Journal of Civil Engineering and Urbanism

Volume 6, Issue 6: ; November 25, 2016 ISSN-2252-0430

Modeling of Urban Vulnerability Probability (Supposition) Using Logistic Regression Model; Case study: Tehran city

Saeed Maleki[1]✉, Abolfazl Meshkini[2], Mohammad Ebrahimi[3], Abdolah Kloraghan[4] Atena Moin Mehr[5], and Morteza Omidipoor[6]

[1]*PhD, Associate Professor, Geography and Urban Planning, Faculty of Earth Sciences and GIS, Shahid Chamran University of Ahvaz, Iran*
[2]*Associate Professor, Department of Geography and Urban Planning, Tarbiat Modares University, Iran*
[3]*MA student of Geography and Urban Planning, Tarbiat Modares University, Iran*
[4]*MA student of GIS, Tarbiat Modares University, Iran*
[5]*MA student of Geography, Tehran University, Iran*
[6]*PhD student of GIS, Tehran University, Iran*

✉Corresponding author's email: malekis@scu.ac.ir

ABSTRACT: Surveying about city seismic vulnerability is one of the urban management requirements. Using an appropriate model with different spatial and non-spatial data and conducting relevant analysis in Geographic Information System and Multi Criteria Decision model can be as useful tools in urban management. The objective of this study is assessing the vulnerability potential of the region 5 of Tehran using logistic regression model. In this study, 14 indicators are used to estimate the vulnerability: Substructure's area of all floors in the block, distance from crisis management centers, Block's area, distance from Gas reducer center and Transmission substation, block's population, Access to communication road, old texture regions, distance from parks and green space, distance from gas and gasoline stations, distance from fault, distance from fire station centers, distance from hospitals and health centers, Average of the block's floor, Slope of the region. Output from 0.83 to 1 class of the SAW model is used as dependent variable to generate the vulnerability distribution map for Logistic regression analysis in IDRISI software. As a result, blocks population (with value of 37.05) and average of the block's floor (with coefficient of 11.6) have the greatest impact on vulnerability. Roc criterion is used for the evaluation of model. The value for this criterion is calculated as 0.996 which shows strong relationship with the values obtained from the logistic regression model.

Key words: Vulnerability, Logistic Regression, Geographic Information System, spatial and non-spatial data

ORIGINAL ARTICLE
PII: S2252043016000012-6
Received 20 Sep. 2016
Accepted 15 Nov. 2016

INTRODUCTION

Casualties, structural damages and social problems have been increased since the mid-twentieth century as a result of events caused by natural disasters (Nuno, 2012). Geographical distribution of natural disasters that lead to major damages, more casualties, as well as the instability of damaged community, is strongly associated with the retardation geography. Undoubtedly, damages of a disaster in a developing country are more than a developed country (Anne-Catherine, 1999). In the past 30 years, many developing countries have witnessed rapid development of urbanization and the spatial expansion of their cities. According to declaration of the human habitat's world conference (Habitat), world population between 1950 to 2010 has increased from 2.5 to 6.9 billion, this increase is in developing countries and their urban population has grown seven times in the 60 years (Oteng-Ababio, 2012). Today, density of population and activity in cities and their growing trend has caused the increasing demand for issue of urban sustainability against natural hazards such as earthquakes. In recent years, the vulnerability has been an important part of the research agenda in many organizations, international scientific plan and a basic analytical tool in the study of global environmental changes, as well as sustainable development (Chunliang et al., 2011). Therefore cities as socio-economic fragile systems, have high vulnerability in natural hazards, and are susceptible sites for occurring of various events. Effectively, decrease in vulnerability of urban areas, leads to decrease in costs of natural hazards (Armas, 2012). Among various levels of physical planning, Mid-level or urbanity is the most efficient level for reducing cities vulnerability against natural hazards. Clearly, until useful understanding of vulnerability's level has not been achieved, the above objective will not be trued. Based on conducted research, casualties and property damages caused by an earthquake in Tehran will be more than similar cities in developed countries and yet seismic countries such as such Japan. In the time of danger's occurrence, Tehran is faced with a lot of different problem in comparison to other cities, because of its features such as the high concentration of buildings, lack

To cite this paper: Maleki S, Meshkini A, Ebrahimi M, Kloraghan A, Moin Mehr A, Omidipoor M. 2016. Modeling of Urban Vulnerability Probability (Supposition) Using Logistic Regression Model; Case study: Tehran city. *J. Civil Eng. Urban.*, 6 (6): 94-100.
Journal homepage http://www.ojceu.ir/main/

of open spaces, lack of urban planning's standards, high population and per capita non-compliance. Region 5 covers an area with more than 5287 hectares of northwest of Tehran. Sensing the degree of vulnerability in this region is so important, because of Hisarak and Baq Feiz faults, non-proper site selection of urban facilities and equipment and increasing population of the region. Evaluation of region's vulnerability is a complex problem. This includes factors such as population density, structure of the building, number of floors, adjacent to hazardous land uses, proximity to fault, etc. These criteria can be modeled by multiple -criteria decision-making (Yi, 2013) . Combining GIS and MCDM techniques provides more accurate assessment about how to distribute the vulnerable areas. This can be used in the Crisis Management. Therefore; this study is conducted based on modeling of vulnerability probability caused by earthquake in form of urban blocks using a logistic regression model in region 5 of Tehran. Logistic regression is a statistical method belonging to generalized linear statistical models and predicts the probability of event occurring by using of independent variables.

Sarris et al. (2003) used vulnerability average index related to topology in order to study vulnerability of Barcelona (Sarris et al., 2003). Lantada et al. (2010) used RISK-UE model and GIS toolbox for assessing seismic risk in Barcelona. Results of this study show that the method is a simple statistical tool to assess the vulnerability of residential buildings and monuments, especially when limited information is available (Lantada, 2010). Sahafy (2008) used Radius model to estimate damages caused by the earthquake in region 13 of Tehran (Amini, 2010). Qadiri (2008) in his thesis studied relation between social structure of cities and degree of vulnerability in the earthquake in Tehran sectors. He concluded that there is a meaningful difference between averages of vulnerabilities degree and socio-economic base, such that a family with a lower socio-economic base has higher degree of vulnerability.

Farajzadeh et al. (2011) by using TOPSIS FUZZY model and indicators such as building materials, age of the buildings, number of floors, quality of buildings, population density, compatibility of land uses and position of buildings in the block, Assessed the region 9 of Tehran's vulnerability. Results showed the vulnerability of region and usefulness of TOPSIS FUZZY model.

AhadnezhadRoshti et al. (2010) used AHP in order to calculated vulnerability in cities against earthquake. Finally they provided earthquake scenarios with different intensity, as well as modeling the vulnerability and assessing casualties and economical damages in Zanjan.

By considering history of the research, this study provides a model of vulnerability with logistic regression method in order to gets more accurate evaluation of vulnerability in region 5 of Tehran. Advantages of modeling with Logistic regression in comparison to other multivariate statistical techniques such as Multivariate regression analysis is that dependent variable can have only two values, one is probability of occurring an event and another is probability of non-occurring an event.

Study area

Region 5 of Tehran is located in the northwest part of the city. It has been bounded from north, east, south and west to southern slopes of the Alborz, Farahzad valley, Karaj Special Road and Can valley respectively. Area of the region is 5287 hectares and population of the region is 679,108. Some Parts of the region, including can village have an old history, while some parts of the region's construction have been formed in recent years. Major physical and demographic changes have occurred in the region during last forty years (1964-2004) (Consultant Engineers, 2005).

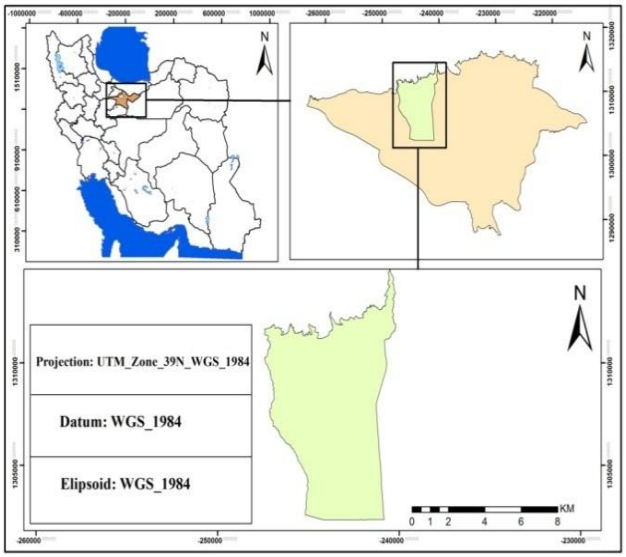

Figure 1. Study area

Urban land use planning is series of objective activities that organize the manufactured environment and meet the need of urban communities in using of lands (Hashemi et al., 2011). Land use of region 5 has been shown in the map number three.

Land distribution should be based on internal divisions of every region. There is an imbalance in distribution of public service in this region. Therefore large areas of the region which form a city block do not benefit from welfare services; however their ages are not more than two decades (Sharmand Consulting Engineers, 2005). Some land uses should be generated in a way that in addition to providing services to the residents, cause minimum damages to buildings, industrial uses, warehouses, and gas and electricity stations in the case of fire and so on. Relief centers, such as fire stations and crisis management centers should have appropriate access to the risk center in order to minimize damages and casualties during and after the crisis.

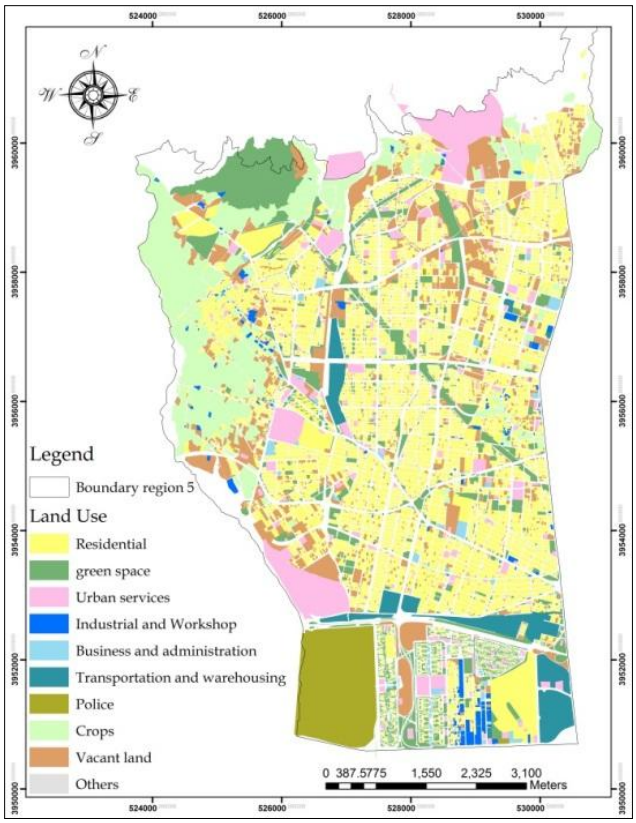

Figure 2. Land use status quo in region five of Tehran

MATERIALS AND METHODS

This study is a descriptive analytical research. Data from descriptive part have been gathered from related organizations and in analytical part, logistic regression model has been used in IDRISI software environment in order to model probability of vulnerability in region 5 of Tehran. One thing that should be noted about logistic regression model is that the independent variable layer (vulnerability) is a dual-mode variable.

In order to achieve the layer, first of all, vulnerability map of the region which is a digital layer with a value between 0 and 1 has been calculated with the implementation of SAW (Simple Additive Weighting method (SAW) is one of the oldest methods of multiple attribute decision making) model in Arc GIS environment. Then values between 0.83 to 1 with the highest degree of vulnerability have been extracted and considered as a dependent variable with value of 1 and 0 (less than 0.83 were considered as zero). Use of independent variables which have an important role in determining dependent variable, creates more complete logistic regression model. For this reason, in the study 14 layers have been used as independent variables (Table 1).

Table 1. Used Variables for vulnerability measurement of region 5 of Tehran city

Old texture regions	Access to communication road	Block's population	Distance from Gas reducer center & Transmission substation	Block's area	Distance from crisis management centers	Substructure's area of all floor in the block
Slope of region	Average of block's floor	Distance from hospitals and health centers	Distance from fire station centers	Distance from fault	Distance from gas and gasoline stations	distance from parks and green space

Logistic Regression Model

Logistic regression model is a multivariate analysis model of generalized linear statistical models. It is used for analyzing presence or absence of a feature or an output based on set of useful independent variables values. The main purpose of logistic regression is modeling probability of occurrence a common binary event, presence or absence of different factors and the significance of this presence or absence. Advantages of modeling with Logistic regression in comparison to other multivariate statistical techniques such as multivariate regression analysis, is that dependent variable can have only two values, one is the probability of occurring an event and another is the lack of it. Regression model is one of the proper statistical models for implementation of zoning based on available data. In the regression model based on available data, an equation will be achieved to predict dependent variable values (Y) based on one or more predictor independent variables (Xi). If values of dependent variable (Y) be qualitative, and have two mode (presence or absence), then obviously normal regression model would not be proper. In such case, probability of occurrence an event (P) is calculated and called logistic regression model. Its equation is as follow:

$$Y = \text{Logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_n X_n \quad (1)$$

Where

$P = \text{Logit}(P) = \text{Ln}(\frac{P}{1-P})$ is probability of vulnerability, β_0 constant number, X1 to Xn independent variables and β1 to βn are coefficients of these variables, respectively. In fact, logit (p) values are calculated in the equation (1) and then p values are extracted based on it. To solve above equation we can use weighted least squares , but this method has problems such as unclearness of P values. Instead of it, maximum similarity method is used. (2) Equation is used to extract maximum amount of similarity coefficients:

$$L = \prod_{i=1}^{N} P_i^{Yi} * (1 + p_I)^{(1-y_I)} \quad (2)$$

Where, Pi is probability of vulnerability and is calculated based on the equation (1), yi is vulnerability calculated values (according to definition is zero or one values). Logistic regression is a linear model.

According to equation (3), standardization for each layer has been done:

$$X_{Standardized} = \frac{\text{X} - \text{Min (x)}}{Max\ (x) - Min\ (x)} \qquad (4)$$

If X is considered as an input variable, then standardized values will be calculated based on above equation. Using above equation, layers are normalized and range of values is between zero and one. Finally, these layers will be entered into the logistic regression model as predictor variables. In order to evaluate best fitness of the model, Chi Square and Pseudo R Square statistic are used (4 relation):

$$\text{Pseudo R Square} = 1 - \left(\frac{\text{Log (L)}}{\text{Log } (L_0)}\right)$$

Thus, if value of Pseudo R Square be equal to one, indicating that the model has perfect fit, and if this value be equal to zero, indicating that there is no relation between selected factors and dependent factor. If value of Pseudo R Square be greater than 0.2 indicating that model has a good fit. The value of Chi Square is calculated in (5) relation:

Chi Square (K) = -2 (Log (Likehood) – Log (L_0))

Above statistic tests null hypothesis (all of the regression coefficients be zero) based on chi-square distribution function with freedom degree K. K is equals to number of variables contained in model. If this statistic be meaningful at 95% level (Chi Square> 14.1), then H0 would be rejected.

Probability of Vulnerability Evaluation and Validation

ROC: It is necessary to note that how well the model can predict dependent variable. IDRISI Software uses ROC criteria for validation of the model (Regmi, 2010). ROC curve is the most efficient way in providing deterministic characteristic, possible identify and forecasting systems that estimating accuracy of the model quantitatively. These values show model's ability to distinguish between positive and negative views in validation sample. ROC value from 0.5 to 1 is variable, 1 indicates full compliance and 0.5 reflects casual compliance (Motevali, 2009).

Discussion and Conclusions

In this study, 14 indicators are used to estimate vulnerability in region 5 of Tehran: Substructure's area of all floors in block, distance from crisis management centers, Block's area, distance from Gas reducer center and Transmission substation, block's population, Access to communication road, Old texture regions, distance from parks and green space, distance from gas and gasoline stations, distance from fault, distance from fire station centers, distance from hospitals and health centers, Average of block's floor, Slope of region. Layers have been prepared in Arc GIS toolbox (Figures of below).

Implementation of logistic regression model

In this study, vulnerability layer of the region is selected as response layer(Y) in order to reach optimal regression coefficients. So that, areas with the highest levels of vulnerability from SAW model (1-0, 83) are placed in the study area of 1 number and the rest of region(less than 0, 83) is assigned for zero number. Other extracted layers (substructure's area of all floors in block, distance from crisis management centers, block's area, distance from Gas reducer center and Transmission substation, block's population, access to communication road, old texture regions, distance from parks and green space, distance from gas and gasoline stations, distance from fault, distance from fire station centers, distance from hospitals and health centers, average of block's floor, slope of region) are considered as predictor variable.

Table 2. Coefficients and standard deviation reached from Logistic Regression

Independent variables	Coefficients	Standard Deviation
Constant number	-24.6386	-
Distance from fire station centers	3.429168	0.997635
Distance from gas and gasoline stations	0.008087	1.012725
Distance from crisis management centers	1.012428	0.993683
Distance from hospitals and health centers	3.371935	0.992438
Old texture regions	1.005451	1.001812
Distance from Gas reducer center & Transmission substation	0.089160	1.006341
Distance from fault	0.504179	1.005255
Access to communication road	-0.573498	0.993772
block's population	37.048960	1.004167
Block's area	-45.452817	1.021497
distance from parks and green space	-6.631314	0.993759
Slope of region	1.692540	0.992201
Average of the block's floor,	+11.643785	0.992198
Substructure's area of all floors in block	-46.885328	1.003381

Table 3. Statistical indicators obtained for evaluation of model

Indicator	Value
Pseudo R-square	0.8866
Chi Square	2931.1804
ROC	0.996

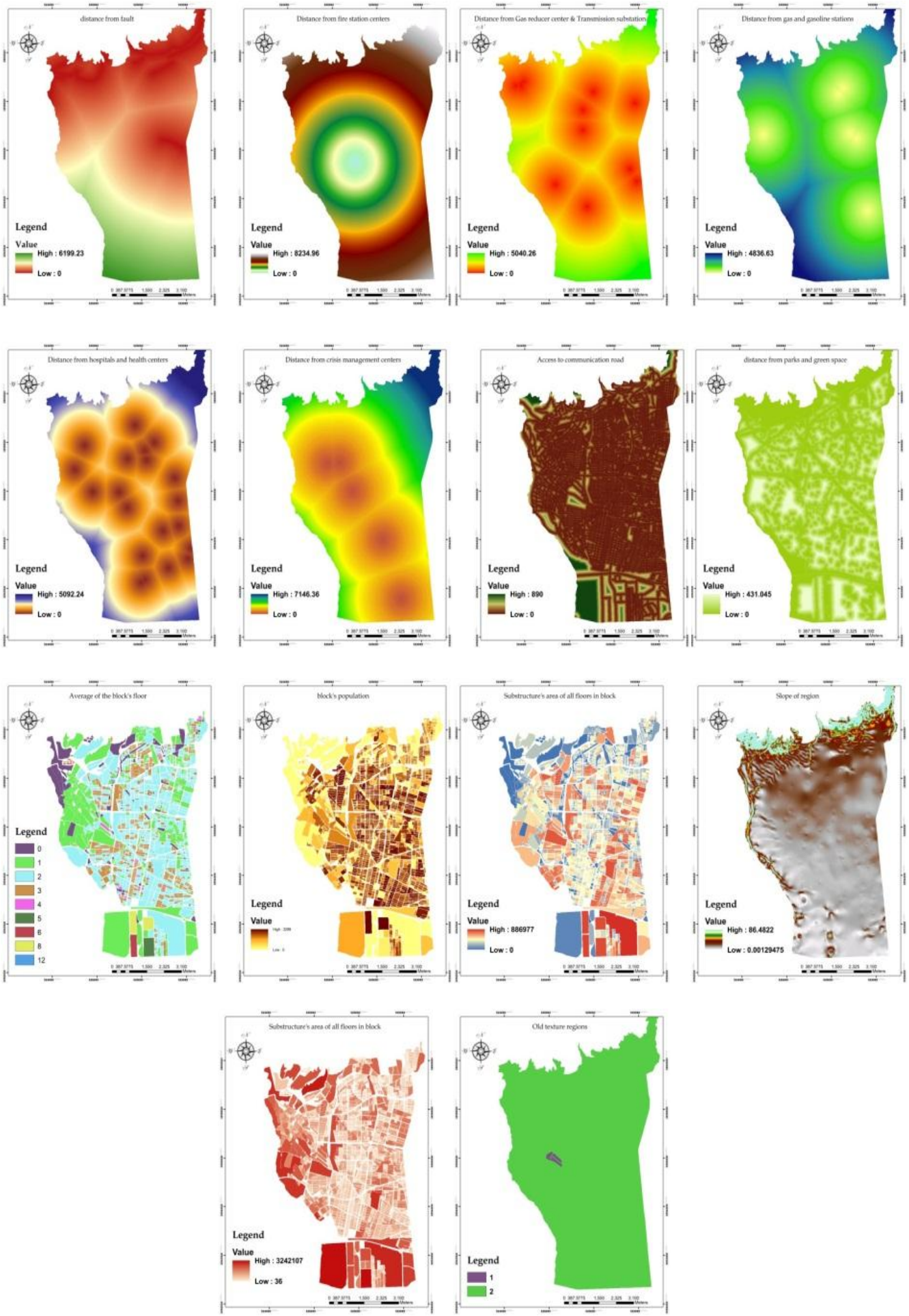

Figure 3. Digitization of layers in Arc GIS

Vulnerability distribution map

There is no standard for numbers of independent parameters in assessing vulnerability degree of an urban area. Therefore with regard to spatial differentiation issue, even two similar regions with a common parameter may have different coefficient which originated from specific features of region. According to vulnerability distribution map extracted from the implementation of logistic regression model, vulnerable regions are around central core and by getting away from the central area the degree of vulnerability will be reduced. In the present model, number of block's population with the highest coefficient is the best variable to predict probability of vulnerability in region.

Coefficients reached from model (Table 2) show that the vulnerability distribution in region has positive relation with parameters such as distance from fire station centers, distance from gas and gasoline stations, distance from crisis management centers, distance from hospitals and health centers, old texture regions, distance of Gas reducer center & Transmission substation, distance from fault, block's population, slope of region, average of block's floor and negative relation with access to communication road, block's area, substructure's area of all floors in block and distance from parks and green space.

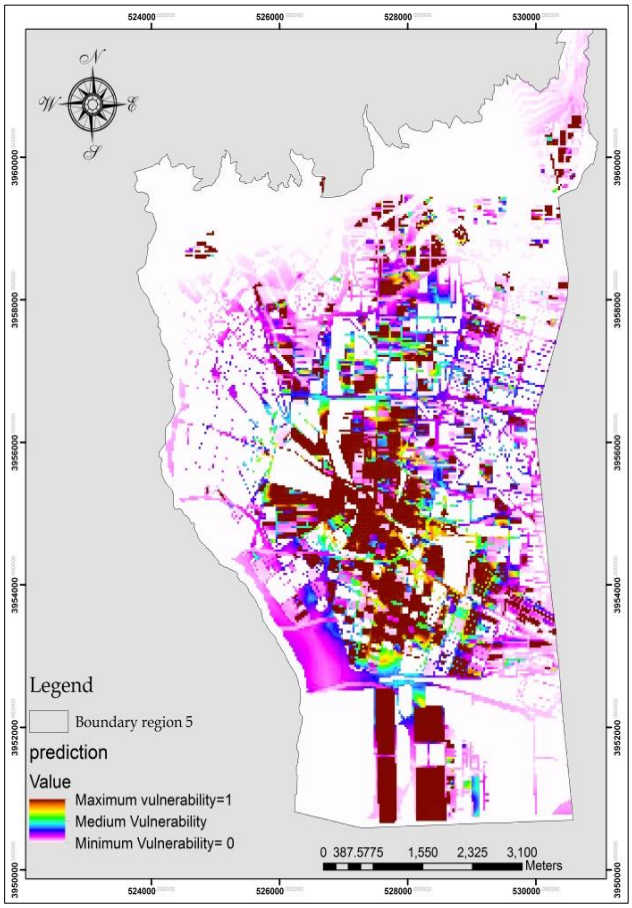

Figure Map No. (8): Probability of vulnerability map using logistic regression model in five region of Tehran

CONCLUSION

All researchers in field of natural hazards believed that occurring events such as earthquake is inevitable but its damages can be minimize by proper management. Recognition of geographical distribution of buildings with different degree of vulnerability has been possible by GIS software. Following from rules of urbanity, distribution and proper site selection of land uses in city and urban areas in the form of urban land use play an important role in minimizing casualties and damages. Today, use of GIS to determine regions with the most vulnerabilities probability has been common. Therefore, correct reorganization of vulnerable buildings and factors influencing vulnerability as well as adopting proper long-term and short-term policies leading to decrease in negative effects of future crisis. In this study, land use indicators are used to estimate vulnerability and effect of every indicator on vulnerability of region 5 of Tehran municipality. Effects of earthquake often manifested in the form of physical destruction. According to the final map, geographical distributions of urban vulnerable blocks are in form of clusters and constitute major parts of central area. Value of coefficients also indicates that the layers such as block's population, average of block's floor, distance from fire station centers, and distance from hospitals and health centers, respectively have the most effect on vulnerabilities probability in 5 region of Tehran. In this study, Pseudo R-Square indicator has been calculated as 0.8866, this value is near to the number one and show acceptable fit of the model. Chi Square value is calculated as 1804.2931, because this value is more than thresholds value, then the hypothesis that all of the coefficients are zero is rejected. The area under the ROC curve is selected as 0.996 which is very high value. This shows there is strong relationship between risk probability and probable value reached from logistic regression model. According to results, logistic regression can be used as an accepted method for vulnerability zoning.

Competing interests

The authors declare that they have no competing interests.

REFERENCES

AhadnezhadRoshti et al. (2010). Modeling of cities building vulnerability to earthquake by using of AHP and GIS (case study: Zanjan city), Journal on Geography & Development, N.19.PP.171-198.

Amini J (2010). Analysis of urban housing vulnerability to earthquake (Case study: Region 8 of Tehran municipality, Dissertation of remote sensing and GIS MA, Tarbiat Modares University, Iran.

Ann Sarris et al. (2010). Earthquake vulnerability and seismic risk assessment of urban areas in high

seismic regions: application to Chania City, Crete Island, Greece. Nat Hazards 54:395–41.

Anne-Catherine Ch (1999). A geographic approach of the global vulnerability in urban area: case of Manizales, Colombian Andes. GeoJournal 49:197–212.

Chunliang K, Stillwell J, Bakema A and Scholten HJ (2010). Modelling Land-Use Change. Progress and Applications.

Dai FC, Lee Cf and Ngai, YY. (2002). Landslide Risk Assessment and Management: an Overview. Engineering Geology, 64: 65-87.

Garcia, MJ, Malplica B and Diaz M (2008). Susceptibility Assessment of Earthquake Triggered Landslide in EI Salvador using Logistic Regression. Geomorphology, 95: 172-191.

Lantada N. et al. (2010). Seismic hazard and risk scenarios for Barcelona, Spain, using the Risk-UE vulnerability index method. Bull Earthquake Eng 8: 201–22.

Lee SA (2004) .Verification of spatial logistic regression for landslide susceptibility analysis: Acasestudy of Korea, Geomorphology, 44 :15-18.

Iuliana A (2012). Multi-criteria vulnerability analysis to earthquake hazard of Bucharest, Romania. Nat Hazards 63:1129–1156.

Motevali, S et al. (2009).Determine of land move sensitive by using of Logistic Regression in water basin region of Mazandaran province, Journal of Physical Geography, N.2 (5).PP.73-83.

Oteng-Ababio M (2012). Neglected vulnerabilities in a rapidly urbanizing city: reflections on earthquake risks in Accra. J House and the Built Environ, 27:187–209.

Pradhan B (2010). Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. Adv. Space Res 45(10):1244–1256.

Qadiri M (2008).Relation of cities social structure and vulnerability level to earthquake (Case study: Neighbourhoods of Tehran, Thesis of PhD of Geography & Urban Planning, Tarbiat Modares University, Iran.

Regmi NR, Giardino JR and Vitek JD (2010). Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA, Geomorphology, and 115: 172–18.

Regmi, NR, Giardino JR and Vitek JD (2010) .Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA, Geomorphology, and 115: 172–187.

Sahafy A (2008). Analysis of urban housing vulnerability to earthquake (Case study: Region 13 of Tehran, Dissertation Geography & Urban Planning, MA. Tehran Modares University, Iran.

V. Nuno Martins Delta Sousa e Silva Pedro Cabral . (2012). Social vulnerability assessment to seismic risk using Multicriteria analysis: the case study of Vila Franca do Campo (Sa˜o Miguel Island, Azores, Portugal). Nat Hazards 62: 385–404.

Yi Peng (2013). Regional earthquake vulnerability assessment using a combination of MCDM methods.