, with a maximum snow depth of two meters snow in

average, 2500 m above sea level. The region of Davos exhibits climatological differences at the five stations used. Since the classification trees were trained without location specific information for where the snow cover variables were generated, the climatological differences were integrated in the learning process; only those rules could be detected, which were valid for all five stations. Thus, it is believed that the methods presented can be applied to many regions without further modifications. However, for regions with other characteristics, a new classification tree must be trained. This is only possible if a similarly large amount of observations are available as was the case for this study. Nowcasting may already have an important value, since information on instability at the present day is rare. In a forecasting operation, the present snow cover would be simulated with measured data first and then the development of the snow cover would be predicted with forecasted meteorological data for the next day. Finally, the predicted snow cover would provide the additional input variables needed for the classification trees. The uncertainty of the forecasted input parameters and its effect on the classification trees was not assessed in this study.

4 Persistence in intra-annual snow depth distribution. Part I: measurements and topographic control

4.1 Introduction

Snow depth distribution in complex Alpine terrain is characterised by a large spatial variability. For example, snow depths across a ridge, compared to a flat field, can be increased by several times in leeward slopes, be less than the flat field case on the upwind slope and might be zero on the ridge line if winds are sustained. This spatial variation can be seen clearly by eye in Alpine terrain, but relatively little quantitative area-wide data exists over larger scales with a few exceptions, which are restricted to maximum snow depth assessment, however (e.g Deems et al., 2006; Trujillo et al., 2007), or are restricted to point measurements (e.g Elder et al., 1991; Anderton et al., 2004; Machguth et al., 2006) or aerial photographs (e.g Blöschl and Kirnbauer, 1992; Winstral and Marks, 2002). The amount of water that is held in the snowpack may vary significantly around drainage basins at the onset of melt, and so modelling snowmelt and runoff will miss essential factors if the variability of snow depth is not included (e.g. Winstral and Marks, 2002). The variability of snow depth over shorter timescales is also of interest; variation in the depth of deposited and redistributed snow after one precipitation event strongly influences avalanche danger (Schweizer et al., 2003a). Similarly, the variability of (new) snow depth in the starting and runout zones of avalanches has been shown to have a considerable influence on calculated runout lengths (Sovilla et al., 2006, 2007), and including this information in avalanche dynamics models would increase their accuracy.

Snow depth distribution in Alpine terrain is influenced by changes in precipitation amounts with elevation and modified by complex wind flows. Physically-based modelling of snow redistribution in Alpine terrain would therefore require that processes, such as saltation, suspension and sublimation, were resolved as well as complex wind fields, requiring considerable computation effort (e.g. Gauer, 2001; Liston et al., 2007; Lehning et al., 2008). While it would be possible to model these processes and the feedback between them, there is a lack of verification snow depth data at the same horizontal scale and with the same height resolution as can be achieved with models.

Because of these difficulties and because of the obvious linkage of processes to terrain, a frequent approach has been to describe snow depth with terrain parameters as a measure of wind shelter or exposure (e.g. Purves et al., 1998; Winstral et al., 2002; Anderton et al., 2004). Based on the terrain parameter developed by Winstral et al. (2002), Winstral and Marks (2002) modelled snow depth quantitatively in moderately steep terrain. However, they could only verify modelled snow

depth indirectly, using aerial photographs of snow covered area during the melting period.
Erickson et al. (2005) observed persistence in topographic controls on the spatial distribution of snow in rugged terrain between multiple years. They found that the terrain parameter developed by Winstral et al. (2002) had the greatest influence on predicted snow depth and was – as potential solar radiation – found to be statistical significant for all the individual years that they investigated. Contrary to that finding, Hodgkins et al. (2005) found a significant variability between two years in both magnitude and spatial distribution of winter accumulation on a glacier in Svalbard. Their results were based on a principal component analysis of simple terrain parameters such as altitude, slope and aspect. Their analysis did not include an index of wind sheltering as developed by Winstral et al. (2002). Furthermore, the limited number of manual measurement points of only 106 and 75 in each respective year may lead to a simplified conclusion. Deems et al. (2008) used high resolution airborne LiDAR data in two different years and in two different areas. They obtained a large inter-annual consistency of snow depth at the end of the accumulation season in many characteristics.
In our study we show how the consistency at the end of the accumulation season is formed by individual storm periods. Motivated by the observed persistence of typical events and its obvious linkage to terrain, we verified a simple model approach based on the terrain parameter developed by Winstral et al. (2002): we demonstrated for the first time with high resolution snow depth data to which extend such a simple model approach is able to reproduce amount, variability and spatial structure of snow depth in complex terrain.
In this paper we present a unique set of snow depth data with a resolution of 1 m obtained using terrestrial LiDAR. This data spans two years and several dozen precipitation events. We tested how deposition patterns can be related to terrain. Finally we compare the amount, variability and spatial structure of snow depth to a terrain parameter proposed by Winstral et al. (2002).

4.2 Methods

4.2.1 Field description and data acquisition

In order to measure snow depth variability in Alpine terrain before and after individual storm or snowfall events, a study area was chosen where the disturbance by skiers can be neglected (Figure 4.1). To assess snow depth distribution at many specific dates during a snow season, airborne laser scanning was considered impractical because of long term planning and high costs. A terrestrial laser scanner was used instead (Riegl LPM-321), for which antecessors were suited to measure snow depth distribution (e.g. Prokop et al., 2008; Schaffhauser et al., 2008) . Prokop et al. (2008) achieved an error estimation of maximum ± 8 cm in comparison to a tachymeter survey with a standard deviation of about 2 cm at a distance of 300 m. Comparisons performed in our study between tachymeter and our TLS device have shown similar results (Grünewald et al., 2010).
Laser scanner, tripod and energy supply was stored in a shelter in the study area. 16 permanent reflectors were mounted in the area of interest covering many vertical and horizontal angles to the measurement positions of the laser scanner. Enough reference points for all measurement positions

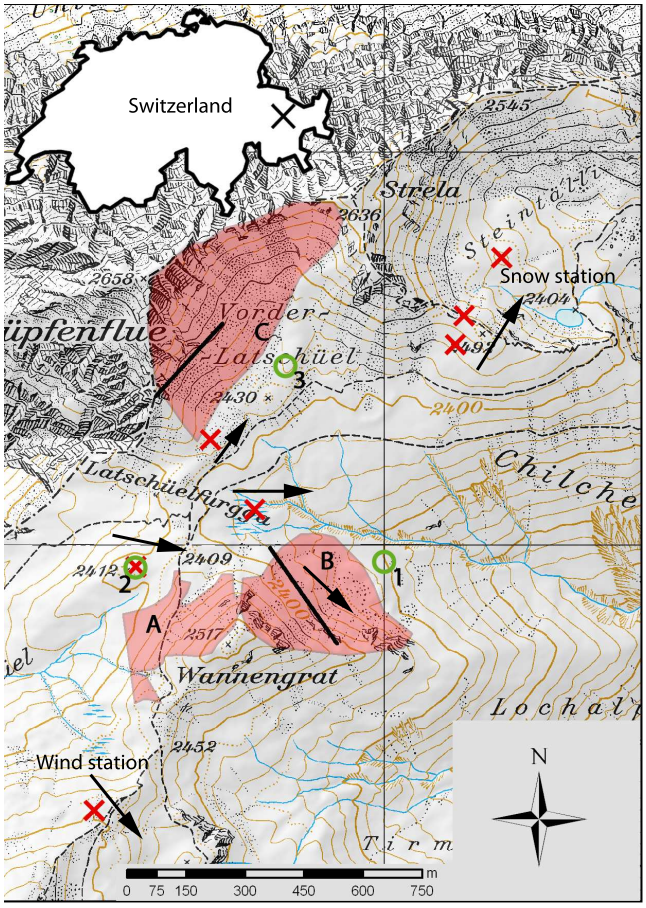

Figure 4.1: Overview of the study area. Shaded (in red) are the three investigated slopes: Under typical NW storms (A) is the windward slope, (B) a cross-loaded slope and (C) a lee slope. The measurement positions of the terrestrial laser scanner are plotted with green circles. Red crosses show the position of seven weather stations. Black arrows show typical wind directions for NW storms situations (including "wind station" and "snow station" used for the model). The direction of the arrow in the cross-loaded slope is deduced from surface structures as sastrugi and dunes. Two black lines mark the position of the two transects shown in Figure 4.7.

are a prerequisite to register the point clouds into a global coordinate system. Measurement positions were chosen to optimise laser scanning requirements such as low angle of incidence, distance to the areas of interest and avalanche safety (Figure 4.1). Areas of interest are marked and the name selection is based on the situation for typical NW storms: a windward slope (A), a cross-loaded slope (B) and a lee slope (C). A measurement campaign took up to three successive days with a horizontal and vertical scanning resolution of 0.06°. The quality of snow storm forecast and weather conditions optimal for scanning did not always allow a two day campaign before and after storms. At least the measurement point 1 covering slope B and C was used (Figure 4.1). Data from measurement point 3 was only used in this study for quality verification having less distance to slope C.

The study area is equipped with seven weather stations. Detailed analysis of the wind stations together with an interpretation of the orthophotos delivered by a digital camera installed next to the laser scanner, which show wind drifts such as dunes, cornices and sastrugies to serve as a proxy for wind directions, will be presented elsewhere (Mott et al., manuscript in preparation, 2010). The weather stations revealed typical storm events: a south storm, often without precipitation but sometimes followed by homogeneous snowfall directly after the storm and a NW storm contributing the major part of the precipitation during a winter season. The wind directions measured at each wind station showed typical patterns for SE and for NW storms (for NW storms see Figure 4.1). The differences in wind direction between the stations are remarkable. The consistent orientations of wind drifts observed in the cross-loaded slope allowed us to determine a typical wind direction there for the NW storm event.

The dataset obtained with the laser scanner includes the summer terrain and snow depth development during the accumulation period 2008/09 including nine snow fall events. Additionally, the dataset includes the maximum snow depth situation in 2007/08 and the first snowfall in 2009/10. The point cloud was triangulated with a limited edge length of 5 m to avoid large data gaps in blind areas to be closed. A regular grid with 1 m resolution was achieved using a linear query of the triangulated surface. Grids for snow depth and snow depth change have been obtained from the regular surface maps (with and without snow) by simply subtracting an earlier surface from a current one.

The temporal evolution of snow depth and snow depth change was quantified with simple statistics such as mean and standard deviation and additionally visualised with snow-depth transects in two sub-areas. Pearson's linear correlation coefficient was computed to quantify the observed persistence of inter- and intra-annual snow depth distribution.

4.2.2 Modelling snow depth with terrain variables

The observed terrain-wind speed interaction motivates the use of a DEM interpretation developed by Winstral et al. (2002). They used a measure of topographic shelter or exposure $\overline{Sx}$ of a grid point to prevailing winds. Firstly, the parameter Sx for a grid point (x_i, y_i) is calculated with

$$Sx_{A,dmax}(x_i, y_i) = max\left(\tan^{-1}\left\{\frac{ELEV(x_v, y_v) - ELEV(x_i, y_i)}{[(x_v - x_i)^2 + (y_v - y_i)^2]^{1/2}}\right\}\right) \tag{4.1}$$

where A is the azimuth of the search direction, $dmax$ the search distance and (x_v, y_v) are the set of all grid points along the line defined with A and $dmax$. Secondly, the Sx value is averaged over several azimuth values A across an upwind window (A_1, A_2) around a prevailing wind direction $\overline{A}$

$$\overline{Sx}_{\overline{A},dmax}(x_i, y_i) = \frac{1}{N} \sum_{A=A_1}^{A_2} Sx_{A,dmax}(x_i, y_i) \tag{4.2}$$

calculated with 5° increments. A 10 m resolution DEM was used for those calculations, which is equal to the lowest resolution used in the original studies of Winstral et al. (2002) and Winstral and Marks (2002). Similarly to Winstral and Marks (2002) a 30° upwind window was used. For $dmax$ a value of 300 m was chosen, which qualitatively delivered best results for measured wind speed relations at sheltered and exposed weather stations.

For a quantitative interpretation of the $\overline{Sx}$ parameter into snow depth distribution, a model approach proposed by Winstral and Marks (2002) was used. In this approach $\overline{Sx}$ fields were applied to receive wind speed and precipitation fields with interpolation between an wind-exposed and wind-sheltered weather station. Subsequently these fields were given to an energy an mass balance model as input. In this study, $\overline{Sx}$ parameter fields only for two wind directions, SE and NW, were considered. These directions correspond to the two peaks of wind direction distribution observed at the wind station (Figure 4.1). For wind speed, measurements of an exposed wind station and a sheltered snow station were used as interpolation points. For precipitation, measurements at wind exposed weather stations are not available in the Swiss Alps. Thus, precipitation measurements were only used at the snow station. As second interpolation point a grid cell in the study area with large erosion during a NW storm was chosen, which will receive no deposition in our model. Similarly, a different point was determined for SE storms. These points can easily be determined by eye in the field. At the snow station, the precipitation amount was calculated with a one-dimensional SNOWPACK simulation (Lehning et al., 1999, 2008) using the weather station as input. SNOWPACK is needed to take the settling during snowfall events into account. Both wind and precipitation fields were limited to the values obtained at the interpolation points.

Wind direction was assumed to be constant for the whole study area, in spite of the observed spatial differences in wind direction for typical SE and NW storm situations. At least for the three slopes investigated (A-C), the chosen wind directions appeared to be reasonable. In section 4.3.5, we discuss possible implications and solutions. In contrast to the original work of Winstral and Marks (2002), the precipitation and wind speed fields were not adjusted with the evolving modelled surface due to snowfall, because terrain parameter fields calculated from measured surface at maximum snow depth situation did not increase the correlation to corresponding measurements compared to those calculated from the summer surface. Winstral et al. (2002) developed a second parameter called drift delineator, which was used to delineate sites of intensive redistribution on lee slopes. As we already obtained too much modelled snow in our lee slope (C) determined only with the $\overline{Sx}$ parameter (see below), we did not use this additional parameter.

The derived wind speed and precipitation fields were used as input for the distributed version of SNOWPACK, Alpine3D (Lehning et al., 2006). Other input parameters were obtained from measurements at the snow station and were assumed to be constant in space (relative humidity)

Table 4.1: Mean μ and standard deviation σ calculated for snow depth at the end of the accumulation season

	2007/08 [m]		2008/09 [m]	
Area	μ	σ	μ	σ
Windward slope (A)	No data	No data	1.0	0.8
Cross-loaded slope (B)	2.7	1.7	2.2	1.8
Lee slope (C)	2.7	0.7	2.6	0.6

or distributed with a fixed altitude gradient (air temperature and longwave radiation). In brief, potential shortwave radiation was modelled in an hourly resolution and the diffuse radiation was assessed with the difference between modelled and measured shortwave radiation at the snow station (Helbig et al., 2009). Shading caused by terrain but no other lateral terrain effects as reflected shortwave radiation were considered, since in our test cases they have turned out to be irrelevant for modelling snow distribution.
The model time step was chosen to be one hour. For each time step wind speed and precipitation field were calculated based on the description above. For the discrimination between both wind directions (NW and SE) the measurements of the wind station were used.

4.3 Results

4.3.1 Overview of observations

The snow depth distribution at the end of the accumulation period of two successive winters 2007/08 and 2008/09 is shown in Figure 4.2. Both dates correspond very well with the date of maximum snow depth (HS_{max} situation) at the snow station, although one month later observed in 2007/08. The snow depth distribution in the three slopes investigated showed different characteristics, which can consistently be observed in both winters: a homogeneous distribution in the lee slope (C), an inhomogeneous distribution in the cross-loaded slope (B) and less snow in the the windward slope (A). This impression is confirmed regarding simple statistics as mean (μ) and standard deviation (σ) as summarised in Table 4.1.
In more detail, the cross-loaded slope (B) was dominated by two accumulation zones with snow depth over 6 m, while mostly no snow between these zones and in the upper area was observed. These accumulation zones developed behind two cross-slope ridges of summer terrain, which are orientated mostly normal to assumed wind direction from NW. In the lee slope, maxima of snow depth were found in thin parallel bands orientated along the largest relief, which corresponds to depressions in summer terrain. The mean snow depth was very similar to the snow depth measured at the sheltered snow station. In the third slope investigated, the windward slope (A), maxima of

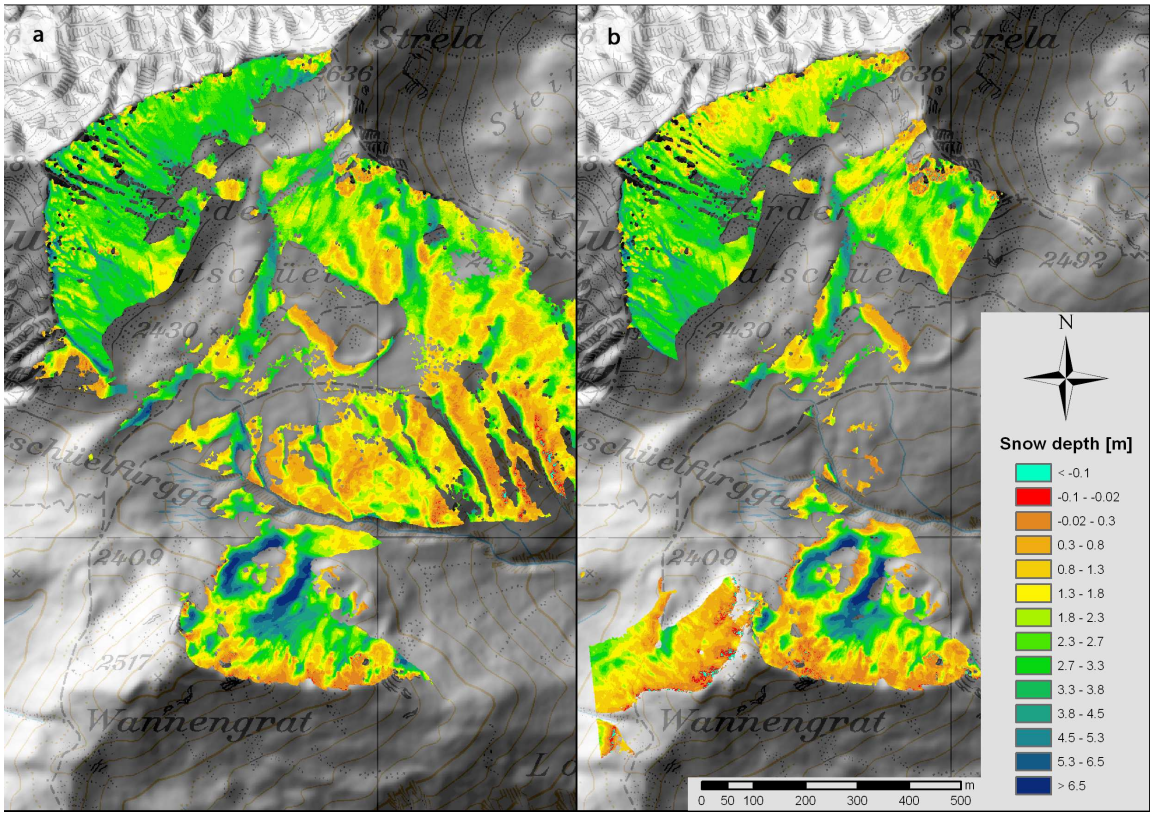

Figure 4.2: Snow depth distribution at the end of the accumulation period in two winters, (a) 26.04.2008 and (b) 27.03.2009.

Table 4.2: Mean μ and standard deviation σ calculated for snow depth change due to the NW storm shown in Figure 4.3d

	17.03.2009 [m]	
Area	μ	σ
Windward slope (A)	-0.04	0.12
Cross-loaded slope (B)	0.17	0.32
Lee slope (C)	0.63	0.32

snow depth were found in the lower part of the slope, where wind speed should be less than in the upper part, and in a depression next to the marked hiking trail.
The snow depth distributions presented in Figure 4.2 suggest a large interannual consistency between both winters. However, the years differed in average snow amount (see also Table 4.1). In the cross-loaded slope half a meter more snow on average was deposited in the first year, while the variance was comparable. The obvious differences in snow depth distribution in the lee slope were caused by a large avalanche in the second year.
Snow depth changes (dHS) caused by individual NW storms are shown in Figure 4.3. These events produce deposition patterns comparable to each other and to the patterns discussed for the HS_{max} situation: a variable cross-loaded slope, in which the previously mentioned accumulation zones are noticeable, homogeneous loading in the lee slope and erosion in the windward slope (Table 4.2). In the upper parts of the cross-loaded and in most parts of the windward slope erosion dominated during NW storms. Snow was deposited in these areas during precipitation periods of different characteristics, which was eroded during NW storms. The snow depth loss in the lee slope of the period shown in Figure 4.3b was the result of the large avalanche mentioned previously with a starting zone width of 600 m. The homogeneous loading of that slope during NW storms might favour the existence of a widespread failure layer, which allowed fracture propagation over such large distances.
The similarity of NW storms and the HS_{max} distribution for both years indicates that the distribution at the end of the accumulation period is dominantly influenced by one typical event. The processes during these NW storms seem to be the major driving factors influencing the snow depth distribution at the end of the accumulation season. Consistency between years was also recognised by Deems et al. (2008) for two study areas. They concluded that these major driving factors are consistent between years. Moreover, the similarity of typical NW storms indicates that processes influencing snow distribution remain similar throughout an accumulation season. Differences in wind distribution, changing snowpack properties and altered surface structure due to snow deposition seem to be underpart.
In addition to typical NW storms presented in Figure 4.3, events with other characteristics could be

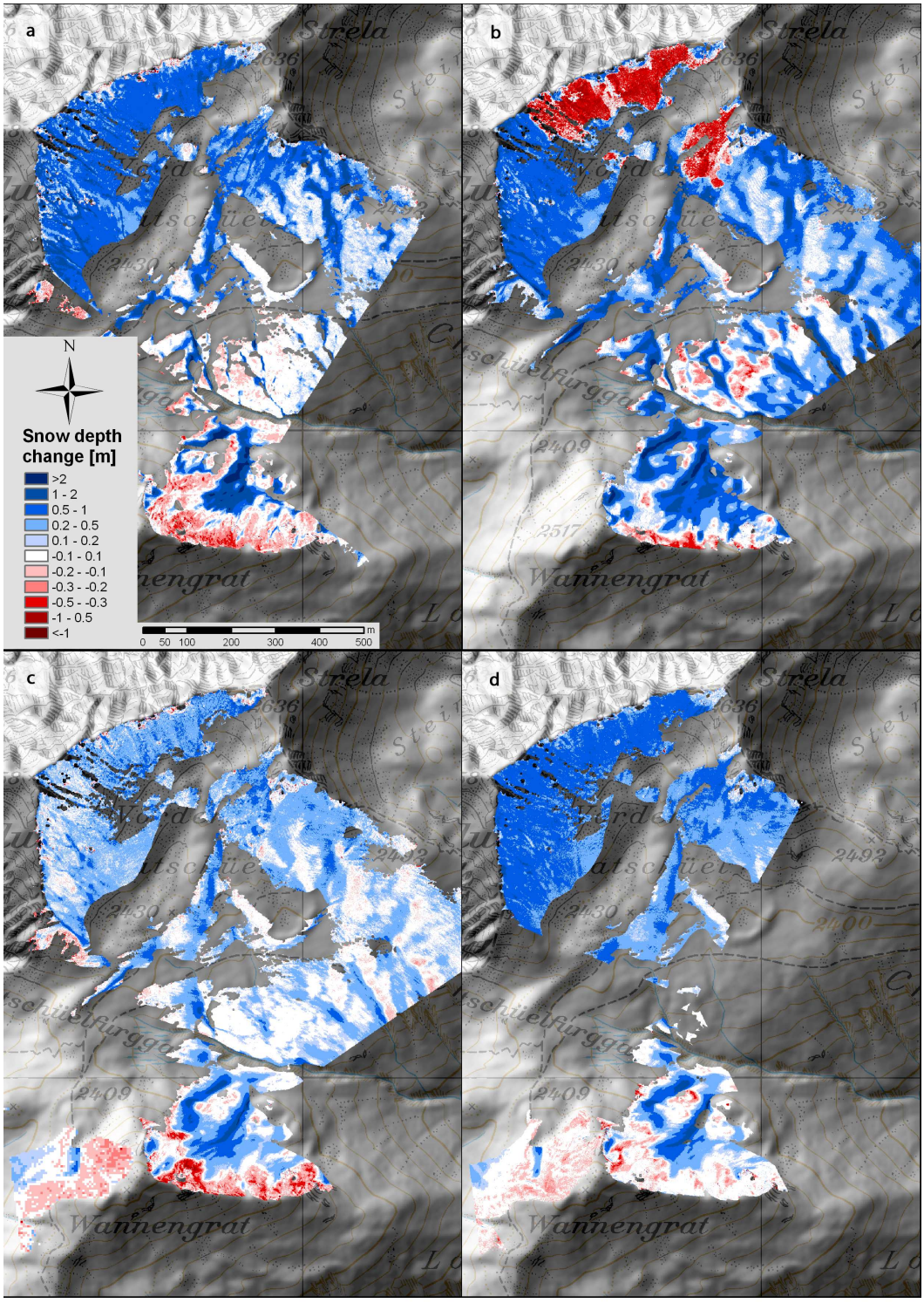

Figure 4.3: Snow depth change of NW storm periods between (a) 20.11.2008 and 27.11.2008, (b) 04.02.2009 and 04.03.2009, (c) 04.03.2009 and 17.03.2009, (d) 17.03.2009 and 27.03.2009.

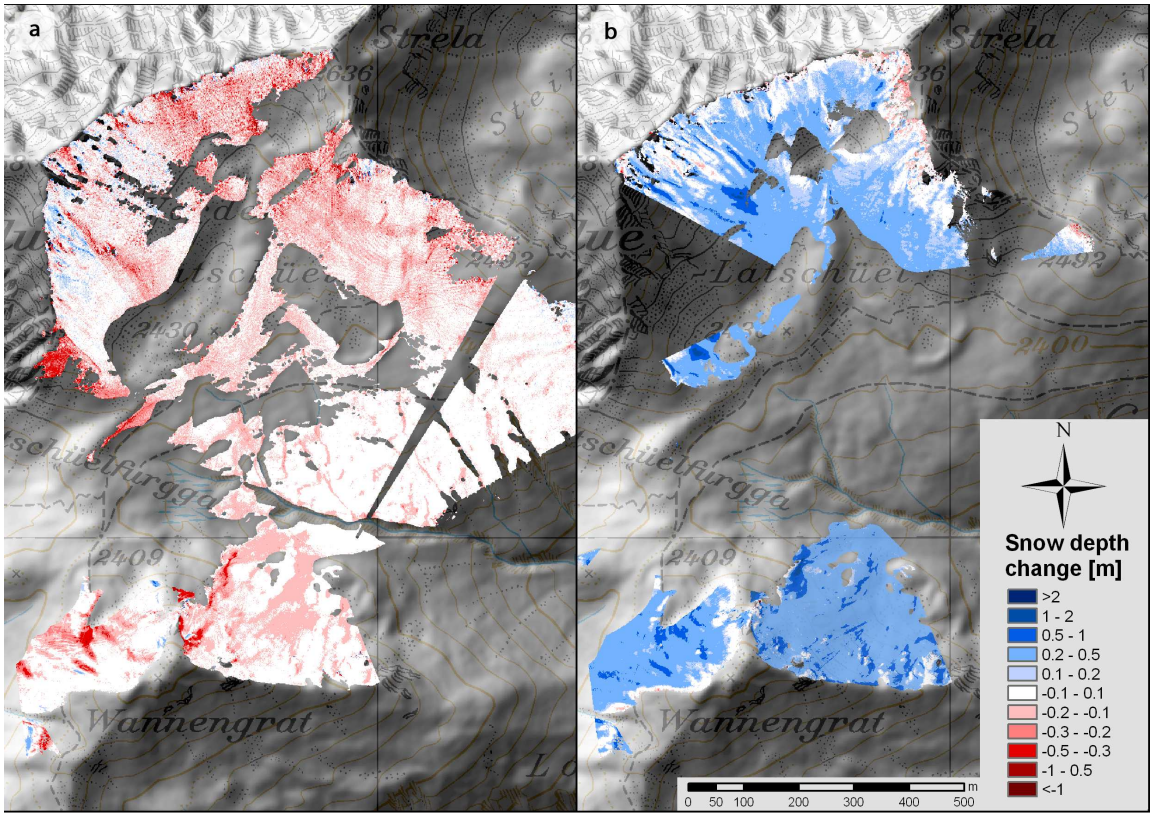

Figure 4.4: Snow depth change of (a) a SE storm between 28.01.2009 and 04.02.2009, and of (b) a homogeneous snow fall with low wind speeds between 14.01.2009 and 22.01.2009.

observed during the accumulation season 2008/09. Figure 4.4 shows two of those events, (a) a SE storm without precipitation and (b) a snowfall during low wind speed conditions. During the SE storm snow depth changes were small and close to the accuracy of the laser scanner. Most erosion was observed in the slope (A), which was consistent with observations of snow drifts as dunes and sastrugi only in this slope. These observations were confirmed by measurements of another SE storm (not shown). During low wind speed conditions (Figure 4.4b), a rather homogeneous snow distribution was observed in comparison to NW storms.

Figure 4.5 shows additional measurements restricted to the cross-loaded slope. In (a) the summer terrain is presented and in (b) and (c) the first snowfall events in two different years. In both cases the large influence of the rough summer terrain is recognisable. However, both events show a very different behaviour. While the first snowfall in 2008/09 (c) is an example for homogeneous loading during low wind speed conditions, the first snowfall in 2009/10 (b) revealed characteristics of a NW storm. The period shown in Figure 4.5e seems to be largely negatively correlated to observations during typical NW storms and to the preceding period shown in (d). As discussed later, possible explanations for the negative values in snow depth change are erosion, sublimation or settling.

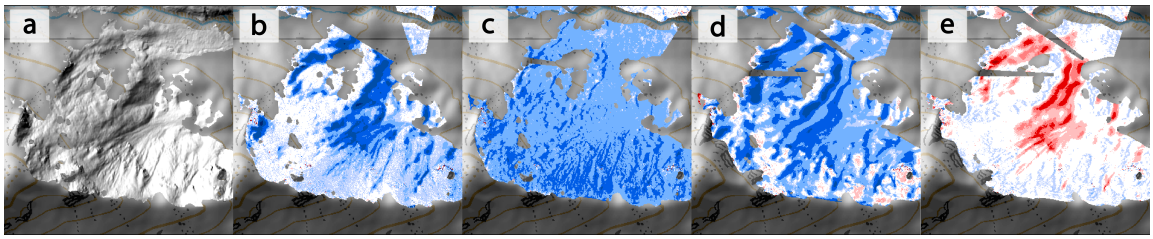

Figure 4.5: (a) Hillshade picture of summer terrain in cross-loaded slope, (b) first snow fall in winter 2009/2010, (c) first snow fall in winter 2008/2009, snow depth change between (d) 27.11.2008 and 23.12.2008, and between (e) 23.12.2008 and 14.01.2009. Same colour coding as in Figure 4.3 and 4.4.

4.3.2 Temporal evolution of snow depth and snow depth change

The evolution of variance and mean of snow depth in the cross-loaded slope (C) is shown in Figure 4.6a in the course of an accumulation season. The variance of snow depth increased during the winter. Significant increase appeared only after NW storms, which are marked with black arrows. A similar development was observed in both other areas investigated (not shown). On the contrary, the variance of snow depth change for NW storms (Figure 4.6b, also marked with black arrows) was decreasing. This finding can be interpreted as a time development reacting on a smoothing surface, or with decreasing amount of snowfall as measured at the snow station. Homogeneous snowfalls associated with low wind speeds (marked with stars in Figure 4.6b) are characterised by low variance and large means. The large mean value of the second NW storm can be explained with the fact that snowfalls with low wind speeds were also included in this snow depth change period, which could be observed at the weather stations. More discussion on the temporal evolutions of structure characteristics will be given in Part II with fractal parameters, which delivers more detailed information as basis for interpretations.

4.3.3 Description of transects

In Figure 4.7 transects showing the temporal evolution of the snow cover are presented for the cross-loaded and lee slopes. The transect locations are plotted in Figure 4.1. In the cross-loaded slope (Figure 4.7a and b, wind direction from the left) the pronounced accumulation zones developed behind ridges. Maxima (of elevation and snow depth) were moving in wind direction between 10 and 20 m. On top of the ridges mostly no snow was deposited. Surface smoothing compared to the summer surface can be observed, but also an increase in slope behind the ridges. Snow accumulation was not able to fill the depression behind the ridges totally during one winter season. This is a prerequisite of similar snow depth change caused by individual NW storms. The increase of the snow depth variance with time discussed previously is visible as well as the decrease of the variance of the snow depth change.
In the lee slope transect (Figure 4.7c and d) a similar large scale depression in summer terrain is

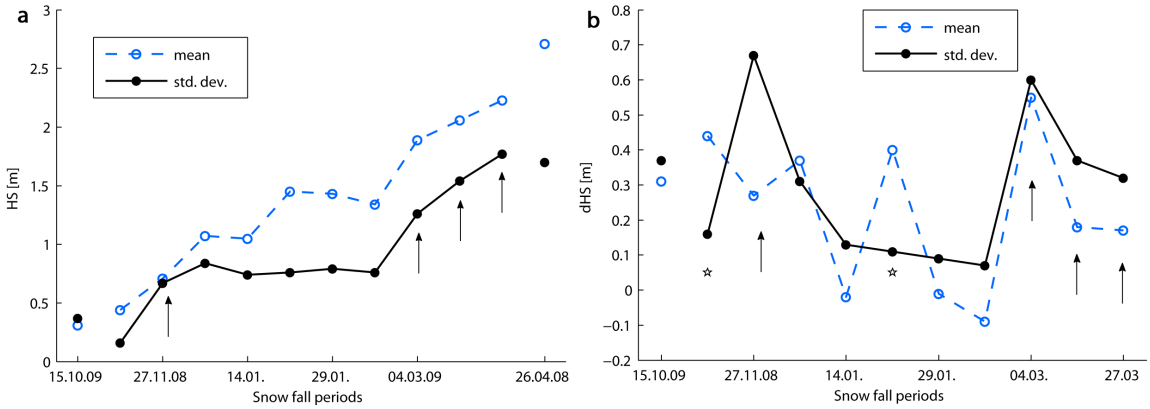

Figure 4.6: (a) Development of mean and standard deviation for snow depth (HS) in cross-loaded slope. For comparison, the first snowfall in 2009/10 and the maximum HS in 2008 are plotted without connective lines. Mean and standard deviation of (b) the snow depth change (dHS) determined from single periods. Labelled is the date at the end of a period. NW storms are marked with an arrow and homogeneous loading with a star.

existent in the transect. However, the low wind speeds predominant in this slope formed no pronounced accumulation zones. Rather shallow snow depths emerged at the depression sides. Maxima of snow depth were located at small scale depressions in summer terrain. Some of them became leveled while others became even more pronounced. Even more obvious as in the cross-loaded slope are the "parallel" snow depth curves indicating a decrease in variance of snow depth change with time, while variance of snow depth was increasing. A similarity to the cross-loaded slope is that the characteristics of snow depth after the first storm influenced strongly the characteristics at the end of the accumulation season.

These results suggest that the interaction between summer terrain and snow deposition seems to act in (at least) two scales: First, wind speed is influenced at a large scale, which is the average aspect of a slope compared to prevailing wind direction. At this scale a slope is defined in windward, lee or cross-loaded characteristics. Second, if wind speeds are large in a slope, in-slope terrain structures become dominant for the distribution of snow depth. Examples are the two accumulation zones behind ridges orientated normal to the wind direction modifying terrain depressions (tens of meters). If wind speeds are low, the effect of small-scale in-slope structure seems to be reduced to the filling of depressions (meters). Examples are the snow depth maxima in the lee slope, which correspond well to such small scale depressions. How this picture can be confirmed with fractal analysis will be discussed in Part II.

4.3.4 Quantitative analysis of inter- and intra-annual consistency

Inter-annual consistency in snow depth distribution was quantified with Pearson's correlation coefficient r. For the cross-loaded slope an amazing large correlation between the HS_{max} situation

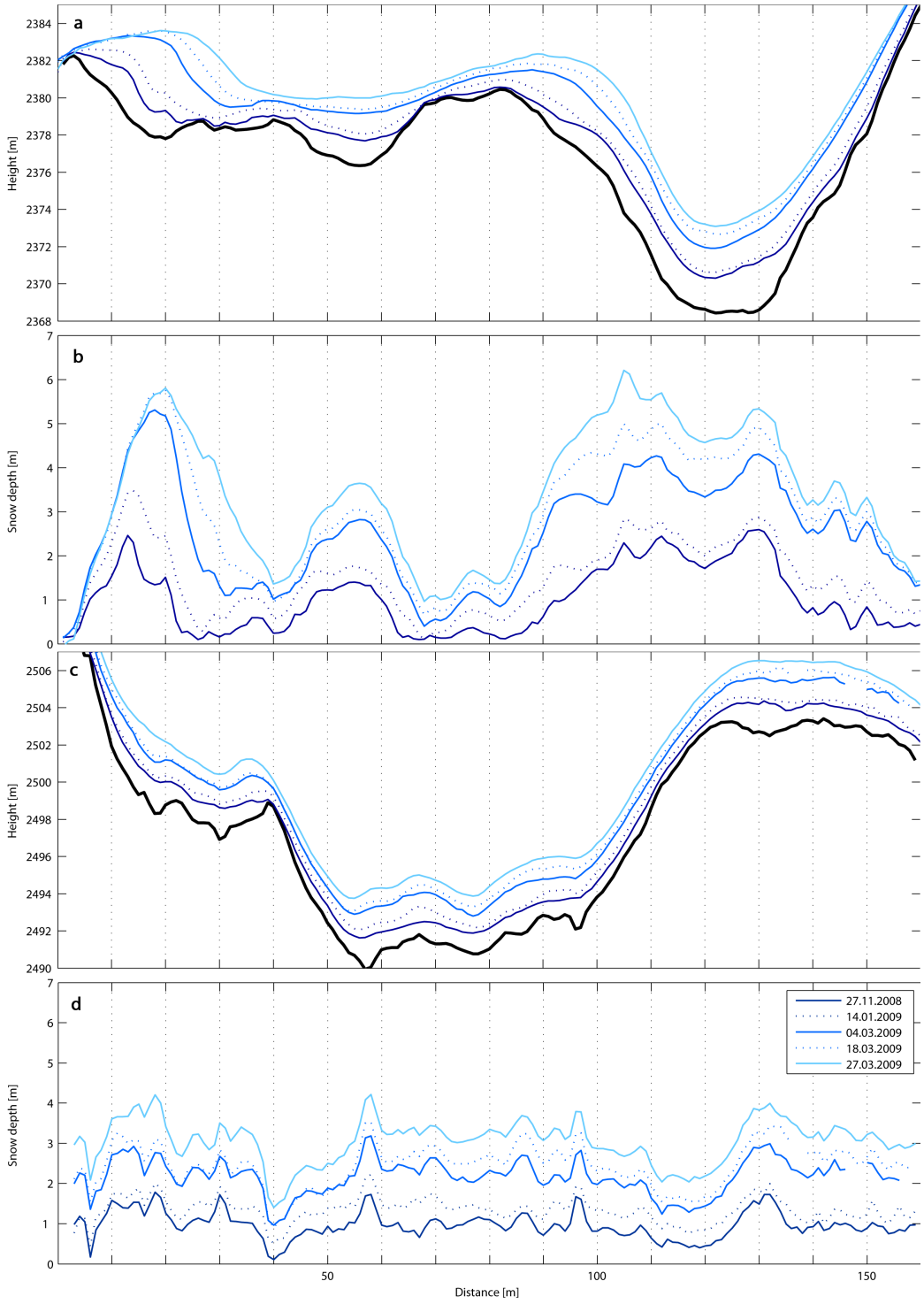

Figure 4.7: Transects of z-values (black line is the summer surface, blue lines are with snow for a selection of time steps) and of snow depth, (a) and (b) in cross-loaded slope and (c) and (d) in lee slope.

of the two winters of $r = 0.97$ was achieved, similarly in the lee slope (restricted to non-avalanche affected area) where a coefficient of $r = 0.93$ was obtained. These results confirm findings of Deems et al. (2008), who reported on a large consistency between two winters with correlation coefficients of $r = 0.88$ for a study area with low rolling topography and $r = 0.92$ for moderate topography. In both areas studied by Deems et al. (2008) vegetation seems to play an important role since both are at least partly characterised by coniferous forests. In our study area terrain is more complex while vegetation is limited to short meadow, and its influence can be neglected. Due to the larger correlations obtained in our study area, and due to the larger correlations in the area described as "moderate" in relation to "low rolling" in the study of Deems et al. (2008), we may formulate the hypothesis of increasing inter-annual consistency of snow depth distribution at the end of the accumulation season with increasing complexity of terrain. This hypothesis is supported by: (i) the larger the complexity of large scale terrain, diverting wind directions might be more affected by terrain in relation to synoptic situations and (ii) if in-slope complexity is large enough that snow depth cannot level all summer terrain depressions (as described previously for our study area), a larger connection to terrain seems plausible. On the other hand, the data basis is still too weak (two winters in a rather small area) to draw final conclusions.

Referring now to snow depth change, in the cross-loaded slope correlation coefficients between the NW storms in 2008/09 winter season ranged between 0.53 and 0.73. These results quantify the observation of intra-seasonal consistency mentioned previously. However, the larger correlation between the HS_{max} situation of two years shows that this inter-season consistency was formed by less similar individual storms. The first storm correlated better with the HS_{max} situations ($r = 0.88$) than the last storm ($r = 0.76$), as already mentioned describing the transects. This first storm was one of the largest snowfall events in this year, which can explain the large influence on the snow depth distribution at the end of the accumulation season. But it might also be the case that a sum of three smaller events converges to a similar distribution, since other effects than NW storms seem not as dominant. The influence of summer terrain may be responsible for the large correlation ($r = 0.77$) of the both first NW storms in 2008/09 (Figure 4.3a) and in 2009/10 (Figure 4.5b). The first and last NW storm in 2008/09 were, in comparison, less correlated ($r = 0.4$).

The negative correlation found in the cross-loaded slope between the situation described in Figure 4.5e and most other NW storms was most pronounced in comparison to the snow depth distribution at the beginning of that period (r = -0.83), but also to the snow depth change of the two periods before (Figure 4.5d and 4.3a), with correlation coefficients of -0.64 and -0.76 respectively. Low wind speeds during the whole period and only little snowfall amount observed at the snow station lead us to believe that settling was the most important process for the negative values in snow depth change. Large snow masses, especially when fresh, will experience more total settling. In the 22 days during that period a change of snow depth of -12 cm was measured at the snow station, while in the cross-loaded slope a maximum of -50 cm could be observed in the same period. In the lee slope, the NW storms did not show large intra-annual correlation (-0.22$< r <$0.4), although the area affected by the avalanche was excluded from the analysis. Despite the low correlation, these storms were persistent during the accumulation season in terms of showing always a homogeneous loading. As already discussed previously, the characteristics of the first NW storms

were dominant with a correlation of 0.76 for the HS_{max} situation while the last storm reached only a value of 0.35. Thus, a similar but even more pronounced conclusion can be drawn when the lee slope: the large inter-annual consistency of the HS_{max} situations was formed by not as consistent individual storms. Furthermore, the characteristics of the first storm were largely preserved through the entire accumulation season.
In the windward slope the largest correlation between snow storms was negative (r = -0.6). It was observed between the periods described in Figure 4.4a and b, which were a SE storm and a rather homogeneous snowfall during low wind speeds. The correspondence between maxima of snow depth loss and snow depth gain between both periods may not only be explained with settling as argued previously for the period in the cross-loaded slope shown in Figure 4.5e. Erosion appears to have been dominant in the areas of largest snow depth loss. The interesting question remains why erosion is largest, where accumulation during a snowfall with low wind speeds is largest? Are sheltered regions exposed to wind when the wind direction is changing? Since the SE storm appeared right after the rather homogeneous loading during low wind speeds, one possible explanation is that more snow could be eroded at those places at which previously more and maybe less bounded snow was deposited.

4.3.5 Modelling snow depth with terrain variables

In Figure 4.8a the $\overline{Sx}$ values are shown for a wind direction of 315°. The $\overline{Sx}$ parameter characterises the local reduction in wind speeds, and is – as a result – a qualitative measure of snow depth (Winstral et al., 2002; Dadic et al., 2010). At first glance, a remarkable agreement with the measured HS_{max} situation or with individual NW storms is visible. The pronounced accumulation zones in the cross-loaded slope were well described. Furthermore, main characteristics such as a variable cross-loading, a homogeneous loading in the lee slope, and less snow in the windward slope were reproduced (Table 4.3). Especially for the cross-loaded slope a good correlation (r = 0.63) could be achieved with snow depth at the end of the accumulation season. Not as convincingly explained is the in-slope variance of the windward slope with a correlation coefficient of only 0.45. Correlation is even worse in the lee slope. As discussed using the transect (Figure 4.7), much of the variance in the lee slope seems to result from small scale terrain effects, which cannot be captured by a 10 m resolution DEM and thus not by the terrain parameter.
The correlations discussed above are somewhat better than the values reported in the original work by Winstral et al. (2002), who obtained values for correlation coefficients less than 0.44 in comparison to snow probe measurements. The better values in our study can be explained with the higher resolution of the measurements, which fits better to the scaling properties of snow depth structure than sparse punctual snow probing (Deems et al., 2006; Trujillo et al., 2007, Part II). Additionally, the smaller extend of the slopes investigated in comparison to Winstral et al. (2002) might result in less spatially diverting wind directions. Hence better correlation with a parameter assuming spatially constant wind directions seems plausible. The topic of diverting wind directions is discussed below in more detail.

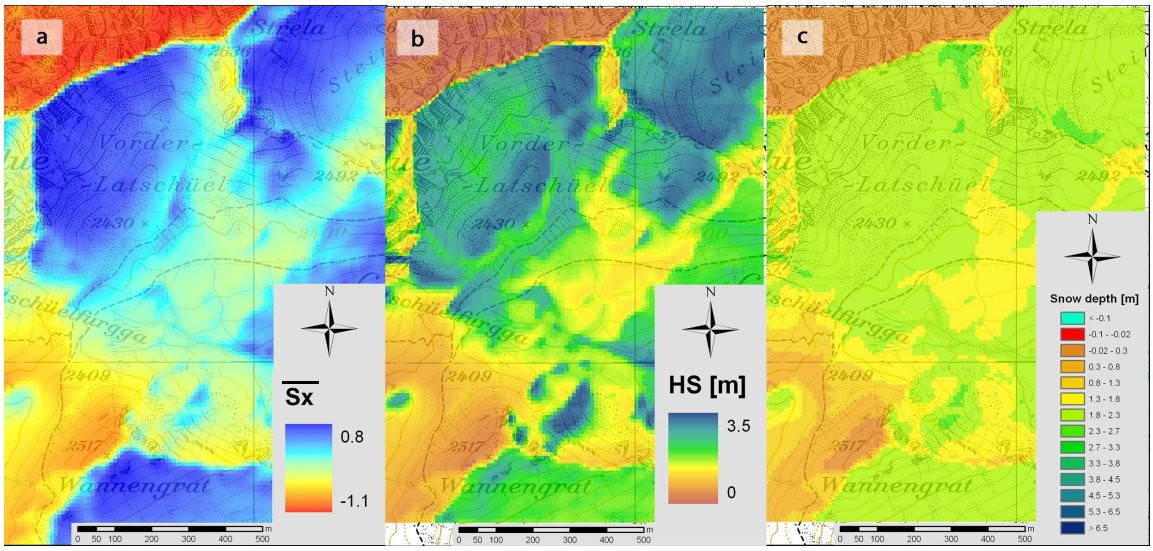

Figure 4.8: (a) Terrain parameter $\overline{Sx}$ with NW wind direction and 300 m separation distance and 10 m resolution, (b) modelled snow depth with Alpine3D using two $\overline{Sx}$ parameters to distribute wind and precipitation, (c) same as (b) but with the same colour scale as used in Figure 4.1.

Table 4.3: Mean μ and standard deviation σ calculated for $\overline{Sx}$ and for modelled snow depth (HS) at the end of the accumulation season

	$\overline{Sx}$		modelled HS [m]	
Area	μ	σ	μ	σ
Windward slope (A)	-0.37	0.13	0.6	0.3
Cross-loaded slope (B)	0.12	0.15	1.8	0.3
Lee slope (C)	0.64	0.07	2.0	0.04

In order to interpret the qualitative measure of wind speed as quantitative snow depth information, distributed wind and precipitation information derived with the terrain parameter $\overline{Sx}$ were implemented in Alpine3D. Before discussing results some general remarks on the possibilities and limitations of the model approach described in the section 4.2.2 will be given: No erosion is possible. This model approach only allows distributed precipitation. Furthermore, modelled homogeneous snowfalls are not possible during low wind speeds, since we did not use a second precipitation measurement at a wind exposed station. As mentioned previously a grid point which will never receive snow during NW storms was used as a second interpolation point. Hence the observed interaction of erosion during NW storms and deposition during low wind speed situations cannot be reproduced. The proposed model approach is more than a one-to-one translation of the terrain parameter field shown in Figure 4.8a. Additional processes potentially influencing the snow distribution at the end of the accumulation season are included. Such are spatially distributed solar radiation, and shading from terrain. Spatially varying settling rates are caused by the interpolated precipitation and by resulting differences in snowpack development. Furthermore, the interpolated wind fields result in spatially varying sensible and latent heat fluxes.
Figure 4.8b shows the modelled snow distribution at the end of the accumulation season 2008/2009, which correspond to the measurements presented in Figure 2b. In Figure 4.8c the same modelled result is shown but with the same colour scale as used for the measurements in Figure 4.2b. As anticipated from the good correlations of the $\overline{Sx}$ parameter to measurements, spatial structures of modelled snow depth were good represented: Clearly visible are the two accumulation zones in the variant cross-loaded slope. The mean value of snow depth was too low, but relatively well captured compared to the variance (Table 4.3). Similar relationships were achieved in the lee slope. However, the measured relationships of mean and variance between cross-loaded and lee slope were well reproduced with a more variant cross-loaded slope with less average snow depth. Similarly, in the windward slope the trend of too low variance was continued, while the measured relations between slopes were reproduced. In comparison to the terrain parameter $\overline{Sx}$ (Figure 4.8a), correlations with measurements increased. Especially for the windward slope an increase from previously mentioned $r = 0.45$ to 0.65 was observed.
We could show that in complex terrain the variance of snow depth distribution at the end of the accumulation season was with such a model approach much too low reproduced (Figure 4.8c), while good relations between slopes as well as in-slope structure at least in two of three areas could be achieved (Figure 4.8b). These results are additional to the work of Winstral and Marks (2002), who verified a similar model approach with aerial photographs during the ablation period. Thus, in their study it was not possible to validate modelled snow depth distribution in terms of variance. In spite of remarkable agreements to measurements, some critical remarks will be given. In complex terrain this method seems not to be robust to the input parameter wind direction. In Figure 4.9, a typical Sx course is plotted for several angles in the cross-loaded slope. If the local horizon is decreasing in such a manner, only small changes in wind direction would result in large differences of the averaged $\overline{Sx}$ values. If the assumed wind direction would be changed from 315° to 292.5°, the $\overline{Sx}$ value for that grid point would rise from 0.1 to 0.4. As comparison, the standard deviation in the entire cross-loaded slope was 0.15 for the direction of 315°. An increase in the upwind window

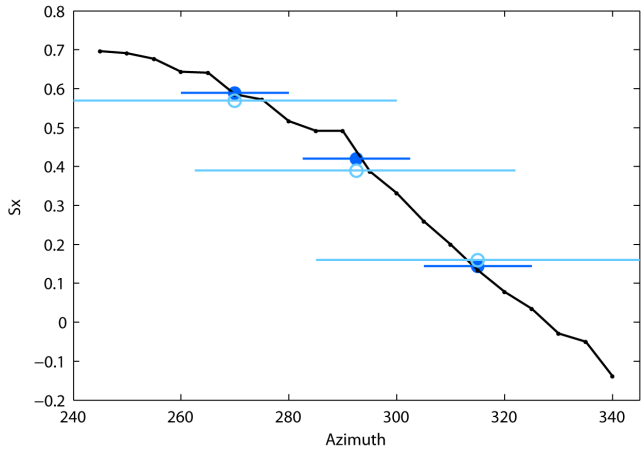

Figure 4.9: Sensitivity of terrain parameter: Sx values of a pixel in the cross-loaded slopes in 5 azimuth degrees increments (black). $\overline{Sx}$ values determined with a 30° (dark blue) and a 60° window (bright blue) for 270°, 292° and 315°. The horizontal lines mark the extent of the average window.

from 30° to 60° does not diminish the strong dependence on wind direction because of the linear decrease of the Sx values at that grid point. As a result, if the wind direction was changed from 315° over 292.5° to 270°, the correlation coefficients between calculated $\overline{Sx}$ values and measured snow depth at the end of the accumulation season decreased strongly from above mentioned 0.63 to 0.46 and 0.26. Especially when wind direction is changing strongly in a study area (up to 90° in our case, cp. Figure 4.1), a good correlation can only be achieved with locally adjusted wind directions. As an example, at the station on the ridge south of the snow station, a large cornice was built each winter consistently to the SW wind direction observed at this station. Calculation with a NW wind direction, for which good results in the cross-loaded slope were achieved, fail for an area less than 1 km apart.
One possible solution would be an integration of diverting wind directions. Purves et al. (1998) used a similar approach adjusting a shelter/exposure parameter for wind speed with diverting wind directions as suggested by Ryan (1977). The diverted wind direction A_d is given by

$$A_d(x_i, y_i) = -0.225\psi(x_i, y_i)\sin(2[\varphi(x_i, y_i) - \overline{A}]) \tag{4.3}$$

where φ is the aspect of grid point (x_i, y_i), ψ is the inclination in percent, and $\overline{A}$ the prevailing wind direction. We implemented $A_d(x_i, y_i)$ instead of a spatially constant wind direction $\overline{A}$ for the $\overline{Sx}$ values calculation shown in Eq. (4.1) and (4.2). If now the prevailing wind direction is again changed from 315° over 292.5° to 270°, in the cross-loaded slope the correlations between adjusted $\overline{Sx}$ parameter and measured snow depth at the end of the accumulation season were 0.65, 0.66 and 0.58. These findings suggest that the implementation of Eq. (4.3) might increase the robustness of the terrain parameter calculation at least for this case study. However, if all three slopes investigated were included, no better overall correlation could be achieved. This can be explained by the fact that too little divergence was achieved with Eq. (4.3) in comparison to the measurements at our

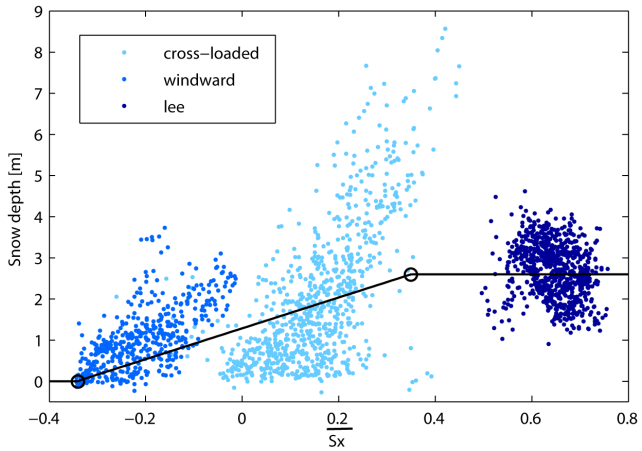

Figure 4.10: $\overline{Sx}$ values (wind direction 315°, maximum search distance of 300 m) vs. measured snow depth at HS_{max} 2009. The different colours indicate the three slopes investigated. The black circles mark the values determined at the snow station and at the used snow free grid point. The line between these points represents an average interpolation used in the modelled winter season.

stations. Theoretically, Ryan's formula allows a divergence of ±22.5° in slopes with an inclination up to 45°, which is too small to reproduce our observations. Additional calibration of that formula would be needed.

Another critical point is that the variance could not be reproduced with this linear interpolation of the $\overline{Sx}$ parameter. In Figure 4.10 the $\overline{Sx}$ parameter with corresponding measured snow depth at the end of the accumulation season is plotted. Also plotted are the both interpolation points for the NW storms. The linear interpolation gives a hint to the average interpolation used at each time step during the model season. Site dependent relationships can be observed: the windward slope revealed a smaller linear dependency than the cross-loaded slope. Also the smaller linear correlation quality for the windward slope is recognisable. Both can be explained if the windward slope is classified in an erosion dominated area in comparison to the cross-loaded slope. The same increase in wind speed may have less effect on erosion than on accumulation, which may explain the lower inclination of a linear dependency. The lower quality can be explained that erosion might be less coupled to wind speed since erosion is also a function of snow-surface properties. Possible non-linearities between the terrain parameter and snow depth were already mentioned by Winstral et al. (2002) and might point to different processes dominant in those slopes (e.g. saltation, preferential deposition). The errors resulting from these non-linear dependencies may exceed the improvements caused by a better described wind direction with Eq. (4.3) and may explain that those improvements can only be achieved locally.

In the lee slope the sheltering effect interpreted with the terrain parameter is too large in comparison to the cross-loaded slope (Figure 4.10). Thus, one overall linear fit would not capture this large variance in the cross-loaded slope. Reasons for the large sheltering in the lee slope could be the above mentioned changing wind directions. However, all other reasonable wind directions did not

alter the mean value substantially.
A reason for the missing variance reproduction in the cross-loaded slope might also be found in the resolution of the used DEM as suggested by Winstral et al. (2002). But also the usage of a DEM generated locally with our summer scans with a resolution of 2 m did not increase the variance of the $\overline{Sx}$ value substantially. For us, the differences in variance between lee and cross-loaded slope point to the possibility that processes which resulted from higher wind speeds in the cross-loaded slope (i.e. saltation) intensified the effects of sheltering and exposure. This is confirmed by findings of Dadic et al. (2010), who found a linear correlation between measured snow depth and modelled wind speeds as a hint to the dominance of preferential deposition in a large glaciered basin. For saltation, on the other hand, a non-linear dependency can be assumed, since Doorschot et al. (2004) found a highly non-linear dependency between particles in saltation and horizontal wind speeds.
In general, we explain the good results achieved with the method developed by Winstral et al. (2002) by the consistency of this method with our process understanding: Accumulation due to preferential deposition and saltation should be influenced by wind speed fluctuations, which are strongly controlled by shelter and exposure to upwind topography. The good agreement of the terrain interpretation to measurements confirms that topography is one of the most important factors influencing the snow depth distribution in complex terrain.

4.4 Conclusion

In this first paper, an in depth discussion of a series of TLS measurements of snow depths in a mountain catchment has been presented. A first significant result is that individual storms create diverse accumulation patterns, but the snow distribution at the time of maximum snow depth (HS_{max}) is mainly shaped by a few "master" NW storms. This is also seen as one reason that the HS_{max} snow distribution at the two investigated years is very similar and in fact more similar than the distribution created by any two pairs of individual storms. Thus it can be concluded that the terrain forces different storm patterns to converge to a similar final snow distribution.
The catchment features several sub-domains, three of which show clearly distinct snow accumulation patterns and have been named windward slope, lee slope and cross-loaded slope. With transects of snow depth a tendency of development towards a smoother surface during the accumulation phase was shown but that the larger scale terrain features remain intact. This interesting observation is a pre-requisite to achieve the snow depth distribution consistency found in our data.
The persistence of snow depth distribution motivates the use of simple non-local terrain parameters, which predict a snow distribution pattern based on wind speed and total available precipitation. The Winstral parameter captured the observed snow distribution qualitatively with a remarkable reproduction of the spatial pattern. It failed to reproduce the magnitude of the observed variability and therefore, when applied to a simple reconstruction of the HS_{max} situation using Alpine3D, also failed to reproduce the total snow depth distribution. This has been attributed to the fact that the simple sheltering effect captured by the terrain parameter produces a linear response to wind speed variations, while it is known that snow redistribution is a highly non-linear function of wind speed.

We conclude that a simple model approach for a very complex interaction of wind, terrain and snow depth distribution supposed by Winstral and Marks (2002) have large explaining potential for the snow depth distribution at the end of the accumulation season and can be seen as a benchmark for more physically orientated models using terrain influenced wind fields as input for saltation, suspension and sublimation models and their interaction with modelled snow cover properties (e.g. Lehning et al., 2008). The physical model approach allows investigating individual processes and is therefore complementary to the pure description of snow depth variability presented here. One next step is therefore to apply the process analysis as presented by Mott and Lehning (accepted)to the more complex setting in this new TLS laser scanner data set.
From our TLS measurements, snow distribution data is also existent for the ablation period together with runoff information. Therefore, a comparison of model performance in the ablation period will soon be presented elsewhere. The subject of this further work will be on the influence of the different energy fluxes to local snow melt with a particular focus on later energy fluxes caused by the heterogeneous snow surface.

5 Persistence in intra-annual snow depth distribution. Part II: fractal analysis of snow depth development

5.1 Introduction

Scale issues have always been a major subject in natural sciences. For example, geology students learn to place a pencil or hammer in every photograph of a geological feature since without this information it would hardly be possible to determine the photograph's scale. The mathematical concept behind these phenomena is the concept of scale invariance or self-similarity, commonly referred to as "Fractals". This concept was initially used to describe landforms by Mandelbrot (1977, 1982), and research has continued as new data becomes available for a wider range of spatial scales (e.g. Abedini and Shaghaghian, 2009). However, a consensus has emerged that natural phenomena only exhibit statistical self-similarity in limited regions and over limited ranges of scale (Xu et al., 1993; Sun et al., 2006). Furthermore, the temporal development of scaling parameters could be assessed for example for seasonal changes in surface roughness due to the evolving vegetation (Pachepsky and Ritchie, 1998) or for changes in an urban heat island effect across several years and between different seasons of a year in a Chinese city (Weng, 2003).
The scaling behaviour of snow depth was studied with manual probing, e.g. Shook and Gray (1996); Kuchment and Gelfan (2001); Arnold and Rees (2003), who found statistical self-similarity at least at certain scales or observed a break in scaling behaviour at approximately tens of meters. An important step further was possible through high-resolution airborne laser scanning data (Deems et al., 2006; Trujillo et al., 2007). Both studies found a multiscaling behaviour with a stronger spatially correlated structure before a scale break, which was consistent with the distance found in the previous studies. In total, both studies analysed six areas with an extent of 1 km^2 each with different vegetation and relief characteristics. Both studies established a link between scaling behaviour of snow depth and vegetation due to similar break distances. Trujillo et al. (2007) could group their five study areas into areas more, or less, dominated by wind. They noticed that in wind-dominated areas the scale break found in the vegetation height data was lower than in corresponding snow depth data, while it was at a similar distance in less wind-affected areas. No scale breaks were observed for the sum of vegetation height plus topography. On the contrary, Deems et al. (2006) found scale breaks in vegetation height plus topography in distances of the same magnitude as found for snow depth. However, they found no scale break for vegetation height alone. It appears that the difference in methods could be the only explanation for the disagreement since the same data sets

were used. Furthermore, Deems et al. (2006)related differences in break distances between regions to overall relief, while Trujillo et al. (2007) could not detect such a dependency. At least, both studies found an anisotropic scaling behaviour, which indicated a larger spatial autocorrelation in dominant wind direction after the scale break.
In an additional study, Trujillo et al. (2009) confirmed their hypothesis of a relation between break distances and wind speeds in two adjacent sub-areas. Deems et al. (2006) found previously in a dataset including these two sub-areas a "global" scaling behaviour. Both findings suggest that such "global" scaling behaviour can be partitioned in a number of clusters of different self-similar scaling regions due to different processes dominating in each region. Deems et al. (2008) observed a notable inter-annual consistency in scaling behaviour after two different accumulation seasons and concluded that there must be a consistent process relationships controlling snow accumulation and redistribution during years.
However, none of these investigations looked at the temporal development of scaling characteristics of snow depth or of snow depth change caused by individual storms during an accumulation season. No research has been done in regions of rugged, unvegitated high Alpine terrain.
This paper explores how the inter-annual consistency at the end of the accumulation season compares to the observed intra-annual consistency of individual snow storms. It extends the more qualitative description presented in Part I. Since fractal parameters have proved to be useful for distinguishing between surfaces that have resulted from different processes (Burrough, 1993), we want to use this method not only as a description of snow depth structure, but also to relate scale, region and temporal dependent variations observed to different formation processes.

5.2 Methods

5.2.1 Field description and data acquisition

Snow depth was measured before and after individual storm events in Alpine terrain with a terrestrial laser scanner (LiDAR). Three slopes of 300 m length were investigated, which are named in this study after their typical loading behaviour during dominant NW storms: windward, lee and cross-loaded. Part I presents a detailed introduction to the areas and measurement methods. In summary, the dataset used in this study contains the summer digital elevation model, nine snowfall periods including the first snowfalls in 2008/09 and 2009/10 as well as the snow depth at the end of the accumulation season for 2007/08 and 2008/09. The snow depth and snow depth change data are available in a 1 m resolution grid. Meteorological data was available from seven weather stations in the catchment.

5.2.2 Fractal analysis

Background

"Clouds are not spheres, mountains are not cones, coastlines are not circles ..." (Mandelbrot, 1982). Euclidean geometry is of hardly any help to describe the irregular form of natural patterns

and features (Sun et al., 2006). Instead, nature has been shown to generate structures and patterns that can be better described using the concept of scale invariance or self-similarity. For fractal objects, a relationship exists between length and measuring scale, which can be described by a power law with exponent $(1 - D)$, where D is the fractal (or Hausdorff-Besicovitch) dimension. A fractal is defined formally as a set for which the fractal dimension exceeds the integer dimension in Euclidean geometry (Mandelbrot, 1982). A curve's dimension can take any non-integer value between 1 and 2, depending on the degree of irregularity of its form. Similarly, a surface's dimension may be a non-integer value between 2 and 3 (Sun et al., 2006). An infinitely rugged surface that fills the third dimension has a limiting fractal dimension of 3, while a perfectly smooth surface has a fractal dimension of 2 (Goodchild and Mark, 1987). Thus, D can be used as a measure of irregularity. Typical interpretations of D in literature are given for example by Burrough (1993), who associates low D values with fractals for which long-range fluctuations dominate while large D values describe the dominance of short-range fluctuations. For Lam and De Cola (1993) low D values characterise smooth and persistent structures, and large D values rugged and antipersistent structures. Because of the quite long history of the fractal concept, review papers are available and are referred here for further readings (e.g. Xu et al., 1993; Sun et al., 2006).

Estimating fractal parameters

Because of typical data gaps in the data delivered by the terrestrial laser scanner, the variogram method was used to estimate fractal parameters, which can be applied for irregular data (e.g. Mark and Aronson, 1984; Sun et al., 2006). The surface investigated must fulfil the conditions of a fractal Brownian surface (Xu et al., 1993). However, researchers have suggested that if this condition is not satisfied, this method could still be used for extracting information (Sun et al., 2006). The variogram for a parameter z is given by

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j)\in N(h)} (z_j - z_i)^2, \tag{5.1}$$

where $N(h)$ the number of point pairs (i, j) in each distance class h (Webster and Oliver, 2007). Omnidirectional variograms, which include all point pairs independently of their direction, and directional variograms, which include only point pairs within a specified directional angular class, were calculated with evenly spaced distance classes after a log-transformation. To obtain enough data points for the linear regression, 50 distance classes were chosen. The smallest distance class is given by the resolution of the grid (1 m). The maximum distance was chosen to be approximately half of the maximum point-pair distance in the slopes investigated (Sun et al., 2006). This resulted in a few missing distance classes at lower distances due to the quadratic nature of the grid (e.g. Figure 5.1) but optimised both demands of enough data points and evenly spaced classes. Directional variograms were calculated for 16 angular classes of 22.5∘.

Since many studies reported that statistical self-similarity could only be observed in limited scales (Xu et al., 1993) and also since this was observed by all snow-related studies mentioned in the introduction, a multifractal model was chosen. Such multifractal behaviour can be a result of variations in processes operating at different scales (e.g. Mark and Aronson, 1984). The model given

in Eq. 5.2 was fitted to the variograms solving least squares,

$$\log(\hat{\gamma}(h)) = \begin{cases} \alpha_1 \log(h) + \beta_1, & \text{if } \log(h) \leq \log(L) \\ \alpha_2 \log(h) + \beta_2, & \text{if } \log(h) > \log(L) \end{cases} \tag{5.2}$$

with the continuity constraint,

$$\alpha_1 \log(L) + \beta_1 = \alpha_2 \log(L) + \beta_2.$$

Model parameters are break distance L, slopes of the log-log transformed variograms α_1 and α_2 as well as ordinal intercepts β_1 and β_2. The slope α of a log-log transformed variogram is for surfaces related to the fractal dimension D

$$D_{s,l} = 3 - \frac{\alpha_{1,2}}{2}. \tag{5.3}$$

Consistency of the fractal dimension with people's roughness intuition was confirmed by a study of Pentland (1984). They used a simplified set-up, however, with a constant ordinal intercept while D was changing. A well-known example, where perception and fractal dimension are rather inconsistent, is mentioned by Burrough (1981): a smooth airport runway has a relatively high D, since variations of long distances are low in amplitude. Similarly and more specific to snow, Fassnacht et al. (2009) obtained a large D value for a snow surface previously characterised as "smooth". The corresponding roughness was objectively low in magnitude. Therefore it seems useful to refer to the magnitude of the roughness together with D, which is also suggested in literature (Klinkenberg and Goodchild, 1992; Sun et al., 2006). This magnitude can be assessed with the ordinal intercept (β_1 in Eq. 5.2) and stands for the expected differences for point pairs a unit distance apart (Klinkenberg and Goodchild, 1992). In literature the abbreviation γ has been consolidated for the ordinal intercept, a notation, which we choose to adopt here.
A discussion on artifacts influencing D is given in Xu et al. (1993) and Sun et al. (2006). D is a function of methods and choice of the methods' input parameters. Furthermore, smoothing effects resulting from the limited size and resolution of data will affect the estimated D value. Anisotropy effects such as the distance dependent point density, footprint and accuracy of the laser scanner used in this study will certainly have an influence. We discuss these issues in the results section.

5.3 Results

The parameters discussed are the fractal dimension before and after a scale break (D_s, D_l) the distance of this scale break L, the ordinal intercept γ and anisotropies of these parameters. These fractal parameters were calculated for snow depth and snow depth change in three different slopes to study their time development in one season and the inter-annual comparison of the HS_{max} situation of two years. In order to structure this large parameter set, we first show results calculated with omnidirectional variograms. Snow depth structure at the end of the accumulation season, differences between slopes and similarities between two years are presented first. Subsequently, the time development of the fractal parameters (D_s, D_l, L, γ), both calculated for snow depth and

snow depth change, is presented in detail for the cross-loaded slope, with selected results from the other two types of slope. Finally, anisotropy effects are shown with directional variograms.

5.3.1 Omnidirectional variograms

Snow depth at the end of the accumulation season

Variograms calculated for snow depth at the end of the accumulation season are presented in Figure 5.1. Scale dependent variations are apparent in all three slopes investigated, in that scale break ranging between meters and tens of meters was consistently observed. Fractal dimension values before the scale break (D_s) were lower than after the scale break indicating a more persistent structure before than after the scale break. This is consistent with all previous studies observing fractal behaviour of snow depth as discussed in the introduction.
Region dependent variations between slopes could be detected: the lee slope was characterised by the shortest break distance and larger D values before and after the scale break. These values indicate a structure where short-range fluctuations were dominant. The largest ordinal intercept γ indicates that the variance at one meter distance was largest. Overall variance was lowest, which was presumably responsible for our prior rating of that slope to have a more homogeneous structure (Part I). Different characteristics were observed in the cross-loaded slope. The long break distance demonstrates that the shorter, more persistent scale extends to larger distances in comparison to the lee slope. The smaller short-range D_s values indicate that this scale was also more spatially persistent. The intercept γ was lower than in the lee slope, but overall variance was largest. The windward slope is characterised by a smaller variance at all scales in comparison to the cross-loaded slope.

Time development in cross-loaded slope

Inter-annual consistency in all parameters could be detected for the cross-loaded slope (Figure 5.2). A similar consistency was observed for the lee slope, while no data in a second year was available for the windward slope. These findings confirm the large correlations mentioned above and are consistent to the study of Deems et al. (2008), who concluded that there must be inter-annually consistent process relationships among the major driving factors controlling snow accumulation and redistribution.
In the following section we will show how this inter-annual consistency is formed during an accumulation season, and see if such consistent process relationships can also be observed during individual snowfall events.

Changes in snow depth Figure 5.3 shows the temporal development of short-range fractal dimension D_s and of break distance L. Most apparent is the increase in break distance for NW storms, especially at the end of the accumulation season (marked with arrows). These NW storms

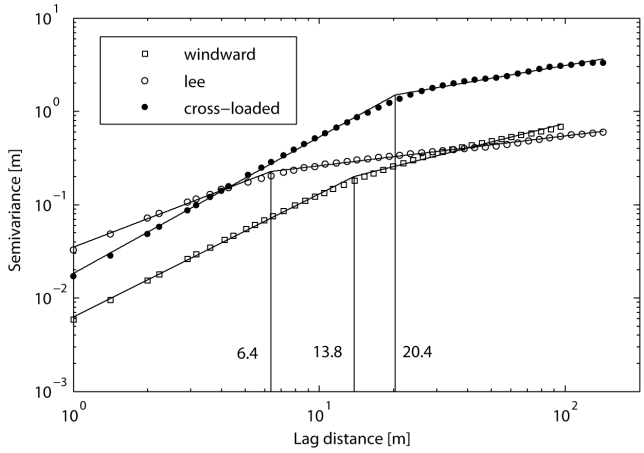

Figure 5.1: Omnidirectional variograms for snow depth at the end of the accumulation season 2008/09 for the three different slopes investigated. The vertical lines and corresponding numbers mark the distance of the scale breaks.

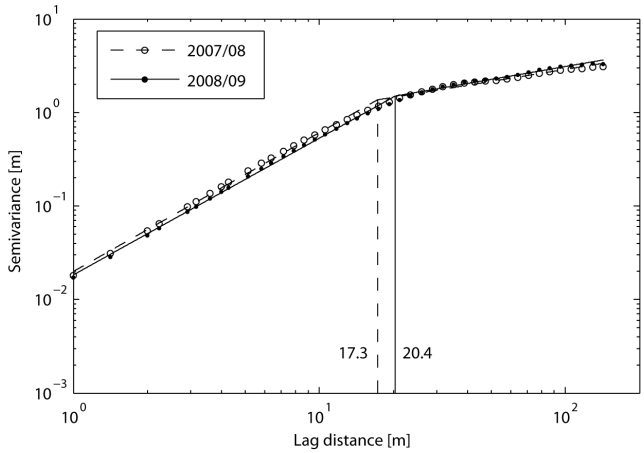

Figure 5.2: Omnidirectional variograms for snow depth at the end of the accumulation season in two different years in the cross-loaded slope (2008/09 same as in Figure 5.1). The vertical lines and corresponding numbers mark the distance of the scale breaks.

were characterised by large amounts of precipitation and wind speeds. Increasing the break distance indicates that the scale of the more persistent range was extended. We interpret this increase with successive filling of summer terrain depression: filling small scale terrain structure results in a somewhat smoothed surface and allows processes as saltation to inherit a larger process scale (see also the transect description in Part I). This increase in scale break is visible in both accumulation zones (drifts), which were observed to increase in size during all NW storms in the cross-loaded slope. Their appearance changed from a segmented structure showing various sub-maxima at the first NW storm (Figure 5.4a) into a more homogeneous structure at the last NW storm (Figure 5.4b).

The ordinal intercept γ (Figure 5.5) decreased during the accumulation season for NW storms in the cross-loaded slope. This decrease in magnitude of the roughness which was described with D confirms the interpretation of summer terrain filling and surface smoothing at small scales. A similar decrease was observed with the overall variance (see Part I). However, the decrease of the overall variance can be explained instead of a time development with the amount of snowfall reported at a flat field snow station during NW storms: larger total variance might be explained with larger snowfall amount. γ seems to show a more apparent time dependency than the overall variance. For the second NW storm, which had a large snowfall amount just as the first NW storm, γ was already much smaller and more similar to subsequent NW storms with lower snowfall amounts. The decrease of γ with time confirms the thesis of summer surface smoothing during the accumulation season: the more the summer structure is smoothed, the smaller are the magnitude of the short-range fluctuations of snow depth change.

The fractal dimension and break distance is able to distinguish between typical NW storms (marked with arrows) and other periods; small D_s values and long break distances are typical features of NW storms in the Wannengrat catchment (Figure 5.3). During these events, accumulation was concentrated in both accumulation drift zones. This resulted in long-range dominated fluctuations (below the scale break) and explains the low D_s values. Also after the scale break a larger spatial persistence was observed with long-range D_l values of approximately 2.8 while they had values around 2.95 for other periods.

In comparison to the NW storms, after both homogeneous snowfalls during low wind speeds short break distances and large D_s values were observed (marked with stars in Figure 5.3). Particularly for the first snowfall in the season 2008/09, the dominance of short range fluctuations might be explained by small scale summer terrain filling (Figure 5.4d). For example, this first snowfall had the largest magnitude γ that was observed during the whole accumulation season, which points to the effect of summer terrain. This explanation might also be valid for the second homogeneous snowfall, since the NW storm in between was able to erode most of the upper area in this slope (see Figure 5.3a in Part I). However, the magnitude γ was substantially lower, which indicates the possibility that some small scale depression remained filled with snow. This interpretation is consistent with visual observations.

Confirming the effect of summer terrain, large γ values were obtained for the first snowfall in the season 2009/10 as well. Based on wind data from the AWS and on the patterns of snow depth distribution (see Figure 4.5b, pronounced accumulation zones) this first snowfall revealed typical characteristics of a NW storm, however falling on bare ground. This explains the longer break

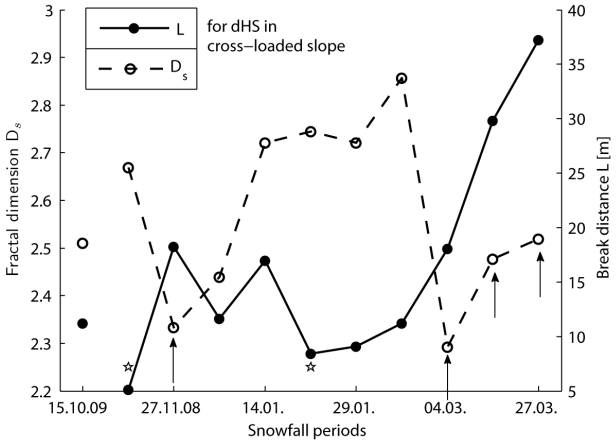

Figure 5.3: Temporal development of break distance L and short-range fractal dimension D_s determined for snow depth change during the winter 2008/09 in the cross-loaded slope. The values for the first snowfall in 2009/10 are plotted without connective lines. NW storms are marked with an arrow and homogeneous loading with a star.

distances and lower D_s values in comparison to the first homogeneous snowfall during low wind speeds the year before. Lower break distances and larger D_s values in comparison with other NW storms reveal the influence of the summer terrain.

Absolute Snow Depth The development of short-range D_s and the break distance L of the absolute snow depth is shown in Figure 5.6. An increase in break distance after NW storms (marked with arrows) is clear. Similar to the development of snow depth change, the maximum (20 m) was reached at the end of the accumulation. However, this value was much smaller than for snow depth change (38 m, Figure 5.3) and more comparable to the first NW storm (20 m, Figure 5.3). This finding confirms the observation determined with the correlation coefficients in Part I that the snow depth distribution at the end of the accumulation season inherits much or most of the characteristics of the first NW storm. This similarity is also shown in Figure 5.4: whereas both pronounced accumulation zones were homogeneously structured in the last NW storm (b), the structure of the first NW storm (a), with many sub-maxima, is still visible at the end of the accumulation season (c).
The fractal dimension D_s decreased after NW storms. During these storms both accumulation zones became more and more pronounced. This may explain the increasing dominance of long-range fluctuation (before the scale break), which results in decreasing D_s values.
The ordinal intercept γ increased during the season (Figure 5.5) after NW storms, similarly to the development of the overall variance shown in Part I. The decrease in D_s values at the same time implies a trend to more long-range dominated fluctuations and results together with the increasing γ in a larger variance for all distances. Indeed, the variance is also dependent on the fractal dimension after the scale break, but values were comparable during the season.

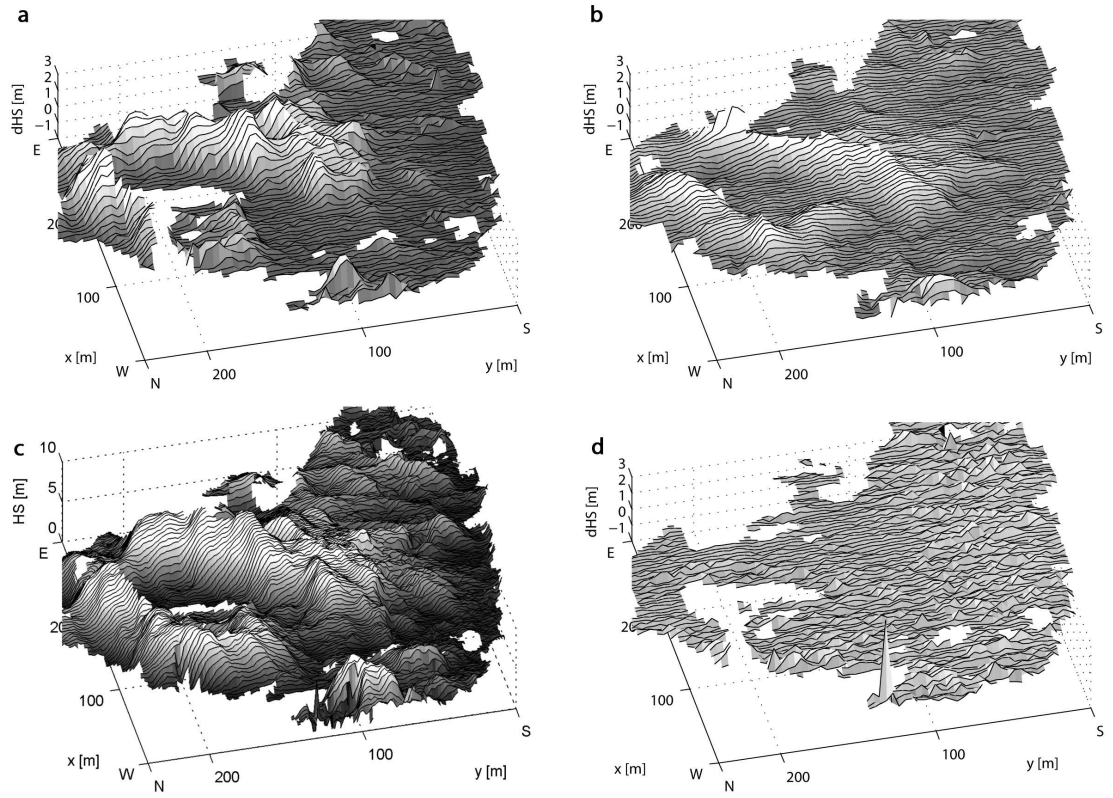

Figure 5.4: Change in snow depth (dHS) during (a) the first NW storm and (b) the last NW storm in 2008/09, while (d) is an example of an snowfall during low wind speeds and also the first snowfall in 2008/09. (c) shows the distribution of snow depth (HS) at the end of the accumulation season in 2008/09.

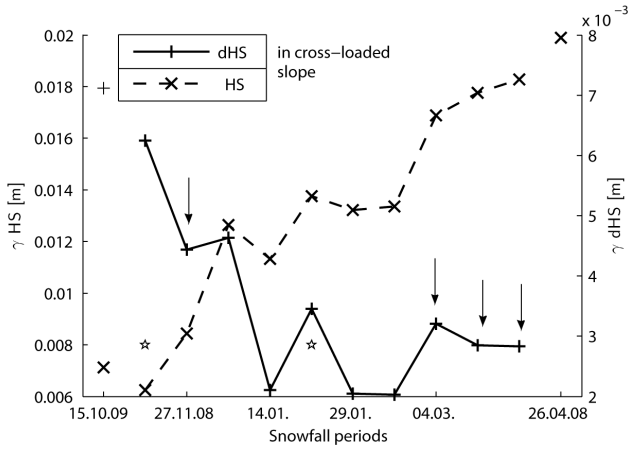

Figure 5.5: Temporal development of γ determined for snow depth (HS) and snow depth change (dHS) during the winter 2008/09 in the cross-loaded slope. For comparison, γ for the maximum HS in 2008 is plotted without connective lines. NW storms are marked with an arrow and homogeneous loading with a star.

For comparison, values for the HS_{max} situation the year before are plotted in Figures 5.5 and 5.6, which point to the similarity between both years. Differences in time development are larger than the differences between two years at the end of the accumulation season. At the end of February, only one NW storm was observed in this season. The increase in break distance and decrease in D_s at the end of the accumulation season indicate that the last three NW storms could influence the spatial structure of snow depth substantially. In comparison, the periods between the first and the three last NW storms do not alter the structure of snow depth to the same degree. These periods were SE storms without precipitation or snowfall during low wind speeds (see also Part I). Hence, snow depth structure actually inherited the characteristics of NW storms with lower D_s and longer break distances. This indicates that the development of snow depth does not converge to a somehow predefined status at the end of the accumulation season, and that the inter-annual consistency of snow depth in this area is highly dependent on the frequency of NW storms in one accumulation season.

Time development in lee slope

Changes in snow depth The differences between the lee and the cross-loaded slope mentioned previously for the HS_{max} situation in Figure 5.1 were also seen for snow depth change during each individual snowfall, i.e. lower break distances (up to eight meters), substantially larger fractal dimensions before (between 2.7 and 2.8) and after the scale break (2.9) and larger ordinal intercept γ. This indicates a rough, non-persistent structure. A significant time development could only be observed for the ordinal intercept γ (Figure 5.7), which decreases similarly as in the cross-loaded slope during NW storms. Observed differences between the cross-loaded slope and lee slope persist during the accumulation season, at least for NW storms. The exceptions to this are snowfall periods

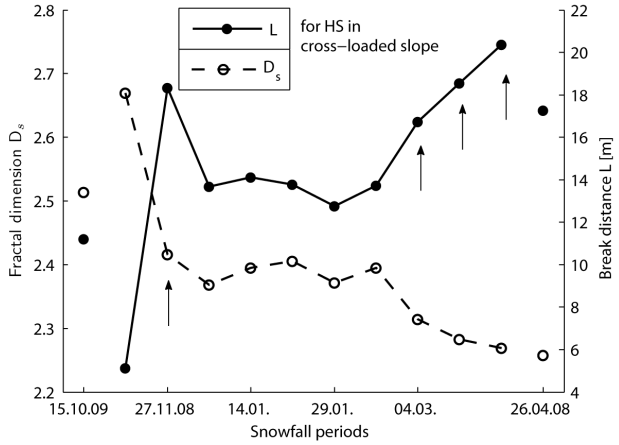

Figure 5.6: Temporal development of break distance L and D_s determined for snow depth during the winter 2008/09 in the cross-loaded slope. For comparison, parameters for the first snowfall in 2009/10 and for the maximum HS in 2008 are plotted without connective lines. Situations after NW storms are marked with an arrow.

during low wind speeds in the cross-loaded slope. In those periods the fractal parameters are very similar in both areas. Also, in the cross-loaded slope, a more antipersistent structure with shorter break distances and larger D_s values was seen as mentioned above. In brief, depositions during low wind speeds appear to reveal a typical region-independent characteristic. For the cross-loaded slope, we explained short break distances and large fractal dimensions with the influence of summer terrain even for the second snowfall, since NW storms were able to erode a large part of the snow fallen before, but no substantial erosion during storms was recorded in the lee slope. However, the transect description in Part I revealed that the picture of smoothing surfaces and filling small scale depressions was partly wrong and some small scale terrain features could survive the full accumulation season. In Part I, a relation of that observation to lower wind speeds was suggested. Furthermore, no increase in break distance with time could be observed in the lee slope, which is an additional indication that some terrain features were not altered to the same degree as in the cross-loaded slope. In addition to the effect of small scale terrain, small random eddies in wind flow developing in the lee slope might be a reason for the less persistent deposition structure.
Similarly, Trujillo et al. (2007, 2009) found shorter break distances in less wind dominated regions. They related shorter break distances in areas with lower wind speeds to existing vegetation height due to similar scaling behaviour. Without disagreeing on the influence of vegetation, we could observe a relation of shorter break distances and lower wind speeds independent of vegetation since in our area the influence of vegetation can be neglected. However, Trujillo et al. (2009)found in the wind dominated area less persistent structure in comparison to an adjacent less wind dominated and forested area, which is contrary to our observations for low wind speeds: low wind speeds without vegetation may be characterised by an antipersistent structure, while interception of trees may counteract that observation.

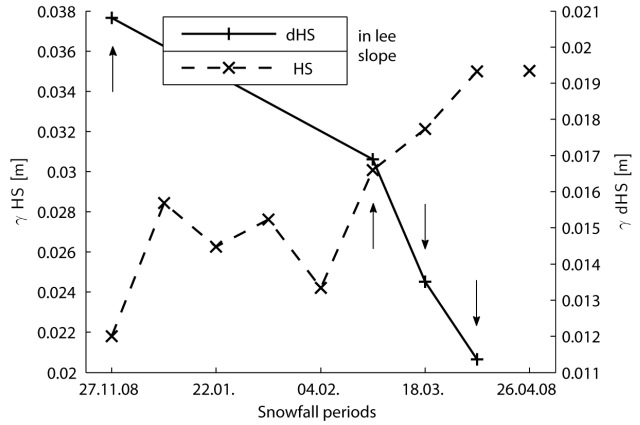

Figure 5.7: Temporal development of gamma determined for snow depth (HS) and snow depth change (dHS) during the winter 2008/09 in the lee slope. For comparison, gamma for the maximum HS in 2008 is plotted without connective lines. NW storms are marked with an arrow.

A simpler explanation for the differences between cross-loaded slope and lee slope can be found in the larger errors of the scanner due to the longer distances to the scan position for the lee slope. More noise could result in the observed larger D values. The larger footprint of the laser beam as well as the coarser resolution of the point cloud delivered should have an opposite influence. For one period the same situation could be measured from a position much close to the area of interest (see Figure 4.1). All characteristic differences, i.e. larger fractal dimensions, shorter break distances and larger ordinal intercepts can be reproduced from this more reliable dataset. The difference in break distance was less than 1 m, in D_s less than 0.1, D_l less than 0.01. All differences are in the range of the intra-annual variation observed in that slope, but much smaller than the differences to other areas. In brief, the influences articulated in the Methods section cannot be neglected, but the characteristic differences could be reproduced and can be interpreted as natural instead of an artifact.

Absolute snow depth Similar temporal developments of fractal parameters for snow depth as in the cross-loaded slope were obtained. These included decreasing short-range fractal dimension (2.65 to 2.5), time-independent long-range fractal dimension, increasing ordinal intercept and a consistency in those parameters for the maximum accumulation situations of different years (Figure 5.7 and 8). However, no increase in break distance could be observed (Figure 5.8).

Time development in windward slope

In the wind dominated windward slope increasing break distances for snow depth change (12 - 28 m) were observed. This confirms the relation of larger wind speeds leading to larger scale breaks. The increase of scale break distances during the accumulation season supports the hypothesis that in areas with larger wind speeds summer terrain can be altered to a larger degree. For snow depth

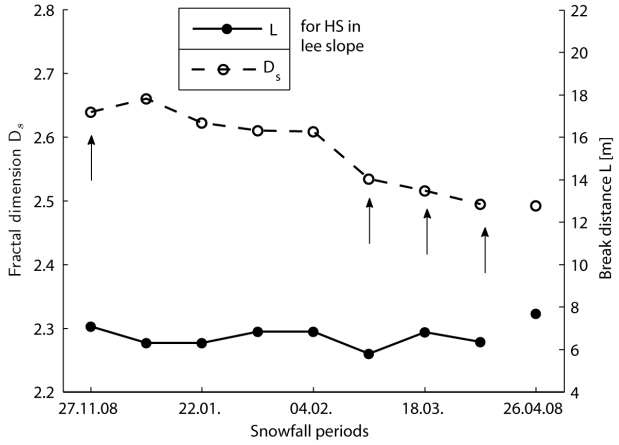

Figure 5.8: Temporal development of break distance and D_s determined for snow depth during the winter 2008/09 in the lee slope. For comparison, parameters for the maximum HS in 2008 are plotted without connective lines. Situations after NW storms are marked with an arrow.

an increase in break distance was observed at a lower level in comparison to snow depth change. These findings are consistent with the observations in the cross-loaded slope. Accordingly, break distances found for the HS_{max} situation (14 m) were most similar to those of the first NW storm, which shows the dominance of the first storm also in this slope. The windward slope showed the same influence of time on the fractal parameters as already seen on the cross-loaded slope, including decreases in D_s and increases in γ with time and no apparent influence of time on D_s.

5.3.2 Directional variograms

As before, we describe the directional variograms in detail for the cross-loaded slope and use the lee and windward slope to illustrate similarities and differences arising over time and from different storms.

Anisotropy in cross-loaded slope

Changes in snow depth During NW storms the break distances showed a substantial anisotropy (Figure 5.9). The direction of the anisotropy (along the semimajor axis of an assumed ellipse) of the first storm was WNW-ESE orientated, which is mostly parallel to dominant wind direction. However, the anisotropy of the later storms was mostly normal to dominant wind direction, which is consistent with the orientation of the accumulation zones.
Figure 5.10 shows anisotropy in short and long-range fractal dimension (D_s and D_l) during the first and last storm in the cross-loaded slope. The anisotropy in D_s was not pronounced for the first storm, while for the last storm anisotropy normal to wind direction can be recognised. Beyond the scale break, however, a more pronounced anisotropy during all NW storms was observed. Normal to wind direction the structure showed a larger spatial persistence. During low wind speeds the

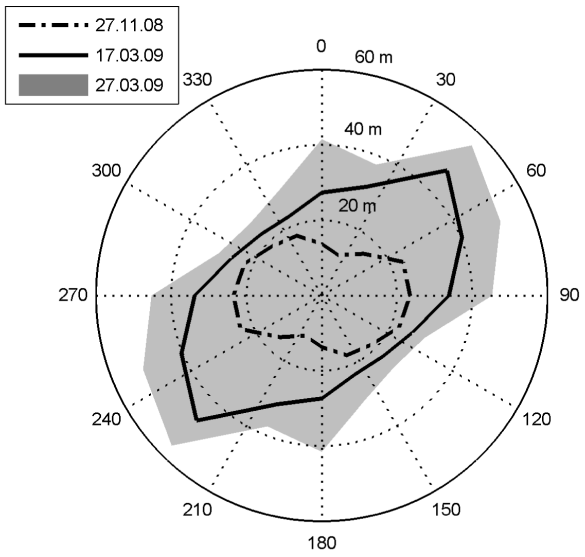

Figure 5.9: Scale break anisotropy of snow depth change for two selected NW storms in cross-loaded slope.

snow depth change structure did not reveal anisotropy in D_l (not shown).
In Figure 5.11 the anisotropy of the ordinal intercept γ is shown in the cross-loaded slope. A larger magnitude of the roughness which was described with D_s was observed parallel to dominant wind direction. While values were decreasing with time (as mentioned above), the anisotropy remained similar.

Absolute snow depth For the scale break anisotropy, a less pronounced shift in direction was observed, similarly to for snow depth change during NW storms. At the end of the accumulation season the direction was normal to dominant wind direction.
After the last three NW storms a small anisotropy in D_s can be observed, also normal to wind direction. This finding is consistent with Deems et al. (2006), but somewhat less pronounced. A pronounced anisotropy in D_l was observed throughout the accumulation season, which was along dominant wind direction and again similarly to snow depth change during NW storms. The relation to dominant wind direction was also found by Deems et al. (2006), and Trujillo et al. (2007) for wind dominated areas. Thus, it appears that the anisotropy of D_l, determined for example for snow depth at the end of the accumulation season, is a good indicator for dominant wind direction at the slope scale.

Summary A typical relation to wind direction could be observed both for snow depth and snow depth change for most of the measured time steps. D_l and γ were orientated along dominant wind

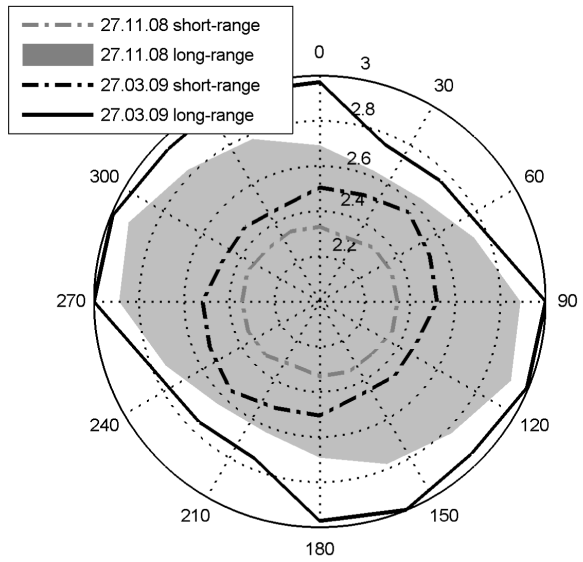

Figure 5.10: Fractal dimension anisotropy of snow depth change for two selected NW storms in cross-loaded slope.

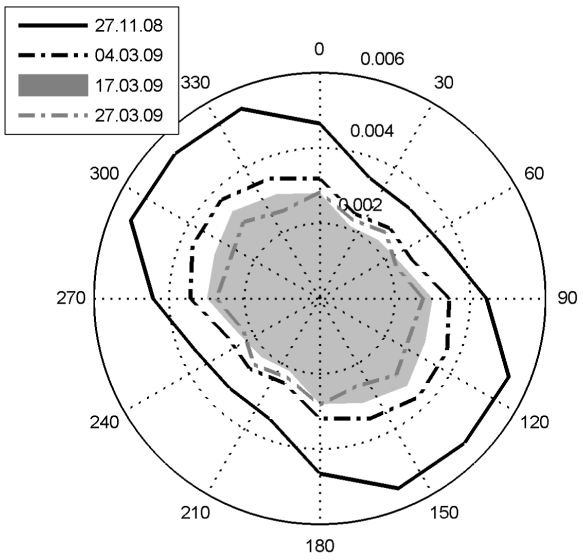

Figure 5.11: Ordinal intercept γ anisotropy of snow depth change for all NW storms in cross-loaded slope.

direction. The scale break, however, was typically normal to dominant wind direction, especially at the end of the accumulation season. In comparison, Trujillo et al. (2007) found longer break distances along the dominant wind direction for snow depth at the end of the accumulation season for two wind dominated regions. Presumably, in that case, vegetation and other point obstacles protruding out of the snow surface caused snow accumulation downstream, which are longer in the wind direction than undisturbed deposition. Only for snow depth change during the first NW storm observed in our study a similar relation was found (Figure 5.9). This finding suggests that snow depth change during this first storm was similarly influenced by point obstacles not already covered by snow.

Anisotropy in lee slope

We did not find a good fit to snow depth change data from the lee slope using directional variograms and Eq. 5.2. We suspect that this is a result of data quality due to the longer distance of that area to the scan position, the lower signal to noise ratio in comparison to snow depth and the lower number of point pairs in comparison to the omnidirectional variograms for the same slope. We therefore report only anisotropies determined for snow depth for the lee slope.
The behaviour of anisotropies seen on the cross-loaded slope was also consistently observed on the lee slope throughout the accumulation season. The scale break anisotropy was aligned in NW-SE direction, while anisotropy in D_l and in γ was normal to that direction (i.e. NE-SW). D_s did not show a pronounced anisotropy. The weather station at the bottom of that slope (see Figure 4.1) as well as some sporadic drift observations indicate that at least in the lower part of that slope the wind was blowing from SW during the typical NW storms. Thus, all anisotropy directions are similarly related to dominant wind direction as in the cross-loaded slope.

Anisotropy in windward slope

Anisotropies were found for many parameters. Most obvious was the consistent orientation of snow depth γ with time. Referencing to wind direction was difficult due to uncertainty in the dominant wind direction. Also, the typical perpendicular and parallel direction of anisotropies mentioned for the two other slopes could not be found, in particular for snow depth at the end of the accumulation season. We suggest that this is the result of two or more dominating situations on the snow depth structure. Additional to NW storms, the relatively large influence of the SE storms in this slope was discussed in Part I.

5.4 Conclusion

Fractal modelling of surface roughness and structure has historically helped to formulate general laws for complicated quantities. The complex features of snow depth distribution and their changes in time in Alpine terrain as described in Part I are a typical field for applying such fractal modelling. We therefore attempted to analyse the scaling behaviour of snow depth and snow depth changes for three distinct sub-areas in the Wannengrat catchment in SE Switzerland. This is an Alpine site

with no vegetation.

In Part I of this paper we showed that the snow depth development is predominantly shaped by north-west storms and leads to a similar snow depth distribution at maximum accumulation. In this part, we have shown that fractal parameters are able to distinguish snow depth structure in wind-protected and wind-exposed areas and the structure of snow depth change during more and less wind-influenced snowfall periods. In the two wind-exposed sub-areas the development of snow depth change can be summarised with a tendency towards larger break distances in the course of the winter and towards smaller intercept γ and fractal dimension D.

While former studies tried to relate the observed scale break to vegetation influence, we interpret the scale break in a different way since vegetation influence can be neglected in our study area. It can be interpreted as the roughness scale of summer terrain which is modified by snow fall and hence dominantly influences snow deposition at a slope scale. At distances larger than the scale break the signal of summer terrain persists. These interpretations are consistent with the visual observation that small scale variations in topography get smoother during the winter but that the roughness scale that produces the dominant snow drifts (i.e. the break distance) remain intact. This is a prerequisite to obtain persistent characteristics of NW storms throughout an accumulation season described in Part I.

An increase in break distance can be seen as a result of successive smoothing of summer terrain, which allows processes as saltation and suspension to act at a slightly increasing scale (20 to 40 m). A trend towards smaller intercept γ suggests that small scale roughness is reduced, similar to the decrease of fractal dimension D. This is a prerequisite to obtain persistent characteristics of NW storms throughout an accumulation season described in Part I.

The snow depth on the (nominal) lee slope has larger intercept γ, larger D and a smaller break distance than the cross-loaded slope. The break distance does not change noticeably during the winter. We suggest that if winds are low, summer surface smoothing is restricted to the filling of depressions (small scale, i.e. below the break distance).

The time development shows the dominance of NW storms on the snow depth structure at the end of the accumulation season. Other events as SE storms without precipitation and snowfall during low wind speed were not able to influence snow depth structure equivalently. This dominance of NW storms was already quantified in Part I with good correlations between snow depth change due to NW storms and the maximum snow depth data. As a result of correlation coefficients up to $r = 0.97$ between the HS_{max} situations of two different winters it was proposed in Part I that snow depth development converge to a similar final snow distribution. However, time development of fractal parameters suggests that even at the end of the accumulation season snow depth structure was highly altered by individual NW storms. Hence, the inter-annual consistency observed between two years might be strongly dependent on the frequency of dominant NW storms in an accumulation season.

Up to now, three areas – one in this study and two areas investigated by Deems et al. (2008) – show a large interannual consistency. Future work will address to the question if years or areas with less interannual consistency can be found in Alpine terrain and if reasons can be for those differences can be given, e.g. the dominance of not only one precipitation event. A large avalanche was observed in

the lee slope (see Part I). The observed snow depth structure (small break distance) might favour fracture propagation over large distances. More studies are needed in areas where large avalanches were recorded to establish a relationship between snow depth structure and the occurrence of large avalanches.

Future work will involve snow depth scaling investigation for larger areas and all possible sub-areas such as steep rock walls and other surface features. An automatic procedure to generate consistent sub-areas using clustering will be attempted. The ultimate goal is to describe mean snow depth and snow depth variations based on simple terrain parameters and overall precipitation information. This would be a major step forward in the field, in which the basic question on "how much snow is on the mountain, where, when and why" is currently unanswered. Our research will complete our picture of the earth's surface roughness without snow (e.g. Mark and Aronson, 1984; Perron et al., 2008; Abedini and Shaghaghian, 2009) and with snow. In particular, we want to characterise snow surface roughness from the scale of the individual snow grain (Manes et al., 2008) to the scale of snow drifts investigated here.

This work has shown that it may be possible to build a coherent fractal description of snow distribution in Alpine terrain based on terrain models and knowledge of prevailing winds. This would be an extremely valuable tool for hazard warning, remote sensing and mountain hydrology.

6 Conclusions and outlook

6.1 Avalanche danger and snowpack stability

The object in chapter 1 was to analyse whether modelled variables of the snow cover model SNOWPACK improve the performance of numerical avalanche forecasting, in comparison to models that are based only on measured weather data. The findings presented suggest that for a balanced performance between all four danger levels and for good overall accuracy modelled snow stratigraphy data as provided by SNOWPACK are needed. SNOWPACK is in particular needed for the important days when the danger level changes. Model performance did not further improve using additional SNOWPACK simulations at an automated weather stations and/or virtual slopes simulations. This result indicates that the limited modelled snow cover variability was not related to regional avalanche danger. This result does not mean that spatial variability does not influence local to regional avalanche danger.

Although a certain quality was detected for the statistical methods, a notable discrepancy remained. A similar imperfect model performance was also recognised in a study of Schweizer and Föhn (1996), although they had more reliable observations than in the presented study, i.e. observed (instead of modelled) snow cover variables and verified (instead of forecasted) danger levels. Two possible conclusions remain: (1) additional information, which is not formalised at present enters the decision process such as the experience and intuition of the individual; (2) the forecasted danger level is not a good target variable, since it might be erroneous due to incorrect data at the time of the forecast or due to variations in human perception (McClung and Schaerer, 2006).

Due to (2), the target variable was changed from forecasted avalanche danger to observed stability (chapter 3). Similarly, this parameter lacks consistency since it is also an expert rating of observations. A study by Schweizer et al. (2008b) showed that a level of uncertainty exists in the detection of unstable slopes with such stability interpretations. Moreover, the extrapolation from point observations to a regional stability assessment is difficult. However, both forecasted danger levels and observed stability are the most reliable target variables describing avalanche danger. Additionally, together with avalanche observations, said variables are the only proxy data which is available for several years.

It could be shown that the stability estimate Sk_{38} as implemented in SNOWPACK is applicable for regional stability assessment. A combination of important modelled variables did show a better classification than the estimate Sk_{38}. These results confirm that snowpack modelling delivers an additional benefit to the measured input variables when snow stability needs to be estimated.

Both for avalanche danger and snow stability evaluation, the models presented could be used by avalanche warning services as nowcast or forecast. Nowcasting may already have an important

value, since information on instability at the present day is rare. In a forecasting operation, the present snow cover would be simulated with measured data first and then the development of the snow cover would be predicted with forecasted meteorological data for the next day. Finally, the predicted snow cover would provide the additional input variables needed for the statistical methods.
In Murphy (1993) identified three characteristics from which the quality of a forecast could be determined. The studies presented within this thesis are mainly focused on the correspondence of a forecast to observations, which Murphy (1993) called quality or type 2 goodness. In the case of the classification trees used in this investigation, it was possible to analyse the correspondence between the forecast and expert judgement. This was labelled by Murphy as consistency or type 1 goodness. The last type mentioned by Murphy, the value, is the additional benefit realised by decision makers through the use of the forecasts and was not studied here. Only the decision makers themselves are able to achieve this value of a forecast system. Hence, the models would need to be integrated into the operational Swiss avalanche warning service. Similarly to a study presented by Heierli et al. (2004), the forecasters could then define daily the additional benefits of including the models over a test region. In addition to the models presented here, other helpful tools are available. These models predict avalanche days with meteorological data (e.g. Buser, 1983) or avalanche danger levels using measured snow stratigraphy observations (Schweizer and Föhn, 1996). All such models are presently not integrated in the operational warning service in Switzerland. Therefore, the proposed value achievement could be expanded and would certainly benefit from comparisons between those models. However, in Switzerland it seems that a preceding step is necessary: the warning service will not accept models verified solely on measures of quality and consistency.

6.2 Snow depth variability

In chapter 3, an in depth discussion of a series of TLS measurements of snow depths in a mountain catchment has been presented. The first significant result is that individual storms create diverse accumulation patterns but that the snow distribution at the time of maximum snow depth (HS_{max}) is mainly shaped by a few "master" NW storms. This is also seen as one reason as to why the HS_{max} snow distribution for the two investigated years is very similar and in fact more similar than the distribution created by any two pairs of individual storms. Thus it can be concluded that the terrain forces different storm patterns to converge to a similar final snow distribution.
With transects of snow depth a tendency of development towards a smoother surface during the accumulation phase was shown but that the larger scale terrain features remain intact. This interesting observation is a pre-requisite to achieve the snow depth distribution consistency found in our data.
The persistence of snow depth distribution motivates the use of simple terrain parameters, which predict a snow distribution pattern based on wind speed and total available precipitation. The Winstral parameter captured the observed snow distribution qualitatively with a notable reproduction of the spatial pattern. It failed to reproduce the magnitude of the observed variability. This

has been attributed to the fact that the simple sheltering effect captured by the terrain parameter produces a linear response to wind speed variations, while it is known that snow redistribution is a highly non-linear function of wind speed. We conclude that a simple model approach for a very complex interaction of wind, terrain and snow depth distribution suggested by Winstral and Marks (2002) have large explaining potential for the snow depth distribution at the end of the accumulation season.

In chapter 4, fractal modelling was attempted to analyse the scaling behaviour of snow depth and snow depth changes. It was shown that fractal parameters are able to distinguish snow depth structure in wind-protected and wind-exposed areas, and the structure of snow depth change during more and less wind-influenced snowfall periods. In the two wind exposed sub-areas the development of snow depth change can be summarised with a tendency towards larger break distances in the course of the winter and towards smaller intercept γ and fractal dimension D. The break distance can be interpreted as the roughness scale of summer terrain which is modified by snow fall and hence dominantly influences snow deposition at a slope scale. An increase in break distance can be seen as a result of successive smoothing of summer terrain, which allows processes as saltation and suspension to act at a slightly increasing scale (20 to 40 m). A trend towards smaller intercept γ suggests that small scale roughness is reduced, similar to the decrease of fractal dimension D. These results are consistent with the visual observation that small scale variations in topography get smoother during the winter but that the roughness scale that produces the dominant snow drifts (i.e. the break distance) remain intact. This persistence in surface structure is a prerequisite to obtain persistent characteristics of NW storms throughout an accumulation season.

In the wind-protected lee slope, snow depth structure is less persistent or rougher. The break distance does not change much during the winter. These results indicate that, if there is not much wind, summer surface smoothing is restricted to the filling of depressions (small scale, i.e. below the break distance). The time development shows the dominance of NW storms on the snow depth structure at the end of the accumulation season. Other events such as SE storms without precipitation and snowfall during low wind speed were not able to influence snow depth structure equivalently. It was proposed in chapter 3 that snow depth development converges to a similar final snow distribution. However, time development of fractal parameters suggests that even at the end of the accumulation season snow depth structure was highly altered by individual NW storms. Hence, the inter-annual consistency observed between two years might be strongly dependent on the frequency of dominant NW storms in an accumulation season.

Up to now, three areas – one in this study and two areas investigated by Deems et al. (2008) – show a large interannual consistency. Future work will address the question if years or areas with less interannual consistency can be found in Alpine terrain and if reasons for those differences can be given, e.g. the dominance of not only one precipitation event.

Furthermore, future work will involve snow depth scaling investigation for larger areas and all possible sub-areas such as steep rock walls and other surface features. An automatic procedure to generate consistent sub-areas using clustering will be attempted. The ultimate goal is to describe mean snow depth and snow depth variations based on simple terrain parameters and overall precipitation information. This would be a major step forward in the field, in which the basic question on

"how much snow is on the mountain, where, when and why" is currently unanswered. Our research will complete our picture of the earth's surface roughness without snow (Mark and Aronson, 1984; Perron et al., 2008; Abedini and Shaghaghian, 2009, e.g.) and with snow. In particular, we want to characterise snow surface roughness from the scale of the individual snow grain (Manes et al., 2008) to the scale of snow drifts investigated here.
From a modelling point of view, future work will address on more physically orientated models using terrain influenced wind fields as input for saltation, suspension and sublimation models and their interaction with modelled snow cover properties (e.g. Lehning et al., 2008). The physical model approach allows investigating individual processes and is therefore complementary to the pure description of snow depth variability presented here. One next step is therefore to apply the process analysis as presented by Mott and Lehning (accepted manuscript, 2010) to the more complex setting in this new TLS laser scanner data set.
From our TLS measurements, snow distribution data is also existent for the ablation period together with runoff information. Therefore, a comparison of model performance in the ablation period will soon be presented elsewhere. The subject of this further work will be on the influence of the different energy fluxes to local snow melt with a particular focus on later energy fluxes caused by the heterogeneous snow surface. Uniform melting of a random snow depth distribution was shown to be sufficient to reproduce simple statistics of snow depth development during an ablation season (Egli et al., manuscript submitted, 2010). It is now possible to investigate whether the observed snow depth structure is a prerequisite for reliable runoff modelling.

6.3 Spatial variability and avalanche danger

To revisit the prediction of avalanche danger, two aspects on how spatial variability of modelled snow parameters can be included in future work will be given. Long term modelling, which is necessary for statistical learning methods, can firstly be achieved with the simple model approach using Winstal's terrain parameter. However, the observed dependency on prevailing wind direction would need to be addressed, especially for areas in which wind is found to be largely divergent. Coupling the terrain parameter with mesoscale wind fields and/or recalibrating empirical formulas as used here (e.g. Ryan, 1977) to more complex terrain conditions could be a solution to this problem. Secondly, long term modelling can also be achieved with more physically based models, since the studies presented have shown that snow depth distribution is dominated by a few "master" storms. This fact greatly simplifies model demands.
For understanding the causes of spatial variability and its effect on avalanche formation other parameters, such as snow depth, can now be measured faster and more objectively. For example, a snow stability interpretation of the SnowMicroPen signal is now available (Bellaire et al., 2009). It would be of special interest if stability or other related snow parameters obtain a similar scaling behaviour as snow depth.
The snow depth structure found in the slope in which a large avalanche was released might favour

fracture propagation over large distances. It would thus be of interest to show if in other slopes, where such large avalanches occurred, a similar scaling behaviour of snow depth can be obtained. Thus, a mapping of especially dangerous slopes may possible based on snow depth structure at the end of the accumulation season. This is of particular interest since in the coming years area-wide snow depth measurements will be available.

List of Tables

List of Figures

References

Abedini, M., Shaghaghian, M. (2009): Exploring scaling laws in surface topography. *Chaos, Solitons and Fractals*, 42 (4), 2373–2383.

Anderton, S., White, S., Alvera, B. (2004): Evaluation of spatial variability in snow water equivalent for a high mountain catchment. *Hydrological Processes*, 18 (3), 435–453.

Arnold, N. S., Rees, W. G. (2003): Self-similarity in glacier surface characteristics. *Journal of Glaciology*, 49 (8), 547–554.

Bartelt, P., Lehning, M. (2002): A physical SNOWPACK model for the Swiss avalanche warning. Part I: numerical model. *Cold Regions Science and Technology*, 35 (3), 123–145.

Bellaire, S., Pielmeier, C., Schneebeli, M., Schweizer, J. (2009): Instruments and Methods Stability algorithm for snow micro-penetrometer measurements. *Journal of Glaciology*, 55 (193), 805–813.

Bishop, C. (2006): Pattern Recognition and Machine Learning. Springer, New York.

Blöschl, G., Kirnbauer, R. (1992): An analysis of snow cover patterns in a small Alpine catchment. *Hydrological Processes*, 6, 99–109.

Brabec, B., Meister, R. (2001): A nearest-neighbor model for regional avalanche foresting. *Annals of Glaciology*, 32, 130–134.

Breiman, L., Friedman, J., Olshen, R., Stone, C. (1998): Classification and Regression Trees. CRC Press, Boca Raton, U.S.A.

Burrough, P. (1981): Fractal dimensions of landscapes and other environmental data. *Nature*, 294 (5838), 240–242.

Burrough, P. (1993): Fractals and geostatistical methods in landscape studies. In: Lam, N., De Cola, L. (editors), Fractals in Geography, 87–121, PTR Prentice Hall.

Buser, O. (1983): Avalanche forecast with the method of nearest neighbours: an interactive approach. *Cold Regions Science and Technology*, 8 (2), 155–163.

Campbell, C., Jamieson, B. (2007): Spatial variability of slab stability and fracture characteristics within avalanche start zones. *Cold Regions Science and Technology*, 47 (1-2), 134–147.

Chen, Y.-W., Lin, C.-J. (2006): Combining SVMs with various feature selection strategies. In: Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L. (editors), Feature Extraction: Foundations and Applications, 319–328, Springer, Berlin.

Dadic, R., Mott, R., Lehning, M., Burlando, P. (2010): Wind influence on snow depth distribution and accumulation over glaciers. *Journal of Geophysical Research*, 115, F01012.

Deems, J., Fassnacht, S., Elder, K. (2006): Fractal distribution of snow depth from LiDAR data. *Journal of Hydrometeorology*, 7 (2), 285–297.

Deems, J., Fassnacht, S., Elder, K. (2008): Interannual consistency in fractal snow depth patterns at two Colorado mountain sites. *Journal of Hydrometeorology*, 9 (5), 977–988.

Doorschot, J., Lehning, M., Vrouwe, A. (2004): Field measurements of snow-drift threshold and mass fluxes, and related model simulations. *Boundary-Layer Meteorology*, 113, 347–368.

Doswell, A., Davies-Jones, R., Keller, D. (1990): On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, 5, 576–585.

Durand, Y., Giraud, G., Brun, E., Mérindol, L., Martin, E. (1999): A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. *Journal of Glaciology*, 45 (151), 469–484.

Elder, K., Dozier, J., Michaelsen, J. (1991): Snow accumulation and distribution in an alpine watershed. *Water Resources Research*, 27 (7), 1541–1552.

Elman, J. (1990): Finding structure in time. *Cognitive Science*, 14 (2), 179–211.

Elsner, J., Schmertmann, C. (1994): Assessing forecast skill through cross validation. *Weather and Forecasting*, 9, 619–624.

Erickson, T., Williams, M., Winstral, A. (2005): Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States. *Water Resources Research*, 41, W04014.

Fassnacht, S., Williams, M., Corrao, M. (2009): Changes in the surface roughness of snow from millimetre to metre scales. *Ecological Complexity*, 6 (3), 221 – 229, special Section: Fractal Modeling and Scaling in Natural Systems.

Gauer, P. (2001): Numerical modeling of blowing and drifting snow in Alpine terrain. *Journal of Glaciology*, 47 (156), 97–110.

Goodchild, M., Mark, D. (1987): Review Article: The Fractal Nature of Geographic Phenomena. *Annals of the Association of American Geographers*, 77 (2), 265–278.

Grünewald, T., Schirmer, M., Mott, R., Lehning, M. (2010): Spatial and temporal variability of snow depth and SWE in a small mountain catchment. *The Cryosphere Discussions*, 4, 1–30.

Guyon, I., Elisseeff, A. (2006): An introduction to feature extraction. In: Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L. (editors), Feature Extraction: Foundations and Applications, 1–28, Springer, Berlin.

Harvey, S., Zweifel, B. (2008): New trends of recreational avalanche accidents in Switzerland. In: Campbell, C., Conger, S., Haegeli, P. (editors), ISSW 2008 Proceedings. (International Snow Science Workshop, Whistler, BC Canada, 21-27 September), 900–906, Whistler.

Heierli, J., Purves, R., Felber, A., Kowalski, J. (2004): Verification of nearest neighbours interpretations in avalanche forecasting. *Annals of Glaciology*, 38 (5), 84–88.

Helbig, N., Löwe, H., Lehning, M. (2009): Radiosity approach for the shortwave surface radiation balance in complex terrain. *Journal of the Atmospheric Sciences*, 66 (9), 2900–2912.

Hodgkins, R., Cooper, R., Wadham, J., Tranter, M. (2005): Interannual variability in the spatial distribution of winter accumulation at a high-Arctic glacier (Finsterwalderbreen, Svalbard), and its relationship with topography. *Annals of Glaciology*, 42 (6), 243–248.

Jamieson, B., Johnston, C. D. (2001): Evaluation of the shear frame test for weak snowpack layers. *Annals of Glaciology*, 32 (11), 59–69.

Jamieson, J., Johnston, C. (1993): Rutschblock precision, technique variations and limitations. *Journal of Glaciology*, 39 (133), 666–674.

Jamieson, J., Johnston, C. (1998): Refinements to the stability index for skier-triggered dry-slab avalanches. *Annals of Glaciology*, 26, 296–302.

Klinkenberg, B., Goodchild, M. (1992): The fractal properties of topography: a comparison of methods. *Earth Surface Processes and Landforms*, 17 (3), 217–234.

Kuchment, L., Gelfan, A. (2001): Statistical self-similarity of spatial variations of snow cover: verification of the hypothesis and application in the snowmelt runoff generation models. *Hydrological Processes*, 15 (18), 3343–3355.

Lam, N., De Cola, L. (1993): Introduction to fractals in geography. In: Lam, N., De Cola, L. (editors), Fractals in Geography, 3–22, PTR Prentice Hall.

Lehning, M., Bartelt, P., Brown, B., Fierz, C. (2002a): A physical SNOWPACK model for the Swiss avalanche warning. Part III: meteorological forcing, thin layer formation and evaluation. *Cold Regions Science and Technology*, 35 (3), 169–184.

Lehning, M., Bartelt, P., Brown, B., Fierz, C., Satyawali, P. (2002b): A physical SNOWPACK model for the Swiss avalanche warning. Part II: snow microstructure. *Cold Regions Science and Technology*, 35 (3), 147–167.

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stöckli, U., Zimmerli, M. (1999): SNOWPACK model calculations for avalanche warning based upon a network of weather and snow stations. *Cold Regions Science and Technology*, 30 (1-3), 145–157.

Lehning, M., Fierz, C. (2008): Assessment of snow transport in avalanche terrain. *Cold Regions Science and Technology*, 51 (2-3), 240–252.

Lehning, M., Fierz, C., Brown, B., Jamieson, B. (2004): Modeling snow instability with the snow-cover model SNOWPACK. *Annals of Glaciology*, 38 (8), 331–338.

Lehning, M., Löwe, H., Ryser, M., Raderschall, N. (2008): Inhomogeneous precipitation distribution and snow transport in steep terrain. *Water Resources Research*, 44, W07404.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., Zappa, M. (2006): ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes*, 20, 2111–2128.

Li, L., Wineberg, C., Darden, T., Pedersen, L. (2001): Gene selection for sample classification based on gene expression: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics*, 17, 1131–1142.

Liston, G. E., Haehnel, R. B., Sturm, M., Hiemstra, C. A., Berezovskaya, S., Tabler, R. D. (2007): Instruments and Methods Simulating complex snow distributions in windy environments using SnowTran-3D. *Journal of Glaciology*, 53 (16), 241–256.

Machguth, H., Eisen, O., Paul, F., Hoelzle, M. (2006): Strong spatial variability of snow accumulation observed with helicopter-borne GPR on two adjacent Alpine glaciers. *Geophysical Research Letters*, 33, L13503.

Mandelbrot, B. (1977): Fractals: Form, Chance and Dimension. W.H. Freeman and Company, San Francisco, CA.

Mandelbrot, B. (1982): The Fractal Geometry of Nature. W.H. Freeman and Company, New York.

Manes, C., Guala, M., Löwe, H., Bartlett, S., Egli, L., Lehning, M. (2008): Statistical properties of fresh snow roughness. *Water Resources Research*, 44 (11), W11407.

Mark, D., Aronson, P. (1984): Scale-dependent fractal dimensions of topographic surfaces: An empirical investigation, with applications in geomorphology and computer mapping. *Mathematical Geology*, 16 (7), 671–683.

McClung, D., Schaerer, P. (2006): The Avalanche Handbook. The Mountaineers, Seattle, Washington, 3 edition.

Meister, R. (1995): Country-wide avalanche warning in Switzerland. In: Proceedings of the International Snow Science Workshop (ISSW 1994), 30 October-3 November 1994, 58–71, Snowbird, Utah.

Murphy, A. (1991): Forecast verification: its complexity and dimensionality. *Monthly Weather Review*, 119 (7), 1590–1601.

Murphy, A. (1993): What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8, 281–293.

Murphy, A., Winkler, R. (1987): A general framework for forecast verification. *Monthly Weather Review*, 115 (7), 1330–1338.

Pachepsky, Y., Ritchie, J. (1998): Seasonal changes in fractal landscape surface roughness estimated from airborne laser altimetry data. *International Journal of Remote Sensing*, 19 (13), 2509–2516.

Pentland, A. (1984): Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 661–674.

Perron, J., Kirchner, J., Dietrich, W. (2008): Spectral signatures of characteristic spatial scales and nonfractal structure in landscapes. *Journal of Geophysical Research*, 113.

Pozdnoukhov, A., Purves, R., Kanevski, M. (2008): Applying machine learning methods to avalanche forecasting. *Annals of Glaciology*, 49 (7), 107–113.

Prokop, A., Schirmer, M., Rub, M., Lehning, M., Stocker, M. (2008): A comparison of measurement methods: terrestrial laser scanning, tachymetry and snow probing for the determination of the spatial snow-depth distribution on slopes. *Annals of Glaciology*, 49 (7), 210–216.

Purves, R., Barton, J., Mackaness, W., Sugden, D. (1998): The development of a rule-based spatial model of wind transport and deposition of snow. *Annals of Glaciology*, 26, 197–202.

Purves, R., Morrison, K., Moss, G., Wright, D. (2003): Nearest neighbours for avalanche forecasting in Scotland – development, verification and optimisation of a model. *Cold Regions Science and Technology*, 37 (3), 343–355.

Rabiner, L. R. (1989): A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77 (2), 257–286.

Ryan, B. (1977): A mathematical model for diagnosis and prediction of surface winds in mountainous terrain. *Journal of Applied Meteorology*, 16 (6), 571–584.

Schaffhauser, A., Adams, M., Fromm, R., Jörg, P., Luzi, G., Noferini, L., Sailer, R. (2008): Remote sensing based retrieval of snow cover properties. *Cold Regions Science and Technology*, 54 (3), 164 – 175.

Schirmer, M., Lehning, M., Schweizer, J. (2009a): Statistical forecasting of regional avalanche danger using simulated snow-cover data. *Journal of Glaciology*, 55 (193), 761–768.

Schirmer, M., Mitterer, C., Schweizer, J. (2009b): On sample size and different interpretations of snow stability datasets. *Geophysical Research Abstracts*, 11, 13413.

Schirmer, M., Schweizer, J., Lehning, M. (2009c): Regional stability evaluation with modelled snow cover data. In: Schweizer, J., van Herwijnen, A. (editors), ISSW 2009 Proceedings. (International Snow Science Workshop, Davos, Switzerland, 27 September-2 Ocotber), 306–312, Davos, Switzerland.

Schölkopf, B., Smola, A. (2002): Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond. MIT Press, Cambridge, Massachusetts.

Schweizer, J., Bellaire, S., Fierz, C., Lehning, M., Pielmeier, C. (2006): Evaluating and improving the stability predictions of the snow cover model SNOWPACK. *Cold Regions Science and Technology*, 46 (1), 52–59.

Schweizer, J., Camponovo, C. (2001): The skier's zone of influence in triggering slab avalanches. *Annals of Glaciology*, 32 (7), 314–320.

Schweizer, J., Föhn, P. (1996): Avalanche forecasting - an expert system approach. *Journal of Glaciology*, 42 (141), 318–332.

Schweizer, J., Jamieson, B. (2003): Snowpack properties for snow profile interpretation. *Cold Regions and Science Technology*, 37 (3), 233–241.

Schweizer, J., Jamieson, B., Schneebeli, M. (2003a): Snow avalanche formation. *Review of Geophysics*, 41 (4), 1016.

Schweizer, J., Jamieson, J. (2007): A threshold sum approach to stability evaluation of manual snow profiles. *Cold Regions Science and Technology*, 47 (1-2), 50–59.

Schweizer, J., Kronholm, K. (2007): Snow cover spatial variability at multiple scales: characteristics of a layer of buried surface hoar. *Cold Regions Science and Technology*, 47 (3), 207–223.

Schweizer, J., Kronholm, K., Jamieson, J., Birkland, K. (2008a): Review of spatial variability of snowpack properties and its importance for avalanche formation. *Cold Regions Science and Technology*, 51 (2-3), 253–272.

Schweizer, J., Kronholm, K., Wiesinger, T. (2003b): Verification of regional snowpack stability and avalanche danger. *Cold Regions Science and Technology*, 37 (3), 277–288.

Schweizer, J., Lütschg, M. (2001): Characteristics of human-triggered avalanches. *Cold Regions Science and Technology*, 33 (2-3), 147–162.

Schweizer, J., McCammon, I., Jamieson, J. (2008b): Snowpack observations and fracture concepts for skier-triggering of dry-snow slab avalanches. *Cold Regions Science and Technology*, 51 (2-3), 112–121.

Schweizer, J., Wiesinger, T. (2001): Snow profile interpretation for stability evaluation. *Cold Regions Science and Technology*, 33 (2-3), 179–188.

Schweizer, M., Föhn, P., Schweizer, J. (1994): Integrating neural networks and rule based systems to build an avalanche forecasting system. In: Hamza, M. (editor), Int. Conf. on Artificial Intelligence, Expert Systems and Neuronal Networks. International Association of Science and Technology for Development (IASTED), Zurich.

Shook, K., Gray, D. (1996): Small-scale spatial structure of shallow snowcovers. *Hydrological Processes*, 10 (10), 1283–1292.

Sovilla, B., Burlando, P., Bartelt, P. (2006): Field experiments and numerical modeling of mass entrainment in snow avalanches. *Journal of Geophysical Research*, 111, F03007.

Sovilla, B., Margreth, S., Bartelt, P. (2007): On snow entrainment in avalanche dynamics calculations. *Cold Regions Science and Technology*, 47 (1-2), 69–79.

Sun, W., Xu, G., Gong, P., Liang, S. (2006): Fractal analysis of remotely sensed images: a review of methods and applications. *International Journal of Remote Sensing*, 27 (21-22), 4963–4990.

Trujillo, E., Ramírez, J., Elder, K. (2007): Topographic, meteorologic, and canopy controls on the scaling characteristics of the spatial distribution of snow depth fields. *Water Resources Research*, 43, W07409.

Trujillo, E., Ramírez, J., Elder, K. (2009): Scaling properties and spatial organization of snow depth fields in sub-alpine forest and alpine tundra. *Hydrological Processes*, 23 (11), 1575–1590.

Webster, R., Oliver, M. (2007): Geostatistics for Environmental Scientists. Wiley.

Weng, Q. (2003): Fractal analysis of satellite-detected urban heat island effect. *Photogrammetric Engineering and Remote Sensing*, 69 (5), 555–566.

Wilks, D. (1995): Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego.

Winstral, A., Elder, K., Davis, R. (2002): Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *Journal of Hydrometeorology*, 5, 524–538.

Winstral, A., Marks, D. (2002): Simulating wind fields and snow redistribution using terrrain based parameters to model snow accumulation and melt over a semi-arid mountain catchment. *Hydrological Processes*, 16, 3585–3603.

Xu, T., Moore, I. D., Gallant, J. C. (1993): Fractals, fractal dimensions and landscapes – a review. *Geomorphology*, 8 (4), 245–262.

Acknowledgements

I wish to thank all the supervisors of my dissertation. Firstly, Michael Lehning at the WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland, who provided continuous support on the ongoing work. Secondly, Jürg Schweizer at the SLF, who gave valuable comments, especially to snow topics, statistical methods and field work. Thirdly, Christoph Schneider at the RWTH Aachen, who was chief supervisor of the present dissertation.
I am very grateful to all my colleagues at the SLF for stimulating working environment, critical feedback and great team work (in no particular order): Rebecca Mott, Vanessa Wirz, Barbara Landl, Nora Helbig, Christine Groot Zwaaftink, Ingrid Reiweger, Yvonne Schaub, Thomas Grünewald, Christoph Mitterer, Luca Egli, Jan Magnusson, Andy Clifton, Jake Turner, Nick Dawes, Henning Löwe, Mathias Bavay, Andreas Stoffel, Michel Bovey, Roland Meister, Alec von Hervijnen and Christoph Gromke. Especially, I want to acknowledge Charles Fierz for his support in snowpack modelling and for interesting discussions. Furthermore, my thanks are due to Martin Hiller, Franz Herzog, Andreas Moser, Christian Simeon and Flury Michel for their great work on equipping the study area with weather stations.
I owe special thanks to Alexander Prokop at the BOKU in Vienna, who initiated terrestrial laser scanning in my dissertation, and to Tobias Sauter at the RWTH Aachen, who helped me with artificial neural networks.
Finally, I extend my thanks to all my friends, especially those in Davos for their company in various sports activities, which were important to get away from hard work and to come back, relaxed.