GIS-based multivariate statistical analysis for landslide susceptibility zoning: a first validation on different areas of Liguria region (Italy)

Roberto Marzocchi[1,2], Agnese Rovegno[2], Bianca Federici[2], Rossella Bovolenta[2], Riccardo Berardi[2]

[1] Gter srl Innovazione in Geomatica Gnss e Gis, (University of Genoa Spin-Off company) – Via Greto di Cornigliano 6r, 16152 Genova (Italy)

[2] Dipartimento di Ingegneria Civile Chimica ed ambientale (DICCA), Università degli studi di Genova, Via Montallegro 1, 16145 Genova

Abstract

Several approaches have been developed worldwide in order to assess landslide hazards. The most diffuse methods are heuristic analysis, statistical or process-based methods (Fell et al., 2008). The present research consists in applying univariate and multivariate statistical analysis for landslide susceptibility zoning, considering the influence of various factors on the occurrence and triggering of slides and flows, and critically reviewing the choice of the calibration area in terms of extension and characteristic of influencing factors. The analysis was applied to different areas of Liguria region (Italy).

Keywords

Landslide susceptibility zoning, statistical methods, GRASS GIS

1. Introduction

Almost 10% of the Italian territory is affected by hydraulic and geological risks (floods and landslides) and the 6.9% of the territory is catalogued as landslide by the IFFI national inventory of landslide (ISPRA , 2013). Recently, various procedures have been developed for the analysis of spatial distribution of future landslides, at different scales and with different approaches.
As well described in literature (Varnes and IAEG 1984; Hutchinson 1995, Fell et al., 2008) the hazard analysis is based on the assumption that landslides occur in the same geological, geomorphological, hydrogeological and climatic conditions as in the past and are normally controlled by these identifiable physical factors. To assess landslide hazard several approaches have been developed worldwide; the main diffuse methods are heuristic analysis, statistical or process-based methods (Fell et al., 2008). Each method has its advantages and disadvantages, but in general the choice of the method depends on objective, scale of the study and finally also on the available data.

Heuristic methods are based on the judgement of the expert carrying out the analysis, both for the individuation of the perimeter of occurred landslides or for the creation of landslide susceptibility maps. Often these methods are based on geological and geomorphological criteria and use local observations which are collected in the studied area.
Instead, statistical methods usually correlate an inventory of landslides occurred in the past with factors which are supposed to be responsible of slope failure (Cascini, 2008). The statistical methods are considered more suitable to be applied to wide and differentiated zones and usually this approach uses GIS environment to integrate spatial variables; nevertheless statistical approaches sometimes require complex calculation procedures (Guzzetti et al., 1999).
Concerning the landslide risk management, in authors' opinion, the availability of an instrument able to perform hazard analysis on a wide area, quickly and with low resources, is really appreciable. Within the activities of a national research program, entitled "Mitigation of landslide risk using sustainable actions", we are analysing landslide susceptibility zoning through statistical methods, focusing the attention on the influence of various factors on the occurrence and triggering of landslides, and critically reviewing the choice of the calibration area in terms of extension and characteristic of influencing factors.
The type of landslides involved in this study are slides and flows, according to the UNESCO WP/WLI (1993). The used GIS software is GRASS 7.0.
On the basis of field experience and our previous research (Natali et al., 2010; Bovolenta et al, 2011) and on the basis of an accurate analysis of the pertinent literature, we have deduced that geomorphological, geological, anthropic and climatic parameters are the most important factors. Consequently, we have considered and analysed them from a statistical point of view in order to understand their influence on the landslide susceptibility and their spatial variability, calibrating the statistical model on the entire Liguria region (in North-western Italy) or on its four administrative provinces (i.e. Imperia, Savona, Genova, La Spezia).
Final purpose of the mentioned research is to develop a guideline for the landslide susceptibility zoning, giving an instrument to non-GIS and non-statistical expert users for the choice of factors to be considered.

2 Methods

Statistical methods usually correlate an inventory of landslides occurred in the past with factors which are supposed to be responsible of slope failure. Hence, an accurate analysis of the pertinent literature and field experiences were necessary to identify the main predisposing factors to landslides. Then a logistic multiple regression was chosen to analyse them.

2.1 Predisposing factors to landslide

Starting from a literature analysis (Dai & Lee,2002; Pauditš & Bednárik, 2002; Ayalew & Yamagishi, 2005; Lee, 2007; Dahal et al., 2008; Cencetti et al., 2010; Natali et al., 2010; Pradhan & Lee, 2010) we identified several predisposing factors to landslide, as summarized in table 1. Some of them, as lithology, slope and land use, are considered by the majority of the references, because they are proven factors of instability. Other factors are taken into account depending on the features of the study area.

	Geological					Geomorphological						Anthropogenic			Climatic	
	Lithology	stratigraphy	Distance from tectonic alignments	Soil type	Soil depth	Slope	Morphological parameters	Aspect	Elevation above sea level	Distance from the drainage network	Accumulation	Land use	Vegetation	Distance from the road network	Climatic aggressivity F_{FAO}	Mean monthly rain
Natali et al. (2010)	x					x						x				
Pradhan & Lee (2010)	x		x	x		x		x		x		x	x			x
Pauditš & Bednárik (2002)	x					x			x							
Lee (2007)	(x)		x	x		x		x		x		x	x			
Dahal et al. (2008)	x			x	x	x		x	x		x	x		x		
Cencetti et al. (2010)	x	x				x						x			x	
Dai & Lee (2002)	x					x	(x)	x	x	(x)		x				
Ayalew & Yamagishi (2005)	x	x	x			x		x	x					x		

Table 1: Literature analysis of predisposing factor to landslide [(x) indicates a weak suggestion of such factor]

In our work, combining literature analysis with some considerations on cartographic data availability, we considered the following factors:

- *Geological factor*: we use a *geo-lithological* vector map (1:100'000), for a lithological description of soils.
- *Geomorpological factors*: we use digital elevation models (DEM), which describe the surface topography, to obtain *elevation*, *slope*, *aspect* maps.

Moreover surface topography controls flow sources, flow direction and soil moisture concentration, that are important factors correlated to the density and spatial extent of landslides. Hence we use the DEM to calculate the *accumulation* map.

- *Anthropogenic factors*: several studies demonstrated that the cuts of the slope caused by the roads are usually causes of anthropologically induced instability; this information could be extracted from either a *land-use cover* map and a *road* map.
- *Climatic factor*: to take into account the rain occurrence and its temporal distribution, which may affect the landslide occurrence; in particular we use the "*climatic aggressivity*" (F_{FAO}) named also "*Fournier index*". It was defined by Arnoldus (1977) as:

$$F_{FAO}=\frac{\sum_{i=1}^{12} p_i^2}{P} \quad (1)$$

where p are the mean monthly precipitation and P the mean annual precipitation.

2.2 Logistic multiple regression

Logistic regression is an example of generalized linear model (GLM). It is a flexible generalization of ordinary linear regression, which is appropriate for response variables that have error distribution models other than a normal distribution (i.e. landslide occurrence). It is a well-working method for dichotomous data, that means that the dependent variable can have only two values (event occurring or not occurring), and predicted values can be interpreted as probabilities since they are constrained to fall in the interval between 0 and 1. In the present study, the dependent variable is a binary variable representing the presence or absence of landslides. In particular logistic regression is known to be more flexible than the alternative methods for dichotomous data, such as the linear regression.
The technique of logistic multiple regression yields coefficients for each variable based on data derived from samples taken across a study area. These coefficients serve as weights in an algorithm which can be used in the GIS environment to produce maps depicting the probability of landslide occurrence.
Using the multiple regression model, the relationship between the occurrence and its dependency on several variables can be expressed as following:

$$logit(p)=\ln\left(\frac{p}{1-p}\right)=b_0+\sum_{j=1}^{n} b_j X_j \quad (2)$$

where X are the variables, b the coefficients and p the landslide occurrence probability obtained by a landslide inventory map.
In the present work the calculation of b regression coefficients has been done using the *r.regression.multi* GRASS GIS command. This tool is implemented for the linear regression model, hence the inputs of the model are the maps of independent variables X_j and the logit of the landslide map.
The outputs are:

- a map with estimate of logit(p)

- a map with the errors (optional)
- a file with coefficient values, the Akaike Information Criterion (AIC) index and the R^2 correlation coefficient for each explaining variable.

The AIC is a measure of the relative quality of a statistical model for a given set of data. If the adding of a variable has a positive effect on the model, the values of AIC have to decrease. Adopting a stepwise method we tried to understand the influence of adding a variable or not.

The R^2 correlation coefficient for each explaining variable represents the additional amount of variance when including the variable compared to when excluding that. It gave us a measure of the influence of a single factor in each model.

Preliminary to the application of the logistic multiple regression, we have to analyse each variable singularly through a bivariate statistical approach. In this way we are able to understand the real influence of each variable on the physical phenomenon and to correctly classify the variable map in order to have a direct or inverse proportionality with landslide occurrence. This preliminary step is especially necessary in case of no-numerical variable, as the geolithological and the land-use map, or in case of numerical variable not characterized by a regular proportionality with landslide occurrence (i.e. aspect map). In fact, the adoption of a logistic regression model involves the calculation of the value of logit through a linear function of the independent variables X; therefore a monotonic (increasing or decreasing) behaviour needs to be observed also in the classification of the X variables.

Using a bivariate statistical approach the landslide occurrence is considered as dependent on a single variable and the conditional probability is obtained using the Bayes's theorem:

$$P(\mathrm{landslide}|X)=\frac{P(\mathrm{landslide}\cap X)}{P(X)} \tag{3}$$

On the basis of bivariate analysis it is possible to correctly reclassify the geolithological, land-use and aspect maps using an increasing or decreasing rule, that otherwise is not possible.

3. Study area and data sources

The susceptibility statistical analysis described in the previous section was calibrated on the whole Liguria region (north-western Italy) or on its four administrative provinces (Imperia, Savona, Genova, La Spezia in figure 1).

Data used were:

- The landslide inventory map, available from Liguria region geo-portal. Its development starts with the IFFI project (Inventory of Landslide Phenomena in Italy). IFFI project was carried out by ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) and the Regions and Autonomous Provinces since 1998. It supplies a detailed picture of the distribution of landslide phenomena within Italy. Actually the inventory is periodically updated from each region and the version used for calibration is updated to 2013. The type of landslides involved in this study are slides and flows, according to the UNESCO WP/WLI (1993).

- A raster DEM of the entire Italy downloaded from ISPRA, with a cell size of 20x20m.
- A geo-lithological vector map (scale 1:100'000) coming from the Liguria Region Environmental Service.
- A land-use vector map (scale 1:10'000) updated to 2012.
- A road network vector map (scale 1:10'000) updated to 2003.
- A map with the ARPAL rain gauges with daily rain data available for at least 10 years (ARPAL, 2012). From the daily rain data available in several rain gauges on the Ligurian territory (figure 1), we calculated the monthly and annual mean precipitation and consequently the F_{FAO} as expressed by the equation 1. The F_{FAO} values in correspondence of the rain gauges were interpolated, after having tested various interpolation methods, as IDW (Inverse Distance Weighted) with 3 or 4 points, RST (Regularized Splines with Tension), Voronoi and Natural Neighbours. We have chosen the Watson algorithm for Sibson natural neighbour interpolation (Sibson, 1980, Sibson, 1981; Watson, 1992) implemented in the GRASS GIS *r.surf.nnbathy* add-on command, because this algorithm is reliable for large dataset with an irregular sampling. Nevertheless the result is available only inside a convex polygon which includes the known data. In order to obtain a complete result on all the Liguria region the Voronoi or Thiessen polygons are used in the boundary region, where the natural neighbour interpolation does not return results. F_{FAO} interpolation is shown in Figure 2.

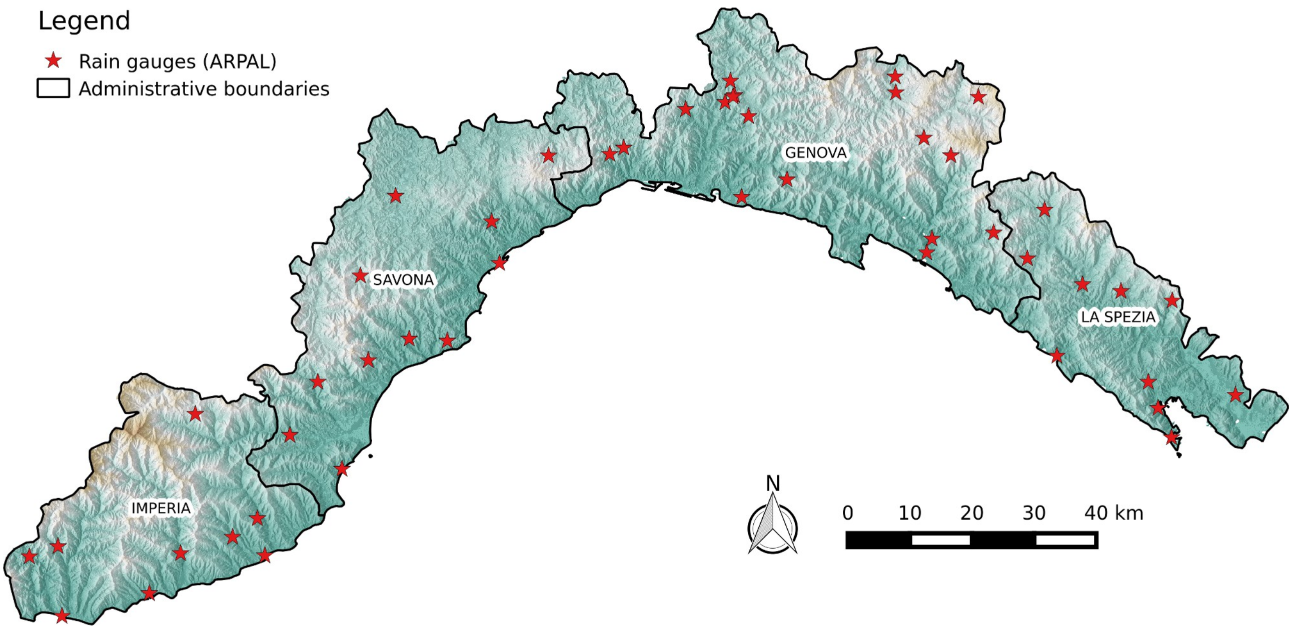

Figure 1:The Liguria region, with administrative boundaries of the four provinces and location of rain gauges used for calculation of F_{FAO}

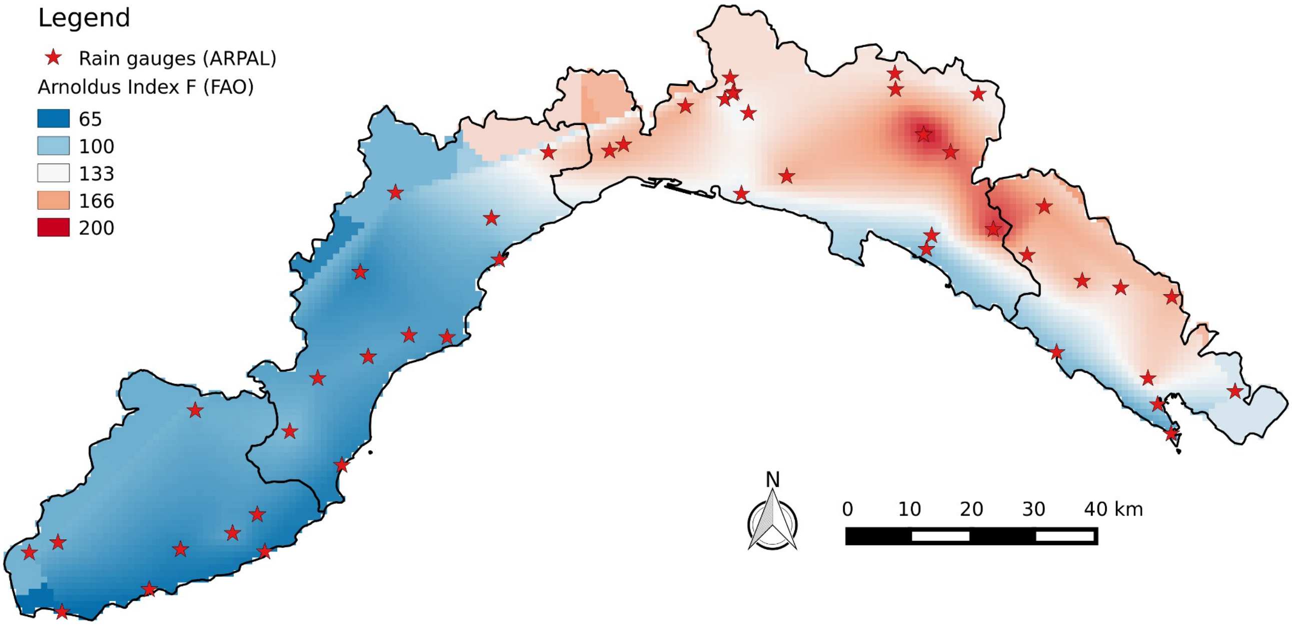

Figure 2: Result of interpolation of Arnoldus or Fournier Index (F_{FAO}).

3. Results and discussion

The 5 maps previously listed were used to calculate the 8 maps of the analysed predisposing factors to landslide: slope, land use, lithology, accumulation, distance from road, elevation, aspect, F_{FAO} . Such 8 maps were conveniently classified in 3 to 9 classes, in function of the results of the bivariate statistical analysis of each variable singularly considered and the landslide inventory map, applied on all the Liguria territory (figure 3). The meaning associated to the classes for the non-numerical variables, i.e. lithology and land cover, or for numerical maps as log(accumulation) and F_{FAO} are reported in table 2.

F_{FAO}		Log(accumulation)		Lithology		Land cover	
1	<100	1	≤ 1 ridge	1	Deposits	1	other
2	100-200	2	1 – 2	2	Magmatic rocks	2	buildings
3	>200	3	2 - 3	3	Metamorphic rocks	3	Agricultural areas
		4	3 – 4	4	Sedimentary rocks	4	wood
		5	4 – 5	5	Other	5	sparse vegetation
		6	> 5 humid zone			6	Open space with little or no vegetation

Table 2: Classification of explaining variables (see figure 3)

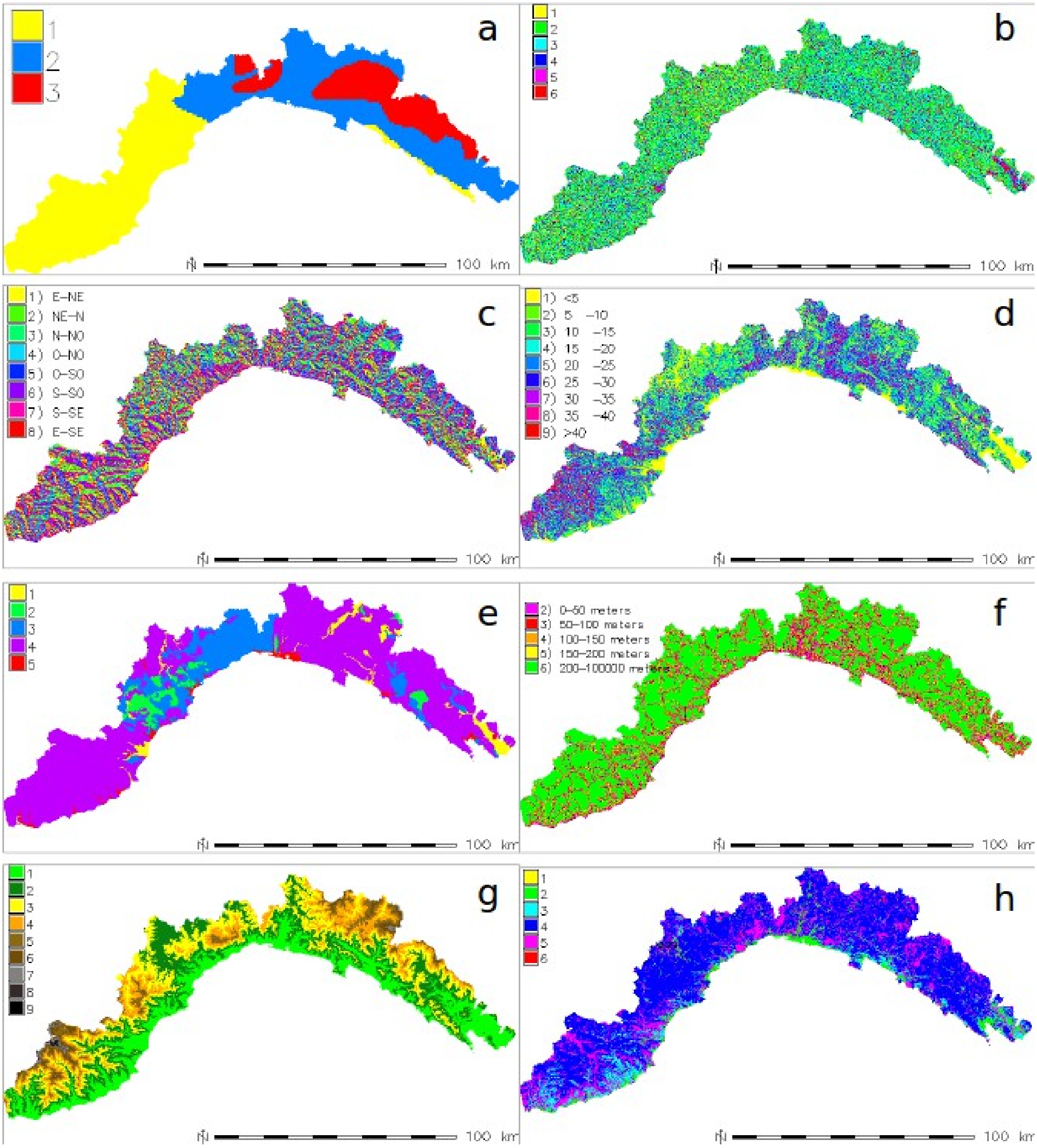

Figure 3: The reclassified maps of explaining variables: a) F_{FAO}; b) Log(accumulation); c) aspect; d) slope; e) lithology; f) distance from road; g) elevation [a class every 250 m a.s.l.]; h) land cover.

Then we have applied the logistic multiple regression and analysed the variability either of AIC index related to the number of analysed explaining variables, and the b and R^2 correlation coefficients for each best model. The calibration has been done using all the Liguria region and the four provinces separately shown in figure 1.

	Liguria		Imperia		Savona		Genova		La Spezia	
Number of variable with lower AIC index	8		7		8		8		7	
Parameter	**b**	**R^2 (10^{-3})**	**b**	**R^2 (10^{-3})**	**b**	**R^2 (10^{-3})**	**b**	**R^2 (10^{-3})**	**b**	**R^2 (10^{-3})**
slope	0.029	0.68	0.040	1.40	0.015	0.19	0.028	0.41	0.07	0.12
land use	0.059	0.32	0.042	0.19	0.030	0.08	0.104	0.71	0.013	0.07
lithology	-0.044	0.22	0.020	0.00	-0.045	0.29	-0.079	0.47	0.010	0.05
accumulation	0.021	0.14	0.033	0.36	0.016	0.09	0.019	0.08	0.004	0.02
aspect	0.006	0.03	0.010	0.11	-0.002	0.00	0.009	0.05	0.003	0.05
elevation	-0.012	0.05	0.015	0.12	-0.005	0.01	-0.086	1.39	-0.002	0.00
distance from road	-0.022	0.19	0.001	0.00	-0.026	0.25	-0.028	0.23	-0.001	0.00
F_{FAO}	0.019	0.04	-	-	-0.038	0.05	0.141	0.61	-	-

Table 3: Results of the logistic multiple regression.

For the Liguria region the AIC value decreases until the eighth explaining variable and the same behaviour was observed for the Savona and Genova provinces (figure 4). For Imperia and La Spezia provinces the F_{FAO} variable has a quasi-uniform class in the calibration regions, therefore its influence is negative on the model as well demonstrated by the increase of the AIC parameter for the model with 8 variables.

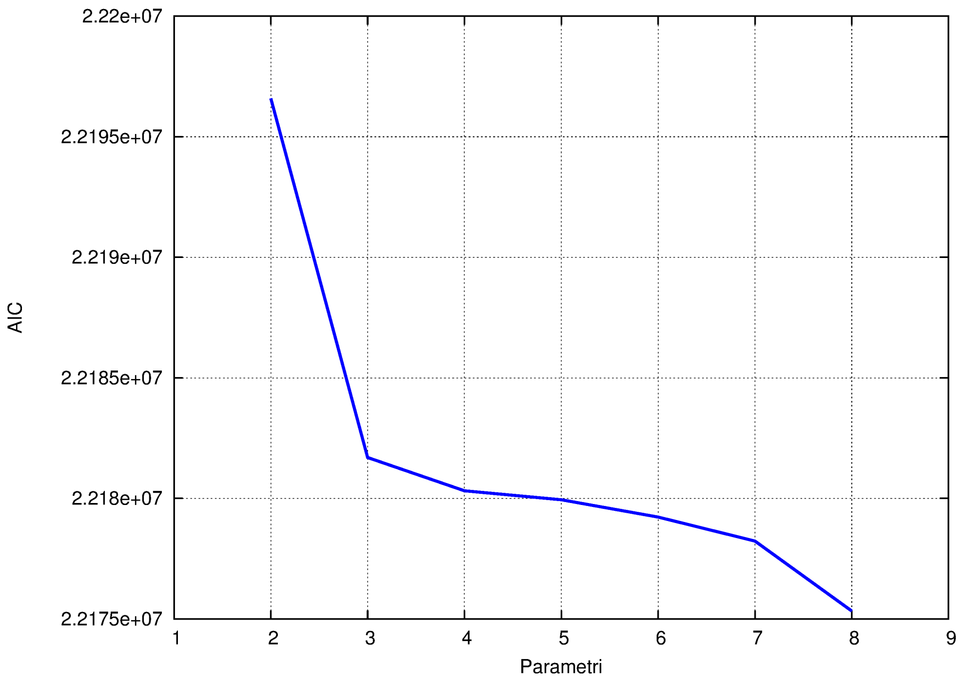

Figure 4: AIC index for the calibration on the entire region.

R^2 gives an idea of the influence of each explaining variable on the landslide susceptibility: higher the value higher the influence. Also the coefficient b is an indicator of the influence of each variable on the landslide occurrence: if it is

positive the variable is a predisposing factor, otherwise it is not-predisposing.
A coherence of behaviour in the various calibration areas is an indicator of a greater reliability of the factor, as it happens for example for slope, land use and accumulation. But some factors, as elevation, aspect, F_{FAO} and lithology, show a very variable influence on the occurrence of landslides, with positive and negative b values if obtained from different calibration areas, and hence lower reliability.
In particular, the coefficients applied to slope, land use and accumulation maps are always positive for all the different calibration areas, meaning that their increasing have an effect also on the increasing of landslide occurrence probability. The coefficient applied to the map "distance from road" assumes values almost always negative, even if very low in absolute terms, because lower distances to the road have higher effect on landslide occurrence.
The elevation variable, seems to have, at least for the Liguria region, a rather low influence on the landslide occurrence, even if it can be strongly related to other physical phenomena hence statistically significant also on the landslide occurrence.
Concerning aspect variable, we observed a different behaviour for the West coast (Imperia and Savona provinces) and East coast (La Spezia province) and a different classification is probably more suitable and will be tested in further analyses.
With regard to F_{FAO} it seems to be a good parameter in analysis on large areas with different climatic conditions while it is not a good parameter when applied to small calibration area characterized by similar climatic conditions.
Finally, lithology map seems to be poorly or highly related to landslide occurrence in the different areas of calibration, as demonstrated from the high variability of R^2 correlation coefficient in the different provinces. This is probably due to the low map scale affecting also the spatial variability, that, in some provinces is very low (e.g. Imperia). For this reason, in our opinion, the use of this lithology map is appropriate only for wide areas (entire region). Obviously the availability of a more accurate vector geo-lithological maps (e.g. 1:25'000) could solve the problem, and the use of this map could become more appropriate from a statistical point of view.
For all these last four parameters, this first validation (conducted on different calibration areas) highlights the importance of bivariate analysis prior to the logistic multiple regression in order to understand if there is a correlation. In positive case, the next step is to correctly classify the map, or exclude the map from the multiple regression analysis in negative case.

5. Conclusions

This first validation performed in the Liguria region (considering different areas for extension, morphological and climatic characteristics) was surely interesting to understand few critical issues of the multivariate statistical analysis.
The next steps of our research are:

- a more accurate analysis of aspect and elevation maps comparing the Liguria region with other areas;
- an extension of the validation area to the Piemonte region in order to add different morphological conditions to the statistical analysis (a more

extended alpine area, the Po valley, the Langhe and Monferrato hills, etc.).

The final aim of our research consists in providing guidelines for not expert users which may clearly explain the requested steps to obtain a susceptibility map with a statistical multivariate approach in order to avoid any mistakes.

Acknowledgement

This work is financed by the national research program (PRIN 2010-2011) called "*Mitigation of landslide risk using sustainable actions*" coordinated by University of Salerno.

References

- ✔ Arnoldus, H.M.J. (1977). Methodology used to determine the maximum potential average annual soil loss due to sheet and rill erosion in Morocco. *FAO Soil bulletin,* 34, 39-48.
- ✔ ARPAL (2012). Atlante climatico della Liguria (http://www.arpal.gov.it/homepage/meteo/analisi-climatologiche/atlante-climatico-della-liguria.html)
- ✔ 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.
- ✔ Bovolenta, R., Federici, B., & Natali, D. (2011). Realizzazione automatica di carte di pericolosità da frana superficiale utilizzando il GIS open souce GRASS. - X Giornata Mondiale dell'Acqua. Frane e Dissesto Idrogeologico: consuntivo. *Atti dei Convegni Lincei*, Accademia Nazionale dei Lincei, Roma, ISSN 0391805; ISBN 9788821810367, 169-177.
- ✔ Cascini, L. (2008). Applicability of landslide susceptibility and hazard zoning at different scales. *Engineering Geology,* 102, 164-177.
- ✔ Cencetti, C., De Rosa, P., Fredduzzi, A., & Minelli, A. (2010). La valutazione della suscettibilità da frana tramite i metodi dell'Analisi Condizionale e degli Indici di Franosità: un'applicazione tramite GRASS al bacino del T. Ventia (Umbria settentrionale). *Geomatic Workbooks*, 9, 41-54.
- ✔ Dahal, R.K., Hasegawa, S., Nonomura A., Yamanaka, M., Masuda, T., & Nishino K. (2008). GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environmental Geology,* 54(2), 314-324, doi:10.1007/s00254-007-0818-3
- ✔ Dai, FC, & Lee, CF. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology,* 42, 213-22.
- ✔ Fell, R., Corominas, J., Bonnard, CH., Cascini, L., Leroi, E., & Savage, B. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102, 85-98, ISSN 0013-7952
- ✔ Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology,* 31, 181–216.
- ✔ Hutchinson, J.N., (1995). Landslide hazard assessment. In: *Proc VI Int Symp*

on Landslides, Christchurch, 1, 1805-1842.

- ✔ ISPRA (2013). Linee guida per la valutazione del dissesto idrogeologico e la sua mitigazione attraverso misure e interventi in campo agricolo e forestale. ISPRA, Manuali e Linee Guida 85 ISBN 978-88-448-0586-9
- ✔ Lee, S. (2007). Application and verification of fuzzy algebraic operators to landslide susceptibility mapping. *Environmental Geology.* 52, 615-623.
- ✔ Natali, D., & Federici, B. (2010). Realizzazione di carte della pericolosità da frana su aree estese: una procedura automatica in GRASS. *Geomatic Workbooks,* 9, 121-126.
- ✔ Pauditš, P, & Bednárik, M. (2002). Using GRASS in evaluation of landslide susceptibility in Handlovská kotlina Basin. *Proceedings of the Open source GIS - GRASS users conference 2002,* Trento, Italy, 11-13 September.
- ✔ Pradhan, B., & Lee, S. (2010). Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides,* 7, 13–30.
- ✔ Sibson, R. (1980). A vector identity for the Dirichlet tesselation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, 87, 151–155.
- ✔ Sibson, R. (1981). A brief description of natural neighbour interpolation. In Barnett V. (ed.), *Interpreting Multivariate Data*, Wiley, New York, USA, 21–36.
- ✔ Varnes, D.J. and the International Association of Engineering Geology Commission on Landslides and Other Mass Movements (1984). Landslide hazard zonation: A review of principles and practice. *Natural Hazards,* 3, Paris, France. UNESCO, 63p.
- ✔ Watson, D.F. (1992). *Contouring: A Guide to the Analysis and Display of Spatial Data*. Pergamon Press, Oxford, UK.
- ✔ WP/WLI-International Geotechnical Societies UNESCO Working Party on World Landslide Inventory (1993). *Multilingual Landslide Glossary*. Bitech Publisher, Richmond, British Columbia, Canada, 59 pp.