

Statistical forecasting of regional avalanche danger using simulated snow-cover data

Michael SCHIRMER, Michael LEHNING, Jürg SCHWEIZER

*WSL Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, CH-7260 Davos Dorf, Switzerland
E-mail: schirmer@slf.ch*

ABSTRACT. In the past, numerical prediction of regional avalanche danger using statistical methods with meteorological input variables has shown insufficiently accurate results, possibly due to the lack of snow-stratigraphy data. Detailed snow-cover data were rarely used because they were not readily available (manual observations). With the development and increasing use of snow-cover models this deficiency can now be rectified and model output can be used as input for forecasting models. We used the output of the physically based snow-cover model SNOWPACK combined with meteorological variables to investigate and establish a link to regional avalanche danger. Snow stratigraphy was simulated for the location of an automatic weather station near Davos, Switzerland, over nine winters. Only dry-snow situations were considered. A variety of selection algorithms was used to identify the most important simulated snow variables. Data mining and statistical methods, including classification trees, artificial neural networks, support vector machines, hidden Markov models and nearest-neighbour methods were trained on the forecasted regional avalanche danger (European avalanche danger scale). The best results were achieved with a nearest-neighbour method which used the avalanche danger level of the previous day as additional input. A cross-validated accuracy (hit rate) of 73% was obtained. This study suggests that modelled snow-stratigraphy variables, as provided by SNOWPACK, are able to improve numerical avalanche forecasting.

INTRODUCTION

Regional avalanche forecasting attempts to predict current and future snow stability, relative to a given triggering level on the scale of a mountain range or a considerable fraction thereof (e.g. McClung and Schaerer, 2006). Forecasts are issued on a daily basis to warn the public about the level of avalanche danger. These public bulletins play a key role in the prevention of avalanche fatalities. Adequate avalanche warnings, combined with avalanche education and efficient rescue, have probably prevented an increase of avalanche fatalities in parallel with the increased recreational use of avalanche terrain, at least in some countries (Harvey and Zweifel, 2008). Reliable and consistent avalanche forecasts are therefore very much needed. To assess the avalanche danger level, most avalanche warning services rely on a combination of manual observations, automatic weather stations, weather forecasts (including model output) and snow profiles (Meister, 1995). For the locations of the automatic weather stations in the Swiss Alps, the amounts of new snow and drifting snow are additionally derived from the numerical snow-cover model SNOWPACK (Lehning and others, 1999; Lehning and Fierz, 2008). Based on all these data, the forecaster uses experience, intuition and local knowledge of the mountain range to estimate and describe the avalanche danger in the public bulletin.

Over the past decades there have been many attempts to create an objective process of danger evaluation, which may also work as a support tool for the avalanche warning service. The French model chain *SAFRAN/Crocus/MÉPRA* (SCM) provides automated avalanche danger prediction for virtual slopes (Durand and others, 1999) and is the only real operational derivation of a risk level on the basis of physical snow modelling. Durand and others (1999) published a contingency table between modelled risk and

avalanche activity with a hit rate of 75%. Since Murphy (1991) described the difficulties of comparison between two forecasting systems under different conditions (e.g. different datasets), the published accuracy measures of the studies described below are not further reviewed here. Several studies have used observations of avalanche activity as an indicator of the avalanche danger (e.g. Buser, 1983; Heierli and others, 2004; Pozdnoukhov and others, 2008). As Schweizer and others (2003) pointed out, the problem with this target parameter is that it does not distinguish between lower danger levels and that observations may be inconsistent, mainly due to limited visibility during times of avalanche activity. Schweizer and Föhn (1996) forecasted the avalanche danger level instead of the avalanche activity using commercial decision-making software. Input variables included snow-cover information observed manually. They trained their models using a verified danger level instead of the forecasted one. The verification was based on additional data and observations, including data not yet available at the time of the forecast. The cross-validated hit rate was 63%, but adding knowledge in the form of expert rules to the system, the performance improved to a hit rate of ~70%. However, their models did not run fully automatically but required manual input of the snow-cover information. Brabec and Meister (2001) used a nearest-neighbour method with only meteorological variables and snow information, such as penetration depth or surface characteristics, so that the model could be run for the whole area of the Swiss Alps. The accuracy was only ~52%. The absence of snow-cover information was given as a reason for the poor results.

Model output from snow-cover models, such as SNOWPACK, can provide snow-cover information with the required resolution in space and time. This study explores whether the performance of data-based forecasting models can be improved with modelled snow-cover data as additional inputs.

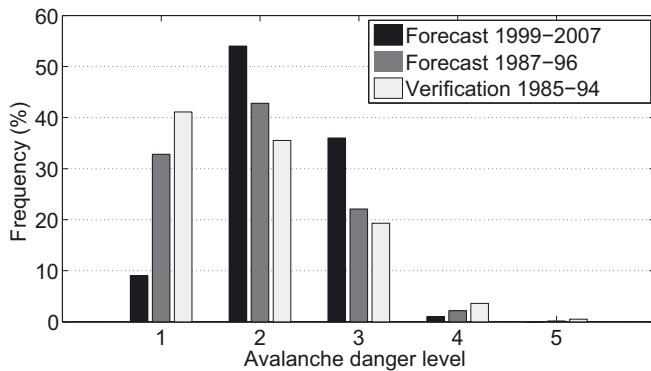


Fig. 1. Relative frequency of the avalanche danger levels in the region of Davos. Black columns show the forecasted levels during the time period considered in this study. Dark grey columns show the forecasted levels for the study of Brabec and Meister (2001) and light grey show the verified levels used in the study of Schweizer and Föhn (1996).

METHODS

Data

In order to establish a link between regional avalanche danger and modelled snow-cover variables, a test region was chosen where both modelled snow-cover data and an estimate of the regional avalanche danger were available. We selected the region of Davos in the eastern Swiss Alps (225 km²), which covers typical avalanche-release zones with an elevation range of 1600–2800 m a.s.l. The study plot in this test region indicates that the automated weather station Weissfluhjoch, at 2540 m a.s.l., shows an average maximum snow depth of ~2.2 m for the nine winters (1999–2007; 1229 days) when the required data were available.

The forecast of the regional avalanche danger level (Fig. 1), which is issued every day at 0800 h, was used as a proxy target parameter since an accurate measure of the avalanche danger is not available (Meister, 1995). The frequency of the danger levels in the chosen time period (black columns) can be seen in Figure 1 (1: 'Low', 2: 'Moderate', 3: 'Considerable', 4: 'High' and 5: 'Very high'; 'Very high' did not occur in this time period). The danger levels 'Moderate' and 'Considerable' were most frequent. The frequency in the chosen time period was substantially different from that in the periods used in the studies of Brabec and Meister (2001) (dark grey) and Schweizer and Föhn (1996) (light grey). For the danger levels 'Low' to 'Considerable' the avalanche danger was characterized by very high persistence. The probability that the avalanche danger tomorrow will be the same as today was ~80%.

As input variables, modelled snow-cover data were generated for the location of Weissfluhjoch. The relevant processes influencing the regional avalanche danger (e.g. new snow, wind or weak layer formation) are assumed to be represented in the chosen study plot. The snow-cover model SNOWPACK was used to model settling and layering of the snow cover, as well as its energy and mass balance (Bartelt and Lehning, 2002; Lehning and others, 2002a,b). This model requires meteorological data as input. We focused on dry-snow situations since dry-snow avalanches are the main threat over most of the winter. For forecasting wet-snow avalanches models need to be trained separately. For each winter a date was determined for the beginning of wet-snow conditions (i.e. 2 weeks before the snow cover modelled

for the Weissfluhjoch study plot became isothermal), which is usually around the beginning of April. We furthermore restricted the dataset to days with snow depth >75 cm, because we think that the stability part of SNOWPACK produces more reliable results with this restriction. This restriction excluded only a limited number of days because the avalanche danger forecast for our region started at about the time when that snow depth was reached – for most winters by the end of November.

A stability index (Schweizer and others, 2006) defined the potential weak layer interface in the modelled snow cover. Motivated by a study which evaluated stability from observed snow stratigraphy (Schweizer and Jamieson, 2003), characteristics of the weak and adjacent layer and of the slab were considered. These model variables were completed with measured and calculated meteorological and snow-surface variables (e.g. wind velocity, outgoing longwave radiation or surface albedo).

All statistical methods described here were tested on a dataset in order to assess their classification capability. This test dataset (Fig. 1) was used for the study of Schweizer and Föhn (1996) and contained a verification instead of a forecast of the regional avalanche danger level and observed, rather than modelled, snow-cover data as input variables. This dataset was more reliable than that covering the years 1999–2007, and therefore adequate to test the statistical methods for this specific problem. The test dataset covered ten winters between 1985 and 1994. Unfortunately no modelled SNOWPACK data can be produced for that time period, because no automated weather station existed before the 1990s. This meant that modelled and measured snow-cover data and their explanatory power could not be directly compared.

Performance measures

The quality of a method was assessed by a cross-validated hit rate (HR) for all danger levels and by the true skill score (TSS) for each of the four danger levels separately (Doswell and others, 1990; Wilks, 1995).

Since both input and target variables were autocorrelated (e.g. a weak layer might have an influence over a long time period), random cross-validation turned out not to be a useful method, giving unrealistically high hit rates. Therefore, each winter was forecasted using a model and a variable selection based on the remaining eight winters, i.e. both model parameters and input variables could change in this cross-validation scheme.

It seems useful to draw special attention to days on which the danger level changed: first, these days might be the most important to predict correctly; second, because of the high persistence of the target parameter, an obviously useless forecast which always predicts the level of the previous day, would show good hit rates for all days. Introducing HR and TSS for only those days on which the danger level changed helped to identify methods which tended to a persistent forecast. Similarly, the HR and TSS were also considered for the days on which the modelled danger level increased or decreased.

Variable selection

Since the snow-cover model delivers a large number of variables at high temporal resolution, variable selection is useful, firstly for data reduction, which increases the speed of the final statistical forecasting model, secondly to achieve

an overall improved performance and thirdly to understand which variables are important (Guyon and Elisseeff, 2006).

For many variables, it may make sense to also consider their sum, mean, extreme values or rate, for different time intervals or time lags. For example, the 24 hour change of the air temperature may be more correlated with the regional avalanche danger than the midday value, or the previous day's snowfall more correlated than that of the current day. This leads to a rapid increase in the number of possible variables. A simple univariate variable selection, a Fisher's discriminant analysis (e.g. Bishop, 2006), was performed to determine for each variable the two most important derived variables. After this step, a large number (300) of variables still remained. Subsequently a rating of this variable set was carried out with Fisher's discriminant analysis. An alternative ranking was obtained with univariate classification trees (Breiman and others, 1998). To avoid overfitting, the best tree was determined by a pruning algorithm based on cross-validation, and the number of terminal nodes was limited to ten. The cross-validated HR of these univariate pruned trees delivered a ranking between the variables. Only variables that were not pairwise linearly correlated ($r^2 < 0.6$) were considered. Since correlated variables are not redundant per se, the redundancy of excluded variables was visually examined with scattergraphs (Guyon and Elisseeff, 2006). For support vector machines (SVM) a variable selection algorithm was used, based on the Fisher's discriminant analysis in combination with SVM (Chen and Lin, 2006).

Additionally, for the nearest-neighbour method (KNN) a combination with a genetic algorithm (GA) was used to rank variable relevancy (Li and others, 2001). Several variable subsets (each of ten variables) were tested with the KNN method. 'Good variable subsets' (see below for the selection criteria) were stored in a final pool. Once the final pool was filled to a certain threshold, the cardinality of each single variable was interpreted as a relevancy ranking: the more often a single variable was selected as a member of a good variable subset, the more relevant the single variable was assumed to be. Since testing all possible ten-member variable subsets of 300 different variables would cost too much computing time, a GA was used as a search tool to create 'good parameter subsets': the GA maximized a fitness function which depended on the selected variable subsets. The fitness function was a combination of the three different hit rates described above (i.e. for all days, for the days on which the avalanche danger level changed and for the days on which the modelled danger level changed). In more detail, the GA was initiated with 100 randomly selected ten-member variable subsets. The GA then determined a maximum by continuous mutation and stored it in the final pool. This procedure was reiterated until the final pool contained 100 variable subsets, which was enough to achieve reproducible results.

For all methods variables were scaled linearly to $[-1, C - 1]$, except for categorical variables which were translated to Booleans for each category. As input for the statistical methods, subsets of the 5, 10 and 30 most important variables of each algorithm were tested.

Statistical methods

In order to find an optimal link between input variables and predicted regional avalanche danger, we used the following statistical methods:

- classification trees (TREE),
- artificial neural networks (ANN),
- nearest-neighbour methods (KNN),
- support vector machines (SVM),
- hidden Markov models (HMM).

In the following each method is described in more detail.

A simple classification tree with only one measured variable, the 3 day sum of the new snow (*HN3d_meas*), performed very well and was therefore used as a benchmark for more complex models with more (especially modelled) input variables. For this tree and more complex trees using more input variables, generalization was achieved with a pruning algorithm using ten-fold cross-validation on the training sets.

Artificial neural networks, both feed-forward and recurrent (e.g. Elman, 1990; Bishop, 2006), with different set-ups (hidden layer size, training algorithms, early stopping or Bayesian regularization) were trained for each winter. Since results depend on the initial weights, the networks were initialized five times and the mean of the results was considered. The results of ANNs discussed in the next section were gained with a recurrent network which used adaptive learning, 100 hidden neurons and 100 passes through the sequence.

The nearest-neighbour approach as used by Brabec and Meister (2001) was applied for a direct comparison to previous work. In addition, this method was modified in two ways. (1) The avalanche danger level of the previous day was used as an additional input to predict the current day. With this information the training dataset was reduced: only those days that showed the same avalanche danger level on the previous day were considered as possible nearest neighbours for the current day. This implicated an even more unbalanced classification problem due to the persistence of the target parameter. (2) The classification was modified. A decision boundary was used to determine whether the danger level should change to a new value on a particular day. Using ten nearest neighbours, three of these days must show a higher (or lower) danger level. This decision boundary was obtained by optimizing the TSS (Heierli and others, 2004). Subsequently the danger level was determined with a majority vote between the neighbours showing higher (or lower) danger levels. For breaking a tie the nearest neighbour among them was used. Weights in KNN methods allow certain variables to have more influence while calculating the distance between nearest neighbours. Optimal weights were determined with a GA (Purves and others, 2003).

SVM (Schölkopf and Smola, 2001) were applied to the problem using the 2001 LIBSVM software package of C.-C. Chang and C.-J. Lin (software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>). Gaussian radial basis functions with radius γ were used as kernel functions. Cross-validation on the training sets was performed to obtain γ and the penalty variable, C (Chen and Lin, 2006).

In addition, hidden Markov models were used, since they are able to predict time series (Rabiner, 1989; Bishop, 2006). To obtain a discrete input for the HMM, the continuous input vectors were mapped into a discrete codebook index with K-means vector quantization.

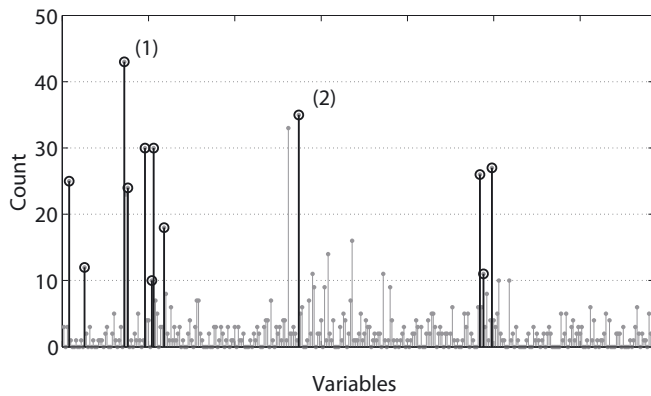


Fig. 2. Variable selection with the GA/KNN method for winter 1999/2000. Frequency (y axis) for each single variable (x axis) in the final pool. The 12 variables selected are marked; the two most important were (1) *HN* and (2) the 3 hour rate of outgoing longwave radiation.

RESULTS

Variable selection

Since the variables were selected for each forecasted winter separately, we present in this section a summary of the results obtained for each winter. In Table 1 the most important variables are listed, selected by the three methods: Fisher's discriminant analysis (Fisher); univariate classification trees (TREE); and the combination of genetic algorithm and nearest neighbour (GA/KNN) described above. In the last column the sign of correlation with the avalanche danger level is given to allow a plausibility check. Except for the variables strain rate of the weak layer, 3 hour rate of crust thickness and net longwave radiation, a physical interpretation can easily be given. Important variables were the new-snow depth (*HN*, *HN3d*) and the other new-snow variables with larger values at higher avalanche danger levels. Surface

albedo, 3 hour rate of the slab thickness and relative humidity can also be related to new-snow situations, although they are not pairwise linearly correlated. The snow transport index (Lehning and Fierz, 2008) was also selected by all three algorithms with larger index values (more drifting and blowing snow) at higher levels. Higher maximum wind speeds in the last 24 hours were also correlated with higher avalanche danger levels. The deformation index (Lehning and others, 2004) was also found to be important, despite the fact that a bug has recently been discovered in the description of the temperature dependence, although the index should still describe a relation between critical and actual stress in the bonds. As the definition suggests, small index values are correlated with high avalanche danger. The variable crust is positively correlated to the number of elapsed days without snowfall and therefore negatively correlated to avalanche danger. The modelled profile type 'Four' (Schweizer and Lütschg, 2001) was related to the danger level 'Considerable'. The profile type 'Four' is the most frequent in the dataset which describes a weak base. Although a weak base is not by itself conclusive (Schweizer and Wiesinger, 2001), it makes sense that a weak base is related to the danger level 'Considerable' as a sign of structural instability. Small bond size and low density of the weak layer might also be considered as signs of structural instability.

A result of a variable selection with the combination of GA/KNN methods can be seen in Figure 2, which shows every single variable on the x axis and their frequency in the final pool on the y axis. Variables with a high pick-frequency were interpreted as important. This method selected, amongst others, the 3 hour rate of the outgoing longwave radiation, which was highly correlated with the rate of the snow-surface temperature. Strong warming of the snow surface before midday was related to lower avalanche danger. It is worth noting that this method chose especially meteorological or snow-surface variables. This may be because the avalanche danger level of the previous

Table 1. Overview of the variable selection results

Selected variable	Measured/modelled	Selected by algorithm			Sign of correlation with target
		Fisher	TREE	GA/KNN	
24 hour new-snow sum (<i>HN</i>)	Modelled	×	×	×	+
72 hour new-snow sum (<i>HN3d</i>)	Modelled	×	×	×	+
24 hour rate of <i>HN3d</i>	Modelled	×			+
3 hour new-snow sum on day 1	Modelled	×		×	+
Deformation index	Modelled	×	×	×	-
Strain rate of weak layer	Modelled	×	×		-
Thickness of surface crust on south slopes (crust)	Modelled	×			-
3 hour rate of crust thickness	Modelled	×			+
Snow transport index	Modelled	×	×	×	+
Profile type 'Four'	Modelled	×			3*
Bond size of weak layer	Modelled		×		-
Density of weak layer	Modelled		×		-
3 hour rate of slab thickness on day 1	Modelled		×		+
3 hour rate of outgoing longwave radiation	Modelled			×	-
Net longwave radiation	Modelled			×	-
Surface albedo	Modelled			×	+
24 hour maximum wind speed	Measured			×	+
Relative humidity	Measured			×	+

*To distinguish between 'Considerable' and the other danger levels.

Table 2. Input variables used in the BRABEC model (Brabec and Meister, 2001)

Description	Weight
New-snow depth (cm)	5
Snow depth (cm)	1
Weather and intensity (code)	1
East component of wind (kt)	3
North component of wind (kt)	3
Air temperature (°C)	2
Snow temperature (°C)	2
Snow surface (code)	1
Penetration depth (cm)	2
Density of new snow (kg m ⁻³)	2
72 hour new-snow sum (cm) (<i>HN3d_meas</i>)	2
72 hour rate of air temperature (°C)	3

day was introduced as an additional input. This already provides a certain level of stability to the system so that snow-stratigraphy variables become less important. Accordingly, only weather or surface properties were important because they describe the change in danger level. To confirm this hypothesis, the GA/KNN method was performed without this additional information. However, the classification power of the KNN method without this additional information of the danger level on the previous day was too poor to reach a conclusion. Better methods were too computationally demanding for combination with the GA.

Statistical methods

The a priori capability of the statistical methods was compared with the cross-validated HR of the test dataset (verified danger levels). Classification trees had the lowest performance, while SVM, ANN, HMM and the KNN method using the avalanche danger level of the previous day as an additional input achieved HRs of ~60%. These results were comparable to the DAVOS4 model (63%) used by Schweizer and Föhn (1996) and the Kohonen neural network (61%) used by Schweizer and others (1994).

Since the distribution of the forecasted avalanche danger level of the dataset covering the years 1999–2007 (Fig. 1) was substantially different to previous studies (Schweizer and Föhn, 1996; Brabec and Meister, 2001), results were not directly comparable. Therefore the method used by Brabec and Meister (2001) with measured meteorological data as input (BRABEC) was applied to the time period of this study.

In the following, results will be given for a selection of models, which represent the range of model quality. No results for SVM are presented since they delivered no additional improvements. The selected models and their input variables are:

BRABEC (Brabec and Meister, 2001; input variables and their weights are presented in Table 2).

Classification tree with only the measured variable *HN3d_meas* as input (TREE).

Hidden Markov model (HMM) with the five best variables selected by the Fisher's discriminant analysis.

Recurrent artificial neural network (ANN) with the five best variables selected by the Fisher's discriminant analysis.

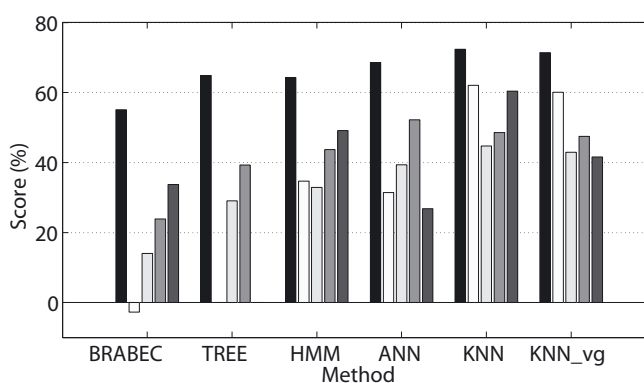


Fig. 3. Comparison of the various model performances for all days. Abbreviations are explained in the text. The black column is the cross-validated HR of all danger levels, the grey columns the cross-validated TSS value for each of the four danger levels (lightest to darkest indicate 'Low' to 'High').

Nearest-neighbour method with the 12 best variables selected by the GA/KNN combination and the avalanche danger level of the previous day as an additional input (KNN).

Same KNN method as above, but with the same input variables as used in BRABEC (i.e. no modelled variables) and the avalanche danger level of the previous day as an additional input (KNN_vg).

Figure 3 shows the performance of the selected methods considering all days. All new methods reached a higher HR than the BRABEC method (55%). Even the simple TREE method achieved a remarkable HR of 65%. However, the TREE method produced only the most frequently used danger levels 'Moderate' and 'Considerable'. Thus the TSS values of the other danger levels were 0%. For most of the nine winters the split value for 3 day sum of new snow (*HN3d_meas*) was 12 cm, which discriminated between the danger levels 'Moderate' and 'Considerable'. The variable *HN3d_meas* was also amongst the input variables for the BRABEC method (Brabec and Meister, 2001). Using prior probabilities, classification trees are, in general, also able to handle unbalanced target parameters as exist in our case (Fig. 1). Whereas this improved the TSS value for the danger levels 'Low' and 'High', it produced insufficient overall hit rates. Also more variables did not improve the results.

The other methods achieved better hit rates and better true skill scores for each single danger level. With an HR of 73%, the KNN method performed best.

Considering all days, the KNN_vg method (as KNN, but with measured, not modelled, input variables) achieved similar results to the KNN method. However, Figure 4 indicates the weakness of that method; it shows the same methods and performance measures as in Figure 3, but only for those days on which the target parameter changed. The low values in HR and TSS indicate that the KNN_vg method did not predict these important days satisfactorily. All other methods showed similar hit rates. For the KNN method it is interesting that the TSS values were greater for the higher avalanche danger levels. It seems to be easier for this model to predict an increase in avalanche danger than a decrease. The HMM and ANN methods best predicted a decrease in the avalanche danger, which can be seen by the more balanced

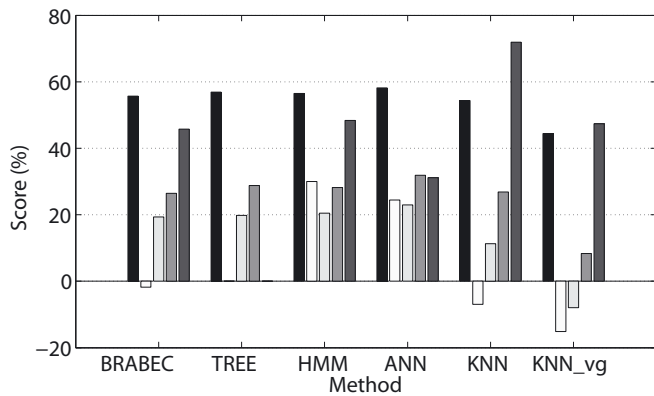


Fig. 4. Comparison of the various model performances for days on which the danger level changed. Abbreviations are explained in the text. The black column is the cross-validated HR of all danger levels, the grey columns the cross-validated TSS value for each of the four danger levels (lightest to darkest indicate 'Low' to 'High').

TSS values between each danger level. Also the simple TREE method was reasonably able to predict the days on which the target parameter changed. As the variable *HN3d_meas* is used as input, it may not be surprising that the TREE method was able to predict an increase in danger level, but a decrease was also predicted correctly in ~50% of all cases.

Since the detection of days on which the avalanche danger decreased or increased is important information, models were trained for the binary criteria 'Increase or not' and 'Decrease or not'. However, no improvement was achieved for these, at first glance, simpler classification problems compared to the methods which forecast the four avalanche danger levels. In the case of decreasing danger, one possible explanation is that it might be very difficult for human forecasters to decide consistently whether the avalanche danger level should decrease on a certain day and not on the day(s) before or after. Therefore, it might also be difficult to find reasons in the parameter space of the input variables presented to the statistical methods. In the case of increasing danger, the reasons for the change in danger level might differ between danger levels, especially between an increase from 'Low' to 'Moderate' and from 'Considerable' to 'High'. This information is lost if the models are trained only on the binary criteria.

As we saw an improvement when adding the avalanche danger level of the previous day as additional information to the KNN method, this procedure was also implemented for the other methods. Due to the persistence of the target value, this implicated a highly unbalanced classification problem. Although most methods provided a formal method to include this additional information, (e.g. prior probabilities (Breiman and others, 1998) or different penalties in the training error function), none of the methods showed an improvement.

A final test was whether the spatial variability of the input variables had additional classification power. Simulations of virtual slopes of the four main aspects and the flat field at two automated weather stations (Weissfluhjoch (2540 m.a.s.l.) and Hahnengretji (2490 m.a.s.l.)) in the test region were used as input for the statistical methods. The variability was implemented either (1) as the maximum difference between the ten simulations of each variable or (2) with additional categorical data describing the station and the aspect of the

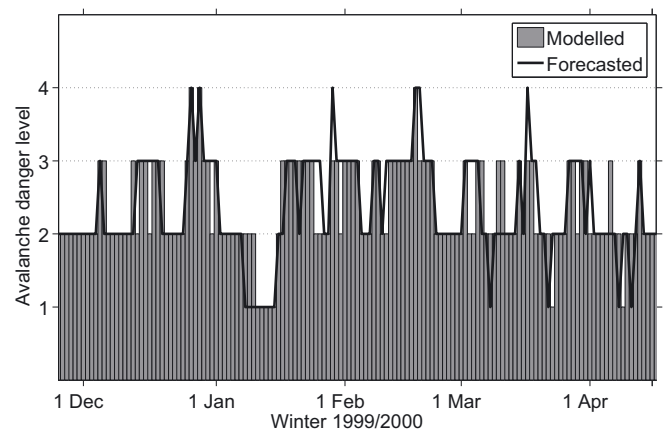


Fig. 5. The avalanche danger forecasted and modelled with the KNN method for winter 1999/2000.

simulations. Neither option improved the performance of the statistical methods.

Figure 5 shows a comparison between the KNN method and the forecasted avalanche danger for a representative winter.

DISCUSSION AND CONCLUSION

The object of this paper was to analyse whether modelled variables of the snow-cover model SNOWPACK improve the performance of forecasting models that use statistical methods, in comparison to models that are based only on measured weather data. The simple classification tree which uses only one measured (not modelled) variable (*HN3d_meas*) distinguishes well between the two most frequent danger levels of 'Moderate' and 'Considerable'. With measured meteorological input variables, as used by Brabec and Meister (2001), a more balanced model performance for all danger levels resulted, but with a strong decrease in the hit rate (HR), independent of the statistical method used. Using additional SNOWPACK variables increased the overall hit rate and produced a balanced performance for all danger levels at the same time. The best results were achieved with a nearest-neighbour method which used the avalanche danger level of the previous day as additional input. A cross-validated HR for all days of 73% was obtained. Evaluating the accuracy of the models for the days only on which the target parameter changed (e.g. the avalanche forecast changed from 'Considerable' to 'High') showed that it is only possible to obtain reasonable results with additional SNOWPACK variables. These findings suggest that for a balanced performance between all danger levels and for good overall accuracy, especially for days when the danger level changes, modelled snow-stratigraphy data, as provided by SNOWPACK, are needed.

Model performance did not improve using SNOWPACK simulations of two automated weather stations and/or virtual slope simulations. This may lead to the conclusion that the modelled snow-cover variation (though limited) was not related to regional avalanche danger. Similarly, Schweizer and others (2003), who found typical measured point stability distributions for the regional danger levels, did not find a relation between mean stability and stability variation. However, we think that considering spatial variability is

important for assessing local to regional avalanche danger. Our conclusion only concerns the situation in which we tried to include more than one meteorological station of the region in the analysis. This approach does not deal with the important local variability, which is given by variations in local energy balance (Helbig and others, in press), or preferential deposition and drifting snow (Lehning and others, 2008).

Nonetheless, a hit rate of $\sim 70\%$ reveals a remarkable discrepancy between the operational avalanche danger forecast by human experts and a statistical model. Since we have tested a wealth of methods with a huge range of complexity, we feel that the main problem is that it seems impossible to reproduce the human decision on the avalanche danger with the input used in our analysis. This missing link was recognized as a problem in the study of Schweizer and Föhn (1996), despite the fact that they used observed snow-cover variables. A study by Schweizer and others (2008) showed that a level of uncertainty exists in the detection of unstable slopes with rutschblock tests and snow profiles (probability of detection of 70%). Such snow profile interpretations are important arguments for the human prediction or verification of the regional avalanche danger level. Two possible conclusions remain: (1) additional information, which is not formalized at present, enters the decision process, such as the experience and intuition of the individual; (2) the forecasted danger level is not a good target variable, since it might be erroneous due to incorrect data at the time of the forecast or due to variations in human perception (McClung and Schaerer, 2006).

An operational prediction of the avalanche danger level with a statistical model augmented with SNOWPACK variables as additional input variables would include the following steps. First, the present snow cover is simulated with measured data; second, the development of the snow cover is predicted with forecasted meteorological data for the next day. This predicted snow cover provides the additional input variables for the statistical methods. This would be a fully automated process which could be applied to the whole area of the Swiss Alps.

In comparison to the study of Durand and others (1999), who compared the output of the French SCM chain to avalanche activity (which can be forecasted reliably without modelled snow-cover parameters; e.g. Pozdnoukhov and others, 2008), our study presented models trained and tested on the regional avalanche danger, as forecasted in the public avalanche bulletin. Using this target parameter, the models with modelled snow-cover input variables performed better than models that used mainly meteorological input variables.

Our study also showed that the uncertainty in the prediction of the avalanche danger level needs to be quantified. For example, the shift in the danger level distribution to the intermediate danger levels that obviously occurred during previous years should be clarified.

ACKNOWLEDGEMENTS

This work was partially funded by the Swiss National Science Foundation and the Swiss Federal Office of the Environment. We acknowledge the valuable comments provided by two anonymous reviewers and by the scientific editor T.H. Jacka. This study would not have been possible without the work of the avalanche warning service. We are very grateful to

T. Sauter, J. Heierli and C. Fierz for insightful comments on the ongoing work.

REFERENCES

- Bartelt, P. and M. Lehning. 2002. A physical SNOWPACK model for the Swiss avalanche warning. Part I: numerical model. *Cold Reg. Sci. Technol.*, **35**(3), 123–145.
- Bishop, C.M. 2006. *Pattern recognition and machine learning*. New York, Springer.
- Brabec, B. and R. Meister. 2001. A nearest-neighbor model for regional avalanche forecasting. *Ann. Glaciol.*, **32**, 130–134.
- Breiman, L., J. Friedman, C.J. Stone and R.A. Olshen 1998. *Classification and regression trees*. Boca Raton, FL, CRC Press.
- Buser, O. 1983. Avalanche forecast with the method of nearest neighbours: an interactive approach. *Cold Reg. Sci. Technol.*, **8**(2), 155–163.
- Chen, Y.-W. and C.-J. Lin. 2006. Combining SVMs with various feature selection strategies. In Guyon, I., M. Nikravesh, S. Gunn and L.A. Zadeh, eds. *Feature extraction: foundations and applications*. Berlin, etc., Springer, 315–328.
- Doswell, C., J. Davies and D.L. Keller. 1990. On summary measures of skill in rare event forecasting based on contingency tables. *Weather Forecast.*, **5**(4), 576–585.
- Durand, Y., G. Giraud, E. Brun, L. Mérindol and E. Martin. 1999. A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. *J. Glaciol.*, **45**(151), 469–484.
- Elman, J.L. 1990. Finding structure in time. *Cognitive Sci.*, **14**(2), 179–211.
- Guyon, I. and A. Elisseeff. 2006. An introduction to feature extraction. In Guyon, I., M. Nikravesh, S. Gunn and L.A. Zadeh, eds. *Feature extraction: foundations and applications*. Berlin, etc., Springer, 1–28.
- Harvey, S. and B. Zweifel. 2008. New trends of recreational avalanche accidents in Switzerland. In Campbell, C., S. Conger and P. Haegeli, eds. *Proceedings of the International Snow Science Workshop, 21–27 September, Whistler, British Columbia, Canada*. Whistler, B.C., 2008. International Snow Science Workshop, 900–906. CD-ROM.
- Heierli, J., R.S. Purves, A. Felber and J. Kowalski. 2004. Verification of nearest-neighbours interpretations in avalanche forecasting. *Ann. Glaciol.*, **38**, 84–88.
- Helbig, N., H. Löwe and M. Lehning. In press. Radiosity approach for the shortwave surface radiation balance in complex terrain. *J. Atmos. Sci.*
- Lehning, M. and C. Fierz. 2008. Assessment of snow transport in avalanche terrain. *Cold Reg. Sci. Technol.*, **51**(2–3), 240–252.
- Lehning, M., P. Bartelt, B. Brown, T. Russi, U. Stöckli and M. Zimmerli. 1999. SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations. *Cold Reg. Sci. Technol.*, **30**(1–3), 145–157.
- Lehning, M., P. Bartelt, B. Brown, C. Fierz and P. Satyawali. 2002a. A physical SNOWPACK model for the Swiss avalanche warning. Part II: snow microstructure. *Cold Reg. Sci. Technol.*, **35**(3), 147–167.
- Lehning, M., P. Bartelt, B. Brown and C. Fierz. 2002b. A physical SNOWPACK model for the Swiss avalanche warning. Part III: meteorological forcing, thin layer formation and evaluation. *Cold Reg. Sci. Technol.*, **35**(3), 169–184.
- Lehning, M., C. Fierz, B. Brown and B. Jamieson. 2004. Modeling instability for the snow-cover model SNOWPACK. *Ann. Glaciol.*, **38**, 331–338.
- Lehning, M., H. Löwe, M. Ryser and N. Radeschall. 2008. Inhomogeneous precipitation distribution and snow transport in steep terrain. *Water Resour. Res.*, **44**(W7), W07404. (10.1029/2007WR006545.)
- Li, L., C.R. Weinberg, T.A. Darden and L.G. Pedersen. 2001. Gene selection for sample classification based on gene expression

- data: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics*, **17**(12), 1131–1142.
- McClung, D. and P. Schaerer 2006. *The avalanche handbook. Third edition*. Seattle, WA, The Mountaineers.
- Meister, R. 1995. Country-wide avalanche warning in Switzerland. In *Proceedings of the International Snow Science Workshop, 30 October–3 November 1994, Snowbird, Utah, USA*. Snowbird, UT, International Snow Science Workshop, 58–71.
- Murphy, A.H. 1991. Forecast verification: its complexity and dimensionality. *Mon. Weather Rev.*, **119**(7), 1590–1601.
- Pozdnoukhov, A., R.S. Purves and M. Kanevski. 2008. Applying machine learning methods to avalanche forecasting. *Ann. Glaciol.*, **49**, 107–113.
- Purves, R.S., K.W. Morrison, G. Moss and D.S.B. Wright. 2003. Nearest neighbours for avalanche forecasting in Scotland: development, verification and optimisation of a model. *Cold Reg. Sci. Technol.*, **37**(3), 343–355.
- Rabiner, L.R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, **77**(2), 257–286.
- Schölkopf, B. and A.J. Smola. 2001. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Cambridge, MA, MIT Press.
- Schweizer, J. and P.M.B. Föhn. 1996. Avalanche forecasting – an expert system approach. *J. Glaciol.*, **42**(141), 318–332.
- Schweizer, J. and J.B. Jamieson. 2003. Snowpack properties for snow profile analysis. *Cold Reg. Sci. Technol.*, **37**(3), 233–241.
- Schweizer, J. and M. Lütsch. 2001. Characteristics of human-triggered avalanches. *Cold Reg. Sci. Technol.*, **33**(2–3), 147–162.
- Schweizer, J. and T. Wiesinger. 2001. Snow profile interpretation for stability evaluation. *Cold Reg. Sci. Technol.*, **33**(2–3), 179–188.
- Schweizer, J., K. Kronholm and T. Wiesinger. 2003. Verification of regional snowpack stability and avalanche danger. *Cold Reg. Sci. Technol.*, **37**(3), 277–288.
- Schweizer, J., S. Bellaire, C. Fierz, M. Lehning and C. Pielmeier. 2006. Evaluating and improving the stability predictions of the snow cover model SNOWPACK. *Cold Reg. Sci. Technol.*, **46**(1), 52–59.
- Schweizer, J., I. McCammon and J.B. Jamieson. 2008. Snowpack observations and fracture concepts for skier-triggering of dry-snow slab avalanches. *Cold Reg. Sci. Technol.*, **51**(2–3), 112–121.
- Schweizer, M., P.M.B. Föhn and J. Schweizer. 1994. Integrating neural networks and rule-based systems to build an avalanche forecasting system. In Hamza, M.H., ed. *Proceedings of the IASTED International Conference on Artificial Intelligence, Expert Systems and Neuronal Networks, 4–6 July 1994, Zürich, Switzerland*. Zürich, International Association of Science and Technology for Development; Calgary, AB, Acta Press, 1–4.
- Wilks, D.S. 1995. *Statistical methods in the atmospheric sciences*. San Diego, CA, Academic Press.

MS received 15 January 2009 and accepted in revised form 13 May 2009