Semi-Supervised Support Vector Machine for Natural Hazards Forecasting. Case Study: Snow Avalanches

A. Pozdnukhov

R. S. Purves

M. Kanevski

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Semi-Supervised Support Vector Machine for Natural Hazards Forecasting. Case Study: Snow Avalanches

A. Pozdnoukhov[a], **R. S. Purves**[b], **M. Kanevski**[a]

[a] *Institute of Geomatics and Analysis of Risk, University of Lausanne, Switzerland (Alexei.Pozdnoukhov@unil.ch)*

[b] *Department of Geography, University of Zurich, Switzerland*

Abstract: This paper explores the use of Support Vector Machine (SVM) as a predictive engine for natural hazards forecasting. It particularly discusses the issues of incorporating this classification method into a decision-support system for operational use in avalanche forecasting. The recent developments concerned with semi-supervised and transductive SVM-based learning targeted at applications in natural hazards forecasting on geomanifolds are presented. The real case study on spatio-temporal avalanche forecasting deals with the development of a predictive engine for the decision support system used at the avalanche-prone site of Ben Nevis, Lochaber region in Scotland.

Keywords: environmental data mining, support vector machine, avalanche forecasting, semi-supervised and transductive learning.

1. INTRODUCTION

Amongst different natural hazards the events like snow avalanches are of particular interest. These events can be characterized by relatively low frequency, complex non-linear relationships with meteorological conditions, geomorphology and a large variety of other factors including human activity on the site. In terms of data-driven modeling, the avalanche forecasting can be considered as a classification problem, where one needs to find a decision boundary in the feature space of factors which discriminate the "safe" and "dangerous" conditions.

In this paper we explore the use of Support Vector Machine (SVM), a method from the field of Machine Learning, as a predictive engine for natural hazard forecasting. We discuss the issues of incorporating the developed model into a decision-support system for operational use in avalanche forecasting, and present the recent achievements. The real case study on the application of SVM is devoted to temporal and spatio-temporal avalanche forecasting at the avalanche-prone site of Ben Nevis, Lochaber region in Scotland, where avalanche forecasts are produced daily in winter months.

The paper is organized as follows. First, in the next section, we introduce the data-driven classification as an approach to decision support. We present there the basic features of a particular machine learning classification method, SVM, including the probabilistic interpretation of its outputs. Next, in Section 3, we motivate the use of semi-supervised and transductive learning in environmental data-driven modelling and describe the related contemporary approaches. We finally review the recent results on the application of SVMs for decision support in avalanche forecasting and provide the preliminary results on the use of semi-supervised and transductive SVM learning (Section 4). The paper is summarized with directions to the further developments and the conclusions in Section 5.

2. BINARY CLASSIFICATION AND DECISION SUPPORT

A wide range of numerical models and tools have been developed over the last decades to support the decision making process in environmental applications ranging from physical models, through expert systems, to a variety of statistically-based methods. In operational forecasting a mixture of all three approaches are often used, with process chains involving physical models and statistical or expert systems being relatively common.

As model complexity has increased, so to has our ability to collect real time, spatially distributed data describing a wide range of parameters through technological advances in sensor networks and automated environmental monitoring, and one can thus expect data-driven models to become increasingly important. Binary classification problems (the task to find a decision rule to discriminate the data into two classes such as “dangerous” and “safe” based on available empirical data), are widely met in environmental decision support. Interestingly, this target-oriented approach to decision support (direct inference from data to binary decisions without considering the intermediate modelling steps which complicate the model and bring uncertainty) is justified by the Occam Razor principle.

Below we present one of the most powerful data-driven classifiers, the Support Vector Machine, and discuss its use in decision support including an important issue of the interpretation of the data-driven forecasts produced by SVM.

2. 1 Support Vector Machine

SVM is a machine learning approach derived from Statistical Learning Theory aimed to deal with data of high dimensionality by approaching the nonlinear problems in a robust and non-parametric way. An interested reader is kindly asked to refer to some of the profound introductions to the theory of SVMs and related algorithms [Vapnik, 1998], [Scholkopf and Smola, 2002]. Here we only mention the main principles of SVMs which will find important applications in their applications in decision-oriented forecasting.

Suppose we deal with the linearly separable data (x_1, y_1), … (x_N, y_N), where x are the input features and $y \in \{+1, -1\}$ are the binary labels. By “linearly separable” we mean data that can be discriminated into two classes by a hyperplane. The idea of SVM is to separate this dataset by finding the hyperplane that is, roughly speaking, the farthest apart from the closest training points. The minimal distance between the hyperplane and the training points is called the margin, which is maximized by the SVM algorithm (Figure 1).

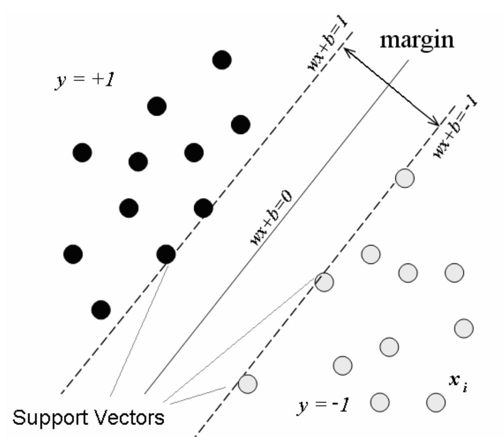

Figure 1. Margin maximization principle: the basic idea of Support Vector Machine.

It is proven in scope of Statistical Learning Theory that the maximum margin principle prevents over-fitting in high-dimensional input spaces, thus leads to good generalization abilities.

The decision function used to classify the data is a linear one, as follows:

$$f(x, w) = w \cdot x + b\,, \qquad (1)$$

where coefficient vector w and threshold constant b are optimized in order to maximize the margin. This is a quadratic optimization problem with linear constraints which has unique solution. Moreover, w is a linear combination of the training samples, many of them having zero weights α_i:

$$w = \sum_{i=1}^{N} y_i \alpha_i x_i \ . \qquad (2)$$

The samples with non-zero weights are the only ones which contribute to this maximum margin solution. They are the closest samples to the decision boundary and called Support Vectors.

To make this classifier non-linear, the so-called kernel trick is used. Kernel is a symmetric semi-positive definite function K(x,x'). According to the Mercer theorem, this implies that it corresponds to a dot product in some space (Reproducing Kernel Hilbert Space, RKHS). Generally, given a (linear) algorithm, which includes data samples in the form of dot products only, one can obtain a (non-linear) kernel version of it by substituting the dot products with kernel functions. This is the case for linear SVM, where the decision function (1) relies on the dot products between samples, as clearly seen by substituting (2) into (1). The final classification model is a kernel expansion:

$$f(x,\alpha) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \qquad (3)$$

The choice of the kernel function is an open research issue. Using some typical kernels like Gaussian RBF, one takes into account some knowledge like distance-based similarity of the samples. The parameters of the kernel are the hyper-parameters of SVM and have to tuned using cross-validation or a testing dataset.

2.2 Probabilistic Post-processing

Though the SVM is specifically constructed to solve the classification task, that is, to *discriminate the binary events*, the outputs of SVMs can be probabilistically interpreted by post-processing. To introduce a characterization of uncertainty, the values of the decision function (1) or (3) can be transformed into the smooth confidence measure, $0<p(y=1|x)<1$. This is done, for example, through taking a sigmoid transformation of $f(x,\alpha)$ [Platt, 1999]:

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

where a and b are constants. These constants are tuned using a maximum likelihood (usually, the negative log-likelihood to simplify the optimization) on the testing dataset. The value of a is negative, and if b is found to be close to zero, then the default SVM decision threshold $f(x)=0$ coincides with a confidence threshold level of 0.5.

The major advantage of the latter interpretation is a possibility to introduce a decision threshold for the smooth confidence outputs $p(y=1|x)$. This threshold may later be tuned to satisfy the desired forecast quality measures.

2.3 Interpretation for Decision Support

The output of the binary classification system can be characterized by several basic measures, which are shown in the Table 1, known as contingency table. Concerning natural hazards, different possible forms and interpretations of the forecast are usually considered. Firstly in *categorical forecasts* a decision boundary is directly used to classify the region\time as being either dangerous or not. Secondly, in *probabilistic forecasts* the output of the system has to be interpreted as the probability of an event in the temporal or spatio-temporal domain of the forecast. Such forecasts can be used, for example, for risk assessment. Thirdly, a so-called *descriptive forecast* is often desirable, since experts wish to interpret and incorporate, for instance, a detailed list of similar events into their decision-making process. Concerning the last category, the Nearest Neighbour methods and their variations commonly named as the "methods of analogues" are extensively used in a number of applications, with their probable roots in early atmospheric predictions [Lorenz, 1969].

Table 1. Basic measures for binary forecasts (contingency table or confusion matrix).

	Forecast	
Observed	Yes (+1)	No (-1)
Yes (+1)	Hits	Misses
No (-1)	False Alarms	Correct Negative

Support Vector Machines are well-suited to produce the abovementioned forms of the forecasts [Pozdnoukhov et.al., 2008]. The categorical forecast is just the predicted class, then the probabilistic interpretation can be used for predicting the event probability. The descriptive forecasts can be produced by providing the corresponding Support Vectors (which are the most valuable discriminative events in the past), though it still needs to be properly verified in a dialogue with a forecaster.

The Table 2 below provides some conventional forecast quality measures which can be used to tune the optimal hyper-parameters of SVM and the decision threshold.

Table 2. Forecast verification measures [Wilks, 1995].

Forecast accuracy measures	
POD - Probability of detection	The probability that the event was forecast when it occurred.
	POD = Hits/(Hits+Misses)
SR - Success rate	The probability that the event occurred when it was forecast.
	SR = Hits/(Hits+False Alarms).
HR - Hit rate	The proportion of correct forecasts.
	HR = (Hits+Correct Negative)/(Total Number of Days)

3. SEMI-SUPERVISED AND TRANSDUCTIVE LEARING

The problem of using unlabeled data is of increasing attention in Machine Learning. By unlabeled data, we mean those data samples which consist of the input values only, while the desired output value is unknown. The methods making use jointly of labelled and unlabeled data are called semi-supervised. When predictions have to be made to given unlabeled locations only, this particular situation is called transductive learning. Most real-life learning problems are actually semi-supervised, which gives rise to the developments of large-scale semi-supervised methods nowadays. For example, the semi-supervised setting of the problem for forecasting natural hazards such as snow avalanches is illustrated in Figure 2. The available data often consists of historical observations of avalanche activity (the list of events registered at given location under given meteorological conditions), the locations where avalanches were not observed under these conditions and the locations where the inputs (location and meteorological situation) are known but no information on avalanche activity is available. The ratio between the amount of available labelled and unlabelled data will always be in favour of the last.

The information one obtains from the unlabeled part of the dataset can be of different nature. A common approach is to consider the *manifold assumption*. This implies that data actually belong to some lower dimensional manifold in high dimensional input space. A large body of literature is devoted to the exploration of such an approach; see [Belkin, 2003], [Chapelle et.al., 2006] and references therein. Another use of unlabelled data is the so-called *cluster assumption*, which implies that the data are structured (clustered) in the input space. This structure can not be observed given the limited amount of labelled data, though the large amount of unlabelled samples would help exploring it. Concerning SVMs, the use of unlabelled data is illustrated in Figure 3. The methods developed to exploit the cluster assumption include Trasductive SVMs and Low Density Separation methods. The overview on semi-supervised learning methods and particularly the approaches to Semi-Supervised SVMs (S^3VMs) can be found in [Chapelle et.al., 2006].

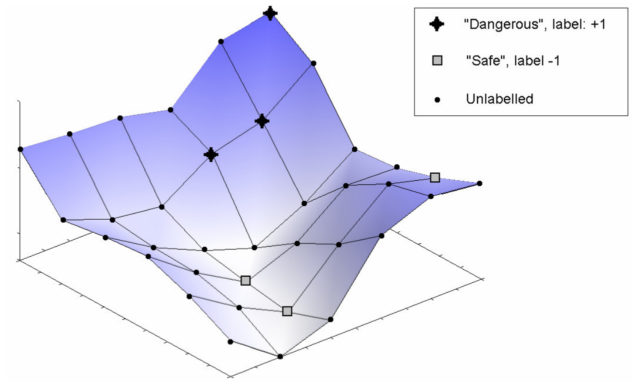

Figure 2. An example of typical natural hazards data. Snow avalanches on the digital elevation model of the terrain: the dangerous (avalanches were observed in the past under similar conditions), certainly safe (expert knowledge or no avalanches observed in the past under similar conditions) and unlabelled samples (those where input features are known but the outcome is unknown or/and has to be predicted).

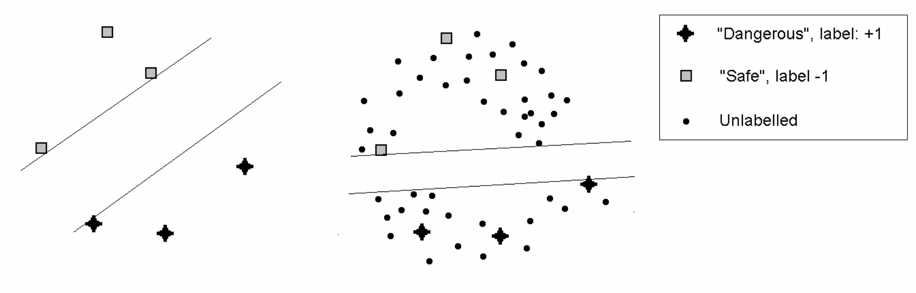

Figure 3. Unlabelled data can precise the data classification.

Semi-supervised learning methods become particularly useful in approaching environmental modelling in data-driven manner. With the finite set of available measurements or observations, the amount of related information (Digital Elevation Models and remote sensing images) one may consider is almost unlimited. It should be noted that the manifold assumption often can be considered to be satisfied while the likelihood of cluster assumption is less evident. The first one can be successively used both in classification and regression problems, while the second one is specific to classification. For both approaches, the amount of unlabelled data to model a manifold or to use with the methods relying on cluster assumption is almost exhaustive. We briefly introduce the model and the implementation of the S^3VMs which we use in the case study below.

3.1 Transductive Support Vector Machine

Several attempts were reported to implement the idea presented in Figure 3. The general approach is to formulate the margin maximization problem including the penalty given if the unlabelled sample appears to be inside the margin. This constraint is similar to the constraint applied to the labelled samples in standard formulation of the SVM. However, the optimization problem complicates significantly. In the experiments below we have used the approach of [Collobert et.al., 2006]. It formulates the optimization problem using the ramp loss as a Constrained Concave-Convex Programming (CCCP), and has two major advantages, first of them being reasonably fast computational speed. The second one is that the noisy and mislabelled data samples do not become the support vectors of the model as it is the case in standard SVM formulation. This model increases the number of hyper-parameters to tune, including the width of the ramp loss and the trade-off cost constant which specifies the sensitivity of the model to misfits on unlabelled data samples. Despite of these practical difficulties, this implementation is used in the experiments below.

4. CASE STUDY: SNOW AVALANCHES

Concerning snow avalanche forecasting, different approaches have being proposed and being used in operational practice. To name some, these are the interpretations of the physical models of the development of the snowpack [Durand et al., 1999], expert systems which attempt to integrate expert knowledge [Schweizer and Föhn, 1996], and Nearest Neighbor methods, [Purves et.al., 2003]. Nearest Neighbour methods accord well with conventional inductive avalanche forecasting processes and are thus relatively popular with forecasters. In machine learning this is a relatively simple pattern classification technique. Moreover, both through theoretical considerations and in forecasting practice [McCollister et.al., 2003] it has been noted that such methods may be prone to over-fitting when dealing with highly-dimensional data.

4.1 Avalanche Forecasting in Lochaber Region

Avalanche forecasts are produced daily in the Lochaber region of Scotland, and the Nearest Neighbour based system is currently used there for decision support [Purves et.al., 2003]. The original data on avalanche activity in the region consist of daily measurements of 10 meteorological and snowpack variables starting from the winter seasons of year 1991. There is a database for this period with 712 registered avalanche events. These were happening at the particular 49 avalanche paths located at mainly north-east oriented slopes and gullies. The Digital Elevation Model of the region is available (Figure 4.1).

By using these data, the binary classification problem was formulated. The input feature vector contains several spatial features (coordinates of the events, elevation, slope, aspect, and convexity of the path) and about 30 temporal features, including meteorological observations for the previous 3 days for each event and derived “expert features”, constructed in a dialogue with a local avalanche forecaster who was asked to list important indicators of avalanche activity. The next step in identifying suitable features used recursive feature elimination in conjunction with a SVM to filter redundant features. This feature selection method iteratively omits the variables with the smallest influence on the decision surface of the SVM classifier. The final feature vector included a total of 26 variables and was used to produce the temporal forecasts with SVM [Pozdnoukhov et. al., 2008].

4.2 Spatially Variable Forecasts

An important step in producing the spatially variable forecasts consists in the spatialization of the meteorological data over the forecasting region. It can be done using physical models, heuristics, or data-driven approaches and is a matter of profound independent research not considered in this paper.

While it was relatively straightforward to put the registered avalanche events into a dataset as a class representing avalanche events, it is much harder to describe the “safe” conditions. Here lies an important issue - the samples of the “safe” class have to be both discriminative and have a proper label. In other words, to include a sample of the “safe” class one has to be sure that the snowpack at the given slope is stable under given conditions, while still representing a “non-trivial” data sample (as the slope with no inclination or without any snow at all). Here the unlabelled samples come into play - if no one is completely sure about the stability of the slope (no avalanches were actually observed due to bad weather conditions but the avalanche activity was suspected), we suggest including the sample as the one with known inputs but no output - the unlabelled one. As for the days with good visibility when no avalanche events were observed and no new snowfalls registered, the “safe” samples were constructed by combining the spatial features of all the potential avalanche paths and the current meteorological features. The intuition behind this was to provide the boundary and most discriminative samples to the system: those which are “safe” but still closest to turning into dangerous ones given changing weather conditions. It resulted in 712 positively labelled samples, about 30000 negatively labelled and about 12000 unlabelled samples. The problem is thus very unbalanced. In supervised SVM, it is usually approached by modifying the misclassification costs of the different classes [Lin et al., 2002].

To approach this problem we have first considered the performance of the fully supervised SVM classifier with a Gaussian RBF kernel was applied to the labelled part of the data. It has produced 600 and 1400 Support Vectors for the "dangerous" and "safe" classes correspondingly. It means that most of the generated "safe" samples are lying far from the decision surface and do not contribute to the discrimination. Thus, the problem is reasonably balanced and the modification of the costs of labelled samples can be avoided. Moreover, the semi-supervised implementation of the SVM foresees the cost given to *unlabelled* samples as a specific parameter. It was tuned with cross-validation resulting in a lower value of C^{unlab}=0.1 considered to the labelled samples misclassification cost, which was found to be C^{lab}=3.

Note that in the semi-supervised setting the model selection is a non-trivial task. Concerning the fully supervised models, the cross-validation is usually applied. However, the choice of unlabelled sample acts as a user-defined parameter of the semi-supervised algorithm, and this issue is not yet properly explored.

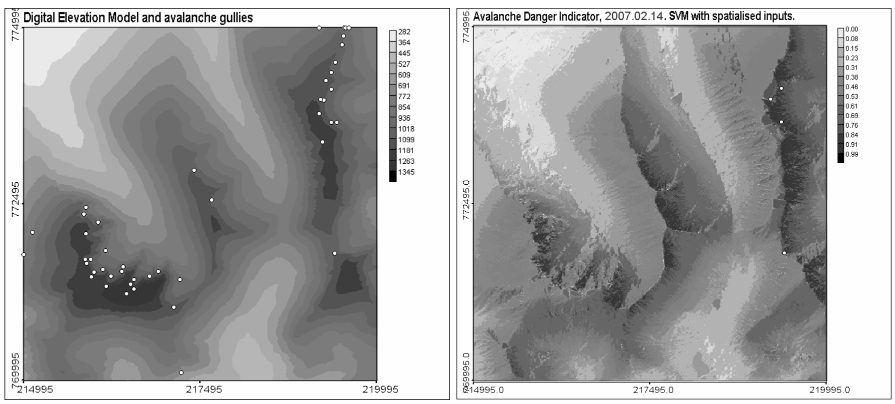

Figure 4. Left: Digital Elevation Model of the region and the locations of avalanche gullies (marked with circles). Right: The output of the semi-supervised SVM. The actual observed validation events are shown with circles.

An example of the produced forecast is presented in Figure 4, right. The performance curves obtained on the validation data (the data of the 2006-2007 seasons were reserved for the latter) are presented in Figure 5. While the overall behaviour of the systems is different, one can notice just slight improvement in the performance obtained with the use of unlabelled data. More efforts need to be done to properly validate the system.

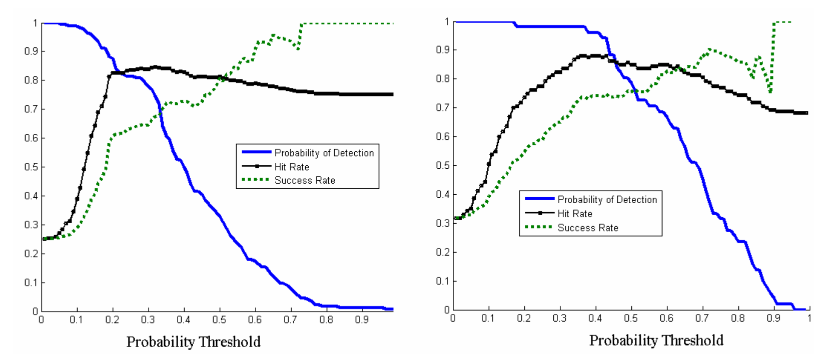

Figure 5. Performance curves of the baseline (left) and semi-supervised (right) systems, obtained on validation data for years 2004-2007. The hit rate of the semi-supervised system is just slightly higher.

5. DISCUSSION AND CONCLUSIONS

The recent developments in semi-supervised and transductive machine learning find exciting applications in natural hazards forecasting. The uncertain nature of these phenomena and availability of information make it possible to formulate the problem as a data-driven prediction based on labelled and unlabelled data. Particularly, in this paper we have illustrated the application of a semi-supervised Support Vector Machine to the spatio-temporal forecasting of the snow avalanche activity in Lochaber region of Scotland.

Bringing such systems into operational use for real-life decision support is a difficult undertaking, though the temporal forecasting with nearest neighbour models is an already accepted practice. Concerning the spatially varying forecasting, the output of physics-based models, such as the Safran-Crocus-Mepra tool [Durand et al., 1999] appears to be useful. Though, with the growing amounts of available data at finer spatial resolutions the data-driven modelling may be more advantageous in regional avalanche forecasting, at least due to the lower computational costs. The aim of the presented research was to make the first steps in this direction by considering whether the latest achievements from machine learning are suited to provide a predictive engine for this problem.

The focus of the further work will be the development of the proper validation schemes in order to investigate whether the semi-supervised approaches can produce useful improvements to the spatio-temporal forecasting of natural hazards. Then, the descriptive forecasts, highly required in an operational use, will be approached by the exploration of the set of support vectors responsible for the predictions.

ACKNOWLEDGEMENTS

This work was partially supported by Swiss National Science Foundation through the project N°200021-113944, “GeoKernels: kernel-based methods for geo and environmental sciences”.

REFERENCES

Belkin, M. (2003). Problems of Learning on Manifolds. Ph.D. thesis, Chicago, 2003.

Chapelle O., Scholkopf B., Zien A., (Eds.) Semi-Supervised Learning. MIT Press, 2006.

Collobert R., Sinz F., J. Weston, and L. Bottou. Large scale transductive SVMs. Journal of Machine Learning Research, 7:1687–1712, 2006.

Durand, Y., Giraud, G., Brun, E., Me´rindol, L., Martin, E., 1999. A computer-based system simulating swowpack structures as a tool for regional avalanche forecast. J. Glaciol. 45 (151), 466–484.

Lin Y., Lee Y., and Wahba G. Support vector machines for classification in non-standard situations. Machine Learning, 46:191–202, 2002.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues, Journal of the Atmospheric Sciences, 26, 636-646.

McCollister C., Birkeland K., Hansen K., Aspinall R. and Comey R. 2003. Exploring multi-scale spatial patterns in historical avalanche data, Jackson Hole Mountain Resort, Wyoming. Cold Regions Science and Technology, 37, 3, 299-313.

Platt J., Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In Advances in Large Margin Classiers, MIT Press, 1999.

Pozdnoukhov A., Purves R., and Kanevski M. Applying Machine Learning Methods to Avalanche Forecasting, to appear in Annals of Glaciology, 2008.

Purves R.S., K.W. Morrison, G. Moss and D.S.B. Wright. 2003. Nearest neighbours for avalanche forecasting in Scotland – development, verification and optimization of a model. Cold Regions Science and Technology, 37, 343-355.

Scholkopf, B. and Smola, A. 2002., Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge, MA.

Schweizer, J. and Föhn, P. M. B. 1996. Avalanche forecasting – an expert system approach. Journal of Glaciology, 42, 141, 318-332.

Vapnik V. N., Statistical Learning Theory. John Wiley & Sons, Inc., New York, 1998.

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