Predictions in avalanche forecasting

D. M. McClung

Department of Geography, University of British Columbia, 1984 West Mall, Vancouver, British Columbia V6T 1Z2, Canada

ABSTRACT. Verification of avalanche forecasts depends on the spatial and temporal scale of the forecast, and the classes of informational entropy of data implicit in the forecast. First I present a classification system for avalanche forecasts based on these parameters. Verification of models in avalanche forecasting may consist of two stages. Often, the first stage is to ensure that the model matches the scales (space and time) and the classification of forecast and that redundant variables and parameters are eliminated. Once that is achieved, verification can proceed to the second stage, testing the model against relevant field data and situations. I provide an example based on the public-danger scale bulletin used for warnings in the back country in North America and Europe. Using data on deaths and accidents from Alpine Europe with Bayesian statistics, I conclude the danger scale has more classes than necessary for back-country applications. This could be a first stage prior to actual verification of this experience-based model.

INTRODUCTION

Avalanche forecasts vary according to spatial and temporal scale and data collected or information available before making a forecast. A systematic approach to verification can profit from a classification system for forecasts. Such a system is presented for the first time in this paper, based on time and spatial scales and data available.

In avalanche forecasting, errors increase if model constraints and data input do not match the scales (time and spatial) or if redundant variables or prediction categories are retained. Given the variety and classes of forecasts there can be two stages to verification: (1) ensuring that data available and model constraints match the scales (time and spatial) and that only variables or portions of the model which are necessary are retained, and (2) verification of the corrected model using data from field testing. It is not useful to proceed to verify a model which is not a physical match for the problem attempted. In this paper, I discuss the process and present an example of the first stage of verification for the experience-based public-danger scale for back-country warnings. The analysis results in a simplified scale which takes into account human perception, the root cause of most back-country avalanche accidents.

CLASSIFICATION OF AVALANCHE PREDICTIONS

To provide a framework to discuss verification of avalanche forecasts, it is useful to classify forecasts in relation to data used. The discussion here is in two parts: classification of the data used in avalanche forecasting, and the classification scheme in terms of data predicted or measured/observed. The classification scheme follows the logic of Lambe (1973) for geotechnical predictions. To frame the discussion, it is useful to have the definition and goal of avalanche forecasting. I propose these as:

Definition: Avalanche forecasting is the prediction of current and future snow instability in space and time relative to a given triggering (deformation energy) level.

Goal: The goal of avalanche forecasting is to minimize the uncertainty about instability introduced by the temporal and spatial variability of the snow cover (including terrain influences), any incremental changes in snow and weather conditions and any variations in human perception and estimation.

Data used in forecasting

Data used to forecast avalanches consist of two general types: singular or case data about the specific situation at hand, and distributional data and information about similar situations in the past. The classification system below is stratified according to measurements or observations of singular data, but both singular and distributional data will normally be used to make a forecast. Distributional data (information) are not measured to make a forecast, but are contained in experience, rule-based expert systems or calculations such as nearest neighbours (Buser and others, 1987) from computer models.

Individual data elements can be further classified according to their informational entropy: the relevance and ease of interpretation for prediction of avalanche occurrences (LaChapelle, 1980, 1985; McClung and Schaerer, 1993). Three general classes are identified:

III. Snow and weather data: mostly numerical, measured at, near or above the snow surface

II. Snowpack factors: data from the snowpack including snow structure, layering, snowpack parameters: mostly non-numerical

I. Stability factors: data such as stability tests which give direct information about avalanches (including avalanche occurrences).

McClung and Schaerer (1993) provide a comprehensive

Table 1. Classification for avalanche predictions

Type	When prediction made	Typical singular data classed by informational entropy (excluding avalanche occurrences)	Results at time forecast
A	Before incremental changes[1] measured or sensed	III: predicted II: predicted I: predicted	True forecast
B	Incremental changes have taken place or are taking place	III: mostly measured or observed II: predicted or partially[2] observed I: predicted or partially observed	Avalanche occurrences not yet searched
B1	Incremental changes have taken place or are taking place	III: mostly measured or observed II: predicted or partially observed I: predicted or partially observed	Avalanches searched for
C	Incremental changes have taken place	III: measured or observed II: observed I: observed	Current instability or "Nowcast" based on data sampling and observations including avalanche occurrences

[1] Incremental changes refer to snow and weather.

[2] Partially observed refers to data observed, for example in a snow profile, rather than extensive data sampling.

discussion of the factors. In general, the higher the class, the higher the informational entropy (uncertainty) with respect to prediction.

Classification scheme

Following Lambe (1973), I place predictions into three main categories in terms of uncertainty in data and measurements available. The categories and descriptions are given in Table 1. The aim is to present an approximate method to class forecasts to frame discussion of verification.

Almost all initial (early-morning) forecasts in ski or highway operations are type B since snow and weather observations are taken and sometimes stability tests are performed in a snow profile but often avalanche occurrences are not yet scanned for. Type A forecasts depend heavily on the accuracy of weather forecasts including snowfall, and, as such, they contain the inherent uncertainty of weather forecasts as well as the effects on snowpack instability. Good snow and weather forecasting, combined using SAFRAN, Crocus and MÉPRA models in France (Durand and others, 1998), provides a type A forecast. Prediction of the public-danger scale for 48 and 72 hours in advance (Cagnati and others, 1998) is also a type A forecast. Type C forecasts are typically made in helicopter skiing operations or back-country travel after avalanche occurrences are scanned for and stability tests and skiing have been performed.

Typically, as the forecast progresses from A to C, the spatial scale of the forecast reduces and more information is sought. A type A forecast is usually targeted toward the synoptic or meso-scale, whereas type B forecasts are usually targeted to-

ward the meso-scale. Type C forecasts are usually sought in back-country skiing including helicopter skiing where interest is on the micro-scale and extensive information is sought to make predictions for particular ski runs or terrain features. In general, as the scale of forecasts decreases, the difficulty increases (higher accuracy is sought) and people compensate by seeking more information, particularly of low entropy. The reader is referred to McClung and Shearer (1993) for a discussion of spatial scales in avalanche forecasting.

Avalanche forecasting is not an event, but an evolutionary process by which the forecast is updated as more information is collected. Table 1 reflects this evolutionary process as time proceeds. Back-country travellers often begin with only synoptic-scale information (e.g. type A or B such as public-danger scale bulletins), and this must be updated with low-entropy data to produce an optimal forecast relevant to the micro-scale (e.g. type C).

The reduction method (Munter, 1997) is a proposal to bypass low-entropy data collection such as stability tests or other information from snow profiles for micro-scale forecasting, but such a proposal can be easily discounted under conditions of highest probability of involvement (see Analysis section below) when avalanche occurrences may not be available. In these cases, other low-entropy data are of vital importance (Föhn and Schweizer, 1995) and variations in human perception are greatest. There are no short-cuts to good avalanche forecasting.

VERIFICATION OF AVALANCHE FORECASTS

Verification of avalanche forecasts is linked to the spatial scale that forecasting is made for. McClung and Schaerer (1993) discuss three scales for forecasting: synoptic (a significant portion of a mountain range: order of $10\,000\,km^2$), meso (spatial scale such as a highway avalanche area or a typical ski area: about $100\,km^2$) and micro (for an individual ski run or terrain feature including back-country skiing: $<1\,km^2$). As the scale decreases, the need for accuracy increases, the forecast progresses from the general to the specific, and the verification problem changes from general to specific.

Verification at the synoptic scale (e.g. Elder and Armstrong, 1987) may consist of looking at the general picture of avalanche occurrences. Verification at the micro-scale requires extensive data sampling including test skiing, test profiles and avalanche-occurrence information.

Statistical models including neural networks, parametric discriminant analyses, non-parametric discriminant analyses (nearest neighbours) and expert systems are nearly all built using distributional data, and these models are normally applied at the meso-scale. Any such model built on distributional data cannot legitimately be verified by use of the data it is built on. Also, such models are valid only for the location from which training data are taken. They provide mostly type B forecasts.

Experience with computer-assisted forecasting (McClung, 1994; Föhn and Schweizer, 1995; Schweizer and Föhn, 1996) has shown that maximum expected accuracy is about 60–65% unless the expert help is applied (Bayesian statistics or interactive expert system), in which case accuracy improves to 70–75%. This is because data input consists largely of class III information which must necessarily give an incomplete picture of instability. For example, the expert system MODUL of Schweizer and Föhn (1996) pro-

vides a type B forecast using class III, class II and class I (usually unavailable) information to achieve 70–75% accuracy. Similarly, the study of McClung (1994) on verification of a numerical meso-scale forecasting model showed that the forecaster's input of judgemental information (e.g. Bayesian prior), which can include class II and I information, must be combined with numerical predictions to achieve 70–75% accuracy.

ANALYSIS: PUBLIC DANGER-SCALE BULLETINS FOR BACK-COUNTRY APPLICATIONS

When verification of a model is attempted, it is often assumed that the model provides an adequate picture of reality and that the physical definition of the model is correct. The five-part public-danger scale was developed from extensive practical experience and data on avalanche occurrences, and there are now enough data on fatalities and accidents to perform simple analyses of the scale with respect to a parsimonious model for the number of prediction classes. Föhn and Schweizer (1995) and Cagnati and others (1998) looked at verification of the five-part danger scale. I prefer first to use data on statistics of deaths and accidents to assess whether the model might be simplified or improved for back-country use: the first stage in model verification.

In this section, I present a simple analysis for the five-part public-danger scale as used in North America and Europe, with statistics on fatalities and accidents from Switzerland and France used to assess the model. The data include statistics on the deaths recorded (Switzerland: ten winters of data, 255 fatalities, 1981–91), accidents (involvements: including death, injury or avalanche release affecting people) recorded (France: five winters of data, 166 accidents, 1993–98) in each portion, $D_k (k = 1,2,3,4,5$: low, moderate, considerable, high, very high or extreme) of the danger scale and the fraction of time (exposure) that the danger scale was applied for each class (for typical descriptions of danger-scale classes and how they are applied in practice from data and experience, see Meister, 1994; Cagnati and others, 1998). The result of the analysis (given below) is that for back-country use a simpler model is possible by reducing the danger scale to four classes, with adjustment to account for variations in human perception.

Avalanche forecasting, including decisions, is part of a risk analysis. As such, there is a probabilistic nature to the problem which shapes the nature of forecasts including the number of reasonable levels or categories for the danger scale. When the danger level is specified by a forecaster, the result is a function of human perception of the forecaster which is conditioned by observations and judgemental estimates about the temporal and spatial state of instability of the snow cover. A back-country traveller using the danger-scale forecast makes use of it but adds his or her own perception of the situation. The result is that decisions are filtered through human perception from at least two sources. The number of levels in the danger scale and the wording attached to those levels can therefore have significant effects upon decisions. The number of levels should be consistent with an order-of-magnitude probabilistic analysis (not too many levels), and the descriptors attached to the levels should take into account human perception and its variations as contained in the goal of avalanche forecasting. The great majority of avalanche accidents and fatalities in western Europe and North America now occur in the back

Table 2. Posterior probability for deaths ($P(D_k:D)$) and accidents ($P(D_k:A)$) and likelihood for deaths ($L(D_k:D)$) and accidents ($L(D_k:A)$)

Swiss death statistics: 255 deaths; 10 winters

D_k	$P(D_k)$	$P(D_k:D)$	$L(D_k:D)$
1	0.300	0.047	0.16
2	0.380	0.325	0.86
3	0.240	0.475	1.98
4	0.050	0.094	1.88
5	0.030	0.059	1.97

French accident statistics: 166 accidents; 5 winters

D_k	$P(D_k)$	$P(D_k:A)$	$L(D_k:A)$
1	0.154	0.000	0.00
2	0.393	0.084	0.21
3	0.344	0.422	1.22
4	0.094	0.422	4.49
5	0.015	0.072	4.80

Note: The likelihood is proportional to the ratio of posterior probability to $P(D_k)$ and is evaluated up to an arbitrary constant.

country (see McClung and Schaerer, 1993), and most of these are caused by human triggering. Thus, a major source of fatalities and accidents is failure in human perception: people perceived the state of instability of the snow cover to be something other than what it was. Any system for forecasting avalanches in the back country which does not take into account human perception is incomplete.

Since avalanche forecasting is related to a probabilistic risk analysis, it is useful to analyze data about the danger scale in a probabilistic sense. For the danger scale, I define a temporal exposure probability $P(D_k)$ as the fraction of time for which danger-scale level D_k is applied. I define a conditional probability $P(D:D_k)$ as the probability that death (D) occurs given that level D_k is applied, and similarly $P(A:D_k)$, with A (accident) replacing D (death). The quantities $P(D:D_k)$ and $P(A:D_k)$ are both defined only if death or accident is assumed to occur, and they are both related to the likelihood of death or accident. Similarly, I define the conditional probabilities $P(D_k:D)$ and $P(D_k:A)$ as the probabilities that death or accident occur in D_k (given death or accident). The latter probabilities define posterior probabilities in a given danger level D_k in Bayes' theorem as posterior probability $\propto$ likelihood $\times$ exposure. By application of Bayes' theorem, the posterior probability is obtained:

$$P(D_k:D) = \frac{P(D_k)P(D:D_k)}{\sum_{k=1}^{5} P(D_k)P(D:D_k)} = \frac{P(D_k)L(D_k:D)}{C},$$

where C is a normalization constant, $C = P(D)$. From the above equation, the likelihood $L(D_k:D)$ is proportional to $P(D_k:D)/P(D_k)$ up to an arbitrary constant (Edwards, 1992). The equation gives the elements of a probability mass function (pmf) for the posterior probability, $P(D_k:D)$, combining likelihood and temporal exposure in danger level k if death occurs. An analogous expression for accident can be written by substituting A for D. The probability and likelihood elements are given in Table 2.
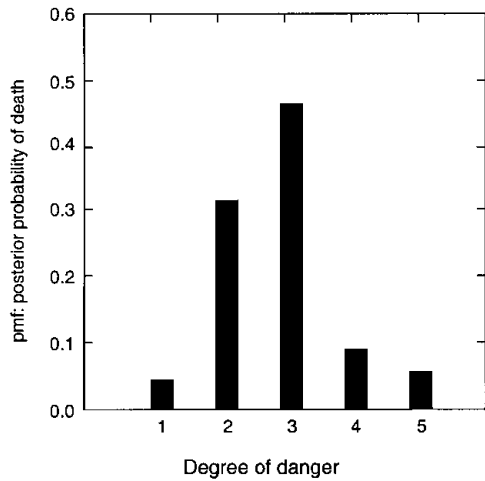
Fig. 1. Probability mass function (pmf): posterior probability of death for public-danger scale levels 1 (low), 2 (moderate), 3 (considerable), 4 (high) and 5 (very high or extreme). Data are from fatalities in Switzerland.

ANALYSIS RESULTS

Table 2 and Figures 1 and 2 show that maximum posterior probability is around the considerable $(k = 3)$ range. Of further interest is that, for the posterior probability, the fraction of the pmf contained in danger level 5 is 0.059 (Swiss deaths) and 0.072 (French accidents). This suggests that overall posterior probability in danger level 5 (which combines time exposure in a danger-scale class k and likelihood in the class k) is a small proportion of the total pmf. I also analyzed data from Tyrol, Austria (46 fatalities; 5 years of data, 1993–98), and the results are similar to those produced by the Swiss data: >80% of the deaths occur in the moderate–considerable range, with none of the posterior probability or fatalities in danger level 5.

Given that accident or death occurs, the likelihood expresses the likelihood of its occurring in D_k up to an arbitrary constant. Figure 3 shows the likelihoods for the death and accident statistics. For the present data, from Swiss death statistics, there is essentially no support for a hypothesis that deaths are more likely in level 5 than in level 4 or 3. French accident statistics show that accidents are slightly more likely in level 5 than level 4. Both datasets show that the likelihood increases sharply from the moderate $(k = 2)$ to the considerable $(k = 3)$ level.

The posterior probability and likelihood results can be combined to provide a suggestion for a simplified scale. The relatively low posterior probability in level 5 and the lack of sensitivity of likelihood in going from level 4 to 5 suggests that for back-country use levels 4 and 5 could be combined into one level, resulting in simpler scale: low and moderate $(k = 1$ and 2) could remain the same, with considerable $(k = 3)$ called high to represent high probability and the jump in likelihood, while high $(k = 4)$ and very high $(k = 5)$ could both be called very high.

Both likelihood $(L(D_k : D$ or $A))$ and posterior probability $(P(D_k : D$ or $A))$ are important for avalanche forecasting. On a given day with a danger level specified by a forecast, one is most interested in the likelihood of death or accident for the day. The estimates in Table 2 (column 3: posterior probability) contain implicit information in terms not just of number of people exposed and time of exposure but also of variations in human perception. Both the posterior probability and the likelihood estimates are of interest for model assessment.

Both likelihood and posterior probability contain information (not precisely known) about the number of people exposed, since specification of the danger level influences that number. The unknown number of people exposed prevents a full, formal risk analysis but does not affect the Bayesian analysis above since the assumptions for application of Bayes' theorem are not violated.

LIKELIHOOD-FUNCTION MODELS

I made numerical models of the likelihood values, $P(D_k : D)/P(D_k)$ and $P(D_k : A)/P(D_k)$, and the expressions are:

$$P(D_k : D)/P(D_k) = 1.33 \ln(D_k + 0.15)$$
$$(\% \text{ variance explained} = 98)$$
$$P(D_k : A)/P(D_k) = 2.63 \ln(D_k - 0.22)$$
$$(\% \text{ variance explained} = 87).$$

For the Swiss data on fatalities a similar expression was derived:

$$P(D_k : D) = 1.43 \ln(D_k + 0.10)P(D_k),$$

and a χ^2 (chi-square) goodness-of-fit test to fatality num-
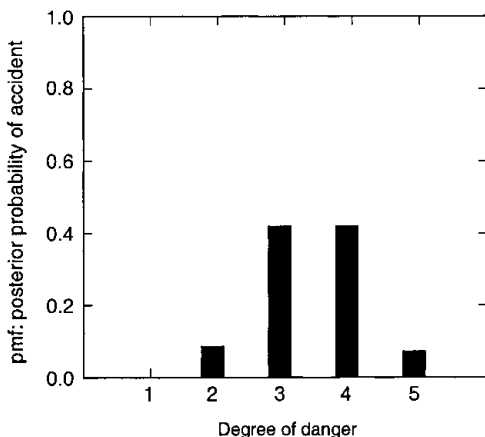

Fig. 2. Probability mass function (pmf): posterior probability of accidents for public-danger scale levels similar to Figure 1. Data are from accidents in France.
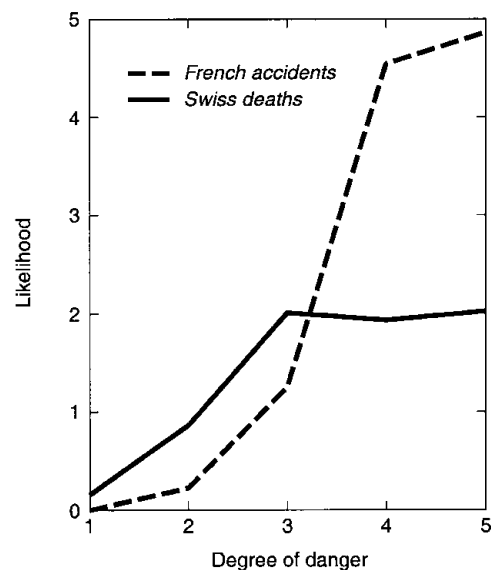

Fig. 3. Likelihood values that accidents or deaths occur per day of exposure vs danger-scale level.

bers gave $\chi^2 = 9.5$ with 4 degrees of freedom ($\alpha = 0.05$), indicating a good fit. These results indicate that likelihood increases approximately in a logarithmic fashion with D_k, and they mirror the lack of sensitivity of the ratios to D_k for $k = 4$ and 5, as implied by the data. The logarithmic dependence is encouraging since it is consistent with order-of-magnitude estimates typical of formal risk analyses.

DISCUSSION

Avalanche forecasting is analogous to a probabilistic analysis to gauge the snowpack instability in space and time relative to a given triggering level. When human activity is contemplated and decisions are included, it becomes a risk analysis. Risk analyses normally proceed by order-of-magnitude estimates with a limited number of useful risk classes.

Any model gives an imperfect picture of reality in avalanche forecasting, and no model can be expected to predict accurately effects which are not contained in data input (entropy classes). The first stage in model verification may consist of correcting model deficiencies and/or illuminating inadequacies. Table 1 provides a framework for discussing forecast type based on scale in space and time and on data available. Once model input and predictive capability matches the physical problem, including scale and type, a second stage of verification can be appropriate. It is not useful to proceed with verification for models with a proliferation of parameters or mismatch of scale, data input, physical problem and model output.

The only data readily available to assess model characteristics for the public danger scale are low-entropy data from accident and death statistics and avalanche occurrences (or lack of them). In common with other aspects of avalanche-forecasting modelling, both high- and low-entropy data are used to specify D_k, including human experience, and verification must also involve high- and low-entropy data (Föhn and Schweizer, 1995).

The public danger scale was developed mainly from human experience and data on avalanches, but quantitative analysis suggests a simplified scale for application in the back country. The overall posterior probability (time exposure and likelihood in a danger class) contained in category 5 is fairly low (about 6–7% of the pmf) and the likelihood is nearly independent of D_k for categories 4 and 5. These results suggest that the scale could be simplified by calling D_3 high and both D_4 and D_5 very high. In addition to a simpler scale, such modification would mean that the category with highest overall posterior probability and sharply increasing likelihood (from moderate to considerable) would be called high instead of considerable, giving a stronger warning signal to the public. Since most deaths and accidents in Western countries are caused by human triggering, the root cause of such accidents is a failure in human perception, which is something the danger scale should take into account.

For back-country applications, the same level of caution for human activities should be used for categories 4 and 5 so that decision changes would be minimal for these levels. However, the proposal here is for a higher level of caution for category 3 where it is most needed: accidents and fatalities are most prevalent and likelihood

sharply increasing. In category 3, more people are exposed than at level 4 or 5 and variations in human perception of instability in the snow cover are expected to be greatest. The goal of avalanche forecasting includes reducing such variations.

From data analyzed thus far, the likelihood of death or accident increases approximately with the logarithm of D_k. This is encouraging since the forecasting process should mirror a probabilistic risk analysis with order-of-magnitude changes between danger levels.

The analysis in this paper applies only to danger-level warnings for back-country applications, not facilities or villages.

ACKNOWLEDGEMENTS

REFERENCES

Buser, O., M. Bütler and W. Good. 1987. Avalanche forecast by nearest neighbour method. *International Association of Hydrological Sciences Publication* 162 (Symposium at Davos 1986 — *Avalanche Formation, Movement and Effects*), 557–569.

Cagnati, A., M. Valt, G. Soratroi, J. Gavaldà and C. G. Séllès. 1998. A field method for avalanche danger-level verification. *Ann. Glaciol.*, **26**, 343–346.

Durand, Y., G. Giraud and L. Mérindol. 1998. Short-term numerical avalanche forecast used operationally at Météo-France over the Alps and Pyrenees. *Ann. Glaciol.*, **26**, 357–366.

Elder, K. and B. Armstrong. 1987. A quantitative approach for verifying avalanche hazard ratings. *International Association of Hydrological Sciences Publication* 162 (Symposium at Davos 1986 — *Avalanche Formation, Movement and Effects*), 593–603.

Föhn, P. M. B. and J. Schweizer. 1995. Verification of avalanche danger with respect to avalanche forecasting. *In* Sivardière, F., *ed. Les apports de la recherche scientifique à la sécurité neige, glace et avalanche. Actes de Colloque, Chamonix 30 mai – 3 juin 1995.* Grenoble, Association Nationale pour l'Étude de la Neige et des Avalanches (ANENA), 151–156.

LaChapelle, E. R. 1980. The fundamental processes in conventional avalanche forecasting. *J. Glaciol.*, **26**(94), 75–84.

LaChapelle, E. R. 1985. *The ABC of avalanche safety. Second edition.* Seattle, WA, The Mountaineers.

Lambe, T. W. 1973. Predictions in soil engineering. *Géotechnique*, **23**(2), 149–202.

McClung, D. M. 1994. Computer assisted avalanche forecasting. *In* Hipel, K. W. and Fang Liping, *eds. Statistical and stochastic methods in hydrology and environmental engineering. Vol. 4.* Dordrecht, etc., Kluwer Academic Publishers, 347–358.

McClung, D. M. and P. A. Schaerer. 1993. *The avalanche handbook.* Seattle, WA, The Mountaineers.

Meister, R. 1994. Country-wide avalanche warning in Switzerland. *In ISSW'94. International Snow Science Workshop, 30 October – 3 November 1994, Snowbird, Utah. Proceedings.* Snowbird, UT, P.O. Box 49, 58–71.

Munter, W. 1997. *3 × 3 Lawinen Entscheiden in kritischen Situationen.* Garmisch-Partenkirchen, Agentur Pohl and Schellhammer.

Schweizer, J. and P. M. B. Föhn. 1996. Avalanche forecasting — an expert system approach. *J. Glaciol.*, **42**(141), 318–332.