Large scale landslide susceptibility assessment using the statistical methods of logistic regression and BSA – study case: the sub-basin of the small Niraj (Transylvania Depression, Romania)

S. Roșca¹, Ş. Bilașco¹,2, D. Petrea¹, I. Fodorean¹, I. Vescan¹, S. Filip¹, and F.–L. Măguț¹

¹“Babeș-Bolyai” University, Faculty of Geography, 400006 Cluj-Napoca, Romania
²Romanian Academy, Cluj-Napoca Subsidiary Geography Section, 9, Republicii Street, 400015 Cluj-Napoca, Romania

Received: 7 November 2015 – Accepted: 14 November 2015 – Published: 27 November 2015

Correspondence to: S. Roșca (rosca_sanda@yahoo.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

The existence of a large number of GIS models for the identification of landslide occurrence probability makes difficult the selection of a specific one. The present study focuses on the application of two quantitative models: the logistic and the BSA models. The comparative analysis of the results aims at identifying the most suitable model. The territory corresponding to the Niraj Mic Basin (87 km$^2$) is an area characterised by a wide variety of the landforms with their morphometric, morphographical and geological characteristics as well as by a high complexity of the land use types where active landslides exist. This is the reason why it represents the test area for applying the two models and for the comparison of the results. The large complexity of input variables is illustrated by 16 factors which were represented as 72 dummy variables, analysed on the basis of their importance within the model structures. The testing of the statistical significance corresponding to each variable reduced the number of dummy variables to 12 which were considered significant for the test area within the logistic model, whereas for the BSA model all the variables were employed. The predictability degree of the models was tested through the identification of the area under the ROC curve which indicated a good accuracy (AUROC = 0.86 for the testing area) and predictability of the logistic model (AUROC = 0.63 for the validation area).

1 General consideration

One of the main natural hazards affecting the territory of Romania is represented by landslides which have a high spatial and temporal frequency and cause damages to transport infrastructure and buildings and determine environmental changes (Bălteanu and Micu, 2009; Bilaşo et al., 2011; Năsui and Petreuş, 2014).

EEA European Directive from 2004 underlines the need to mapping and identification areas with vulnerability to landslides using indirect techniques in European and national
context (Guzetti, 2006; Van Westen et al., 2006; Magliulio et al., 2008; Polemio and Petruci, 2010).

Thus, the studies determining their probability of occurrence are highly valuable in the process of reducing their potential negative effects. Among the methods used for determining the spatial probability of landslides, statistical methods are recommended by very good results and high validation rates (Zezere et al., 2004; Petrea et al., 2014; Roşca et al., 2015a, b).

Considering the increase in the number of possibilities for data processing and the evolution of methods developed in the GIS environment, various methods of landslide susceptibility assessment have been developed, out of which the logistic regression and bivariate statistical analysis methods is one of the most frequently used (Harrell, 2001; Kleinbaum and Klein, 2002; Ayalew and Yamagishi, 2004, 2005; Dai and Lee, 2002; Lee, 2010; Cuesta et al., 2010; Chiţu, 2010; Mancini et al., 2010; Wang et al., 2011; Guns and Vanacker, 2012; Jurchescu, 2013; Măguţ et al., 2013; Akbari et al., 2014; Van den Eeckhaut et al., 2010). This analysis starts from the hypothesis that the combination of factors which led to the occurrence of landslides in the past will have the same effect in the future (Crozier and Glade, 2005).

Among the advantages of this method one must take into consideration the possibility of simultaneously integrating both quantitative and qualitative data in the model and the testing of represent dependent variables while their triggering and preparing factors are the independent (explanatory) variables.

The purpose of this study is to identify the large scale susceptibility of landslide occurrence by applying the logistic model in the sub-basin of the Small Niraj (Fig. 1). The database included a complete landslide inventory and the descriptive data of 16 causing factors used for generating the model. These factors describe the morphometrical, geological and the hydroclimatic characteristics of the territory under analysis.
2 Study area

The study area is located in the north-east of Transylvania Depression, Romania, and has recorded important economical and environmental losses over the last two years: 67 persons, 45 houses, 115 ha of land and a country road were affected by landslides. The catchment area is found between 24°47′52″ and 24°58′32″ E longitude and 46°30′53″ and 46°37′42″ N latitude, totalizing an area of 68 km² and including the territories of ten settlements. The Small Niraj represents the main river of the area.

Based on the Romanian National Meteorological Administration Institute the mean temperature varies between −4.2°C in January and 17.9°C in August. The mean annual rainfall is around 622 mm yr⁻¹, while the maximum precipitation falls between May (73.5 mm) and June (81.5 mm).

3 Database and methodology

GIS spatial analysis models are built upon complex structures and databases generated from varied sources. One of the main problems to solve during the building of a spatial analysis model that localizes the areas with different landslide susceptibility values is represented by the identification of its actual format along with the building and the integrated management of the model input data.

The large variety of databases serving as input data in the complex identification model concerning landslide susceptibility, makes it that the different model structures have a resolution dependent on the model scale. Bearing in mind that the scale for the models fits within the large scale category, the authors have built a database both vector (landslide areas, geology, seismicity, land use) and raster data (slope angle, aspect, fragmentation depth, fragmentation density, elevation, CTI, SPI, plan and profile curvature etc.) (Table 1).

The spatial distribution of the 16 factors included in the model was determined using GIS functions of spatial analysis included in the ArcGis software.
The different database sources made their validation mandatory so as to ensure an accurate representation. The validation of the databases was done using the comparison technique (the database was compared to field data) as well as using observation (by visual identification of the correspondence existing between the cartographic representation and the existing situation in the field). Having the certainty that a valid and accurate database is used, the logical schemas of the BSA and logistic model were subsequently completed in order to be used for determining the probability of landslide occurrence.

The landslide susceptible areas are identified through the BSA model by considering the statistic value specific to each class of the factors included in the initial database, without taking into account the importance of the factor within the informational flux of the model. The statistical model based on the bivariate probability analysis was applied to predict the spatial distribution of landslides by estimating the probability of landslide occurrence based on the assumption that the prediction should start from the existing landslides (Chung et al., 1995; Dhakal et al., 2000; Saha, 2002; Sarkar and Kanungo, 2004; Magiulio et al., 2008; etc.).

The statistical value of each factor class included in the bivariate model was calculated using the equation proposed by Yin and Yan (1988), as well as Jade and Sarkar (1993):

\[ I_i = \log \frac{S_i}{N_i}, \]  

(1)

where: \( I_i \) = statistical value of the analysed factor, \( S_i \) = area affected by landslides for the analysed variable, \( N_i \) = area of the analysed variable, \( S \) = total landslide area in the analysed basin and \( N \) = area of the analysed basin.

By using Eq. (1), the statistical value of each variable is identified, the insignificant variables (characterised by negative values) being integrated with an equal weight in the model structure, occasionally reducing the susceptibility class values.
In order to predict landslide susceptibility at pixel level in the study area the model of logistic regression was also taken into consideration. This method was mathematically described by Harrel (2001): represents the set of points (pixels from the study area); Y represents the binary variables (0 for pixels without landslides and 1 for pixels with landslides); X₁,...Xₙ represent independent variables, in this study the 15 factors included in the model, each classified in various categories and represented with the help of dummy variables, out of which one class was not included in the model in order to be used as a control value (Van den Eeckhaut et al., 2006).

Thus, the probability of occurrence for a new landslide event is represented by:

\[ P = \frac{1}{1 + e^{-Z}}, \]  
where: \( Z = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n \), \( X_1 \ldots X_n \) – preparing and triggering factors, \( \beta_0 \) – constant and \( \beta_1 \ldots \beta_n \) – multiplication coefficients.

One can notice that the probability of occurrence becomes a linear function for each variable included in the model (Kleimbaum and Klein, 2002). In order to estimate the parameters, a logarithmic transformation of the odds ratio was necessary (represented by the ratio of the probability of success and the probability of failure) which changes the variation interval from \((0, 1)\) to a sigmoid curve, in the interval \((-\infty, +\infty)\) (Thiery, 2007; cited by Jurchescu, 2013). The main methodological stages are described in Fig. 2.

The \( \Omega \) study area was divided into two random sub-categories: \( \Omega_1 \) and \( \Omega_0 \). Hence, 500 points were used in the modelling process, 250 points generated at a minimum distance of 60 m in the landslide areas and 250 points at a minimum distance of 80 m in the non-landslide areas. A number of 40 landslides were randomly selected for the training stage and 15 landslides were included for the validation of the model. The validation set of points included a total of 200 randomly generated points at a minimum distance of 40 m (100 points inside the landslides and 100 points outside them). The importance of this stage which relies on a division of the study area in two sets of samples has been repeatedly emphasised by numerous authors with respect to the
independence of the validation set of data used to test the results of the logistic regression for landslide susceptibility assessment (Van den Eeckaut et al., 2006, 2010; Mancini et al., 2010; Mărgărint et al., 2013; etc.).

The coefficient values \((X_1, \ldots, X_n)\) of each landslide factor were necessary in order to determine the probability of landslide occurrence for each pixel, these coefficients being considered as representative for \(\Omega_1\) and \(\Omega_0\). In order to preserve the independence of the input factors, the 16 variables were transformed into dummy variables, resulting in a total of 73 variables, as each input factor was classified in different categories necessary for the comparative analysis. For each factor, one of the dummy variable was kept for reference (Hilbe, 2009).

The multiplication coefficient of each variable was determined by applying the logistic regression (Table 2). The \(\beta_0, \ldots, \beta_n\) parameters were estimated using the maximum likelihood ratio (i.e. inverse probability) (Harrel, 2011). This stage identifies the difference between the model which does not include the \(X_1\) parameter in the input database and the model which includes in its input database the \(X_n\) parameter. The variables with the highest influence were identified with the help of the AIC criterion which indicates the statistical significance of the variable.

A value below 0.05 is considered optimal, representing the threshold for the data acceptable within the model database. A statistical threshold value of < 0.1 determines the elimination of that specific variable from the present database, as it would raise multicollinearity issues (Cuesta et al., 2010). The coefficients resulting from the logistic regression were implemented in a GIS environment using the Raster Calculator functions, by multiplying them with the raster variables which represent the landslide preparing and triggering factors.

The goodness of fit was determined by generating the area under the ROC curve using the training data, while the prediction capacity of the model was identified using the validation data set (Hosmer and Lemeshow, 2000; Guzzetti, 2006). The quality of the information included in the input variables for the landslide susceptibility model.
as well as the number of variables need to be considered in the process of variable selection, in order to reduce redundancy (Chițu, 2010).

The 16 variables (elevation, slope angle, average precipitation, slope aspect, drainage density, drainage depth, hydrological soil classes, distance to streams, distance to roads and settlements, Stream Power Index (SPI), land use, lithology, plan curvature and profile curvature, Topographic Wetness Index (CTI) were included in the model, their selection being performed according to their statistical relevance in the logistic regression.

4 Results, validation and discussion

The establishing of the research methodology applied in the present study needs a comparative approach of the methods and of the results obtained through the implementing of the previously mentioned models.

The comparison of the spatial analysis methods integrated within the two models emphasises the difference among the necessary databases, as well as the complexity and implementation possibility of the models. The comparative approach of the results on the different levels of the modelling process as well as of the final results shows the practical utility of such databases within each model, as well as the accuracy of the representation.

4.1 Applied logistic regression to landslide susceptibility assessment

The statistical correlation between the mapped landslides from the Niraj River Basin and their causing factors was determined for the logistic model using the statistical software R. The training variables were included in the logistic regression and the AIC was used to perform an automated stepwise selection of the best model, namely the combination of variables which best explains the occurrence of landslides in the analysed territory.
The model with the best AIC value (AIC = 524) is given by the following expression:

$$
\text{fit3} = \text{glm}(\text{alunec} \sim \text{lndse}_8 + \text{spi}_1 + \text{dst}_h5 + \text{as}_10 + \text{as}_7 + \text{dst}_dr6 + \text{lndse}_3 \\
+ \text{dns}_f4 + \text{as}_6 + \text{slop}_4 + \text{pp}_2 + \text{dst}_dr7 + \text{dst}_lc7, \text{family} = \text{binomial}),
$$

(3)

According to the values of the multiplication coefficients (Table 2), the landslides from the Small Niraj River Basin are due to the following combination of favourable factors: slope angles ranging between 10 and 15° (Slop_4: 0.675), predominantly south-western and southern slope aspect (As_7: 1.374, As_6: 0.818), drainage density ranging between 1.5 and 2 m km$^{-2}$ (Dns_4: 1.017) and distance to streams ranging between 200 and 400 m (Dst_h5: 1.123). The negative coefficient values are caused by a reduced landslide density in the respective factor classes, thus being interpreted as restrictive classes for landslide occurrence.

For the interpretation of the results, the odds difference plays a very important role (Table 2). For example, keeping all the input variables constant while the average precipitation value is set at 650 mm yr$^{-1}$, the probability of landslide occurrence is by 29% higher than in the case of the reference value of precipitation (525 mm).

Thus, the highest increase in probability for landslide occurrence is recorded when comparing the south-western slopes with the reference class of level areas (195%) indicating a powerful dependency relationship between landslide occurrence and south-western slopes.

The resulting coefficients were multiplied with their corresponding 13 raster files using Raster Calculator according to Eq. (4):

$$
\text{Mdl_fit3} = \exp(-1.1381 + -2.0400 \cdot [\text{lndse}_8] + -1.3942 \cdot [\text{spi}_1] \\
+ 1.1238 \cdot [\text{dst}_h5] + -1.5113 \cdot [\text{as}_10] + 1.3744 \cdot [\text{as}_7] \\
+ 0.9694 \cdot [\text{dst}_dr6] + -2.3552 \cdot [\text{lndse}_3] + 1.0179 \cdot [\text{dns}_f4] \\
+ 0.8183 \cdot [\text{as}_6] + 0.7655 \cdot [\text{slop}_4] + 0.8281 \cdot [\text{pp}_2] \\
+ -0.7583 \cdot [\text{dst}_dr7] + 0.8739 \cdot [\text{dst}_lc7]).
$$

(4)
The landslide susceptibility map was generated by applying the odds ratio Eq. (5) representing the landslide susceptibility in the interval 0–1 (Fig. 3).

\[ S = \frac{p}{(1 - p)}, \]  

(5)

where \( S \) – susceptibility, \( P \) – probability.

The goodness of fit and the predictability of the model were determined using the ROC curve for the model sample and the testing sample, respectively. The sensitivity of the model represents the true positive rate (pixels with a high probability of landslide occurrence being validated by real landslides), while the model specificity represents the probability that the areas identified as highly susceptible to landslides to be invalidated by the lack of any landslides (false positive rate) (Hosmer and Lemeshow, 2000).

The area under the ROC (Relative Operational Curve) is 0.86 for the training data set and 0.63 for the testing (validation) data set, the first value indicating the goodness of model fit while the second represents the predictability of the model, or its capacity to predict future events (Fig. 4).

The large area under the ROC indicates a high sensitivity of the model as well as a low false positive rate which account for a satisfying precision of the results. The smaller ROC area in the case of the validation data, though still above the threshold of 0.5, is due to a smaller landslide set available for validation.

The classification of the results in the final susceptibility classes was based on the success rate (Chung and Fabbri, 1999, 2003, 2008; Van Westen et al., 2003; Remondo et al., 2003), resulting the map in Fig. 5.

4.2 Applied bivariate probability analysis (BSA) to landslide susceptibility assessment

The processing of the derived and modelled database by means of the ArcGis software using the specific functions of conversion, analysis and spatial integration has led to the generation of landslide susceptibility maps and their corresponding raster databases according to the statistical values of each coefficient class.
The results of the models are included in a raster database which highlights the probability of landslide occurrence for each pixel of the analysed area with a statistical value ranging from $-6.727$ to $+2.756$. The final susceptibility map was classified using the Natural Breaks method in five susceptibility classes (very low, low, medium, high and very high) (Fig. 5).

When analysing the classified susceptibility map one can note the vast expansion of the high and very high susceptibility classes (65% of the analysed area) which correspond to the slopes from the upper river basin of the Small Niraj (in the administrative territory of the Șirea Nirajului settlement), as well as in the hilly sector of the lower river basin (in the administrative territories of Miercurea Nirajului, Drojdi and Maia).

The validation of the results was performed in a first stage using the percentage of the landslide areas in each class (Fig. 6). Thus, there is a very good validation of the results as the largest proportion of the active landslides (71.23%) are included in the very high susceptibility class which also represents the second largest area in the Small Niraj River Basin (28.3 km$^2$).

By comparing the two databases it becomes obvious that 92.8% of the active landslides overlay the high and very high susceptibility areas and only 6.55% are included in the medium susceptibility class. This high degree of model fit is represented by the large area under the ROC (0.983) which indicates a good correlation between the model results and the landslides in the field (Fig. 6).

4.3 Comparison of results

The spatial distribution of the susceptibility classes in the case of the map generated with the help of the logistic model highlights a similar distribution in for the middle slope sectors from the lower and middle river basin, in the administrative territory of Miercurea Nirajului, Eremitu and Maia, but on the western slope of Măgherani Hill there are some obvious differences (Fig. 7).

The results differ between the application of the BSA model and the logistic model (Fig. 8). By applying the BSA model in which all the classes of the 16 factors were
included in the model, namely all the 72 dummy variables, there is an overestimation of the high susceptibility class (32.7%) and of the very high susceptibility class (32.5%). By applying the logistic model, these values decrease to 15.2% for the high susceptibility class and to 10.9% for the very high susceptibility class, as the variables corresponding to statistically insignificant classes were eliminated.

When comparing the input databases for the two models, there is a decrease in the initial number of variables (16) in the case of the logistic regression due to the application of the likelihood test (Table 6.21). Hence, the variable classes with a very reduced spatial expansion were excluded from the model as they would lead to additional errors (for example: the territories ranging between 700 and 800 m, slope angle values between 25 and 30°, territories at less than 50 m from settlements and at 25–50 m from the street network, a lithology dominated by sands, gravels alternating with marl and vineyards land use).

Another series of variable classes were excluded from the analysis, for example the territories with a drainage density between 0.5–1 m km⁻², a drainage depth between 51–100 m, the territories situated at 25–50 m from streams, pastures as well as the slopes with positive values of the plan curvature due to their low statistical significance.

As a result of the landslide susceptibility assessment performed with the help of the two quantitative models (bivariate statistical analysis and logistic regression) the areas with a high probability of landslide occurrence were highlighted in the study area as well as the stable territories. These results are considerably superior to previous analyses (surse) which used the legislative semi-quantitative Romanian methodology (H.G. 447/2003) (Rosca et al. 2015a). However, there is still the necessity of increasing the quality of the databases corresponding to the causing factors and the number of the landslides included in the modelling processes, as well as a more thorough analysis of the relationships between the parameters.
5 Conclusions

The two models under analysis in the present study, the logistic and the BSA models, have shown the high complexity of the databases involved, the multiple correlation between several factors determining landslide activation as well as the obvious practical utility of the logistic model in future similar studies.

The use of the logistic model has allowed the testing of variable interdependencies leading to a reduction of the input data, hence a shorter modelling time. The BSA model operates with all databases, 16 variables represented as 72 dummy variables, hence it takes longer for the model to be implemented and leads to an increased redundancy of the data, while the database management is slower and needs better software and hardware resources. One needs to consider that the database quality is essential for creating the model and that the inventory list of active landslides used in this study needs to be completed in order to successfully validate the BSA model in a similar way with the validation of the logistic model performed at this point.

However, the better validation results given by the BSA model (0.98), as compared to the 0.86 value resulted from the logistic model, indicates a better model fit of the BSA model. This fact is explained by the use within the BSA model of input data consisting of all the active digitised landslides which were also used to determine the landslide density for each of the existing classes of the variables, namely their statistical value. This can be analysed from a two-point perspective: it can be seen as an advantage when evaluating the ability of the model to correctly determine the existence or inexistence of the phenomenon, although with a slight overestimation of the results, and it can be seen as a disadvantage when a prediction is desired, just like in the case of the present study.
References


### Table 1. Database structure.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Database</th>
<th>Structure type</th>
<th>Source/resolution</th>
<th>Database type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Contour lines</td>
<td>vector</td>
<td>Topographic maps, 1 : 25 000</td>
<td>primary</td>
</tr>
<tr>
<td>2.</td>
<td>DEM</td>
<td>Raster (grid)</td>
<td>20 m</td>
<td>modelled</td>
</tr>
<tr>
<td>3.</td>
<td>Slope</td>
<td>Raster (grid)</td>
<td>degrees</td>
<td>derived</td>
</tr>
<tr>
<td>4.</td>
<td>Lithology</td>
<td>vector</td>
<td>Geological map, 1 : 200 000</td>
<td>primary</td>
</tr>
<tr>
<td>5.</td>
<td>Aspect</td>
<td>Raster (grid)</td>
<td>20 m</td>
<td>derived</td>
</tr>
<tr>
<td>6.</td>
<td>Drainage Density</td>
<td>Raster (grid)</td>
<td>m km$^{-1}$</td>
<td>derived</td>
</tr>
<tr>
<td>7.</td>
<td>Drainage Depth</td>
<td>Raster (grid)</td>
<td></td>
<td>derived</td>
</tr>
<tr>
<td>8.</td>
<td>Hydrological soil classes</td>
<td>Raster (grid)</td>
<td>Soil Map, 1 : 200 000</td>
<td>derived</td>
</tr>
<tr>
<td>9.</td>
<td>Distance to settlements</td>
<td>Raster (grid)</td>
<td>Derived from Ortofotoplans</td>
<td>derived</td>
</tr>
<tr>
<td>10.</td>
<td>Distance to roads</td>
<td>Raster (grid)</td>
<td>Derived from Ortofotoplans</td>
<td>derived</td>
</tr>
<tr>
<td>11.</td>
<td>Distance to hydrography</td>
<td>Raster (grid)</td>
<td>Derived from Ortofotoplans</td>
<td>derived</td>
</tr>
<tr>
<td>12.</td>
<td>Stream Power Index</td>
<td>Raster (grid)</td>
<td>20 m</td>
<td>modelled</td>
</tr>
<tr>
<td>13.</td>
<td>Profile curvature</td>
<td>Raster (grid)</td>
<td>20 m</td>
<td>derived</td>
</tr>
<tr>
<td>14.</td>
<td>Plan curvature</td>
<td>Raster (grid)</td>
<td>20 m</td>
<td>derived</td>
</tr>
<tr>
<td>15.</td>
<td>Compound Topografic Index (CTI)</td>
<td>Raster (grid)</td>
<td>20 m</td>
<td>modelled</td>
</tr>
<tr>
<td>16.</td>
<td>Precipitation data</td>
<td>Raster (grid)</td>
<td>Interpolation with a statistical model</td>
<td>modelled</td>
</tr>
<tr>
<td>17.</td>
<td>Seismicity</td>
<td>vector</td>
<td>Seismic zonation map, 1 : 200 000</td>
<td>primary</td>
</tr>
<tr>
<td>18.</td>
<td>Land use</td>
<td>vector</td>
<td>Ortofotoplans, 1 : 5000; Conversion – 20 m</td>
<td>primary</td>
</tr>
<tr>
<td>19.</td>
<td>Landslide areas</td>
<td>vector</td>
<td>Spot Images, orthophotograps, GPS points</td>
<td>derived</td>
</tr>
<tr>
<td>20.</td>
<td>Landslide probability map</td>
<td>Raster (Grid)</td>
<td>Conversion – 20 m</td>
<td>derived</td>
</tr>
</tbody>
</table>

Equations of spatial analysis (20 m resolution).
Table 2. Regression coefficients of the input variables. The bolded data represents the variables considered representatives.

<table>
<thead>
<tr>
<th>Regression coefficients</th>
<th>Coefficient symbols</th>
<th>Coefficient values</th>
<th>Probability (Odds difference)</th>
<th>Reference variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>−1.1381</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad leaved forests</td>
<td>Indse_8</td>
<td>−2.0400</td>
<td>−0.87 %</td>
<td>Indse_6</td>
</tr>
<tr>
<td>0 &lt; SPI &lt; 5</td>
<td>spi_1</td>
<td>−1.3942</td>
<td>−0.75 %</td>
<td>spi_2</td>
</tr>
<tr>
<td>201 m &lt; Distance to streams &lt; 400 m</td>
<td>dst_h5</td>
<td>1.1238</td>
<td>108 %</td>
<td>dst_h7</td>
</tr>
<tr>
<td>Northern aspect</td>
<td>as_10</td>
<td>−1.5113</td>
<td>−0.78 %</td>
<td>as_1</td>
</tr>
<tr>
<td>South-western aspect</td>
<td>as_7</td>
<td>1.3744</td>
<td>195 %</td>
<td>as_1</td>
</tr>
<tr>
<td>401 m &lt; Distance to roads &lt; 800 m</td>
<td>dst_dr6</td>
<td>0.9694</td>
<td>63 %</td>
<td>dst_dr8</td>
</tr>
<tr>
<td>Vineyards</td>
<td>Indse_3</td>
<td>−2.3552</td>
<td>−0.90 %</td>
<td>Indse_6</td>
</tr>
<tr>
<td>1.5 m km⁻² &lt; Drainage density &lt; 2 m km⁻²</td>
<td>dns_f4</td>
<td>1.0179</td>
<td>77 %</td>
<td>dns_f5</td>
</tr>
<tr>
<td>Southern aspect</td>
<td>as_6</td>
<td>0.8183</td>
<td>27 %</td>
<td>as_1</td>
</tr>
<tr>
<td>10,1° &lt; Panta &gt; 15°</td>
<td>slop_4</td>
<td>0.7655</td>
<td>15 %</td>
<td>slop_1</td>
</tr>
<tr>
<td>Average precipitation = 650 mm year⁻¹</td>
<td>pp_2</td>
<td>0.8281</td>
<td>29 %</td>
<td>pp_1</td>
</tr>
<tr>
<td>801 &lt; Distance to roads &lt; 1600</td>
<td>dst_dr7</td>
<td>−0.7583</td>
<td>−0.53 %</td>
<td>dst_dr8</td>
</tr>
<tr>
<td>801 &lt; Distance to settlements &lt; 1600</td>
<td>dst_lc7</td>
<td>0.8739</td>
<td>40 %</td>
<td>dst_lc8</td>
</tr>
</tbody>
</table>
Table 3. Spatial distribution of susceptibility classes.

<table>
<thead>
<tr>
<th>Susceptibility class</th>
<th>Statistical value</th>
<th>Area (km²)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Very low</td>
<td>0–0.128</td>
<td>21.489</td>
<td>24.70</td>
</tr>
<tr>
<td>2. Low</td>
<td>0.128–0.306</td>
<td>23.116</td>
<td>26.57</td>
</tr>
<tr>
<td>3. Medium</td>
<td>0.306–0.528</td>
<td>19.594</td>
<td>22.52</td>
</tr>
<tr>
<td>4. High</td>
<td>0.528–0.749</td>
<td>13.26</td>
<td>15.24</td>
</tr>
<tr>
<td>5. Very high</td>
<td>0.749–0.990</td>
<td>9.528</td>
<td>10.95</td>
</tr>
</tbody>
</table>
Table 4. Spatial distribution of susceptibility classes.

<table>
<thead>
<tr>
<th>Susceptibility class</th>
<th>Statistical value</th>
<th>Area (km²)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Very low</td>
<td>−6.727 . . . −3.231</td>
<td>4.410</td>
<td>5.07</td>
</tr>
<tr>
<td>2. Low</td>
<td>−3.231 . . . −1.743</td>
<td>9.353</td>
<td>10.76</td>
</tr>
<tr>
<td>3. Medium</td>
<td>−1.743 . . . −0.516</td>
<td>16.372</td>
<td>18.83</td>
</tr>
<tr>
<td>4. High</td>
<td>−0.516 . . . 0.524</td>
<td>28.486</td>
<td>32.76</td>
</tr>
<tr>
<td>5. Very high</td>
<td>0.524 . . . 2.756</td>
<td>28.330</td>
<td>32.58</td>
</tr>
<tr>
<td>Criterion/symbol</td>
<td>Variable classes</td>
<td>Statistical value (BSA)</td>
<td>Regression coefficients (Logistic Regression)</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>1. ELEVATION</td>
<td>Mde_1 338–400 m</td>
<td>−0.306</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Mde_2 401–500 m</td>
<td>0.135</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Mde_3 501–600 m</td>
<td>0.008</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Mde_4 601–700 m</td>
<td>0.018</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Mde_5 701–800 m</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mde_6 801–900 m</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Mde_7 901–1000 m</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Mde_8 1001–1081 m</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>2. ASPECT</td>
<td>As_1 Horizontal</td>
<td>−0.015</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>As_2 N</td>
<td>0.075</td>
<td>−1.511</td>
</tr>
<tr>
<td></td>
<td>As_3 NE</td>
<td>0.215</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>As_4 E</td>
<td>0.047</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>As_5 SE</td>
<td>−0.123</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>As_6 S</td>
<td>0.147</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td>As_7 SV</td>
<td>0.308</td>
<td>1.374</td>
</tr>
<tr>
<td></td>
<td>As_8 V</td>
<td>−0.828</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>As_9 NV</td>
<td>0.055</td>
<td>−</td>
</tr>
<tr>
<td>3. SLOPE ANGLE</td>
<td>Slop_1 0–2 °</td>
<td>−0.216</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Slop_2 2.1–5 °</td>
<td>−0.402</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Slop_3 5.1–10 °</td>
<td>−0.106</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Slop_4 10.1–15 °</td>
<td>0.264</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>Slop_5 15.1–20 °</td>
<td>0.209</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Slop_6 20.1–25 °</td>
<td>0.14</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Slop_7 25.1–30.4 °</td>
<td>−0.789</td>
<td>0</td>
</tr>
<tr>
<td>4. DRAINAGE DENSITY</td>
<td>Dns_f1 0.1–0.5 m km⁻²</td>
<td>0.35</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Dns_f2 0.5–1 m km⁻²</td>
<td>0.249</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dns_f3 1.1–1.5 m km⁻²</td>
<td>−0.328</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Dns_f4 1.5–2 m km⁻²</td>
<td>0.728</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td>Dns_f5 2.1–2.51 m km⁻²</td>
<td>0.001</td>
<td>−</td>
</tr>
<tr>
<td>5. DRAINAGE DEPTH</td>
<td>Ad_f1 &lt; 50 m</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Ad_f2 51–100 m</td>
<td>−0.0001</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Ad_f3 101–150 m</td>
<td>0.026</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Ad_f4 151–200 m</td>
<td>0.055</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Ad_f5 201–255 m</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>6. HYDROLOGICAL SOIL CLASSES</td>
<td>Gr_sol1 A</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Gr_sol2 B</td>
<td>0.039</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Gr_sol3 C</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Gr_sol4 D</td>
<td>−0.041</td>
<td>−</td>
</tr>
<tr>
<td>7. DISTANCE TO SETTLEMENTS</td>
<td>Dist lc1 0–25 m</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dist lc2 26–50 m</td>
<td>−1.401</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dist lc3 51–100 m</td>
<td>−0.394</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Dist lc4 101–200 m</td>
<td>−0.268</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Dist lc5 201–400 m</td>
<td>−0.096</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Dist lc6 401–800 m</td>
<td>0.003</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Dist lc7 801–1600 m</td>
<td>0.225</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>Dist lc8 1601–3200 m</td>
<td>−0.186</td>
<td>−</td>
</tr>
</tbody>
</table>
Table 5. Continued.

<table>
<thead>
<tr>
<th>Criterion/symbol</th>
<th>Variable classes</th>
<th>Statistical value (BSA)</th>
<th>Regression coefficients (Logistic Regression)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. DISTANCE TO STREAMS</td>
<td>Dst_h1 0–25 m m</td>
<td>-0.694</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_h2 26–50 m m</td>
<td>-0.419</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dst_h3 51–100 m</td>
<td>-0.216</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_h4 101–200 m</td>
<td>-0.009</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_h5 201–400 m</td>
<td>0.127</td>
<td>1.123</td>
</tr>
<tr>
<td></td>
<td>Dst_h6 401–800 m</td>
<td>0.025</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_h7 801–1600 m</td>
<td>-0.108</td>
<td>-</td>
</tr>
<tr>
<td>9. LITHOLOGY</td>
<td>Lit_1 Conglomerates</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lit_2 Marly clays, gravel</td>
<td>0.078</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Lit_3 Gravel, sand</td>
<td>-0.495</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Lit_4 Marly clays, gravel</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>10. LAND USE</td>
<td>Lnduse_1 Urban and rural area</td>
<td>-0.823</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lnduse_2 Predominantly agricultural areas</td>
<td>-0.02</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lnduse_3 Vineyards</td>
<td>-0.158</td>
<td>-2.355</td>
</tr>
<tr>
<td></td>
<td>Lnduse_4 Orchards</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Lnduse_5 Pastures</td>
<td>0.376</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Lnduse_6 Areas with complex use</td>
<td>0.358</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lnduse_7 Heterogeneous agricultural territories</td>
<td>0.125</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lnduse_8 Broad leaved forests</td>
<td>-0.683</td>
<td>-2.040</td>
</tr>
<tr>
<td></td>
<td>Lnduse_9 Coniferous forests</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lnduse_10 Natural pastures</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Lnduse_11 Bush transit areas</td>
<td>-0.61</td>
<td>-</td>
</tr>
<tr>
<td>11. CTI</td>
<td>Cti_1 0–5</td>
<td>-0.109</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cti_2 5...10</td>
<td>0.053</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cti_3 10...15</td>
<td>-0.14</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cti_4 15...17</td>
<td>-0.384</td>
<td>-</td>
</tr>
<tr>
<td>12. STI</td>
<td>Spi_1 0–5</td>
<td>-0.443</td>
<td>-1.394</td>
</tr>
<tr>
<td></td>
<td>Spi_2 5...10</td>
<td>0.157</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Spi_3 10...15</td>
<td>-0.031</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Spi_4 15...21</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>13. DISTANCE FROM ROADS</td>
<td>Dst_dr1 0–25</td>
<td>-1.147</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_dr2 26–50</td>
<td>-1.319</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dst_dr3 51–100</td>
<td>0.085</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_dr4 101–200</td>
<td>-0.663</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_dr5 201–400</td>
<td>-0.064</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dst_dr6 401–800</td>
<td>0.18</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>Dst_dr7 801–1600</td>
<td>-0.062</td>
<td>-0.758</td>
</tr>
<tr>
<td></td>
<td>Dst_dr8 1601–3200</td>
<td>0.26</td>
<td>-</td>
</tr>
<tr>
<td>14. AVERAGE PRECIPITATION</td>
<td>Pp1 525</td>
<td>0.206</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Pp2 650</td>
<td>-0.118</td>
<td>0.828</td>
</tr>
<tr>
<td>15. PLAN CURVATURE</td>
<td>Crb_pl1 1.64</td>
<td>-0.007</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Crb_pl2 0–2.24</td>
<td>0.011</td>
<td>-</td>
</tr>
<tr>
<td>16. PROFILE CURVATURE</td>
<td>Crb_pr1 0–0.31</td>
<td>-0.524</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Crb_pr2 0.31–2.3</td>
<td>0.083</td>
<td>0</td>
</tr>
</tbody>
</table>

0 – excluded classes due to low sample size.
0 (bold) – excluded classes due to lack of statistical significance.
Bold values represent the classes included in the model due to their statistical significance.
The italic values (ex. -0.758) are used as reference classes due to their vast spatial expansion in the study area.
Figure 1. Geomorphological map of the Small Niraj catchment and geographical position of the study area (1 – flood plain, 2 – slopes and connecting surfaces, 3 – slopes with complex modelling, 4 – active landslides, 5 – permanent hydrographic network, 6 – temporary hydrographic network, 7 – watershed divide, 8 – settlements).
Figure 2. Applied methodological flow chart.
Figure 3. Landslide susceptibility map generated using the logistic model.
Figure 4. Area under the ROC curve for the training data (left panel) and the testing data (right panel).
Figure 5. Landslide susceptibility map generated using the BSA model.
Figure 6. Percentage distribution of active landslide on the probability classes and ROC curve value.
Figure 7. Regional differences of susceptibility classes obtained through BSA model or by applying logistic model.
Figure 8. Comparative percentage distribution on susceptibility classes obtained by applying BSA model (a) and logistic model (b).