

# ANALYSIS OF LANDSLIDE HAZARD MAPPING METHODS: REGRESSION MODELS VERSUS WEIGHT RATING

A.Erener,<sup>a</sup> HSB. D üzgül.<sup>b</sup>

<sup>a</sup>Middle East Technical University, Ankara, Turkey (e-mail: erener@metu.edu.tr)

<sup>b</sup>Middle East Technical University, Ankara, Turkey (e-mail: : duzgun@metu.edu.tr)

**KEY WORDS:** Susceptibility, Quantitative Method, Qualitative Method, Spatial Regression, Logistic Regression, Heuristic Approach, ROC Curve

## ABSTRACT:

Hazards due to landslides cause huge sufferings to human and substantial economic and environmental losses throughout the world. For the assessment of the landslide hazard, different methodological approaches are proposed. They are mainly grouped into two: qualitative and quantitative methods. The quantitative methods eliminate the problem of subjectivity of the qualitative methods. On the other hand, qualitative methods have simpler methodological procedures. Among the quantitative methods, statistical techniques are one of the most popular ones. This study aims at providing a comprehensive comparison of quantitative and qualitative landslide hazard assessment maps, as well as proposing a new approach to enhance the performance of qualitative methods. The conventional statistical analysis can only produce, average and global parameter estimates for the whole study region. However, there are local variability in the factors causing landslides due to spatial correlation; i.e. geographically nearby units will have similar characteristics, while the ones, which are farer apart will have less similar properties. Therefore, as regression methods based on ordinary least square estimation, they are lack in modelling the spatial correlation structure of the parameters. For this reason, in this study ordinary logistic regression (OLR) is extended by spatial regression (SR) techniques for landslide hazard assessment. Moreover, a heuristic approach is developed for obtaining the susceptibility map based on assigning weights to landslide causing factors. In this approach a weight rating system is adapted, where weights were assigned to different causitive factors. In both of the quantitative and qualitative analyses, slope, aspect, curvature, which are derived from Digital Elevation Model (DEM), lineaments and vegetation index, which are extracted from Landsat Thematic Mapper (TM) satellite images, elevation which is obtained from the contour map are considered to be the landslide influencing factors. The proposed models are implemented to a case study area from More and Romsdal region of Norway. Hazard maps from SR,OLR and susceptibility map from heuristic approach are prepared and compared. It is concluded that SR model have better explanation capability for landslide occurrences than OLR model and heuristic model.

## 1. INTRODUCTION

The instability of a slope is governed by a various interrelated terrain, geological, hydrological and land use characteristics (van Westen et. al., 1997). In order to estimate the landslide prone areas for a given region, different methods (qualitative or qualitative) are proposed considering the landslide causing factors. Usually the product of qualitative methods (e.g. Carrara et.al, 1995) is in the nature of susceptibility maps as they do not provide information about the probability of landsliding. Hazards maps, on the other hand, can be obtained through the use of quantitative approaches. Examples of studies using these methods in literature are Carrara, 1983, Carrara 1999, Sakellariou, and Ferentinou, 2001, Lee 2003. One of the major problems associated with hazard and/or susceptibility maps, is that it is hardly possible to assess the reliability of the products. One way would be the implementation of various methods to the same area and observation of the differences. For this purpose, qualitative and quantitative hazard and susceptibility maps are produced and compared for More and Romsdal region of Norway in this study. Qualitative method involves the use of a heuristic technique in which the landslide causing factors are weighed according to their importance. Regression methods form the basis of quantitative methods, where a mathematical

relation between the landslide occurrences and landslide causing factors is established. In this study it is also proposed a regression approach, in which the spatial correlation of the regression parameters are taken into account by extending ordinary logistic regression (OLR) to spatial regression (SR).

## 2. METHODOLOGY

For the application of statistical methods and heuristic methods, a spatial database that considers landslide-related factors is constructed. In heuristic analysis, instability factors are ranked and weighted according to their assumed or expected importance in causing mass-movement. Then a susceptibility map is created by the application of different weights to each parameter.

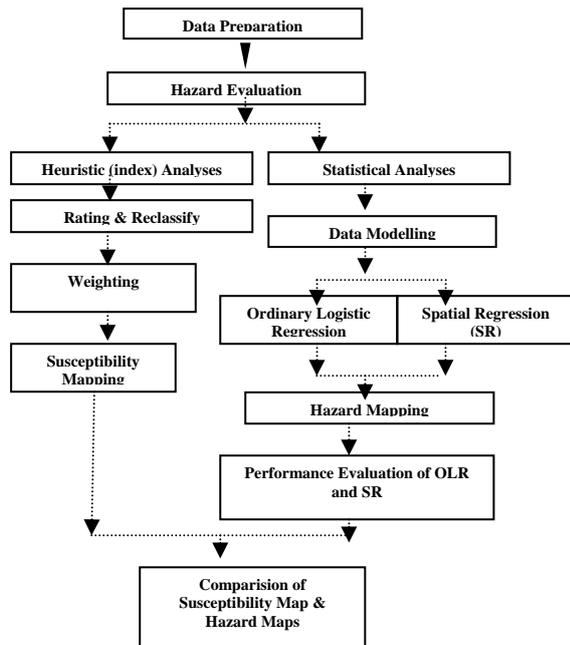


Figure 1. Methodology of the study

In the statistical methods, a spatial database is constructed by assigning factor values into each mapping unit, which are grid cells in this study. Then the OLR and SR models are established. In OLR relationships between the landslide occurrences and each landslide-related factor are analyzed. However the spatial correlation among the parameters is not considered. In SR on the other hand spatial autocorrelation among the regression parameters are taken into account. After application of statistical models hazard maps are produced and the relative operating characteristics (ROC) curve is used for performance evaluation of OLR and SR. Then the susceptibility map result from heuristic method and hazard maps result from statistical methods are compared.

### 2.1. Study Area And Data Preparation

The study area occupies approximately 12,168 km<sup>2</sup> in the western part of Norway, More and Romstal region located on 62° 57' 0" North, 7° 1' 0" East (Figure2).

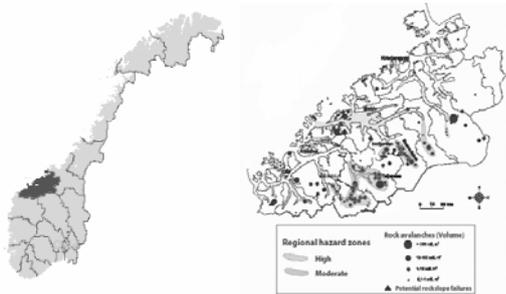


Figure 2. Study region

It is of primary importance to recognize the conditions that caused instability of the slope and the processes that triggered the movement (Popescu, 1994). Rock falls and avalanches inventory in More & Romsdal is obtained from Norwegian Geological Survey (NGU). The landslide influencing factors are

divided into three categories: topographical, land cover and structural geology. Topographical parameters are slope, aspect, curvature and elevation. Elevation is obtained from the contour map of the region and the rest of the topographical parameters, the slope, aspect and curvature values are calculated from a 30m resolution Digital Elevation Model (DEM), which is constructed from TIN of contour maps. Vegetation index and lineaments constitute the land cover and structural geology, respectively. The land cover and structural geology parameters are extracted from Landsat Enhanced Thematic Mapper (ETM) images with a 30m x 30m resolution. PCI Geomatica V9.1 is used for automatic extraction of lineaments from Band 7; since this band is useful for discrimination of lineaments and is also sensitive to vegetation moisture content (Sabins, 1996). Normalized Difference Vegetative Index which is a measure of the vegetative cover is used to determine the density of green. For Landsat imagery, Bands 4 and 3 are used to generate this measure, since these bands best highlight chlorophyll absorption and provide good contrast between vegetation and soil. Table1 represents the properties and scale of the influencing parameters used in this study.

CLASSIFICATION	SUB_CLASSIFICATION	GIS DATA TYPE	SCALE
Topographical	Historical landslides	Point Coverage	1:500000
Topographical	Elevation	Line Coverage	1:100000
Topographical	Slope	GRID	30*30
Topographical	Aspect	GRID	30*30
Topographical	Curvature	GRID	30*30
Structural Geology	Lineament	Line Coverage	1:100000
Land Cover	Vegetation Index	GRID	30*30

Table 1. Data properties and scale

Evaluation of landslide hazard requires the preliminary selection of a suitable mapping unit. Selection of the terrain-unit largely influences all the subsequent analyses and modelling. In this study the region is divided into grid-cells. Each cell is assigned a value for each factor and are treated as a case or sampling unit.

## 3. SUSCEPTIBILITY AND HAZARD MAPPING

### 3.1. Qualitative susceptibility mapping by index-based assessment

Seven layers of data are superimposed to create the landslide susceptibility map. Slope is given the most emphasis, followed by, density of vegetation, density of lineament, aspect and curvature. A numerical rating system is applied and each of the 7 factors is grouped into three categories, and each category is assigned a value between 1 and 3, with 1 being least susceptible and 3 most susceptible to landslides. Based on their relative importance to slope instability in the study area, the seven factors are assigned weights between 0.0 and 1.0 (collectively adding to 1.0). Slope is assigned the highest weight, followed by density of lineament density of vegetation, aspect and curvature. A raster-based GIS is used to overlay the five 30 x 30 m resolution grids and calculate a Landslide Susceptibility Index (LSI) for

each individual cell. The final map shows areas of low, medium, and high landslide susceptibility.

The rating operation is applied to all factors taken into consideration for hazard assessment and they are ranked depending on their importance to slope instability. Before application of rates and classification, the water region is extracted from the study region by masking the water area. Slope is considered to be the most important factor to affect the landslide occurrence. The slope map is classified into three slope angle classes (0-15= 'low', 15-25= "moderate" and 25-60.87= "high'), and each class is assigned a value between 1 and 3. The aspect map classified into three classes (270-360= 'low', 90-270= "moderate" and 0-90= "high'), and each class is assigned a value between 1 and 3. Curvature values represent the morphology of topography. From an applied viewpoint, output of curvature can be used to describe the physical characteristics of a drainage basin in an effort to understand landslide processes. The profile curvature affects the acceleration and deceleration of flow, and therefore influences landslide phenomena. A positive curvature indicates that the surface is upwardly convex at that cell. A negative curvature indicates that the surface is upwardly concave at that cell. A value of zero indicates that the surface is flat. The negative and positive curvature has positive effect to landslide occurrences from the result of statistical modelling for that reason only curvature is ranked for the study. The curvature map is then threshold into three classes ("−62" - "−2" and "+2" - "+53" =moderate" and "−2" - "+2" = "low"), and each class is assigned a value between 1 and 3, with the range close to 0 value being least susceptible and the negative and positive values larger than 0 being the most susceptible to landslides. The NDVI (Normalized Difference Vegetation Index) result map is grouped into three class and dense vegetated areas given low risk and low dense areas given the high risk. The lineament density is grouped into three class low dense lineament considered as least susceptible areas and scored with (1); Medium dense lineament between scored with (2), high dense lineament being most susceptible to landslides scored with (3). Equation 1 is used for susceptibility map generation, where slope and lineament density have higher weights as they have relatively higher influence on slope instability in rock falls and avalanches.

$$[\text{Aspect}] * 0.15 + [\text{Slope}] * 0.25 + [\text{NDVI}] * 0.15 + [\text{Curvature}] * 0.15 + [\text{DEM}] * 0.1 + [\text{Lineament}] * 0.2$$

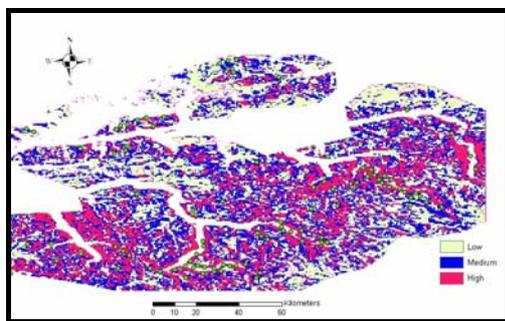


Figure 3. Susceptibility map produced from heuristic method

### 3.2. Quantitative hazard mapping by regression models

#### 3.2.1. Ordinary Logistic regression

Landslide hazardous areas are analyzed and mapped using the landslide occurrence factors by logistic regression model. Using a logistic multiple regression model, the spatial relationships between the landslide location and each landslide-related factor are analyzed, and a formula of landslide occurrence possibility is obtained. The logistic regression involves fitting the dependent variable using an equation in the following form (Begueria & Lorente 1999)

$$\ln\left(\frac{\pi_1}{1-\pi_1}\right) = \beta_0 + \beta_1 * X_1 + \dots + \beta_n * X_n \quad (1)$$

where  $\pi$  is the probability that the event Y occurs,  $p(Y=1)$ ,  $\pi/(1-\pi)$  is the "odds ratio" or likelihood ratio,  $\ln[\pi/(1-\pi)]$  is the log odds ratio, or "logit",  $\beta_0$  is the intercept, and  $\beta_1, \beta_2, \dots, \beta_n$  are coefficients that measure the contribution of independent factors ( $X_1, X_2, \dots, X_n$ ) to the variations in Y.

The dependent variable represents presence or absence of landslides and the 7 independent variables are considered in the model as: MAXSLP (Maximum slope), MAXASP (Maximum aspect), VEGDEN (Vegetation density), NEGCUR (Negative curvature), POSCUR (Positive Curvature), ELEV (Maximum elevation) and LINMEN (Lineament density). The OLR model based on dependent and independent variables are given in Equation 2 and the statistics of coefficients are listed in Table 2.

$$f(x) = -5.0548 + 0.1089 * [\text{MAXSLP}] - 0.0014 * [\text{MAXASP}] - 0.9625 * [\text{VEGDEN}] + 0.0055 * [\text{NEGCUR}] + 0.0205 * [\text{POSCUR}] - 0.0120 * [\text{ELEV}] - 0.0010 * [\text{LINMEN}] \quad (2)$$

Variable	Parameter ( $\beta$ )	Sig	Exp( $\beta$ )
MAXSLP	0.1089	0	1.115
MAXASP	-0.0014	0.4834	0.9986
VEGDEN	-0.9625	0.0368	0.3819
NEGCUR	0.0055	0.9822	1.0055
POSCUR	0.0205	0.2881	1.0207
ELEV	-0.012	0	0.9881
LINMEN	-0.001	0.01	0.999
Constant	-5.0548	0	
Cox & Snell-R <sup>2</sup>	.073		
Nagelkerke R <sup>2</sup>	.466		
Chi-Square	438.325		
Overall success rate	98.48%		

Table 2. Coefficients value for logistic regression

The logit of F(x) function, P(L), is calculated for all of the pixels of the study based in Eq. 2. Depending on the equation 3, the hazard map is obtained and given in Figure 4.

$$P(L) = \frac{1}{1 + \text{Exp}^{-f(x)}} \quad (3)$$

The landslide probability values obtained from Eq. 3 are classified as low, medium and high hazard zones. Red colour that represents the 0.4% to 28% of hazard is determined as the northern part of the region.

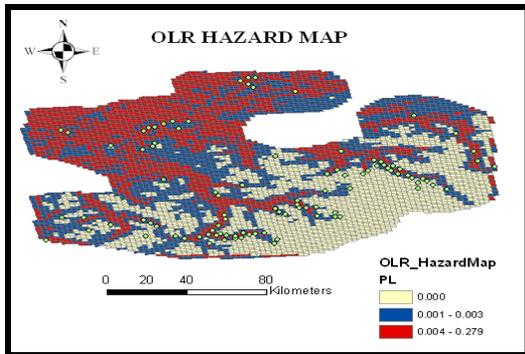


Figure 4. Hazard Map of Ordinary Logistic Regression

### 3.2.2. Spatial Regression

The spatial regression is modification of the regression equation (Eq. 1) by using contiguity matrix (proximity matrix or geographic weights matrix), in which the neighborhood information about the spatial zones are characterized and accounted for spatial autocorrelation. In matrix notation SR is expressed by:

$$Y = X\beta + \rho Wy + \varepsilon \quad (4)$$

Where,

- $\varepsilon$ : Vector of errors with zero mean and constant variance  $\sigma^2$
- W: Proximity matrix
- $\rho$ : Interaction parameter or spatial autoregressive coefficient
- $\beta$ : Parameter to be estimated due to relationship between the variables

In this study k-nearest neighbour weights are used to construct W matrix. Local pseudo r-square show that nearly 0.13 of the variance in landslide is explained by the model. The coefficients are given at the below Table 3.

Variables	Coefficient	Std.Error	t-Statistic	Probability
Constant	-0.00244	0.009892	-0.24628	0.805339
Slope	0.001106	0.000208	5.305055	0.0000001
Aspect	-6.05E-07	2.80E-05	-0.02158	0.985179
Vegetation	0.009201	0.013066	0.704187	0.481306
PosCurvature	0.001246	0.000359	3.473535	0.000517
Neg Curvature	0.000203	0.001236	0.16464	0.868857
Elevation	-2.34E-05	4.42E-06	-5.29933	0.0000001
Lineament	-1.70E-05	2.61E-06	-6.53329	0

Table 3. Spatial regression coefficients

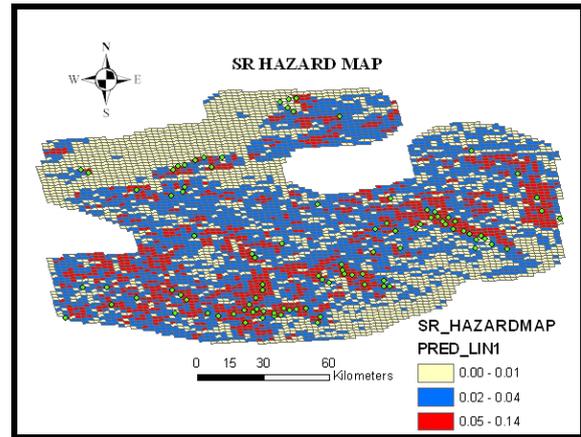


Figure 5. Hazard Map of Spatial Regression

From the estimated landslide occurrences values a hazard map is produced shown on Figure.5.

### 3.2.3. Quality Determination of SR and Logistic Regression Models

The relative operating characteristics (ROC) curves give how well the two models predict landslide. Hence the ROC curves of OLR and SR are constructed for evaluating the performance of the two models (figure 5). As can be seen from Figure SR has better predictive performance that OLR since in SR plot ROC curve have larger area under the curve.

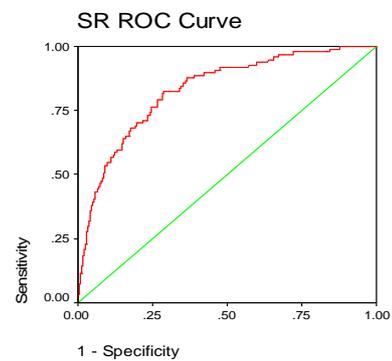
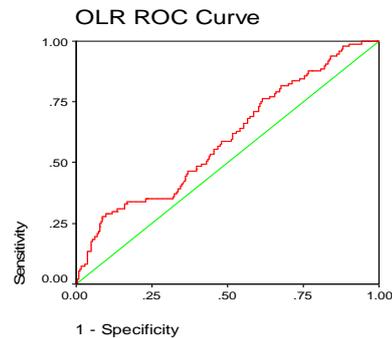


Figure 6. SR and OLR Model comparison using ROC curve

### 3.3 Comparison of Susceptibility and Hazard Maps

The landslide probability values obtained from OLR and SR models are classified in to 3 zones as low, medium and high. The hazard map produced by OLR considers the northern part of the region all represented as high hazard which indicates high degree of overestimation. Compared to the hazard map produced from OLR, the hazard map produced from SR model is more realistic. Also the  $R^2$  which show the variance explained by the model increased from 0.073 to 0.13 with SR. Hence by explicit modelling of spatial effects and by measuring the spatial dependence and heterogeneity in the dataset the ability of the model to explain the probability of occurrence of the landslide is increased.

Susceptibility map produced from heuristic method also resembles to the hazard map produced from the SR model. However, because of the subjectivity on ranking and weighting operation, which may change depending on the knowledge of the expert, the heuristic method doesn't provide reproducible results. On the other hand, heuristic method is simpler than statistical methods and it may give similar results to quantitative methods if there is an expert(s) who knows the study region well.

### 4. CONCLUSION

In this study the techniques of OLR, SR and heuristic method are employed for landslide hazard assessment. Heuristic approach is developed for obtaining the susceptibility map based on assigning weights to landslide causing factors. In this approach a weight rating system is adapted, where weights are assigned to different causative factors.

On the other hand logistic regression analysis can only produce, 'average' and global parameter estimates. For a given region, a single regression model might be fitted to data on landslide occurrence and its influencing factors along the whole region.

As the landslide occurrences have local scale variations, use of global models cause the ignorance of local scale variations in hazard assessment. To assess the probability of landslide within a specified period of time and within a given area, it is important to understand the spatial correlation between landslide occurrences and influencing factors. For this reason at regional scale, landslide hazard studies applied with the spatial regression model to consider non-stationary in the parameters. SR model allows different relationships to exist at different points in space. Therefore, local rather than global parameters can be estimated and spatial- non stationary can be examined by the weighting system depend on the location in geographical space. In this way, the main shortcomings of non-spatial regression, which are assumptions of identically and independently distributed

explanatory variables and uncorrelated error terms, are tried to be eliminated by relaxing the regression method with the allowance of spatial autocorrelation. By including these relationships the predictive ability of the developed model increases. As a result the landslide hazard assessment is extended by applying spatial regression models by considering that the surface of the earth is physically continuous every where, the value in a geographic cell must be strongly related to values in geographic nearby cells. In this respect non-stationary models handled in this study show that SR is a suitable choice for the assessment of landslide hazard.

### REFERENCES

- Beguiria S. & Lorente A. 1999. Hazard mapping by multivariate statistics: comparison of methods and case study in the Spanish pyrenees. Debrisfall assessment in mountain catchments for local end-users. *Instituto Pirenaico de Ecología, CSIC*, 202, 50080-Zaragoza, Spain.
- Carrara A., 1983. Multivariate models for landslide hazard evaluation. *Mathematical Geol.*, 15, 403-426
- Carrara A., Cardinali M., Guzetti F., Reichenbach P., 1995. GIS technology in mapping landslide hazard. *Geographical Information Systems in assessing Natural Hazards*, 135-175
- Lee S., 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sensing* preview article, Taylor & Francis Ltd.
- Popescu, M. E., 1994. A suggested method for reporting landslide causes, *Bulletin of International*
- Sabins, F. F., 1996. *Remote Sensing: Principles and Interpretation*, 3rd Ed.: W. H. Freeman and Company", New York, 494 pages.
- Soeters, R., Van Westen, C.J., 1996. Slope instability recognition, analysis, and zonation. In: Turner, K.A., Schuster, R.L. (Eds.), *Landslides: investigation and mitigation, Transport Research Board Special Report*, 247, 129-177.
- Sakellariou M. G, And Ferentinou M. D., 2001. GIS-Based Estimation of Slope Stability, *Natural Hazards Rev.*, 2(1), 12-21
- Van Westen, C.J., Rengers, N., Terlien, M.T.J. And Soeters, R. 1997. Prediction of the occurrence of slope instability phenomena through GIS based hazard zonation. In: *Geol. Rundsch.* 86, 404-414

