

Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping

Natan Micheletti · Loris Foresti · Sylvain Robert · Michael Leuenberger · Andrea Pedrazzini · Michel Jaboyedoff · Mikhail Kanevski

Received: 4 March 2013 / Accepted: 23 November 2013 / Published online: 20 December 2013
© International Association for Mathematical Geosciences 2013

Abstract This paper explores the use of adaptive support vector machines, random forests and AdaBoost for landslide susceptibility mapping in three separated regions of Canton Vaud, Switzerland, based on a set of geological, hydrological and morphological features. The feature selection properties of the three algorithms are studied to analyze the relevance of features in controlling the spatial distribution of landslides. The elimination of irrelevant features gives simpler, lower dimensional models while keeping the classification performance high. An object-based sampling procedure is considered to reduce the spatial autocorrelation of data and to estimate more reliably generalization skills when applying the model to predict the occurrence of new unknown landslides. The accuracy of the models, the relevance of features and the quality of landslide susceptibility maps were found to be high in the regions characterized by shallow landslides and low in the ones with deep-seated landslides. Despite providing similar skill, random forests and AdaBoost were found to be more efficient in performing feature selection than adaptive support vector machines. The results of this study reveal the strengths of the classification algorithms, but evidence: (1) the need for relying on more than one method for the identification of relevant variables;

N. Micheletti (✉) · M. Leuenberger · M. Kanevski
Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment,
University of Lausanne, 1015 Lausanne, Switzerland
e-mail: natan.micheletti@unil.ch

L. Foresti
Royal Meteorological Institute of Belgium, 1180 Brussels, Belgium

S. Robert
Seminar for Statistics, Swiss Federal Institute of Technology, 8092 Zürich, Switzerland

A. Pedrazzini · M. Jaboyedoff
Institute of Earth Sciences, Faculty of Geosciences and Environment,
University of Lausanne, 1015 Lausanne, Switzerland

(2) the weakness of the adaptive scaling algorithm when used with landslide data; and (3) the lack of additional features which characterize the spatial distribution of deep-seated landslides.

Keywords Landslide susceptibility mapping · Support vector machines (SVM) · Adaptive scaling SVM · Random forests · AdaBoost · Multiscale terrain features · Object-based validation

1 Introduction

Territorial planning and natural hazards assessment make extensive use of landslide susceptibility (LS) maps. Geotechnical, physical and statistical approaches can be designed to predict the occurrence of slope failures (Montgomery and Dietrich 1994; Soeters and van Westen 1996; van Westen et al. 2005, 2008). However, geotechnical methods are more suited to study specific events and are often applied to single slopes (Daia and Lee 2002; Dietrich et al. 1995; Terlien et al. 1995). For regional-scale LS mapping, machine learning and statistical parametric and non-parametric classification methods are more appropriate due to their flexibility in incorporating the collected observational data (Brenning 2005; Guzzetti et al. 1999, 2006).

Statistical LS mapping deals with the description of the topographic, hydrological, geological and land cover conditions which correlate with the spatial distribution of historically known landslides. In practice, a digital elevation model (DEM) is used to extract continuous features to account for the topographical-hydrological context of hillslope, while geological maps are used to extract a set of categorical features representing the main geological facies. In a second phase, a statistical classification model, for example logistic regression (LR) (Atkinson and Massari 1998; Ayalew and Yamagishi 2005; Ohlmacher and Davis 2003), discriminant analysis (Adrizzone et al. 2002; Carrara 1983) or generalized additive models (Goetz et al. 2011; Park and Chi 2008), is applied to discriminate landslide from non-landslide pixels/objects in the multidimensional space of topographic, geological and land use features. Due to the complexity of topographic and geological conditions associated with landslide occurrence, more flexible nonlinear methods such as machine learning algorithms have also been considered, in particular artificial neural networks (Ermini et al. 2005; Lee et al. 2004; Melchiorre et al. 2008; Neaupane and Achet 2004; Yilmaz 2010b), support vector machines (SVM) (Ballabio and Sterlacchini 2012; Yao et al. 2008; Yilmaz 2010a), Gaussian processes (Gallus 2010) and random forests (RF) (Liess et al. 2011; Stumpf and Kerle 2011). Great effort has been placed upon comparing these different approaches both for LS mapping (Brenning 2005) and in a broader context for geomorphological and landform mapping (Brenning 2009; Marmion et al. 2008, 2009; Nefeslioglu et al. 2008; Pradhan and Lee 2010; Yesilnacar and Topal 2005). Once trained, the classifiers are employed to predict the membership to the landslide class over the region of interest. The membership function can be further post-processed to compute class probabilities or categorized into variable hazard levels to obtain maps relating to the landslidesusceptibility.

The prominence of linear over nonlinear methods for LS mapping is mostly due to their simplicity and easy interpretability. In fact, logistic regression and discriminant analysis enable a direct analysis of the relevance of input topographic and geological predictors by looking at the size of the regression coefficients (Lee 1974; Ohlmacher and Davis 2003) or the components of the weight vector representing the hyper-plane separating the two classes (Carrara 1983; Fisher 1936). When working with nonlinear methods, the task is more complicated as it is not possible to obtain the weights representing the separating hyper-plane. The linear hyper-plane only exists in an implicit (the reproducing kernel Hilbert space for SVM, Schölkopf and Smola 2002) or a more explicit embedding space (the hidden layer representation in artificial neural networks, Haykin 1999). In these particular configurations, the direct analysis of the relevance of input features is not achievable. To overcome the inability of directly extracting the weights of input features, several alternatives were proposed in the machine learning community, known as feature selection (Guyon and Elisseeff 2003; Guyon et al. 2006). Feature selection deals with the analysis of the relevance of features when used to solve a particular prediction problem. Simple strategies involve the repeated application of a prediction algorithm to test several subsets of features and select the one providing maximum accuracy (Guyon and Elisseeff 2003; Guyon et al. 2006). Other methods, such as the mean decrease in accuracy in the random forest algorithm (Breiman 2001), exploit the intrinsic properties of the algorithms and allow obtaining both a prediction model and the information about the relevance of features (Foresti et al. 2011; Grandvalet and Canu 2003).

The aim of this paper is to apply three modern machine learning methods, that is support vector machines, random forests and AdaBoost (AB), both for LS mapping and for the detection of relevant features in Canton Vaud, Switzerland. The analysis is not limited on scarp zones, but focuses on the landslide affected area. Three sub-regions of Canton Vaud are independently analyzed to account for the different landslide types and geological context. Feature selection is applied to eliminate the irrelevant variables and to obtain simpler regional scale LS maps. Particular attention is given to the extraction of morphological features at multiple spatial scales to account for the characteristic size of landslides, as proposed by Kalbermatten et al. (2011). Similarly to Brenning (2005), Foresti et al. (2012), and Muchoney and Strahler (2002), an object-based sampling procedure was implemented to reduce the high spatial autocorrelation characterizing the datasets used for the analysis. This approach provides better conditions for training the machine learning models and allows obtaining more reliable estimations of the generalization skill, that is the performance in predicting the occurrence of new unknown landslides. The paper gives a comprehensive description of the machine learning methods used by focusing on their main distinctive characteristics and key free parameters. This understanding is fundamental for an optimal training of the models and successive interpretations of the obtained results. Moreover, the cross-comparison of methods highlights their respective strengths and allows drawing more robust conclusions about the relevance of features. Section 2 presents SVM, RF and AB. Section 3 describes the data and the experimental setup. Section 4 illustrates the results of feature selection and LS mapping. The interpretation of results is discussed in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Review of Machine Learning Algorithms

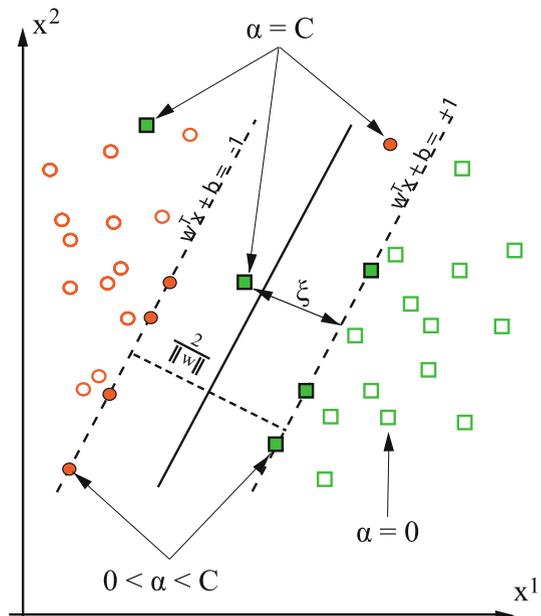
In the present study, machine learning algorithms are used to perform a supervised classification for landslide susceptibility analysis, that is to model the unknown dependence between a number of potential landslide conditioning factors (input features) and the presence or absence of landsliding (binary output). This section presents the three machine learning algorithms used for LS mapping: adaptive support vector machines, random forests and AdaBoost.

2.1 Support Vector Machines

Support vector machines are non-parametric kernel-based techniques derived from the statistical learning theory (Vapnik 1998). They are particularly appealing for solving nonlinear high-dimensional classification, regression and density estimation problems in a robust way by controlling model complexity (Cherkassky and Mulier 2007; Moguerza and Munoz 2006; Schölkopf and Smola 2002).

Consider a set of L observations $\{(\mathbf{x}_i, y_i)\}_{i=1}^L$, where $\mathbf{x}_i \in \mathbb{R}^D$ is the input D -dimensional vector of features and $y_i \in \{+1, -1\}$ is the labeled output. To ensure high generalization skills, SVM retrieves the hyperplane $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$, which maximizes the margin between the two classes (Fig. 1), which is formulated as an optimization problem. The optimization can be relaxed by allowing some misclassification and using a hyper-parameter C controlling the complexity of the model (balance between margin maximization and classification error). SVM is solved in its dual formulation as a convex quadratic programming problem (Cherkassky and Mulier 2007; Moguerza and Munoz 2006; Vapnik 1998)

Fig. 1 Principles of support vector machines. The margin is the distance between the two dashed lines. Support vectors are defined by non-zero Lagrange multipliers α_i and drawn with filled symbols. The parameter C controls the total amount of accepted misclassification (outliers), which is the sum of slack variables ξ over the training samples



$$\left\{ \begin{array}{l} \max_{\alpha} \quad \sum_{i=1}^L \alpha_i - \frac{1}{2} \sum_{j=1}^L \sum_{i=1}^L \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} \quad \sum_{i=1}^L \alpha_i y_i = 0 \\ \quad \quad \quad 0 \leq \alpha_i \leq C \quad \forall i \end{array} \right. , \tag{1}$$

where α_i are Lagrange multipliers non-zero only for support vectors (the most discriminative observations lying on the margin) and $K(\mathbf{x}_i, \mathbf{x}_j)$ is a kernel function used to map the input data into a higher dimensional space where linear separability applies. One of the most used kernels in environmental sciences (Kanevski et al. 2009) is the Gaussian radial basis function, which enables controlling the smoothness of the SVM solution by varying the width σ of the kernel. The simple linear SVM is obtained by replacing the kernel with a dot product in Eq. (1). Finally, the class label of a new incoming sample \mathbf{x} is evaluated by its position relative to the hyperplane, thus by applying the sign to the decision function (Kanevski et al. 2009; Vapnik 1998)

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

where b is an offset (derived from the support vectors). A probabilistic interpretation, that is providing a landslide membership probability between 0 and 1, can be derived by applying a logistic transformation to re-scale the decision function (Platt 1999). This transformation is optimized by maximum likelihood using the algorithm of Lin et al. (2007).

The direct evaluation of the weights of features is not possible with the isotropic Gaussian kernel. The use of an anisotropic kernel would require the optimization of several parameters, which is a nonlinear and computationally hard problem. A solution resides in the adaptation of the scaling properties of the input space before applying the isotropic Gaussian kernel. Grandvalet and Canu (2003) proposed transforming the input space by applying a linear transformation to each input feature as follows

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{S} = \begin{pmatrix} x_1^1 & \dots & x_1^D \\ \vdots & \ddots & \vdots \\ x_L^1 & \dots & x_L^D \end{pmatrix} \cdot \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_D \end{pmatrix}, \tag{3}$$

where \mathbf{X} is the $L \times D$ matrix of input data and $\mathbf{S} = \text{diag}(s)$ is the $D \times D$ matrix of scaling factors. In this configuration, the classic SVM dual problem in Eq. (1) is modified with an additional constraint over scaling factors to favour sparse solutions, where only a small number of input features contributes to the solution (Grandvalet and Canu 2003). The optimization procedure of scaling factors is based on a gradient descent and is explained in detail in Grandvalet and Canu (2003). The iterative gradient descent is carried out for a given complexity parameter C and scale parameter s_0 (which replaces the σ of the classic SVM). These two hyper-parameters are optimized by a grid search using a calibration dataset. The classic SVM is trained with the library

LibSVM (Chang and Lin 2001), while the adaptive scaling SVM is available in the SVM and kernel methods Matlab toolbox (Canu et al. 2005).

2.2 Random Forests

Random forests (Breiman 2001) are very powerful and flexible ensemble classifiers based upon decision trees. Each individual tree is constructed on a bootstrapped sample of the data using classification and regression trees methodology (CART), with a random subset of variables selected at each node. The size of the variables subset m is usually much smaller than the total number of variables in the input space. The different trees are decorrelated due to the random selection of features at each node, which boosts the power of the ensemble classifier by reducing its variance without harming the bias (Breiman 2001). Each node partitions the data according to the maximal linear separability. The final decision of class membership (output) is based on the majority vote among all decision trees.

RFs have many interesting features such as accurate prediction power (comparable to or even outperforming ANN and SVM), low tendency to overfitting, low computational cost and the ability to work with very high-dimensional data (Caruana and Niculescu-Mizil 2006). In addition, because each tree in the forest is trained on a bootstrapped sample of the data, not all observations appear in each tree. The prediction performance achieved on the samples which are left out (called out-of-bag) can be used to analyze the relevance of features. Importance factors of variables, in this case the mean decrease in accuracy (Breiman 2001), are evaluated by comparing the out-of-bag performance with the one obtained by randomizing the order of observations of one variable at a time (shuffling), which is known as permutation-based variable importance assessment. According to Nicodemus (2011), the use of the mean decrease in accuracy, compared with the Gini variable importance, is well founded and more stable, in particular in the presence of potential correlation among environmental variables.

2.3 AdaBoost

Adaboost is a powerful machine learning algorithm developed by Freund and Schapire (1997) which improves the performance of simpler algorithms, or weak learners (e.g., linear classifiers on one variable) by combining them together. In a nutshell, AB iteratively trains a sequence of classifiers by focusing more and more on difficult (i.e., misclassified) data points and takes the final decision of class membership by combining their predictions and weighting them according to their performance. Furthermore, AB can be used to evaluate the relevance of variables by looking at how often they are selected by the weak learners. In the present study, a more recent and more efficient implementation based on a stochastic gradient descent was used (Friedman 2001, 2002). The packages `gbm` (Ridgeway 2013) and `randomForest` (Liaw and Wiener 2002) of the programming language R (R Core Team 2013) were used to carry out the experiments.

3 Methodology

3.1 Study Site and Data Preparation

This section presents the landslide inventory of Canton Vaud and the extraction of topographic and geological features.

3.1.1 Landslide Inventory

The landslide inventory of Canton Vaud (Fig. 2a) was generated from different mandates of the Canton Vaud administration. A first study was part of the DUTI project of the Swiss Federal Institute of Technology in Lausanne (Noverraz 1994). The DUTI-NOVERRAZ mandate had as its objective the generation of a complete inventory of sub-stabilized and active landsliding phenomena in the Canton Vaud. The identification methods were based upon a morphological and geological analysis of the field (Noverraz 1994). Morphological features were used for identification, delimitation and evaluation of the deepness of landslides, while geological analyses were used as verification tools and to explain the landsliding mechanisms. The Noverraz (1994) workflow included three approaches: (1) coarse identification of landslide zones using 1:10,000 topographical maps; (2) aerial photographs observation; and (3) field confirmation of all spotted landsliding zones. In the first step, topographical maps were used to identify steep zones uphill and detachment niches, followed by a depressed zone downhill and a convex zone, rivers deviation and enhanced erosion signs on the opposite valley side. Aerial photographs (1:25,000 scaled, or 1:15,000 for infra-red imagery from special airborne surveys) were used to refine LS zones. In particular, the state and degradation of vegetation, roads and geometric deformations were used to confirm the alleged pres-

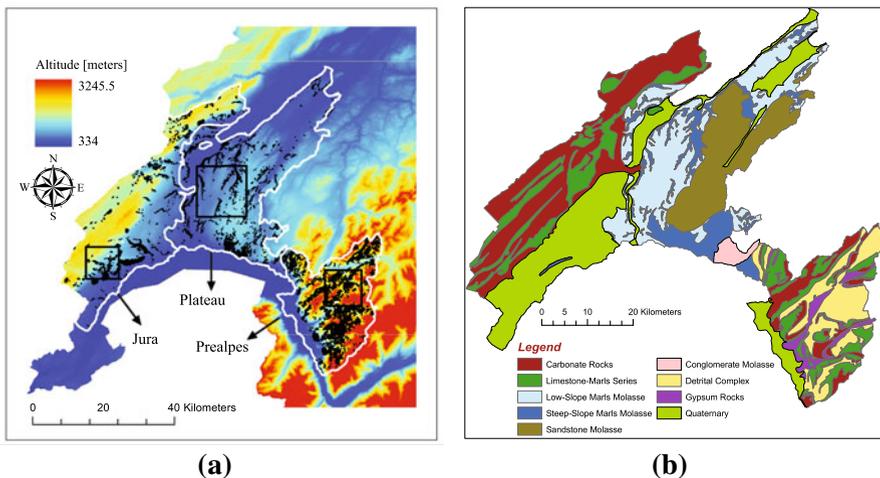


Fig. 2 **a** Landslide map of Canton Vaud overlaid on the $25 \times 25 \text{ m}^2$ DEM. The three sub-regions are delineated by white lines while the three regions of interest are marked with black boxes. **b** Geological map of Canton Vaud (9 main lithotypes)

Table 1 Size of landslides for the three geological domains of Canton Vaud (m^2)

Domain	$Q_{.1}$	$Q_{.25}$	$Q_{.50}$	$Q_{.75}$	$Q_{.90}$	Mean
Plateau	1,275	3,123	7,924	19,468	45,949	20,311
Jura	1,159	4,495	11,532	28,919	68,299	35,296
Prealps	1,740	4,985	14,079	36,274	83,768	38,164

ence of a LS. Fieldwork and quantitative evaluation of unstable zones completed the cartographic analysis, detecting sub-stabilized zones (with movements of a few centimeters per year) and active landslides (greater than 2 cm per year) (Noverraz 1994).

A second, recent mandate from Canton Vaud was assigned to the University of Lausanne (Pedrazzini et al. 2008). The objective was the correction and completion of the LS inventory of (Noverraz 1994) to establish an indicative mapping of natural hazards of Canton Vaud, including not only landslides, but also rockfalls, debris flows, etc. A detailed geomorphological analysis of the original inventory with new LiDAR DEM data (1 m cell size) unveiled spatial, extensional and morphological inconsistencies (Pedrazzini et al. 2008). The authors explain these errors as technical and natural limitations; in the first case, the low-resolution used for the analysis limited the precision of the mapping, while in the second case the aerial imagery analysis was judged inefficient in the presence of high anthropological disturbance and forest. The work of Pedrazzini et al. (2008) allowed: the evaluation, verification and, in the case of large errors, the correction of the DUTI-NOVERRAZ inventory; the identification of additional LS zones, and the evaluation of an indicative hazard map for the Canton Vaud. The early inventory is found to be extremely complete in the Plateau and in the Prealps. LiDAR DEM analyses allowed detecting new unstable areas in the Plateau and Jura regions. The final database included digitalized LS perimeters of the two mandates with detailed information, such as deepness, estimated movement velocity, genesis type, etc. It is important to note that the landslide inventory is composed of landslides showing clear morphological evidences but a variable activity in terms of present-day deformation rate. As suggested by Guzzetti et al. (2009), the principle that the past and the present are keys to the future can be adopted (Aleotti and Chowdhury 1999; Carrara et al. 1991; Guzzetti et al. 1999; Varnes 1984). This principle implies that slope failures in the future will be more likely to occur under similar geological and topographical conditions which led to past and present instabilities. Landslide inventory maps, even if incomplete, are fundamental data sources to predict the potential future occurrence of landslides.

In this study, the landslide inventory corrected by Pedrazzini et al. (2008) was employed (Fig. 2a). 85 % of the landslide features were detected since the early inventory (Noverraz 1994). Finally, the database includes 3,270 sliding bodies in the Plateau, 1,670 in the Jura chain and 3,576 in the Prealps, corresponding respectively to 1,326, 665 and 996 LS after aggregation of contiguous polygons. Different sliding bodies could be identified within the same LS to denote different properties (e.g., difference in deepness) or parts of the landslide. However, this study generally employs the whole database as unstable class, thus without differentiating between these factors. Descriptive statistics for each domain of Canton Vaud are presented in Table 1. Given the log-normal distribution of LS sizes, quantiles are used to describe this distribution instead of mean and standard deviation.

3.1.2 The Geological Setting of Canton Vaud

The nature of landslides in Canton Vaud differs considerably depending on the geographical area (Noverraz 1994; Pedrazzini et al. 2008). The three main geological regions of Canton Vaud (Plateau, Jura and Prealps) have been employed as separate study zones to account for the different nature of landslides which are, for instance, mostly deep-seated and complex in the Prealps area and shallow and roto-translational in Plateau and Jura regions. Consequently, it was assumed that the pattern linking controlling factors and LS is similar within each sub-area and thus the relevant features are expected to be associated with the most common landslide type in each sub-region.

The Plateau is mainly composed of Molasse sediments (mostly sandstone and marls) related to Alpine erosion and by the deposition of moraines. In the Vaud Plateau, hydrological processes are a major controlling factor of landslide occurrence due to the erosion of river sides. The topography is hilly and marked by steep slopes in the presence of rivers. The Jura chain has a SW–NE orientation and is characterized by thin skinned tectonics. Outcropping lithologies are mainly composed of Mesozoic sedimentary rocks, mostly limestone and marls (Trumpy 1980). The study area is moderately deformed with a homogeneous trend of smooth folds and faults. The mountainous zone of the Prealps is formed by a complex allochthonous tectonic klip-pens (Mosar et al. 1996). They are formed by limestones, dolomites, marls and shales. These can be divided into three underlying units: the Median Prealps, the Helvetic and the Ultrahelvetic geological unit. The type of morphology is generally fluvio-glacial, containing steep and rugged mountain sides (Trumpy 1980). In the Prealps region, landslide triggering is mainly related to intense rainfall and rapid snow melt (Bollinger et al. 2012; Bonnard 2006; Tullen 2000). In the Plateau and Jura region, triggering mechanisms are mainly intense rainfall and progressive erosion on the foot of the slope (Noverraz and Bonnard 1990). To account for the geological conditions of Canton Vaud, 9 main lithotypes have been derived by aggregating the 16 classes of a 1:25,000 geological map provided by the Security and Environment Department of Vaud. Figure 2b illustrates the spatial distribution of these 9 lithological classes in the three geographical sub-regions of Canton Vaud.

3.1.3 Topographic Features

Digital elevation models have become inexpensive sources of information and therefore a database for the computation of morphological attributes that are related to landslide occurrence (Brenning 2005; Kalbermatten et al. 2011). For the present case study, a 25 m × 25 m resolution DEM (Fig. 2a) was employed. The DEM was derived by the Swiss Federal Office of Topography (swisstopo) in 2009 from the Swiss National Map 1:25,000 in the project MNT25-matrix model. Its precision is 1.5 m for the Plateau and the Jura and of 2 m for the Prealps. The selection of the set of morphological features to extract resulted from discussions with landslide experts and the common practices in the field (Brenning 2005; Yao et al. 2008; Yesilnacar and Topal 2005). The following morphological and hydrological features were extracted from the DEM using the

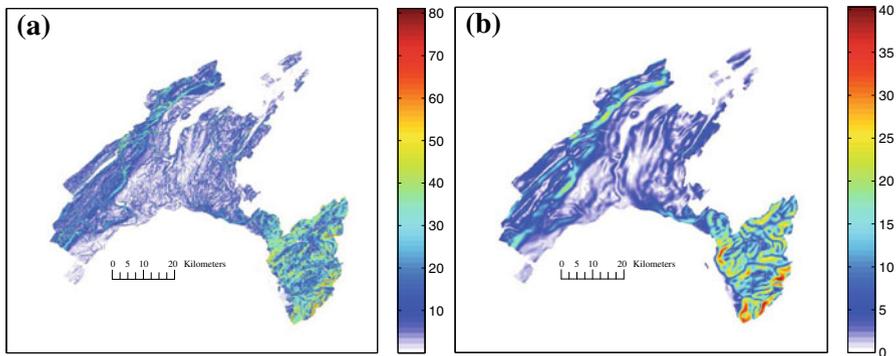


Fig. 3 Multiscale representation of slope (**a** slope, **b** filtered slope) [degrees]

softwares TauDEM (Tarboton 2005), ArcGIS and Matlab: terrain aspect, slope, plan curvature, profile curvature, topographic wetness index (TWI) and surface curvature. Since terrain aspect is a circular variable, its inclusion is made using two uncorrelated variables as proposed in Brenning and Trombotto (2006): North-exposedness, corresponding to the cosine of aspect, and West-exposedness, corresponding to minus the sine of aspect angle.

Due to the scale-dependent relationship between the size of landslides and the underlying topography, some features (slope, plan, surface and profile curvatures) were completed by adding their multiscale representations. The procedure consists in filtering the DEM with Gaussian convolution matrices prior to computing terrain slope and curvature (refer to Foresti et al. 2011, 2012) for more details on the methodology). Based on visual empirical observation of the smoothed variables, convolution matrices of 125 and 605 pixels with bandwidths σ of 21 and 101 (measured in number of pixels) respectively have been applied. These values correspond to 525 and 2,525 m, respectively. The resulting features are identified by the coding G.21 and G.101 and an example is shown in Fig. 3. The variables composing the resulting datasets are presented in Table 2. They are used to perform the supervised classification to discriminate between absence or presence of LS.

3.2 Experimental Setup

SVM, RF and AB were trained based on the input set of geological and topographic features (including terrain altitude) to predict the occurrence of landslides. The models are separately trained and evaluated for each sub-region, that is Plateau, Jura and Prealps. A region-of-interest (ROI) has been extracted from each sub-region to compute landslide susceptibility maps for visualization purposes (Fig. 2a).

Random and object-based sampling strategies were considered to analyze the impact of calibrating and testing the model on spatially independent landslides. The object-based sampling is performed in order to reduce the spatial autocorrelation between the landslide pixels composing the training, calibration and test sets and to obtain more realistic estimates of prediction accuracy (Brenning 2012). The random

Table 2 List of morphological and hydrological variables used

Variable	Min	Max	Mean	STD
Altitude	372	3,200	901	443
Northness	−1	1	−0.067	0.697
Westness	−1	1	−0.043	0.713
Slope	0	81.10	11.53	11.70
Slope G101	≈0	40.27	6.43	6.42
Plan C.	−357.39	192.50	−0.0033	1.98
Plan C. G21	−68.87	79.28	−0.0032	0.58
Plan C. G101	−85.59	105.34	−0.0025	0.19
Profile C.	−11.98	5.91	0.0046	0.20
Profile C. G21	−0.75	0.47	0.0022	0.04
Profile C. G101	−0.06	0.07	≈0	0.0088
Surface C.	2.36	3.67	≈0	0.0651
Surface C. G21	−0.08	0.12	≈0	0.0084
Wetness Index	1.44	18.47	7.00	1.92

strategy consists of sampling the required number of pixels for training, calibration and testing purposes without any spatial constraint. Consequently, training, calibration and testing pixels can come from the same landslide. Due to the high spatial autocorrelation of topographic and geological features, the use of training pixels to predict calibration and testing pixels located on the same landslide would lead to unrealistic overestimations of prediction accuracy. Further insights about sampling strategies and the issue of spatial autocorrelated data in accuracy assessments in the field are presented in [Nefeslioglu et al. \(2008\)](#) and [Yilmaz \(2010b\)](#). The object-based sampling consists of two main steps. The first step considers landslides as areas of contiguous pixels (see area statistics in Sect. 3.1.1). The landslide objects are randomly split into training, calibration and test groups and to each landslide is assigned a label. This step is performed only once. In a second step, the required number of pixels for training, calibration and testing sets is randomly sampled from the set of labeled landslides. This object-based procedure avoids using data from the same landslide for different purposes, for instance training of the model, parameter selection (calibration), and estimation of the generalization skill (testing).

Balanced training and calibration sets of 2,000 points each (1,000 pixels from landslides and 1,000 pixels randomly sampled from non-landslide areas) were constructed to ensure geographical representativeness while keeping the training dataset at a size that is computationally tractable by the machine learning algorithms. For both approaches, the random sampling part is iterated 10 times in order to analyze the sensitivity of machine learning algorithms to different training and calibration sets, and an independent and balanced testing set of 10,000 pixels has been constructed and used for estimating the generalization skills. The use of the object sampling strategy ensures the presence of examples for Plateau, Jura and Prealps extracted from 348, 193 and 323 landslides not employed for model training or calibration respectively. Before training the different models, input features were standardized to 0 mean and unit variance to ease the training and comparison of machine learning methods and to

provide equal scaling to every feature prior their assessment. Generalization skills of all models were assessed by the testing set using AUC. Maps that relate to the landslide susceptibility were derived using the probabilistic SVM interpretation explained in Sect. 2.1.

The selection of the hyper-parameters of support vector machines was done by grid search in the range of $C = \{1; 10; 100; 1,000\}$ and of $\sigma = \{1; 1.5; \dots; 10\}$ (same range for s_0 in the adaptive SVM). The best parameter combination is the one that maximizes the area under the receiver operating characteristics ROC (Egan 1975) curve (referred to as AUC) evaluated on the calibration set. The generalization skills of all models were also assessed by the AUC. This scalar is a measure of performance which considers the difference between the true positive and false positive rates integrated over the whole range of thresholds of the decision/membership function of a classifier. An AUC of 1 corresponds to perfect accuracy while an AUC of 0.5 occurs for a random pattern (no discrimination). For the adaptive SVM, the model which had the maximum calibration AUC but at least 5 null scaling factors was selected. This procedure was employed to eliminate the degenerated solutions (no convergence of the algorithm). The final scaling factors, which are also an estimation of feature relevance, were divided by s_0 to ease the comparison of features relevance within the 10 different trials. The relevance of features for the linear SVM is simply analyzed by the magnitude of the components of the weight vector defining the separating hyperplane, which is computed as

$$\mathbf{w} = \sum_{i=1}^L \alpha_i y_i \mathbf{x}_i, \quad (4)$$

and where negative values of the vector \mathbf{w} represent a negative relationship between the input features and the output.

Random Forests requires two parameters to be tuned by the user: (1) the number of trees T , (2) the number of variables m to be randomly selected from the available set of features. It is recommended (Breiman 2001) to pick a large number of trees and the square root of the dimensionality of the input space for m . Following these guidelines, the number of trees in RF has been fixed to 500 after a preliminary analysis and the number m of variables sampled at each node has been chosen to be 4 to analyse the joint contribution of subsets of features while keeping a fast convergence during iterations. No calibration set is required to tune the parameters.

The parameters to tune in AdaBoost are the number of iterations and the learning rate. Generally, the slower the learning rate, the better the performance but the longer the computing time. Here a default learning rate of 0.001 was fixed. The number of iterations was optimized on the calibration set and is about 6,000. A linear classifier was used as weak learner.

4 Landslide Susceptibility Mapping Results

This section presents the results of feature selection, model performance and landslide susceptibility mapping in the three sub-regions of Canton Vaud.

4.1 Feature Selection

Figure 4 summarizes feature selection results of adaptive SVM, RF and AB by constructing boxplots from the final distribution of scaling and importance factors evaluated on the 10 different splits using the object-based sampling approach. The boxplots are only used for visualization purposes to analyze the sensitivity of algorithms to data sampling and are not intended to give a robust estimation of statistics, which would require running the models hundreds of times.

The first observation is that the three methods present different levels of sparseness, as they are more or less parsimonious in selecting or eliminating variables. While the adaptive SVM explicitly seeks for sparse solutions within its optimization scheme, the results are not as good as expected. In fact, the scaling factors of features which are shrunk to 0 vary a lot within the 10 random splits. This sensitivity to data splitting denotes a certain instability of the adaptive SVM optimization. It is not yet clear if this is related to the presence of categorical variables in the input space (which complicates the numerical estimation of the gradient), the consequences of extending the simple isotropic SVM to account for more complex anisotropies via input scaling, the restrictive conditions over the scaling factors or the estimation of the gradient as a combination of training samples (which can vary between different random splits). Even though RF and AB do not force an explicit sparseness, they highlight a lower number of features with respect to the adaptive SVM, the AB having the highest parsimony. The linear SVM also shows quite variable results but has the strength of indicating the type of dependence (positive or negative). With the exception of terrain slope, which is among the most important variables controlling landslide occurrence in every region, different features are highlighted.

In the Plateau region, AB selects only the terrain slope, surface curvature G.21 and steep-slope marl molasse while discarding all other features (Fig. 4c). RF confirms the selection of the lithology and slope (Fig. 4b). The weighting of adaptive scaling SVM (Fig. 4a) presents a high degree of variation and its interpretation is difficult. However, four variables are characterized by weights outliers and are occasionally evaluated as relevant by the algorithm. These variables are westness, terrain slope, surface C. and steep-slope marl molasse. The linear SVM provides similar results for slope, surface curvature G.21 and steep-slope marl molasse but shows a high variance in the plan curvature G.21 (Fig. 5a).

A higher number of features is highlighted in the mountain range of Jura. AB selects altitude, terrain slope and surface curvature G.21 (Fig. 4f). RF identifies the same variables plus terrain slope G.101 (Fig. 4e). The solution for the adaptive SVM is more sparse compared to the Plateau and part of the results are in contrast with AB and RF. In particular, the plan curvature at small scale is selected without important sampling variation together with the three lithologies (carbonate, limestone-marl and quaternary), while surface curvature and profile curvature are eliminated (Fig. 4d). The linear SVM leads to similar conclusions and shows the uncertainty related to the impact of geological features in the models (Fig. 5b).

Due to the complex nature of landslides in the Prealps, more features are expected to contribute to the solution. AB selects mainly altitude, terrain slope, slope G.101, limestone-marl and clearly eliminates plan-profile-surface curvatures at small scale,

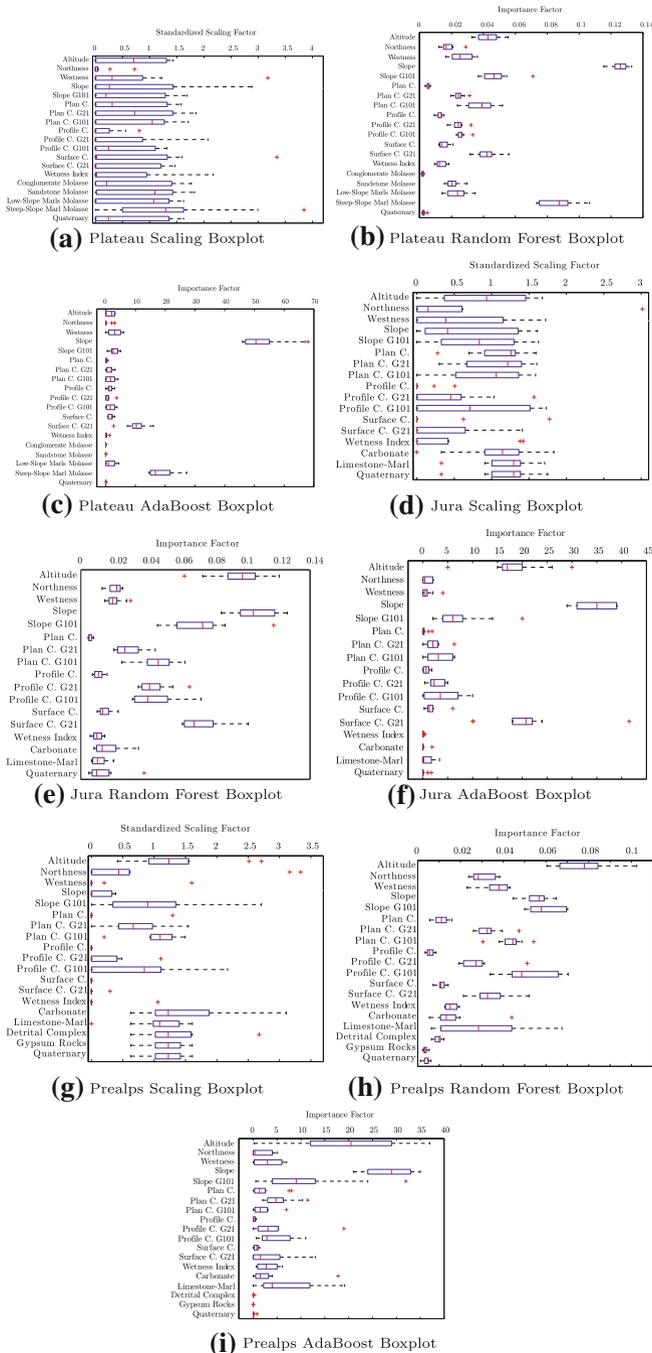


Fig. 4 Boxplots summarizing the scaling and importance factors evaluated by adaptive SVM, RF and AB in the three sub-regions (Plateau, Jura and Prealps). Sampling variation is assessed by 10 random splits using the object-based sampling procedure

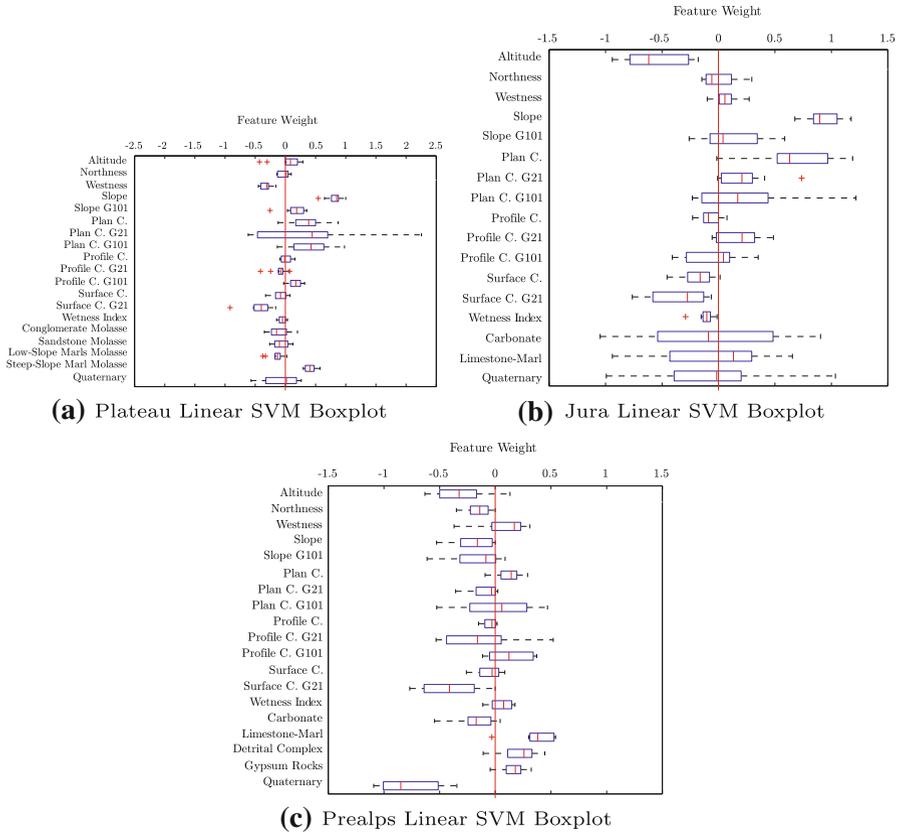


Fig. 5 Linear SVM weights boxplots assessed by 10 random splits using the object-based sampling procedure

detrital complexes, gypsum rocks and quaternary formations (Fig. 4i). The relevance of other features is hard to assess since they all have relatively low values but without having a null weight in many experiments. As expected RF tends to increase the importance of such variables by giving a less sparse solution (Fig. 4h). The results of adaptive scaling confirm the above-mentioned convergence issues when using heterogeneous data (Fig. 4g). In fact, the scaling factors of the five lithotypes remain close to one without showing significant differences between them. Again, it is interesting to point out the increasing importance of features with increasing spatial scales. This tendency was also found in the Plateau and Jura regions and confirms the utility of filtering the DEM at appropriate scales before computing terrain features. Results of the linear SVM are hard to evaluate but demonstrate the relevance of geology in controlling the occurrence of landslides (Fig. 5c).

The results of feature selection methods can be summarized by keeping the following features:

Plateau Westness, terrain slope, surface curvature G.21 and the steep marl-molasses, for a total of 4 features out of 19.

Jura Altitude, terrain slope, terrain slope G.101, plan curvature, surface curvature G.21, the 3 lithotypes, for a total of 8 features out of 17.

Prealps Altitude, terrain slope, terrain slope G.101, plan curvature G.101, profile curvature G.101, the 5 lithotypes, for a total of 10 features out of 19.

The final selection was done to accommodate the results of every single method and is rather conservative in the sense that a larger number of features is kept. In fact, because of the large variability of results, some relevant features may be discarded if selected based only on the results of a single method.

4.2 Model Assessment

Table 3 summarizes the area under the ROC curve evaluated for the independent testing set using both the random and object-based approaches for all the models considered. Plateau, Jura and Prealps present increasing complexity of conditioning factors associated with landslides: the corresponding AUC are respectively 0.88 to 0.92, 0.83 to 0.93 and 0.63 to 0.83. Compared with random sampling, the object-based sampling causes a decrease of 0.01 to 0.04 in the Plateau, 0.01 to 0.06 in Jura and of 0.09 to 0.12 in the Prealps. The performances obtained using random sampling are likely to be overestimated due to the spatial autocorrelation of the datasets, while the object-based sampling provides a more realistic estimate for the generalization error. The loss of performance in the Prealps can be attributed to the difficulty of predicting the occurrence of spatially independent landslides characterized by specific and localized factors. This reduction in performance is also more accentuated for models using the Gaussian instead of the linear kernel. This may be due to the increased risk of overfitting when using Gaussian kernels, which are unable to extrapolate the decision function for the prediction of landslides having conditioning factors that are too different from the ones in the training set. The linear kernel is more appropriate for extrapolating the decision function beyond the range of values of conditioning factors covered by the training set.

Except for the linear SVM, all methods provide similar accuracies using the object-based sampling. However, the performances obtained by RF and AB slightly exceed the ones of the Gaussian SVM in the Prealps. An experiment employing a logistic regression procedure provided performances of 0.87, 0.82 and 0.57 AUC in Plateau, Jura and Prealps, respectively (object-sampling strategy). These values are 0.01 to 0.02, 0.02 and 0.12 to 0.15 inferior to SVM, RF and AB results, a tendency observed in similar studies (Nefeslioglu et al. 2008; Yesilnacar and Topal 2005; Yilmaz 2010a).

Table 3 also shows the performances obtained using the subset of selected features described in Sect. 4.1. The result of adaptive SVM considers the scaled input space without applying a hard feature selection. Unexpectedly, using an isotropic kernel on the scaled input spaces reduces the AUC by 0.06–0.12 compared to using the complete standardized input space. Therefore, it is advised to use adaptive SVM as an exploratory tool to reveal the relevant variables, but to apply classic SVM in prediction mode. In general, the use of a reduced input space does not change much the performance of SVM. The decrease of the AUC is limited to 0.01 to 0.02 in all regions.

Table 3 Mean testing AUC together with sampling variation (standard deviation evaluated over 10 splits) for all models considered

Model	Sampling	Features	Plateau	Jura	Prealps
SVM _{gauss}	Random	All	0.92 (0.001)	0.90 (0.002)	0.81 (0.003)
SVM _{gauss}	Object	All	0.88 (0.009)	0.84 (0.013)	0.69 (0.012)
SVM _{lin}	Random	All	0.89 (0.001)	0.84 (0.002)	0.72 (0.002)
SVM _{lin}	Object	All	0.88 (0.015)	0.83 (0.013)	0.63 (0.018)
RF	Random	All	0.92 (0.001)	0.93 (0.003)	0.83 (0.004)
RF	Object	All	0.88 (0.006)	0.84 (0.013)	0.72 (0.015)
AdaBoost	Random	All	0.91 (0.001)	0.89 (0.001)	0.81 (0.003)
AdaBoost	Object	All	0.89 (0.005)	0.84 (0.011)	0.72 (0.016)
SVM _{adaptive}	Object	Scaled	0.76 (0.029)	0.73 (0.045)	0.63 (0.046)
SVM _{gauss}	Random	Subset ^a	0.90 (0.002)	0.90 (0.003)	0.80 (0.003)
SVM _{gauss}	Object	Subset ^a	0.89 (0.006)	0.84 (0.014)	0.70 (0.023)
SVM _{lin}	Random	Subset ^a	0.89 (0.001)	0.85 (0.011)	0.71 (0.001)
SVM _{lin}	Object	Subset ^a	0.89 (0.004)	0.84 (0.009)	0.63 (0.028)
RF	Random	Subset ^a	0.88 (0.002)	0.91 (0.003)	0.82 (0.002)
RF	Object	Subset ^a	0.86 (0.008)	0.85 (0.012)	0.70 (0.017)
AdaBoost	Random	Subset ^a	0.90 (0.001)	0.88 (0.002)	0.79 (0.003)
AdaBoost	Object	Subset ^a	0.89 (0.003)	0.84 (0.012)	0.72 (0.013)
LR	Object	All	0.87 (0.026)	0.82 (0.012)	0.57 (0.093)

^a Reduced input space described in Sect. 4.1

Table 4 Fraction of support vectors in the training set with corresponding standard deviation

Model	Sampling	Plateau (%)	Jura (%)	Prealps (%)
SVM _{gauss}	Random	44.39 (3.85)	54.56 (7.69)	69.30 (6.44)
SVM _{gauss}	Object	37.97 (3.90)	52.32 (4.98)	60.70 (15.16)
SVM _{lin}	Random	41.88 (1.80)	54.01 (1.35)	74.54 (1.20)
SVM _{lin}	Object	35.85 (3.32)	53.85 (5.32)	72.90 (1.97)

In some cases, a 0.01 AUC increase is observed, probably because of the elimination of noisy and uninformative features and the reduction of the space dimensionality. In conclusion, it is possible to construct models depending on few features while keeping acceptable performances in the Plateau and Jura. On the other hand, while performances are maintained at the same level, the Prealps probably require a larger set of features to characterize the complexity of unknown pre-conditioning factors.

The fraction of support vectors (Table 4) in the training set for SVM models is a good measure of the generalization skill and robustness of the model (Vapnik 1998). A robust model will depend only on a small set of support vectors, leaving the solution basically unchanged when adding new uninformative samples to the training set. In particular, the fraction of support vectors depicts how many training samples are needed to separate the two classes. As expected, this number is lower for the Plateau (36–44 %) compared with the Jura (52–55 %) and the Prealps (60–75 %). Also, the

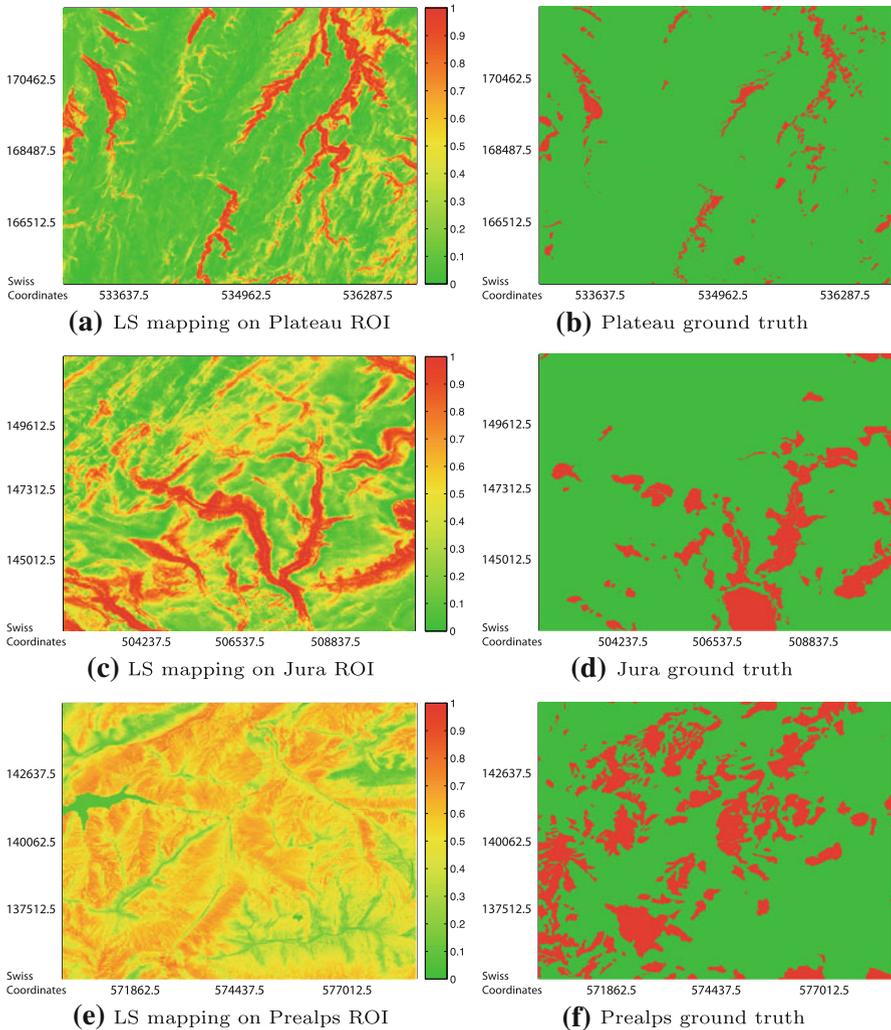


Fig. 6 **a, c, e** Probabilistic landslide susceptibility maps obtained by Gaussian SVM. Results are averaged over 10 splits using the object-based sampling approach. **b, d, f** Corresponding ground truth (known landslides)

fraction of support vectors slightly decreases when using the object-based sampling up to values of 8 %, which indicates the reduction of overfitting effects.

4.3 Landslide Susceptibility Mapping

Figure 6 shows the Gaussian SVM-based landslide susceptibility maps for the three ROI considered within each region and presented in Fig. 2a.

LS mapping in the Plateau (Fig. 6a) is characterized by a high discrimination of stable and unstable zones since there are few uncertain areas where the decision

function is around 0.5. In fact, most of areas are either close to 1 (unsafe) or 0 (safe). The high detection rate of landslides is demonstrated by actual observations (Fig. 6b). As already demonstrated by feature selection results (Fig. 4), landslides occupy mostly riverside slopes.

The mapping of landslide susceptibility in the Jura (Fig. 6c) has similar patterns but the regions of uncertainty are more spread. The ground truth validates most of the results except for the large landslide in the bottom of Fig. 6d, which is mainly a down-slope accumulation zone. To avoid the integration of data from these areas, seed cell (Suzen and Doyuran 2004; Yesilnacar and Topal 2005) or scarp (Yilmaz 2010b) sampling methods could be used to select more representative pixels at the head or scarp of the landslide.

LS mapping in the Prealps (Fig. 6e) is characterized by a higher uncertainty since few regions can be declared as safe or unsafe. Nevertheless, a reasonable correspondence between the areas with low susceptibility and the absence of landslides at regional level is observed (Fig. 6f). In particular, there is a higher influence of binary lithological features (also demonstrated by Fig. 4) in the top left and right corners of the map, which results in sharp transitions of the SVM decision function.

5 Discussion

The generalization skills of machine learning algorithms applied to discriminate landslide zones in the Plateau, Jura and Prealps were assessed using both random and object-based sampling to construct training, calibration and test sets. By reducing the mutual spatial autocorrelation between pixels, the object-based approach provides more reliable estimates of model performances with respect to random sampling. The consequence is that object-based approaches give slightly lower performances, a fact also observed in previous studies employing alternative sampling methods, for example seed cells sampling (Nefeslioglu et al. 2008). On the other hand, the first step of the object-based sampling applied in this study is not reiterated (Sect. 3.2) and the sampling variation is only assessed by randomly sampling pixels within the predefined set of disjoint landslides. A possible improvement would be to iterate the random splitting of landslide at the object and not pixel level, which would give a better estimation of the sampling variation and the representativeness of individual landslides.

The area under ROC curves evaluated on the spatially independent testing set (object-based sampling) for the different models (SVM, RF, AB) were 0.88 to 0.89 for the Plateau region, 0.83 to 0.84 for the Jura chain and 0.63 to 0.72 for the Prealpes. These values are comparable to previous studies employing similar machine learning approaches (Nefeslioglu et al. 2008; Yesilnacar and Topal 2005; Yilmaz 2010a). The results confirm what was expected by prior knowledge. In fact, the occurrence of landslides in the Plateau is principally correlated to the presence of riverside slopes (Pedrazzini et al. 2008), which easily described by a small set of morphological features. The lower class separability and high uncertainty in the Prealpes is due to the presence of complex and deep-seated landslides, for example the one of La Frasse (Tacher et al. 2005) compared to the shallow landslides of the Plateau and Jura regions. In this region, the performance of the Gaussian SVM is considerably higher than linear

one, which demonstrates that assuming linear data dependencies is an oversimplification to describe such complex cases. For these reasons, more research is needed to improve the statistical characterization of deep-seated landslides. In particular new image processing algorithms need to be developed for extracting more informative features accounting for the tectonic context (presence of faults, layer inclination) and the complexity of landslide conditioning factors. Up to date, the study of individual landslides with geotechnical approaches remains more appropriate in the Prealps, although affected by the over-prediction of shallow landslide occurrence (Pedrazzini et al. 2008).

Feature selection outputs are sometimes difficult to compare because the nature of landslides and their relation with conditioning factors can differ from case to case. Nevertheless, the terrain slope at one or more scales was found relevant in every region of Canton Vaud, which confirms the conclusions of other authors (Lee et al. 2004; Pradhan and Lee 2010; Yesilnacar and Topal 2005; Yilmaz 2010a). Furthermore, the terrain altitude results important in two out of three zones, which implicitly accounts for the climatological increase of precipitation with height, a factor that is relevant for landslide triggering (Pradhan and Lee 2010; Nefeslioglu et al. 2008; Yesilnacar and Topal 2005; Yilmaz 2010a). Another result that is confirmed by previous studies is the relevance of terrain curvature (Lee et al. 2004; Pradhan and Lee 2010; Yesilnacar and Topal 2005) and lithology (Nefeslioglu et al. 2008; Pradhan and Lee 2010). It is worth mentioning that other authors also studied the contribution of vegetation activity, for instance the timber diameter and age (Lee et al. 2004) and the NDVI (Yilmaz 2010a).

Another open issue is the simultaneous use of different types of features. In fact, input features are composed of continuous (topographic and hydrological) and categorical (geological) variables. Distance-based methods such as the SVM suffer from the scaling of input variables and it is mandatory to scale the data to 0 mean and unit variance before the analyses. Also in this case, the mixing of continuous and categorical variables still complicates the definition of distances and the assessment of relationships between variables (Otey et al. 2006). Convergence issues and instability of the adaptive scaling SVM may be explained in part by the use of heterogeneous inputs. Random forests and AdaBoost are less affected since they are developed using decision-trees and are not based on explicit Euclidean metrics. It should be noted, however, that, according to Strobl et al. (2007), the original random forest variable importance measure (permutation-based and Gini variable importance measure) may be slightly biased when used with variables having different measurement scales (continuous and categorical) or a different number of categories. One solution could be to summarize categorical variables into one ordinal variable by subjectively assessing the landslide susceptibility potential of each geological type. The design of machine learning algorithms could also be targeted to handle heterogeneous data either by introducing special metrics or by optimally mixing the properties of different methods, like the use of random forest based on conditional inference trees (Strobl et al. 2007).

The sensitivity to data resampling is expected to give information about the stability of the algorithms in providing consistent results. For example, it is clear that additional improvements on the optimization of adaptive SVM are required as results are quite variable depending on data splitting. The present study is a first experimentation of

adaptive SVM application, which was shown to work well when applied to simulated datasets (Grandvalet and Canu 2003). However, when applied to a real case study of LS mapping, the algorithm was affected by a higher variation of results, lower performances and in some cases convergence issues. These facts demonstrate that the application of machine learning algorithms to real case studies instead of simulated ones is fundamentally different, which reveals the weaknesses of the algorithms and can provide guidelines for further improvements.

The application of several methods for LS mapping has multiple advantages. First, since the classification task is formulated and solved in a different way, different sets of features are identified by the algorithms. This allows having different viewpoints to the feature selection problem and enables a cross-comparison of methods. Second, as shown in Sect. 4.2, machine learning approaches show higher performances compared to logistic regression, in particular in the Prealps. This tendency was already observed in Yilmaz (2010a) for ANN and SVM versus LR and conditional probability and in Nefeslioglu et al. (2008) and Yesilnacar and Topal (2005) for ANN versus LR. The disadvantage is the increased complexity of the algorithms and the difficulty in the interpretation of the relevance of features.

6 Conclusion

The paper studies the application of three machine learning methods for landslide susceptibility mapping from a set of morphological and geological features. The landslide inventory of Canton Vaud, Switzerland, was used to develop and test the different machine learning classification models. Topographic features, such as terrain slope and curvature, were extracted from the DEM at multiple spatial scales to correlate better with the characteristic size of landslides in three selected sub-regions. A set of categorical features was employed to characterize the geological conditions. In addition to a random sampling approach, an object-based strategy was proposed to obtain more realistic estimates of prediction accuracy. Support vector machines, random forests and AdaBoost were used to discriminate between absence or presence of LS, with performances peaking at AUC of 0.92, 0.93 and 0.83 for the three sub-regions of Canton Vaud. The machine learning algorithms were adapted to perform feature selection in order to reveal the variables which contribute most in determining the spatial distribution of landslides. The methods highlight the most important features to be selected but present different levels of sparseness and sensitivity to data resampling. Every algorithm is trained on the same set of features, but, as the classification problem is formulated and optimized in a different way, the final set of relevant features may be different. This can give the user the ability to analyze the problem under a different view, which is not possible by relying only on one single method. The results showed that the uncertainty of LS maps is smaller when most of landslides are shallow and strongly controlled by morphological features. On the other hand, LS maps computed over regions with many deep seated landslides are much more uncertain. This demonstrates the importance to integrate both statistical, physical and field based approaches to have a better overview of the landslide distribution in the different regions of the Canton Vaud, which present fundamentally different sliding mechanisms.

Acknowledgments This study was partially funded by the Swiss National Science Foundation projects *Geokernels: kernel-based methods for geo- and environmental sciences. Phase II (No. 200020-121835/1)* and *rockslides in Rhône valley (No. 200021-118105)*. We thank Prof. Stuart Lane for the interesting comments provided and Pierrick Nicolet for his valuable help. We also are grateful to the two anonymous reviewers, who provided us with constructive comments and helped in improving the quality of the manuscript.

References

- Adrizzone F, Cardinali M, Carrara A, Guzzetti F, Reichenbach P (2002) Impact of mapping errors on the reliability of landslide hazard. *Nat Hazard Earth Sys* 2:3–14. doi:[10.5194/nhess-2-3-2002](https://doi.org/10.5194/nhess-2-3-2002)
- Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. *B Eng Geol Environ* 58:21–44. doi:[10.1007/s100640050066](https://doi.org/10.1007/s100640050066)
- Atkinson PM, Massari R (1998) Generalised linear modelling of susceptibility to landsliding in Central Apennines, Italy. *Comp Geosci* 24:373–385. doi:[10.1016/s0098-3004\(97\)00117-9](https://doi.org/10.1016/s0098-3004(97)00117-9)
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15–31. doi:[10.1016/j.geomorph.2004.06.010](https://doi.org/10.1016/j.geomorph.2004.06.010)
- Ballabio C, Sterlacchini S (2012) Support vector machines for landslide susceptibility mapping: the Staffora River Basin Case Study, Italy. *Math Geosci* 40:47–70. doi:[10.1007/s11004-011-9379-9](https://doi.org/10.1007/s11004-011-9379-9)
- Bollinger D, Hegg C, Keusen HR, Lateltin O (2012) Ursachenanalyse der Hanginstabilitäten 1999. *Bull Angew Geol* 5:5–38
- Bonnard C (2006) Evaluation et prédiction des mouvements des grandes phénomènes d’instabilité de pente. *Bull Angew Geol* 11:89–100
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)
- Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation. *Nat Hazard Earth Sys* 5:853–862. doi:[10.5194/nhess-5-853-2005](https://doi.org/10.5194/nhess-5-853-2005)
- Brenning A (2009) Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sens Environ* 113:239–247. doi:[10.1016/j.rse.2008.09.005](https://doi.org/10.1016/j.rse.2008.09.005)
- Brenning A (2012), Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package *sperrorest*. In: International geoscience and remote sensing symposium (IGARSS), IEEE, International, pp 5372–5375. doi:[10.1109/IGARSS.2012.6352393](https://doi.org/10.1109/IGARSS.2012.6352393)
- Brenning A, Trombotto D (2006) Logistic regression modeling of rock glacier and glacier distribution: topographic and climatic controls in the semi-arid Andes. *Geomorphology* 81:141–154. doi:[10.1016/j.geomorph.2006.04.003](https://doi.org/10.1016/j.geomorph.2006.04.003)
- Canu S, Grandvalet Y, Guigue V, Rakotomamonjy A (2005) SVM and Kernel Methods Matlab toolbox. Perception Systèmes et Information, INSA de Rouen, Rouen, France
- Carrara A (1983) Multivariate models for landslide hazard evaluation. *Math Geol* 15:403–426. doi:[10.1007/BF01031290](https://doi.org/10.1007/BF01031290)
- Carrara A, Cardinali M, Detti R, Guzzetti F, Pasqui V, Reichenbach P (1991) GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf Proc Land* 16:427–445. doi:[10.1002/esp.3290160505](https://doi.org/10.1002/esp.3290160505)
- Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd international conference on machine learning, pp 161–168. doi:[10.1145/1143844.1143865](https://doi.org/10.1145/1143844.1143865)
- Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27
- Cherkassky V, Mulier F (2007) Learning from data: concepts, theory, and methods. Wiley, New York
- Daia F, Lee C (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* 42:213–228. doi:[10.1016/S0169-555X\(01\)00087-3](https://doi.org/10.1016/S0169-555X(01)00087-3)
- Dietrich WE, Reiss R, Hsu ML, Montgomery DR (1995) A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. *Hyrol Process* 9:383–400. doi:[10.1002/hyp.3360090311](https://doi.org/10.1002/hyp.3360090311)
- Egan J (1975) Signal detection theory and ROC analysis. Academic Press, New York
- Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* 66:327–343. doi:[10.1016/j.geomorph.2004.09.025](https://doi.org/10.1016/j.geomorph.2004.09.025)
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann Eugen* 7:179–188

- Foresti L, Tuia D, Kanevski M, Pozdnoukhov A (2011) Learning wind fields with multiple kernels. *Stoch Env Res Risk A* 25:51–66. doi:[10.1007/s00477-010-0405-0](https://doi.org/10.1007/s00477-010-0405-0)
- Foresti L, Kanevski M, Pozdnoukhov A (2012) Kernel-based mapping of orographic rainfall enhancement in the Swiss Alps as detected by weather radar. *IEEE T Geosci Remote* 99:1–14. doi:[10.1109/TGRS.2011.2179550](https://doi.org/10.1109/TGRS.2011.2179550)
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119–139. doi:[10.1006/jcss.1997.1504](https://doi.org/10.1006/jcss.1997.1504)
- Friedman J (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232. doi:[10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451)
- Friedman J (2002) Stochastic gradient boosting. *Comput Stat Data An* 38:367–378. doi:[10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gallus D (2010) Gaussian processes for classification of spatial data in context of an early warning chain. Dissertation, Karlsruhe Institute of Technology
- Goetz JN, Guthrie RH, Brenning A (2011) Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* 129:376–386. doi:[10.1016/j.geomorph.2011.03.001](https://doi.org/10.1016/j.geomorph.2011.03.001)
- Grandvalet Y, Canu S (2003) Adaptive scaling for feature selection in SVMs. *Adv Neur In* 15:553–560
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
- Guyon I, Gunn S, Nikravesh M, Zadeh L (eds) (2006) Feature extraction: foundations and applications. Springer, New York
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216. doi:[10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)
- Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2006) Estimating the quality of landslide susceptibility models. *Geomorphology* 81:166–184. doi:[10.1016/j.geomorph.2006.04.007](https://doi.org/10.1016/j.geomorph.2006.04.007)
- Guzzetti F, Ardizzone F, Cardinali M, Rossi M, Valigi D (2009) Landslide volumes and landslide mobilization rates in Umbria, Central Italy. *Earth Planet Sci Lett* 279:222–229. doi:[10.1016/j.epsl.2009.01.005](https://doi.org/10.1016/j.epsl.2009.01.005)
- Haykin S (1999) *Neural Networks: a comprehensive foundation*, 2nd edn. Prentice Hall, Upper Saddle River
- Kalbermatten M, Van De Ville D, Turberg P, Tuia D, Joost S (2011) Multiscale analysis of geomorphological and geological features in high resolution digital elevation models using the wavelet transform. *Geomorphology* 138:352–363. doi:[10.1016/j.geomorph.2011.09.023](https://doi.org/10.1016/j.geomorph.2011.09.023)
- Kanevski M, Pozdnoukhov A, Timonin V (2009) *Machine Learning For Spatial Environmental Data: Theory, Applications and Software*. EPFL Press, Lausanne
- Lee E (1974) A computer program for linear logistic regression analysis. *Comput Prog Biomed* 4:80–92. doi:[10.1016/0010-468X\(74\)90011-7](https://doi.org/10.1016/0010-468X(74)90011-7)
- Lee L, Ryu J, Won J, Park H (2004) Determination and application of the weights for landslide susceptibility mapping using and artificial neural network. *Eng Geol* 71:289–302. doi:[10.1016/S0013-7952\(03\)00142-X](https://doi.org/10.1016/S0013-7952(03)00142-X)
- Liaw A, Wiener M (2002) Classification and regression by random forest. *R News* 2(3):18–22
- Liess M, Glaser B, Huwe B (2011) Functional soil-landscape modelling to estimate slope stability in a steep Andean mountain forest region. *Geomorphology* 132:287–299. doi:[10.1016/j.geomorph.2011.05.015](https://doi.org/10.1016/j.geomorph.2011.05.015)
- Lin HT, Lin CJ, Weng RC (2007) A note on Platt's probabilistic outputs for support vector machines. *Mach Learn* 68:267–276. doi:[10.1007/s10994-0075018-6](https://doi.org/10.1007/s10994-0075018-6)
- Marmion M, Hjort J, Thuiller W, Luoto M (2008) A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. *Earth Surf Proc Land* 33:2241–2254. doi:[10.1002/esp.1695](https://doi.org/10.1002/esp.1695)
- Marmion M, Hjort J, Thuiller W, Luoto M (2009) Statistical consensus methods for improving predictive geomorphology maps. *Comp Geosci* 35:615–625. doi:[10.1016/j.cageo.2008.02.024](https://doi.org/10.1016/j.cageo.2008.02.024)
- Melchiorre C, Matteucci M, Azzoni A, Zanchi A (2008) Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology* 94:379–400. doi:[10.1016/j.geomorph.2006.10.035](https://doi.org/10.1016/j.geomorph.2006.10.035)
- Moguerza JM, Munoz A (2006) Support vector machines with applications. *Stat Sci* 21:322–336. doi:[10.1214/088342306000000493](https://doi.org/10.1214/088342306000000493)
- Montgomery DR, Dietrich WE (1994) A physically based model for the topographic control on shallow landsliding. *Water Resour Res* 30:1153–1171. doi:[10.1029/93WR02979](https://doi.org/10.1029/93WR02979)
- Mosar J, Stampfli GM, Girod F (1996) Western Préalpes Médiannes Romandes: timing and structure. A review. *Eclogae Geol Helv* 89:389–425

- Muchoney D, Strahler A (2002) Pixel- and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sens Environ* 81:290–299. doi:[10.1016/S0034-4257\(02\)00006-8](https://doi.org/10.1016/S0034-4257(02)00006-8)
- Neaupane K, Achet S (2004) Use of backpropagation neural network for landslide monitoring: a case study in the higher himalaya. *Eng Geol* 74:213–226. doi:[10.1016/j.enggeo.2004.03.010](https://doi.org/10.1016/j.enggeo.2004.03.010)
- Nefeslioglu H, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for preparation of landslide susceptibility maps. *Eng Geol* 97:171–191. doi:[10.1016/j.enggeo.2008.01.004](https://doi.org/10.1016/j.enggeo.2008.01.004)
- Nicodemus KK (2011) Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures predictors from random forest variable importance measures. *Brief Bioinform* 12:369–373. doi:[10.1093/bib/bbr016](https://doi.org/10.1093/bib/bbr016)
- Noverraz F (1994) Carte des instabilités de terrain du Canton de Vaud. Rapport conclusif et explicatif des travaux de levé de cartes. Swiss Federal Institute of Technology, Lausanne
- Noverraz F, Bonnard C (1990) Mapping methodology of landslide and rockfall in Switzerland. In: ALPS 90, Alpine landslide practical seminar, Milano, pp 43–53
- Ohlmacher GC, Davis JC (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng Geol* 69:331–343. doi:[10.1016/S0013-7952\(03\)00069-3](https://doi.org/10.1016/S0013-7952(03)00069-3)
- Otey ME, Ghoting A, Parthasarathy S (2006) Fast distributed outlier detection in mixed-attribute data sets. *Data Min Knowl Disc* 12:203–228. doi:[10.1007/s10618-005-0014-6](https://doi.org/10.1007/s10618-005-0014-6)
- Park NW, Chi KH (2008) Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model. *Int J Remote Sens* 29:247–264. doi:[10.1080/01431160701227661](https://doi.org/10.1080/01431160701227661)
- Pedrazzini A, Surace I, Horton P, Loye A (2008) Cartes Indicatives de Danger des Mouvements de Versants du Canton de Vaud. Faculty of Geosciences and Environment, University of Lausanne
- Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola AJ, Bartlett P, Schölkopf B, Schuurmans D (eds) *Advances in large margin classifiers*. MIT Press, Cambridge, pp 61–74
- Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: back-propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Modell Softw* 25:747–759. doi:[10.1016/j.envsoft.2009.10.016](https://doi.org/10.1016/j.envsoft.2009.10.016)
- R Core Team (2013) R: A language and environment for statistical computing. Vienna, Austria. <http://www.R-project.org/>. Accessed 17 January 2013
- Ridgeway G (2013) gmb: Generalized Boosted Regression Models. R package version 2.1. <http://CRAN.R-project.org/package=gmb>. Accessed 17 January 2013
- Schölkopf B, Smola A (2002) *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge
- Soeters R, van Westen CJ (1996) Slope instability recognition, analysis, and zonation. In: Turner AK, Schuster RL (eds) *Landslide: investigations and mitigation*. National Academy Press, Washington D.C., pp 129–177
- Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform* 8:25. doi:[10.1186/1471-2105-8-25](https://doi.org/10.1186/1471-2105-8-25)
- Stumpf A, Kerle N (2011) Object-oriented mapping of landslides using random forests. *Remote Sens Environ* 115:2564–2577. doi:[10.1016/j.rse.2011.05.013](https://doi.org/10.1016/j.rse.2011.05.013)
- Suzen M, Doyuran V (2004) Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Eng Geol* 71:303–321. doi:[10.1016/S0013-7952\(03\)00143-1](https://doi.org/10.1016/S0013-7952(03)00143-1)
- Tacher L, Bonnard C, Laloui L, Parriaux A (2005) Modelling the behaviour of a large landslide with respect to hydrogeological and geomechanical parameter heterogeneity. *Landslides* 2:3–14. doi:[10.1007/s10346-004-0038-9](https://doi.org/10.1007/s10346-004-0038-9)
- Tarboton DG (2005) Terrain analysis using digital elevation models (TauDEM). <http://hydrology.usu.edu>. Accessed 21 November 2012
- Terlien M, van Westen CJ, van Asch T (1995) Deterministic modelling in GIS-based landslide hazard assessment. In: Carrara A, Guzetti F (eds) *Geographical information systems in assessing natural hazards*. Kluwer, Dordrecht, pp 55–77
- Trumpy R (1980) *Geology of Switzerland, a guide book. Part A, an outline of the geology of Switzerland*. *Earth Sci Rev* 17:3
- Tullen R (2000) Glissement de la Chenolette (Bex-Les Plans, VD). *Bull Géol Appl* 5:39–45

- Vapnik V (1998) *Statistical learning theory*. Wiley, New York
- Varnes DJ (1984) *Landslide hazard zonation: a review of principles and practice*. Commission of Landslide of IAEG, UNESCO, Natural Hazards, Paris
- van Westen CJ, van Asch T, Soeters R (2005) Landslide hazard and risk zonation: why is it still so difficult? *B Eng Geol Environ* 65:167–184. doi:[10.1007/s10064-005-0023-0](https://doi.org/10.1007/s10064-005-0023-0)
- van Westen CJ, Castellanos Abella EA (2008) Spatial data for landslide susceptibility, hazards and vulnerability assessment: an overview. *Eng Geol* 102:112–131. doi:[10.1016/j.enggeo.2008.03.010](https://doi.org/10.1016/j.enggeo.2008.03.010)
- Yao X, Tham L, Dai F (2008) Landslide susceptibility mapping based on support vector machines: a case study on natural slopes of Hong Kong, China. *Geomorphology* 101:572–582. doi:[10.1016/j.geomorph.2008.02.011](https://doi.org/10.1016/j.geomorph.2008.02.011)
- Yesilnacar E, Topal T (2005) Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Eng Geol* 79:251–266. doi:[10.1016/j.enggeo.2005.02.002](https://doi.org/10.1016/j.enggeo.2005.02.002)
- Yilmaz I (2010a) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61:821–836. doi:[10.1007/s12665-009-0394-9](https://doi.org/10.1007/s12665-009-0394-9)
- Yilmaz I (2010b) The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. *Environ Earth Sci* 60:505–519. doi:[10.1007/s12665-009-0191-5](https://doi.org/10.1007/s12665-009-0191-5)