GEOGRAPHIC KNOWLEDGE DISCOVERY TECHNIQUES FOR EXPLORING HISTORICAL WEATHER AND AVALANCHE DATA

by

Christopher Michael McCollister

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Earth Sciences

MONTANA STATE UNIVERSTIY
Bozeman, Montana

May 2004

APPROVAL

Of a thesis submitted by

Christopher Michael McCollister

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Katherine J. Hansen
(Co-Advisor)

Karl W. Birkeland
(Co-Advisor)

Approved for the Department of Earth Sciences

David R. Lageson

Approved for the College of Graduate Studies

Bruce R. McLeod

STATEMENT OF PERMISSION TO USE

Christopher Michael McCollister

May 14, 2004

ACKNOWLEDGEMENTS

I would like to thank my advisors, Kathy Hansen, Karl Birkeland, and Richard Aspinall for their patience, support, insight, and motivation. I would also like to thank the Jackson Hole Mountain Resort for its cooperation and indirect financial support, particularly Tom Spangler, Corky Ward, the entire ski patrol, and the forecasters at the Bridger Teton National Forest Avalanche Center. I would like to express my extreme gratitude to Bob Comey and the Bridger Teton National Forest Avalanche Center interns for their tireless work of entering data. I would like to express my gratitude to my family for their financial assistance, without which this is research could not have taken place. The Association of American Avalanche Professionals, the Milton J. Edie scholarship, and the Montana State University Earth Sciences Department also provided financial support. I would also like to thank Dr. Gary Harkin of the Montana State University Computer Science Department for his insightful discussions on pattern recognition. Finally, I extend a hearty thanks to Gary Poulson, Jim Kanzler, Larry Livingood, and all the unnamed avalanche workers who have diligently recorded data at Jackson Hole Mountain Resort over the past three decades.

TABLE OF CONTENTS

TABLE OF CONTENTS-CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES-CONTINUED

ABSTRACT

Many ski areas, backcountry avalanche centers, highway departments, and helicopter ski operations record and archive daily weather and avalanche data. The objective of this thesis is to present probabilistic techniques that allow avalanche forecasters to better utilize weather and avalanche data by incorporating a Geographic Information System with a modified meteorological nearest neighbors approach. This nearest neighbor approach utilizes evolving concepts related to visualizing geographic information stored in large databases. The resulting interactive database tool, Geographic Weather and Avalanche Explorer, allows the investigation of the relationships between specific weather parameters and the spatial pattern of avalanche activity. In order to validate these new techniques, two case studies are presented using over 10,000 individual avalanche events from the past 23 years that occurred at the Jackson Hole Mountain Resort.

The first case study explores the effect of new snowfall, wind speed, and wind direction on the spatial patterns of avalanche activity. Patterns exist at the slide path scale, and for groups of adjacent slide paths, but not for either the entire region as a whole or when slide paths are grouped by aspect. Since wind instrumentation is typically located to measure an approximation of the free air winds, specific topography around a given path, and not aspect, is more important when relating wind direction to avalanche activity.

The second case study explores the spatial variability of hard slab and dry loose avalanches, and characterizes these avalanche types with respect to their geographic location and associated weather conditions. I analyzed these data with and without the incorporation of three weather parameters (wind speed, 24-hour maximum temperature, and new snow density). Slide paths near each other often had similar proportions of hard slabs and a higher proportion of hard slabs occurred on exposed ridges. The proportion of loose avalanches also was similar for adjacent slide paths, and these paths were typically sheltered from strong winds. When I incorporated the three weather parameters I found significant increases in the average proportion of hard slabs with increases in new snow density, but not for changes in the 24-hour maximum temperature or wind speed. When I analyzed the proportion of loose avalanches associated with the three weather parameters I found a more direct relationship than with hard slabs. Changes in both wind speed and density significantly changed the average proportion of loose avalanches, with low wind and low density resulting in higher proportions of loose avalanches. My results quantify what operational avalanche forecasters have long known: Geographic location and weather are both related to the proportion of hard slab and dry loose avalanches.

INTRODUCTION

Avalanches are dangerous. Besides threatening roads, structures, and ski areas, avalanches have killed a total of 703 people in the United States from the winter season of 1950-51 to 2003-04 (Northwest Avalanche Center, 2004). From the 1910s to the end of the current season (2003-04) the state of Wyoming, where this research took place, has experienced a total of 54 deaths caused by avalanches (Figure 1). The cause of these fatal avalanches often involves human-error and they may have been avoidable, however some occur due to a lack of scientific understanding about avalanche processes and the timing and location of avalanches. The avalanche community needs to better understand the spatial and temporal patterns of avalanches in order to improve its ability to forecast their occurrence and avoid the dangers they pose.

Forecasting today utilizes direct observations, field tests, and the analysis of meteorological data. However, improvements are needed in our ability to visualize and therefore understand these spatial patterns of avalanches. This study provides those improved techniques and a variety of visual tools for forecasters and researchers. Additionally, current data analysis techniques are not user-friendly. Therefore, while many snow safety operations have collected weather and avalanche data, most are not analyzing them. I have created a program to address these problems.

A scientific understanding of avalanches, as well as knowledge of the local patterns of avalanche activity (gained through experience) is crucial for avalanche forecasters (McClung, 2002a). The former can be taught, but the latter is much more difficult to teach, communicate, or even define. For example, how new snow, wind speed, and wind

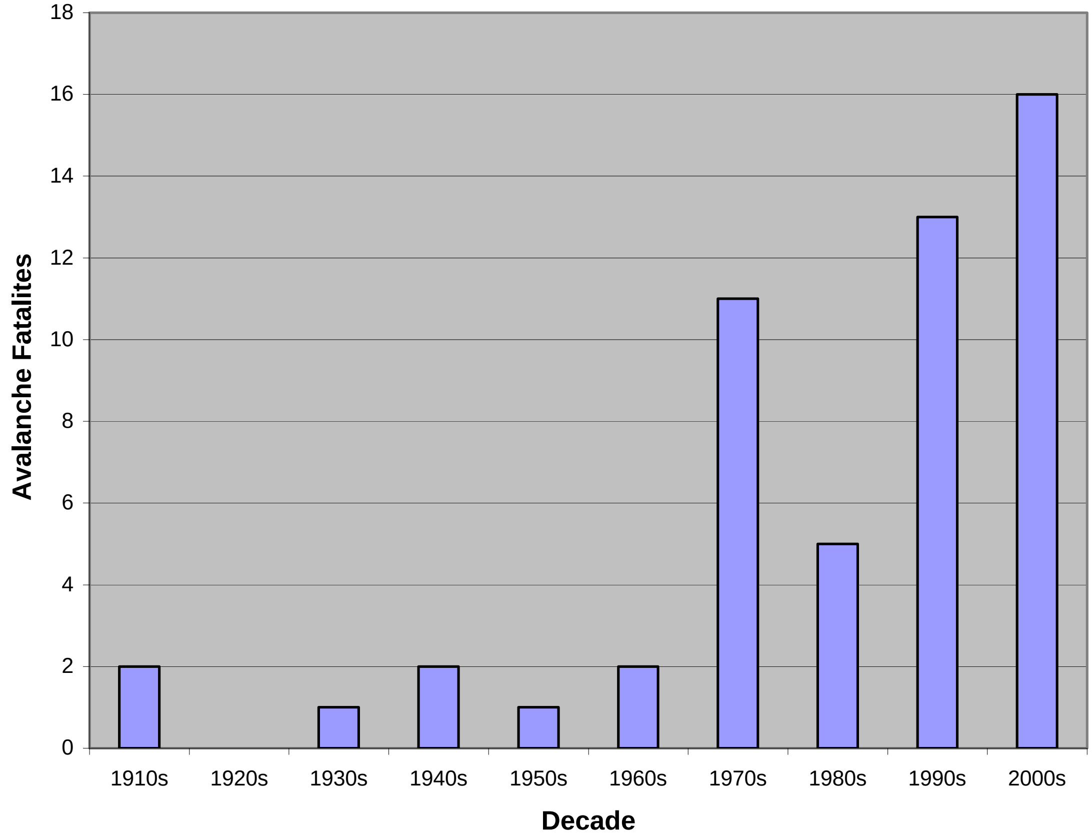


Figure 1: Wyoming Avalanche Fatalities by Decade from 1911-2003.

direction conceptually lead to selective wind loading and the formation of slab avalanches is relatively easy to teach and understand. However, an understanding of which *particular* slide paths load under which conditions of new snow, wind speed, and wind direction requires additional knowledge that may require decades of local individual observations and experience.

Objective and Hypothesis

This thesis presents new methods to better utilize historical weather and avalanche data to enable the visualization of avalanche probabilities and the generation of hypotheses about the spatial and temporal pattern of avalanches. My hypothesis is that meteorological variables play a predictive role in creating regional spatial patterns of avalanche activity, and that different weather conditions create different patterns of avalanche activity. For example, freezing lines may segregate wet avalanches from dry avalanches, or changes in wind direction may load different slopes.

To address this objective I have created a software program to analyze historical weather and avalanche data by combining a meteorological nearest neighbors technique with a Geographic Information System (GIS) using Geographic Knowledge Discovery (GKD) concepts. The result is a user-friendly database tool to visualize data via dynamic maps and graphs, thereby taking advantage of the pattern recognition capacity of humans. This is only a tool for the avalanche forecaster or researcher to help digest and interact with large amounts of data. It is not a replacement for an avalanche forecaster.

To verify these methods I present two case studies. In the first study I investigated relationships between new snowfall, wind speed, and wind direction on the avalanche activity at the Jackson Hole Mountain Resort, Wyoming for the entire ski area, for sub-regions, for groups with similar aspect and elevation, and for individual avalanche paths. I chose the three variables of new snowfall, wind speed, and wind direction to define the factor of wind loading. These variables are known to play a major role in wind loading. Wind loading is a primary factor for avalanches, especially in a ski area setting where daily disruption of snow pack layers occurs via skier compaction and avalanche reduction measures.

The second study builds upon the first. It compares two avalanche types (hard slab avalanches and loose avalanches) with wind speed, 24-hour maximum temperature, and new snow density. These three weather variables may contribute to the formation, or lack there of, of hard slabs. An increase in wind speed may increase hardening of the new snow leading to an increase in hard slabs. Warmer temperatures increase settlement rates, and may thereby increase the formation of hard slabs. Finally, an increase in new snow density may relate to denser harder slabs. Both case studies are presented here in their publication format. They stand as separate case studies; however, there is some repetition between the two.

LITERATURE REVIEW

Geographic Pattern Exploration

Geographic pattern exploration is a major feature of this research. It provides probabilities of avalanches in a geographic, spatial manner. Five areas of exploration have been utilized in the past, including exploratory data analysis (EDA), scientific visualization (SciVis), geographical visualization (GVis), knowledge discovery in databases (KDD), and geographic knowledge discovery (GKD). Tukey (1977) described EDA as numerical detective work. EDA had its roots in statistics with the fundamental goal of identifying interesting relationships in data (Wachowicz, 2001). Today, EDA techniques are very visual in nature (Andrienko *et al*., 2001) and share similarities with scientific visualization (SciVis). "One of the more interesting ways to define scientific visualization is to define it as computationally intensive visual thinking" (Rhyne, 2000, p. 20). GVis was an extension of SciVis with a geographic (spatial) component, while KDD was a formalization of the process to extract meaningful data from large databases (Fayyad *et al*., 1996a). GKD was the geographic extension of KDD and is also a process. As these fields of exploration matured, they have become more interrelated. An estimated 80% of all digital data has a spatial component (MacEachren and Kraak, 2001), so computer scientists have starting to address the need for spatial location to be incorporated in their modeling techniques (Han *et al*., 2002; Fayyad, 1996b). Likewise, integrating the concepts of EDA and KDD with the spatial concepts of geography have become primary research agendas that have led to several journal issues completely devoted to these concepts (MacEachren and Kraak, 1997; 2001; Kraak and MacEachren, 1999).

These fields of exploration, particularly GVis, KDD, and GKD, also share a primary goal of finding patterns and relationships in large spatial datasets. GVis and KDD have several underlying concepts in common (MacEachren *et al*. 1999). First, both fields involve the *interaction* of computers and humans and see this interaction as a process, attempting to capitalize on the strengths of both (Miller and Han, 2001; MacEachren *et al*. 1999; Andrienko and Andrienko, 1999; Ramakrishnan and Grama, 1999; Fayyad *et al*., 1996b; Hibbard and Santek, 1989). Second, both GVis and KDD utilize *iteration,* which allows visualization of patterns with different attributes, at different times, or at different scales that may illuminate trends that would not be obvious in a static view (Andrienko *et al*., 2001; Ramakrishnan and Grama, 1999; MacEachren *et al*. 1999). Iteration is also familiar to avalanche forecasters, who typically use iteration while forecasting to reduce uncertainty and improve forecast accuracy (LaChapelle, 1980). Third, these share a high interactivity between the user and computer allowing the user to pose "*what if*" questions for hypotheses generation (Gahegan *et al*., 2001; MacEachren *et al*. 1999). Finally, commonalities of *multiple perspectives* allow the user to view the data at different scales, measures, or different factors (Andrienko *et al*., 2001; MacEachren *et al*. 1999).

GVis is different from the other exploratory analyses in that the data must have a geographic component and representations of the data employ the human eye-brain ability to visually recognize and identify patterns. MacEachren (1992, p. 101) defines

GVis as "the use of concrete visual representations – whether on paper or through computer displays or other media – to make spatial contexts and problems visible, so as to engage the most powerful of human information processing abilities, those associated with vision." In contrast, KDD is a process, consisting of data selection, preprocessing, transformation, data mining methods and algorithms, interpretation, and evaluation, with the underlying goal of extracting meaningful patterns from large databases (Fayyad *et al.*, 1996a). GKD is the geographic extension of KDD (Miller, 2001), and is also a process of finding interesting patterns in data, but with the added complexity of geographic relationships being embedded in the data. Many of the data mining methods used in traditional KDD assume that all variables are independent. This is not the case with spatial data, where spatial autocorrelation is an intrinsic part of the system.

Avalanche Forecasting

Avalanche forecasting utilizes inductive and deductive reasoning along with data and knowledge from experience to reduce the uncertainty of the avalanche hazard for a given area (LaChapelle, 1980; McClung, 2002a; 2002b). McClung (2000) groups avalanche forecasting into three types based on the size of the forecast area. Type A forecasts provide more general information for large forecast areas at the mountain range scale. Type B forecasts are more specific and are typically at the scale of a ski area or highway operations. At the slope scale, type C forecasts are the most specific of the three, and are made by heli-guides, backcountry guides, and backcountry skiers.

LaChapelle (1980) categorized avalanche-forecasting data as direct stability data (Class I), snow-pack structure data (Class II), or meteorological data (Class III).

Direct stability data are explicit evidence of avalanches or failures within the snow pack. This type of data includes observations of avalanche events, failure of the snow pack without avalanches, commonly known as "whomping", and direct stability tests such as rutschblock (Fohn, 1987) or stuff block tests (Birkeland *et al.,* 1996). Additional tests include the compression test (Jamieson and Johnston, 1996) and the quantified loaded column stability test (Landry *et al*., 2001). These data are the most relevant of the three types of data for slope scale avalanche forecasting, and lead to the lowest uncertainty forecast because it is results from direct observations. However, considerable uncertainty can also exist in these data (Landry *et al*., in press). These data are also used in Type A and B forecasts, but may be predicted instead of observed.

Snow pack structure data reveal the physical structure of the snow pack and have been traditionally defined as the data measured in snow pit profiles. Some studies that have assessed structure have used temperature profiles (Deems, 2003), snow hardness profiles (Kozak *et al.,* 2003), snow crystal size (Cooperstein, *et al.* in preparation) and type profiles (Birkeland, 1998 and Birkeland *et al*., 1998). These data are used to find weak layers in the snow pack such as surface hoar layers, depth hoar, and near surface faceting, resulting in a forecast with more uncertainty than with direct stability data. All three forecast types use these data, but it is often predicted for type A forecasts, and observed for type B and C forecasts. Additional structure data include outputs from snow pack models such as SNOWPACK (Lehning *et al.,* 1998) or SAFRAN-CROCUS-

MÉPRA (Durand *et al.,* 1999). The usefulness of these data has come into question in recent years. Several studies investigating the spatial variability of the snow pack have found high variability in the snow pack over short distances (Landry *et al.,* in press; Birkeland *et al.,* in press; Kozak *et al.,* 2003; Birkeland, 2001; and Birkeland *et al.,* 1995). This leads to the question of how relevant a snow pit is for determining the stability of a nearby slope. These data become even less reliable as the scale goes from slope scale to mountain range scale with the inherent increase in physical distances between a snow pit and an avalanche path.

Meteorological data include, but are not limited to, precipitation (new snow, snow water equivalent, and snow depth), wind (speed and direction, and maximum gust), temperature (maximum, minimum, and mean), relative humidity, and solar input. These weather measurements are usually taken at multiple locations and are often automated. Typically, these data are used in real time and are incorporated into the day's forecast. Of the three types of data, meteorological data leads to avalanche forecasts with the most uncertainty due to the inherent uncertainty of weather forecasts and the actual weather's effect on the snow pack. Meteorological data are typically integrated into an avalanche forecast by applying learned knowledge about the consequences of certain weather events. The data are usually used for all three forecast types and are typically predicted for type A forecasts and observed for type B and C forecasts. Meteorological data are most important for mountain scale, type A forecasts with many varying geographic zones (i.e. freezing elevations, snow accumulations, etc). When these data are recorded and archived, they can be analyzed to gain intrinsic knowledge about the local area.

Meteorological data are well suited for avalanche research because they are directly related to historical avalanche data via the date. Additionally, they are readily available and highly abundant. The volume of these data is increasing exponentially as a function of time and amount of data being recorded, due primarily to the automation of data collection. Each year more data are being recorded by increasing the types of measurements, adding new data collection site locations, and increasing the rate of taking measurements. For example, today the Bridger Teton National Forest Avalanche Center acquires up to 5664 daily weather measurements from automated remote sites alone. For the daily forecast, the automated data are reduced to 174 pieces of summary data, such as minimums, maximums, ranges, and time of maximums. Finally, 111 pieces of data are obtained manually, including verification of precipitation measurements and several subjective pieces of data, such as the amount of snow available for transport and new and old snow surface types for a variety of aspects and elevations. Combining these three types of data results in 5949 pieces of data recorded each day. It is nearly impossible to process and analyze this volume of data manually. The methods developed by this research allow the application of a computer to help process these large amounts of data.

Numerical Analysis Methods for Meteorological Data

A number of techniques have been applied towards determining the relationship between weather and avalanches including Classification and Regression Trees (CART), discriminant analysis, cluster analysis, and nearest neighbors analysis. Davis *et al.*

(1996) presented an example of CART, and Obled and Good (1980) presented an overview and comparison of the last three methods.

The classification and regression tree method (Elder and Davis, 2000; Davis *et al*., 1999; Davis *et al*., 1996; Davis and Elder, 1994) is a series of binary questions ultimately resulting in a yes or no answer about the possibility of an avalanche occurrence. The binary questions are a tree-structured set of connected nodes. The nodes are either a threshold function (i.e. Max 24 hr temp < 15.1º C) or an end node, which designates the classification (a day with avalanches or a day without avalanches). The classification process starts at the trunk and each fork of a branch represents the next threshold function until a leaf (end node) is reached designating the specific classification. To create the tree, a learning dataset is needed. It is possible to perfectly predict the learning set if an unlimited number of nodes are used. However, this perfect tree will not predict a test dataset, and needs to be pruned, which is simply limiting the number of nodes.

Discriminant analysis is a method to partition, or discriminate between two classes, in this case, a day with or without avalanches. The variables with the most discriminating ability are determined and become the components of the discriminant axis. The discriminant axis is a multi-variable vector used to partition the two classes.
Obled and Good (1980) use a non-parametric form of discriminant analysis to create the discriminant axis (first eigenvector) by using a calibration data set, which was then tested using a separate test dataset.

Cluster analysis is a two-step process where days are first grouped into clusters typically related to snow and weather conditions. A unique discriminant analysis is then performed on each cluster to differentiate between the two classes (avalanche vs. non-avalanche).

Nearest neighbors is an intuitive search method to find historical days in a database that are similar to a day of interest, or a target day. The target day is defined by a set of variables chosen as search variables, which can be weighted. Next, a distance measurement is performed between the target day and all historical days. Typically this distance measurement is a simple Euclidean distance. Ordering of the historical days is based on this distance, and a set number (k) of nearest days becomes a subset associated with the target day. This technique is called the k-nearest-neighbor classifier, first formalized by Cover (1967). A probability of avalanches can be determined by the percentage of days with an avalanche, or a Boolean classification can be used based on the percentage of avalanche days (i.e. an avalanche day is a day with > 30% avalanche days).

Nearest neighbor methods are non-parametric, memory-based techniques (Hand, *et al*. 2001). They are non-parametric in the sense that they do not have any underlying assumptions about normality. They are memory based because they are a direct output of historical data, as opposed to deterministic techniques, where a function is developed to describe the relationship numerically. Dasarathy (1991) presented a historical review of the relevant nearest neighbor literature.

Some of the avalanche models that use nearest neighbors optimize the variable weights. Gassner *et al*. (2000) do this by using a local expert to set the weights (based on perceived importance), and then create a measure of correctness to compare different weighting schemes. Purves *et al*. (2002) used a genetic algorithm to determine optimal weights.

Common Analysis Problems

These methods all have drawbacks. First, they do not account for the geographic component of slide paths, which experience has shown to be quite important. Second, they typically do not analyze the data at the individual slide path scale, which is of primary importance to ski patrollers and others doing avalanche hazard reduction work. Finally, they usually treat a day as either a day with avalanches or without. As a result of this type of classification, most are not probabilistic in nature. This is misleading because some days may have many large and significant avalanches, while another day may have only one small slide, and yet both these days would be classified as equivalent 'avalanche days'. Probabilities give much more information to the forecaster than this simple Boolean classification.

Nearest Neighbor Analysis Unique Problems

Although nearest neighbor techniques are the most widely used numerical avalanche forecasting tools, they have their own set of unique problems. One problem is in determining how many of the nearest days to incorporate. Many nearest neighbor programs use ten nearest historical days (Purves *et al*., 2002; Brabec and Meister, 2001); Gassner *et al*., 2000; Kristensen and Larsson, 1994; and Buser, 1983), but those ten days may or may not be good representations of the target day. For example, if I am searching for the nearest neighbors of a target day with rare conditions, its neighbors will be very distant and unrepresentative of target day. In contrast a target day with common conditions will have closer neighbors and be more representative.

Another possible problem is the number of variables considered in the analysis. If too many variables are included, the analysis may become altered. Gahegan (1999) discussed the problem with using a large number of variables - as is tempting with the large dimensionality (high number of potential variables) of some spatial databases. In these large databases, calculations quickly become unmanageable. This is known as the *curse of dimensionality* (Hand *et al.,* 2001). Nearest neighbor techniques also suffer dimensionality problems. As the number of dimensions (variables) gets large, the data become more and more spread out. This leads to the distance of *nearest* neighbor becoming similar to the distance of the *most distant* neighbor, with both eventually approaching each other as the number of dimensions gets large (Hand *et al*., 2001; Hinneburg *et al*., 2000; Beyer *et al*., 1999). When the distance of the nearest neighbor equals the distance of the most distant neighbor, the definition of neighbor becomes meaningless. Beyer *et al*. (1999) found this to occur with as few as 10 to 15 variables under conditions of independent and identical distributions (IDD). Most previous nearest neighbor applications for analyzing avalanche datasets use between 11 and 22 variables (e.g., Purves *et al*. (2002) used 11, Brabec and Meister (2001) used 12, Gassner *et al*. (2000) used 12 and 20, Kristensen and Larsson (1994) used 22, and Buser (1983) used 13).

GEOGRAPHIC EXPLORATION AND AVALANCHE FORECASTING

Effect of High Dimensions on Weather and Avalanche Data

To demonstrate the possible effect of high dimensions on a weather and avalanche data set, I performed a simple test by varying the number of dimensions (variables), and calculated the ratio of the most distant neighbor to the nearest neighbor as done by Beyer *et al*. (1999). If there was high contrast, this ratio was large. If there was no contrast, this ratio was one. The mean of each variable was used as that variable's target value. Comparing the ratios of the most distant neighbor to the most similar neighbor for different dimensions demonstrated how increased dimensionality decreased the distance between the nearest and most distant neighbor. There was a full order of magnitude of difference between one and five dimensions, and the difference leveled off between five and 20 dimensions with a ratio of around 10-15 to 1 (Figure 2).

Differences from Traditional Nearest Neighbor Methods

My approach differed from previous studies in both the goal and in the specific methodology. My primary goal was to improve avalanche forecasting by enhancing the forecaster's interaction with large datasets, and by creating a tool to visualize, explore, and ask questions of the data in order to find spatial patterns. The ideal tool would incorporate geography, be probabilistically based, and be useful for analyzing avalanche data at different scales (ranging from an individual slide path to the entire region). In

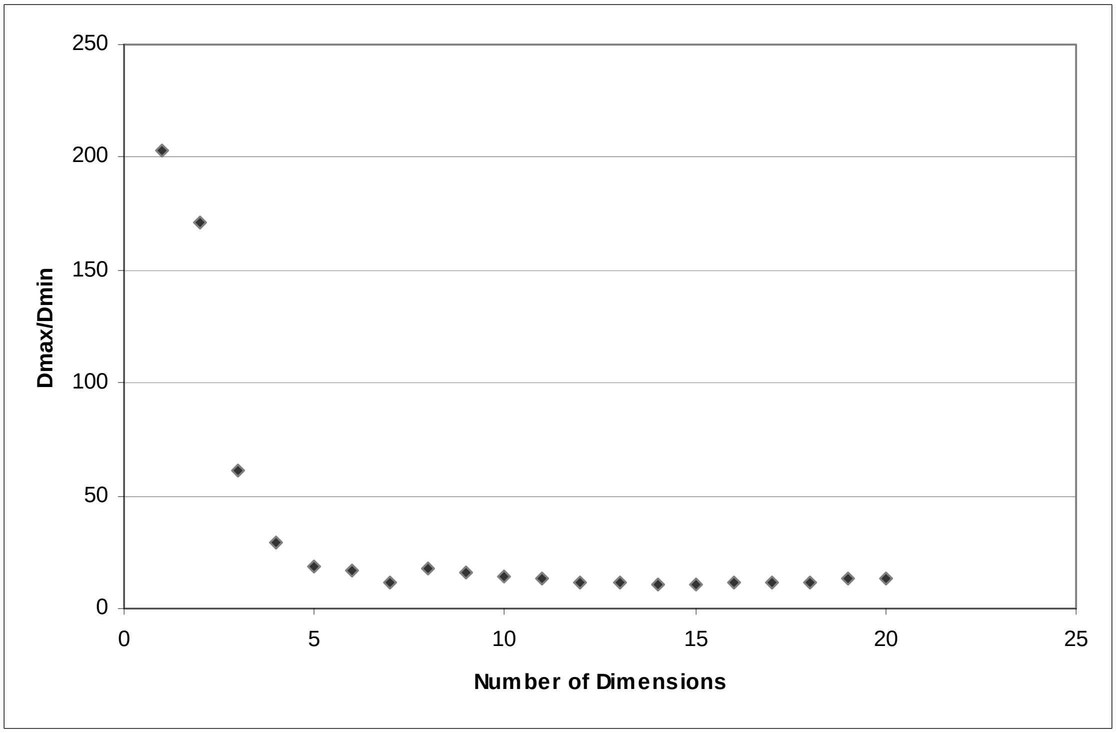


Figure 2: Ratio of the Furthest Neighbor Distance to the Nearest Neighbor Distance.

addition, my secondary goal was to facilitate hypothesis generation and testing which was used to produce the two case studies.

My methods differed from other nearest neighbor techniques used in avalanche forecasting in four ways. First, I used only three variables to combat the effects of high dimensionality. This had the additional advantage of allowing me to visualize my entire data space in three dimensions. Second, I did not optimize the feature weights used in our nearest neighbor search because I was analyzing the system at the slide path scale. Third, I introduced inverse distance weighting using the nearest neighbor distance to weight more similar days more heavily in the calculation of individual slide path probabilities. This was critically important since it allowed me to calculate avalanche probabilities rather than a binary “avalanche” or “no avalanche” classification. Finally, I produced several visualization techniques to enable the use of multiple perspectives.

I used only three variables to decrease high dimensional problems, simplify visualization, and to create a new factor, such as wind loading. Using only three variables, as I did in this study, allowed me at least four times the amount of differentiation than if I had used five to twenty variables. An additional advantage of using only three variables was the relative ease in my ability to graph, display, and visualize the entire data space. This allowed me to have a mental picture of the data that helped understanding and decreased mental misrepresentations. Finally, the variables were combined to explore a factor. In this case the factor was wind loading as defined by new snow, wind speed, and wind direction. Wind loading is related to the amount and type of new snow, the wind speed, and the wind direction. By analyzing only these three variables I was able to explore the effect of wind loading without introducing error or biases from other variables. To analyze other variables, other factors may be conceived. For example, an avalanche forecaster would not be concerned with wet slides during a cold windy winter storm, and therefore would not be particularly concerned with three-day minimum temperatures. Likewise, in the spring, a warm windy event without precipitation will probably not lead to wind loading, but non-freezing temperatures over multiple days could lead to wet slides, and three-day minimum temperature information would be very important to a forecaster in this situation. If one representation of the data was used for both situations, it may not be as differentiating as the two separate representations.

I chose not to optimize weights for three reasons. First, it was computationally expensive. The measure for previous optimizing methods has been based on the presence or absence of an avalanche event on a given day. In this study, I was interested in individual avalanche paths, and any optimization would have to be carried out on each individual avalanche path. Combining this with a large database quickly makes these calculations difficult to manage. Second, most nearest neighbor models define variable weighting as the relative importance of that specific variable. When a weather variable is weighted heavily, near days have less variation around its target value. Instead of viewing weighting as the relative importance of a specific variable, I viewed feature weighting as a method to increase segregation in that variable's dimension. This is why I weighted wind direction twice as heavily as wind speed and new snow, since I was most concerned with understanding the effects that wind direction had on selective wind loading. Finally, I wanted to visualize the effect of weighting on nearest neighbor distances (another viewpoint). A display enabled the user to compare the relative (the component from each variable) and the total distances for all near days.

I used inverse distance weighting to count more similar days more heavily when calculating slide path probabilities for near days. As far as I can ascertain, this was the first use of inverse distance weighting using the nearest neighbor distance metric for avalanche applications. Using the nearest neighbor distance metric as a basis of weighting of nearest neighbors was first introduced by Dudani (1976). Other work followed with modifications to the weighting methods (Baily and Jain, 1978; MacLeod *et al.*, 1987). My method was to create a weighted mean of the near days using a nonlinear function similar to methods described by Zhang *et al*. (1997) and Stanfill and Waltz (1986).

Applying GKD to Avalanche Data

A number of researchers illustrate the strength of combining the concepts of GVis and KDD/GKD (i.e., MacEachren *et al*., 1999; Andrienko *et al*., 2001; Gahegan *et al.*, 2001; Wachowicz, 2001; Miller and Han, 2001). I applied these concepts to historical weather and avalanche data, which were well suited to be analyzed using the concepts of GVis and GKD. Slide paths have a geographic location along with geographic attributes (aspect, elevation, etc.) and can therefore be mapped, analyzed, and viewed with a GIS (Stoffel *et al*. 1998). The k-nearest-neighbors technique has already been used as a searching technique to find similar historical days (Buser 1983; 1989) and was the data-mining algorithm for my KDD/GKD approach.

Recently, there have been other studies aimed towards developing ways to aid data visualization and hypothesis generation. Cornice, a model developed by Purves *et al*. (2002), facilitates both of these goals, while SNOWBASE (Hägeli and Atkins, 2002) focuses on visualization and data storage.

In my approach, the nearest neighbor concept was used as a search method instead of a classifier. Avalanche probabilities for a given set of input variables were calculated for each slide path based on the set of the most similar historical days found by a nearest neighbor search. Both KDD and GVis consider multiple perspectives to be very important, so I viewed the data three different ways. First, a GIS representation of the slide paths was used to display individual slide path probabilities. This was the GVis perspective, and Evans (1997) found similar geographic representations of spatial data to be beneficial to the user. Second, the mean probabilities for aspect and elevation categories were used to relate those geographic attributes to the associated weather variables, which can be viewed using a rose diagram. Finally, a mean avalanche probability was calculated for all slide paths to get an overview of the set of weather variables.

Iteration is also a key concept of KDD and GVis (Andrienko *et al*., 2001; Ramakrishnan and Grama, 1999; MacEachren *et al*. 1999). The input values for a given set of weather variables for the nearest neighbor search were systematically varied to create a series of avalanche probability sets. Each variation was considered an iteration, and each iteration was viewed using one of the perspectives described above. More importantly, a feature of any perspective (individual slide path, aspect-elevation category, or mean probability) could have been analyzed throughout its series. If no relationship existed between the weather variables and the feature (i.e. an avalanche path), the avalanche probability would not drastically change with changes in the nearest neighbor search values. The response of a feature to changes in weather variables was a pattern or signature. Finally, by visualizing probability patterns of slide paths along with viewing different perspectives, I attempted to discover unknown patterns, thereby increasing avalanche knowledge. For example, I could discover if certain slide paths exhibit similar patterns.

GeoWAX

I developed GeoWAX (**Geo**graphic **W**eather and **A**valanche E**X**plorer) to implement the above ideas. GeoWAX is not a model in the sense that models often require some sort of optimal calibration to produce a result. Instead, GeoWAX is a tool for exploring historical data.

Since GeoWAX was developed for the exploration of data, all levels of interconnectivity of the data representations are retained and available to the forecaster. In other words, the user can utilize the lower level data from a higher-level representation. These representations include dynamic maps depicting avalanche events of a historical day, composite maps of near days displaying avalanche probabilities of individual slide paths, aspect-elevation rose diagrams displaying avalanche probabilities of aspect-elevation zones, graphs of nearest neighbor distances and weights, and representations of how avalanche probabilities change with changes in target day search criteria.

GeoWAX was used to analyze two case studies. The first example was published in Cold Regions Science and Technology (McCollister *et al.,* 2003) and explores relationships between geographic attributes of slide paths (aspect), weather variables, and the associated avalanche occurrences. The second example was presented at the International Symposium on Snow and Avalanches 2003 in Davos, Switzerland and explores relationships between weather attributes and avalanche attributes (avalanche type).

CASE STUDY 1: EXPLORING MULTI-SCALE SPATIAL PATTERNS IN HISTORICAL AVALANCHE DATA, JACKSON HOLE MOUNTAIN RESORT, WYOMING

Introduction

Avalanche forecasting utilizes inductive and deductive reasoning along with data and knowledge to reduce the uncertainty of the avalanche hazard for a given area (LaChapelle, 1980; McClung, 2002a; 2002b). Data used for avalanche forecasting can be categorized as meteorological, snow pack structure, or direct stability data (LaChapelle, 1980). These data are typically used in real time and are incorporated into the day's forecast. When these data are recorded and archived, they can be analyzed to gain intrinsic knowledge about the local area. The purpose of this paper is twofold. First, we present a technique for analyzing avalanche and weather data. Second, by implementing that technique using our program GeoWAX (**Geo**graphic **W**eather and **A**valanche E**X**plorer), we investigate relationships between new snowfall, wind speed, and wind direction on the avalanche activity at the Jackson Hole Mountain Resort at three scales: 1) the entire ski area (10^7 m^2), 2) groups of adjacent slide paths and groups based on aspect and elevation (10^3 to 10^4 m^2), and 3) individual avalanche paths (10^2 to 10^3 m^2).

A scientific understanding of avalanches, as well as knowledge of the local patterns of avalanche activity (gained through experience) is crucial for avalanche forecasters (McClung, 2002a). The former can be taught, but the latter is much more difficult to teach, communicate, or even define. For example, how new snowfall, wind speed, and wind direction conceptually lead to selective wind loading and the formation of slab avalanches is relatively easy to teach and understand. However, an understanding of which slide paths load under specific conditions of new snowfall, wind speed, and wind direction requires additional knowledge that may require decades of local individual observations and experience. Our method utilizes historical data to help aid in the visualization of the data, and to generate hypotheses regarding the role that different meteorological variables play in creating spatial patterns of avalanche activity at Jackson Hole Mountain Resort.

This study utilizes meteorological data for two reasons. First, they are directly related to historical avalanche data. Second, they are readily available and highly abundant. In addition, due primarily to the automation of data collection, the volume of these data is increasing exponentially as a function of time and amount of daily data being recorded. Each year more data are being recorded by increasing the number of different weather variables, adding new data collection site locations, and increasing the rate of taking measurements. These typical weather variables include, but are not limited to, precipitation (new snowfall, snow water equivalent, snow depth), wind (speed and direction, maximum gust), and temperature (maximum, minimum, mean).

A number of techniques have been and are being used to help forecast avalanches utilizing historical weather and avalanche data. These include discriminant analysis, cluster analysis, nearest neighbors, and binary decision trees. Obled and Good (1980) present an overview and comparison of the first three methods, Buser (1983, 1989) details the nearest neighbor method, and Davis *et al*. (1996) present an example of binary decision trees. Nearest neighbor and binary decision tree methods are now operationally used by a number of avalanche forecast operations. Recently, other tools have been developed to aid data visualization and hypothesis generation. Cornice, a model currently used by the Scottish Avalanche Warning Service (Purves *et al.,* 2002), facilitates both of these goals. SNOWBASE (Hägeli and Atkins, 2002), a program used by Canadian Mountain Holidays helicopter-skiing operation, focuses on visualization and data storage.

Our methods attempt to build on this past research in three ways. First, we incorporate the geographic component (i.e., the location, aspect, and elevation) of the slide paths. Second, we analyze the data at the individual slide path scale, which is of primary importance to ski patrollers and others doing avalanche hazard reduction work. Finally, instead of treating a day as either a day with avalanches or without, we create a probability of avalanching for each individual slide path, which can be geographically viewed using a GIS. Our primary goal is to create a tool to visualize, explore, and ask questions of weather and avalanche datasets, thereby allowing us to find spatial patterns and facilitate hypotheses generation.

Geographic Visualization and Geographic Knowledge Discovery are two emerging fields that share our primary goal of finding patterns and relationships in large spatial datasets. Both fields have several underlying concepts in common (MacEachren *et al*. 1999). First, both involve the *interaction* of computers and humans and see this interaction as a *process*, attempting to capitalize on the strengths of both (Miller and Han, 2001; MacEachren *et al*. 1999; Andrienko and Andrienko, 1999; Ramakrishnan and

Grama, 1999; Fayyad *et al*., 1996; Hibbard and Santek, 1989). Second, *iteration* allows visualization of patterns with different attributes, at different times, or at different scales that may illuminate trends that would not be obvious in a static view (Andrienko *et al*., 2001; Ramakrishnan and Grama, 1999; MacEachren *et al*. 1999). Iteration is also familiar to avalanche forecasters, who typically use iteration while forecasting to reduce uncertainty and improve forecast accuracy (LaChapelle, 1980). Third, high interactivity between the user and computer allows the user to pose *"what if"* questions for hypothesis generation (Gahegan *et al*., 2001; MacEachren *et al*. 1999). Finally, *multiple perspectives* allow the user to view the data at different scales, measures, or even different concepts (Andrienko *et al*., 2001; MacEachren *et al*. 1999). Purves *et al.* (2002) emphasize the importance of multiple perspectives for avalanche forecasting tools.

The rest of this paper will outline the study area for our project, the methods we used to develop GeoWAX, and provide an example of how we used GeoWAX to investigate of the role of new snow, wind speed and wind direction on the spatial patterns of avalanching at Jackson Hole Mountain Resort at a variety of scales.

Study Site

This study uses historical data recorded by the Jackson Hole Mountain Resort, which is located on Rendezvous Mountain in the southern end of the Teton Range in northwestern Wyoming, USA (Figure 3). The base elevation of the mountain is 1923 m, rising to a summit elevation of 3185 m. The Jackson Hole Mountain Resort is situated at 43º 36' north latitude and is roughly 1000 km from the nearest moisture source (Pacific

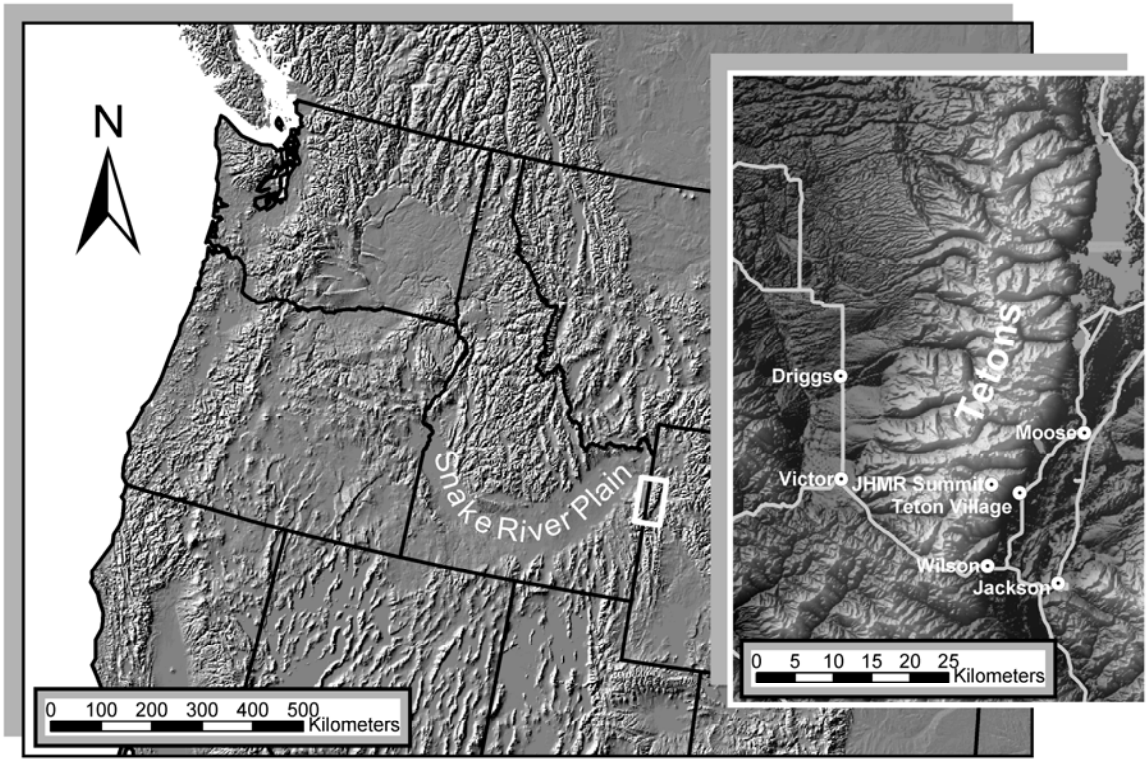


Figure 3: Jackson Hole Mountain Resort at Teton Village, Wyoming, USA.

Ocean), giving the area an intermountain climate (Mock and Birkeland, 2000). In the winter, precipitation is mainly in the form of snow. Mid latitude cyclones from the Pacific are intensified by orographic uplift as they encounter the western side of the Teton Range, especially when they travel along the relatively low and flat Snake River Plain to the west. The yearly mean snowfall for the study plot at the top of the resort is 12.8 m of snow containing 1.5 m of snow water equivalence, while the base receives 2.6 m and 0.5 m, respectively (Kozak, 2002). The predominant wind direction for most storms affecting Jackson Hole Mountain Resort is west-southwest (Birkeland *et al*. 2001). The town of Jackson, WY, has kept climatic records since 1948. The yearly mean high and low temperatures are 12.1 and –5.1 degrees C, with the coldest month being January (-2.8 and –15.0 C) and the warmest month being July (27.2 and 4.4 degrees C).

Methods

Applying Geographic Knowledge Discovery to Historical Avalanche and Weather Data

We applied the concepts of Geographic Visualization and Geographic Knowledge Discovery to historical weather and avalanche data. Slide paths have a geographic location along with geographic attributes, such as aspect and elevation, and can therefore be mapped, analyzed, and viewed with a GIS (Stoffel *et al*. 1998). The k-nearest-neighbors technique is our data-mining algorithm, and is used to generate avalanche probabilities (Appendix A). We base these avalanche probabilities, which are related to a given set of input variables for each slide path, on a set of the most similar historical days found by the nearest neighbor search. Multiple perspectives of the data included a GIS representation of the slide paths to display individual slide path probabilities for each slide path (Figure 4), a rose diagram to relate mean probabilities for aspect and elevation categories to the search variables, and graphical displays of the total nearest neighbor distance, the inverse distance weighting, and the partial distances for the nearest days.

The meteorological variables used for this study include new snowfall, wind speed, and wind direction. We are interested in these variables because the daily disruption of the snow pack due to skier traffic and avalanche hazard reduction activities minimizes the importance of older layers in the snow pack. We also only used three search variables to

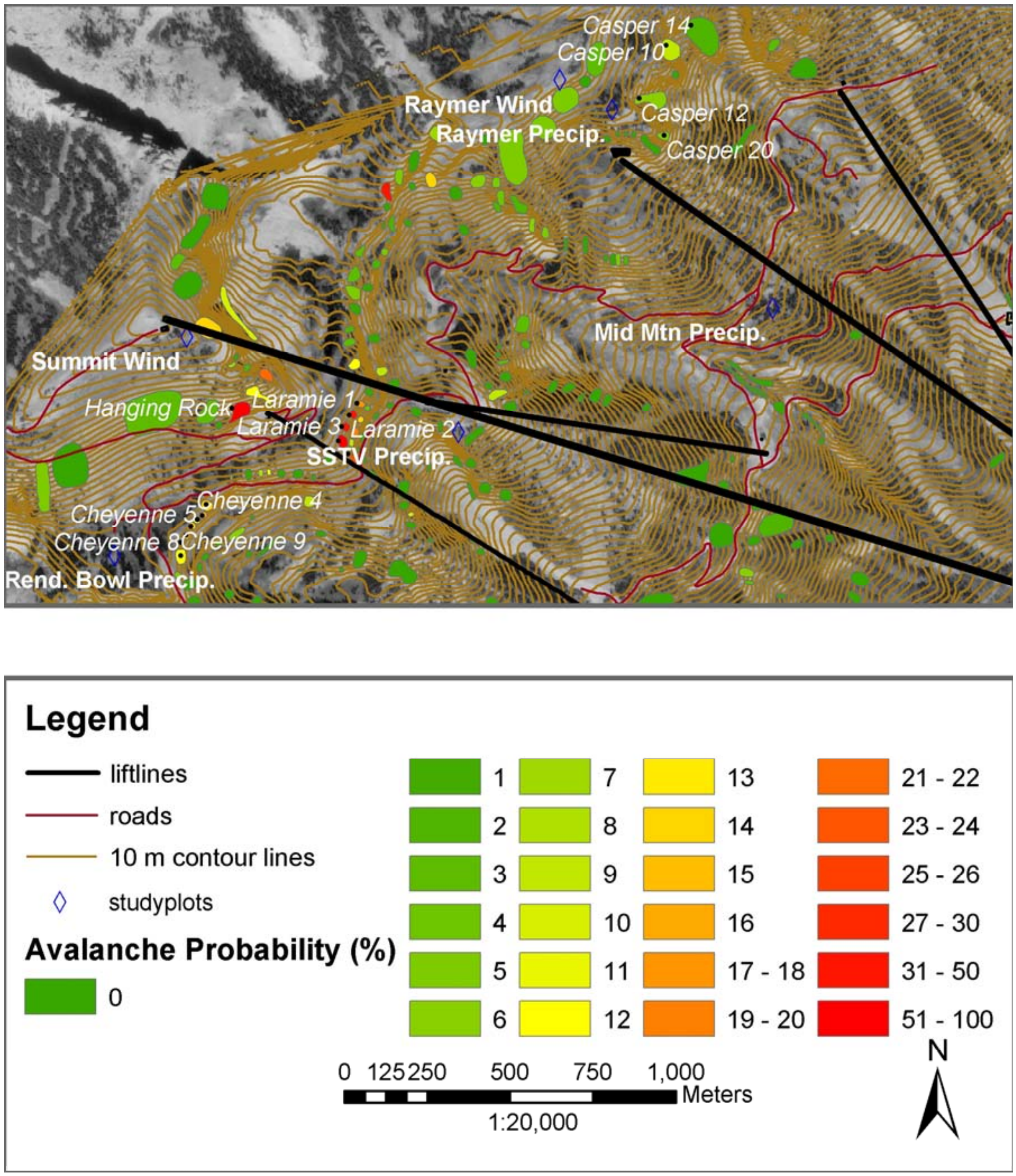


Figure 4: GIS Representation of the Jackson Hole Mountain Resort, Wyoming, USA.

minimize the potential problematic effects of high dimensionality, which can occur in nearest neighbor techniques using as few as 10 to 15 weather variables (Aggarwal *et al*., 2001; Hand *et al*., 2001; Hinneburg *et al*., 2000; Beyer *et al*., 1999).

For each of the three weather variables, we analyzed how they affected the pattern of avalanche activity for individual slide paths, for aspect-elevation categories, and for the mean avalanche probability. After investigating the effects of the individual weather variables, we analyzed them together to find wind loading patterns associated with new snowfall. When analyzed together, a specific pattern is created for each slide path, each aspect-elevation category, and for the mean avalanche probability. This pattern is a specific signature for each feature and can be used to identify similarities and differences between similar feature types such as two slide paths. Next, we analyzed these signatures to identify the scales on which these three weather variables operate. Hägeli and McClung (2000) concluded that the scales of weather variables used in avalanche models might not be representative of the scales of natural processes in this complex earth system.

Data

The data for this project include both historical and geographic data. The historical data are composed of daily weather measurements and the associated avalanche activity from the Jackson Hole Mountain Resort, Wyoming. The historical weather and avalanche data span 23 winter seasons, from 1978-79 to 2001-02, which include 3,304 days and over 10,000 individual avalanche events.

By the 1978-79 season, 34 variables consisting of 204 weather measurements were recorded daily along with the associated avalanche activity. These weather data included measurements from four precipitation sites (Rendezvous Bowl, SSTV, Mid-mountain, and Base, see figure 4) that recorded new snowfall, snow water equivalent (SWE) and total snow depth; three temperature sites recording 6:00 AM, 24 hour minimum, and maximum temperatures; one summit wind site (4 x 6-hour-mean speed and direction); and numerous subjective variables such as snow available for wind transport and daily warming. Throughout subsequent seasons those original weather variables have been recorded along with new additional weather variables. Today, over 50 variables, consisting of hundreds of individual weather measurements are recorded daily, which include data from five precipitation sites, four temperature sites, and three wind sites, most of which are remote and automatically recorded up to four times per hour. Precipitation measurements are manually verified at each site daily. The historical avalanche data consist of 10,232 avalanche events within the ski area. Avalanche events are recorded using standard U.S. methods (Perla and Martinelli, 1978), which include the date, slide path name, time, type, trigger, depth, U.S. size, and sliding surface as attributes. These data reside at the Jackson Hole Mountain Resort.

The geographic data sources include a one-meter resolution USGS Ortho Quad, a digital elevation model (DEM), and a polygonal representation of the starting zones of 220 in-bounds slide paths. The elevation data for the Jackson Hole Mountain Resort, in an Auto-CAD format (Schriber, 1998), were imported into a GIS (ArcInfo 7.0, ESRI) and oriented using common features in the Ortho Quad. Three-D Analyst, an extension of

Arc-View 3.2 (ESRI), was used to create a 5-meter DEM from the original 10-foot contour data. An aspect grid was created from the DEM using Spatial Analyst 2.0 (ESRI). Using the GIS, the lead avalanche forecaster for the Jackson Hole Mountain Resort digitized the slide path starting zones on-screen using the Ortho Quad and contour data for reference. The slide path starting zones were in a polygonal (vector) format where each starting zone was represented by an enclosed polygon with the attributes of name, mean elevation, and mean aspect of each starting zone. The mean elevation and aspect for each starting zone were calculated by averaging all respective grid cells contained in that starting zone's polygon.

Creating Slide Path Avalanche Probabilities

Creating individual avalanche probabilities for each slide path is a seven-step process (Appendix B). First, a set of weather variables along with a set of values is chosen as a basis for searching the historical database. These criteria constitute a *target day*. An example of a target day might be the following: new snowfall = 25 cm, mean wind direction = 270°, mean wind speed = 5 m/s. Second, an optional filter is applied to limit the historical days used. For example, we might only consider days with new snowfall greater than 15 cm but less than 35 cm. Third, as in other nearest neighbor approaches (i.e. Buser, 1983; 1989), all variables and target day values are normalized by their standard deviation. In step four, optional variable weights can be chosen to increase differentiation of a specific weather variable.

Some nearest neighbor models optimize the variable weights. Gassner *et al.*, (2000) did this by using a local expert to set the weights and then created a measure of correctness to compare different weighting schemes. Purves *et al*. (2002) used a genetic algorithm to determine optimal weights. Though this can be useful, we chose not to optimize weights for two reasons. First, it was computationally expensive for our analyses since we were analyzing individual avalanche paths. Second, most nearest neighbor models define weight as relative importance, and therefore when a weather variable is weighted heavily, the nearest days have less variation around the heavily weighted target value.

Step five involves the calculation of the nearest neighbor distance for all days in the filtered, standardized database. This technique creates a distance measurement for each day in the historical database based on its similarity to the target day. The more similar a historical day is to the target day, the shorter the distance measurement. In step six similar days are found in the historical database by ordering the historical days by their nearest neighbor distances.

In the final step, slide path probabilities are calculated based on the actual avalanche activity of the most similar days as defined by their nearest neighbor distance. First, the user chooses the maximum number of days to used. For example, if we consider the 100 nearest days, the number of avalanches is summed and the mean is calculated for each slide path over those 100 nearest days. If one slide path had ten avalanches during those 100 nearest days, its avalanche probability is 10%. Likewise, a slide path with 50 avalanche events out of 100 nearest days has an avalanche probability of 50%.

Additionally, the nearest days can be optionally weighted by an inverse function of the nearest neighbor distance to count more similar days more heavily. Our method creates a weighted mean of the nearest days using a nonlinear function similar to methods described by Zhang *et al*. (1997) and Stanfill and Waltz (1986) by weighting a day with the inverse of the nearest neighbor distance (NNDist) plus a zero distance value (ZDV) to avoid dividing by zero, all raised to the inverse distance exponent (IDE). An IDE value of zero would count each nearest day equally, while a IDE value of 1 would be traditional inverse distance weighting. The numerator of equation 1 is the summation of weighted avalanche events where days with no avalanches receive a 0, and days with an avalanche receive a 1. The denominator of equation 1 is the 100% maximum probability of avalanching equaling the weighted summation of an avalanche event on each of the nearest days.

$$Slidepath\ Avalanche\ P = \frac{\displaystyle\sum_{MND}^{Day_i=1} \frac{Avalanche\ Event_{Day_i} \ni : \{0,1\}}{(NNDist_{Day_i} + ZDV)^{IDE}}}{\displaystyle\sum_{MND}^{Day_i=1} \frac{1}{(NNDist_{Day_i} + ZDV)^{IDE}}}$$

The resulting set of slide path avalanche probabilities allowed the creation of the GIS representation (Figure 4).

Creating Avalanche Probabilities for Aspect-Elevation Categories

Creating avalanche probabilities for the aspect-elevation categories is a two-step process (Appendix B). In the first step, the combined geographic attributes of the aspect and elevation of slide paths are related to weather variables for the entire ski area rather than for individual avalanche paths, with each slide path being categorized based on its mean elevation and aspect. Low (1829-2286 m), middle (2286-2743 m), and high (2743+ m) are used as three elevation categories along with eight aspect categories (N, NE, E, SE, S, SW, W, NW) for a total of 24 possible categories. Next, the mean slide path probabilities are calculated for all slide paths based on their aspect-elevation category, and are viewed using a rose diagram.

Creating Series Signatures

The combination of the target day and the set of resulting output (slide path avalanche probabilities, aspect-elevation probabilities, and the mean slide path probability) constitute what we define as a Nearest Neighbor Avalanche Probability Profile (NNAPP). A NNAPP encapsulates the total response of the system for a set of search variables (Appendix B).

The effects of weather variables on avalanche activity can be visualized as a multi-dimensional space where each weather variable is represented by a different dimension. New snowfall, wind direction, and wind speed define a three-dimensional space. To explore the response to changes of new snowfall, wind direction, and wind speed, a NNAPP is created for each set of search variables by systematically varying one weather variable at a time, eventually creating a NNAPP to populate each location (variation of variables) in the three-dimensional series space. We call this a series signature. The NNAPP attribute avalanche probability now constitutes a fourth dimension. Two of the three weather variables and an avalanche probability are graphed, visualized, and analyzed. Examples of series signatures are shown below.

<u>GeoWAX</u>

We developed GeoWAX to implement the previous methods using Microsoft's Visual Basic 6.0 along with ESRI's Map Control in ArcView 8.1 to implement the embedded map. GeoWAX is an interactive program to enable avalanche forecasters to explore their historical data and aid in visualization of data and hypotheses generation. The forecaster can vary the search variables used in the nearest neighbors search and the variable weights, and can filter the weather data based on a range of each search variable or any set of fixed values. When creating the slide path probabilities, the forecaster can also vary the number of nearest days to be used along with the nearest neighbor distance weighting function. Since GeoWAX was developed for the exploration of data, all levels of interconnectivity of the data representations are retained and available to the forecaster. For example, when viewing a series signature, all of the NNAPPs are retained and can be viewed (GIS representation of slide paths, aspect-elevation rose-diagram, and mean avalanche probabilities). Likewise, the actual weather and avalanche events for all nearest days can be viewed along with a GIS representation of a day's avalanche events.

<u>Wind Loading of New Snowfall</u>

We chose new snowfall, wind speed, and wind direction to explore their effect on avalanche activity for the Jackson Hole Mountain Resort. New snowfall (Rendezvous Bowl precipitation) values ranged from 0 to 35 cm in 5 cm increments for a total of 8 steps in the new snowfall dimension. Wind direction (summit wind) was varied from 0° to 360° in 20° increments for a total of 19 steps and was weighted twice as heavily as new snowfall and wind speed to help differentiate the different wind direction categories. The wind speed dimension had three categories: 5 m/s (low), 10 m/s (moderate), and 15 m/s (high). All variables were normalized with their standard deviation to normalize distance measurements. Days were filtered with ranges based on the target values. New snowfall ranged ± 15cm, wind speed ± 4 m/s, wind direction ± 30° around their respective target values and the inverse of the square root of the nearest neighbor distance was used to weight more similar days. A minimum of 10 days and a maximum of 100 days were used to create the 456 NNAPPs. Every slide path, aspect/elevation category, and the mean probability were available for analysis, producing individual, unique series signatures.

Statistical Analyses

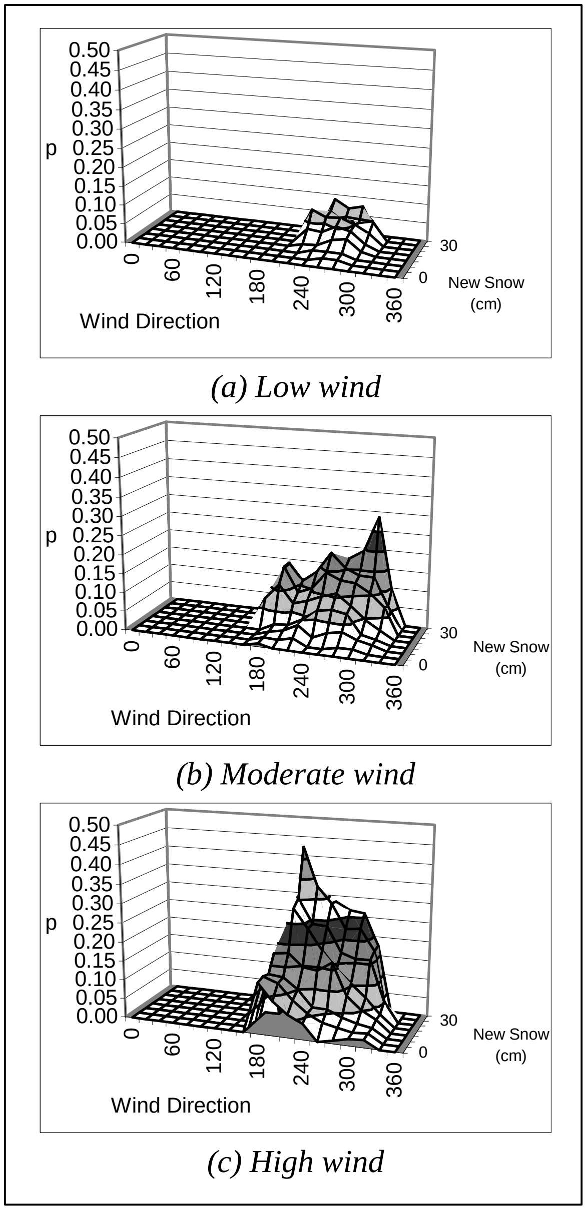
The goal of our statistical analyses is to compare the pattern observed for one series signature (for an avalanche path or groups of paths) to another series signature. We use two types of non-parametric statistics to analyze our data. First, we use a Mann-Whitney U test to compare the means of two series signatures. This test is applicable when we are interested in the effect of only one variable, such as wind speed. When our primary interest is in the pattern observed, we use Spearman's rho, a non-parametric correlation analysis similar to Pearson's r, to compare the avalanche probabilities in one series signature to the other series signature.

Results and Discussion

Individual Weather Variables

Our investigation focuses on how snowfall, wind speed and wind direction affect the spatial patterns of avalanche activity at Jackson Hole Mountain Resort. An increase in new snowfall leads to an increase in the avalanche probability at all scales, from individual paths to the entire ski area. More new snowfall results in more stress added to buried weak layers or interfaces, thereby increasing the probability of avalanche activity (McClung and Schaerer, 1993), and the effect of this can be seen at all scales, ranging from individual paths to the entire resort.

In contrast, the effect of wind speed differs depending on the scale of observation. At the scale of individual avalanche paths, considerable variability exists. Most slide paths exhibit an increase in avalanche probability with an increase in wind, with a few paths displaying a large increase, such as Buffalo Bowl, a middle elevation (2404 m) slide path. The series signatures for low, moderate, and high wind situations for Buffalo Bowl show this large increase in avalanche probability for increasing wind speed (Figure 5), and all were significantly different (P values < 0.001). In contrast, some slide paths, such as Broadway, decrease in avalanche probability with an increase in wind; perhaps the higher wind speeds scours those paths. Others, such as Cajun Couloir, increase in avalanche probability under certain wind directions, and decrease at other directions (Figure 6). At the scale of the entire ski area there is a general increase in avalanche probability between low and moderate wind, but not between moderate and high wind,



(a) Low wind

(b) Moderate wind

(c) High wind

Figure 5: Series Signatures for Buffalo Bowl.

for both the overall mean and the aspect-elevation categories. These results demonstrate how much variability exists at the scale of single paths within the overall mean for the ski area.

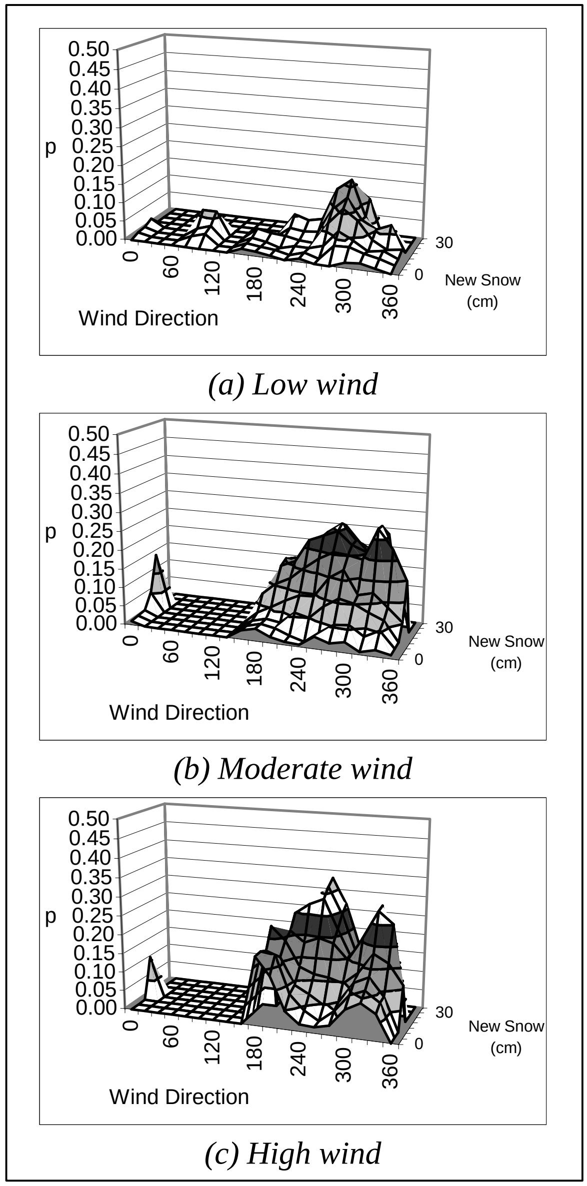


Figure 6: Series Signatures for Cajun Couloir.

The effect of wind direction also differs for different scales. At the scale of individual slide paths, changes in wind direction change the probability of avalanche activity. Although changes in wind direction also lead to changes in avalanche activity for slide paths grouped by aspect and elevation, these changes are similar to each other and the overall mean computed for the entire ski area. The responses of the individual avalanche paths may cancel each other out and "smooth" the data.

Series Signature Patterns

Our data exploration with GeoWAX shows that many slide paths exhibit similar series signatures. Additionally, slide paths with similar signatures are often in the same geographic area. Examples of grouped spatial similarity include the Cheyenne group, the Laramie group, and the Casper group, which gives us insight into possible scales of avalanche processes that may exist at the Jackson Hole Mountain Resort.

Slide paths in the Cheyenne group include Cheyenne 3-9, The Snag, and Roadcut, and all exhibit similar series signatures, with high avalanche activity associated with winds out of 240-260º. When the slide paths within the Cheyenne group are compared to each other the Spearman's correlations (rho) range from 0.746-0.983 (p values < 0.001), showing strong inter-group similarities. The series signature for Cheyenne 3 is a typical series signature for this group (Figure 7).

A similar situation exists for the slide paths in the Laramie group, which include Laramies 1-5. These are some of the most active slide paths on the mountain, with series signatures displaying high avalanche probabilities with winds from 180 to 360 degrees (Figure 8). Their correlation values range from 0.801-0.962, with p values < 0.001 (Figure 9), again showing strong similarities within the group. There are also some smaller scale (individual slide path) trends. Laramie 5, an east northeast-facing starting

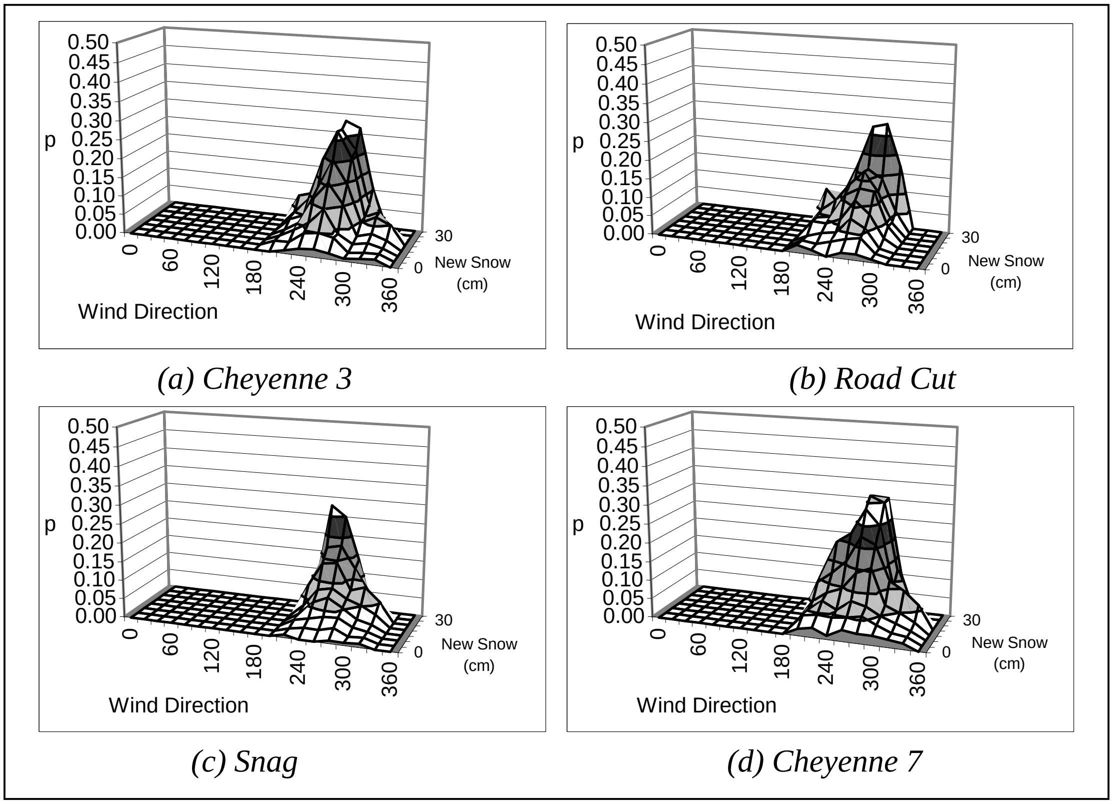
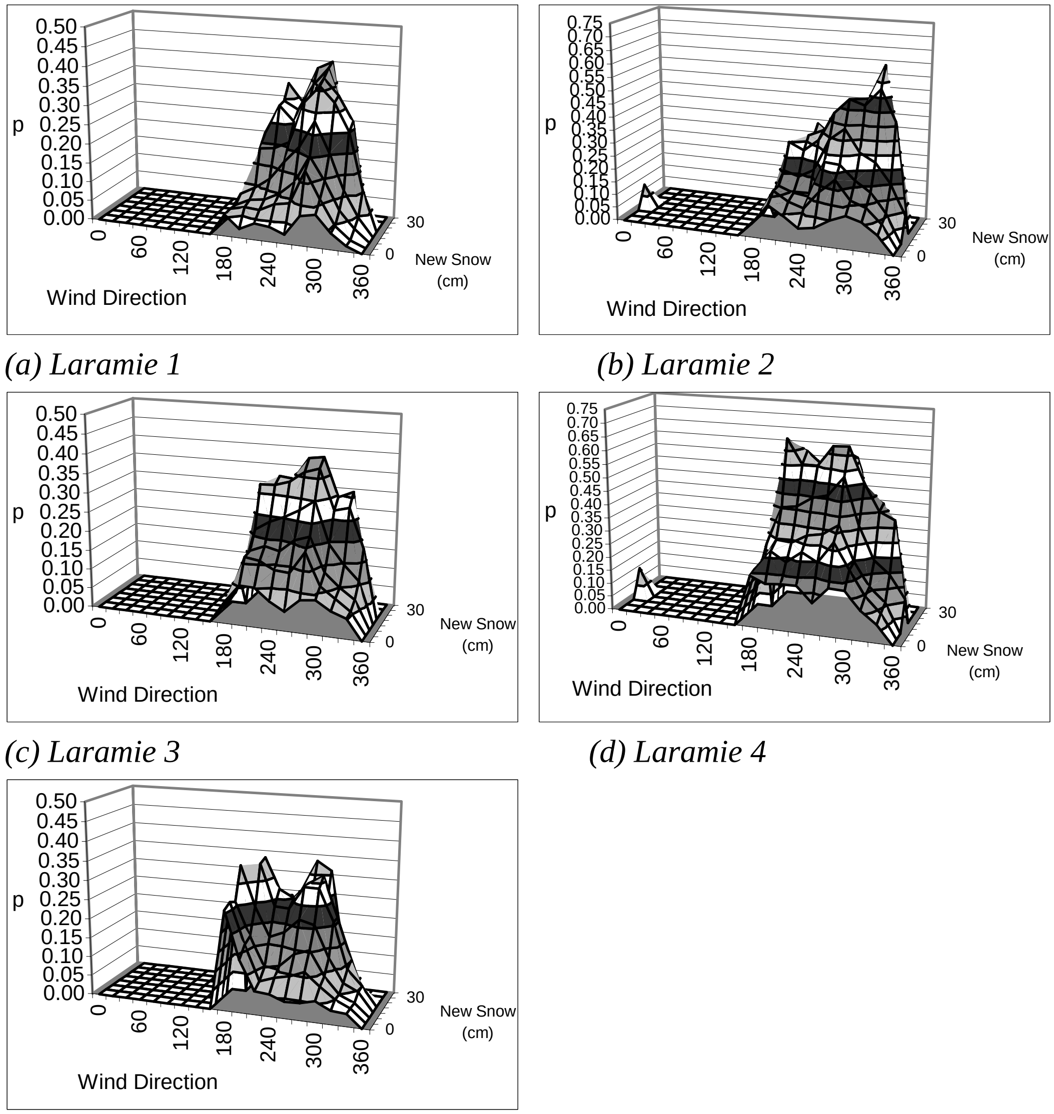


Figure 7: Series Signature for the Cheyenne Group in High Wind.

zone just on the lee side of a small ridge, had its highest avalanche probabilities with southerly winds. In contrast, Laramies 1-4 had their highest avalanche probabilities with more westerly winds.

In contrast to both the Cheyenne and Laramie groups, the slide paths in the Casper group (Caspers 10, 12, 14, 20) all experience their highest avalanche activity with winds either more southerly or northerly than the predominant west southwest winds (Figure 10). Although the correlation values are all significantly correlated with p values < 0.001, the amount of similarity is much less, ranging from 0.478 to 0.863 (Figure 11). The two slide paths that are most dissimilar to each other are also the farthest apart in



(a) Laramie 1 *(b) Laramie 2*

(c) Laramie 3 *(d) Laramie 4*

(e) Laramie 5

Figure 8: Series Signatures for the Laramie group in High Winds.

distance (Casper 20 and Casper 14). Here a sizable difference between slide paths in the same group exists. Both Casper 12 and Casper 20 experience high avalanche activity with southerly winds while Casper 10 and Casper 14 become more active with northerly

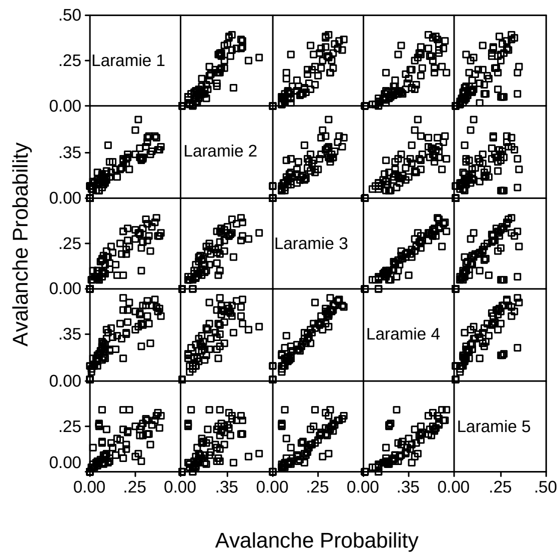


Figure 9: Scatter Plot for Corresponding Series Signatures for the Laramie Group.

winds. We suspect this is due to the geographic location of the slide paths. Caspers 12 and 20 are both situated at the southern end of Casper bowl, and are leeward of a ridge with southerly winds. In contrast, Caspers 10 and 14 are in the center of Casper bowl and may be more sheltered from southerly winds. Similar to the Laramie group, the Casper group also shows some differences at the slide path scale, yet still had similarities at the group scale.

After finding similar series signatures for different groups, we created a mean series signature for each group and compared these group means with each other using their series signatures (Figure 12 and 13). The *Cheyenne group* and the *Laramie group* are quite similar with a correlation of rho = 0.923. In contrast, the *Cheyenne group* and the *Laramie group* are more poorly correlated to the *Casper group* with rho values of 0.496 and 0.591, respectively (P values < 0.001).

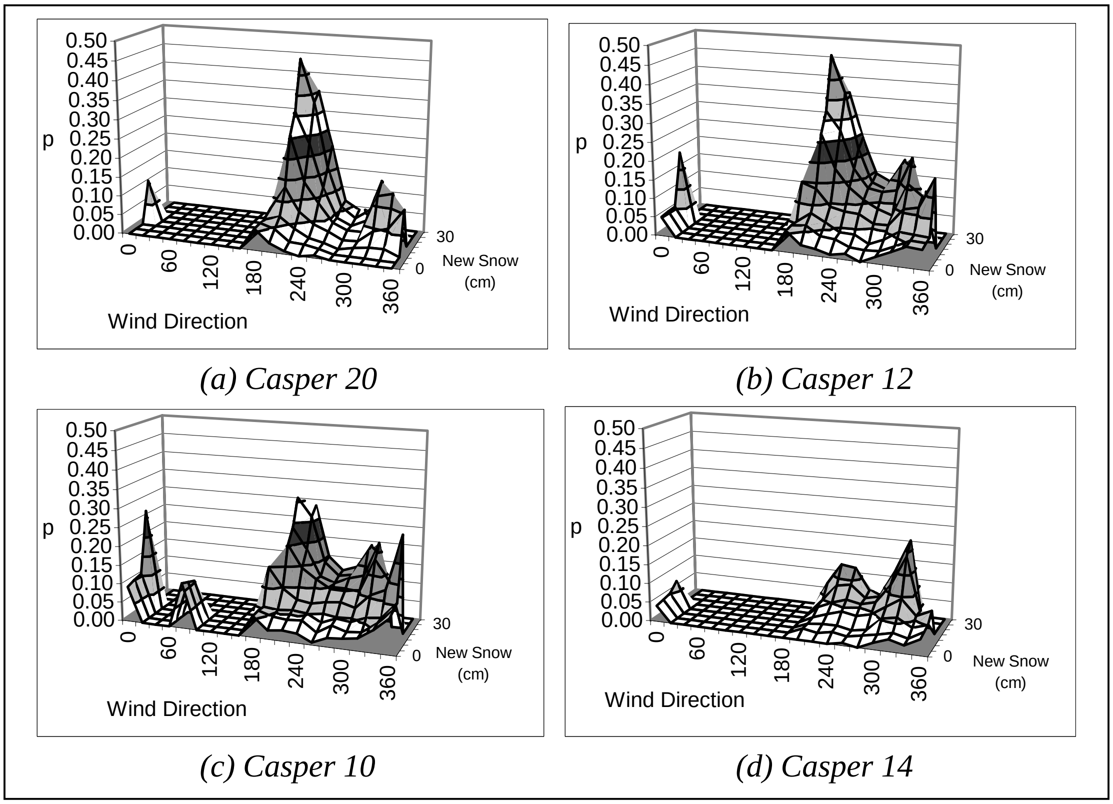


Figure 10: Series Signatures for the Casper Group in High Wind.

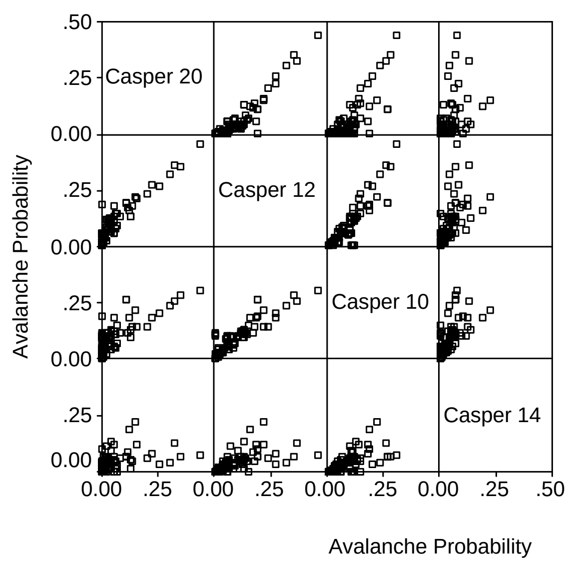
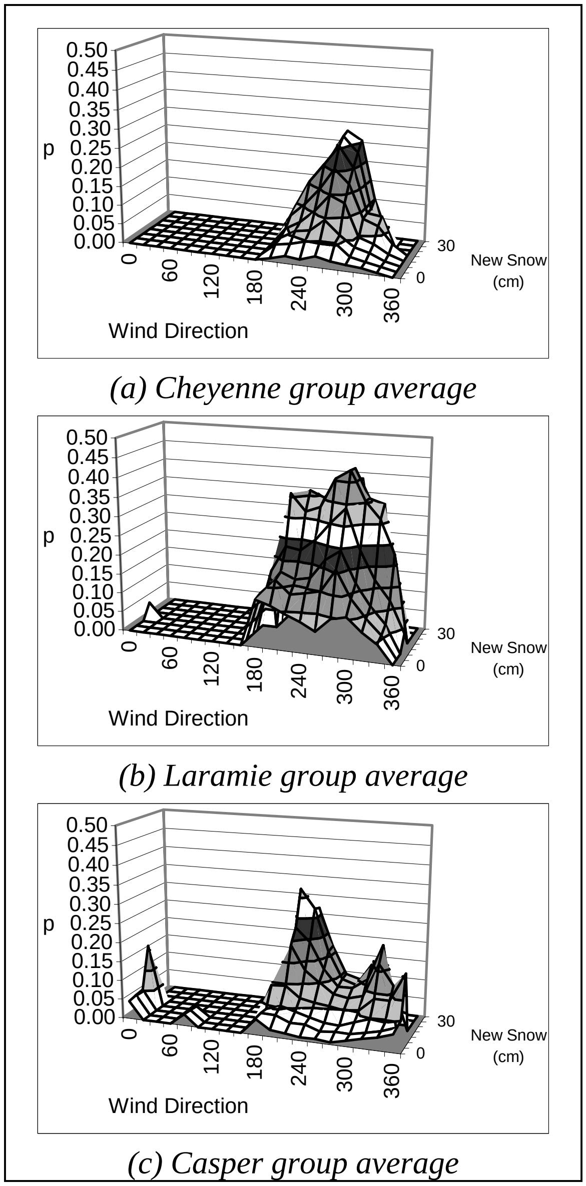


Figure 11: Scatter Plot for Corresponding Series Signatures for the Casper Group.



(a) Cheyenne group average

(b) Laramie group average

(c) Casper group average

Figure 12: Series Signatures for the Cheyenne, Laramie, and Casper Groups in High Winds.

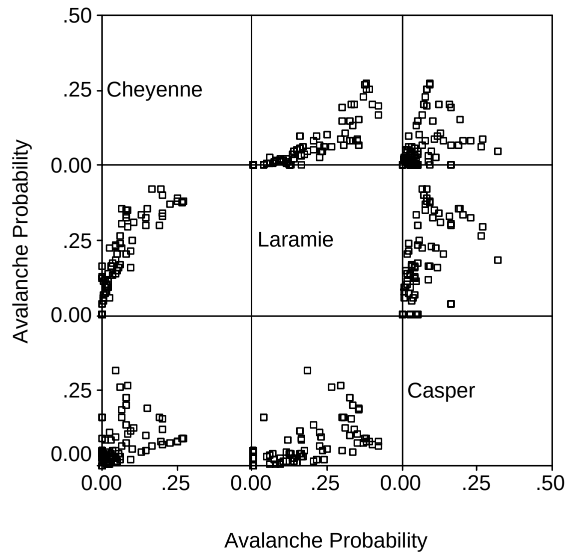
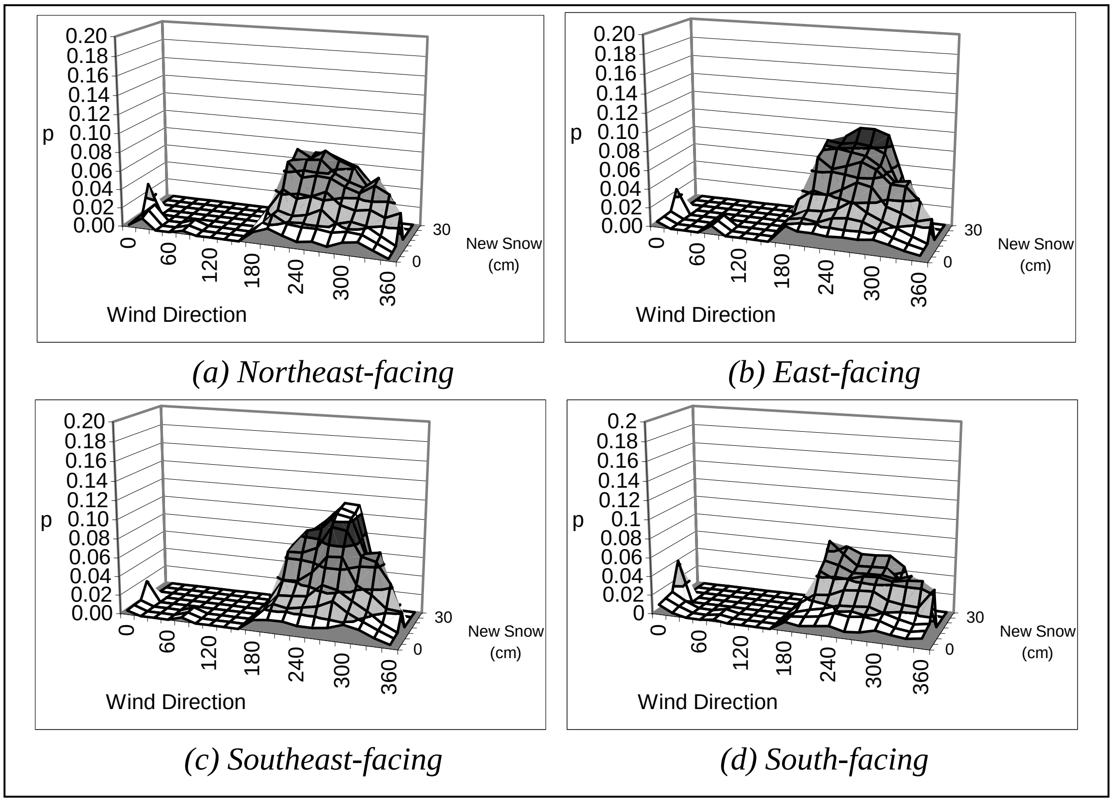


Figure 13: Scatter Plot for Corresponding Series Signatures for the Cheyenne, Laramie, and Casper Groups in High Winds.

<u>Compare Aspect-Elevation Series Signatures</u>

Enlarging the scale of our analysis to sets of avalanche paths grouped by aspect, rather than by geographic location, gives us different results. A look at four high-elevation aspect categories (northeast-facing, east-facing, southeast-facing, and south-facing) shows that their series signatures appear similar with no obvious relationship to wind direction (Figure 14). Further, a correlation and scatter plot analysis shows that they all correlate well with each other, with rho = 0.89-0.97 (Table 1; Figure 15). Thus, while sizable differences exist between some groups of slide paths based on their location within the ski area (i.e., comparing the Cheyenne and Casper groups), those sizable differences do not exist between sets of avalanche paths grouped by aspect and elevation. The differences at the scale of individual slide paths and groups of slide paths shown in



(a) *Northeast-facing* (b) *East-facing*

(c) *Southeast-facing* (d) *South-facing*

Figure 14: Series signatures for Four High Elevation Aspect Zones in High Winds.

the previous section must cancel each other out when based on the mean aspect-elevation categories. These results do not support a direct relationship between wind direction collected at a central location and slide path aspect for numerous slide paths in complex terrain.

Table 1: Similarity Among Grouped Slide Paths and Between Grouped Slide Paths.

	Spearman's rho	p (2-tailed)
Cheyenne group	0.746-0.983	**0.000**
Laramie group	0.801-0.962	**0.000**
Casper group	0.478-0.863	**0.000**
Cheyenne vs. Laramie	0.923	**0.000**
Cheyenne vs. Casper	0.496	**0.000**
Laramie vs. Casper	0.591	**0.000**
NE facing vs. E facing	0.952	**0.000**
NE facing vs. SE facing	0.927	**0.000**
NE facing vs. S facing	0.899	**0.000**
E facing vs. SE facing	0.966	**0.000**
E facing vs. S facing	0.909	**0.000**
SE facing vs. S facing	0.888	**0.000**

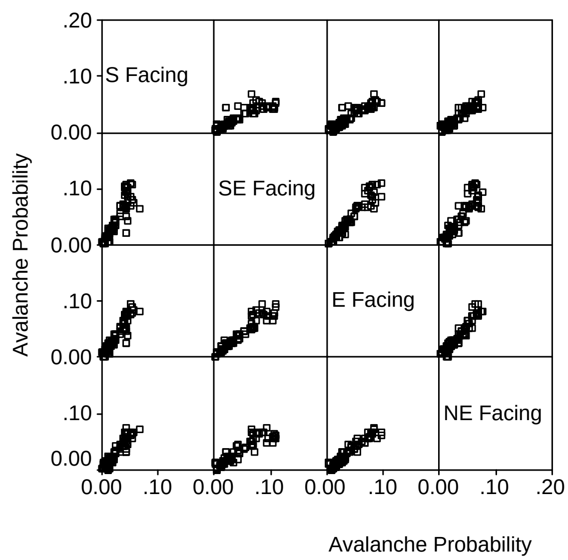


Figure 15: Scatter Plot for Corresponding Series Signatures for Four High Elevation Aspect Zones in High Winds.

Wind Correlation Between Summit and Raymer Wind Sites

Our results indicate that patterns emerge due to changes in wind direction. We suspect specific wind flow patterns cause the observed slide path differences. To explore the possibility that specific wind flow patterns exist, we plotted the hourly mean wind direction for our two wind sites over two seasons (Figure 16). When the two seasons were plotted separately, the same distinct pattern was observed. These distinct wind patterns suggest specific wind flow patterns develop around the mountain according to specific upper air wind directions. Font *et al*. (2001) found similar results when they created aeolian susceptibility maps that categorized small-scale wind patterns by the aeolian features created by different local wind directions and then related this to a centralized wind station. In their work specific centralized wind directions led to consistent patterns of wind erosion and deposition as determined by the maps, which is consistent with our findings.

Conclusion

Each of the three weather variables we investigated affected the avalanche probabilities differently. New snowfall increases avalanche activity at all scales. However, it does not play a significant role in differentiating avalanche activity between individual slide paths. In contrast, wind speed does have a differentiating effect, depending on the avalanche path location. For example, high wind is important in the creation of avalanches at lower elevations, which may be due to a wind threshold needed for slab development that only occurs at lower elevations with high summit winds. Of

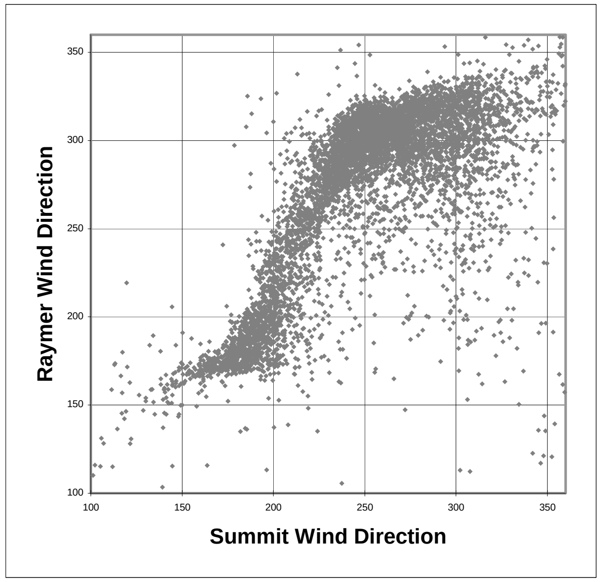


Figure 16: Direction-Direction Scatter plot for the Summit and Raymer Wind Sites.

the three weather variables, wind direction is the most important for differentiating individual slide path avalanche probabilities, probably because winds are being redirected by topography, and are selectively loading specific avalanche paths.

The combination of our three weather variables, along with their series signature representations, provides new knowledge about selective wind loading at a variety of scales, from individual avalanche paths to groups of paths. Analyzing series signatures was critical for our analyses, which resulted in a high correlation between adjacent slide paths and relatively low correlations between different groups of slide paths. In addition to this interpolated knowledge, we can use the series signatures for a given path, or groups of paths, to extrapolate the wind loading effect for highly unusual situations. For example, we would be much more concerned with avalanche paths in the Casper group than the Cheyenne group if we had high winds out of 140-160 degrees (SSE) associated with a large storm.

All of our high-elevation aspect categories exhibit similar series signatures. At the scale of the entire ski resort (10^7 m^2) there is no obvious relationship between avalanche activity based on aspect and wind direction. This result is important to demonstrate that wind direction measured at a central, high elevation location does not necessarily directly relate to the specific aspects being wind loaded. We are not implying that aspect with respect to wind direction does not play a role in avalanche development; clearly, at the scale of individual paths, wind direction is critically important. However, since wind instrumentation is typically located to measure an approximation of the free air winds, specific topography around a given path, and not simply aspect, is more important when relating wind direction to avalanche activity.

CASE STUDY 2: EXPLORING THE SPATIAL VARIABILITY OF HARD SLAB AND DRY LOOSE AVALANCHES, JACKSON HOLE, WYOMING, U.S.A.

Introduction

The avalanche type is one of the data variables recorded for avalanche events using standard U.S. methods (Perla and Martinelli, 1978). These avalanche types include hard slabs, soft slabs, wet slabs, dry loose, and wet loose avalanches. The objectives of this study are twofold. First, we explored the spatial distribution of the two types of dry avalanches - hard slab and loose avalanches - irrespective of weather. Next, we incorporated three weather variables (wind speed, 24-hour maximum temperature, and new snow density) to explore the relationship between those three weather variables and the types of avalanches observed. Understanding where different types of avalanche occur is important since slab avalanches are more dangerous than loose snow avalanches, and hard slab avalanches can be particularly difficult to predict and mitigate (Richmond, 1994).

Avalanches can be broken down into two main types: slab avalanches or loose snow avalanches (McClung and Schaerer, 1993). Slab avalanches result from the shear fracture of a weak layer underlying a relatively more cohesive slab and are particularly dangerous because people can trigger the avalanche well below the fracture line (Schweizer, 1999). In the U.S. classification, slab avalanches are broken down into dry and wet avalanches, and dry slab avalanches are further subdivided into either hard slabs or soft slabs (Perla and Martinelli, 1978). This classification is somewhat subjective, but hard slabs generally consist of harder and denser snow (greater than 300 kgm^{-3}), and angular blocks of the slab tend to be preserved over long distances in the avalanche, depending on the ruggedness of the path. In contrast, soft slabs are less cohesive and disintegrate into loose material shortly after the avalanche starts. Perla and Martinelli (1978) state that if it is not clear whether an avalanche is a hard or soft slab that it should be classified as a soft slab.

Loose avalanches do not release as a cohesive unit like slab avalanches. Instead, they start when a small amount of snow slips out of place and moves down slope, encountering and entraining other cohesionless snow. Loose snow avalanches typically consist of less snow than slab avalanches, they are not as large or destructive, and they only rarely catch people because they usually release below the trigger.

We use 23 seasons (1978-79 to 2001-02) of historical weather with 10,232 associated avalanche events from the Jackson Hole Mountain Resort (JHMR) to investigate the spatial patterns of hard slab and loose snow avalanches and their relationship to wind speed, new snow density, and 24-hour maximum temperature. Based on our field observations, we hypothesize that 1) more hard slabs are observed in wind affected areas and more loose snow avalanches occur in protected areas, and 2) more hard slabs are observed with an increase in wind and more loose snow avalanches are associated with cold, calm conditions.

Study Site

This study uses historical avalanche and weather data recorded by the ski patrol at the Jackson Hole Mountain Resort which is located on Rendezvous Mountain in the southern end of the Teton Range in northwestern Wyoming, USA (43º 36' N, 111˚ W). The resort ranges in elevation from 1923 m to 3185 m. The Jackson Hole Mountain Resort is roughly 1000 km from the Pacific Ocean, its nearest moisture source, resulting in an intermountain climate (Mock and Birkeland, 2000). McCollister et al. (2003) present a more complete description of this study site (Figure 3).

Methods

Spatial Distribution

Using ArcView 8.1, we created three maps to visually analyze the total count, the percent hard slab, and the percent loose avalanche for each slide path. While soft slabs are by far the most common type of avalanche event, we were interested in the occurrences of the rarer hard slab and loose avalanche events. The maps were visually analyzed to determine if similar avalanche types occurred in close proximity. Mapping the total count for each slide path allowed us to visualize the spatial distribution of the common and uncommon avalanche paths. The percent hard slab map displayed the ratio of hard slabs to the total number of avalanches for each slide path. Like the hard slab map, the loose avalanche map depicted the percentage of loose avalanches and was calculated the in the same manner as the percent hard slab map.

Relationship to Weather Variables

We used methods similar to McCollister *et al.* (2003) to relate avalanche occurrences to weather variables using the program GeoWAX (Geographic Weather and Avalanche Explorer). Geographic Visualization and Geographic Knowledge Discovery have the primary goal of finding interesting patterns and relationships in large spatial datasets (Miller and Han, 2001; MacEachren *et al*. 1999). GeoWAX, an iterative, interactive program that displays multiple perspectives of the data, utilizes these geographic concepts, allowing the user to explore spatial historical weather and avalanche data for visualization, pattern discovery, and hypothesis generation. Stoffel et al. (1998) were among the first to demonstrate the usefulness of Geographical Information Systems (GIS) to analyze avalanche data. The data-mining algorithm we used is a nearest neighbor approach similar to Buser (1983, 1989); see McCollister *et al.* (2003) for a more complete discussion of GeoWAX.

In this study, we only considered avalanche occurrences if the desired avalanche type occurred. For example, consider an avalanche path that slid on 25 days out of 100 near days. The overall mean probability is 25%. If three of those avalanches were hard slabs, 20 were soft slabs, and two were loose avalanches, there would be a 3%, 20%, and 2% chance of encountering the respective avalanche type, and a 75% chance of the slide path not sliding. Using the same example, the slide path can be characterized by the proportion of slides for a given avalanche type. In this example, 12% of the avalanches (3% of 25%) would be hard slabs, 80% (20% of 25%) would be soft slabs, and 8% loose avalanches (2% of 25%). These methods can be used to describe individual slide paths, slide paths grouped by aspect and elevation, and for the mean of all slide paths based on specific weather variable values.

For this study we analyzed the slide path mean for hard slabs and loose avalanches using wind speed, 24-hour maximum temperature, and new snow density. Days were only considered if they received at least 10 cm of new snow. All three variables were weighted equally for the nearest neighbor search, and near days were weighted more heavily using the inverse of the square root of the nearest neighbor Euclidean distance. A set of series signatures (McCollister *et al.,* 2003) was created using three different wind speeds (10 m/s, 20 m/s, and 30 m/s), five different 24-hour maximum temperature (-12.2º, -9.4º, -6.7º, -3.9º, and -1.1º C), and four density (20, 60, 100, and 140 kgm^{-3}) target values.

Hard slabs and loose avalanches are relatively rare compared to soft slabs, which comprise the vast majority of avalanche events at the JHMR. Hard slabs comprise 4.63% (474 of 10,232)) of all the avalanches in our database, and dry loose snow avalanches comprise 3.73% (382 of 10,232). Because we were examining rare events, our sample size was effectively reduced. To minimize this effect, we used the mean of all slide paths to calculate the proportion of different avalanche types for hard slabs and loose avalanches.

Statistical Analysis

We performed two non-parametric statistical tests for the overall mean of both avalanche types. First, we performed a Mann-Whitney U test to compare the medians of two different target variables of one variable. For example, consider the target value comparison of snow with a density of 20 kgm^{-3} to snow with a density of 60 kgm^{-3}. All target values of wind speed and 24-hour maximum temperature would be pooled, creating a sample size of 15 (3 x 5) for each avalanche type. The second non-parametric test was a paired comparison of medians using the Wilcoxon Signed Rank test. The pairing was based on corresponding target variables of wind speed and 24-hour maximum temperature for the two avalanche types. Using the target value comparison of 20 kgm^{-3} snow with 60 kgm^{-3} snow earlier, this test would compare 15 pairs.

Results and Discussion

Spatial Distribution

The frequency map (Figure 17) enabled us to spatially view the relative frequency of all slide paths and identify common individual slide paths and common groups of slide paths. The most commonly observed slide path was Laramie 4 with a total of 620 avalanche events. The next most common slide path was Hanging Rock with 564 avalanche events. Some slide paths in similar geographic areas have similar total counts, but other areas do not. Although this type of analysis is subjective in nature, it revealed some interesting trends. Areas with more uniform terrain, such as the Laramie and Cheyenne group, seemed to have less variation in their avalanche counts. Unlike the Laramie and Cheyenne groups, the slide paths in Casper Bowl and the Cirque have more varied terrain, cover a larger area, and have more variation in their avalanche counts. This would also apply to Cheyenne Bowl as a whole.

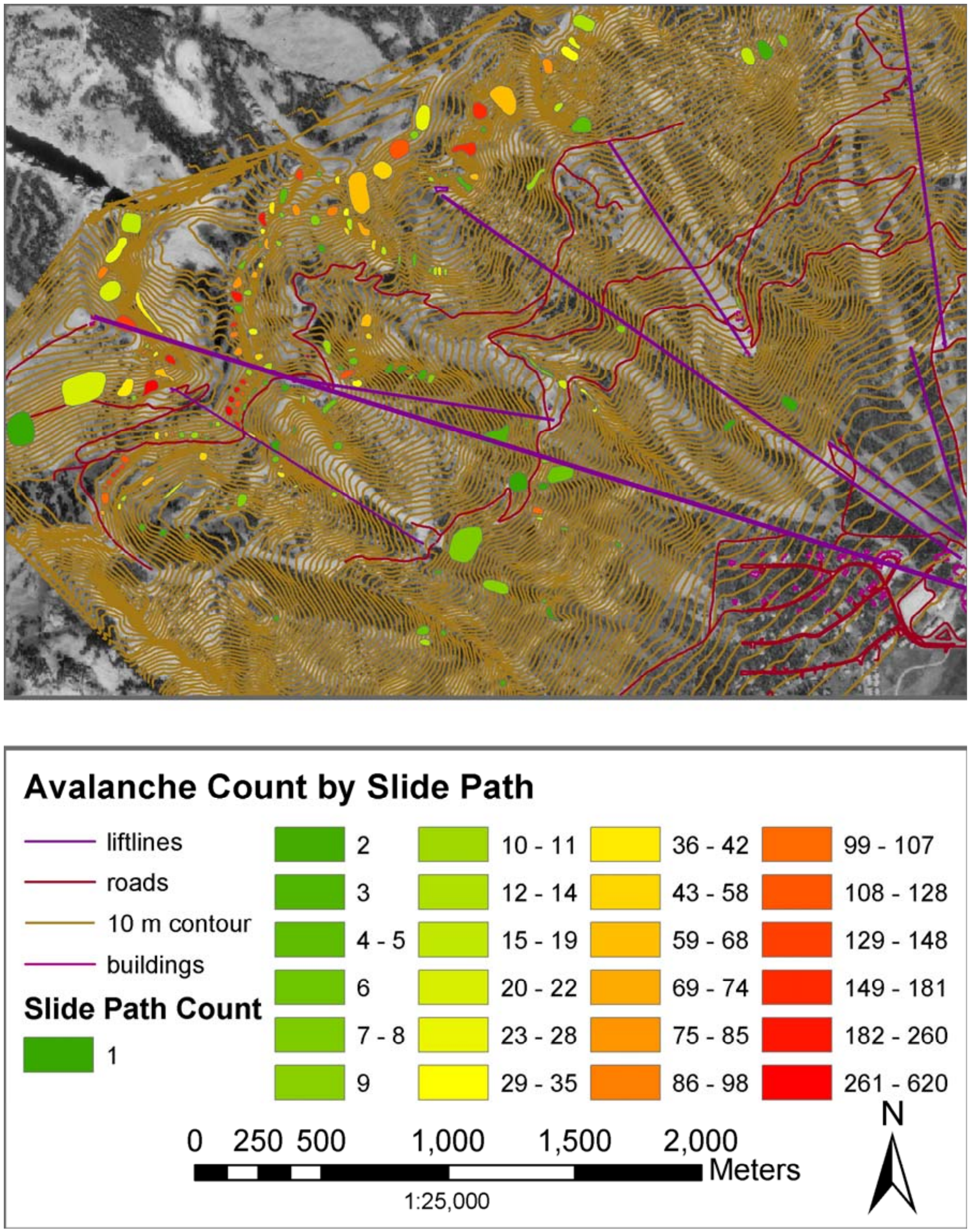


Figure 17: Avalanche Count by Slide Path.

Similar to the frequency map, the percent hard slab map (Figure 18) enabled us to identify slide paths and groups of slide paths that had high and low frequencies of hard slabs. There appeared to be less spatial variation (i.e., more similarity between adjacent slide paths) with percent hard slabs than with the total count. In other words, even when two slide paths had different overall frequencies, they still had commonalities with nearby slide paths in regards to their proportion of hard slabs. For example, most slide paths in the Cirque rarely released as hard slabs, with the few exceptions being paths that slid only rarely (e.g., Cirque 3S with one hard slabs out of five total avalanches). Slide paths in the upper Laramie group also all have a low proportion of hard slabs. In contrast, slide paths in the Cheyenne group have a higher proportion of hard slabs than slide paths in the cirque, and all have fairly similar proportion of hard slabs. There also appeared to be some geographic location relationships. Some of the more exposed ridges such as the Headwall, North Ridge, and the Far drift all have a high proportion of hard slabs. Increased exposure to wind may have increased the amount of hard slabs for these particular slide paths. This effect may be artificially high for slide paths in the North Ridge because they are typically controlled on the second day after a storm. The additional time between the deposition of the snow and the avalanche mitigation measures might allow more time for more settlement and consolidation, thereby resulting in more hard slabs.

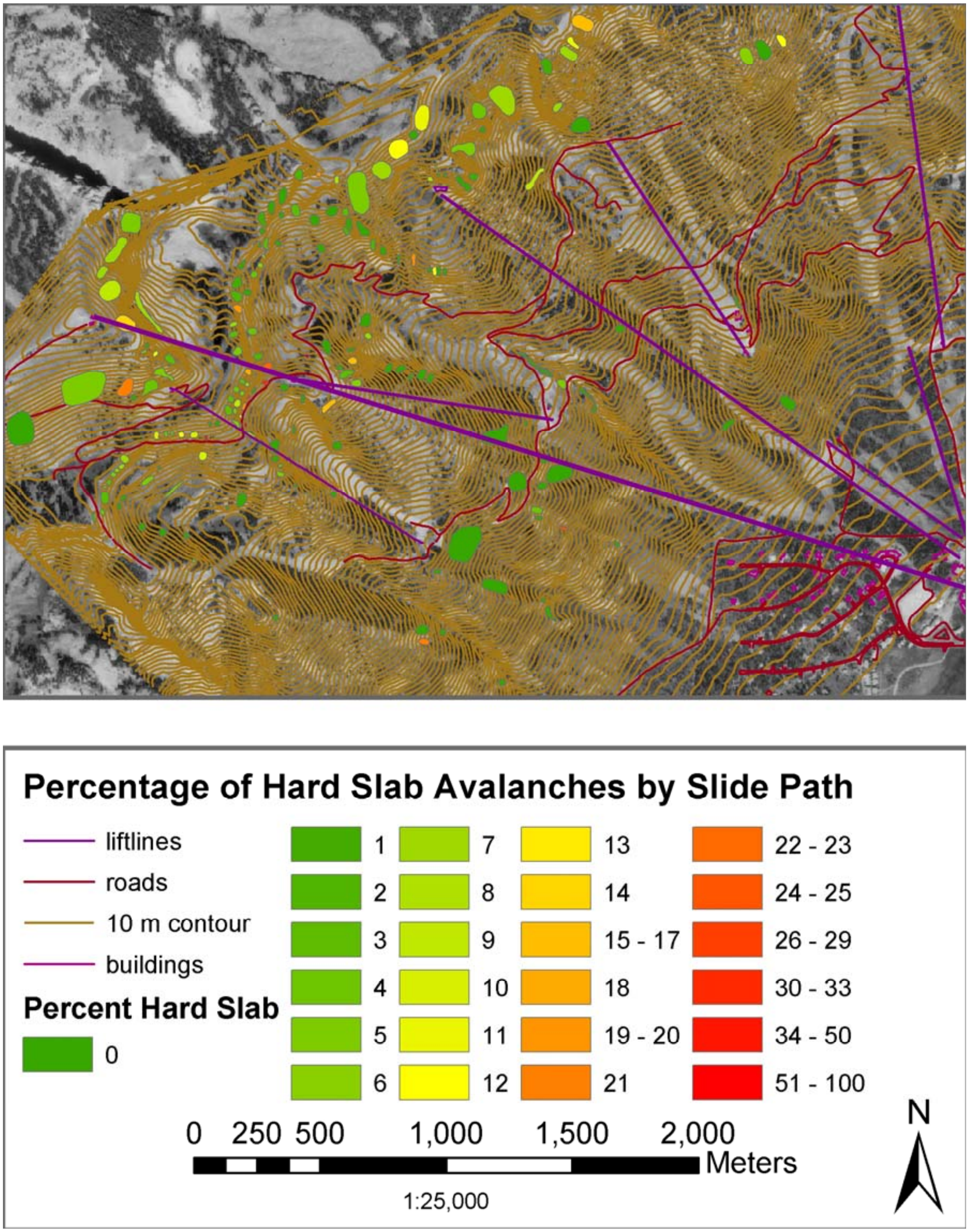


Figure 18: Percentage of Hard Slabs by Slide Path.

We noted two other trends for paths with a higher proportion of hard slabs: some slide only rarely, and others are paths that have gentler slope angles. This first relationship can be seen by comparing the upper Laramie group, which commonly avalanche, to the rarer sliding lower Laramie A and Laramie B groups. The Laramie group had a low proportion of hard slabs, while the Laramie A and B groups had a higher proportion of hard slabs. A group of adjacent slide paths in Cheyenne Bowl also exhibited this relationship. Old Reliable and Mudslide 1 were relatively common and had a lower proportion of hard slabs than the surrounding Fox's slide, mudslide 2, and mudslide 3 which were relatively rare and had a higher proportion of hard slabs. The second apparent trend with some slide paths, such as Dean's Slide, was a decrease in slope angle resulting in a higher proportion of hard slabs. Richmond (1994) calls these gentler slopes "stubborn" and the combination of the gentler slope angle and the higher proportion of hard slabs makes them especially dangerous. With over 21% of its 42 avalanche events being hard slabs, Dean's slide had the highest proportion of hard slabs for all slide paths. In contrast, the adjacent Hanging Rock, with similar wind loading features had a much lower proportion of hard slabs (4%). Hanging Rock is a commonly occurring slide path, and is steeper, and these two factors may relate to the low proportion of hard slabs. Our result has practical implications. Avalanches are sometimes difficult to assess on Dean's slide path, which has caught several avalanche workers, and killed one experienced ski patroller.

There are several geographic patterns for loose avalanches and these patterns are also often opposite to the patterns observed for hard slabs (Figure 19). Slide paths in the

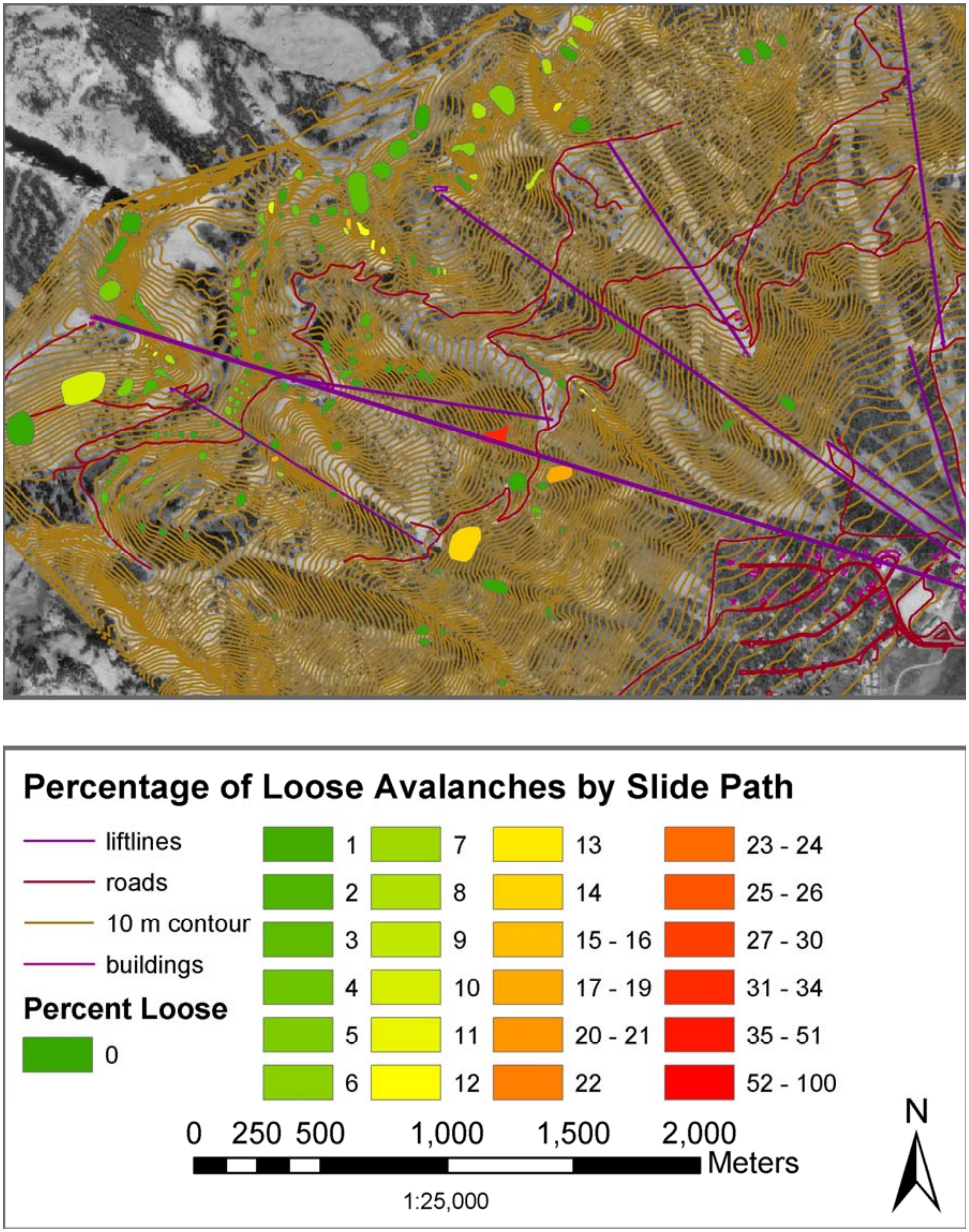


Figure 19: Percentage of Loose Avalanches by Slide Path.

Cheyenne group exhibited a relatively low proportion of loose avalanches. In contrast, slide paths in the Laramie group had a higher proportion of loose avalanches. Higher still is a group of sheltered slide paths in the Cirque which include Cirques 20, 20a, 21, 21a, 21b. All of these slide paths are further from their nearest ridge (Headwall) than other nearby slide paths, and are mostly sheltered by trees. Slide paths in Casper Bowl also exhibit a high proportion of loose avalanches. The ridges that had a high proportion of hard slabs (Headwall, North Ridge, Far Drift) all had a low proportion of loose avalanches. Similar to rare slide paths with high hard slab proportions, rare slide paths may also have artificially high loose proportions such as Riverton, Rawlins, and Lander Bowls.

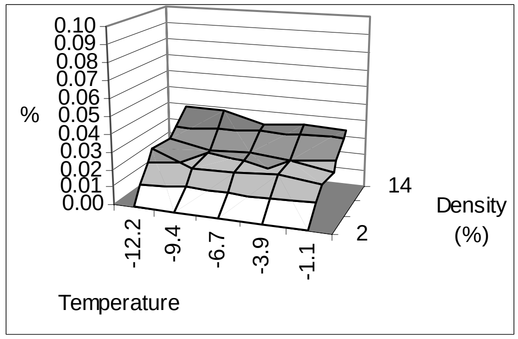
Relationship to Weather Variables

When we related wind speed, 24-hour maximum temperature, and density to the mean proportion for all slide paths we found significant trends for two weather variables with loose avalanches, and one for hard slabs. Slide paths had a significant increase in the mean proportion of hard slabs with a 40 kgm^{-3} increase in density (Table 2, Figure 20). Conversely, there was not a significant trend for wind speed or 24-hour maximum temperature with either test. We suspect the observed density trend may be due to two processes. First, denser snow may have resulted in denser, harder slabs. This may have been the case with small (class 1 and 2) new snow hard slab avalanche events. A second possibility is the rapid increase in weight associated with denser snow may have resulted in the triggering of deeper, harder layers. This kind of hard slab avalanche

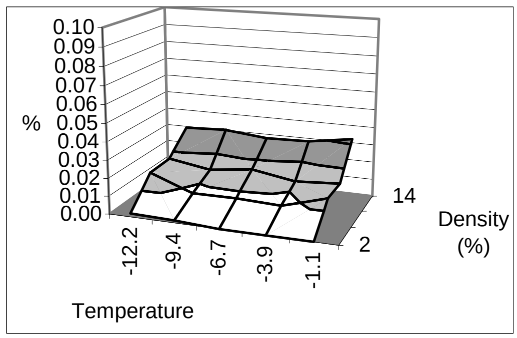
Table 2: Hard Slab Avalanche Statistics

Hard Slabs		
Wind Speed	**Mann-Whitney U**	**Wilcoxon Signed Rank**
Low vs. Moderate	0.165	**0.001**
Low vs. High	0.565	0.199
Moderate vs. High	0.738	0.184
Density		
2% vs. 6%	**0.000**	**0.001**
2% vs.10%	**0.000**	**0.001**
2% vs. 14%	**0.000**	**0.001**
6% vs. 10%	**0.038**	**0.036**
6% vs. 14%	**0.000**	**0.001**
10% vs. 14%	**0.000**	**0.001**
Temperature		
-12.2º vs. -9.4º	0.630	**0.010**
-12.2º vs. -6.7º	0.378	**0.003**
-12.2º vs. -3.9º	0.242	**0.004**
-12.2º vs. -1.1º	0.219	**0.010**
-9.4º vs. -6.7º	0.843	0.110
-9.4º vs. -3.9º	0.514	0.075
-9.4º vs. -1.1º	0.671	0.213
-6.7º vs. -3.9º	0.843	0.386
-6.7º vs. -1.1º	0.713	0.508
-3.9º vs. -1.1º	0.932	0.678

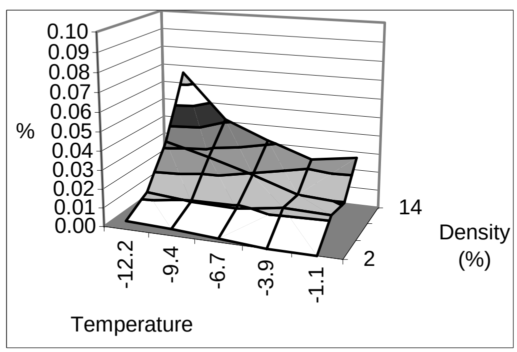
event may have been classified as a hard slab due to the hard old snow layers and may be the case with large (class 4, 5) avalanche events. To further explore the differences in size between soft and hard slabs, the mean crown depth was calculated for each avalanche type (loose avalanches could not be used because depth is typically not possible to record). Soft slabs had a mean depth of 35 cm, while hard slab avalanches had a mean depth of 75 cm, and were significantly different (t-test, $p < 0.000$). This large difference in depth between soft and hard slabs suggests that most hard slabs are larger



(a) Low wind



(b) Moderate wind



(c) High wind

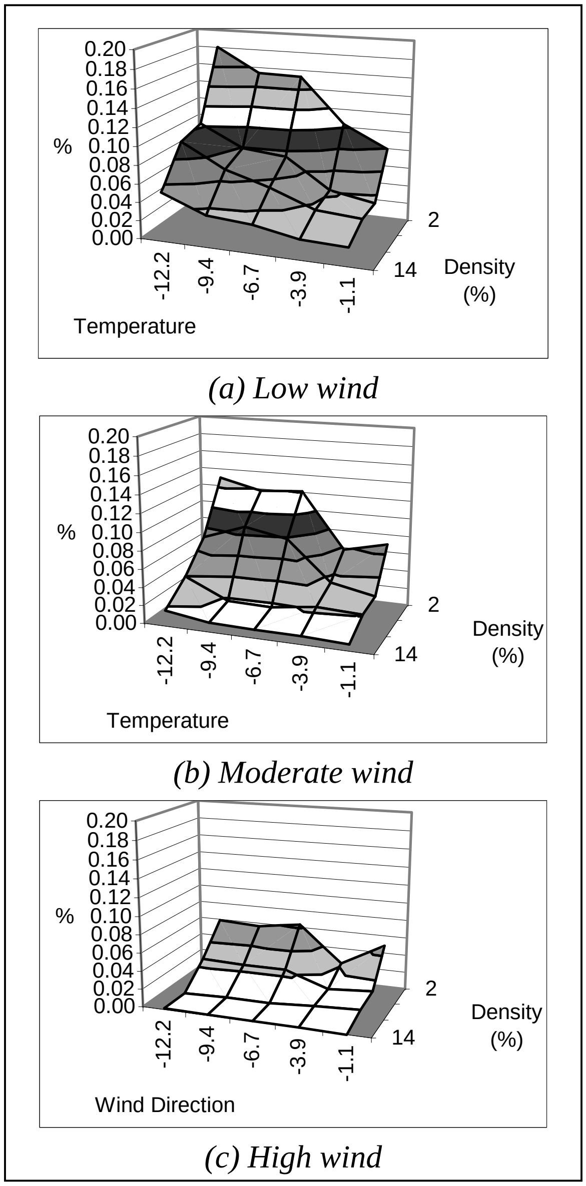
Figure 20: Increase of Density Produces Significant Increase of Hard Slab Avalanches.

events. Additionally, because soft slabs and hard slabs might be mis-classified, the difference between the two types is blurred or “evened-out”, which would result in less significant findings. Despite this fact, I still found very significant differences between the depth of soft slabs and hard slabs.

Dry loose avalanches are typically new snow events in a ski area setting because un-skied powder snow is often preferred by skiers, resulting in daily skier compaction of slide paths, which disrupts layer formation. In contrast to hard slabs, slide paths experienced a significant increase in the proportion of loose avalanches with both a decrease in wind and new snow density based on both tests (Table 3, Figure 21). A decrease in the 24-hour maximum temperature also resulted in a significant increase in loose avalanches in eight of 10 tests using the Wilcoxon Signed Rank- test (p = 0.004-0.028), though significance was not found with the Mann-Whitney U test. Low wind

Table 3: Loose Avalanche Statistics

Loose Avalanches		
Wind Speed	**Mann-Whitney U**	**Wilcoxon Signed Rank**
Low vs. Moderate	**0.043**	**0.000**
Low vs. High	**0.000**	**0.000**
Moderate vs. High	**0.004**	**0.000**
Density		
2% vs. 6%	**0.011**	**0.001**
2% vs.10%	**0.000**	**0.001**
2% vs. 14%	**0.000**	**0.001**
6% vs. 10%	**0.041**	**0.001**
6% vs. 14%	**0.000**	**0.001**
10% vs. 14%	0.098	**0.001**
Temperature		
-12.2º vs. -9.4º	0.671	**0.019**
-12.2º vs. -6.7º	0.630	**0.028**
-12.2º vs. -3.9º	0.114	**0.004**
-12.2º vs. -1.1º	0.101	**0.004**
-9.4º vs. -6.7º	0.729	0.084
-9.4º vs. -3.9º	0.248	**0.012**
-9.4º vs. -1.1º	0.219	**0.012**
-6.7º vs. -3.9º	0.299	**0.015**
-6.7º vs. -1.1º	0.326	**0.012**
-3.9º vs. -1.1º	0.908	0.638



(a) Low wind

(b) Moderate wind

(c) High wind

Figure 21: Decrease of Wind Speed and Density Produce Significant Increase of Loose Avalanches.

speeds, low 24-hour maximum temperatures, and low new snow density resulting in higher proportions of loose avalanches is consistent with our observations of snow and avalanches. Cold snow results in slower bond formation between new precipitation particles. With low density snow there is a smaller number of bonds for a given volume than denser snow. Finally, low wind results in less mechanical breakdown of precipitation particles and slows the formation of new bonds between particles.

Conclusions

Specific slide paths with a high frequency of avalanche activity have a lower proportion of hard slabs. These slide paths may have a lower threshold of release, resulting in more small soft slab avalanches, which does not allow for the build up of snow and large hard slab avalanches. In contrast, low frequency slide paths may have an increased build up of new snow between avalanche events, which can result in large events if the failure layer is a deep layer. This is particularly true for slide paths with gentler slope angles. Slide paths near each other often have similar avalanche type proportions. This trend is most apparent with slide paths that have that have similar characteristics, such as absence or presence of trees, similar shape, aspect, and wind loading patterns. Slide paths that rarely slide have low sample sizes. Rare events have the potential to be biased by random variation, and this randomness can have a large impact on ratios for individual slide paths.

Our results of high-density snow leading to increased proportion of hard slabs, and low wind speeds and cold, light density snow leading to a higher proportion of loose avalanches, are not surprising. However, since this matches well with field observations, these results do help confirm the validity of our methods. The broader impact of our work is that these same methods can also be used to explore relationships between other avalanche event attributes (i.e., depth, sliding surface, trigger, run-out distance) and specific weather variables. Similarly, the spatial component of the slide paths (aspect, slope, elevation, tree-cover, and distance from ridge) could also be incorporated.

CONCLUSIONS

This thesis presented two case studies to help verify the new methods created in developing GeoWAX. The first case study related patterns of avalanching to weather by investigating the effect of new snowfall, wind speed, and wind direction on the spatial location of avalanche formation at Jackson Hole Mountain Resort. New snowfall increased avalanche activity at all scales. However, it did not play a significant role in differentiating avalanche activity between individual slide paths. Adjacent slide paths often behaved similarly with respect to wind loading, and there was no direct relationship between wind direction and the aspect of wind-loaded slopes. These results suggest that different wind directions result in specific repeatable patterns of wind loading, but those patterns are not as simple as westerly winds loading east-facing slopes. Instead, specific avalanche patterns likely result from the interaction of free air winds with complex topography resulting in path specific patterns of wind loading.

The second case study builds on the methods of the first and relates weather to avalanche types by exploring the spatial variability of hard slab and dry loose avalanches occurring at the Jackson Hole Mountain Resort over the last 23 seasons. These avalanche types are characterized with respect to their geographic location and associated weather conditions. It is important to understand where the different types of avalanches occur and under what weather conditions. Slab avalanches are much more dangerous than loose avalanches because slab avalanches are often triggered well down slope of the crown, and therefore are much more likely to catch the person that triggered the avalanche. I analyzed these data with and without the incorporation of three weather variables (wind speed, 24-hour maximum temperature, and new snow density). Hard slab avalanches increased significantly with increasing density and loose avalanches increased with decreases in temperature and new snow density. Specific slide paths with a high frequency of avalanche activity have a lower proportion of hard slabs, while low frequency slide paths have a higher proportion of hard slabs; slide paths near each other often have similar avalanche type proportions.

Both case studies confirm inherent knowledge about avalanches that forecasters have learned through experience. In doing so, these results help to verify the methods developed in the GeoWAX program. The greatest contribution of this project is that the techniques used in developing GeoWAX have been shown to be useful in visualizing, exploring, and analyzing both geographic and avalanche attributes by using a combination of a traditional meteorological forecasting technique (Nearest Neighbors) with Geographic Knowledge Discovery concepts. The combination has produced visualization tools that are useful in four ways (Figure 22). First, the tools can be used for research to better understand the dynamic spatial and temporal pattern of snow and avalanches under a variety of weather conditions. Second, GeoWAX could be used for daily forecasting, specifically for Type B forecasts (ski areas and highway operations). Third, these tools could be used to help refresh the memory of seasoned avalanche professionals for specific areas. Finally, GeoWAX could be used as an educational tool for new avalanche professionals with little or limited experience for a specific forecast area.

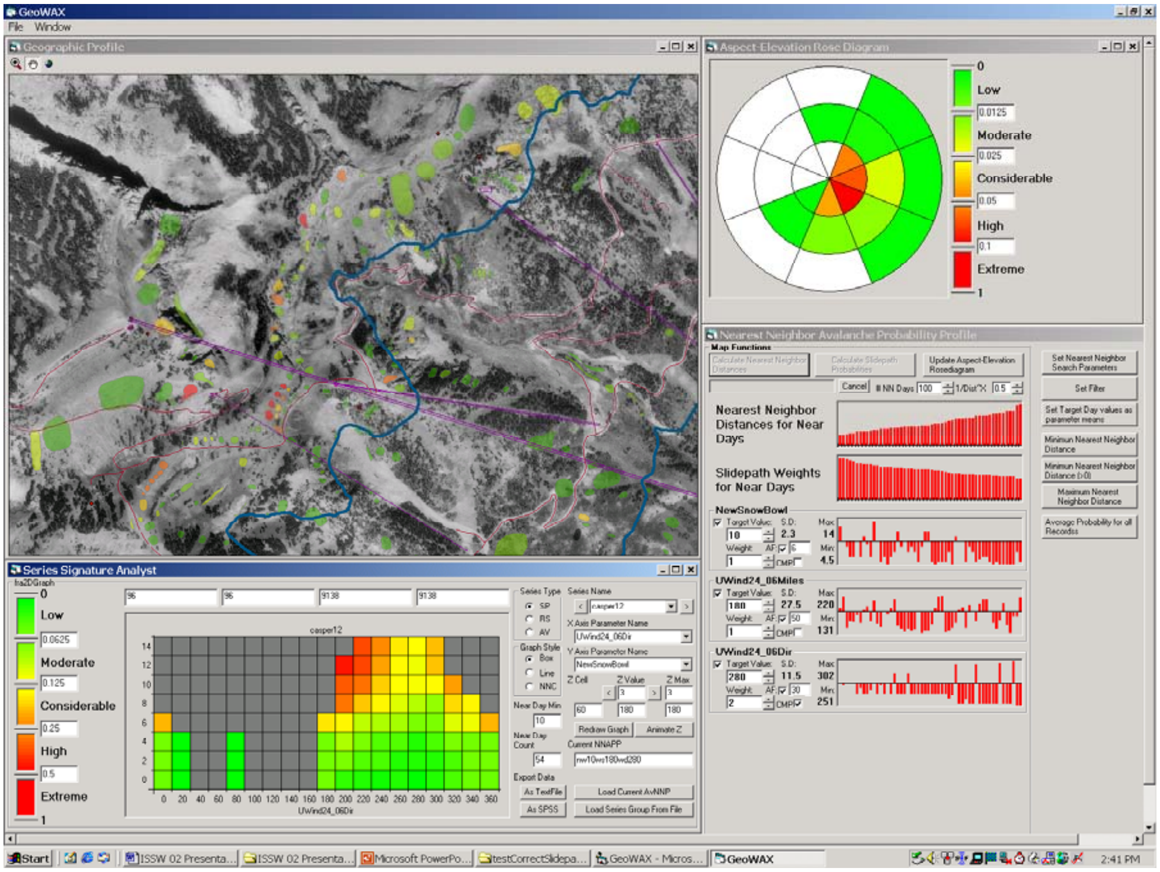


Figure 22: Example of GeoWAX.

Project Difficulties and Recommendations for the Future

Digitizing Historical Data

One of the biggest hurdles for this project was managing the massive amounts of data. The first difficulty was digitizing the data. At the onset of this project all of the data were still in a paper format. Specific programs were developed to facilitate data entry using Microsoft Access. I recommend that ski areas, highway departments, and other snow safety operations that collect this type of data immediately initiate the process of digitally archiving the data, even if they are not currently analyzing the historical data.

<u>Naming Issues</u>

The next hurdle was to create a consistent naming format for the slide paths. Twenty-three years of historic data recorded by many different people led to multiple names for many of the slide paths. Originally, all of the known slide paths were named with a group and number (i.e. Laramie 4). However, today many of the slide paths have been renamed. Some of this renaming was done for ease of interpretation, such as Laramie 23 becoming Alta 3. Others have been renamed after a ski patroller has been caught by a specific avalanche, such as Dean's slide, Jeff's slide, Frosty's slide, and Cooke's Knob. Finally, poor spelling and typos introduced a lot of problems. For example, AMP ROCKS, amp. Rock, AMP. ROCKS, AMPHITHEATER ROCKS, AMPHITHEATER ROCK, AMPHITHEATRE RCKS all designate Amphitheater Rocks. In this study I used 188 slide paths that were spelled 3566 different ways. To work through this problem I again created specific programs to help deal with the naming convention. It is recommended that a consistent naming scheme be developed and adhered to. Ideally, if the data were being archived digitally, the avalanche events would be archived by selecting slide paths from a list instead of being manually entered, which would eliminate spelling and typing issues.

Accuracy of Historical Data

A third problem was the accuracy of the original data. This is usually human-caused error. Typically, the avalanche forecaster diligently records the data. However, these avalanche data are relayed to the forecaster by the individual ski patrollers, creating an opportunity for miscommunication. Large avalanches almost always get communicated. However, smaller avalanches, particularly class 1 avalanches, are not always communicated, due to their high abundance. I recommend that snow safety operations implement a standard protocol where all avalanche workers would enter their own data immediately after avalanche control, which would then be rechecked by the forecaster before becoming part of the permanent record.

Accuracy of Geographic Data

The final potential problem is the accuracy of the geographic data. The 5-meter Digital Elevation Model that was used was created from a 10-foot contour map created by the Engineering Department of the Jackson Hole Mountain Resort. Without these data, I would have had to use a 10-meter or 30-meter USGS DEM. As the grid size increases, it becomes smoothed and less accurate, especially for smaller slide paths. This becomes critical when creating rates of change, such as slope or aspect. Slope is even more problematic than aspect because aspect has of range of 360º, while slope has a range of only 90º. Making matters even more complicated is that the critical range of avalanches is between 30º and 45º, which creates a practical range of only 15º for slope. It is recommended that a larger scale map be developed for ski areas or where forecasts are being made, and that careful analyses be done in utilizing the slope and aspect data. Perhaps a secondary check of these accuracies should be a part of the standard use of these tools.

Ideas for Future Research

There are many research possibilities using the current functionality of GeoWAX by simply varying which variables to use in nearest neighbors searches. Climatic studies could also be performed. These include investigating the role of El Niño years versus La Niña years on the pattern of avalanches, or how the patterns of avalanching change for different seasonal classifications (coastal, intermountain, or continental) for a specific location using the scheme developed by Mock and Birkeland (2000).

There are also at least three research topics associated with nearest neighbors. First, should rare target days and common target days be handled the same way? This reflects on what are the optimal number of days, and are they always the same? A second possible research topic is the use of r-nearest neighbors versus k-nearest neighbors. In this study I used k-nearest neighbors, which is a fixed number of nearest neighbors. The r-nearest neighbors technique finds all of the nearest neighbors in a distance radius r, and all neighbors within this radius are used. This approach might be a solution to the first problem, but would lead to the question of the optimal r. The third nearest neighbor research possibility would be to explore different distance measurements. One in particular is the manalohobis distance, which incorporates covariance between variables. This would be particularly important when using highly correlated variables such as snow water equivalence and new snow. Other distance measurements include varying the power of the individual variable distance differences. For example, the Euclidean distance is squared, but other exponents could be used, such as one, or even less than one.

Finally, the techniques in the case studies could be combined to relate geographic attributes to avalanche attributes. For example they could be use to address the questions of whether wet slides are more prevalent on sunlit aspects or lower elevations, or are loose avalanches more prevalent on steeper slopes.

In summary, this research has produced many contributions to both the snow science community and the emerging field of Geographic Knowledge Discovery. This work:

- Combined variables to form a new factor, such as wind loading, which cannot be directly measured.
- Developed new methods to manage, explore, and analyze large weather and avalanche data sets, which will become much more common in the future.
- Introduced the benefits of Geographic Knowledge Discovery for the snow science community via the creation of an iterative tool that demands high interaction between the user and the computer by using multiple perspectives of the data to take advantage of natural pattern recognition ability of human beings.
- Provided a simplified system for analyzing ski area data, resulting in fewer confounding factors and more direct information on the specific processes under evaluation.

- Introduced of the potential harmful effects of using high dimensions in nearest neighbor searches with use of weather and avalanche data sets.
- Introduced the use of inverse distance weighting for nearest neighbors distance metric with the use of weather and avalanche data sets.
- Provided a probability analysis at the slide path scale, which produces spatial patterns of avalanche activity.
- Created a tool that can be used to test other hypotheses, such as the changes of the temporal and spatial pattern of avalanches for different seasonal weather classifications.
- Provided a better tool for utilizing the available data for both avalanche forecasters and for the snow science community.
- Provided a better understanding of wind loading, demonstrating the occurrence of patterns at different scales.
- Provided a better description and understanding of hard slabs and loose avalanches.

In the end, the major contributions of this research are twofold. Firstly, it has resulted in the development of a useful operational avalanche-forecasting tool that can also be utilized to explore weather and avalanche datasets and to generate and test hypotheses. Secondly, this work has provided new insights into processes leading to the development of patterns and scales of avalanche behavior. Hopefully others can utilize some of the techniques developed here for further research.

REFERENCES CITED

Aggarwal, C.C., A. Hinneburg, and D.A. Keim. 2001. On the surprising behavior of distance metrics in high dimensional spaces. *Proceedings of the 8th International Conference on Database Theory.* ICDT 2001, London, UK, Jan. 4-6, 2001, 420-434.

Andrienko, G.L. and N.V. Andrienko. 1999. Interactive maps for visual data exploration. *International Journal of Geographic Information Science,* 13(4), 355-374.

Andrienko, N., G. Andrienko, A. Savinov, H. Voss, D. Wettschereck. 2001. Exploratory analysis of spatial data using interactive maps and data mining. *Cartographic and Geographic Information Science*, 28(3), 151-165.

Baily, T. and A.K. Jain. 1978. A note on distance-weighted k-nearest neighbor rules. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC 8(4), 311-313.

Beyer, K., J. Goldstein, R. Ramakrishnan, and U. Shaft. 1999. When is "nearest neighbor" meaningful? *Proceedings of the 7th International Conference on Database Theory,* Jerusalem, Israel, 217-235.

Birkeland, K.W., R.F Johnson, and D. Herzberg. 1996. The stuffblock snow stability test. U.S. Forest Service Missoula Technology Development Center publication 9623-2836-MTDC, Missoula, Montana.

Birkeland, K.W. 1998. Terminology and predominant processes associated with the formation of weak layers of near-surface crystals in the mountain snowpack. *Arctic and Alpine Research,* 30(2), 193-199.

Birkeland, K.W., R.F. Johnson, and D.S. Schmidt. 1998. Near-surface faceted crystals formed by diurnal recrystallization: A case study of weak layer formation in the mountain snowpack and its contribution to avalanche formation. *Arctic and Alpine Research,* 30(2), 199-204.

Birkeland, K.W. 2001. Spatial patterns of snow stability throughout a small mountain range. *Journal of Glaciology,* 47(157), 176-186.

Birkeland, K.W., C.J. Mock, and J.J. Shinker. 2001. Avalanche extremes and atmospheric circulation patterns. *Annals of Glaciology*, 32, 135-140.

Birkeland, K., K. Kronholm, M. Schneebeli, and C. Pielemeier. *In Press.* Changes in the shear strength and micro-penetration hardness of a buried surfaces hoar layer. *Annals of Glaciology.*

Brabec, B. and R. Meister. 2001. A nearest neighbor model for regional avalanche forecasting. *Annals of Glaciology,* 32, 130-134.

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

Buser, O. 1989. Two years experience of operational avalanche forecasting using the nearest neighbors method. *Annals of Glaciology,* 13, 31-34.

Cooperstein, M.S., K.J. Hansen, and K.W. Birkeland. *In press.* The effects of slope aspect on the formation of surface hoar and diurnally recrystalized near-surface faceted crystals and implications for avalanche forecasting. *Proceedings of the 2004 International Snow Science Workshop, Jackson Hole, WY.*

Cover, T.M. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.

Dasarathy, B.V. (Ed.). 1991. *Nearest Neighbor (NN) Norms: NN pattern classification techniques.* Los Alamitos, CA, IEEE Computer Society Press.

Davis, R.E., and K. Elder. 1994. Application of classification and regression trees: selection of avalanche activity indices at Mammoth Mountain. *Proceedings of the 1994 International Snow Science Workshop,* Snowbird, UT, 285-294.

Davis, R.E., K. Elder, D. Howlett and E. Bouzaglou. 1996. Analysis of weather and avalanche records from Alta, Utah and Mammoth Mountain, California using classification trees. *Proceedings of the 1996 International Snow Science Workshop*, Revelstoke, BC, Canada, 14-19.

Davis, R.E., K. Elder, D. Howlett and E. Bouzaglou. 1999. Relating storm and weather factors to dry slab activity at Alta, Utah and Mammoth Mountain, California using classification and regression trees. *Cold Regions Science and Technology,* 30, 79-89.

Deems, J.S. 2003. Topographic effects on the spatial and temporal patterns of temperature gradients in a mountain snowpack. M.S. Thesis, Department of Earth Sciences, Montana State University. 85 pp.

Dudani, S.A. 1976. The distance-weighted k-nearest neighbor rule. *IEEE Transactions on Systems, Man, and Cybernnetics,* SMC 6(4), 325-327.

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

Elder, K., and R.E. Davis. 2000. Decision trees predicting avalanche response: tools for training? *Proceedings of the 2000 International Snow Science Workshop,* Big Sky, Montana, 140-146.

Evans, B.J. 1997. Dynamic display of spatial data-reliability: does it benefit the map user? *Computers and Geosciences,* 23(4), 409-422.

Fayyad, U.M., G. Piatetsky-Shapiro, P. Smyth. 1996a. From data mining to knowledge discovery. In *Advances in knowledge discovery and data mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, *Eds*. AAAI/MIT Press, Cambridge, Mass.

Fayyad, U.M., G. Piatetsky-Shapiro, P. Smyth. 1996b. The KDD process for extracting useful knowledge from volumes of data. *Communications of ACM,* 39(11), 27-34.

Fohn, P.M.B. 1987. The "rutschblock" as a practical tool for slope stability evaluation. In *Avalanche Formation, Movement, and Effects.* International Association of Hydrological Sciences Publication No. 162, pp.223-228.

Font, D., G. Furdada, J.M. Vilaplana. 2001. Aeolian susceptibility maps: methodology and applications. *Annals of Glaciology,* 32, 306-310.

Gahegan, M. 1999. Four barriers to the development of effective exploratory visualization for the geosciences. *International Journal of Geographic Information Science,* 13(4), 289-309.

Gahegan, M., M. Wachowicz, M. Harrower, T-M Rhyne. 2001. The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic Information Science*, 28(1), 29-44.

Gassner, M., H-J Etter, K. Birkeland, and T. Leonard. 2000. NXD2000: An improved avalanche forecasting program based on the nearest neighbor method. *Proceedings of the 2000 International Snow Science Workshop*, Big Sky, Montana, 52-59.

Hägeli, P. and D.M. McClung. 2000. A new perspective on computer-aided avalanche forecasting: scale and scale issues. *Proceedings of the 2000 International Snow Science Workshop,* Big Sky, Montana, 66-73.

Hägeli, P. and R. Atkins. 2002. Storage and visualization of relevant avalanche information at different scales. *Proceedings of the 2002 International Snow Science Workshop,* Penticton, BC, Canada, 32-38.

Han, J., R.B. Altman, V. Kumar, H. Mannila, D. Pregibon. 2002. Emerging scientific applications in data mining. *Communications of the ACM*, 45(8), 54-58.

Hand, D., H. Mannila, P. Smyth. 2001. *Principles of Data Mining*. The MIT Press, Cambridge, MA.

Hibbard, W. and D. Santek. 1989. Visualizing large data sets in the earth sciences. *Computer*, 22(8), 53-57.

Hinneburg, A., C. C. Aggarwal, and D.A. Keim. 2000. What is the nearest neighbor in high dimensional space? *Proceedings of The International Conference on Very Large Data Bases,* Cairo, Egypt, 506-515.

Jamieson, J.B. and C.D. Johnston. 1996. The compression test for snow stability. *Proceedings of the 1996 International Snow Science Workshop,* Banff, Alberta, Canada, pp. 118-125.

Kozak, M.C. 2002. The spatial and temporal variability of snow layer hardness. MS Thesis, Department of Earth Resources, Colorado State University, Fort Collins, Colorado, 162 pp.

Kozak, M., K. Elder, K. Birkeland and P. Chapman. 2003. Variability snow layer hardness by aspect and prediction using meteorological factors. *Cold Regions Science and Technology,* 37(3), 357-371.

Kraak, M-J., and A. MacEachren. 1999. Visualization for exploration of spatial data. *International Journal of Geographic Information Science,* 13(4), 285-287.

Kristensen, K. and C. Larsson, 1994. An avalanche forecasting program based on a modified nearest neighbor method. *Proceedings of the 1994 International Snow Science Workshop,* Snowbird, Utah, 22-30.

LaChapelle, E.R. 1980. The fundamental processes in conventional avalanche forecasting. *Journal of Glaciology,* 26(94), 75-84.

Landry, C.C., J.J. Borkowski, and R.L. Brown. 2001. Quantified loaded column stability test: mechanics, procedure, sample-size selection, and trials. *Cold Regions Science and Technology,* 33(2-3), 103-121.

Landry, C., K. Birkeland, K. Hansen, J. Borkowski, R. Brown and R. Aspinall. *In press.* Variations of snow strength and stability on uniform slopes. *Cold Regions Science and Technology.*

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

MacEachren, A.M. and M-J. Kraak. 1997. Exploratory cartographic visualization: advancing the agenda. *Computers and Geosciences,* 23(4), 335-343.

MacEachren, A.M., M. Wachowicz, R. Edsall, D. Haug and R. Masters. 1999. Constructing knowledge from multivariate spatiotemporal data: integrating geographic visualization and knowledge discovery in database methods. *International Journal of Geographic Information Science,* 13(4), 311-334.

MacEachren, A.M. and M-J. Kraak. 2001. Research challenges in geovisualization. *Cartographic and Geographic Information Science*, 28(3), 3-12.

MacLeod, J.E.S., A. Luk, and D.M. Titterington. 1987. A re-examination of the distance weighted k-nearest neighbor classification rule. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC 17(4), 689-696.

McClung, D.M. and Schaerer. 1993. *The Avalanche Handbook.* Seattle, WA, The Mountaineers

McClung, D.M. 2000. Predictions in Avalanche Forecasting. *Annals of Glaciology* 31(2), 377-381.

McClung, D.M. 2002a. The elements of applied avalanche forecasting Part I: The human issues. *Natural Hazards,* 26, 111-130.

McClung, D.M. 2002b. The elements of applied avalanche forecasting Part II: The physical issues and the rules of applied avalanche forecasting. *Natural Hazards,* 26, 131-146.

McCollister, C., K. Birkeland, K. Hansen, R. Aspinall, and R. Comey. 2003. Exploring multi-scale spatial patterns in historical avalanche data, Jackson Hole Mountain Resort, Wyoming. *Cold Regions Science and Technology,* 37(3), 299-313.

Miller, H.J. and J. Han. 2001. Geographic data mining and knowledge discovery. In *Geographic data mining and knowledge discovery*, H.J Miller and J. Han, *Eds.* Taylor and Francis, New York, NY.

Mock, C.J. and K.W. Birkeland. 2000. Snow avalanche climatology of the western United States mountain ranges. *Bulletin of the American Meteorological Society,* 81(10), 2367-2392.

Northwest Avalanche Center. 2004. http://www.nwac.noaa.gov/1950-2001_united_states_avalanche_fataliti.htm

Obled, C and W. Good. 1980. Recent developments of avalanche forecasting by discriminant analysis techniques: a methodological review and some applications to the Parsenn area (Davos, Switzerland). *Journal of Glaciology,* 25(92), 315-346.

Perla, R. and M. Martinelli. 1978. *Avalanche Handbook.* Agriculture Handbook 489, rev. ed., USDA Forest Service, Washington, D.C.

Purves, R., K. Morrison, G. Moss, B. Wright. 2002. Cornice – development of a nearest neighbors model applied in backcountry avalanche forecasting in Scotland. *Proceedings of the 2002 International Snow Science Workshop,* Penticton, BC, Canada, 117-122.

Ramakrishnan, N. and A.Y. Grama. 1999. Datamining: from serendipity to science. *Computer*, 32(8), 34-37.

Rhyne, T-M. 2000. Scientific visualization in the next millennium. *IEEE Computer Graphics and Applications*, 20(1), 20-21.

Richmond, D.P. 1994. Repeated mistakes by avalanche professionals. *Proceedings of the 1994 International Snow Science Workshop*, Snowbird, Utah, 449-458.

Schriber, B. 1998. Personal Communication. Director of Engineering for the Jackson Hole Mountain Resort.

Schweizer, J. 1999. Review of dry snow slab avalanche release. *Cold Regions Science and Technology,* 30(1-3), 43-58.

Stanfill, C. and D. Waltz. 1986. Toward memory-based reasoning. *Communications of ACM,* 29(12), 1213-1228.

Stoffel, A., R. Meister and J. Schweizer. 1998. Spatial characteristics of avalanche activity in an alpine valley – a GIS approach. *Annals of Glaciology,* 26, 329-336.

Tukey, J.W. 1977. *Exploratory data analysis*. Addison-Wesley Publishing Company, Reading, Massachusetts.

Wachowicz, M. 2001. GeoInsight: An approach for developing a knowledge construction process based on the integration of GVis and KDD methods. In *Geographic data mining and knowledge discovery*, H.J Miller and J. Han, *Eds.* Taylor and Francis, New York, NY.

Zhang, J., Y-S Yim, J. Yang. 1997. Intelligent selection of instances for prediction functions in lazy learning algorithms. *Artificial Intelligence Review,* 11, 175-191.

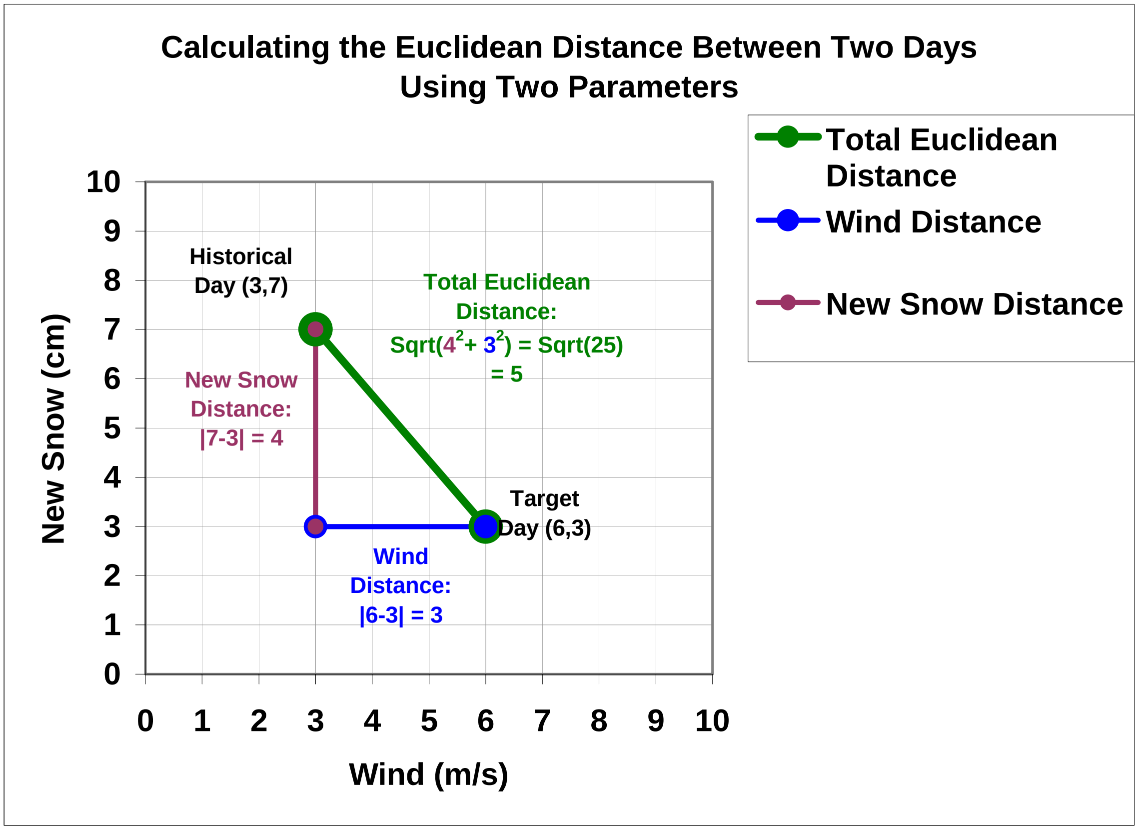
APPENDICIES

APPENDIX A

NEAREST NEIGHBOR GRAPH DEFINITIONS

Generating a Nearest Neighbor Distance Between a Target Day and a Historical Day

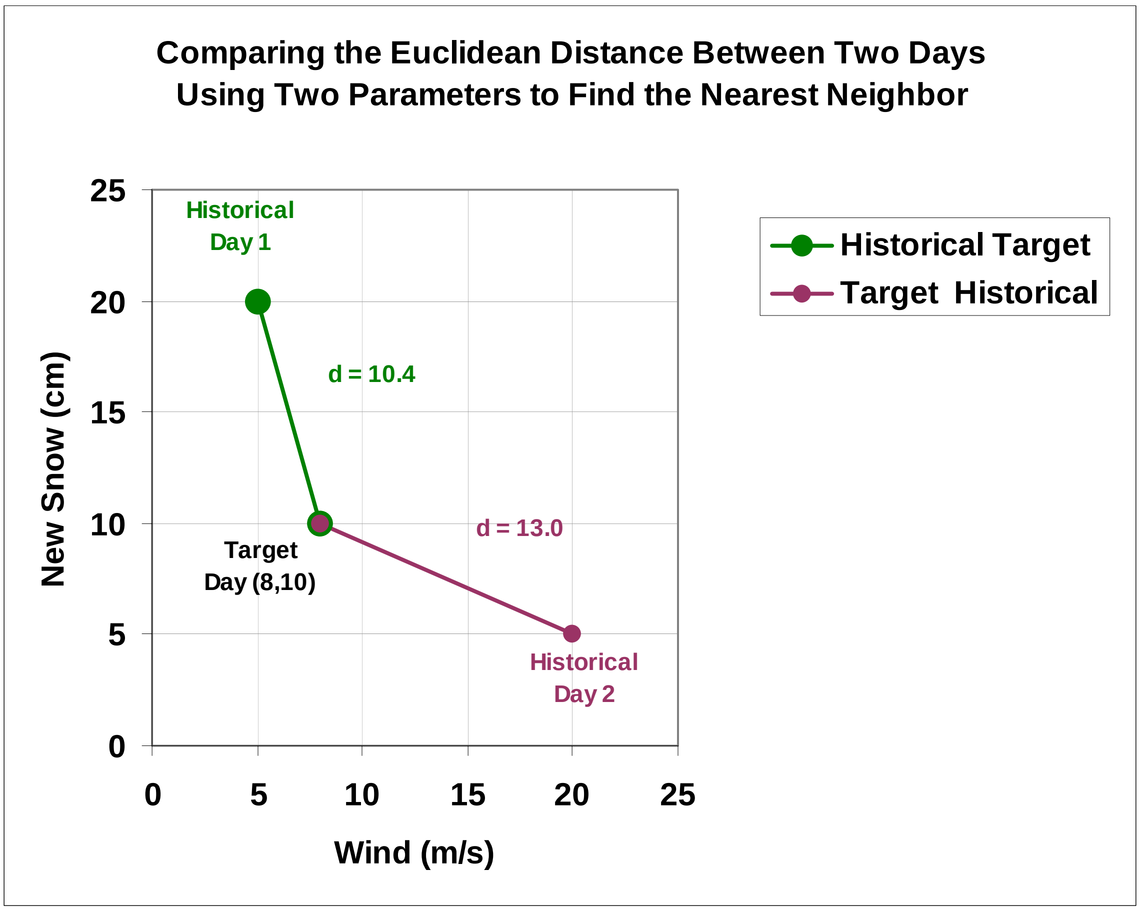
	Historical Day	Target Day
Wind	3	6
New Snow	7	3



The Euclidean Distance between two days using two parameters (wind and new snow) can be visualized as a right triangle. The side and the base of the right triangle are the distances for the individual parameters. The hypotenuse is the total distance and can be calculated using the Pythagorean theorem. The Pythagorean theorem states that $a^2 + b^2 = c^2$, where a and b are the side and base of a right triangle, and c is the hypotenuse. In this example a is the new snow distance, b is the wind distance, and c is the total Euclidean distance. It is easy to expand this idea to multiple parameters. For example, if three parameters are used the Euclidean distance would be $a^2 + b^2 + c^2 = d^2$. This can be visualized in three dimensions as a diagonal between two opposing corners of a cube. More than three parameters are difficult to visualize, but can easily be calculated using the general formula: Euclidean Distance = Sqrt{ Sum[$(t_i - h_i)^2$] }, where t_i is the i^{th} parameter of the Target Day and hi is the i^{th} parameter of the Historical Day.

Comparing Nearest Neighbor Distances between Two Historical Days

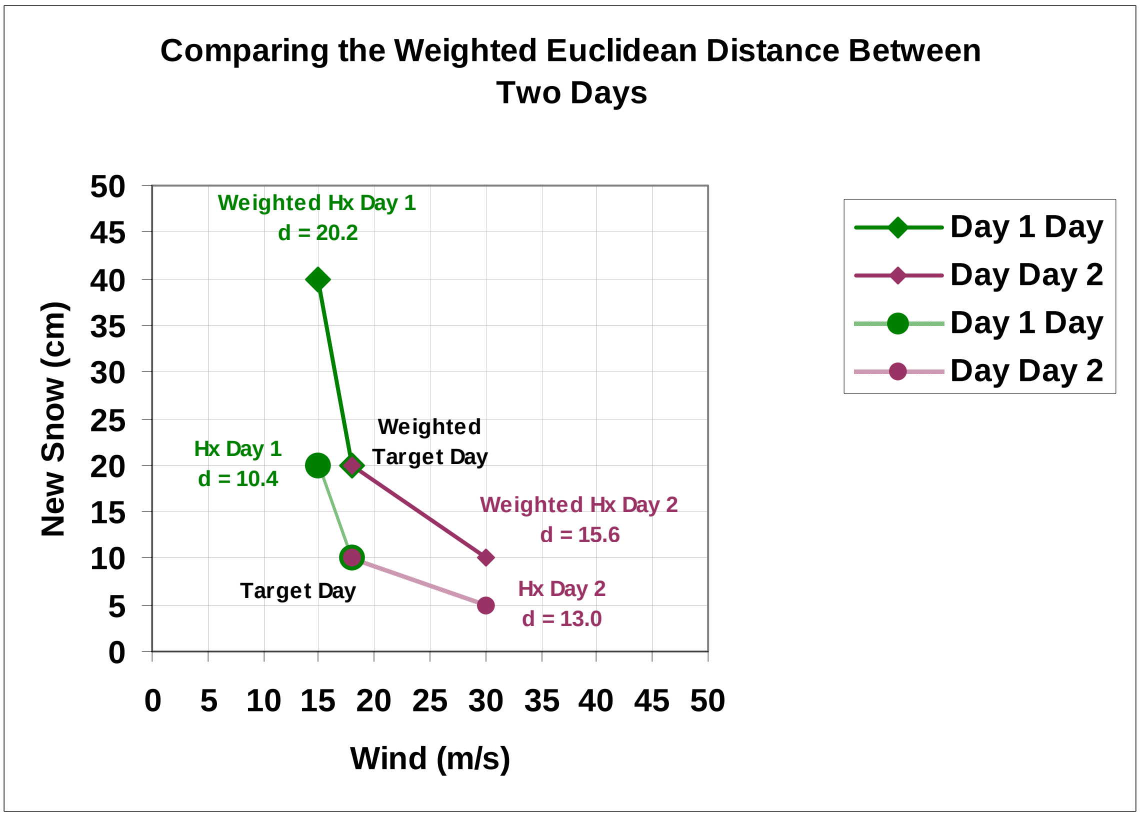
	Historical Day	Target Day	Historical Day
Wind	5	8	20
New Snow	20	10	5
Distance	10.4		13



Using the Pythagorean theorem, the distances from a Target Day to two Historical Days can be compared. The Target Day is defined as having 8 m/s average wind and 10 cm of new snow. Historical Day 1 had 5 m/s average wind and 20 cm of new snow, while Historical Day 2 had 20 m/s average wind and 5 cm of new snow. The distances from the Target Day to Historical Day 1 and Historical Day 2 were 10.4 and 13.0, respectively. The distance from the Target Day to Historical Day 1 is less than the distance to Historical Day 2, making Historical Day 1 the Nearest Neighbor.

Changes in Nearest Neighbor Distance with Weighting

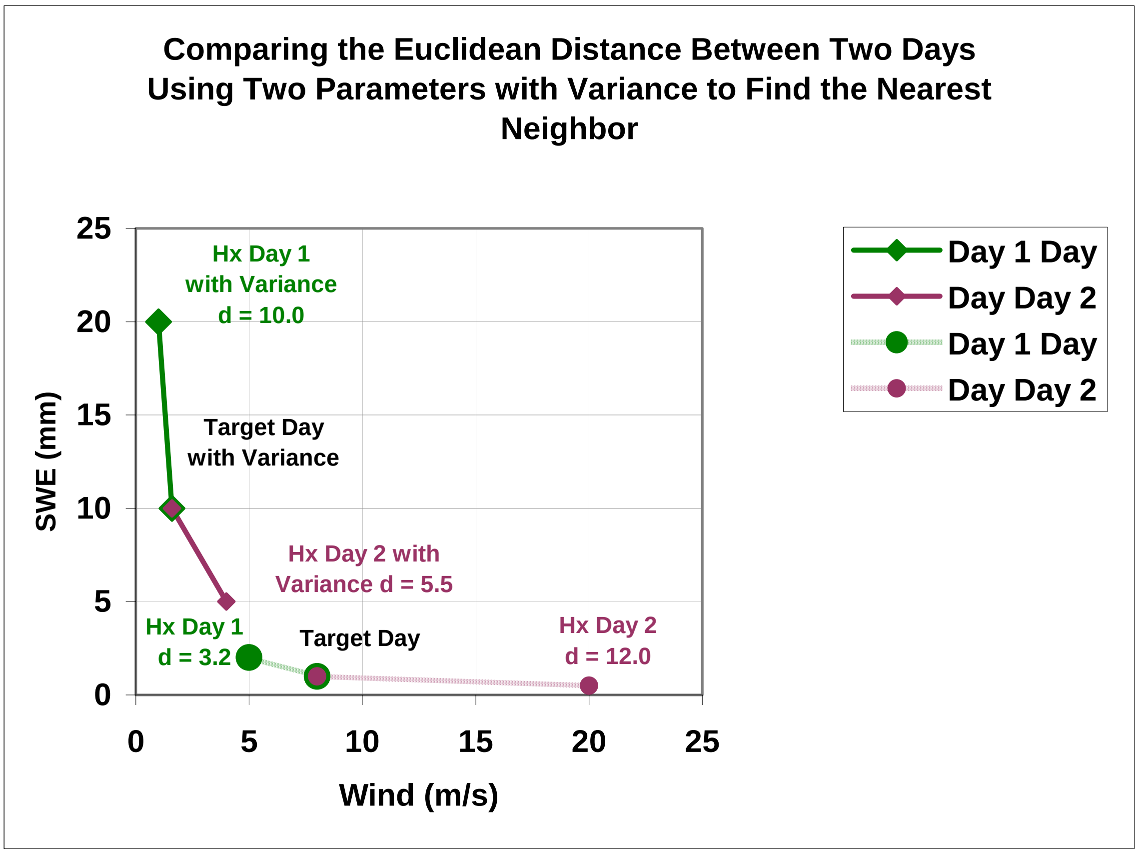
		Non-weighted			Weighted		
	Parameter	Historical	Target	Historical	Historical	Target	Historical
	Weight	Day 1	Day	Day 2	Day 1	Day	Day 2
Wind	1	15	18	30	15	18	30
New Snow	2	20	10	5	40	20	10
Distance		10.4		13.0	20.2		15.6



Often times the forecaster will want one parameter to be weighted more heavily than another. The Target Day has 18 m/s average wind and 10 cm of new snow. Historical Day 1 (Hx Day 1) had 15 m/s average wind and 20 cm of new snow, and Historical Day 2 (Hx Day 2) had 30 m/s average wind and 5 cm of new snow. Prior to weighting Historical Day 1 is the Nearest Neighbor (distance of 10.4 vs. 13.0). If the forecaster was interested in weighting New Snow twice as much as the Wind, then the New Snow parameter would be multiplied by two for all days (Target Day, Historical Day 1, and Historical Day 2). This has the effect of stretching the new snow parameter, which changes the distance measurements. With the new snow parameter being weighted twice that of the wind parameter, Historical Day 2 is now the Nearest Neighbor (distance of 15.6 vs. 20.2).

Changes in Nearest Neighbors Distance when Incorporating Variance

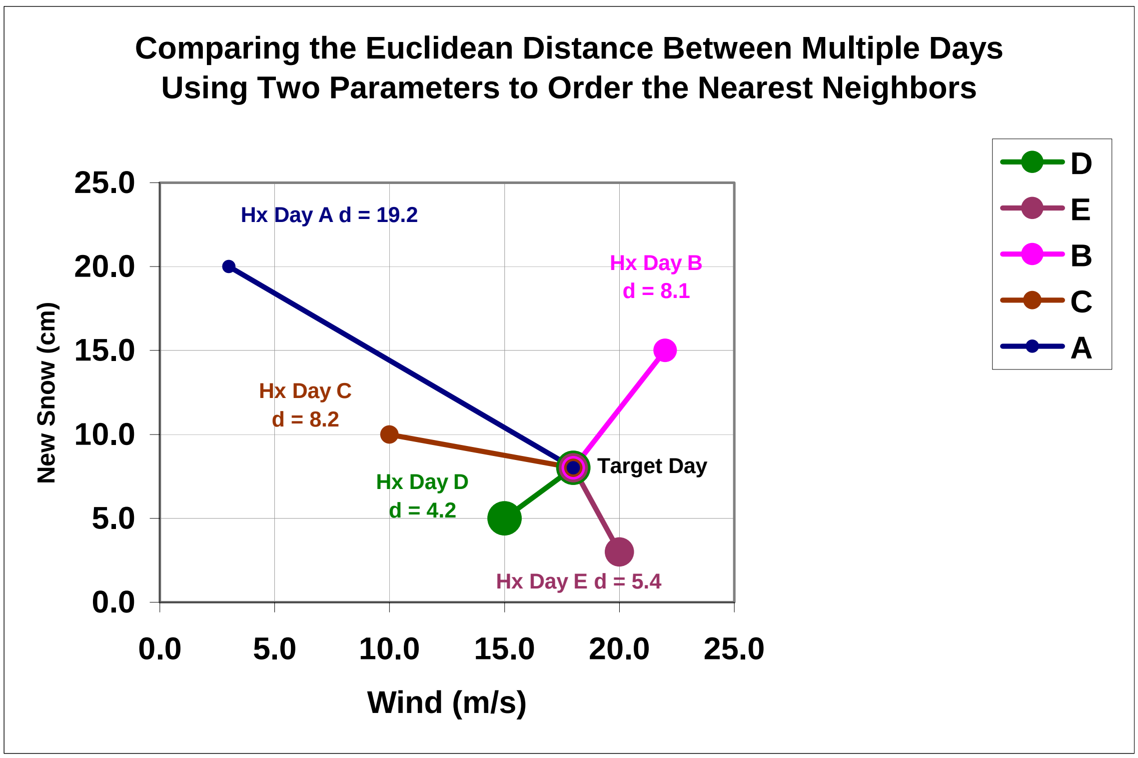
		Variance Not Incorporated			Variance Not Incorporated		
	Parameter	Historical	Target	Historical	Historical	Target	Historical
	Variance	Day 1	Day	Day 2	Day 1	Day	Day 2
Wind	5	5	8	20	1	1.6	4
SWE	0.1	2	1	0.5	20	10	5
Distance		3.2		12.0	10.0		5.5



Parameter variances are used to normalize variables. When the variance is not incorporated, parameters with relatively small numbers, such as snow water equivalent contribute very little to the total distance. In this example, historical day 1 is closer when the variance is not used. When the variance is incorporated, historical day 2 is closer.

Ordering Nearest Neighbors

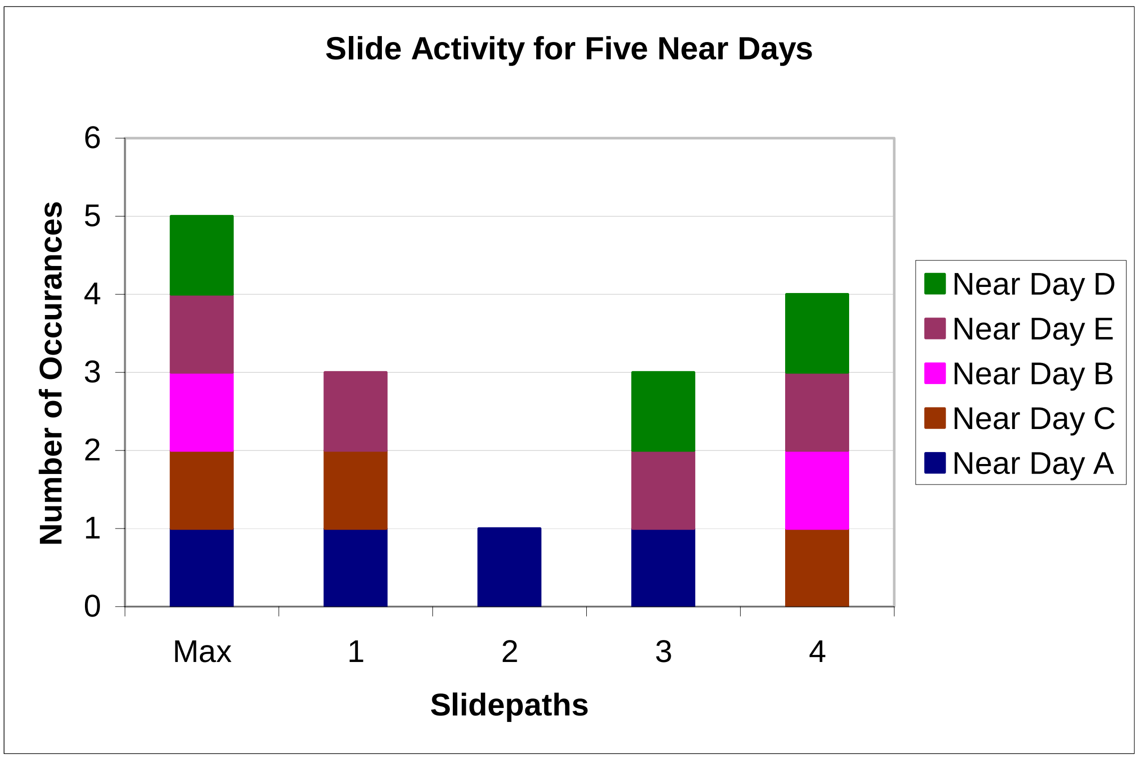
	Historical Day	Target Day Values	Historical Day Values	Parameter Distance	Total Distance
Wind	A	18.0	3.0	15.0	19.2
New Snow	A	8.0	20.0	-12.0	
Wind	B	18.0	22.0	-4.0	8.1
New Snow	B	8.0	15.0	-7.0	
Wind	C	18.0	10.0	8.0	8.2
New Snow	C	8.0	10.0	-2.0	
Wind	D	18.0	15.0	3.0	4.2
New Snow	D	8.0	5.0	3.0	
Wind	E	18.0	20.0	-2.0	5.4
New Snow	E	8.0	3.0	5.0	



Near days are ordered using their nearest neighbor distance. These days and distances will be used in remainder of the nearest neighbor examples.

Combining Near Neighbors with Historical Avalanche Data

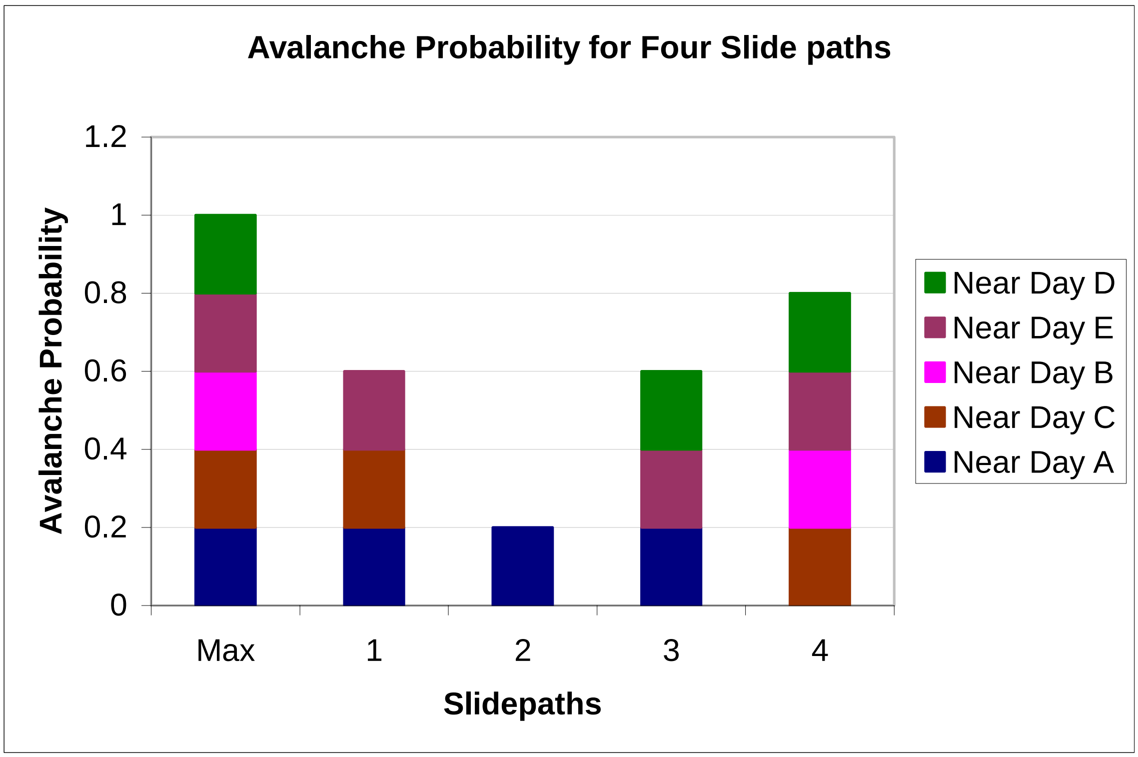
Near Day	Color	Distance	1/Distance	Max	1	2	3	4
Near Day D	Green	4.2	0.2380952	1	0	0	1	1
Near Day E	Plum	5.4	0.1851852	1	1	0	1	1
Near Day B	Pink	8.1	0.1234568	1	0	0	0	1
Near Day C	Brown	8.2	0.1219512	1	1	0	0	1
Near Day A	Dark Blue	19.2	0.0520833	1	1	1	1	0
Cumulative				5	3	1	3	4



For each of the near days, avalanche activity is summed by slide path. The max column is the maximum number of slides possible for a slide path, which is also the number of days, and will be used to create a probability.

Avalanche Probability Not Using Inverse Distance Weighting

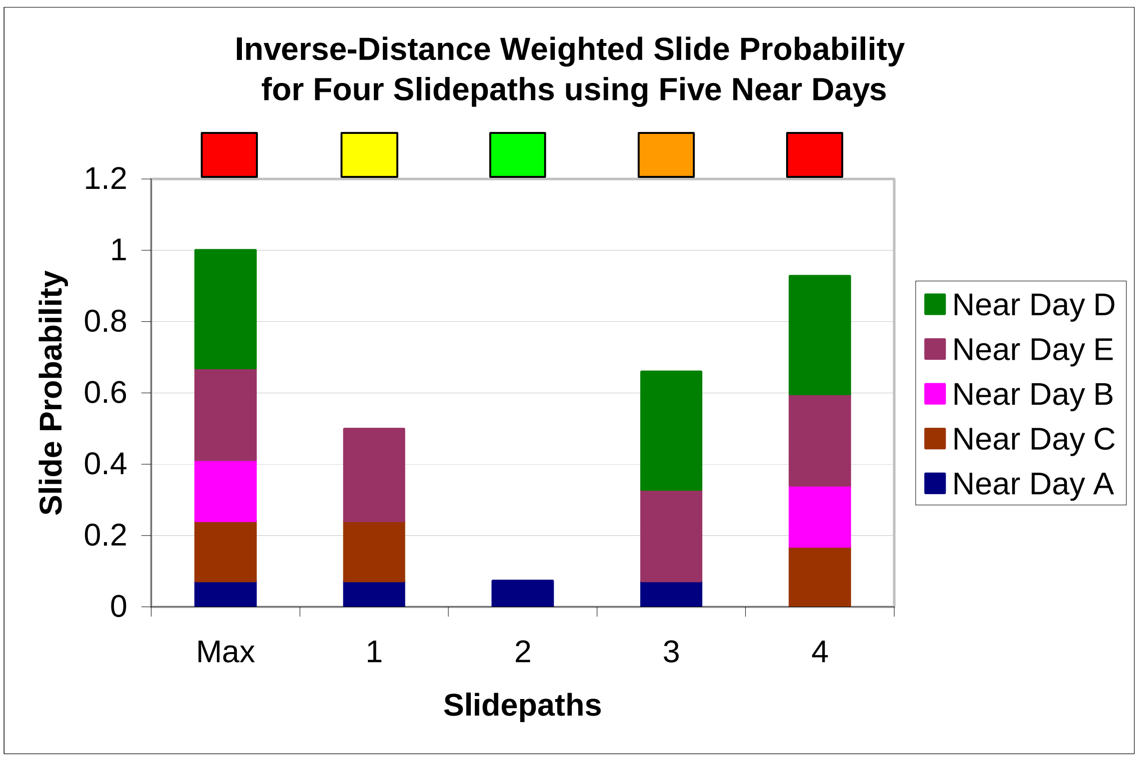
Near Day	Color	Distance	Max	1	2	3	4
Near Day D	Green	4.2	0.2	0	0	0.2	0.2
Near Day E	Plum	5.4	0.2	0.2	0	0.2	0.2
Near Day B	Pink	8.1	0.2	0	0	0	0.2
Near Day C	Brown	8.2	0.2	0.2	0	0	0.2
Near Day A	Dark Blue	19.2	0.2	0.2	0.2	0.2	0
Cumulative			1	0.6	0.2	0.6	0.8



When not using inverse distance weighting, the slide path avalanche probabilities are simply the number of occurrences divided by the maximum possible (an avalanche on all near days).

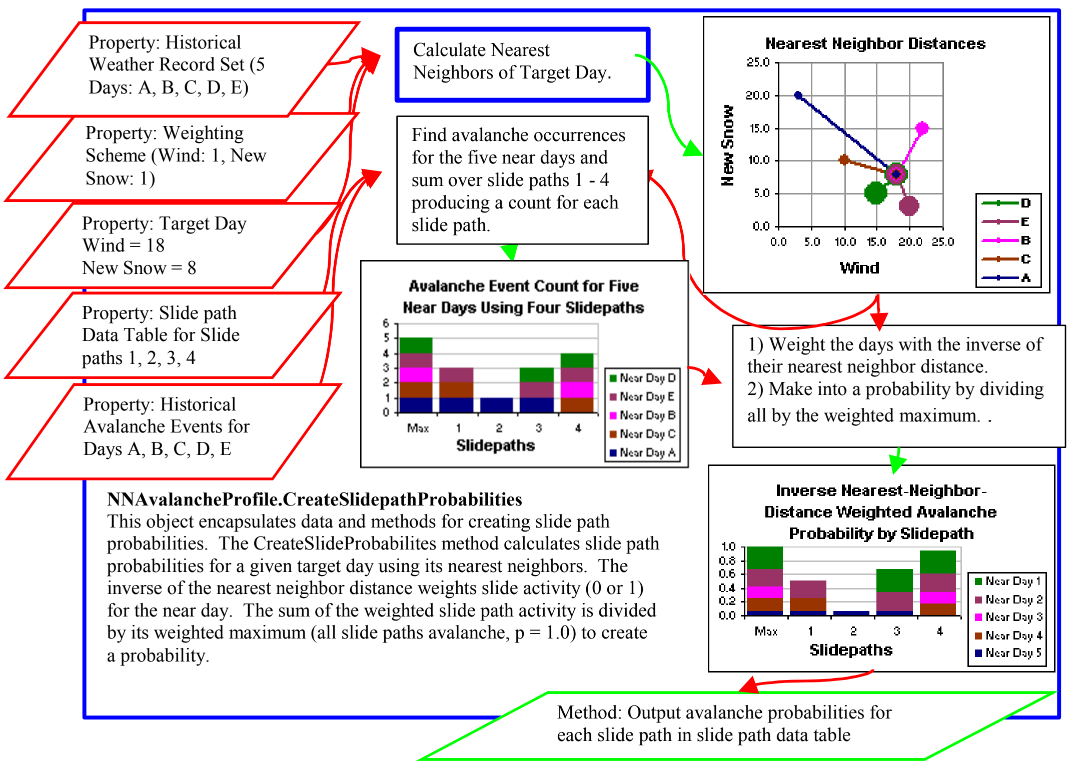
Inverse-Distance Weighted Slide path Avalanche Probability

Near Day	Color	Distance	1/Distance	Max	1	2	3	4
Near Day D	Green	4.2	0.238	0.330	0.000	0.000	0.330	0.330
Near Day E	Plum	5.4	0.185	0.257	0.257	0.000	0.257	0.257
Near Day B	Pink	8.1	0.123	0.171	0.000	0.000	0.000	0.171
Near Day C	Brown	8.2	0.122	0.169	0.169	0.000	0.000	0.169
Near Day A	Dark Blue	19.2	0.052	0.072	0.072	0.072	0.072	0.000
Sum				1.000	0.498	0.072	0.660	0.928



When weighting individual days with the inverses of the nearest neighbor distance, the days that are nearer (more similar to the target day) are counted more heavily. This can be seen by comparing the effect that a given day has on the avalanche probability. For example, both slide paths 1 and 3 avalanched three of five days, but slide path 3 has a higher probability because the avalanche activity occurred on nearer days. These probabilities are then be color-coded to display avalanche hazard which can then be dynamically displayed in the form of a map.

Flow Chart For Creating Inverse-Distance Weighted Avalanche Probabilities



APPENDIX B

PROGRAM PSUEDO CODE

Variable Definitions

td = The *target day vector* defines a set of weather variables (**p**) and holds a specific values for each weather variable.

HxWx = The *historical weather matrix* is composed of **j** days and **p** weather variables.

HxAv = The *historical avalanche matrix* is composed of **i** avalanche events and the set of avalanche attributes (i.e. date, type, size, etc.) for each avalanche event $\mathbf{HxAv_i}$.

σ = The historical weather standard deviation vector contains the standard deviation of $\mathbf{HxWx_p}$ for each weather variable **p**.

i = Avalanche counter and identifier

j = Historical day counter and identifier

p = Weather *variable* vector counter and identifier

a = Represents an *aspect* category.

e = Represents an *elevation* category.

nnd = Nearest neighbor distance vector ($\mathbf{nnd_j}$ is the nearest neighbor distance for $\mathbf{j}^{th}$ day)

ndm = The *near day maximum* is the maximum number of near days to use to calculate slide path avalanche probabilities.

zdv = The *zero distance value* is a minimum value used to weight days that have a **nnd** equal to zero (historical day is identical to target day, or $\mathbf{HxWx_{jp} = td_p}$) to avoid divide by zero errors.

GeoAv = The *geographic avalanche matrix* contains all slide paths with a geographic (GIS) representation along with their attributes (avalanche probability, mean starting zone aspect, and elevation).

pGeoAv = The *probability geographic avalanche vector* contains the probability vector from the **GeoAv** matrix.

avWS = The avala*nche weighted sum vector* holds an inverse nearest neighbor distance weighted sum of avalanche occurrences for each avalanche path **i**.

avWMax = The *avalanche weighted maximum* holds an inverse nearest neighbor distance weighted sum of the maximum possible avalanche occurrences (an avalanche on every near day).

ide = The *inverse distance exponent* allows more similar days (smaller $\mathbf{nnd_j}$) to be weighted more heavily. Values for ide are greater than zero. An ide value of zero counts all near days evenly while an ide value of 1 would result in more similar days being weighted by their inverse $\mathbf{nnd_j}$. The value of ide is not limited to integers.

f = Indicates that **HxWx** has been filtered ($\mathbf{HxWx^f}$).

s = Indicates that **HxWx** has been standardized by its standard deviation ($\mathbf{HxWx^s}$).

o = Indicates that **HxWx** has been ordered by its **nnd** ($\mathbf{HxWx^o}$).

seg = Indicates that **pGeoAv** has been segregated by aspect category **a** and elevation category **e** ($\mathbf{pGeoAv^{seg}}$). $\mathbf{pGeoAv^{seg}_{ae}}$ is the subset of avalanche probabilities that are in aspect category **a** and elevation category **e**.

AERD = The *aspect-elevation rose diagram matrix* contains the mean avalanche probabilities for all slide paths based on aspect category **a** and elevation category **e** ($\mathbf{AERD_{ae}}$).

n = number of variables p

$\mathbf{v_p}$ = variation vector for variable p

w = weighting vector

NNAPP = The *nearest neighbor avalanche probability profile* is a combination of the set of avalanche probabilities (**pGeoAv**), the aspect-elevation rose diagram matrix (**AERD**), and the mean avalanche probability as defined by the target day.

NNAPP = The *nearest neighbor avalanche probability profile **n**-dimensional **matrix*** contains a NNAPP for all variations **v** of variables **p**.

p(v) = The set of variations ($\mathbf{v_{p1}}$-$\mathbf{v_{pn}}$) for variable **p**.

$\mathbf{GeoAv_i SeriesSignature}$ = An n-dimensional matrix holding avalanche probabilities for a single avalanche path **i** ($\mathbf{GeoAv_i}$).

$\mathbf{AERD_{ae} SeriesSignature}$ = An n-dimensional matrix holding avalanche probabilities for a single aspect elevation category with aspect **a** and elevation **e** ($\mathbf{AERD_{ae}}$).

MeanAvSeriesSignature = An n-dimensional matrix holding avalanche probabilities for the mean of all slide path series signatures.

Creating Slide Path Avalanche Probabilities.

*Define the target day (**td**) with a set of weather variables **p** along with specific search values.*

td = ($\mathbf{p_1}$: new snow = 25 cm; $\mathbf{p_2}$: wind speed = 5 m/s; $\mathbf{p_3}$: wind direction = 270°)

*Apply an optional filter **f** to the historical weather matrix **(HxWx)***

$\mathbf{HxWx^f}$ = filter**(HxWx)**

Calculate the standard deviation vector **σ** *for each weather variable **p** (*σ_p*) in* $\boldsymbol{HxWx^f}$

Standardize $\boldsymbol{HxWx^f}$ *by standard deviation **σ** for each weather variable **p** so all variables are represented in a measurement system with **σ** = 1.*

For each day **j** ($\mathbf{HxWx^f_j}$) in $\mathbf{HxWx^f}$

For each weather variable **p** ($\mathbf{HxWx^f_{jp}}$) in $\mathbf{HxWx^f_j}$

$\mathbf{HxWx^{fs}_{jp}} = \mathbf{HxWx^f_{jp}} / \boldsymbol{\sigma_p}$

Next $\mathbf{HxWx^{fs}_{pj}}$

Next $\mathbf{HxWx^f_j}$

Standardize ***td*** *by standard deviation* $\boldsymbol{\sigma}$ *for each weather variable* ***p*** *so all variables are represented in a measurement system with* $\boldsymbol{\sigma}$ *= 1.*

For each weather variable **p** ($\mathbf{td_p}$) in **td**

$$\mathbf{td^s_p = td_p / \sigma_p}$$

Next $\mathbf{td_p}$

Calculate nearest neighbor distance ($\boldsymbol{nnd_j}$*) between the target day and each day in* $\boldsymbol{HxWx^{fn}}$*.*

For each day **j** ($\mathbf{HxWx^{fs}_j}$) in $\mathbf{HxWx^{fs}}$

$$\mathbf{nnd_j}^* = [\textstyle\sum_\mathbf{p} \mathbf{w_p}(\mathbf{td^s_p - HxWx^{fs}_{jp}})^2]^{1/2}$$

*Note: if the variable is a direction measurement, the difference can never exceed |180º| (before standardization). If a difference greater than |180º| was found, then it was subtracted from 360º. This was done prior to standardization.

Next $\mathbf{HxWx^{fs}_j}$

Order (minimum to maximum) $\boldsymbol{HxWx^{fs}}$ *by the day's nearest neighbor distance (*$\boldsymbol{nnd_j}$*).*

$\mathbf{HxWx^{fso}}$ = order($\mathbf{HxWx^{fs}}$)

*Calculate an avalanche probability for each avalanche path **i** ($pGeoAv_i$) in **GeoAv**.*

For each day **j** ($\mathbf{HxWx^{fso}_j}$) in the first **ndm** days of $\mathbf{HxWx^{fso}}$

Find all avalanches ($\mathbf{HxAv_j}$) in **HxAv** that occurred on day $\mathbf{HxWx^{fso}_j}$

For each avalanche **i** ($\mathbf{HxAv_{ij}}$) in $\mathbf{HxAv_j}$ that occurred on day $\mathbf{HxWx^{fso}_j}$

$\mathbf{avWS_i} = \mathbf{avWS_i} + 1 / (\mathbf{nnd_j} + \text{zdv})^{\text{ide}}$

Next $\mathbf{HxAv_{ij}}$

$\text{avWMax} = \text{avWMax} + 1 / (\mathbf{nnd_j} + \text{zdv})^{\text{ide}}$

Next $\mathbf{HxWx^{fso}_j}$

For each avalanche path probability **i** ($\mathbf{pGeoAv_i}$) in **GeoAv**

$\mathbf{pGeoAv_i} = \mathbf{avWS_i} / \text{avWMax}$

Next $\mathbf{pGeoAv_i}$

<u>Calculating Aspect-Elevation Rose Diagram Values</u>

*Segregate avalanche path probabilities by aspect category **a** and elevation category **e**.*

$\mathbf{pGeoAv^{seg}} = \text{segregate}(\mathbf{GeoAv})$

*Calculate a mean probability for each aspect-elevation category for all avalanche paths with aspect category **a** and elevation category **e** (**pGeoAv**$^{seg}_{ae}$).*

For each aspect category **a** (**AERD**$_{a}$) in **AERD**

For each elevation category **e** (**AERD**$_{ae}$) in **AERD**$_{a}$

AERD$_{ae}$ = mean(**pGeoAv**$^{seg}_{ae}$)

Next **AERD**$_{ae}$

Next **AERD**$_{a}$

Calculating Overall Mean

The mean avalanche probability for a set slide path avalanche probabilities is simply their mean.

Mean avalanche probability = (Σ_i**pGeoAv**$_i$)/ i_{total}

Creating Nearest Neighbor Avalanche Probability Profile (NNAPP)

*A single NNAPP is simply the combination of the set of avalanche probabilities (**pGeoAv**), the set of aspect-elevation avalanche probabilities (**AERD**), and the mean avalanche probability as defined by a specific target day (**td**).*

NNAPP = { **pGeoAv**, **AERD**, mean avalanche probability}

Creating Series Signatures

Create the $\boldsymbol{n}$*-dimensional* ***NNAPP*** *matrix using n variables* $\boldsymbol{p}$ *each with* $\boldsymbol{v_p}$ *variations.*

For each weather variable $\mathbf{p}$ ($\mathbf{td_p}$) in $\mathbf{td}$

    For each variation $\mathbf{v}$ ($\mathbf{td_{p(v)}}$) of $\mathbf{td_p}$

        Create $\mathbf{NNAPP_{p(v)}}$ for variation $\mathbf{td_{p(v)}}$

    Next $\mathbf{td_{p(v)}}$

Next $\mathbf{td_p}$

*Create an **n**-dimensional series signature for each slide path, aspect-elevation rose diagram, and the mean avalanche probability.*

For each weather variable **p** ($\mathbf{td_p}$) in **td**

For each variation **v** ($\mathbf{td_{p(v)}}$) of $\mathbf{td_p}$

For each slide path **i** ($\mathbf{GeoAv_i}$) in **GeoAv**

$\mathbf{GeoAv_iSeriesSignature_{p(v)} = NNAPP(GeoAv_i)_{p(v)}}$

Next $GeoAv_i$

For each aspect category **a** ($\mathbf{AERD_a}$) in **AERD**

For each elevation category **e** ($\mathbf{AERD_{ae}}$) in $\mathbf{AERD_a}$

$\mathbf{AERD_{ae}SeriesSignature_{p(v)} = NNAPP(AERD_{ae})_{p(v)}}$

Next $\mathbf{AERD_{ae}}$

Next $\mathbf{AERD_a}$

$\mathbf{MeanAvSeriesSignature_{p(v)} = NNAPP(MeanAv)_{p(v)}}$

Next $\mathbf{td_{p(v)}}$

Next $\mathbf{td_p}$