Analysis, Modeling and Spatio-Temporal Prediction of Avalanches Using Support Vector Machines

Matasci Giona
Institute of Geomatics and Analysis of Risk, University of Lausanne
Lausanne, VD, 1015, Switzerland
giona.matasci@unil.ch

Kanevski Mikhail
Institute of Geomatics and Analysis of Risk, University of Lausanne
Lausanne, VD, 1015, Switzerland
mikhail.kanevski@unil.ch

Pozdnoukhov Alexei
National Centre for Geocomputation, National University of Ireland Maynooth
Maynooth, Co. Kildare, Ireland
alexei.pozdnoukhov@nuim.ie

1 Introduction

Avalanche forecasting is a crucial task in many winter skiing and mountaineering venues. Snow avalanches are linked to meteorological factors, snowpack conditions and terrain features with a rather complex non-linear relationship [6]. Experts are in the field on a daily basis to collect data, understand the processes leading to the release of avalanches and provide reports assessing the danger under the current conditions. In their activity they can be assisted by numerical methods, based either on physical principles or on statistical approaches. The latter set of techniques aims at predicting a target output related to the avalanche activity, generally a binary indicator of the occurrence of the events, based on several explanatory input variables.

The supervised classification technique of *Nearest Neighbours* (NN) is widely used in the field, mainly due to its ease of interpretation [1, 10]. The method produces forecasts by comparing the current situation with past recordings of the conditions measured on the slopes and the associated avalanche activity. However, such an approach usually only considers the temporal component: the spatialization of the forecasted danger is carried out by the specialists with the identification of critical regions according to particular elevations and aspects highlighted by the model.

Spatial approaches to the analysis of snow avalanches have seldom been taken within the research community. For instance, using a hierarchical Bayesian methodology, the study of the occurrence of the events at the level of the municipality has been tackled in [3].

On the other hand, the constant refinement of the data collection process with more efficient monitoring networks and the availability of high-resolution models of the topography opens promising opportunities for studies integrating the spatial component of the phenomenon.

In this context, this research aims at developing a spatio-temporal model able to efficiently discriminate safe from the active avalanche paths under a given set of conditions summarized by several meteorological and snowpack factors. The spatially varying component of the avalanche activity has been introduced by computing a set of localised features related to the likelihood of avalanches at the level of the single paths.

We applied *Support Vector Machines* (SVM) [11], a powerful machine learning tool designed for classification, which adheres to the guidelines provided by the Statistical Learning Theory [12]. The algorithm builds a robust and non-parametric classifier able to efficiently cope with highly non-linear problems involving datasets of large dimensions such as those encountered in the field of statistical avalanche forecasting.

Promising results have been obtained in terms of predictions of path activity for 2 test winter seasons and mapping of the indicator of the forecast avalanche danger over the whole extent of a mountain range [8]. Additionally, we present here a feature selection approach to identify in a completely automated fashion the features contributing the most to the correct assessment of the likelihood of hazardous events. The routine we tested, named *Recursive Feature Elimination* (RFE) [5], has proved a promising ability in suggesting a ranking of the key factors influencing the avalanche releases.

Data used in the present case study concern 18 years of observations in the region of Lochaber, Scotland, UK.

2 Support Vector Machines

2.1 Linear SVM

Given a training dataset of n points $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i \in \Re^d$ are the d-dimensional input vectors and $y \in \{+1, -1\}$ are the binary labels, SVM seeks to find a hyperplane separating the 2 classes with a maximal margin (distance between the two closest training samples of the opposite classes). The equation of this hyperplane, defined by vector $\mathbf{w}$ and bias constant b, becomes the decision function used to classify a new test data point $\mathbf{x}$ (class membership assigned according to the sign of $f(\mathbf{x})$):

(1) $$f(\mathbf{x}) = \mathbf{w}\mathbf{x} + b\,.$$

In order to maximize the margin $\rho = \frac{2}{\|\mathbf{w}\|}$, parameters $\{\mathbf{w}, b\}$ are optimized by solving a *quadratic programming* problem. The solution of the optimization problem allows the final linear SVM decision function to be formulated as

(2) $$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i \mathbf{x}\mathbf{x}_i + b\,,$$

where the α_i's are the Lagrange multipliers associated with each training sample. Only a subset of them receives a non-zero α_i weight. The latter are called *Support Vectors* and are the only data points contributing to the SVM solution. These examples define the margins of the hyperplane (regions where $f(\mathbf{x}) = \{+1, -1\}$) and are the closest samples to the decision boundary.

2.2 Non-linear soft margin SVM

In order to deal with overlapping classes and noisy data, a “soft” margin adaptation has been introduced. The misclassification of the training samples is allowed and a hyper-parameter C (to be tuned by the user) is added in the optimization as a trade-off between margin maximization and magnitude of the training errors.

Since the decision function (2) is uniquely based on dot products between samples, the introduction of the so-called “kernel trick” can be used to yield non-linear classification models. The dot product $\mathbf{x} \cdot \mathbf{x}_i$ is substituted by a kernel function $K(\mathbf{x}, \mathbf{x}_i)$, which is a symmetric semi-positive definite similarity measure between two vectors. Valid kernels correspond to a dot product in a *Reproducing Kernel Hilbert Space*. The final SVM decision function is therefore provided by

(3) $$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b\,.$$

In the present work we made use of Gaussian RBFs kernel functions, $K(\mathbf{x}, \mathbf{x}_i) = e^{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma^2}}$, which lead to a simple interpretation of Eq. (3): it is a weighted sum (by coefficients α_i and similarities with the test sample $\mathbf{x}$) of the labels y_i associated with the training samples $\mathbf{x}_i$. In fact, kernel values K decrease as the squared Euclidean distance between the examples $\|\mathbf{x} - \mathbf{x}_i\|^2$ increases. The σ parameter of the kernel, the bandwidth of the Gaussian function, has to be tuned by the user.

2.3 SVM output interpretation

Since the range of values taken by the SVM decision function $f(\mathbf{x})$ is not fixed, it is useful to compute a sigmoid transformation to obtain class-conditional posterior probabilities $p\,(y = 1\,|\mathbf{x}\,) \in [0, 1]$ as follows [7]:

$$(4) \qquad p\,(y = 1\,|\mathbf{x}\,) = \frac{1}{(1 + \exp{(a \cdot f(\mathbf{x}) + b)})},$$

where a and b are constants tuned by maximum likelihood on the validation set. These indicative values, however, in the present case are to be used for exploratory data analysis and regarded as a general likelihood of an avalanche release and not as event probabilities.

Once the SVM decision function is rescaled in such an interval, it is possible to tune a SVM threshold $t \in [0, 1]$. This threshold allows us to let predictions be either more liberal ($t < 0.5$) or more conservative ($t > 0.5$) with respect to the standard threshold on the raw decision function $f(\mathbf{x}) = 0$.

Both in the validation (parameters selection) and in the testing (generalization ability assessment) phases, the performance of the model can be evaluated by computing a series of skill scores starting from the binary confusion matrix associated with the considered forecasting task [2]. Table 1 summarizes the statistics frequently used in the field of avalanche forecasting.

Table 1: ***Main forecast verification measures and skill scores [2].***

Measure	Formula	Range and description
Overall accuracy (OA)	$\frac{\text{Hits} + \text{Corr.Negatives}}{\text{Total}}$	$[0, 1]$, the proportion of correct forecasts
Hanssen and Kuipers discriminant (HK)	Hit rate – FalseAlarm rate	$[-1, 1]$, the capacity to discriminate between events and non-events
Heidke skill score (HSS)	$\frac{\text{Hits} + \text{Corr.Negatives} - \text{Chance}}{\text{Total} - \text{Chance}}$	$]-\infty, 1]$, the fraction of correct predictions accounting for correct outcomes due to random guess (Chance)

3 Statistical spatio-temporal avalanche forecasting

3.1 Avalanche data from the Lochaber region

The case study that will be illustrated in the next sections deals with the combination of spatial and temporal information into an avalanche forecasting model based on SVM for the Lochaber region, Scotland [9, 8]. The goal is to produce spatially varying avalanche forecasts at the level of single avalanche paths existing in this mountaineering area. The region includes Ben Nevis (the highest mountain in the UK) and is one of the 5 ski venues in Scotland for which forecasting is carried out on a daily basis during the winter season (see `www.sais.gov.uk`). In the considered area, a nearest neighbor model called *Cornice* is operationally used to assist in forecasting the avalanche days (temporal forecasts only) [10].

Current weather and snowpack conditions are described by a set of 9 variables that are measured or estimated by local forecasters on the slopes or that are registered by an automatic weather station: "snow index" (ordinal index of the precipitation as fresh snow on a day), "rain at 900 m" (binary variable indicating rain at 900 m), "snow drift" (binary variable indicating when experts remark snow drifting), "air temperature" (measured in °C), "wind speed" (24 hours vector mean speed in m/s), "wind direction" (24 hours vector mean wind direction in °), "cloud cover" (as percentage of the sky), "foot penetration" (in the snow, measured in cm at a pit site) and "snow temperature" (measured in °C at a depth of 10 cm at a pit site).

The 9 variables are recorded every day of the winter season (roughly 4 months per year), along with the avalanche events observed in the region. Each release point is described by elevation, slope and aspect. Data from 1991 to 2008 were available for this study: information about 712 avalanche events occurred in 49 different avalanche paths (gullies) was used.

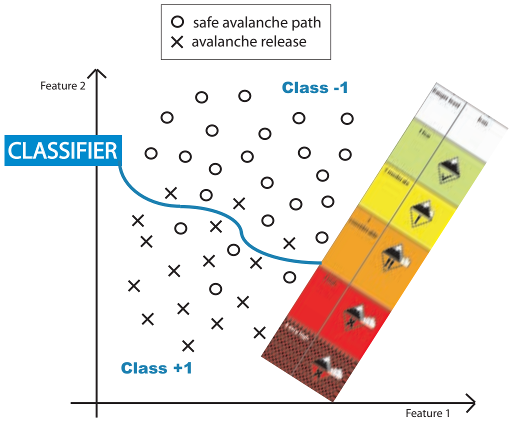

Figure 1: ***An illustration of spatio-temporal avalanche forecasting considered as a supervised classification problem. The European Danger Scale ("1" - low (green) to "5" - very high (red and black)) is shown to align the distribution of avalanching events in a space of conditioning factors and the regional avalanche danger levels.***

3.2 Input features

In order to get the desired spatio-temporal forecast at the level of the avalanche path, the series of daily measurements of meteorological conditions related to snowpack stability described in Sect. 3.1 have to be smartly combined with the spatial description of the terrain morphology available via the DEM. The latter, with its relatively high resolution of 10 meters, provides detailed information about "elevation", "slope" and "aspect" of the paths where the avalanche events could happen (3 initial variables included in the set). A group of spatialized local features with changing values according to the location of the avalanche release point was computed. Examples of these spatially varying features are, for instance, "air temperature", "wind direction", "wind speed", "snow accumulation", etc. The conditioning factors have been interpolated over the whole domain using very simple heuristics (temperature/elevation gradients, basic physical models, etc.). Additionally, global factors describing the general avalanching likelihood over the Lochaber region as a whole were also added to the set. Variables as "snow index", "foot penetration" in the snow cover, etc. were appended. For many of these variables (both local and global), information about conditions recorded in the previous days were also included (2 pre-days at most because of the rapidly changing weather conditions).

The final input vector counted 39 features: 22 spatio-temporal features (describing local conditions at a given gully) and 17 temporal features with global validity (constant values). The complete list, with a brief description of the meaning of each variable is available in [8].

3.3 Set up of the classification problem

From a statistical avalanche forecasting perspective, one can now tackle the prediction of the events as a binary supervised classification problem. In the case of spatio-temporal forecasts, our goal is to discriminate between safe avalanche paths and avalanche paths giving rise to releases under given conditions (Fig. 1). This is an effective exploratory approach to reveal patterns in historical data available for avalanche forecasting.

Initially, we assign to the positive class (+1 label) the vectors characterizing (spatial location, weather and snowpack conditions) the observed avalanche events. To complete the binary classification problem, a negative class (−1 label) is needed as well. The chosen intuitive approach is to let this class be composed by all the 49 gullies (actually the 40 covered by DEM information) susceptible to give rise to an avalanche release on a safe day. Therefore, for every day of the winter season when all the variables listed in Sect. 3.1 could be measured and the visibility allowed avalanche observations but no event was actually documented in the region, we computed all the features describing the conditions at each avalanche path. The purpose was to let the classifier train on a set of critical situations which were close to the "safe/event" decision boundary and likely to cross it under slightly different weather conditions. Additionally, in order to avoid the resulting unbalanced classification task (negative samples outnumbering the positive ones by a factor of ≈ 134 to 1), we decided to include virtual positive examples in the dataset. These virtual positive class data points were generated by drawing 25 additional samples from normal distributions centred on the actual feature values of each real avalanche case. Standard deviations were set to account for instrumental error or observation uncertainties. Moreover, such a procedure is justified by the fact that many more avalanches than those observed and reported actually occur in the considered domain. In fact, avalanches are either randomly spotted by the forecasters on their daily patrolling activity or reported by skiers/mountaineers.

The final dataset consisted of 18'512 positive samples and of 95'115 negative samples spanning the winter seasons from 1991 to 2008. In order to evaluate the generalization ability of the SVM model on an independent subset, the mentioned dataset was split in a training set covering the years from 1991 to 2005 (15'756 samples with label +1 and 89'335 samples with label −1) and in a test set related to seasons from 2006 to 2008 (2'756 samples with label +1 and 5'780 samples with label −1). Note that among the test set positive examples we find 106 real avalanche cases.

4 Avalanche forecasts assessment in space and time

4.1 Release predictions at single avalanche paths

Once the datasets have been built, the SVM needs to be optimized to provide an efficient predictive model able to generalize properly on new unseen data while avoiding the overfitting of the training set. The parameters of the classifier were tuned on a randomly chosen validation subset consisting of 20% of the original training set. For this task we made use of the HK discriminant, HSS and OA measures (see Tab. 1). The SVM hyper-parameters space (C and σ) has been thoroughly searched and the following optimal values were found: $C = 0.5$ and $\sigma = 10$. After having applied the sigmoid transform of Eq. (4), performed with tuned parameters $a = -1.2$ and $b = -0.23$, the SVM threshold t was set to a level of 0.37.

The SVM model was then asked to predict the binary labels for the test dataset, as though we were simulating the forecasts for 2 coming winter seasons (years from 2006 to 2008). The results, as shown by Tab. 2(a), are encouraging. The SVM model correctly predicted 97 of the 106 events observed during these test seasons. The number of false alarms (1150) was quite large but was deemed unavoidable in order to ensure a reasonable number of detections. Forecasting skills are summarized by the following scores: OA = 0.84, HK = 0.72 and HSS = 0.66.

As a matter of comparison, the performance of the proposed model has been related to that of the benchmark method widely used in (temporal) avalanche forecasting: a Nearest Neighbours model. After optimization, such a model was applied by forecasting an avalanche release if 2 or more neighbours out of 20 were avalanche events. The detection rate, as illustrated by Tab. 2(b), is lower than the SVM model, with only 55 correctly predicted observed events. The performance measures are less noteworthy: OA = 0.75, HK = 0.37 and HSS = 0.39.

Table 2: ***Performance of the SVM model compared to a Nearest Neighbours model. Confusion matrices refer to the test set (in brackets: real events).***

(a) SVM model.

		predicted	
		Class +1	Class −1
observed	Class +1	2541 (**97**)	215 (**9**)
	Class −1	1150	4630

(b) Nearest Neighbours model.

		predicted	
		Class +1	Class −1
observed	Class +1	1415 (**55**)	1341 (**51**)
	Class −1	827	4953

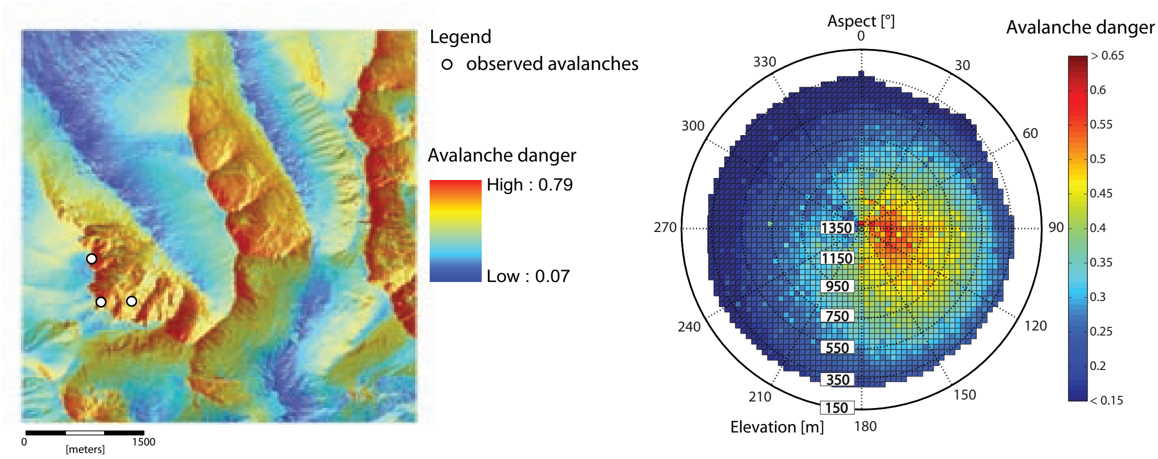

Figure 2: ***Map (left) and corresponding aspect/elevation diagram (right) of the avalanche danger for 15 February 2007.***

4.2 Avalanche danger mapping

Danger maps represent a valuable tool when assessing the threat of natural hazards such as avalanches over a mountain range. In order to compute an "avalanche danger" map over the Lochaber region under a given set of conditions, a prediction grid for the days of interest has to be built. Hence, we extrapolated the avalanching conditions summarized by the 39 input features mentioned in Sect. 3.2 over the whole extent covered by the DEM, obtaining a prediction grid with 10 m × 10 m pixels. Using the retained SVM model, it is possible to compute a decision function value, rescaled by means of the optimal sigmoid, for every point of the map. This can be intended as an indicator of the likelihood of an avalanche release. Its range being always comprised between 0 and 1, such a measure allows a comparison of different maps produced under various conditions.

Another method of displaying the spatial variability of the avalanche danger is by plotting aspect/elevation diagrams. This is a common graphical representation of the hazard in operational avalanche forecasting. For every combination of aspect/elevation values, we plotted the largest avalanche danger found among the cells of the map having these characteristics.

We chose to produce maps and diagrams for some interesting days in the winter of 2007 (independent test set). Figure 2 provides a visualization of the predicted avalanche danger for 15 February 2007, a day where 3 avalanche events have actually been observed on the Ben Nevis sector. The map over the 5 km × 5 km region of Lochaber shows the critical regions where avalanches are likely to be released. As in the previous days quite important snowfalls have occurred and south-westerly winds were blowing, the east-facing aspects are correctly highlighted as the most dangerous (good correspondence with the events reported that day).

The aspect/elevation diagram draws the attention of the user on the same critical regions, confirming thus the main trend noticed on the map. Indeed, the "danger rose" as well suggests that the east aspects above 1000 m present a large risk of avalanching (danger indicator above 0.5).

5 Feature selection via Recursive Feature Elimination

5.1 Method and experimental setup

Feature selection methods provide a classifier with a small subset of variables selected from the initial set so that it can work in a lower dimensional input space with the relevant features only. The application of this kind of algorithm provides the analyst with meaningful information about the real influence and utility of each input feature used in the classification problem.

In the case of avalanche forecasting, it is worth looking for the most useful variables contributing to the prediction at the avalanche paths of our study region. Such a task is usually carried out manually by an avalanche expert. Nevertheless, an objective way to assess variables influences in a forecasting model is advisable since, as pointed out by Purves *et al.* in [10], when dealing with high-dimensional datasets and with long periods of recordings, expert suggestions tend to be too reliant on the recent winters behaviors (special conditions, particular recorded events, etc.). To overcome this drawback the authors use *genetic algorithms* to find the best set of weights for their NN model. Working with SVM allowed us, on the contrary, to make use of a technique that is well suited for the choice of the relevant features: Recursive Feature Elimination [5]. Many methods have been proposed to select the best features or to reduce the input space dimensionality [4]. Hereafter we will briefly present the RFE procedure for a SVM classifier using a kernel expansion.

The intuition behind this procedure consists in identifying, at every step of the algorithm after having trained the SVM, the least influential feature and removing it from the initial set of d features. The process is iterated until the set of features is empty, providing thus a variables ranking accounting for their usefulness.

The importance of every feature in the SVM model, in the linear case, is given by the weighting vector $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$. When dealing with a non-linear SVM it is impossible to directly compute the components of $\mathbf{w}$ because the sample $\mathbf{x}_i$, included in a simple dot product in Eq. (2), turns into the input of the kernel function in Eq. (3). Therefore, the method consists of looking for the smallest change in the square of the length of vector $\mathbf{w}$ when removing feature k. This value, identified with $W(\boldsymbol{\alpha})^2$, is not computed directly as the norm of $\mathbf{w}$, but as

$$W(\boldsymbol{\alpha})^2 = \|\mathbf{w}\|^2 = \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\alpha}^T \mathbf{H} \boldsymbol{\alpha}\,, \tag{5}$$

where the α's (forming column vector $\boldsymbol{\alpha}$) are the weights for each training point found after the optimization task, $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel output reporting the similarity between the training samples x_i and x_j and $\mathbf{H}$ is a matrix consisting of elements $y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$.

As proposed by Guyon *et al.*, at each iteration, the feature g to withdraw according to the final ranking criterion is selected as

$$g = \arg\min_k \left| W(\boldsymbol{\alpha})^2 - W_{(-k)}(\boldsymbol{\alpha})^2 \right| \,, \tag{6}$$

where the notation $(-k)$ denotes that the candidate feature k has not been included in the computation of (5). Since the norm of the weighing vector $\mathbf{w}$ defines the SVM margin ρ, we select the variable whose removal least changes the distance between the strict class boundaries $f(\mathbf{x}) = \{-1, +1\}$.

The main steps of the non-linear RFE procedure are summarized by Algorithm 1.

We run a series of 10 experiments initialized with a different subset of 8000 training samples randomly chosen from the complete set of available training points (see Sect. 3.3). This strategy has been adopted for computational issues (RFE is very demanding if run with a large training set) and in order to account for the effect of the choice of the samples when building the datasets. We adopted a

Algorithm 1 Non-linear RFE

Inputs: Training samples with known class labels $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, $\mathbf{x}_i \in \Re^d$.
repeat
    train the SVM and compute $W(\boldsymbol{\alpha})^2$ according to Eq. (5)
    assess the performance of the SVM on the test set and store the result
    for $k = 1, d$ **do**
        temporarily remove variable k from the set of features
        compute $W_{(-k)}(\boldsymbol{\alpha})^2$
        obtain and store the ranking criterion for feature k as $\left|W(\boldsymbol{\alpha})^2 - W_{(-k)}(\boldsymbol{\alpha})^2\right|$
    end for
    find the feature g to remove according to Eq. (6)
    remove the values of this feature from the initial training data
    update the ranking list
until every feature has been removed from the initial set
Output: ranked features list (first removed → less relevant), SVM performance in test at each iteration.

SVM with Gaussian kernel whose parameters (C and σ and threshold t) were tuned at the beginning of each experiment, in this case, by 5-fold cross-validation using HK as performance score.

5.2 Results and discussion

The application of the RFE procedure yields two main types of results. The first one is a classification accuracy curve created by plotting the selected model success statistic (HK) in test as a function of the iterations (features removal) (Fig. 3). Such a plot informs the user about the actual performance of the model when dealing with the decreasing number of variables to rely on. The second graphical element is a boxplot of the rank (importance) obtained by the 39 features in the 10 experiments (Fig. 4). The expert is provided here with an objective ranking of the features translating the influence of each one of these in the SVM model.

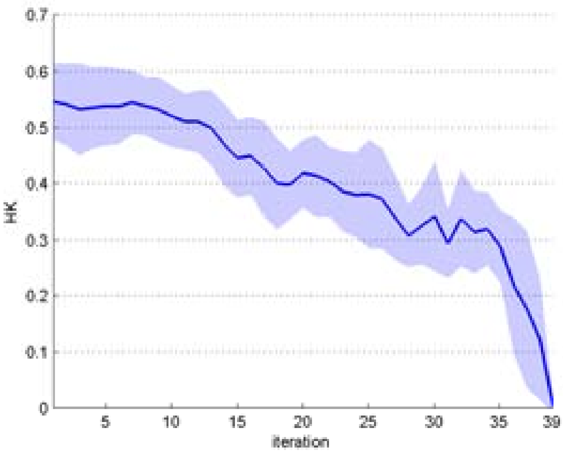

Figure 3: ***Hanssen and Kuipers discriminant along the RFE iterations for the 10 experiments. Mean values are shown with a solid blue line while the shaded surface displays the standard deviation.***

Looking at the average curve of the HK measure one notices that the magnitude of the measure decreases as the features removal procedure progresses. This confirms the intuitive thought that there is a loss of information if we provide the model with only a few features describing the avalanche paths. However, one also notices that this decrease is not at all smooth. By removing up to 7 well chosen features, we are able to keep an average classification of good quality. In fact, HK starts a significant decrease only when the kept subset counts less than 32 variables. The removal of noisy features bringing into play corrupted information about avalanche activity is therefore completely justified. We then have a gradual decrease (HK from 0.54 to 0.32) in the classification quality from iteration 7 to iteration 34, associated with the withdrawal of more important factors affecting the release of avalanches. Finally, when removing the last 5 key features (from iteration 35 onwards) the discrimination between events and non-events suddenly becomes harder (HK < 0.29), revealing the extreme usefulness of the variables ranked as most relevant by RFE.

It is now insightful to have look at the ranking of the features and check if it is in agreement

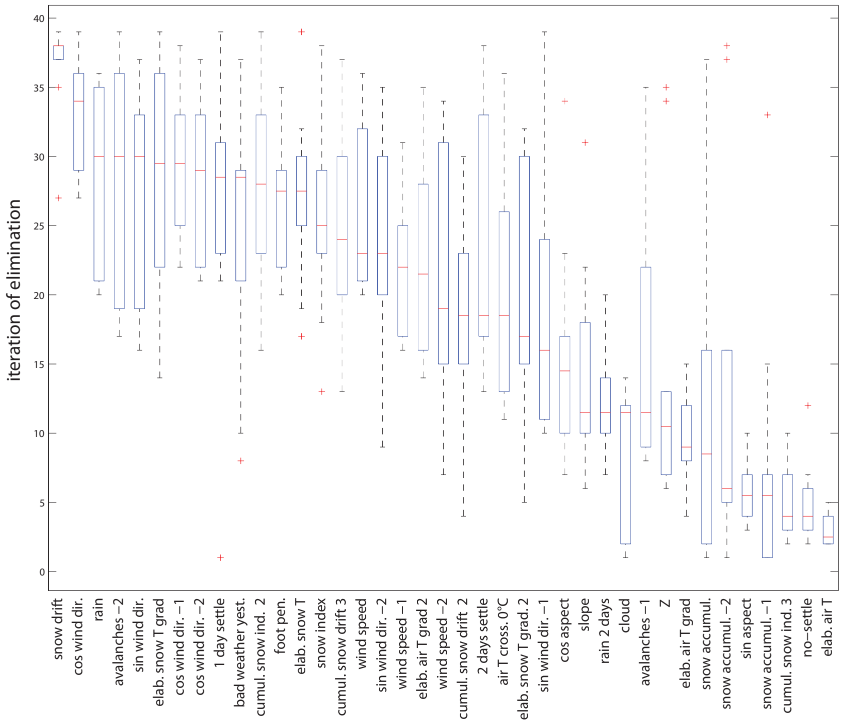

Figure 4: ***Boxplots of feature rankings over the 10 experiments, sorted by decreasing median rank values. The rank (ordinate) corresponds to the iteration at which the feature is eliminated in the RFE procedure, i.e. the larger the rank, the more relevant the feature (variable kept until the end of the process).***

with the interpretation of the influences that the involved variables have on the process affecting an avalanche release.

We remark as the global binary indicator "snow-drift" is detected as the most useful variable (last removed variable on average), mainly because of its benefit in discriminating the critical days from the safe days in terms of avalanche activity. Among the 5 key variables we then find the cosine (2^{nd} rank) and sine (5^{th} rank) transforms of the "wind direction". These projections of the wind direction on the north-south and west-east axes are of decisive importance for correctly predicting the slope aspects where critical snow deposits are likely to take place. In 3^{rd} and 4^{th} position we respectively have the "rain" and "avalanches -2" indicators. Both are global variables. The former reports whether rain is falling, resulting in an additional snowpack load. The latter accounts for a sort of temporal memory of the model by recording the avalanche activity that has been observed over the 2 previous days (possible values $\in \{0, 1, 2\}$).

Among the less important features ranked by the algorithm we find many variables whose information is made available to the system through the features created by combining the pre-days scores or several initial variables. The least relevant variable, "elab. air T" (removed in 5 experiments out of 10 as 2^{nd} feature, *i.e* median rank of 2.5), for instance, contains the information about the current temperature of the air. Such values have already been integrated in the "elab. air T grad." and "elab. air T grad. 2" variables encoding the changes in temperature between current and previous days. The RFE algorithm rejects this variable since it brings in redundant information.

6 Conclusions

The present work illustrated how Support Vector Machines have been applied for the spatio-temporal prediction of avalanches based on historical data. The task has been considered as a binary supervised classification problem where the goal was to discriminate the avalanche paths subject to a snow release from the safe paths. Meteorological as well as snowpack information was combined with the DEM to build a meaningful set of features describing the local avalanching conditions.

SVM proved potential in terms of predictions of avalanche activity at the level of the paths for 2 test winter seasons at the avalanche-prone site of Lochaber. The benchmark technique of NN was outperformed in this study involving both spatial and temporal components of the avalanche danger.

Moreover, we proposed an approach aimed at the extrapolation of the predictions over the entire considered region, leading to maps of the avalanche danger. Such plots, along with the associated aspect/elevation diagrams, can be considered as powerful decision support tools for the avalanche forecasters dealing with the assessment of this natural hazard.

Finally, the selection of relevant features via the Recursive Feature Elimination technique provided us with a fully automatic method able to identify the most influencing factors in the considered prediction task. Helpful insights about which phenomena affect the most the avalanche activity in space and time are thus available for the experts of the field as a complement to their knowledge.

Acknowledgements - The research is funded in part by the Swiss National Science Foundation projects "GeoKernels: kernel-based methods for geo- and environmental sciences (Phase II)" (No 200020-121835/1). A. Pozdnoukhov acknowledges the support of Science Foundation Ireland under the National Development Plan, particularly through Stokes Lectureship Award and Strategic Research Cluster grant (07/SRC/I1168). Graham Moss and the sportscotland Avalanche Information Service are thanked for their assistance in identifying expert features and the provision of data for Lochaber. Dr. Ross Purves (University of Zurich) is acknowledged for the constructive discussions and collaborations we had while working on this case study.

References

[1] *O. Buser,* Two years experience of operational avalanche forecasting using nearest neighbors method, *Annals of Glaciology* ***13*** *(1989), 31–34.*

[2] *C. A. Doswell and D. L. Keller R. Davies-Jones,* On summary measures of skill in rare event forecasting based on contingency tables, *Weather and Forecasting* ***5*** *(1990), 576–585.*

[3] *N. Eckert, E. Parent, L. Bélanger, and S. Garcia,* Hierarchical Bayesian modelling for spatial analysis of the number of avalanche occurrences at the scale of the township, *Cold Regions Science and Technology* ***50*** *(2007), no. 1-3, 97–112, Snow and Avalanches EGU 2006.*

[4] *I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh,* Feature Extraction: Foundations and Applications, *Springer, 2006.*

[5] *I. Guyon, J. Weston, S. Barnhill, and V. Vapnik,* Gene Selection for Cancer Classification using Support Vector Machines, *Machine Learning* ***46*** *(2002), 389–422.*

[6] *D. M. McClung and P. A. Schaerer,* The avalanche handbook, *The Mountaineers (1993).*

[7] *J. Platt,* Probabilistic outputs for Support Vector Machines and comparisons to regularized likelihood methods, *Advances in Large Margin Classifiers, MIT Press, 1999, pp. 61–74.*

[8] *A. Pozdnoukhov, G. Matasci, M. Kanevski, and R. S. Purves,* Spatio-temporal avalanche forecasting with Support Vector Machines, *Natural Hazards and Earth System Science* ***11*** *(2011), no. 2, 367–382.*

[9] *A. Pozdnoukhov, R. S. Purves, and M. Kanevski,* Applying machine learning methods to avalanche forecasting, *Annals of Glaciology* ***49*** *(2008), 107–113.*

[10] *R. S. Purves, K. W. Morrison, G. Moss, and D. S. B. Wright,* Nearest neighbours for avalanche forecasting in Scotland - development, verification and optimisation of a model, *Cold regions science and technology* ***37*** *(2003), 343–355.*

[11] *B. Schölkopf and A. J. Smola,* Support Vector Machines, Regularization, Optimization and Beyond, *The MIT Press, 2002.*

[12] *V. Vapnik,* Statistical Learning Theory, *Wiley, 1998.*