A Logistic Regression Method for Urban growth modeling Case Study: Sanandaj City in IRAN

Sassan Mohammady
GIS MSc student, Dept. of Surveying and Geomatics Eng., College of Eng. University of Tehran, Tehran, Iran

Paper Reference Number: 07-63-1000
Name of the Presenter: Sassan Mohammady

Land use activity is a major issue and challenge for town and country planners. Different approaches have been attempted in spatial modeling. In this research we use logistic regression, that have been very successful in interpreting socio-economic activities. Our case study is Sanandaj and we use Landsat imagery, taken in 2000 and 2006.Our parameters in this study are: distance to principle roads, distance to residence region, elevation, slope, distance to faults, distance to facilities, distance to downtown and the number of urban pixels in an neighborhood with radius of 1 pixel. One of the most important problem in this analysis is large volume of data. We use different method to find training data and consider accuracy of model, and finally we predict land us change for 2012, 2018.

Key words: Regression Logistic, Urban Growth Modelling

1. Introduction

The major tasks of urban planners are urban land allocation to different applications with a special focus on the role and function of the city, its economy, and the ability to simulate the effect of user interaction with each other. Continuing migration of rural population to cities and population increases, many problems of today's cities, including the expansion of urban areas, lack of infrastructure and urban services and environmental pollution are facing these issues are directly related to land and land use. The impact of piecemeal planning in large cities is often a major concern of stakeholders, such as those involved in research modeling, forecasting and policy making related to planning sustainable urban development (Barredo et al., 2004). Dynamic spatial urban models provide an improved ability to assess future growth and to create planning scenarios, allowing us to explore the impacts of decisions that follow different urban planning and management policies (Kaiser et al, 1995; Klostermann, 1999).

2. Data and Material

Remote sensing techniques have already shown their value in mapping urban areas, and as data sources for the analysis and modeling of urban growth and land use change (Batty et al, 2001; Clarke et al., 2002; Donnay et al, 2001; Herold et al, 2001; Jensen et al, 1999). Remote sensing techniques and the availability of free to less expensive data sources of satellite imagery and their temporal frequency has greatly enhanced the potential for monitoring urban growth (Im, J.,2008, Goodchild, M.F.2000), urban land use dynamics (Herold, M,2003), landscape pattern analysis (Li, X,2004), and urbanization (Weng, Y,2007). The study area in this research is the city of Sanandaj, Iran. Data used in our research, come from two satellite images related to years 2000 and 2006 from Landsat satellite and are from TM & ETM^+ sensors which, they have earth pixel size of 28.5 meters. The main road map, attractive areas (such as parks), faults, slope, elevation, land use maps and others also are formatted in Shape file, using software Arc GIS 9.3 ESRI. All processes on satellite images are done using software ENVI 4.7.

2.1 Logistic Regression

Statistical methods can easily recognize the effects of the independent variables and provide the reliability with regard to their contribution. Logistic regression as a statistical technique is more general case of linear regression. In most cases, these models provide a good fit to the spatial processes and land use changes. Our aim is modeling the dynamics of urban growth process and at the other side the ability to interpret the models is very important. Due to the discrete nature of the land use change, a common method to approximate the logistic regression is a function which the development probability for each pixel is determined. In this model, observations are pixels. Binary dependent variables representing urban or non-urban status of the pixels in the corresponding period of the model output. This function is a monotone curvature and output function is between zero and one. The following equation represents the regression function is reasonable.

$$P = \frac{\exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)} \tag{1}$$

P: probability of land use change for each cell

Xi : effective parameters in urban growth

β_0: constant parameter

β_i : Coefficients of each of the independent parameters that must be calculated

The output of the logistic regression, is provider of urban expansion by using variables that are exponential functions of their elements. Various coefficients determine with using the method of least squares.

3. Research Methodology

The Landsat imagery of Sanandaj city is shown in Fig. 1.After omitting geometrical errors and fixing image distortions, the related images are prepared based on Maximum Likelihood method which, is one the classified supervised methods (Fig 2). Our parameters are distance to street, distance to attractive area, distance to fault, distance to downtown, slope, elevation and the number of urban pixels in an neighborhood with radius of 1 pixel. Based on satellite images related to years 2000 and 2006 and using logistic regression method, we have calibrated the model with 10 percent of our data. The purpose of calibration is to establish the relationship between land use change and the factors that affect probability of land conversion (Wu, 2002). To eliminate correlated variables, the covariance between variables were examined. If Variables covariance exceeded 0.8, according to the value of the variable, one of them must be removed. Map of land use change probability is a tool to detect changes in land use. Each cell probability is between zero and one. If probability of cell is closer to 1, that cell is more likely to change. 19762 points (10% our data) were selected randomly for training and 90% another data for checking our model were used.

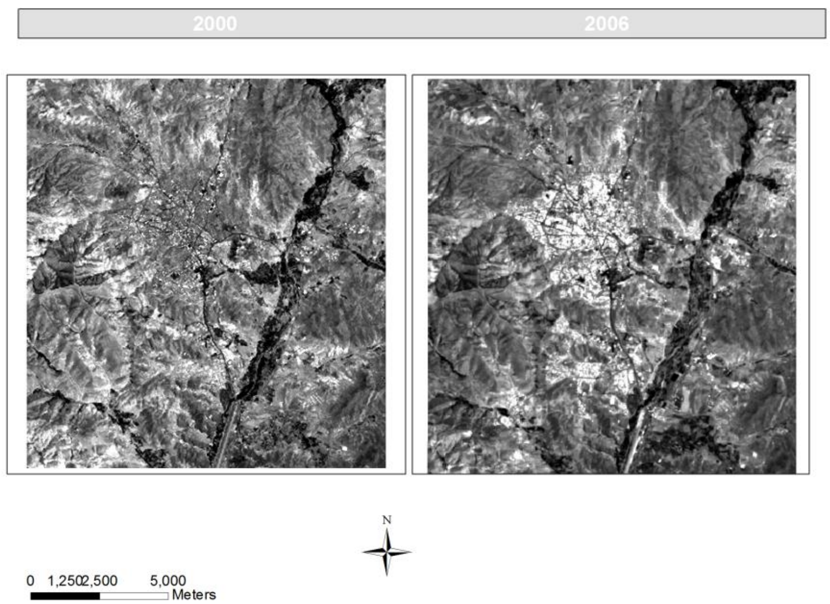

Fig.1 Satellite images of Sanandaj in 2000 and 2006,respectively

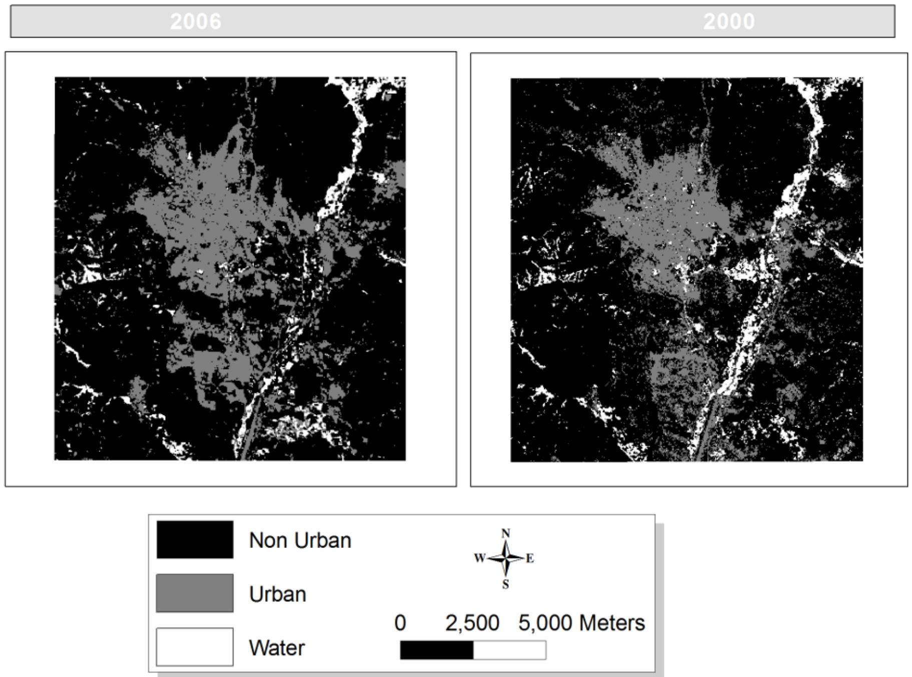

Fig.2 Sanandaj classified land use map in 2006 and 2000,respectively

4. Results and Analysis

The statistical results of each explanatory variable are given in table 1. All the 8 variables are significant in the logistic regression model based on a two-tailed test at 95% Confidence Level (CL). A variable with the highest coefficient is the most significant while a variable with the lowest coefficient is the least significant. Therefore distance to main road is the most significant in the model while distance to fault is the least significant in the model.

Variable	Coefficient	Standard error
Distance to Downtown	4.823	0.506
Distance to Residence region	8.205	0.943
Distance to Facility	-5.479	0.528
Elevation	0.314	0.231
Slope	-0.398	0.189
Distance to Main Road	13.492	0.971
Distance to Fault	0.298	0.309
Number of Urban pixel in 3*3 neighborhood	3.236	0.113

Table1. Logistic regression results for 2000-2006 simulation (* significant at <0.05)

4.1 Percent Correct Match (PCM)

Percent correct match is a way to evaluate models of urban development. In fact, this method only considers the special case of the comparison matrix is as follows.1) In fact, urban development has taken place and model been able to predict urban development. 2) In fact urban development has not occurred and model has been modeled correctly. This method compares only the parameters of the original diameter of the A and D matrices are used to assess the shortcomings of this method is considered. The Percent Correct Match (PCM) is calculated based on Confusion matrix (table2). (Pontius and Schneider, 2001)

Model	Reality		
	Change	Non-change	Total
Change	A	B	A+B
Non-change	C	D	C+D
Total	A+C	B+D	A+B+C+D

$$PCM = \frac{A+D}{A+B+C+D} \quad (2)$$

Table2.Confusion Matrix

4.2 Kappa Statistics

Kappa coefficient, a statistical method for evaluating models of urban development in general is widely used in spatial issues. This coefficient shows the rate of compatibility between simulated and reality. In other words, this factor can be used to measure the spatial distribution of the amount of similarities between the two maps. It is generally considered that Kappa values for map agreement are: >0.8 is excellent; 0.6-0.8 is very good; 0.4-0.6 is good; 0.2-0.4 is poor and <0.2 very poor (Pijanowski et al., 2005). The calculation of Kappa is based on contingency table (Monserud and Leemans, 1992) (Table 3).

	Model					Total
	Class	1	2	…	c	
Reality	1	P_{11}	P_{12}	…	P_{1c}	P_{1T}
	2	P_{21}	P_{22}	…	P_{2c}	P_{2T}
	…	…	…	…	…	…
	c	P_{c1}	P_{c2}	…	P_{cc}	P_{cT}
Total		P_{T1}	P_{T2}	…	P_{Tc}	1

$$P(A) = \sum_{i=1}^{c} P_{ii} \quad (3)$$

$$P(E) = \sum_{i=1}^{c} P_{iT} . P_{Ti} \quad (4)$$

$$KS = \frac{P(A) - P(E)}{1 - P(E)} \quad (5)$$

Table3.The Contingency Table

The simulation of 2006 from the 2000 was executed for the 6 years period. The simulated 2006 map was validated with the actual map of 2006, by comparing the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) matches using both maps. The overlay map of the simulated and actual 2006 maps is presented in Figure 3. A performance matrix derived from the result shown in Figure 3 is given in Table 4.

	Reference Data 2006	
	Developed	**Undeveloped**
Simulated Data 2006		
Developed	**18213 (TP)**	**4119 (FP)**
Undeveloped	**6736 (FN)**	**148795 (TN)**

Table4**.** The Performance matrix

The computed Kappa statistics is 0.7353, and PCM is 93.90%. Based on the earlier work (Pijanowski, et al., 2005) Kappa values calculated for simulated land use map seems very good.

Conclusion

In this paper, we simulated urban growth modeling for Sanandaj city in Iran. The eight land use explanatory variables that were used for the modeling were all significant at 95% CL which implied that all the explanatory variables contributed to the urban expansion of Sanandaj. There isn't a big covariance between these variable. So, all of them are independent from each other. The computed kappa statistics and PCM was 0.7353 and 93.90% respectively. And we can say our simulation results have a very good agreement (Pijanowski et al., 2005) with our reference data (Fig 3).We predict urban bounders for 2012 and 2018 (Fig 4).

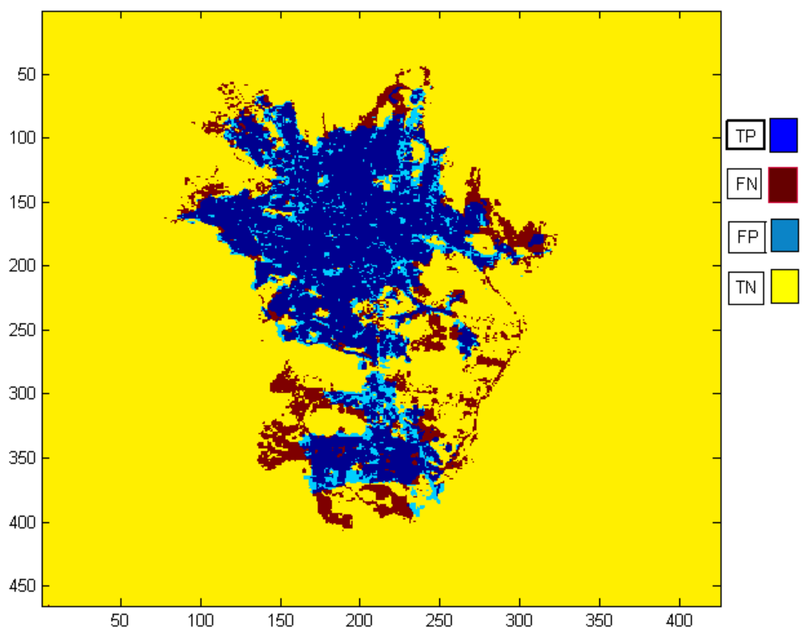

Fig 3. Simulation result of 2006.

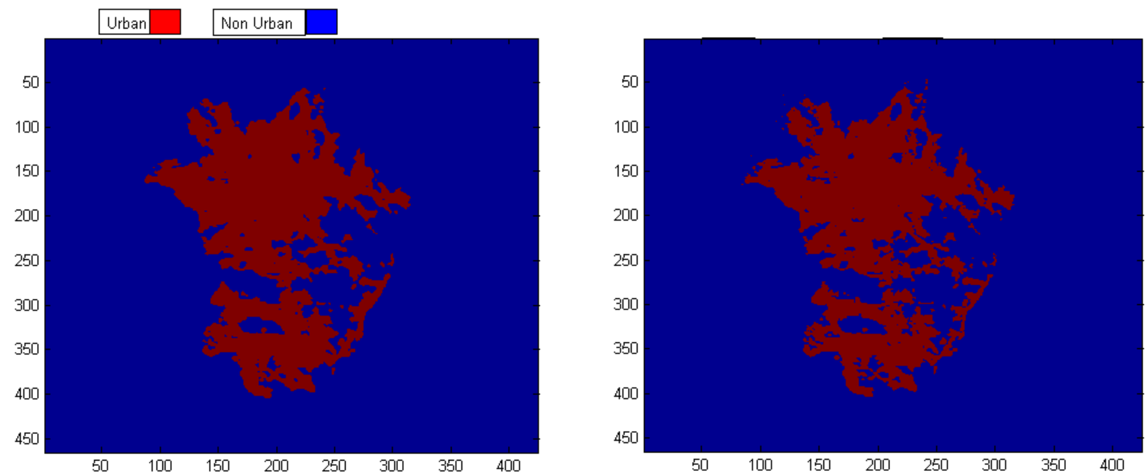

Fig4. Predict for 2018,2012,respectively

Acknowledgements

I would like to thank the following people for their support, without whose help this work would never have been possible: Dr. Delavar and Prof. Pijanowski.

References

Barredo, J.I., Demicheli, L., Lavalle, C., Kasanko, M., McCormick, N., 2004. Modelling future urban scenarios in developing countries: an application case study in Lagos, Nigeria. Environment and Planning B: Planning and Design 32, 65–84.

Batty,M., & Howes, D. (2001). Predicting temporal patterns in urban development from remote imagery. In J. P. Donnay, M. J. Barnsley, & P. A. Longley (Eds.), Remote sensing and urban analysis (pp. 185–204). London: Taylor and Francis.

Clarke, K. C., Parks, B. O., & Crane, M. P. (2002). Geographic information systems and environmental modeling. New Jersey: Prentice Hall.

Donnay, J. P., Barnsley, M. J., & Longley, P. A. (2001). Remote sensing and urban analysis. In J. P. Donnay, M. J. Barnsley, & P. A. Longley(Eds.), Remote sensing and urban analysis (pp. 3– 18). London: Taylor and Francis.

Goodchild, M.F. Spatial analysis: Methods and problems in land use management. In Spatial Information for Land Use Management; Hill, M.J., Aspinall, R.J., Eds.; Gordon and Breach Science Publishers: Singapore, 2000; pp. 39-50.

Hagen, A. (2003). Multi-method assessment of map similarity. International Journal of Geographical Information Science, 17(3), 235-249

Herold, M.; Goldstein, N.; Clarke, K. The spatio-temporal form of urban growth: Measurement, analysis and modeling. Remote Sens. Environ. 2003, 85, 95-105

Herold, M., Menz, G., & Clarke, K. C. (2001). Remote sensing and urban growth models—demands and perspectives. Symposium on remote sensing of urban areas, Regensburg, Germany, June 2001, Regensburger Geographische Schriften, vol. 35, on supplement CD-ROM.

Im, J.; Jensen, J.; Tullis J. Object-based change detection using correlation image analysis and image segmentation. Int. J. Remote Sens. 2008, 29, 399-423.

Jensen, J. R., & Cowen, D. C. (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. Photogrammetric Engineering and Remote Sensing, 65(5), 611 – 622.

Kaiser, E., Godschalk, D., & Chapin Jr., S. F. (1995). Urban land use planning. Urbana, IL: University of Illinois.

Klostermann, R. E. (1999). The what if? Collaborative planning support system. Environment and Planning. B, Planning and Design, 26, 393– 408.

Li, X.; Yeh, A.G.O. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landscape Urban Plan. 2004, 69, 335-354.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. Ecological Modelling, 62, 275-293.

Pijanowski, B. C., Pithadia, S., Shellito, B. A., & Alexandridis, K. (2005). Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. International Journal of Geographical Information Science, 19(2), 197-215. doi: 10.1080/13658810410001713416

Pontius Jr, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment, 85(1-3), 239-248. doi: 10.1016/s0167-8809(01)00187-6

Weng, Y. Spatiotemporal changes of landscape pattern in response to urbanization. Landscape Urban Plan. 2007, 81, 341-353.

Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversions. International Journal of Geographical Information Science, 16(8), 795-818.