Verification of nearest-neighbours interpretations in avalanche forecasting

JOACHIM HEIERLI,[1] ROSS S. PURVES,[2] ANDREAS FELBER,[1] JULIA KOWALSKI[1]

[1]*WSL Swiss Federal Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, CH-7260 Davos-Dorf, Switzerland*
E-mail: heierli@slf.ch
[2]*Department of Geography, University of Zürich-Irchel, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

ABSTRACT. This paper examines the positive and negative aspects of a range of interpretations of nearest-neighbours models. Measures-oriented and distribution-oriented verification methods are applied to categorial, probabilistic and descriptive interpretations of nearest neighbours used operationally in avalanche forecasting in Scotland and Switzerland. The dependence of skill and accuracy measures on base rate is illustrated. The purpose of the forecast and the definition of events are important variables in determining the quality of the forecast. A discussion of the application of different interpretations in operational avalanche forecasting is presented.

INTRODUCTION

Nearest-neighbours (NN) avalanche forecasting compares data describing past avalanche and non-avalanche days with current or forecast data. In NN a distance between days in the dataset and the forecast day is defined to identify previous days which are most "similar" to the forecast day (the nearest neighbours). The nature of events on the nearest neighbours is then used to build hypotheses about the likely resulting avalanches (Buser, 1983, 1989).

Statistically, NN is a non-parametric pattern classification technique which arranges data in a multi-dimensional space and applies a distance metric (usually Euclidean) to define the distance between past and present data (Ripley, 1996).

Various NN forecast techniques are currently used operationally in local avalanche forecasting. All assume that similar events are likely to exhibit similar precursors and that snow and weather factors and/or snowpack factors can be extrapolated over the geographic forecast area (e.g. Buser, 1983, 1989; Gassner and others, 2001; McCollister and others, 2002; Mérindol and others, 2002; Purves and others, 2002).

In this paper, the forecasted event is defined as a day with one or more recorded avalanches in the forecast region (an *avalanche day*). When such an avalanche day is found amongst the nearest neighbours selected by NN, it is called a *positive neighbour*.

The nearest neighbours can be interpreted in a number of ways, with the three most common interpretations being:

Categorial forecast: Here *decision boundaries* are used to classify days into a number of forecast categories. Often these categories are dichotomous (avalanches forecast or not), and an avalanche day is forecast when the number of positive neighbours is greater than or equal to some defined decision boundary. Brabec and Meister (2001) have used NN in multi-categorial form to predict the five categories of the European avalanche-hazard scale.

Probability forecast: The probability of the event is estimated (e.g. "Avalanches are expected today with a probability of 10%"). A probability forecast relies on the ability of NN to produce an estimation of the a posteriori probability of an event. This posterior probability is then used in the forecast as the prior probability of the event. In practice, the number of positive neighbours divided by the total number of nearest neighbours is used to estimate the probability of an avalanche day. McCollister and others (2002) have used such an approach at Jackson Hole ski area, Wyoming, U.S.A.

Descriptive forecast: A detailed list of events and all associated, individual observations recorded in the past are provided by NN to the forecaster. This description is then used by the forecaster as an aide-mémoire characterizing the nature of the associated avalanche days. This information is combined with other available information, and further interpreted by the forecaster. This descriptive scheme, based on hypothesis testing as described by LaChapelle (1980), has been recommended by Buser (1983, 1989) and Purves and others (2002). It has been practised by users of *NXD* (Gassner and others, 2001), *Cornice* (Purves and others, 2002) and *Astral* (Mérindol and others, 2002).

Each interpretation requires adequate verification. The verification is intended to indicate the positive and negative aspects of differing interpretations of NN and to examine the possible influences of different datasets. The latter question was addressed using two datasets with different purposes utilized in operational avalanche forecasting.

The first dataset was used to forecast daily avalanche risk to roads, railway and settlement areas in a region of Valais, Switzerland, where the forecaster must decide whether roads or railways must be closed or endangered habitation evacuated.

Table 1. Summary characteristics for the Swiss and Scottish datasets (d = days, av. = avalanches, wi. = winters)

Dataset	Base rate %	Positive*	Number of events Negative**	Total
Swiss	7	78 d, 295 av.	970 d	1048 d, 6 wi.
Scottish	20	253 d, 525 av.	1031 d	1284 d, 12 wi.

* Positive events: avalanches recorded on that day.

** Negative events: no avalanches recorded on that day.

In the second dataset, the model was used in Lochaber, Scotland, by avalanche forecasters responsible for provision of back-country avalanche forecasts to mountaineers. These forecasts describe the current snow and avalanche conditions and their likely evolution over 24 hours and utilize the European avalanche-hazard scale to describe the degree of hazard.

In both cases, the forecasters utilize the descriptive interpretation of the 10 nearest neighbours. In the Swiss case the NN rule is performed by *NXD* (Gassner and others, 2001) and in the Scottish case by *Cornice* (Purves and others, 2002).

CHARACTERISTICS OF THE DATASETS

Although both datasets are used to describe avalanche events, they differ a great deal in the purpose of the forecasting being carried out and therefore in the nature and frequency of occurrence of the recorded events.

In the Swiss case, only large avalanches which may reach traffic lines or settlements are recorded. These avalanches often occur in conditions of High or Extreme avalanche hazard, and most are triggered naturally. Avalanches with no hazard potential to roads, railways or habitation are not recorded. The *base rate* (i.e. the fraction of all days in the dataset when avalanches were recorded) is 7%. In the Scottish case, a mountaineer might be dislodged or buried by even a small avalanche. Given that most events involving victims are triggered by those victims, then human-triggered avalanches are of particular importance to forecasters. Such conditions often equate to Moderate or Considerable hazard of avalanches on the European avalanche-hazard scale. The base rate is 20% for this dataset. Table 1 summarizes characteristics of each dataset.

VERIFICATION METHODS

Neither the quality of the Scottish and Swiss forecasts, nor *NXD* and *Cornice* are compared since both the underlying datasets and the forecast purposes do not match. Indeed, Murphy (1991) has shown that comparative verification of two forecast systems is a complex and high-dimensional problem compared to the absolute verification considered here.

Verification of the categorial forecast

The *measures-oriented verification* of dichotomous categorial forecasts can be divided into finding *accuracy measures* and *skill measures*. Such measures can be obtained from the joint distribution of observations and forecasts (Table 2), and a

Table 2. Joint distribution of forecasts and observations for binary categorial forecasts (contingency table)

Forecast	Observed Yes	No	Total
Yes	a	b	$a+b$
No	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

selection of such measures is introduced in Table 3. More detail on such measures can be found in Doswell and others (1990) and Wilks (1995, p. 238–250).

Verification of the probability forecast

Probability forecasts are best verified and interpreted by factorizing the joint probability distribution of observations and forecasts into conditional and marginal distributions, called *distributions-oriented verification* (Murphy and Winkler, 1986). Various aspects of forecast quality can be described by factorization. In this paper the following are examined:

> *Reliability*: also called calibration or conditional bias, it is quantified by the weighted average of the squared differences of forecast probabilities and the relative frequencies of the events in each subsample (Wilks, 1995, p. 262).

> *Resolution*: the ability to discern days with different avalanche-day probability.

> *Bias*: the general tendency to under- or over-forecast.

Furthermore, additional aspects of the forecast, such as skill, sharpness, discrimination and uncertainty can be deduced from other factorizations (Wilks, 1995, p. 258–272).

Verification of the descriptive forecast

If the description (event list and associated details) provided by the NN rule is intended to be used by the forecaster, then some adequate verification of this description is required. This verification should characterize the description with respect to its ability to provide the forecaster with meaningful information.

Table 3. Forecast verification measures (Doswell and others, 1990; Wilks, 1995)

Measure	Definition
Forecast accuracy measures (dichotomous forecasts)	
POD (probability of detection)	The probability that the event was forecast when it occurred. POD $= a/(a+c)$
SR (success rate)	The probability that the event occurred when it was forecast. SR $= a/(a+b)$
HR (hit rate)	The proportion of correct forecasts. HR $= (a+d)/(a+b+c+d)$
Forecast skill measures	
HSS (Heidke skill score)	Skill score based on HR. HSS $= (a+d-e)/(a+b+c+d-e)$ *
KSS (Kuipers skill score)	Like HSS, but marginal distribution of reference forecasts equal to base rate. KSS $= (ad-bc)/[(a+c)(b+d)]$

* $e = [(a+b)(a+c)+(b+d)(c+d)]/(a+b+c+d)$ is the expected number of correctly forecast events due to chance.
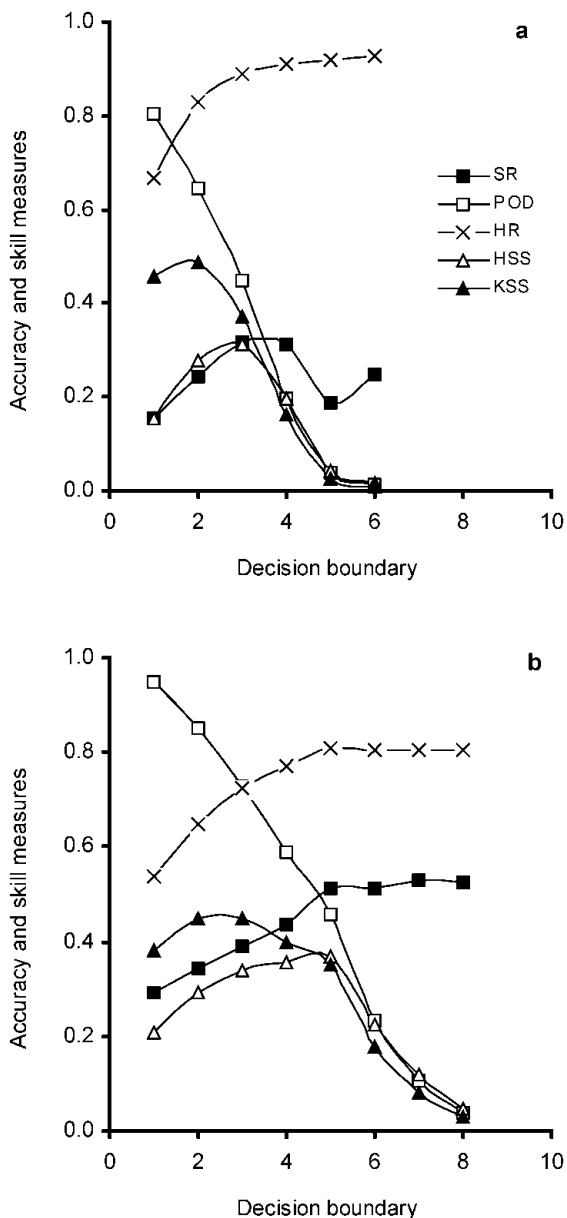
Fig. 1. Dependence of accuracy and skill measures (see Table 3) on the choice of decision boundary (number of positive neighbours of the forecast day). (a) Swiss dataset; (b) Scottish dataset. The forecast on the dataset with the lower base rate (a) exhibits a better HR, although its forecast is generally less accurate than the forecast on the dataset with the higher base rate (b), as evidenced by its better POD/SR pair.
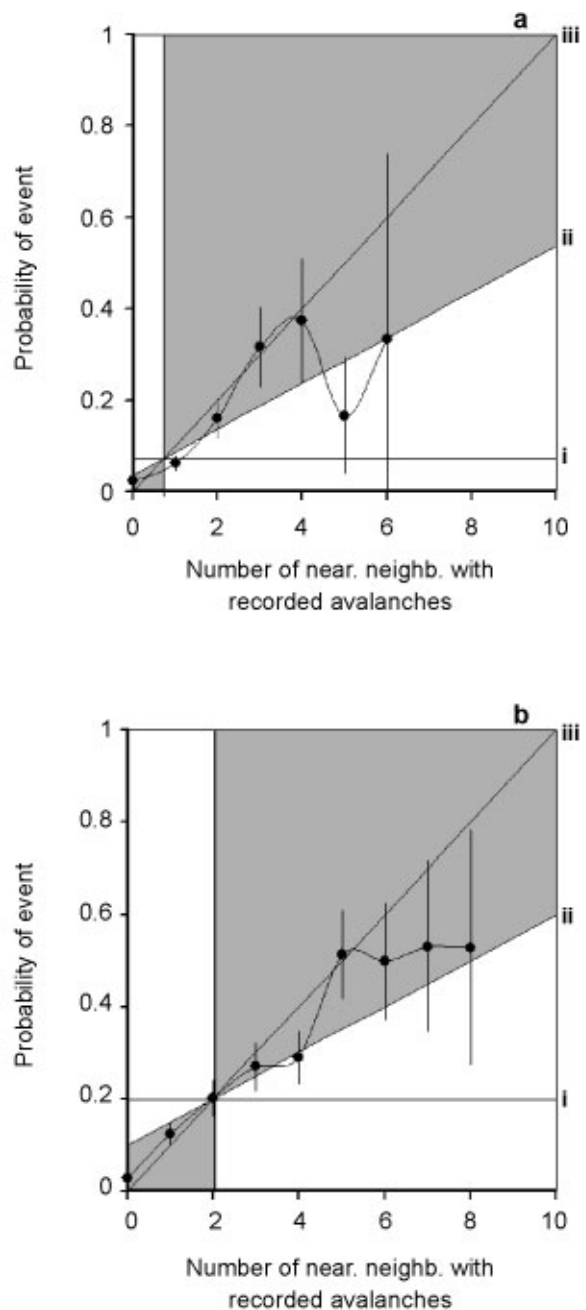
Fig. 2. Attributes diagrams showing the relation between the days (classed by their number of positive nearest neighbours) and the posterior probability of those days being events. (a) Swiss dataset; (b) Scottish dataset. The error bars denote the standard deviation of the Poisson distribution. Points close to line i have the least resolution; points close to line ii have no skill. Points in the grey zone contribute positively to skill, while points in the white zone contribute negatively. Points on line iii have the best reliability and skill.

In this paper, a first approach is presented, whereby the forecaster of the Swiss dataset was asked to perform a critical, subjective post-rating of each day when avalanches occurred in his region. Emphasis was laid on rating the *value of the information* provided by NN, not the quality of his final forecast. The NN description of each forecast day was rated as one of five ordinal categories: "severe misfit", "misleading", "unhelpful" (i.e. neither positive nor negative), "useful" or "very useful".

RESULTS

Here the results obtained from the verification of the three interpretation schemes are presented.

Categorial forecast

A measures-oriented verification was carried out to examine how the accuracy and the skill of the forecasts varied for a range of decision boundaries between 1 and 10 positive neighbours (Fig. 1a and b).

No results are given for decision boundaries above 6 in Figure 1a and above 9 in Figure 1b. No data with these numbers of positive neighbours were available in the respective datasets.

Probability forecast

A distributions-oriented verification was carried out to examine how well NN was able to produce a probability forecast, especially with regard to reliability and resolution as presented in the attributes diagrams (Fig. 2a and b).

Descriptive forecast

A summary of the subjective post-rating of the value of information provided by NN is presented in Figure 3. The histogram bars show the relative frequencies of the classes defined by how helpful the information was on days when avalanches occurred.

DISCUSSION

Categorial forecast

Figure 1 describes the dependency of accuracy (POD, SR, HR) and skill (KSS, HSS) on the choice of decision boundary (k). Various criteria may be used to specify the value of the decision boundary, such as $POD(k) = SR(k)$, $\max[KSS(k)]$ or $\max[HSS(k)]$. While Figure 1 is helpful in quantifying the dependency, the choice of decision boundary should be case-dependent and take account of human factors such as appreciation of risk and the consequences of unforecasted events and false alarms (McClung, 2002).

Despite Murphy's comments on the difficulties of comparison between datasets (Murphy, 1991), some simple comparisons between the Swiss and Scottish datasets can still be drawn. The Swiss data (Fig. 1a; base rate 7%) exhibit a higher HR than the Scottish (Fig. 1b; base rate 20%) while tbeir POD/SR pair is less accurate. It appears that these differences are driven chiefly by the base rate.

Probability forecast

The distributions on the attributes diagrams in Figure 2a and b exhibit several interesting features (Wilks, 1995, p. 266). The Swiss dataset (Fig. 2a) displays "unsteady" behaviour for days with over four positive neighbours, due to insufficient data. These data points result from only 15 out of 1048 days. This suggests that on a dataset with a base rate as low as 7%, 1048 data points still constitute an insufficient database for a definitive verification over the entire range up to ten neighbours. The Scottish dataset is also not entirely sufficient (Fig. 2b). Indeed, the attributes diagram exhibits a decrease in resolution for days with over five positive neighbours, indicated by the flattening of the curve to the right whereby the probability remains constant for an increasing number of positive neighbours.

Next, points with sufficient data are considered: days with zero to four positive neighbours in Figure 2a and with zero to six in Figure 2b. The closer the data come to line iii indicating perfect reliability, the better the forecast in this respect. Both forecasts exhibit good reliability which is a positive feature of NN. Both forecasts also exhibit little bias as shown by the equal distribution of data points above and below line iii.

Descriptive forecast

On 64% of forecasted avalanche days, the descriptive information provided by the NN rule was a posteriori judged
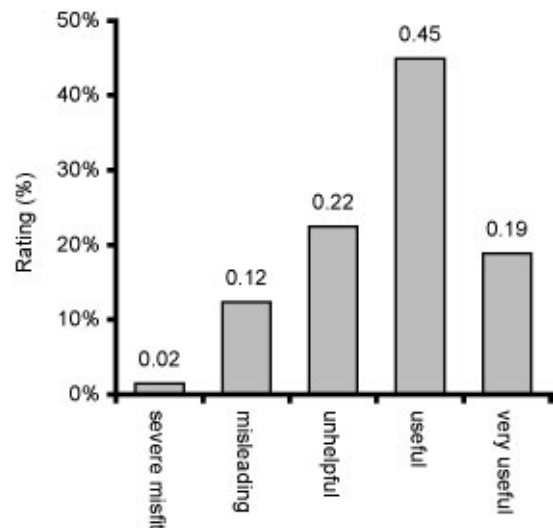

Fig. 3. Subjective a posteriori rating by the forecaster of the value of information provided by the descriptive event list obtained from the NN tool to produce the daily forecast. Most information in the description is helpful, but it is mixed with unhelpful or misleading information (Swiss dataset only).

useful or very useful by the forecaster. Severe misfits were exceptional and limited to 2% of the forecast days, while 12% of the descriptions were misleading and 22% unhelpful (Fig. 3). This indicates that the detailed description of the events in the nearest neighbours provides forecasters with valuable information.

Positive and negative aspects of interpretations of NN

All three interpretations provide some useful information content, but this is dependent on the intended application and the underlying data.

Categorial forecasting provides no room for interpretation by the forecaster: no information on the uncertainty of a forecast is available. Thus, if forecasters wish to utilize categorial forecasting it is key that they understand the implications of the POD/SR pair and the human factors related to false alarms and unforecasted events.

Probability forecasts may be helpful when used in a suitable context, but given that the definition of events in this case study is very broad — from a single avalanche to many in a given area and on a given day — a probability value on its own may be of limited use to the forecaster. Defining the events more precisely will inevitably produce less reliable forecasts due to the reduction of the base rate. This is a serious dilemma in avalanche forecasting, where the requirement is often to produce more precise forecasts (in terms of space, time or avalanche type).

Descriptive forecasts provide the most flexibility for the forecaster to interpret the nearest neighbours and associated avalanches. This interpretation, like any other part of the conventional avalanche-forecasting process, requires considerable knowledge and skill from the forecaster.

CONCLUSION AND FURTHER WORK

Measures-oriented verification quantifies the skill and accuracy of forecasts but does not allow comparative verification. Distribution-oriented verification of forecasts leads to

valuable information on the sufficiency of the database, the reliability of the forecast, its resolution and its bias.

NN apparently produces reliable, unbiased probability forecasts, but this must be verified case-by-case. Forecasters may find difficulty making decisions based only on probability forecasts. A low base rate is a serious limiting factor on the reliability and skill of a NN forecast.

The descriptive interpretation produces useful and interpretable forecasts, and an initial verification is presented in this paper. Many aspects of the value of information using descriptive NN remain unknown and will be investigated in further work.

ACKNOWLEDGEMENTS

REFERENCES

Brabec, B. and R. Meister. 2001. A nearest-neighbor model for regional avalanche forecasting. *Ann. Glaciol.*, **32**, 130–134.

Buser, O. 1983. Avalanche forecast with the method of nearest neighbours: an interactive approach. *Cold Reg. Sci. Technol.*, **8**(2), 155–163.

Buser, O. 1989. Two years experience of operational avalanche forecasting using the nearest neighbour method. *Ann. Glaciol.*, **13**, 31–34.

Doswell, C., R. Davies-Jones and D. L. Keller. 1990. On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, **5**, 576–595.

Gassner, M., K. Birkeland, H.J. Etter and T. Leonard. 2001. NXD2000: An improved avalanche forecasting program based upon the nearest neighbour method. *In ISSW 2000. International Snow Science Workshop, 1–6 October 2000, Big Sky, Montana. Proceedings.* Bozeman, MT, American Avalanche Association, 52–59.

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

McClung, D. M. 2002. The elements of applied avalanche forecasting — Part 1: The human issues. *Nat. Hazards*, **26**(2), 111–129.

McCollister, C., K. Birkeland, K. Hansen, R. Aspinall and R. Comey. 2002. A probabilistic technique for exploring multiscale spatial patterns in historical avalanche data by combining GIS and meteorological nearest neighbours with an example from the Jackson Hole Ski Area, Wyoming. *In* Stevens, J. R., *ed. International Snow Science Workshop 2002, 29 September–4 October 2002, Pentiction, British Columbia. Proceedings.* Victoria, B.C., B.C. Ministry of Transportation. Snow Avalanche Programs, 109–116.

Mérindol, L., G. Guyomarc'h and G. Giraud. 2002. A French local tool for avalanche hazard forecasting: Astral, current state and new developments. *In* Stevens, J. R., *ed. International Snow Science Workshop 2002, 29 September–4 October 2002, Pentiction, British Columbia. Proceedings.* Victoria, B.C., B.C. Ministry of Transportation. Snow Avalanche Programs, 105–108.

Murphy, A. H. 1991. Forecast verification: its complexity and dimensionality. *Mon. Weather Rev.*, **119**, 1590–1601.

Murphy, A. H. and R. L. Winkler. 1986. A general framework for forecast verification. *Mon. Weather Rev.*, **115**, 1330–1338.

Purves, R. S., K. W. Morrison, G. Moss and D. S. B. Wright. 2002. Cornice — development of a nearest-neighbours model applied in backcountry avalanche forecasting in Scotland. *In* Stevens, J. R., *ed. International Snow Science Workshop 2002, 29 September–4 October 2002, Pentiction, British Columbia. Proceedings.* Victoria, B.C., B.C. Ministry of Transportation. Snow Avalanche Programs, 117–122.

Ripley, B. 1996. *Pattern recognition and neural networks.* Cambridge, Cambridge University Press.

Wilks, D. S. 1995. *Statistical methods in the atmospheric sciences.* New York, Academic Press. (International Geophysics Series 59.)