A Nearest Neighbor Model for Regional Avalanche Forecasting

Bernhard Brabec, brabec@slf.ch[1]

[1] Swiss Federal Institute for Snow and Avalanche Research (SLF), Flüelastr. 11, CH-7260 Davos-Dorf, Switzerland

Abstract: This paper describes the application of nearest neighbor algorithms to the task of regional avalanche forecasting. Several problems had to be solved and are described: data used as explanatory variables (snow and weather data), data used as dependent variable (avalanche hazard judgments described with the European Avalanche Hazard Scale) and validation of the model by cross-validation.

Keywords: avalanche hazard; nearest neighbors; time series; cross-validation

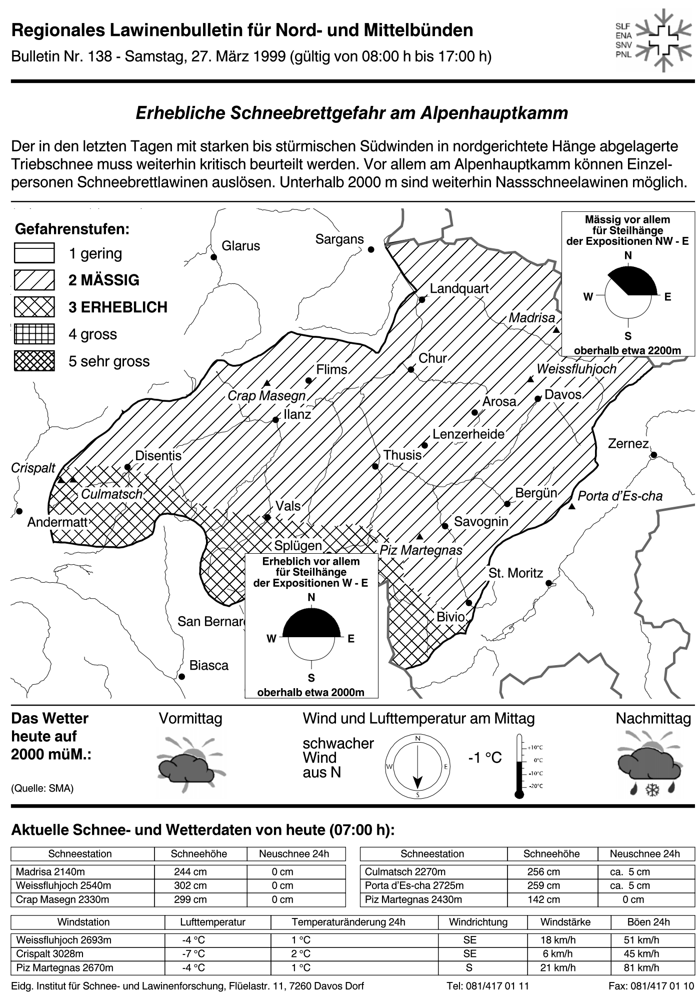
1 Introduction

Since 1945 the Swiss Federal Institute for Snow and Avalanche Research in Davos (SLF) has been in charge of the national avalanche forecast which covers an area of about 27'000 km^2. Analyzing daily

- snow-pack observations and avalanche occurrences from a network of about 40-75 manned stations,
- information about the current weather,
- the weather forecast and
- about 100 snow profiles recorded twice a month

an avalanche bulletin was compiled two to three times a week or whenever the situation changed dramatically. With the introduction of the European Avalanche Hazard Scale in 1993 a common language to describe the snow cover stability and the probability of an avalanche release has been found which is now being used in all European countries.
Until the early 1990s avalanche forecasting at SLF was mainly based on intuition and experience of the forecasters evaluating measurements and observations. While these factors still play an important role today, progress in the development of computer models as well as rapid developments in sensor, communication and information technology during the last 5 to 10 years have opened up new ways in avalanche forecasting. Mathematical

analysis of measurements, numerical simulations of weather and snow-pack data and statistical computations of the avalanche hazard provide new tools for modern avalanche forecasting. In 1997 regional avalanche forecasts have been introduced in order to provide information about the avalanche hazard on a more regional scale (about 5000 km^2) and in a way that it can be understood by people having little avalanche education (see figure 1). The



Regionales Lawinenbulletin für Nord- und Mittelbünden
Bulletin Nr. 138 - Samstag, 27. März 1999 (gültig von 08:00 h bis 17:00 h)

Erhebliche Schneebrettgefahr am Alpenhauptkamm

Der in den letzten Tagen mit starken bis stürmischen Südwinden in nordgerichtete Hänge abgelagerte Triebschnee muss weiterhin kritisch beurteilt werden. Vor allem am Alpenhauptkamm können Einzelpersonen Schneebrettlawinen auslösen. Unterhalb 2000 m sind weiterhin Nassschneelawinen möglich.

Aktuelle Schnee- und Wetterdaten von heute (07:00 h):

Schneestation	Schneehöhe	Neuschnee 24h	Schneestation	Schneehöhe	Neuschnee 24h
Madrisa 2140m	244 cm	0 cm	Culmatsch 2270m	256 cm	ca. 5 cm
Weissfluhjoch 2540m	302 cm	0 cm	Porta d'Es-cha 2725m	259 cm	ca. 5 cm
Crap Masegn 2330m	299 cm	0 cm	Piz Martegnas 2430m	142 cm	0 cm

Windstation	Lufttemperatur	Temperaturänderung 24h	Windrichtung	Windstärke	Böen 24h
Weissfluhjoch 2693m	-4 °C	1 °C	SE	18 km/h	51 km/h
Crispalt 3028m	-7 °C	2 °C	SE	6 km/h	45 km/h
Piz Martegnas 2670m	-4 °C	1 °C	S	21 km/h	81 km/h

Eidg. Institut für Schnee- und Lawinenforschung, Flüelastr. 11, 7260 Davos Dorf Tel: 081/417 01 11 Fax: 081/417 01 10

FIGURE 1. Regional avalanche forecast for Nord- und Mittelbünden, March 27, 1999

construction of these regional forecasts is supported by a semi-automatic computer-system which will receive a module for decision support in the next years. The system is developed as part of the program Avalanche Warning Switzerland 2000 as described in Russi et al. (1998).

2 Concept of a model to forecast the regional avalanche hazard

The SLF has a long tradition in building models to forecast avalanche related events as shown by Föhn (1998). Nevertheless only few of the proposed models have been used by the operational warning service - mostly because of the fact that the models could not be extended for more than a small region (e.g. Parsenn) and because their handling is time-consuming. A regional hazard model useful for SLF has to

- be based on operationally available data,
- calculate an avalanche hazard map (see figure 1),
- cope with missing data,
- allow user interaction for the judgment of available data and
- explain its results.

An approach that has been especially succesful in local avalanche forecasting in the past are nearest neighbors as used by Buser (1983): for a given day the 10 nearest neighbors of the past are calculated based on data about snow and weather and a Euclidean, weighted distance metric (see equation 1).
This approach is now adapted to regional avalanche forecasting. A relational database containing national avalanche bulletins of 10 years is constructed (SLF (1989-1997)). The hazard estimations given in the avalanche bulletins are supposed to be correct for about 70% (see Brabec and Stucki (1998)). In a second database data from manned observation stations and automatic weather stations are collected. The manual data is measured on a daily basis between 6 a.m. and 8 a.m. and contains numerical and categorical data (see table 1 and Gliott and Föhn (1989)). The data of the automatic stations contains only numerical data but is measured every 30 minutes. The locations of the stations are all over Switzerland. For the two types of stations different nearest neighbor models are developed. For each station the weighted mean of its neighboring days is calculated. These locally calculated hazard levels are then interpolated regionally to construct a hazard map.

3 Implementation of the model

The implementation of the model has been done by extending S-PLUS Version 3.4. A model called NEX-BEO has been implemented for the data

variable	description	weight	observation
HN	new snow depth	5	HN
HS	snow depth	1	HS
WI	weather and intensity $WI = 10W + I$	1	W, I
EE	east component of wind $EE = FF * cos(45 * D + 225)$	3	D, FF
NN	north component of wind $NN = FF * sin(45 * D + 45)$	3	D, FF
Ta	air temperature	2	T_a
Ts	snow temperature	2	T_S
Sf	snow surface	1	S_f
ES	penetration depth	2	PS
Di	density of new snow $Di = \frac{HNW}{HN} * 100$	2	HN, HNW
SHN	3 day sum of new snow depth $SHN_t = HN_t + HN_{t-1} + HN_{t-2}$	2	HN
DTa	air temperature difference $DTa_t = Ta_t - Ta_{t-1}$	3	T_a

TABLE 1. transformation between SLF observers' data and NEX-BEO and weights

of the manned stations. The distance metric (see equation 1) is Euclidean and treats categorical data numerically by introducing numeric levels. The distance between day t and day u is calculated as

$$dist^{t,u} = \sum_{i=1}^{nv} w_i (v_i^t - v_i^u)^2 \tag{1}$$

nv: number of variables v

w_i: weight for variable v_i

Some of the input data have also been transformed including data from previous days. The weights w_i have been assigned by experience and rules of thumb given by the forecasters (see table 1). For the calculation of the mean hazard level k nearest neighbors are considered. For the decision which hazard level to forecast different criteria for the decision boundaries are used and compared:

- class means: 1.5, 2.5, 3.5, 4.5
- experience of the forecasters: 1.85, 2.75, 3.65, 4.55. This model is called NEX-MOD.

- optimal boundaries calculated by preserving the probabilities of the classes (densities are estimated by cross-validation). Figure 2 shows the estimated a priori class probabilities and the cumulative density function for the prediction of k=10 for data from Davos Weissfluhjoch. The construction of the decision boundaries (lines) is illustrated at the right side of the graph: cumulative density is shown on the y-axis whereas the resulting values for the boundaries are given on the x-axis.

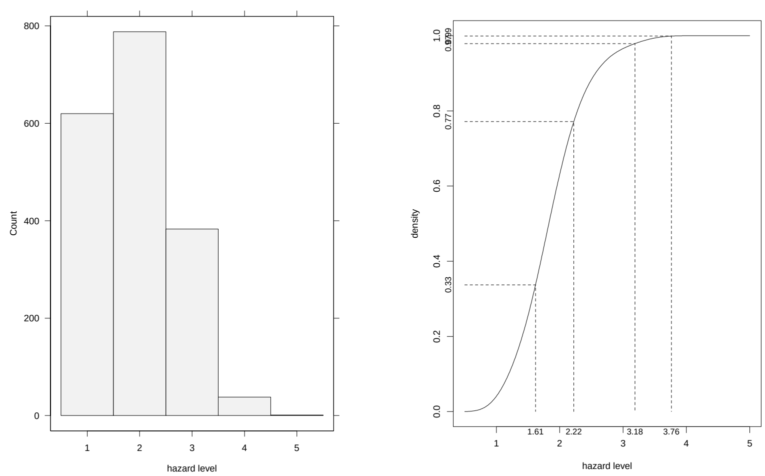


FIGURE 2. A priori class probabilities and cumulative density function for the prediction of k=10 for data from Davos Weissfluhjoch.

Days are then sorted according to the results of the distance function. The hazard level for the station is calculated as the mean over the k nearest days. These local results are then used as input for the construction of the hazard map by interpolation using a simple linear model.

4 Evaluation of the model

For the evaluation of the models cross-validation as proposed by Stone (1974) is applied. Each winter is eliminated from the dataset and then forecasted with the rest of the data.
The number of nearest neighbors k has been varied between 1 and 30. Figure 3 shows results of the comparison for different values of k for data from the area around Davos. Two manned stations have been considered situated about 3.6 km from each other: Weissfluhjoch (5WJ) located at

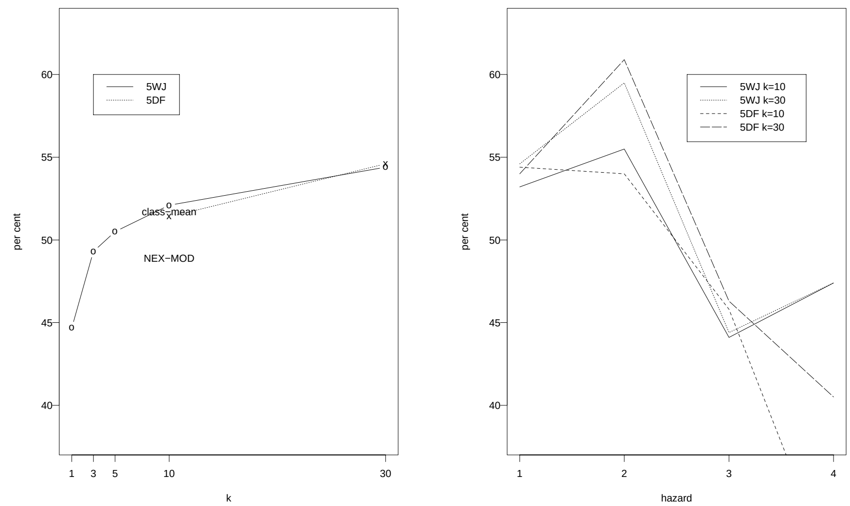


FIGURE 3. Comparison values of k for NEX-BEO. The right graph shows the performance of the 4 best models for all classes.

FIGURE 4. Best models compared for 10 years

2540 m and Davos (5DF) at 1560 m. Overall the performance (correct classification in per cent) is increasing with increasing k for the optimal decision boundaries (marked as `o` and `x`). `NEX-MOD` and boundaries from `class-mean` perform worse for k=10. As the right side of figure 3 shows data from 5WJ is better suitable for predicting higher hazard levels (3 and 4) whereas 5DF shows a drop in performance there. A value of k=30 shows better performance for the lower hazard degrees (1 and 2). For further comparisons only optimal decision boundaries for 5WJ, 5DF and k=10,30 have been considered.
Figure 4 shows the variation of these models for different winters. Especially the winters 1989/**90** and 1990/**91** are badly predicted. Reasons for this decrease in performance might be the size of the dataset (only ten winters) or missing measurements describing the snow cover stability.
The two best models (k=30 in figure 3) have been chosen for the regional averaging procedure. Including local decision boundaries before regional averaging has reduced the performance whereas the calculation of new decision boundaries has shown an overall performance of 55.2% which is slightly better than the two individual local models.

5 Outlook

The model discussed in sections 3 and 4 will be applied for all manned SLF observer stations in Switzerland.
For the automatic stations a new distance metric has to be defined. For each parameter a 5-day time-series is used and the difference is weighted exponentially over time. The sum over the variables is calculated according to equation 2:

$$dist^{t_0,u_0} = \sum_{i=1}^{nv} w_i \sum_{t=0}^{no} \exp^{-a*t}(v_i^{t_0-t} - v_i^{u_0-t})^2 \qquad (2)$$

nv: number of variables v

no: number of observations within 5 days (e.g. 240 in our case)

w_i: weight for variable v_i

a: influence of previous days; after 5 days near 0

The distance function for the model of the automatic stations is then minimized over days.
The result of both types of stations will then be used for the calculation of hazard maps for all over Switzerland. Extensions and further development of the model will include ideas presented in Dasarathy (1991) and Kristensen and Larsson (1994).

Acknowledgments: We would like to acknowledge the support of the Swiss National Science Foundation, the Swiss government and the Board of the Swiss Federal Institutes of Technology. Furthermore, a number of people at SLF and ETH are involved in supporting the development of the models and delivering input data. In particular we would like to thank Walter Ammann, Othmar Buser, Paul Föhn, Martin Gassner, Martin Mächler, Roland Meister, Tom Russi, Martin Schneebeli, Jürg Schweizer, Werner Stahel, Urs Stöckli, Andreas Stoffel and Martin Zimmerli for their contributions and discussions.

Bibliography

B. Brabec and Th. Stucki. Verification of avalanche bulletins by questionnaires. In *25 years of Snow Avalanche Research at NGI*, Voss, Norway, 1998.

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

Belur V. Dasarathy. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, 1991.

Paul M.B. Föhn. An overview of avalanche forecasting models and methods. In *25 years of Snow Avalanche Research at NGI*, Voss, Norway, 1998.

Sievi Gliott and Paul Föhn. Handbuch für Beobachter. Interner Bericht 637, Eidg. Institut für Schnee- und Lawinenforschung, 1989.

Krister Kristensen and Christer Larsson. An avalanche forecasting program based on a modified nearest neighbour method. In *International Snow Science Workshop*, pages 22–30, Snowbird, Utah, 1994.

T. Russi, W. Ammann, B. Brabec, M. Lehning, and R. Meister. Avalanche warning switzerland 2000. In *International Snow Science Workshop*, Sunriver, Oregon, 1998.

SLF. Schnee und Lawinen in den Schweizer Alpen 1987/88-1995/96. Winterberichte des Eidg. Institutes für Schnee- und Lawinenforschung Weissfluhjoch/Davos, 1989-1997. Nr. 52-60.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, 36(1):111–147, 1974.