The differentiation of thaws in connection with slushflow occurrences

Pavel Chernouss, Olga Tyapkina
The Centre of Avalanche Safety of «Apatit» JSC, 33 a, 50 Years of October St., Kirovsk, Murmansk region 184230, Russia

Erik Hestnes, Steinar Bakkehøi
Norwegian Geotechnical Institute, P.O. Box 3930 Ullevaal Hageby, N-0806 Oslo, Norway

ABSTRACT. Classification of meteorological conditions accompanying slushflow occurrences is attempted. Days with mean daily air temperature above 0°C have been chosen for the analysis. Each day has been described with a set of meteorological parameters. Multivariate statistical methods have been used to derive rules for discrimination of "slushflow" and "non-slushflow" situations based on these parameters. The linear discriminant analysis and Bayes' rule have been applied to data for the Khibini Mountains in Russia and mountainous regions of Norway. The results obtained show that snow depth and total water supply have a statistically sufficient difference between the two situations to enable them to be used for relatively reliable recognition of such situations. Classification accuracy is better than 75% for both types of situation. Transformation of the classification method into a forecast method is discussed.

INTRODUCTION

The use of statistical methods for avalanche forecasting is very common. In most cases methods of statistical classification are applied, but they have never been used for slushflow forecasting. The limited data available, due to modest international research on slushflows, is the main obstacle for establishing such methods. There are, however, more or less complete data on slushflow occurrences and accompanying meteorological conditions for the Khibini Mountains in the Kola peninsula, North West Russia and from the fjord and valley districts of Norway. Although the amount of data from each region is not enough for statistical methods, it is alluring to combine data from the two regions to obtain a larger statistical sample.

The main problem is to identify parameters which describe the weather conditions in such a way that it is possible to consider "slushflow situations" and "non-slushflow situations" in different regions as samples from a common parent population. Characteristic parameters for such descriptions were chosen based on previous studies by Hestnes and others (1994). Standard meteorological information was used to describe the situations. The methods were established for the Khibini Mountains and tested on data from West and North Norway.

METHODS

The linear discriminant analysis and direct application of Bayes' rule were chosen for classification of situations with mean daily air temperature above 0°C into "slushflow" and "non-slushflow" situations.

Discriminant analysis

There are some models for avalanche forecasting based on the discriminant analysis technique (Bois, Oblead and Good, 1975; Oblead and Good, 1980; McClung and Tweedy, 1994). The same approach used for classification of situations into "avalanche" and "non-avalanche" by Chernouss (1975) is applied in this work. Each situation is described by a set of parameters - $x_1,...,x_n$ or vector $\mathbf{x}$ - belonging to one of two classes: S - "slushflow" or N - "non-slushflow". It is assumed that parameters in the two classes are normally distributed, have the same variance-covariance matrices - Σ, and different vectors of expectations - $\mu^{(S)}$ and $\mu^{(N)}$.

Discriminant analysis seeks a linear function of the parameters:

$$\ln\frac{P_S(\mathbf{x})}{P_N(\mathbf{x})} = \mathbf{x}'\,\Sigma^{-1}(\mu^{(S)}-\mu^{(N)}) - \frac{1}{2}(\mu^{(S)}+\mu^{(N)})'\,\Sigma^{-1}(\mu^{(S)}-\mu^{(N)}) \quad (1)$$

Where: $P_S(\mathbf{x})$ and $P_N(\mathbf{x})$ are density probabilities for S and N classes respectively.

According to Anderson (1958) the domains of the best classification are determined as:

$$\mathbf{x}\in S \text{ if } \mathbf{x}'\Sigma^{-1}(\mu^{(S)}-\mu^{(N)})-\frac{1}{2}(\mu^{(S)}+\mu^{(N)})'\Sigma^{-1}(\mu^{(S)}-\mu^{(N)})\geq \ln K, \quad (2)$$

$$x \in N \text{ if } x'\Sigma^{-1}(\mu^{(S)}-\mu^{(N)}) - \frac{1}{2}(\mu^{(S)}+\mu^{(N)})'\Sigma^{-1}(\mu^{(S)}-\mu^{(N)}) < \ln K, \quad (3)$$

$$K = \frac{P(N)C(S/N)}{P(S)C(N/S)} \quad (4)$$

Where: P(S) and P(N) are a priori probabilities of "slushflow" and "non-slushflow" situations; and: C(S/N) and C(N/S) are values of losses due to misclassification of situations from N and S classes respectively. "Best classification" means that which gives minimal average losses.

The inequalities (2) and (3) are linear functions of the parameters describing situations

$$\lambda_1 x_1 + \ldots + \lambda_n x_n + b \geq 0 \quad (5)$$

$$\lambda_1 x_1 + \ldots + \lambda_n x_n + b < 0 \quad (6)$$

So, if the evaluations $\mu^{(S)}$, $\mu^{(N)}$, Σ are made and coefficients $\lambda_1 \ldots \lambda_n$ and b obtained on the basis of available samples, each new situation **x** should be related to class S if $\lambda_1 x_1 + \ldots + \lambda_n x_n + b \geq 0$ or to class N if $\lambda_1 x_1 + \ldots + \lambda_n x_n + b < 0$.

Discriminant analysis can easily be carried out using different statistical software, for example with Statgraphics, SPSS, Stadia etc.

Bayes' rule

The method was applied similarly as for diagnostics of avalanche situations (Zuzin, 1989).

Bayes' rule provides an opportunity for reevaluation of probabilities of hypotheses, which form a complete group of incompatible hypotheses, if certain events x_k have occurred. Events x_k can be the result of measurement or calculation. Below this rule is applied for the pair of hypotheses: S - "slushflow situation" and N - "non-slushflow situation" and can be written as:

$$P(S/x_1 \ldots x_n) = \frac{P(S)P(x_1,\ldots,x_n/S)}{P(S)P(x_1,\ldots,x_n/S) + P(N)P(x_1,\ldots,x_n/N)} \quad (7)$$

where: $P(S/x_1,\ldots,x_n)$ is the conditional probability of "slushflow situation" if events $x_1,\ldots,x_n$ have occurred; and: P(S) and P(N) are the a priori probabilities of "slushflow situation" and "non-slushflow situation", respectively; $P(x_1,\ldots,x_n/S)$ and $P(x_1,\ldots,x_n/N)$ are the conditional probabilities of the $x_1,\ldots,x_n$ - event for "slushflow situation" and "non-slushflow situation", respectively.

Substituting P(N) = 1-P(S) and dividing both numerator and denominator into $P(S)P(x_1,\ldots,x_n/S)$ gives:

$$P(S/x_1 \ldots x_n) = \frac{1}{1 + \left(\frac{1}{P(S)} - 1\right)\frac{P(x_1,\ldots,x_n/N)}{P(x_1,\ldots,x_n/S)}} \quad (8)$$

If the events x_k are the results of measurements or calculations of independent continuous parameters, Bayes' rule can be written as:

$$P(S/x_1,\ldots,x_n) = \frac{1}{1 + \left(\frac{1}{P(S)} - 1\right)\prod_{k=1}^{n}\frac{\rho_N(x_k)}{\rho_S(x_k)}} \quad (9)$$

where: $\rho_S(x_k)$ and $\rho_N(x_k)$ are the conditional probability densities of the x_k - parameter for "slushflow" and "non-slushflow" situations, respectively. The values of $\rho_S(x_k)$ and $\rho_N(x_k)$ can easily be estimated with previous observations. The application of the formula (9) does not require a normality of the parameters or any other limits for their distribution laws within classes, but parameters should be statistically independent.

According to theory the quality of the discrimination of the situations should be the same both for the linear discriminant function and the formula (9) if the parameters are statistically independent, normally distributed and have the same variances in classes. The question of which method is the best one for practice can be answered after detailed verification with independent data.

DATA

Since slushflows occur when air temperature is above 0°C, only meteorological information for periods of thaws and corresponding data on slushflow releases were analyzed. The parameters were chosen based on previous studies by Hestnes and others (1994).

Khibinian data

Regular observations of slushflows in the Khibini Mountains have been carried out by the Centre of Avalanche Safety of "Apatit" JSC since the mid fifties. The observations cover almost twenty sites where slushflows occur. All sites are within a relatively small area, a radius of less than 15 km. The heights of the slushflow starting points vary from 300 to 800 m a.s.l.

All the registered slushflows have occurred during spring break-up between 28th of April and 10th of June. Thirty-three slushflow events with a release time determined to an accuracy less than one day, were chosen amongst the total number of registered cases. Standard meteorological information was supported by the meteorological stations located in the valleys and on the mountain tops in the area, within the height range from 250 to 1090 m a.s.l.

To avoid data heterogeneity connected with elevation differences of the slushflow starting zones and the meteorological stations, all data from the meteorological stations were interpolated to the different 100 m step levels between 300 and 800 m a.s.l. The situations for each level were divided into two classes. The class of "slushflow situation" consists of situations where at least one slushflow was released at the considered level within the area. All other situations belong to the class of "non-slushflow situation". The classified situations of the different levels were then combined into two new classes which covered the situations at all levels.

The basic idea of such combining is that the same conditions should lead to the same results. Since slushflows in the Khibini Mountains only occur during spring time, only thaws between the 15th of April and 15th of June were taken into account. Thaws after slushflow releases were excluded.

Mean daily values of air temperature, wind speed, air humidity, cloudiness; daily precipitation and snow height were chosen as primary parameters. Functions of the above mentioned parameters were used as secondary ones: daily sums of melted water, calculated with an energy balance model outlined by Harstveit (1984); daily sums of water income due to snowmelt and precipitation; total water income since the thaw started and some others.

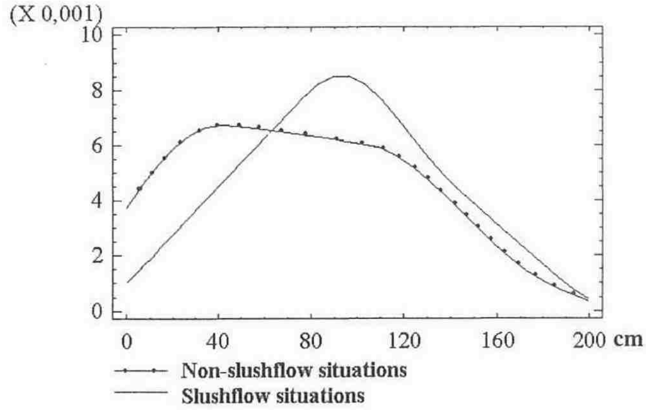

Fig. 1 Smoothed empirical probability densities of snow height for the Khibini Mountains.

Norwegian data

There is no regular registration of slushflow releases in Norway. However, slushflows are reported if they cause damage or some sort of trouble for the population. The most complete list of slushflow events is at the Norwegian Geotechnical Institute.

Meteorological stations close to slushflow starting zones, enabling interpolation of meteorological data at different height levels, are practically impossible to find. As a rule the stations are located at the foot of mountains. Due to this fact, only slushflows released less than 200 m above the nearest meteorological station were chosen for the analysis.

Corresponding meteorological data as for the Khibini Mountains were used, without any height corrections. For this reason and some others, only 25 "slushflow cases" were chosen for further treatment, in spite of a large total number of slushflow events registered in Norway.

The common samples of the two classes were formed the same way as for the Khibini Mountains with one exception. Winters without slushflows were not considered because it was impossible to say definitely that there were no slushflow events these years. Small slushflows, causing no trouble for population, could have occurred. Thus, only the thaws before slushflow release in each winter were taken into consideration as "non-slushflow" ones.

RESULTS

First of all, different sets of parameters describing the Khibinian situations were tested for discrimination of "slushflow" and "non-slushflow" days. The quality of the discrimination with discriminant function was not acceptable if only integral parameters characterizing the whole thaw from its beginning to the day of diagnostics, were used. For example, the average value of such parameters as air temperature, wind speed, precipitation, air humidity, cloudiness and snow height for the period.

The same picture was obtained when trying to discriminate "slushflow" and "non-slushflow" days based on these primary parameters, except for snow height. The difference between snow height probability distributions for the two classes was significant (Fig. 1). It was tested with the Kolmogorov - Smirnov test at 95 % confidence level. Daily sum of water income was found the most informative parameter amongst the secondary parameters for "slushflow" and "non-slushflow" days discrimination. Differences in the probability distributions are evident (Fig. 2).

The graphs of the smoothed empirical probability density for snow height and daily water income were used

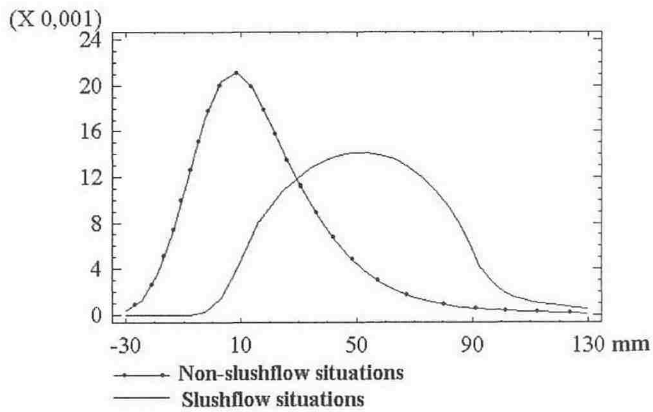

Fig. 2 Smoothed empirical probability densities of daily water income for the Khibini Mountains.

for calculation of values $\rho_N(x_k)/\rho_S(x_k)$ in formula (9). The insignificant influence of the integral parameters tells that the process of slushflow formation is relatively short and taking into account a longer period just decreases the informativeness of the parameters. The discriminant function for "slushflow" and "non-slushflow" days for the Khibini mountains is

$$F(h,i)=0.0221*h + 0.0962*i - 5.11 \qquad (10)$$

Where: i - represents the daily water income (mm) and h - the snow height (cm). The a priori probabilities of the two classes and the values of losses are assumed as equal. If F(x) is equal or more than zero, a day which is tested is recognized as "slushflow" otherwise as "non-slushflow". The results of the classifications are shown in Table 1:

Table 1. Classification matrix for discrimination of "slushflow" and "non-slushflow" days in the Khibini Mountains. 85.4% of all days successfully classified; 78.8% of "slushflow days" successfully classified; 85.5% of "non-slushflow days" successfully classified.

Observed

Predicted	S_o	N_o	Σ
S_p	26	403	429
N_p	7	2374	2381
Σ	33	2777	2810

Table 2. Classification matrix for discrimination of "slushflow" and "non-slushflow" days in mountains of Norway. 87.3% of all days successfully classified; 76.0% of "slushflow days" successfully classified; 88.4% of "non-slushflow days" successfully classified.

Observed

Predicted	S_o	N_o	Σ
S_p	19	29	48
N_p	6	222	228
Σ	25	251	276

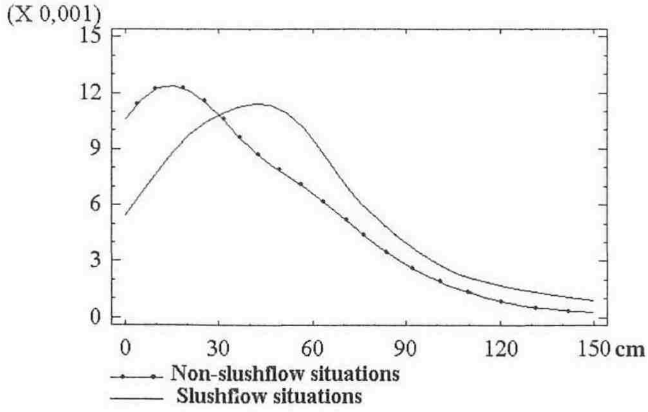

Fig. 3 Smoothed empirical probability densities of snow height for mountains of Norway.

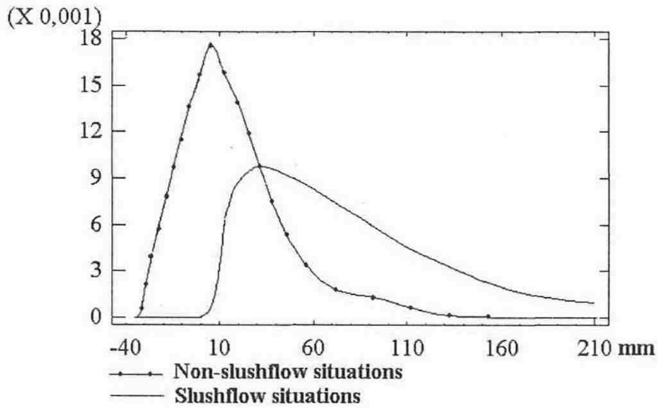

Fig. 4 Smoothed empirical probability densities of daily water income for mountains of Norway.

Since parameters x_k in formula (9) should be statistically independent, the significance of linear correlation between snow height and daily water income was tested. The test showed that there is a statistically significant relationship between them at 99 % confidence level, but the relationship is very weak. Correlation coefficient equals - 0.35.

Figure 5 shows the results of discrimination of the same days with formula (9) depending on the threshold probability used for categorical classification. If the calculated probability is equal or more than the threshold value the situation should be classified as "slushflow" otherwise as "non-slushflow".

A threshold value should be defined by the customer of the classification and related to a defined acceptable risk. For example if the threshold probability is 0.4, which means that if the probability calculated is equal to or more than 0.4 the situation is recognized as "slushflow" if less than 0.4 as "non-slushflow". The accuracy of the categorical classification of both types of situation is close to 80%. If the threshold probability is taken as 0.25 then 90% of "slushflow days" will be recognized correctly, but for "non-slushflow days" this value drops to 65-70%.

Application of the classification functions obtained for the Khibini Mountains to the Norwegian data gave almost the same results (see Figure 5 and Table 2). This is quite natural because the density distribution parameters used are very similar for these two regions. For physical reasons, it can be supposed that other regions will have similar results.

FORECAST SCHEME

Both methods can be easily transformed into forecast methods by using prognostic meteorological values instead of current data. If a one day moving average is used for input data it is possible to get values of discrimination function or release probability, with a resolution better than one day. So, besides the methods being used as pure formal procedures for slushflow forecasting, they could also be used as a powerful tool by forecasters, giving them the opportunity to observe the dynamics of the process and reevaluate the subjective estimations of the probabilities of slushflow releases. This is obviously valuable in data interpretation when the process is controlled by many factors, which can influence in opposite directions.

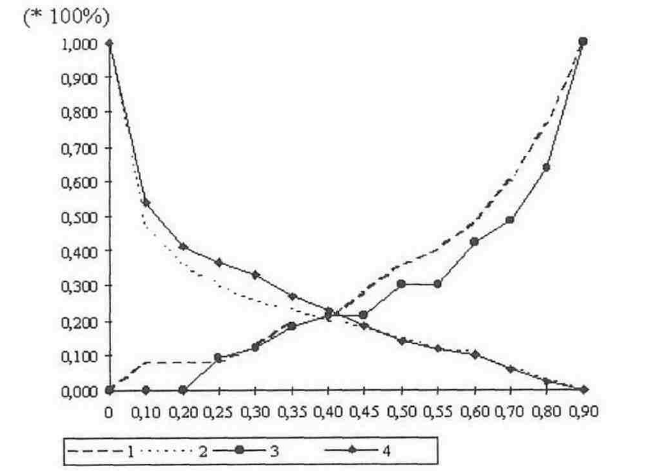

Fig. 5 Empirical values of classification errors depending on threshold probability. 1.3 - percentage of misclassified "slushflow days" for the Khibini Mountains and mountains of Norway respectively, 2.4 - percentage of misclassified "non-slushflow" days for the Khibini Mountains and mountains of Norway, respectively.

The forecasting models should be used within relatively small regions. We would suggest a maximum radius of 30 km around meteorological stations.

CONCLUDING REMARKS

The quality of situation discrimination is very high taking into account that information on slope morphology and dimensions, underlying surface, snow conditions etc. was not used.

The results obtained are confirmation of an idea that statistical methods of slushflow diagnostics and forecasting established for one area, can be successfully applied in other areas if diagnostic parameters are related to the starting zones.

In spite of quite different climatic conditions and slushflow regimes in the Khibini Mountains and in the coastal districts of Norway, there are no significant differences in the results of the classification. Nevertheless it would make sense to test the obtained classification rules in other geographical areas.

Though the methods described are formally two-dimensional they use a synthetic parameter - daily water income - which is calculated on the basis of five meteorological parameters. Data on water supply calculated with regular meteorological information are the most informative for discrimination of "slushflow" and "non-slushflow" situations.

ACKNOWLEDGEMENTS

This work was supported by the Research Council of Norway within the frame of the Nordic Scholarship Scheme, by the Norwegian Geotechnical Institute and by the Centre of avalanche safety of "Apatit" JSC. The support is gratefully acknowledged. The authors also wish to thank the Norwegian Meteorological Institute which placed meteorological data at our disposal.

REFERENCES

Anderson T. 1958. *Introduction to multivariate statistical analysis.* John Wiley and Sons, New York, 500.

Bois P., Obled C. and Good W. 1975. Multivariate data analysis as a tool for day-by-day avalanche forecast. *IAHS Pub.* 114 (Simposium at Grindelwald 1974 - *Snow mechanics*), 391-403.

Chernouss P. 1975. Primenenie mnogomernogo discriminantnogo analiza dlya raspoznovaniya lavinoopasnih situacii. Issledovaniya snega i lavin v Hibinah. [An application of multivariate discriminant analysis for recognition of avalanche-danger situations. *In book: The studies of snow and avalanches in the Khibini Mountains* (The Khibinian Glaciological Conference at Kirovsk 1972)] Leningrad, Gidrometeoizdat, 64-70.

Harstveit K. 1984. Snowmelt modelling and energy exchange between the atmosphere and melting snow cover. Bergen, University of Bergen, Geophysical Institute. *Scientific report* 4, 120.

Hestnes E., Bakkehoi S., Sandersen F. and Andersen L. 1994. Weather and snowpack conditions essential to slushflow release and downslope propagation. *International Snow Science Workshop Proceedings. 1994, Snowbird, Utah,* 40-57.

McClung D. and Tweedy J. 1994. Numerical avalanche prediction: Kootenay Pass, British Columbia, Canada. *J. Glaciol.*, 40(135), 350-358.

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

Zuzin Yu. L. 1989. Opit ispolzovaniya bayesovskogo podhoda dlya indikacii lavinnoi opasnosti v Hibinah. Trudi tretego vsesoyuznogo soveshaniya po lavinam. [An experience of Bayes' approach used for indication of avalanche danger in the Khibini Mountains. *Proceedings of III - All-Union Workshop on Avalanches. September 6-9, 1986, Kirovsk*] Leningrad, Gidrometeoizdat, 185-187.