Estimating the forecasting success of artificially triggering of avalanches with the combination of cluster and discriminant analysis

Reinhard Fromm[1]

[1] Department Natural Hazards and Alpine Timberline, Federal Research and Training Centre for Forest, Natural Hazards and Landscape, Austria

ABSTRACT: Meteorological conditions play an important role for the mechanical stability of the snowpack. Statistical algorithms can be used to determine the likelihood of the occurrence of avalanches which was shown in several previous studies. The common use of two statistical methods offers the possibility to analyze the data set in regard to specific weather conditions. In the first step the k-mean cluster analysis selects days with similar weather conditions; so that each day is assigned to a predefined group. Significant weather conditions are used for the definition of the initial conditions of each cluster. Each group represents a typical weather situation. In the second step the discriminant analysis is used to separate between avalanche and non-avalanche days for each group. Consequently the coefficients of the discriminant functions can differ and the best fit of the discriminant function depending on the predominant weather situation is applied.
Results of the clustering algorithm and the hit rate of the subsequent discriminant analysis are shown.

KEYWORDS: avalanche forecast, cluster analysis, discriminant analysis.

1 INTRODUCTION

Obled and Good (1980) classified three different approaches of statistical avalanche forecasting. The first method (*model I*) consists of a simple discriminant analysis applied to a sample of avalanche days against a sample of non-avalanche days. The second approach (*model II*) tries to take into account different types of avalanche phenomena associated with different types of snow and weather conditions. It requires two-stage decision criteria. The third method (*model III*) aims to find the nearest neighbours of recorded weather conditions and avalanche observations.

Numerous algorithms similar to the type *model I* have been developed using different statistical methods like fuzzy-logic (Terada et al., 1991; Kleemayr et al., 2000), discriminant analysis (Obled and Good, 1980), regression and classification trees (Davis et al., 1999; Howley, 2007). The subdivision for *model II* may be done using the avalanche classification based on de Quervin (1973), applying a non-hierarchical clustering method (Obled and Good, 1980) or a multiple discriminant analysis (Wilks, 1995). The nearest neighbours model (*model III*) is based on hierarchical clustering and has often been applied to avalanche forecasting issues (Obled and Good, 1980; Buser, 1983; Brabec and Meister, 2001; Gassner and Brabec, 2002; Zeidler and Jamieson, 2004).

The approach proposed in this study can be assigned to the type *model II*. The clustering with the k-means algorithm (non-hierarchical) groups days with similar meteorological and snowpack conditions. Using the same data set the discriminant function decides between avalanche or non-avalanche day.

It is proposed that the discriminant analysis can separate avalanche days and non-avalanche days under similar weather conditions.

2 METHODS

2.1 Weather and snow data

The study was carried out in an avalanche controlled area of the ski resort Lech (Vorarlberg, Austria). The release areas extend between 1680m and 2540m a.s.l. Data were collected from 2003 to 2007.

An automatic weather station located in the study area recorded air temperature, snow depth, wind speed and direction. Daily observations of the snowpack and data collected with the Swiss *Rammsonde* are available.

The variables describing the meteorological conditions and the properties of the snowpack were chosen similar to Singh et al. (2005). For the statistical analysis the air temperature, the snow temperature (0.1m below the snow surface), the snow depth, the daily snowfall accumulation, the difference of the maximum air temperate (previous day) and the minimum air temperature (during the last 24h) were used.

2.2 Cluster analysis

The k-means cluster analysis requires a predefined number of k clusters. It generates k different clusters of the greatest possible distinction.
The original k-means cluster analysis algorithm starts its iteration with randomly selected data sets for the initialization. This procedure leads to different solutions; especially conditions which are rarely observed can be underrepresented.
In order to overcome these restrictions the initial conditions are set manually. Specific data sets of the weather data are used to select the initial values for the k-means algorithm. The conditions are: minimum air temperature, maximum air temperature, maximum snow depth change, maximum snow depth, minimum snow temperature, maximum snow temperature, mean snow depth, minimum snow depth, mean air temperature, minimum penetration depth of the *Rammsonde*, maximum penetration depth of the *Rammsonde*, maximum daily air temperature change (table 1).
Each of these data sets represents a day with significant weather conditions. The iterative process of the k-means cluster analysis starts with the assignment of each data set to the cluster under the condition that the Euclidian distance to the centroid of a cluster in the k-dimensional phase space is a minimum. The centroids are recomputed and the iteration is repeated until no changes in any cluster occur.
Finally each day is a member of a cluster and each cluster represents a significant weather or snowpack situation.
The cluster analysis does not aim to differ between avalanche and non-avalanche days. It is supposed that within each cluster the records contain information which forces processes of an avalanche release, but they do not compensate.
For example, rising air temperature forces the settlement of a cold snowpack decreasing the long-term probability of an avalanche release. But an isothermal snowpack loses mechanical stability caused by melting which may result from high air temperatures. Such cases should not be in the same cluster, so that the discriminant analysis is able to separate between avalanche and non-avalanche days effectively.

2.3 Discriminant analysis

Within each cluster the linear discriminant analysis has been applied. The existing data pool has been divided by the discriminant function.
The sign of the distance D from a actual data set to the discriminant function separates avalanche and non-avalanche days.

$$D \begin{cases} \geq 0 & \text{avalanche day} \\ < 0 & \text{non-avalanche day} \end{cases}$$

The coefficients generated by the discriminant analysis refer to each individual cluster.

2.4 Combining procedure

On a new forecasting day new weather and snowpack observations are available and the Euclidian distances to all data sets are calculated. The data set with the minimal distance is the closest. Concurrently it is member of a group of the cluster analysis.
The coefficients of the discriminant function of that specific group are used to calculate the distance D from the actual data set. Both, the sign and the absolute value of D contain information about the likelihood of the occurrence of an avalanche.
The information of the sign of D is binary; the absolute value of D represents the distance to the discriminant function which separates the data set. Unfortunately the distance D in the k dimensional phase space can assume any value which inhibits the interpretation.
Results generated with different discriminant functions from other groups of the cluster analysis are not comparable. Therefore a scaling is applied. The scaling of D is accomplished by a fit with equation (1) which ensures that the scaled avalanche danger S is between $0 \leq S \leq 1$.

$$S = \frac{1}{2}\tanh\left(\frac{D-\mu}{\sigma}\right) + \frac{1}{2} \quad (1)$$

where μ is the median of all D and σ is parameterized, so that the point ($D = 0$, $S = 0.75$) is reached. High avalanche danger $S \geq 0.75$ corresponds to positive distances D (avalanche days) and $S < 0.75$ indicate negative distances D (non-avalanche days). The parameterization generates results which makes the calculations comparable among each other.
The procedure can be applied for both, the regional and the local scale. Avalanches were always artificially triggered at the same positions (release areas). The attempts and the success of an avalanche release had been recorded. The weather conditions at these locations are estimated from the data of the automatic weather station. The air temperatures were calculated using the lapse rate and the altitude. The new snow depth depends on the intensity of snow drift which differs at each release area. A snow drift index was determined at selected release areas. It is based on the estimation of

Table 1: Groups generated by the cluster analysis (details see text).

Cluster		number of		frequency of the avalanche danger				
ID	name (initial conditions)	members	avalanche days	1	2	3	4	5
1	minimum air temperature	0	0	0	0	0	0	0
2	maximum air temperature	21	2	7	12	2	0	0
3	maximum new snow event	25	1	5	10	10	0	0
4	maximum snow depth	5	5	0	0	2	3	0
5	minimum snow temperature	18	10	0	1	14	3	0
6	maximum snow temperature	42	4	7	28	7	0	0
7	mean snow depth	25	1	5	19	1	0	0
8	minimum snow depth	7	0	5	1	1	0	0
9	mean air temperature	19	0	10	8	1	0	0
10	maximum *Rammsonde* penetration	28	13	0	15	10	3	0
11	minimum *Rammsonde* penetration	8	7	0	0	3	5	0
12	maximum daily temperature change	37	5	0	28	8	1	0

experts and the wind speed. The index correlates to the exposition of the slope ($r^2 = 0.71$), so that the snow drift index for all other release areas was estimated by that relationship.

3 RESULTS

The cluster analysis accomplished by the k-means algorithm generated k different clusters or groups. The initial conditions were pre-defined using selected weather conditions. The results from the cluster analysis are summarized in table 1. First, the number of avalanche days and the number of members of each cluster gives an impression on the probability for an avalanche day. The frequency of the avalanche danger (according to the European avalanche danger scale), provided by the avalanche warning center, within each group, is the second information.

The first cluster (ID=1) in table 1 has no members. Both, the minimum air temperature and the minimum snow temperature have been used for the initialization. The two variables are very similar, so that one of these two groups became empty by the clustering algorithm. In contrast both groups 2 and 6, which represent high air and snow temperatures, exist, because the snow temperature can only reach 0°C.

The 5 members of cluster 4 correspond to large snow depths. Avalanches occurred on all of the days and the avalanche danger was 3 or 4.

The groups which had low snow depths (8) or average air temperatures (9) had no avalanche days and the avalanche danger was less or equal 3.

Both groups with high and low penetration depth of the *Rammsonde* (10, 11) include many avalanche days with moderate to high avalanche danger. High penetration depths occurred after large snowfall events where the artificial release of avalanches is easier. A low penetration depth indicates a hard layer at the snow surface which may form a bed surface for slab avalanches (e.g. drifted snow during the day).

The largest group (6, high snow temperature) contains 42 members and 4 avalanche days. The avalanche danger was less or equal 3. It seems that the daily temperature change (12) consists of all, avalanche days with high and low avalanche danger and days without avalanches. High daily temperature change occurs mostly in spring and it has no effect to the avalanche danger on a well settled snowpack.

The discriminant analysis which depicts between avalanche or non-avalanche days was carried out for each group of the cluster analysis. The overall success of the estimated avalanche days is shown by pie charts at each release area (Fig. 1). The hit rate is defined as the ratio of the number of correct estimated avalanche days to the number of avalanche days. A black solid circle represents a hit rate of 100% and a white solid circle corresponds to a total failure. The number of observed avalanche days is plotted right to each pie chart.

In the majority of the cases the discriminant analysis is able to separate between avalanche and non-avalanche days.

The previous separation of the data set by the cluster analysis reduced the sample size of each sub data set which was used for the discriminant analysis. For release areas with more than 20 avalanches the hit rate is between 70% and 85%. If less avalanches were observed the hit rate seems to be rather random. The hit rate does not vary with the exposition of the release areas.

It seems that the proposed snow drift index provides acceptable results, but it should be noted that the micro relief was neglected. The

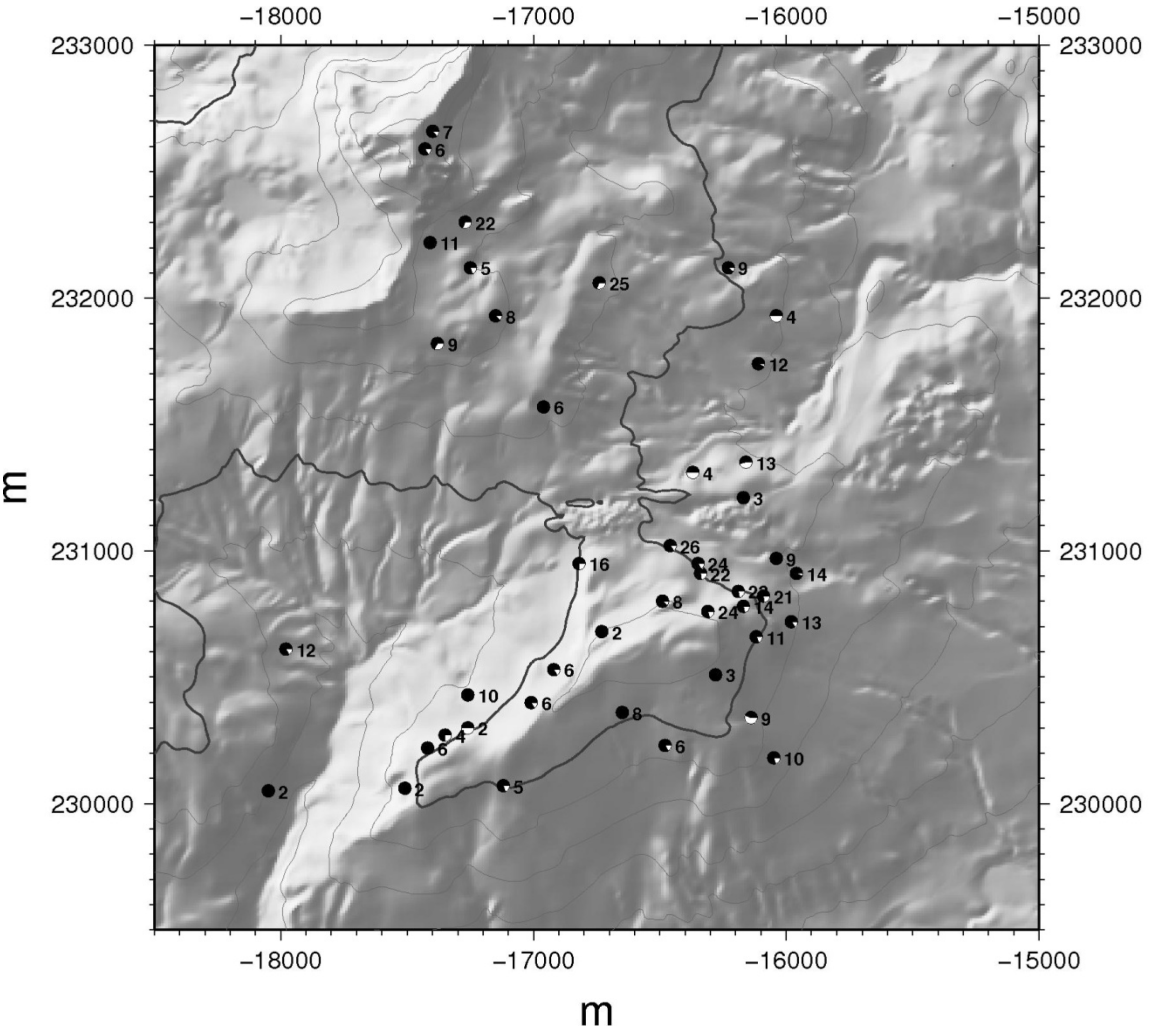

Figure 1. The filled (black) pie charts show the success of calculated avalanches. The number of avalanche days for each release area (right).

solar heating of south facing slopes may play an important role for cluster 2 and 12.
The artificial triggering of avalanches with remote-controlled devices (e.g. Gazex) handicaps the rigorous documentation of the success of avalanche releases.

4 CONCLUSIONS

The basic idea that weather conditions are responsible for the avalanche danger can be confirmed and if numerous observations are available, statistical methods are able to estimate the avalanche danger without a detailed knowledge of complex physical processes which proceed simultaneously. The combination of the two algorithms, cluster analysis and discriminant analysis, permits more detailed statements on the avalanche danger.
Further investigations may optimize the selection of the initial conditions and the number of clusters.
It applies to all statistical methods for the estimation of the avalanche danger or the probability of the occurrence of avalanches that a larger amount of data with a sufficient quality improve the hit rates.

ACKNOWLEDGEMENTS

The project was founded by the Austrian BMVIT (Bundesministerium für Verkehr, Inovation und Technologie) and the ADS GbR Lech. The approach presented in this manuscript has been used in a software package which is in operational use at Lech (Vorarlberg, Austria).
Thanks to Roman Sandner and Paul Dobesberger for the extensive field work and the formulation of the snow drift index.

REFERENCES

Brabec, B. and Meister, R. 2001. A nearestneighbor model for regional avalanche forecasting. Annals of Glaciology, 32: 130–134.

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

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

de Quervin, M. 1973. Eine Internationale Lawinenklassifikation. Zeitschrift f. Gletscherkunde und Glaziologie, IX(1-2): 189–206.

Gassner, M. and Brabec, B. 2002. Nearest neighbour models for local and regional avalanche forecasting. Natural Hazards and Earth System Sciences, 2: 247–253.

Howley, A. 2007. Snow Avalanche Prediction in Grasdalen, Norway. Master's thesis, Faculty of Mathematics and Natural Science, University of Oslo, Norway.

Kleemayr, K., Moser, A., Schöll, J., and Vollsinger, S. 2000. Lawinenprognose mit statistischen und selbstlernenden Verfahren im Projekt NAFT. Schutz des Lebensraumes vor Hochwasser, Muren und Lawinen, Proc. Intern. Symp. Interpraevent 2000, Villach, Austria, pages 411–422.

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

Singh, A., Srinivasan, K., and Ganju, A. 2005. Avalanche forecast using numerical weather prediction in Indian Himalaya. Cold Regions Science and Technology, 43:83–92.

Terada, H., Nakamura, Y., and Chiba, A. 1991. Method for evaluating snow avalanche occurrence on basis of fuzzy theory. Proceedings of the Japan-U.S. Workshop on snow avalanche, landslide, debris flow prediction and control, 95–104.

Wilks, D. 1995. Statistical Methods in the Atmospheric Siences, Volume 59 of International Geophysics Series. ACADEMIC PRESS.

Zeidler, A. and Jamieson, B. 2004. A nearestneighbour model for forecasting skier-triggered dry-slab avalanches on persistent weak layers in the Columbia Mountains, Canada. Annals of Glaciology, 38: 166–172.