~~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)

Sanda, Roşca[1], Ştefan, Bilaşco[1,2], Dănuţ, Petrea[1], Ioan Fodorean[1] , Iuliu Vescan[1] , Sorin Filip [1] & Flavia–Luana Măguţ[1]

[1]"Babeş-Bolyai" University, Faculty of Geography, 400006 Cluj-Napoca, Romania, rosca_sanda@yahoo.com, sbilasco@geografie.ubbcluj.ro dpetrea@geografie.ubbcluj.ro, fioan@geografie.ubbcuj.ro, vescan@geografie.ubbcluj.ro, sfilip@geografie.ubbcluj.ro, luana.magut@gmail.com

[2]Romanian Academy, Cluj-Napoca Subsidiary Geography Section, 9, Republicii Street, 400015, Cluj-Napoca, Romania

Correspondence to: Sanda, ROȘCA (rosca_sanda@yahoo.com)

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).

Keywords: Landslide modelling, Logistic regression, BSA, GIS database, GIS modelling, comparation

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; Bilașco et. al 2011; Năsui and Petreuș 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; Roșca 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; Dai and Lee 2002; Ayalew and Yamagishi 2005; Lee 2005; 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 ~~y~~ 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.

Fig. 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 modellation, 4 – active landslides, 5 – permanent hydrographic network, 6 – temporary hydrographic network, 7 – watershed divide, 8 – settlements)

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 in the last two years: 67 persons, 45 houses, 115 hectares of land and a country road were affected by landslides. The catchment area is found between 24°47´52" and 24°58´32" eastern longitude and 46°30´53" and 46°37´42" northern latitude, totalizing an area of 68 km^2 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/year, 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).

Table 1: Database structure

Nr.	Database	Structure type	Source/resolution	Database type
1.	Contour lines	vector	Topographic maps, 1:25.000	primary
2.	DEM	Raster (grid)	20 m	modelled
3.	Slope	Raster (grid)	degrees	derived
4.	Lithology	vector	Geological map, 1:200000	primary
		Raster (Grid)	Conversion – 20 m	derived
5.	Aspect	Raster (grid)	20 m	derived
6.	Drainage Density	Raster (grid)	m/km	derived
7.	Drainage Depth	Raster (grid)	m	derived
8.	Hydrological soil classes	Raster (grid)	Soil Map, 1:200000	derived
9.	Distance to settlements	Raster (grid)	Derived from Ortofotoplans	derived
10	Distance to roads	Raster (grid)	Derived from Ortofotoplans	derived
11.	Distance to hydrography	Raster (grid)	Derived from Ortofotoplans	derived
12.	Stream Power Index	Raster (grid)	20 m	modelled
13.	Profile curvature	Raster (grid)	20 m	derived
14.	Plan curvature	Raster (grid)	20 m	derived
15.	Compound Topografic Index (CTI)	Raster (grid)	20 m	modelled
16.	Precipitation data	Raster (grid)	Interpolation with a statistical model	modelled
17.	Seismicity	vector	Seismic zonation map, 1:200000	primary
		Raster (Grid)	Geological map, 1:200000	derived
18.	Land use	vector	Ortophotoplans, 1:5000; Conversion – 20 m	primary
		Raster (Grid)	Conversion – 20 m	derived
19.	Landslide areas	vector	Spot Images, orthophotograps, GPS points	primary derived
		Raster (Grid)	Conversion – 20 m	derived

20.	Landslide probability map	Raster (Grid)	Equations of spatial analysis (20 m resolution)	modelled

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:

$$I_{i=log}\frac{Si/_{Ni}}{S/_{N}}, (1)$$

where:
Ii = Statistical value of the analysed factor
Si = Area affected by landslides for the analysed variable
Ni = Area of the analysed variable
S = Total landslide area in the analysed basin
N = Area of the analysed basin

By using formula (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: Ώ 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 representhed 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:

$$P = \frac{1}{1+e^{-z}}, \quad (2)$$

where:
$Z = β0 + β1X1 + … + βnXn$,
X1...Xn – preparing and triggering factors
β0 – constant ,
β1... βn - 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 (-∞, +∞) (Thiery 2007, cited by Jurchescu 2013). The main methodological stages are described in Fig. 2.

Fig. 2: Applied methodological flow chart

The Ω study area was divided into two random sub-categories: Ώ$_1$ and Ώ$_0$. Hence, 500 points were used in the modelling process, 250 points generated at a minimum distance of 60 metres in the landslide areas and 250 points at a minimum distance of 80 meters 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 meters (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 (X1, …Xn) of each landslide factor were necessary in order to determine the probability of landslide occurence for each pixel, these coefficients being considered as representative for Ώ1 and Ώ0. 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 β0 …βn 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 X1 parameter in the input database and the model which includes in its input database the Xn parameter. The variables with the highest influence were identified with the help of the AIC criterium 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 (Chiţu 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:

fit3 = glm(alunec ~ lndse_8 + spi_1 + dst_h5 + as_10 + as_7 + dst_dr6 + lndse_3 + dns_f4 + as_6 + slop_4 + pp_2 + dst_dr7 + dst_lc7, family = binomial, data = model_df2) (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/km2 (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.

Table 2: Regression coefficients of the input variables

Regression coefficients	**Coefficient symbols**	**Coefficient values**	**Probability (Odds difference)**	**Reference variable**
Constant	**-1.1381**			
Broad leaved forests	*lndse_8*	-2.0400	-0.87%	lndse_6
0 < SPI < 5	*spi_1*	-1.3942	-0.75%	spi_2
201 m < Distance to streams < 400 m	*dst_h5*	*1.1238*	108%	dst_h7
Northern aspect	*as_10*	-1.5113	-0.78%	as_1
South-western aspect	*as_7*	*1.3744*	195%	as_1
401 m < Distance to roads < 800 m	*dst_dr6*	*0.9694*	63%	dst_dr8
Vineyards	*lndse_3*	-2.3552	-0.90%	lndse_6
1.5 m/km^2 < Drainage density < 2 m/km^2	*dns_f4*	*1.0179*	77%	dns_f5
Southern aspect	*as_6*	*0.8183*	27%	as_1
10,1° < Panta > 15°	*slop_4*	*0.7655*	15%	slop_1
Average precipitation = 650 mm/year	*pp_2*	*0.8281*	29%	pp_1
801 < Distance to roads < 1600	*dst_dr7*	-0.7583	-0.53%	dst_dr8
801 < Distance to settlements < 1600	*dst_lc7*	*0.8739*	40%	dst_lc8

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/year, 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 formula (4):

Mdl_fit3 = exp(-1.1381 + -2.0400 * [lndse_8] + -1.3942 * [spi_1] + 1.1238 * [dst_h5] + -1.5113 * [as_10] + 1.3744 * [as_7] + 0.9694 * [dst_dr6] + -2.3552 * [lndse_3] + 1.0179 * [dns_f4] + 0.8183 * [as_6] + 0.7655 * [slop_4] + 0.8281 * [pp_2] + -0.7583 * [dst_dr7] + 0.8739 * [dst_lc7]) (4)

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

$$S = p/(1-p), \qquad (5)$$

where S - susceptibility, P – probability

Fig. 3: Landslide susceptibility map generated using the logistic model

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).

Fig. 4: Area under the ROC curve for the training data (left) and the testing data (right)

Table 3: Spatial distribution of susceptibility classes

	Susceptibility class	Statistical value	Area	
			(km^2)	%
1.	Very low	0 – 0.128	21.489	24.70
2.	Low	0.128 – 0.306	23.116	26.57
3.	Medium	0.306 – 0.528	19.594	22.52
4.	High	0.528 – 0.749	13.26	15.24
5.	Very high	0.749 – 0.990	9.528	10.95

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).

Fig. 5: Landslide susceptibility map generated using the BSA model

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).

Fig. 6: Percentage distribution of active landslide on the probability classes and ROC curve value

Table 4: Spatial distribution of susceptibility classes

	Susceptibility class	Statistical value	Area

			(km^2)	%
1.	Very low	-6.727...-3.231	4.410	5.07
2.	Low	-3.231...-1.743	9.353	10.76
3.	Medium	-1.743...-0.516	16.372	18.83
4.	High	-0.516...0.524	28.486	32.76
5.	Very high	0.524...2.756	28.330	32.58

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).

Fig. 7: Regional differences of susceptibility classes obtained through BSA model or by applying logistic model

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.

Fig. 8: Comparative percentage distribution on susceptibility classes obtained by applying BSA model (8.A) and logistic model (8.b)

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/km2, 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.

Table 5: Comparative statistical values (for BSA and logistic regression)

Criterion /*symbol*		**Variable classes**	**Statistical value (BSA)**	**Regression coefficients (Logistic Regression)**
1. ELEVATION	*Mde_1*	338 – 400 m	-0.306	-
	Mde_2	401-500 m	0.135	-
	Mde_3	501-600 m	0.008	-
	Mde_4	601-700 m	0.018	-
	Mde_5	701-800 m	0	0
	Mde_6	801-900 m	0	-
	Mde_7	901-1000 m	0	-
	Mde_8	1001-1081 m	0	-
2. ASPECT	*As_1*	Horizontal	-0.015	
	As_2	N	0.075	*-1.511*
	As_3	NE	0.215	-
	As_4	E	0.047	-
	As_5	SE	-0.123	-
	As_6	S	0.147	*0.818*
	As_7	SV	0.308	*1.374*
	As_8	V	-0.828	-
	As_9	NV	0.055	-
3. SLOPE ANGLE	*Slop_1*	0-2 °	-0.216	-
	Slop_2	2.1-5 °	-0.402	-
	Slop_3	5.1-10 °	-0.106	-
	Slop_4	10.1-15 °	0.264	*0.765*
	Slop_5	15.1-20 °	0.209	-
	Slop_6	20.1-25 °	0.14	-
	Slop_7	25.1-30.4 °	-0.789	0
4. DRAINAGE DENSITY	*Dns_f1*	0.1-0.5 m/km^2	0.35	-
	Dns_f2	0.5-1 m/km^2	0.249	**0**
	Dns_f3	1.1-1.5 m/km^2	-0.328	-
	Dns_f4	1.5-2 m/km^2	0.728	*1.017*
	Dns_f5	2.1-2.51 m/km^2	0.001	-
5. DRAINAGE DEPTH	*Ad_f1*	<50 m	0	-
	Ad_f2	51-100 m	-0.0001	**0**
	Ad_f3	101-150 m	0.026	-
	Ad_f4	151-200 m	0.055	-
	Ad_f5	201-255 m	0	-
6. HYDROLOGICAL SOIL CLASSES	*Gr_sol1*	A	0	-
	Gr_sol2	B	0.039	-
	Gr_sol3	C	0	-
	Gr_sol4	D	-0.041	-

7. DISTANCE TO SETTLEMENTS	*Dst_lc1*	0-25 m	0	0
	Dst_lc2	26-50 m	-1.401	0
	Dst_lc3	51-100 m	-0.394	-
	Dst_lc4	101-200 m	-0.268	-
	Dst_lc5	201-400 m	-0.096	-
	Dst_lc6	401-800 m	0.003	-
	Dst_lc7	801-1600 m	0.225	*0.873*
	Dst_lc8	1601-3200 m	-0.186	-
8. DISTANCE TO STREAMS	*Dst_h1*	0-25m m	-0.694	-
	Dst_h2	26-50 m	-0.419	**0**
	Dst_h3	51-100 m	-0.216	-
	Dst_h4	101-200 m	-0.009	-
	Dst_h5	201-400 m	0.127	*1.123*
	Dst_h6	401-800 m	0.025	-
	Dst_h7	801-1600 m	-0.108	-
9. LITHOLOGY	*Lit_1*	Conglomerates	0	-
	Lit_2	Marly clays, gravel	0.078	0
	Lit_3	Gravel, sand	-0.495	0
	Lit_4	Marly clays, gravel	0	-
10. LAND USE	*Lnduse_1*	Urban and rural area	-0.823	-
	Lnduse_2	Predominantly agricultural areas	-0.02	-
	Lnduse_3	Vineyards	-0.158	*-2.355*
	Lnduse_4	Orchards	0	0
	*Lnduse5_*	Pastures	0.376	**0**
	Lnduse_6	Areas with complex use	0.358	-
	Lnduse_7	Heterogeneous agricultural territories	0.125	-
	Lnduse_8	Broad leaved forests	-0.683	*- 2.040*
	Lnduse_9	Coniferous forests	0	-
	Lnduse_10	Natural pastures	0	-
	Lnduse_11	Bush transit areas	-0.61	-
11. CTI	*Cti_1*	0-5	-0.109	-
	Cti_2	5…10	0.053	-
	Cti_3	10…15	-0.14	-
	Cti_4	15…17	-0.384	-
12. STI	*Spi_1*	0-5	-0.443	*-1.394*
	Spi_2	5…10	0.157	-
	Spi_3	10…15	-0.031	-
	Spi_4	15…21	0	-
13. DISTANCE FROM ROADS	*Dst_dr1*	0-25	-1.147	-
	Dst_dr2	26-50	-1.319	0
	Dst_dr3	51-100	0.085	-
	Dst_dr4	101-200	-0.663	-
	Dst_dr5	201-400	-0.064	-
	Dst_dr6	401-800	0.18	*0.969*
	Dst_dr7	801-1600	-0.062	*-0.758*
	Dst_dr8	1601-3200	0.26	-
14. AVERAGE PRECIPITATION	*Pp1*	525	0.206	-
	Pp2	650	-0.118	*0.828*
15. PLAN CURVATURE	*Crb_pl1*	-1.64	-0.007	-
	Crb_pl2	0-2,24	0.011	-
16. PROFILE CURVATURE	*Crb_pr1*	0-0,31	-0.524	-
	Crb_pr2	0,31-2,3	0.083	**0**

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.

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 all. 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.

4. 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

Ayalew L., Yamagishi H., Ugawa N. (2004) Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Nigata Prefecture, Japan. Landslides 1: 73-81.

Ayalew L., Yamagishi H. (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65: 15-31.

Akbari A., Yahaya F. B. M., Azamirad M., Fanodi M. (2014) Landslide Susceptibility Mapping Using Logistic Regression Analysis and GIS Tools. EJGE, 19:1987-1696.

Bălteanu D., Micu M. (2009) Landslide investigation: from morphodynamic mapping to hazard assessment.A case-study in the Romanian Subcarpathians: Muscel Catchment. Landslide Process from Geomorphologic Mapping to Dynamic Modelling, Malet et al (eds). CERG Edotions, Strasburg, France, 235–241.

Bilașco Șt, Horvath Cs., Roșian Gh., Filip S., Keller I.E. (2011) Statistical model using GIS for the assessment of landslide susceptibility. Case-study: the Someș Plateau. Rom J Geogr 2:91–111,

Chiţu Z. (2010) Spatio-temporal prediction of the landslide hazard using GIS techniques Case Study Sub-Carpathian area of Prahova Valley and Ialomita Valley PhD Thesys, Bucureşti (in romanian).

Chung C. F., Fabbri A. G., van Westen C. J. (1995) Multivariate Regression Analysis for Landslide Hazard Zonation. Geographical Information Systems in Assessing Natural Hazards, editors Carrara, A., Guzzetti, F., Kluwer Academic Publisher, Dordrecht, 107–134.

Chung C.-J.F., Fabbri A.G. (1999) Probalistic prediction models for landslide hazard mapping. Photogrammetric Engineering and Remote Sensing, 65-12: 1389-1399.

Chung C.-J.F., Fabbri A.G. (2003) Validation of spatial prediction models for landslide hazard mapping. Natural Hazards, 30: 451-472.

Chung C.-J.F., Fabbri A.G. (2008) Predicting landslides for risk analysis – Spatial models tested by a cross-validation technique. Geomorphology, 94: 438-452.

Crozier M.J., Glade T. (2005) Landslide Hazard and Risk: Issues, Concepts and Approach. Landslide Hazard and Risk, Edited by Th. Glade, M. Anderson, M J. Crozier, John Wiley & Sons, Ltd, 1-38.

Cuesta M., Jiménez-Sánchez M., Colubi A., González-Rodríguez G. (2010) Modelling shallow landslide susceptibility: a new approach in logistic regression by using favourability assessment. Earth and Science, 99: 661-674.

Dai F., C., Lee C.F., Ngai Y.Y. (2002) Landslide risk assessment and management: an overview. Engineering Geology, 64: 65-87.

Dhakal A.S., Amada T., Aniya M. (2000) Landslide hazard mapping and its evaluation using GIS: An investigation of sampling schemes for a grid-cell based quantitative method. Photogrammetric Eng. & Remote Sensing, 66(8): 981–989.

Guzzetti F., Reichenbach P., Ardizzone F., Cardinali M., Galli M. (2006) Estimating the quality of landslide susceptibility models. Geomorphology, 81: 166–184.

Hosmer D.W., Lemeshow S. (2000) Applied logistic regression, 2th edition, John Wiley & Sons, New York, 392 p.

Guns M., Vanacker V. (2012) Logistic regression applied to natural hazards: rare event logistic regression with replications. Natural Hazards Earth Syst. Sci., 12: 1937-1947.

Guzzetti F. (2006) Landslide hazard and risk assessment, http://hss.ulb.uni-bonn.de/2006/0817/0817.htm

Harrell, F.E. Jr. (2001) Regression Modeling Strategies, available online http://biostat.mc.vanderbilt.edu/wiki/pub/Main/RmS/rms.pdf

Hilbe J. (2009) Logistic regression models, CRC Press INC, 637 pp.

Jade S., Sarkar S. (1993) Statistical model for slope instability classification. Eng Geol.36:71-98.

Jurchescu M. (2013) Olteț morpho-hydrographic basin. Study of applied geomorphology, PhD Thesys, Bucureşti (in Romanian).

Kleinbaum D.G., Klein M. (2002) Logistic Regression A self-learning text, 2nd edition, Springer, Springer Science&Business Media.

Lee S. (2005) Cross-Verification of Spatial Logistic Regression for Landslide Susceptibility Analysis: A Case Study of Korea, A Geosciences Information Center, Korea Institute of Geoscience and Mineral Resources, 30: 305-350.

Magliulo P., Di Lisio A., Russo F., Zelano A. (2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistic: a case study in southern Italy. Nat Hazards 47:411–435.

Mancini F., Ceppi C., Ritrovato G. (2010) GIS and statistical analysis for landslide susceptibility mapping in the Daunia area, Italy. Natural Hazards Earth Syst. Sci., 10: 1851-1864.

Măguţ F., L. (2013) Risk to landslide in Baia Mare Depression, PhD Thesys, Cluj-Napoca (in Romanian).

Mărgărint M.C., Grozavu A., Patriche C.V. (2013) Assessing the spatial variability of weights of landslide causal factors in different regions from Romania using logistic regression. Natural Hazards and Earth System Sciences Discussions, 1: 1749-1774.

Năsui D., Petreuș A. (2014) Landslide susceptibility assessment in the Irigșu Glacis (Baia Mare City, Romania). Carpath J Earth Environ Sci 9:185–190.

Petrea D., Bilașco Șt., Roșca S., Vescan I., Fodorean I. (2014) The determination of the Landslide occurrence probability by spatial analysis of the Land Morphometric characteristics (case study: the Transylvanian Plateau). Carpath J Environ Sci 9:91–110.

Polemio M., Petrucci O. (2010) Occurrence of landslide events and the role of climate in the twentieth century in Calabria, southern Italy. Q J Eng Geol Hydrogeol 43:403–415. doi:10.1144/1470-9236/09-006

Remondo J., Gonzalez A., Diaz de Teran J.R., Cendrero A. (2003) Validation of landslide susceptibility maps, examples and applications from a case study in Northern Spain. Natural Hazards, 30: 437-449.

Roșca S., Bilașco Șt., Petrea D., Fodorean I., Vescan I., Filip S. (2015a) Application of landslide hazard scenarios at annual scale in the Niraj River basin (Transylvania Depression, Romania). Natural Hazards, 77:1573–1592, DOI 10.1007/s11069-015-1665-2.

Roșca S., Bilașco Șt., Petrea D., Vescan I., Fodorean I. (2015b) Comparative assessment of landslide susceptibility. Case study: the Niraj river basin (Transylvania depression, Romania). Geomatics, Natural Hazards and Risk, DOI: 10.1080/19475705.2015.1030784

Saha A.K, Gupta R.P., Arora M.K. (2002) GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. Int. Jour. of Remote sensing, 23(2): 357–369.

Sarkar S., Kanungo D.P. (2004) An integrated approach for landslide susceptibility mapping using remote sensing and GIS. Photogrammetric Engineering and Remote Sensing, 70(5): 617–625.

Van Westen, C.J., Rengers N., Soeters R. (2003) Use of geomorphological information in indirect landslide susceptibility assessment. Natural Hazards, 30: 399-419.

Van Westen C.J., Van Asch T.W.J., Soeters R. (2006) Landslide hazard and risk zonation—why is it still so difficult? Bull Eng Geol Environ 65:167–184. doi:10.1007/s10064-005-0023-0

Van den Eeckhaut M., Vanwalleghem T., Poeses J., Govers G., Verstraeten G., Vandekerckhove L. (2006) Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). Geomorphology, vol.76, 3-4: 392-410.

Van den Eeckhaut M., Hervas J., Jaedicke C., Malet J-P., Picarelli L. (2010) Calibration of logistic regression coefficients from limited landslide inventory data for European-wide landslide susceptibility modelling, In: Malet, J.-P., Glade, T., Casagli, N. (Eds.), Proc. Int. Conference Mountain Risks: Bringing Science to Society, Florence, Italy, 24-26 November 2010. CERG Editions, Strasbourg, pp. 515-521.

Wang L., Sawada K., Moriguchi S. (2011) Landslide Susceptibility Mapping by Using Logistic Regression Model with Neighborhood Analysis: A Case Study in Mizunami City. International Journal of Geomate, 1 (2): 99-104.

Zèzere J.L., Reis E., Garcia R., Oliveira,S., Rodrigues M.L., Vieira G., Ferreira A.B. (2004) Integration of spatial and temporal data for the definition of different landslide hazard scenarios in the Area North of Lisbon (Portugal). Natural Hazards and Earth System Sciences, European Geosciences Union, 4: 133-146. SRef-ID: 1684-9981/nhess/2004-4-133.

Yin K., Yan T.Z. (1988) Statistical prediction models for slope instability of metamorphosed rocks. In: Bonnard C. editor. Vol. 2. Proceedings of the 5th International Symposium on Landslides, Lausanne. 12-69.

***E.E.A. (2004) Impacts of Europe's changing climate - An indicator-based assessment. European Environment Agency Report, 2.

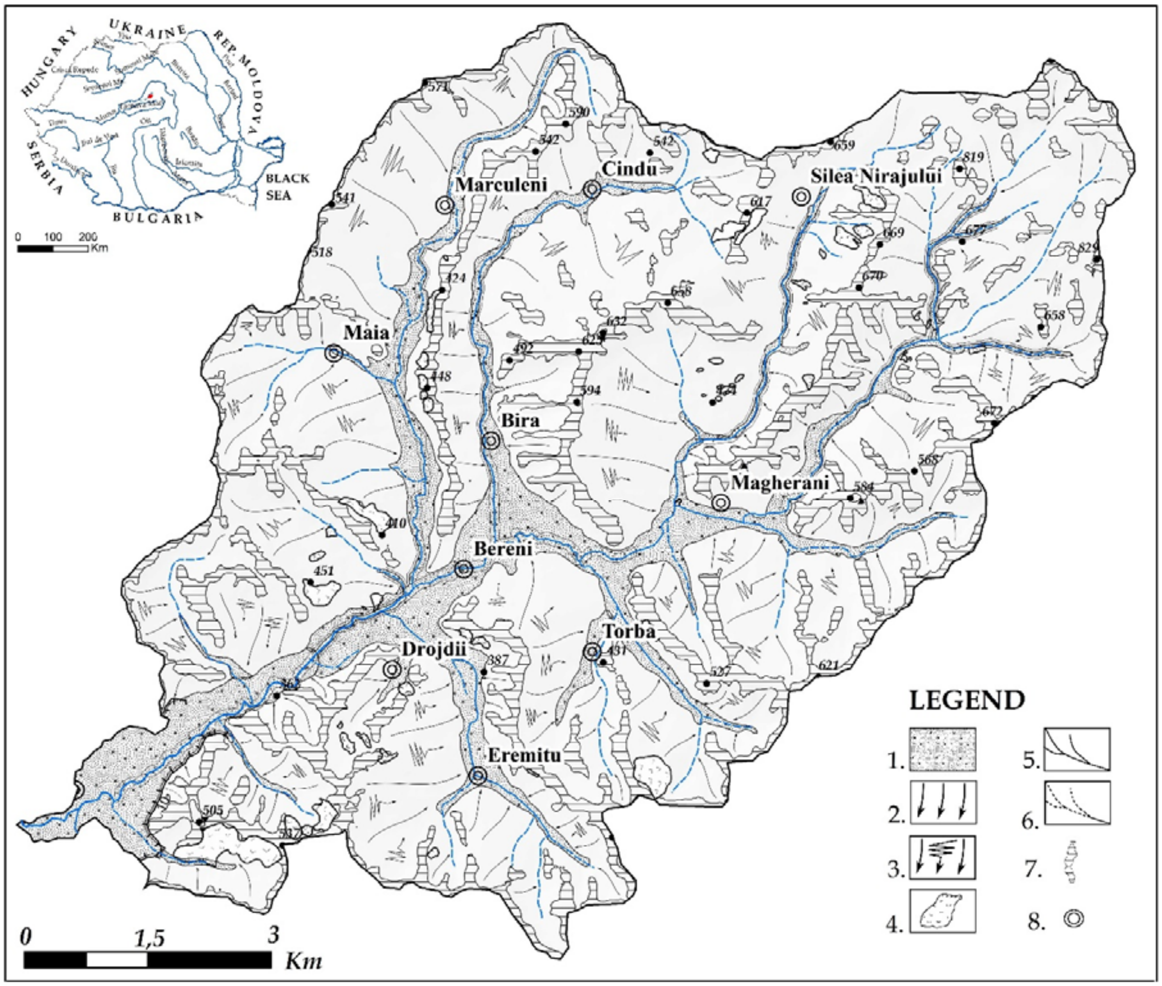

Fig. 1: Geomorphological map of the Small Niraj catchment and geographical position of the study area (1 – flood plain, 2 – slopes and connecting surfaces, 3 – slopes with complex modellation, 4 – active landslides, 5 – permanent hydrographic network, 6 – temporary hydrographic network, 7 – watershed divide, 8 – settlements)

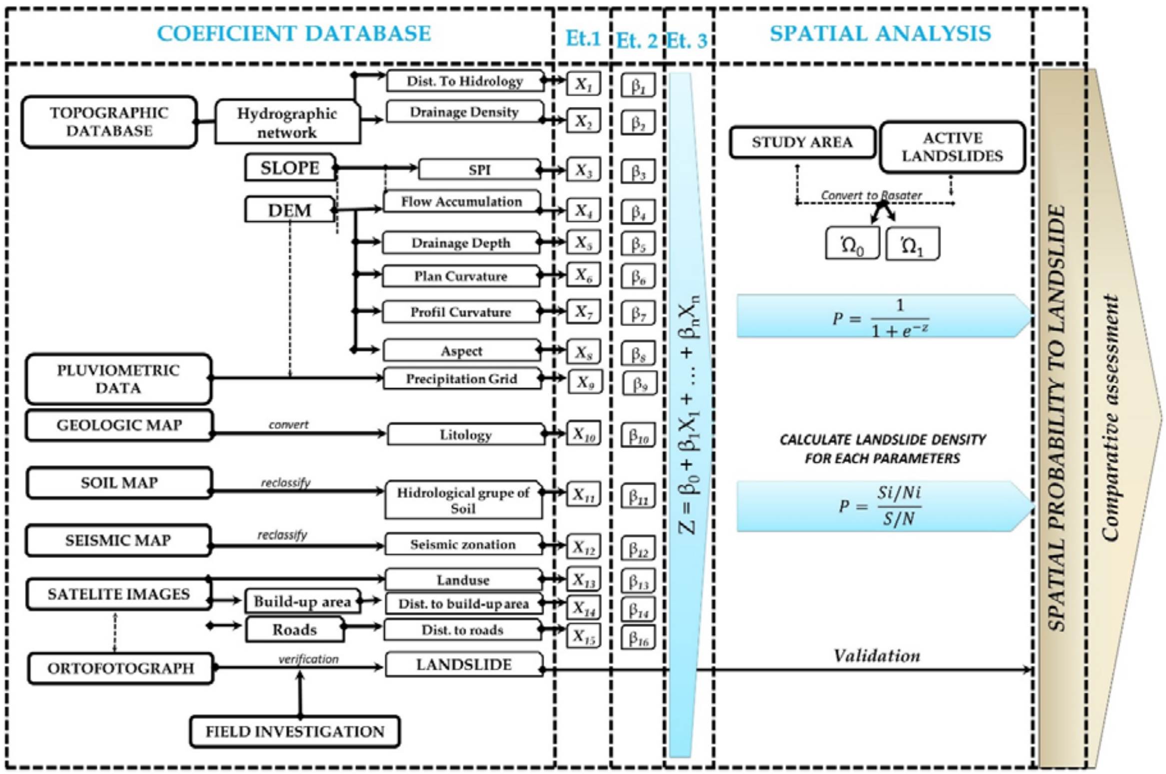

Fig. 2: Applied methodological flow chart

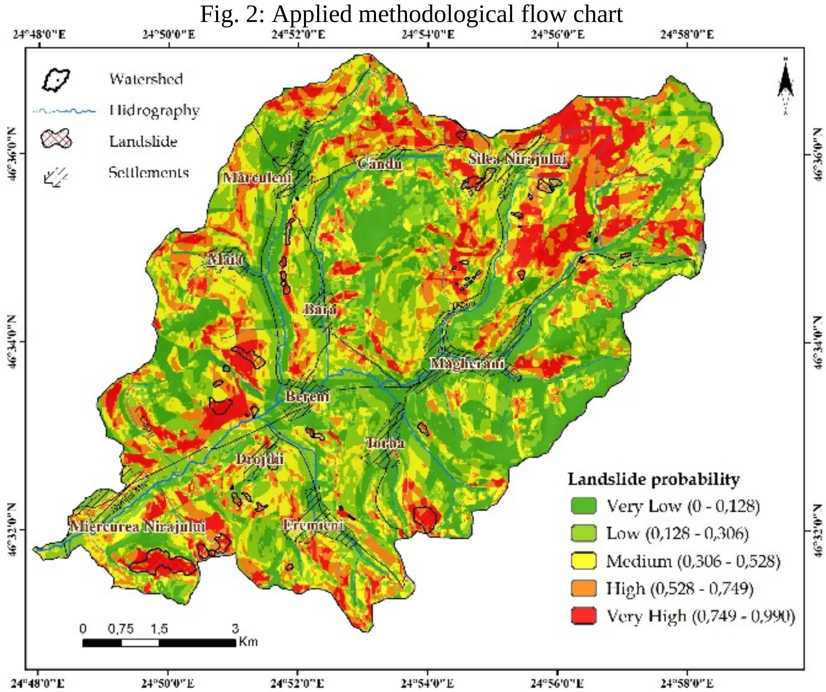

Fig. 3: Landslide susceptibility map generated using the logistic model

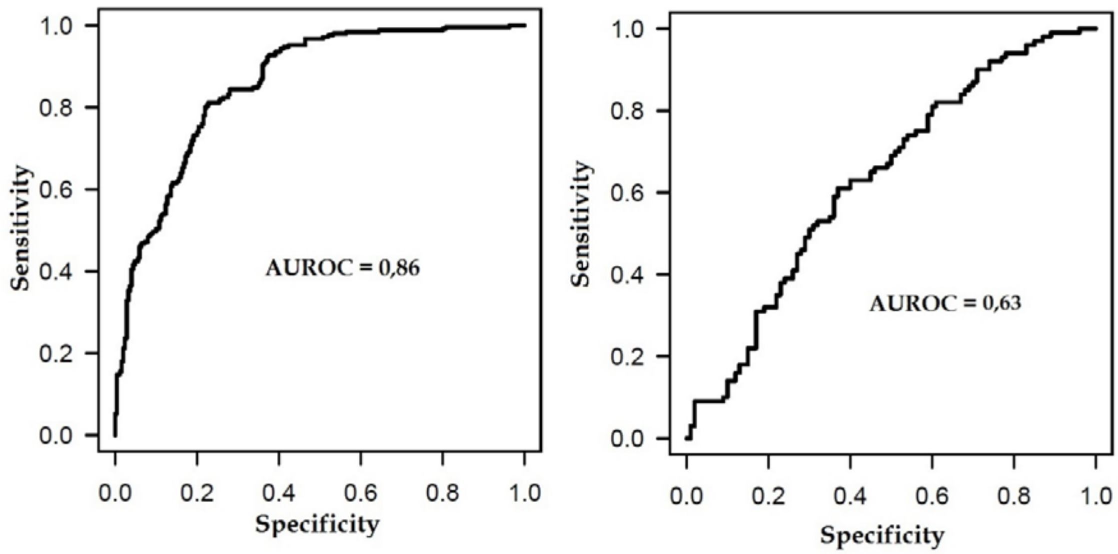

Fig. 4: Area under the ROC curve for the training data (left) and the testing data (right)

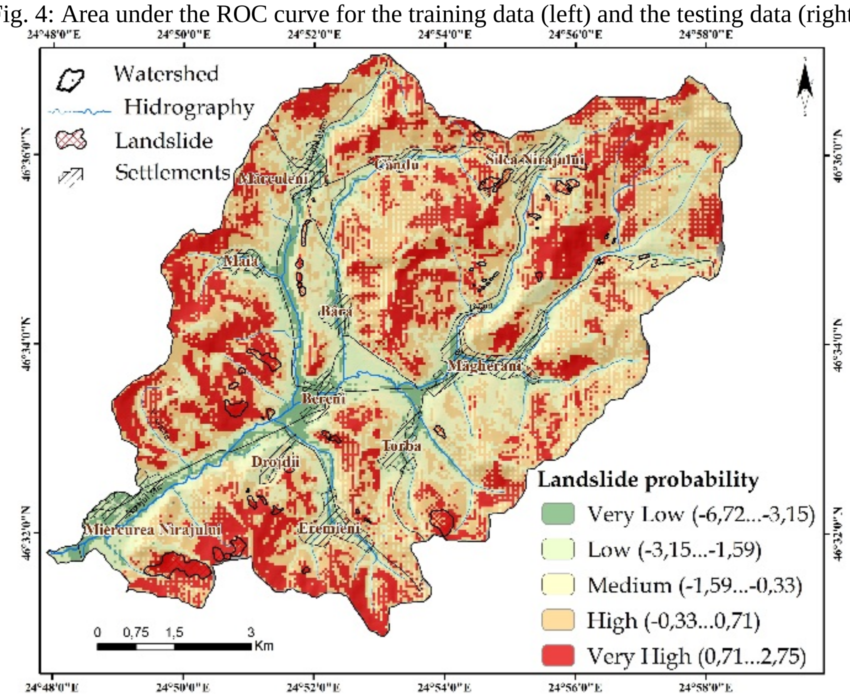

Fig. 5: Landslide susceptibility map generated using the BSA model

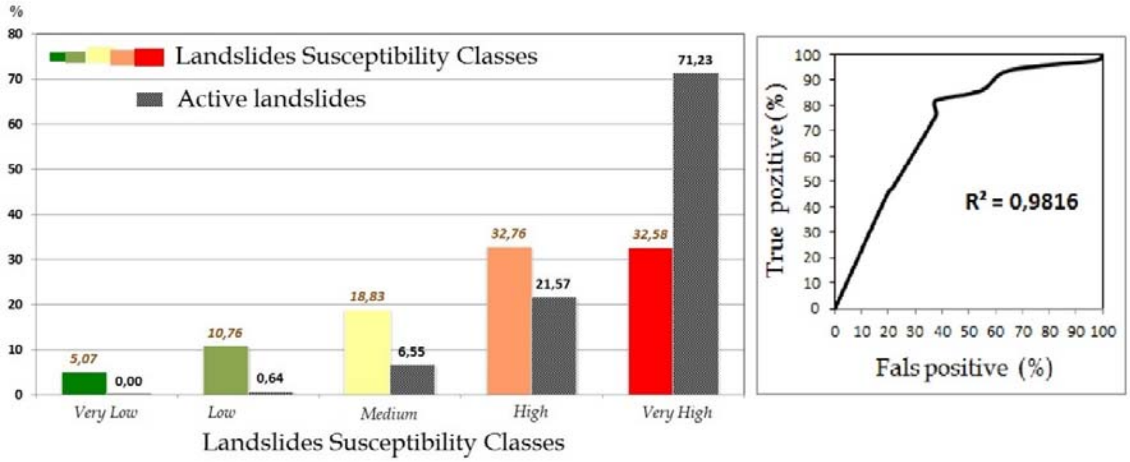

Fig. 6: Percentage distribution of active landslide on the probability classes and ROC curve value

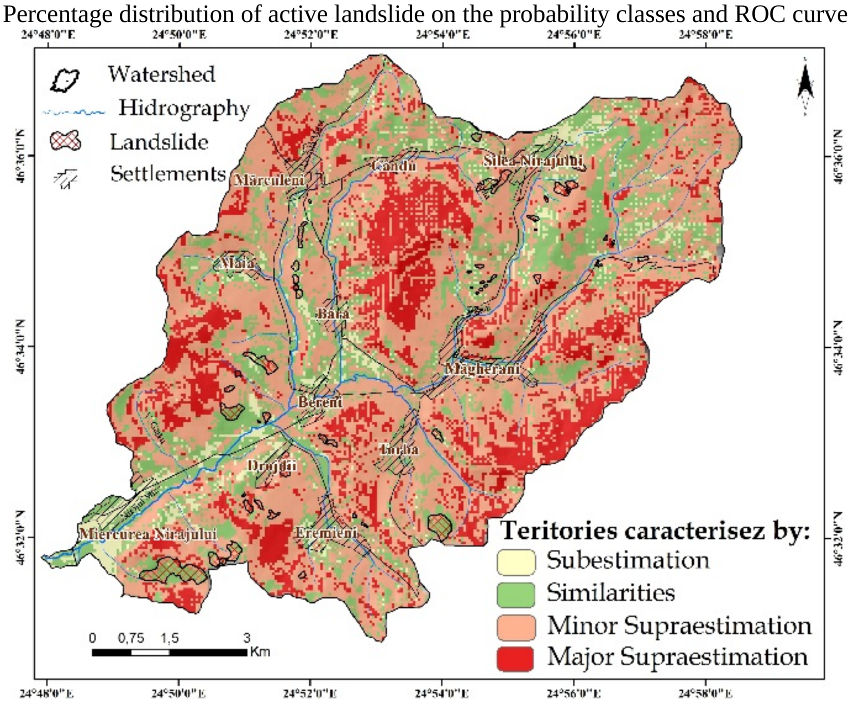

Fig. 7: Regional differences of susceptibility classes obtained through BSA model or by applying logistic model

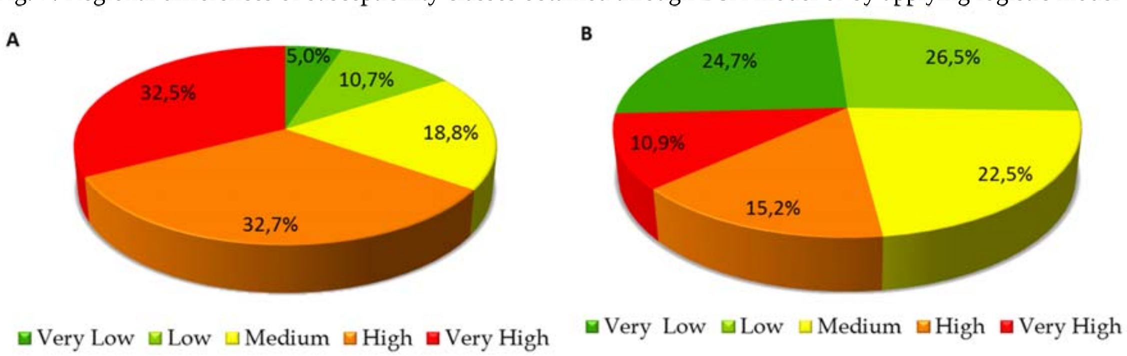

Fig. 8: Comparative percentage distribution on susceptibility classes obtained by applying BSA model (8.A) and logistic model (8.b)