Publications Prepared for the 2017 and 2018 International Conference on Applied Analysis and Mathematical Modeling Special Issue was published by reviewing extended papers

Research Article

MODELING LAND USE SIMULATION OF ISTANBUL FOR 2023 WITH LOGISTIC REGRESSION

Cemre Fazilet ALDOĞAN*[1], Ömer AKIN[2], Ahmet EROĞLU[3], Hande DEMİREL[4]

[1]*Istanbul Technical University, Informatics Institute, ISTANBUL;* ORCID:0000-0003-4502-3479
[2]*Istanbul Technical University, Faculty of Civil Engineering, ISTANBUL;* ORCID:0000-0002-8109-0313
[3]*Istanbul Technical University, Faculty of Civil Engineering, ISTANBUL;* ORCID:0000-0002-8478-4900
[4]*Istanbul Technical University, Faculty of Civil Engineering, ISTANBUL;* ORCID:0000-0003-0338-791X

Received: 17.08.2018 Revised: 13.12.2018 Accepted: 13.12.2018

ABSTRACT

Land cover and land use modelling provides key information for understanding the environmental processes and evaluating the contributory factors leading to the land use changes. Therefore, the possible driving forces of them are required to be revealed thoroughly to give tangible explanations about the nature of land use and land cover change. The study area is Istanbul that is a megacity, which accommodates millions of people, and its population continues to increase as time passes. Due to the fact that population in Istanbul increases, urban areas start to grow rapidly and the number of people who resides in the cities grows as a result. Land management of Istanbul transforms into complex structure. In order to assist decision makers, a land cover/use prediction model of Istanbul was created for quantifying these factors in the model. In this paper, logistic regression was used for predicting land use change and simulating a land use model for the year of 2023. By means of the developed model, quantitative analyses could be incorporated to the decision making process of future city components.
Keywords: Geographic information systems, land use change modelling, logistic regression.

1. INTRODUCTION

Istanbul is considered the most important city of Turkey economically, culturally and historically. Moreover, it is the most crowded city in Turkey. The Istanbul Strait connects the Sea of Marmara and the Black Sea, and divides the city into European and Asian (Anatolian) sides, making it a strategically important spot as well. The population of Istanbul was just 4.7 million in 1980 before growing to 10 million in 2000, and finally to 15 million in 2017. Similar to other mega-cities in developing countries, Istanbul has faced rapid urbanization and motorization in the past three decades in conjunction with population growth, mainly due to internal immigration, and economic growth.

Land use change is a phenomenon, which has a direct impact on the environment and, the balance between human needs and ecosystem is an issue to maintain. Land use scenarios is significant for assessing the change quantitatively and knowing the accurate capacity of a region where is studied. They are the essential part of management processes on preventing urban sprawl

* Corresponding Author: e-mail: aldogan17@itu.edu.tr, tel: (212) 221 48 19

and preserving the resour ces for benefiting them at an optimized level. Since the problem is a spatial problem, it can be seen that the integrated model needed can only be solved by a Geographic Information System (GIS) based model [1, 2]. However, for many cities, there are no land use data for the past years. Similar to these cities, historical land use/land cover data for the metropolitan area of Istanbul is not at the expected sensitivity and resolution. Although land use changes can be monitored by traditional land surveys, remotely sensed imageries provide efficient means of obtaining information on temporal trends and spatial distribution of urban areas in less time, effort and cost [3] and GIS is used to model and analyze the land use/land cover classes obtained from remotely sensed imageries. For the scope of this study, five land use classes which are urban areas, industrial and commercial areas, barren lands, forests and water structures are created from remotely sensed images. Besides seeing the relation between the classes quantitatively, it is also possible to predict future land use/land cover by modeling with reference to past patterns. There are several simulation techniques in the literature to predict the future land use/cover such as linear regression, logistic regression, artificial neural network, agent based modeling [4, 5, 6, 7]. In this study, logistic regression is used to predict land use pattern of 2023.

Prediction of the future land use assists policy makers to make better decisions about the administration of the land depending on the growing population and increasing agricultural areas. High resolution land use maps can be used to analyze the pattern and simulated maps can help the decision makers to plan infrastructure or transportation investments. It is also important to control the growth in the city.

2. STUDIES

Study area comprises of 19 districts that are located in the European part of Istanbul as shown in Figure 1. Administrative boundaries of the study area illustrated in Figure 2 were received from Open Street Map.

Figure 1. Study Area

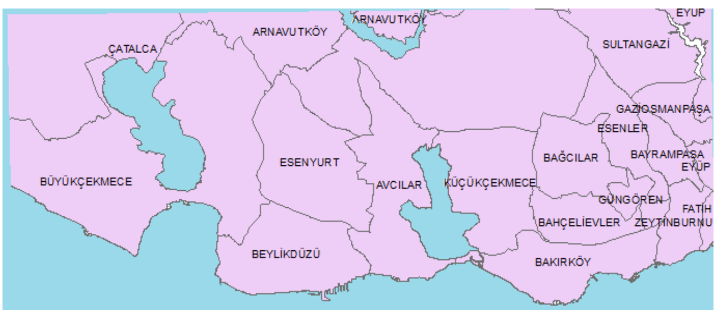
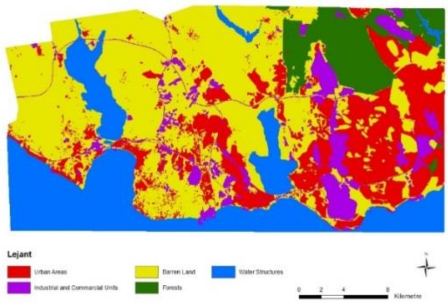
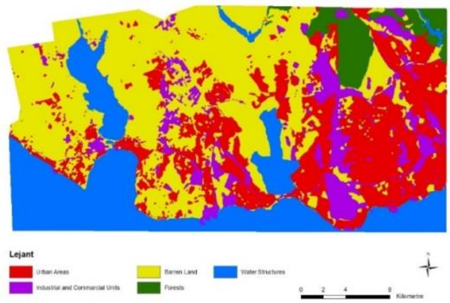


Figure 2. Administrative Boundaries

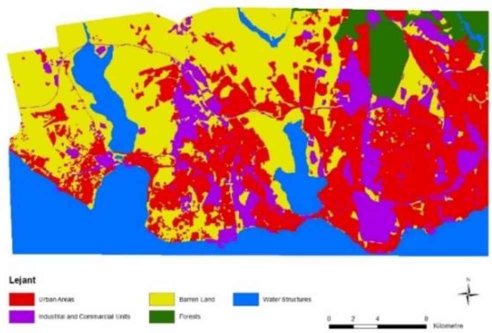
In the study, three remotely sensed mult i spectral imageries illustrating the study area are acquired in three dates that are 1997, 2007 and 2014. The imageries' spectral resolutions are 20*20, 20*20 and 1.5*1.5 respectively. The imageries are geo-referenced using corresponding orthophotos resulting Root Mean Square Errors (RMSE) of 1.7, 1.5 and 0.75 meters respectively. Supervised classification using Maximum likelihood as parametric rule is implemented for all imageries to determine the land use classes. Urban areas, industrial and commercial areas, barren lands, forests and water structures classes are created after the classification process. Results of the classifications and the area covered by each class in three years (1997, 2007 and 2014) are shown in Figure 3 and Table 1 respectively.



a) Land Use Map of 1997



b) Land Use Map of 2007



c) Land Use Map of 2014

Figure 3. Land use dynamics of the study area

Table 1. Areas covered by land use classes

Classes	**Area (km²)**		
	1997	**2007**	**2014**
Urban Areas	157.30	190.50	237.65
Industrial and Commercial Units	58.34	78.47	99.29
Barren Land	314.54	290.82	233.9
Forests	64.27	35.2	32.96
Water Structures	175.95	175.08	165.57
Total Area	770.40	770.07	769.37

As seen from the table, in the years that studied a significant increase was observed in the "urban areas" and "industrial and commercial unit" classes. An opposite trend was observed in forests and barren lands. This trend is observed in previous studies for Istanbul [8].

After achieving the land use patterns of 1997, 2007 and 2014, land use pattern of 2023 was created with the logistic regression technique for the urban and industrial classes separately. Logistic regression is a mathematical model that tries to predict the probability of a situation using the relationship between multiple independent variables and a dependent variable [9]. The dependent variable takes one of the values 1 or 0 according to the probability of the event. Since the dependent variable is binary in the logical regression, the curve to be formed by the model is in the form of "s". When a threshold value is determined on this curve, the dependent variable is 1 at the points that are greater than that value, and the dependent variable is 0 at the points with a value less than the value. The number 1 indicates the change, and the number 0 indicates the possibility of no change. This method is suitable and used to investigate the likelihood of a change of land cover / land use [10, 11, 12].

The relationship between the probability of land class change in each cell in the land cover / use map and the driving factors that may affect this change are explained by the mathematical model in equality 1.

$$\log\left(\frac{P_{ij,A}^{t}}{1-P_{ij,A}^{t}}\right) = \beta_0 + \beta_1 X_{ij,1}^{t} + \beta_2 X_{ij,2}^{t} + .. + \beta_n X_{ij,k}^{t} \quad (1)$$

In the equation, the probability of urbanization in the cell ij is $P_{ij,A}^{t}$. β_0, the equation constant and $\beta_1, \beta_2, .. \beta_n$ are the coefficients of the driving factor $X_{ij,k}^{t}$ at time *t* in cell *ij*.

Land use change modelling was implemented for urban areas and industrial areas separately in this study. The probability of transition from barren land to urban areas and barren land to industrial areas are dependent variables in the model. Independent variables are distance to roads, distance to industrial areas, digital elevation model, slope, population, network density, distance to hospitals, distance to universities and distance to forest. The closeness to the roads and the road density in the regions are important factors affecting the equation because it shows a spatial character that makes it possible to reach from one place to another. If an area is close to the roads, the odds are high. Likewise, factors such as distance to commercial/industrial centers and population can be considered important spatial variables since they make it possible to grow urban in the regions. The closeness to units such as hospitals and universities is also an important indicator that affects the equation because it will lead to the expansion of the urban region established in the region and to urbanization in the non-urban regions. The closeness to such zones is a socio-economic factor associated with individual preference, then decision-making. Persons may wish to reside in a region close to the hospital. The digital elevation model and slope map have also been used as one of the determining factors in the model since the elevation of the region can be considered as an important factor affecting these dynamics.

3. RESULTS

In order to initiate logistics regression, the model designed should be validated from previous years land cover/use maps. For this purpose the land use map of 2014 was acquired using 1997 and 2007 land use maps. Kappa is calculated on the basis of the projected and observed values over the entire region [13]. The equation used is as follows:

$$Kappa = \frac{P_0 - P_c}{1 - P_c} \quad (2)$$

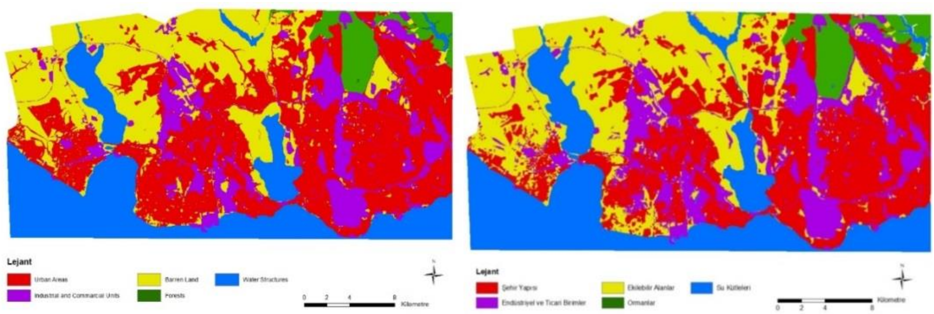
Where *Po* is the correct percentage of the model output, and *Pc* is the expected correct percentage. The value of Kappa ranges from 0 to 1. Although there is no universally accepted standard, a Kappa value greater than 0.8 is considered to be an indication of a strong agreement between the predicted and observed maps [14]. In this study, calculated kappa value is 0.83, which indicates that the model was built correctly. The coefficients of the independent variables of the model are given in Table 2.

Table 2. Coefficients of the independent variables used in logistic regression

Independent Variables	Barren land to Urban	Barren land to Industrial
Constant	-1.7772	-1.8583
Distance to Roads	-0.0087	-0.0074
Distance to Industrial Areas	0.0053	
Distance to Hospitals	-0.0094	-0.0018
Distance to Universities	-0.0012	-0.0016
Distance to Forest		0.0001
Network Density	0.1248	0.0523
Slope	0.0048	-0.055
Digital Elevation Model	0.0082	0.01
Population	-0.0014	0.0041

In the logical regression method, the validity and the sensitivity of the simulation is determined by examining the receiver operating characteristic (ROC) curve. Curve occurs with the ratio of sensitivity to precision in cases where the threshold of discrimination is different in binary classification systems. A ROC value greater than 0.5 indicates that the findings of the change in simulation are compatible with the model produced by logical regression.

After validation step, with the logistic regression analysis implemented, both the change in the urban areas and the change in the industrial areas for 2023 are examined separately and the results are mapped in Figure 4. In Figure 4.a, simulation showing the change from barren land to urban areas has ROC value as 0,87. In Figure 4.b, simulation showing the change from barren land to ındustrial areas has ROC value as 0.85. When the results are analyzed, it is predicted that the urban area is 237,65 km^2 in 2014 and this area will be 286,96 km^2 in 2023. The industrial area is 99.29 km^2 in 2014 and it will be 118,82 km^2 in 2023. The increase in urban areas is 20.75% and the increase in industrial and commercial areas is 19.67%.



a) Simulated Urban Areas in 2023 b) Simulated Industrial Areas in 2023

Figure 4. Land Use Maps of 2023

By means of the developed model land cover/use simulations, quantitative analyses could be incorporated to the decision making process, interacting land cover/use classes could be managed and controlled. This study could be enhanced via introducing constraints such as water-shed areas, green areas, development zones, where simulation is forced not to sprawl over these areas. Further analysis could be performed that enables scenario-based analyses. Overall, this quantitative method could aid policy makers for establishing and controlling valid procedures for urban management.

Acknowledgement

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK), Grant No: CAYDAG-115Y692. We also would like to thank IETT for contributing to the project.

REFERENCES

[1] Horner, M. W., Schleith, D., (2012) Analyzing temporal changes in land-use–transportation relationships: ALEHD-based approach, *Applied Geography*, 35(1-2), 491-498.

[2] Gehrke, S. R., Clifton, K. J., (2016) Toward a spatial-temporal measure of land-use mix, *Journal of Transport and Land Use*, 9(1), 171-186.

[3] Elvidge C.D., Sutton P.C., Wagner T.W., (2004) Land Change Science: Observing, Monitoring, and Understanding Trajectories of Change on the Earth's Surface. *Kluwer Academic Publishers*, Dordrecht, Netherlands.

[4] Overmars K. P., Koning de G. H. J., Veldkamp A., (2003) Spatial autocorrelation in multi-scale land use models, *Ecological Modelling*, 164, 257–270.

[5] Arsanjani J. J., Helbich, M., Kainz W., Boloorani A. D., (2013) Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion, *International Journal of Applied Earth Observation and Geoinformation*, 21, 265–275.

[6] Jiménez, A. A., Vilchez, F. F., González, O. N., Flores, S. M. M., (2018) Analysis of the Land Use and Cover Changes in the Metropolitan Area of Tepic-Xalisco (1973–2015) through Landsat Images, *Sustainability*, 10(6), 1-15.

[7] Tian, G., & Qiao, Z., (2014) Modeling urban expansion policy scenarios using an agent-based approach for Guangzhou Metropolitan Region of China, *Ecology and Society*, 19(3).

[8] Cetin, M., Demirel, H., (2010) Modelling and Simulation of Urban Dynamics, *Fresenius Environmental Bulletin*, 9(10A), 2348-2353.

[9] Cox, D. R., (1958) The regression analysis of binary sequences (with discussion), *Journal of Royal Statistical Society*, 20(2), 215–242.

[10] Verburg, P. H., Soepboer, W., Veldkamp, A., (2002) Modelling the spatial dynamics of regional land use: The CLUE-S Model, *Environmental Management*, 30(3), 391–405.

[11] Nong Y., Du, Q., (2011) Urban growth pattern modeling using logistic regression, *Geo-spatial Information Science*, 14(1), 62-67.

[12] Jiang, W., Chen, Z., Lei, X., Jia, K., Wu, Y., (2015) Simulating Urban Land Use Change by Incorporating an Autologistic Regression Model into a CLUE-S Model, *Journal of Geographical Sciences*, 25(7), 836-850.

[13] Congalton, R. G., and Green, K., (1999) Assessing the accuracy of remote sensed data. *CRC Press*, New York, USA.

[14] Tian, G., Ouyang, Y., Quan, Q., and Wu, J., (2011) Simulating Spatio-temporal Dynamics of Urbanization with Multi-Agent systems - a Case Study of the Phoenix Metropolitan Region, USA. *Ecological Modelling*, 222(5), 1129-1138.