Computers & Geosciences 81 (2015) 1–11

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling

J.N. Goetz [a,b,d,*], A. Brenning [d,b], H. Petschko [c], P. Leopold [a]

[a] Health and Environment Department, AIT-Austrian Institute of Technology GmbH, Konrad-Lorenz-Straße 24, 3430 Tulln, Austria
[b] Department of Geography and Environmental Management, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
[c] Department of Geography and Regional Research, University of Vienna, Universitätsstraße 7, A-1010 Vienna, Austria
[d] Department of Geography, Friedrich Schiller University Jena, Loebdergraben 32, 07743 Jena, Germany

ARTICLE INFO

Article history:
Received 6 June 2014
Received in revised form
14 January 2015
Accepted 16 April 2015
Available online 20 April 2015

Keywords:
Statistical and machine learning techniques
Landslide susceptibility modeling
Spatial cross-validation
Variable importance

ABSTRACT

Statistical and now machine learning prediction methods have been gaining popularity in the field of landslide susceptibility modeling. Particularly, these data driven approaches show promise when tackling the challenge of mapping landslide prone areas for large regions, which may not have sufficient geotechnical data to conduct physically-based methods. Currently, there is no best method for empirical susceptibility modeling. Therefore, this study presents a comparison of traditional statistical and novel machine learning models applied for regional scale landslide susceptibility modeling. These methods were evaluated by spatial *k*-fold cross-validation estimation of the predictive performance, assessment of variable importance for gaining insights into model behavior and by the appearance of the prediction (i.e. susceptibility) map. The modeling techniques applied were logistic regression (GLM), generalized additive models (GAM), weights of evidence (WOE), the support vector machine (SVM), random forest classification (RF), and bootstrap aggregated classification trees (bundling) with penalized discriminant analysis (BPLDA). These modeling methods were tested for three areas in the province of Lower Austria, Austria. The areas are characterized by different geological and morphological settings.

Random forest and bundling classification techniques had the overall best predictive performances. However, the performances of all modeling techniques were for the majority not significantly different from each other; depending on the areas of interest, the overall median estimated area under the receiver operating characteristic curve (AUROC) differences ranged from 2.9 to 8.9 percentage points. The overall median estimated true positive rate (TPR) measured at a 10% false positive rate (FPR) differences ranged from 11 to 15pp. The relative importance of each predictor was generally different between the modeling methods. However, slope angle, surface roughness and plan curvature were consistently highly ranked variables. The prediction methods that create splits in the predictors (RF, BPLDA and WOE) resulted in heterogeneous prediction maps full of spatial artifacts. In contrast, the GAM, GLM and SVM produced smooth prediction surfaces. Overall, it is suggested that the framework of this model evaluation approach can be applied to assist in selection of a suitable landslide susceptibility modeling technique.

1. Introduction

Mitigating the impacts of landslides remains a great challenge for land-use planners and policy makers. Landslide susceptibility models, which are used to derive maps of locations prone to landslides, can support and enhance spatial planning decisions focused on reducing landslides hazards. Currently there is a vast selection of quantitative methods applied for spatial modeling and predicting landslide susceptibility (Chung and Fabbri, 1999; Guzzetti et al., 1999; Dai et al., 2002; van Westen et al., 2003; Brenning, 2005; Goetz et al., 2011; Pradhan, 2013). Quantitative methods for modeling landslide susceptibility can be generalized into physically-based and statistical approaches (Soeters and van Westen, 1996; van Westen et al., 1997). This study focuses on statistical and machine learning techniques, which have become common approaches for modeling landslide susceptibility over large regions (van Westen et al., 1997; Brenning, 2005; Petschko et al., 2014, Micheletti et al., 2014).

The basic assumption of the empirical approach is that future

* Corresponding author at: Department of Geography, Friedrich Schiller University Jena, Loebdergraben 32, 07743 Jena, Germany.
E-mail address: jason.goetz@uni-jena.de (J.N. Goetz).

http://dx.doi.org/10.1016/j.cageo.2015.04.007
0098-3004/

landslides are likely to occur in similar conditions of the past (Varnes, 1984). Typically, a range of predictors (i.e., independent variables) is used to represent landslide preparatory conditions (van Westen et al., 2008). The exact relationship of the predictors to the response (i.e., landslide presence/absence) is not always well known a priori. In most cases, the predictors are proxies for conditions and processes that are difficult to measure across large regions (Pachauri and Pant, 1992; Guzzetti et al., 1999; Goetz et al., 2011). The susceptibility model output is a prediction surface or map that spatially represents the distribution of predicted values, usually as probabilities distributed across grid cells.

The freedom of choice to decide which modeling method is most suitable for a particular application is challenging. Numerous comparisons of susceptibility modeling methods have been conducted; yet no single best method has been identified (Brenning, 2005; Yesilnacar and Topal, 2005; Lee et al., 2007; Yilmaz, 2009, 2010; Yalcin et al., 2011; Goetz et al., 2011; Pradhan, 2013). The search for the optimal susceptibility modeling method is a complicated one and should not only consider model accuracy. Robustness to sampling variation and adequacy to describe processes associated with landslides are also crucial model properties (Frattini et al., 2010).

The simplest approach to select an optimal model for prediction is to compare the error rates estimated from cross-validation, where the modeling method with the lowest error estimate is determined as the best one to use. This assessment on the prediction performance is also viewed as essential for a model to have any practical or scientific significance (Chung and Fabbri, 2003; Guzzetti et al., 2006). There are a variety of measures to assess the accuracy of landslide susceptibility models. Common ones are derived from success rate curves, prediction rate curves (Chung and Fabbri, 2003) or receiver operating characteristic (ROC) curves (Brenning, 2005; Beguería, 2006; Gorsevski et al., 2006; Frattini et al., 2010). It is necessary to carefully select a suitable performance measure. Ideally, this measure should communicate performance in the context of the model application (Brenning, 2012a). Performance should also be assessed using test sets that are independent from the training set used to build the prediction model, resampling-based estimation methods such as cross-validation being the state of the art (Brenning, 2012a): cross-validation utilizes the entire dataset for training and testing the model.

The ability to communicate model behavior is a desirable quality for landslide susceptibility models (Brenning, 2012a). Generally speaking, users feel more comfortable in the practical application of a model if they understand how the model works. The ability of a model to adequately describe the system behavior can be assessed by determining how well the predictors represent the processes associated with landslides (Frattini et al., 2010). In statistical methods, this is relatively straight forward compared to machine learning models. The model coefficients from generalized linear models can be used to evaluate the relative importance of landslide predictors (Dai and Lee, 2002; Ayalew and Yamagishi, 2005). Variable importance has also been estimated for regression models by observing the relative frequencies of variable selection when an automatic stepwise variable selection method has been applied and tested with cross-validation (Brenning, 2009; Goetz et al., 2011; Petschko et al., 2014). In contrast, the internal mechanisms defining the representation of response by the predictors are difficult to interpret in machine learning models because of their 'black box' nature. Micheletti et al. (2014) demonstrated how some feature selection properties of different machine learning techniques can be implemented to assess the relative importance of variables for landslide susceptibility modeling. However, since their approach applied features selection methods only relevant to the corresponding machine learning technique, making comparisons of variable importance with other modeling techniques can be challenging. A standardized approach for comparing the relative variable importance of different modeling statistical and machine learning techniques for geospatial problems was demonstrated by Brenning et al. (2012b). They assessed variable importance using internal estimates of changes in error rates by randomly permuting predictors in out-of-bag samples (Breiman, 2001; Strobl et al., 2007).

There are many criteria that can be considered for model selection in the context of landslide susceptibility (Brenning, 2012a). This study focuses on one particular aspect, which is the predictive performance. Therefore, a rigorous assessment of prediction performance is performed on various statistical and machine learning techniques in an attempt to determine the 'best' predictive model. The modeling techniques include logistic regression (GLM), generalized additive models (GAM), weights of evidence (WOE), the support vector machine (SVM), random forest classification (RF), and bootstrap aggregated classification trees (bundling) with penalized discriminant analysis (BPLDA). The importance of predictor variables in each model is also analyzed to demonstrate how a standard measure of variable importance can be applied to communicate and compare model behavior, even when a model is considered 'black box'. The main objective of this paper is to demonstrate an approach to make a rigorous comparison of landslide susceptibility models for the purpose of spatial prediction.

2. Materials and methods

2.1. Study area

Multiple areas of interest (AOIs) were selected to observe model behavior under different landslide conditions. The modeling techniques were tested on AOIs that were each 50 km^2 and within the province of Lower Austria (Fig. 1). The Molasse AOI (Fig. 1a) is located in a relatively low lying basin ($<$ 300 m a.s.l). It mainly consists of sand and clay sediments, sandstones, claystones, and marls. These bedrock materials can be covered by Quaternary gravels and eolian sediments (loess). Deep-seated and shallow landslides occur in this area. The Austroalpine AOI (Fig. 1b), which includes the Upper Austroalpine lithology units, is made up of predominately steep terrain, and has the highest elevations in Lower Austria (1000–2000 m a.s.l). The lithology is dominated by limestone and dolomite rock, with some interbedded strata of claystone and marl. Landslides in the Austroalpine area are typically shallow. Generally, the Flysch AOI (Fig. 1c) is very susceptible to landslide activity (Petschko et al., 2014). This low mountain region has exceptionally undulating terrain, and consists of sedimentary rocks that are made up of layers of sandstones, marls and claystones. The main triggers for landslides in Lower Austria are intense rainfall and rapid snowmelt events (Schwenk, 1992; Schweigl and Hervás, 2009). For more details on the lithology and geology of Lower Austria, please refer to Gottschling (2006) and Wessely (2006).

2.2. Landslide inventory

The landslides in this analysis are a subset of an inventory for Lower Austria that has been previously published by Petschko et al. (2014), which consists of mapped initiation areas for deep-seated and shallow landslides. These landslides were mapped in a geographic information system (GIS) using topographic derivatives (e.g. hillshade and slope angle maps) of an airborne laser scanner (ALS) digital terrain model (DTM) with a 1 m $\times$ 1 m spatial resolution, which was acquired between 2006 and 2009. The general procedure for mapping these landslides was similar to Schulz (2004, 2007). This inventory consists of points (single grid cells)