Int. J. Architect. Eng. Urban Plan, 26(2): 183-194, December 2016

DOI: 10.22068/ijaup.26.2.183

Research Paper

Analysis of urban growth pattern using logistic regression modeling, spatial autocorrelation and fractal analysis Case study: Ahvaz city

A. Khajeh Borj Sefidi[1], M. Ghalehnoee[2,*]

[1]PhD Candidate in Urban Planning, Urban Planning Group, Art, Architecture & Urban Planning Department, Najafabad Branch, Islamic Azad University, Isfahan, Iran
[2]PhD in Urban Design, Department of Urban Studies, Art University of Isfahan, Iran

Received: 24 September 2014, Revised: 20 August 2016, Accepted: 6 September 2016, Available online: 29 December 2016

Abstract

Transformation of land use-land cover change occurs due to the numbers and activities of people. Urban growth modeling has attracted authentic attention because it helps to comprehend the mechanisms of land use change and thus helps relevant policies made. This study applied logistic regression to model urban growth in the Ahvaz Metropolitan Area of Khuzestan province in IDRISI Selva software and to discover what are driving forces effective on the urban growth of Ahvaz city, and with what intensity? Historical land use and land cover data of Ahvaz were extracted from the 1991and 2006 Satellite images. The following two groups of factors were found to affect urban growth in different degrees as indicated by odd ratios: (1) Constraints Distance to the Bridge, Rural Areas, Planned town and Industry activities (all with odds ratios<1_or coefficient <0); and (2) Number of urban cells within a 5·5 cell window, Distance to the Hospitals, Main Road, High Road, Rail Line, River, CBD and Secondary centers, agriculture areas in distance more than 5km of Urban area and Vacant area (all with odds ratios>1_or coefficient >0). Relative operating characteristic (ROC) value of 0.906 indicates that the probability map is valid. It was concluded logistic regression modeling is suitable for Understanding and measuring of driving forces effect on urban growth. Second, unlike the Cellular Automata (CA) model, the logistic regression model is not temporally explicit; urban growth trend in Ahvaz isn't in the event of infill development strategy. Also, variables of sprawl based agents indicate more power than to compact base agents.

Keywords: *Urban growth, Logistic regression modeling, Spatial autocorrelation, Ahvaz.*

1. INTRODUCTION

The rapid physical and socio-economic restructuring of cities in developing countries, have increasingly been attracting the attention of researchers throughout the world. Land transformation especially Urban growth research in these places tend to explore the various means of land use change and the social, economic, and spatial variables influencing it.

Models are one of the tools that are used by decision-makers for studying the behavior and controlling land use changes and their trends for example urban growth. They are also important tools for exploring the interactions between land use dynamics and the driving factors of change [1]. Several techniques such as exploratory spatial data analysis, logistic and multiple regression analysis, cellular automata (CA), artificial neural networks (ANN) are employed in such urban land use change research.

* *Corresponding author: m.ghalehnoee@aui.ac.ir*
Tell: +989125330253; Fax: +982177240277

Urban Growth as a form of land use change is a complex system. This phenomenon characterized by varied attributes and is influenced by different factors in different regions. Therefore, it is beneficial to test if the urban growth transformation pattern observed affected of which one social, econometric and biophysical factors. A number of mathematical methods in the literature deal with urban growth. The most popular modeling consists; logistic regressions which attempt to examine and forecast urban growth using an econometric formulation [e.g. 2, 3, 4], neural-networks modeling by which the interaction between the different elements of an urban system is studied based on the way biological neural systems develop [e.g. 5, 6, 7], and gravity models which address the interaction between the elements of urban systems by using a similar formulation to the Newton's law of gravity [e.g. 8]. Also, Agent-Based Models (ABM) and Cellular Automata (CA) have become popular for representing the actions, behavior and interactions of individual agents in space and time [9]. In recent years, ABM and CA techniques have been particularly useful in modeling urban expansion [10, 11]. Several studies have endeavored to understand the spatial-temporal pattern of land cover change and its driving force [2, 12]. Simulation-based models such as Cellular Automata (CA) attempts to capture the spatial-temporal pattern of urban change by incorporating spatial interaction effect in to the model however, the poor explanatory capacity of simulation models has limited the detail understanding and interpretability of urban growth dynamics with its potential driving forces [13]. Moreover, most dynamic simulation models are not capable of incorporating adequate socio-economic variables [3].

Empirical models use statistical analysis to uncover the interaction between land cover change and explanatory variables and have much better interpretability than simulation models. For example, regression analysis can help to identify the driving factors of urban growth and quantify the contributions of individual variables and their level of significance [13]. Binominal (or binary) logistic regression is a form regression, which is used to model the relationship between a binary variable and one or more explanatory variable yielding dichotomous outcome [14]. Logistic regression is based on the concepts of binominal probability theory, which does not need normally distributed variable, and in general has no explicit requirement. As an empirical estimation method, logistical regression has been used in deforestation analysis [15, 16], agriculture area changes [12, 17]. In the context of urban growth modeling, logistic regression model was used to study the relationship between urban growth and biophysical driving forces [2, 3, 4, and 18]. The conversion of non-urban to urban land use is considered as state 1, while the no conversion is indicated as state 0 in the same period of time. A set of independent variables are selected to explain the probability of non-urban area to conversion to urban. The main purpose of urban growth modeling is to understand the dynamic processes responsible for the changing pattern of urban landscape, and therefore interpretability of models is the most important aspect the modeling process. The advantage of statistical models is their simplicity for construction and interpretation or their capacity to correlate spatial patterns of urban growth with driving forces mathematically. However, statistical models lack theoretical foundation as they do not attempt to simulate the processes that actually drive the change [19].

In this paper, urban growth using logistic regression is modeled, explained and analyzed. The logistic regression model was applied to study the urban growth in Ahvaz, Khuzestan. The modeling aims to discover the relationship between urban growth and social, econometric and biophysical factors and to predict the future urban pattern. A Markov Chain-CA model has been previously applied to simulate the urban growth of Ahvaz in same period of time [20]. This will allow comparison between these two approaches of modeling applied to Ahvaz Metropolitan area. The steps of the modeling are to: (1) conduct multi-resolution calibration of a series of logistic regression models and find the best resolution of modeling using fractal analysis; (2) refine the model by correcting and covariance analysis for spatial autocorrelation; (3) use the refined model to explain the driving forces of the urban growth; (4) validate the model by Relative operating characteristic (ROC) statistics and kappa index; and (5) analysis and predict the future urban pattern.

2. STUDY AREA

The Ahvaz, Khuzestan metropolitan region is defined here to include urban and Rural areas with a spatial extent of about 540 (Fig. 1). In the recent of the 20th century and present century, Ahvaz has introduced as the premier mineral, industrial and transportation urban area of the southwestern Iran and one of the fastest growing metropolitan areas in the Nation. Continuous high rate of population growth has been an explosive growth of the urban extent. This has resulted in extremely land cover and land use changes within the metropolitan region, wherein urbanization has consumed vast land of vacant and agricultural land adjacent to the city proper and has pushed the rural/urban fringe farther and farther away from the original Ahvaz urban core.

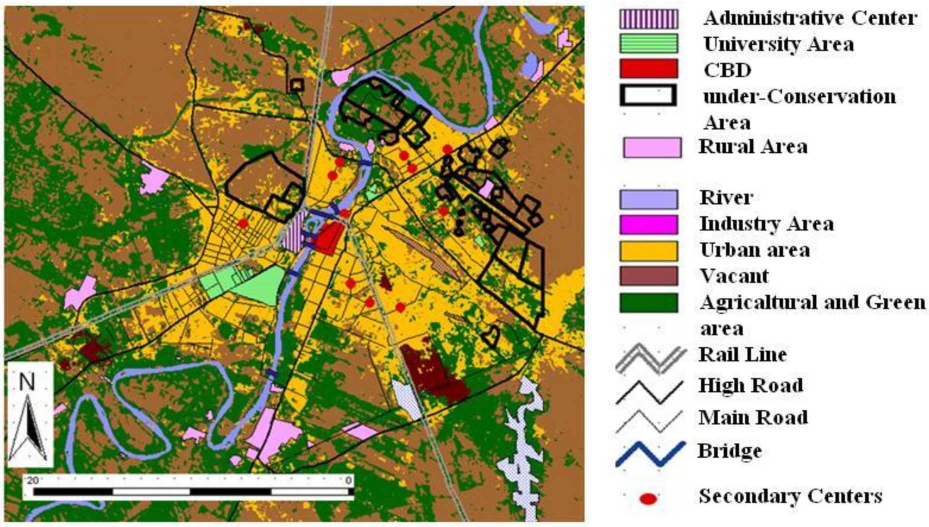

Fig. 1 Main Structure of Ahvaz City (Land use achieved from Aster Satellite -2006)

There has been an unbalanced and polarizing growth in Ahvaz: a dividing line exists between the west and the east of the Karoon River, strongly corresponding with the long-standing residential segregation patterns; for example existence of quite planned residential patterns in east and patterns of market-base residential development in west. This unbalanced growth has many dimensions which are shaping factors of the urban patterns: population, race, income, housing, employment and transportation patterns [20]. Explosive population growth is occurring in the south and outer suburbs of the region. Urban sprawl is also observed. Existent tendencies of urban development have occurred in the northwest (Kianshahr; Kianabad; Padadshahr; Golestan; Mellat, Aghajari district), and in the far southern suburban communities.

The massive urbanization and industry development since the two recent decades had adequate significantly to the spatial change of the urban structure of the Metropolitan Ahvaz. The emergence of these changes has created a cluster and sprawl structure. This state must be evaluated. So initial for compass urban growth to the sustainability is recognition and relative comparison socio-economic and spatial driving forces of urban growth. Modeling urban growth by logistic regression focused on spatial driving force.

3. METHOD AND MATERIAL

A logistic regression model was used to analysis the urban growth with socio-econometric, demographic, and environmental driving forces. In a raster GIS environment in IDRISI Selva software, the data layers are tessellated to form a grid of cells. Binomial logistic regression, in which the input dependent variable must be Boolean in concept, that is, it can have only two values inclusive 0 and 1. Such regression analysis is usually employed in estimating a model that describes the relationship between one or more continuous independent variables to the binary dependent variable. The basic assumption is that the probability of dependent variable takes the value of 1 follows the logistic curve and its value can be calculated with the following equation [21]:

$$P = (Y = 1|X) = \frac{exp \sum_{k=0}^{K} b_k x_{ik}}{1 + exp \sum_{k=0}^{K} b_k x_{ik}} \quad (1)$$

Where: P is the probability of the dependent variable being 1; X is the independent variables, X= ($X_0, X_1, X_2 \dots X_K), X_0 = 1$; B is the estimated parameters, $B = (b_0, b_1, b_2 \dots b_k)$. The state of a cell is dichotomous: either the presence of urban growth or absence of urban growth. It postulated value 1 is used to represent urban growth and value 0 for non-urban growth. Where $P(y = 1|x_k)$ is the probability of the dependent variable Y being 1 given ($x_1, x_2, x_3 \dots x_k$), i.e. the probability of a cell being urbanized; Xi is an independent variable representing a driving force of urbanization, which can be of continuous, ordinal or categorical nature; and b_i is the coefficient for variable x_i. Logistic regression employs Maximum Likelihood Estimation (MLE) procedure to find the best fitting set of coefficients. The maximum likelihood function used is the following [21]:

$$L = \prod_{i=1}^{N} \mu_i^{y_i} * (1 - \mu_i)^{(1-y_i)} \quad (2)$$

Where L is the likelihood, μ_i is the predicted value of the dependent variable for sample i and y_i is the observed value of the urban growth (dependent variable) for sample i. To maximize the aforesaid equation (2), it thus requires the solution for the following simultaneous nonlinear equations [21]:

$$\sum_{i=1}^{N} (y_i - \mu_i) * x_{ij} = 0 \quad (3)$$

Where x_{ij} is the observed value of the independent variable j for sample i. In solving the above equations, it had been used the Newton-Raephson algorithm. Logistic regression modeling, as an empirical estimation approach, allows a data-driven rather than a knowledge-based approach to the choice of predictor variables. Nevertheless, we still made an informed selection of variables. Selection of socio-economical predictor variables was detected by a historical review of urban growth in Ahvaz as reviewed in Section 2. These variables correspond to the important dimensions shaping Ahvaz urban patterns and main structures and function and spatial economic parameters (population, race, policies of big companies for example NIOC, main constraints, employment, and housing). The choice of econometric and biophysical variables imitate to most dynamic simulation modeling practices and local condition for example some variables extracted from the determining factors inclusive slope, land use, exclusion, urban extent, transportation, Hillshade factors as in SLEUTH model [22, 23, 24]. While, the spatial influences of major highways, economic activity centers, existing land use types, and socioeconomic driving force of development and effect of their policies on land development, Distance of bridges as one of agglomeration economies factors in Ahvaz, protect area of oil shaft are arisen local conditions. Correlations may exist between those demographic variables. Logistic regression calibration should check for multi collinear. Model calibration in this study had two stages including initial calibration and refining.

In this research, the land cover maps produced from Ahvaz metropolitan area for the years 1991, 2002 and 2006 satellite images are used [20], which show five categories of land use/cover: urban, industry, grassland, vacant, and water.

The 1991 census data were used for the social variables in model calibration. The 2006 census data were used for model prediction. The model should perform best if predictor data are collected at the year 1991, which lies halfway through the time period considered. A 2002 map was used to detection transition variables. A 2006 Land Cover Data map was used for validation. An interaction term number of urban cells within a neighborhood were calculated as an independent variable to take spatial interaction effects into account.

The complete list of variables is shown in Table 1 and Fig. 2 shows the raster maps of the independent variables. Tree design variables denoted as $X_1, X_{14,} X_{16}$ representing 3 land use/cover classes respectively were generated to distinguish among the 5 categories of land use-cover by recoding the 1991 land use-cover map into a binary map for each land use-cover category. If all the tree design variables take the value of zero, then a cell value in the ‘‘water’’ layer must be one; if any one of the five land use/cover classes takes the value of 1, a cell value in the ‘‘water’’ layer must be zero. Including a ‘‘water’’ variable in the model would be redundant and cause multi-collinear. Initial model calibration used only the first 17 variables.

In last model, $X_4, X_8, X_{13} \& X_{15}$ were eliminated into because of spatial autocorrelation that might exist. Fig. 3 shows the map of urban growth from 1991 to 2006, which serves as the dependent variable Y.

4. MODEL CALIBRATION

We used multi resolution and covariance method for calibration the best resolution for modeling. Moellering and Tobler (1972) argue that geographic processes operate at different scales. There are methods to find at what resolutions new patterns may emerge and when the performance of models takes a significant turn. These turning points should be those at which the resolutions approach dominant operational scales.

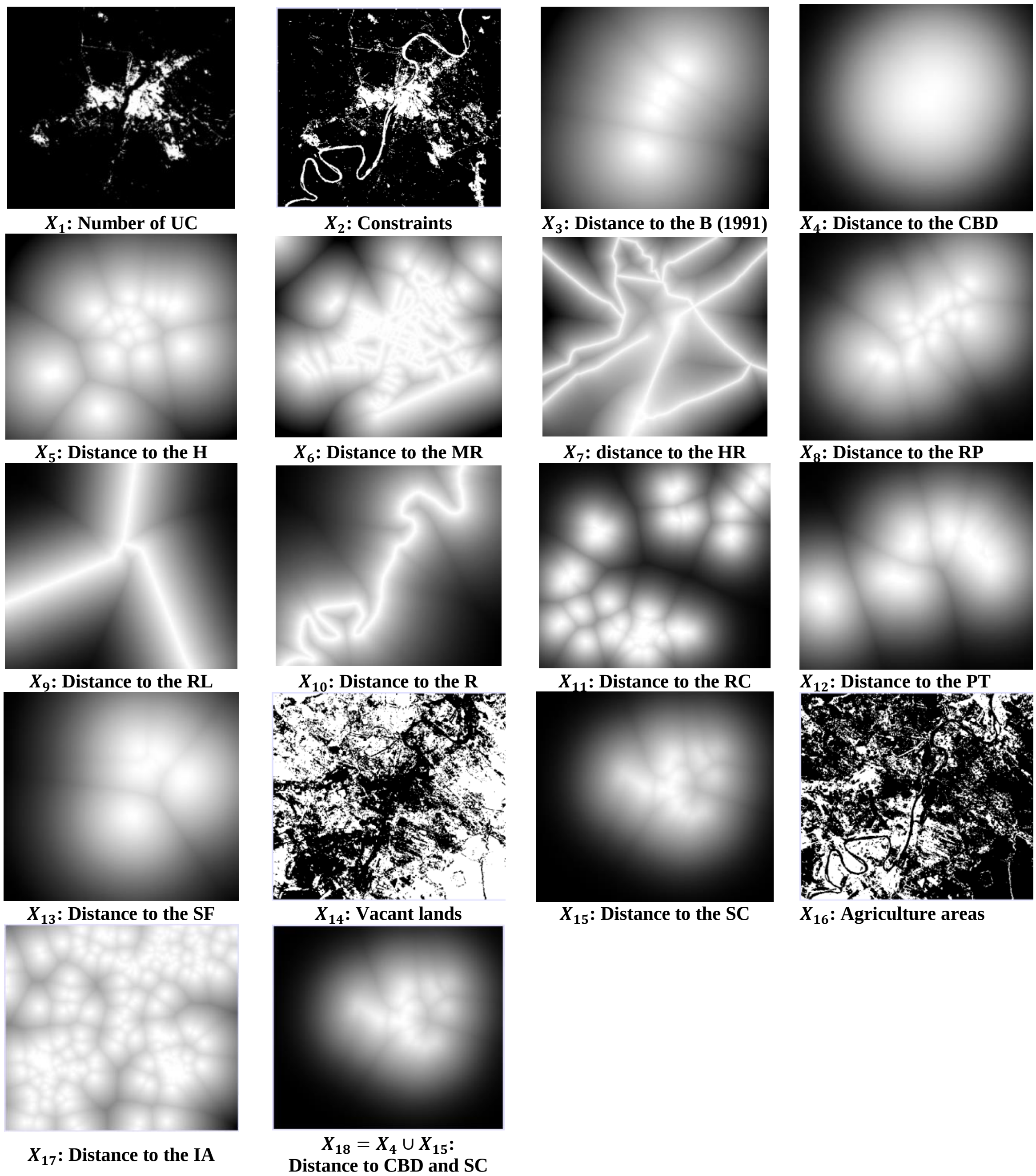

Fig. 2 Raster layers of independent variables. Black represents 0 and white represents 1

These resolutions are where modeling should be conducted [25, 26]. Previous studies in environmental modeling using simple linear regression emphasis that R-square values are higher at coarser scales [3, 27, and 28]. Because of this statistic does not mean what R-square means in ordinary least square (OLS) regression (the proportion of variance explained by the predictors), logistic regression does not have an equivalent to the R-square. So, this statistic is not suitable for recognizing the best resolution of modeling [3, 21].Goodness of fit values might not be used to determine the best resolution due to a lack of a turning point.

We used fractal analysis same Hu and Lo research (2007) to determine the optimal scale of modeling. Fractal dimensions were calculated for the probability surface maps predicted using the logistic regression model calibrated in different resolutions (from 50 m to 400 m). The triangular prism surface area [29, 30] and linear method [21] for calculation the fractal dimension exists that in this was used second method. Most common means of measuring fractional dimension is to measure the length of a section of that feature with a measuring instrument of varying precision. The fractal dimensions in this study were calculated using the software package IDRISI SELVA.

Fig. 4 shows the change of fractional dimension with the resolution of modeling. Fractal dimension increases almost linearly with the change of resolution from 50 m to 400 m, and then decreases at the turning point of 350 m. This suggests that the urbanization probability surface does not demonstrate the property of self-similarity of real fractals since self-similar objects must have constant fractional dimension. Previous studies have demonstrated that true fractals with self-similarity at all scales are unfrequented [31] and most real-world curves and textures are not genuine fractals with a constant fractal dimension. The change of fractal dimension across scale, defined by Mandelbrot (1983), can be expounded positively and used to summarize the scale changes of the spatial phenomena(3). The scale at which the highest fractal dimension is measured may be the scale at which most of the processes operate [31, 33, and 34] and the model performs best. To test if the model in deed performs best at the turning point of 350 m, a series of urban growth probability maps generated from the logistic regression was compared against an actual urban growth map and ROC values, which validate the model running, were calculated.

The highest ROC value was achieved at the resolution of 350 m. Thus the resolution of 350 m was selected as the optimal scale at which the logistic model best represents the dynamics of urbanization and the underlying processes.

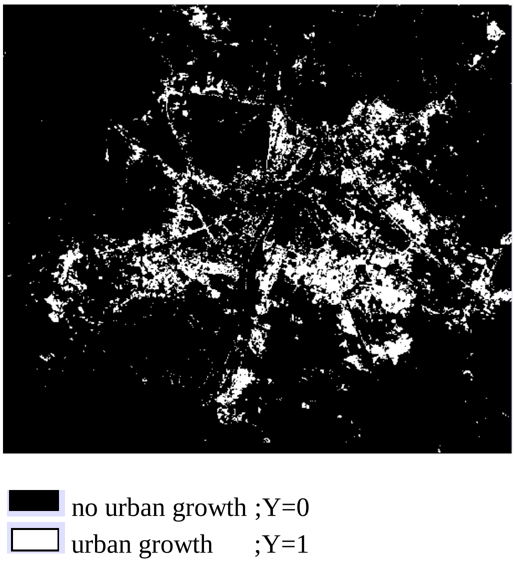

Fig. 3 Dependent variable Y_ Urban Growth 1991_2006

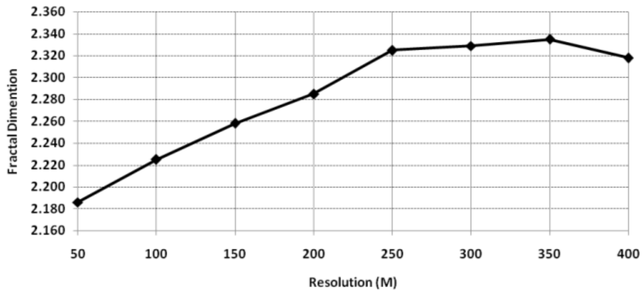

Fig. 4 Lumped fractal dimension of logistic regression model predicted probability surface plotted against resolution.

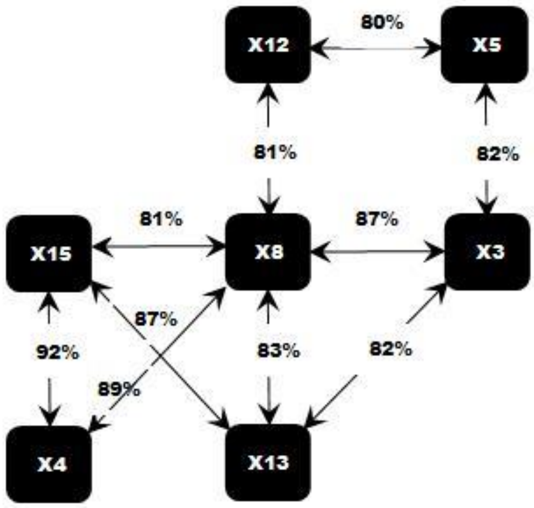

Fig. 5 Covariance analysis and significant correlation between variables

Table 1 List of variables included in the logistic regression model: These factors are extracted from three main sources: 1) Factors that were selected form the main sources of research as basis. 2) Factors from the main structure of Ahvaz city and 3) factors arising from major developments affecting the original structure from the main changes the original structure affecting

Variables	Meaning	Nature of Variable
Dependent		
Y	0: no urban growth; 1: urban growth	Dichotomous
Independent		
X_1	Number of urban cells within a 5 * 5 cell window	Design
X_2	Constraints: inclusive River area, Swamps, wetland, protected area of Oil Shaft, Built up 1991)	Continuous
X_3	Distance to the Bridges (B) in 1991 (km)	Continuous
X_4	Distance to the CBD (km)	Continuous
X_5	Distance to the Hospitals (H) (km)	Continuous
X_6	Distance to the Main Road (MR) (km)	Continuous
X_7	Distance to the High Road (HR) (km)	Continuous
X_8	Distance to the Regional Parks (RP) (km)	Continuous
X_9	Distance to the Rail line (RL) (km)	Continuous
X_{10}	Distance to the River (R) (km)	Continuous
X_{11}	Distance to the Rural Areas (RA) (km)	Continuous
X_{12}	Distance to the Planned Towns (PT) (km)	Continuous
X_{13}	Distance to the Sporty Facilities (SF) (km)	Continuous
X_{14}	Vacant land	Design
X_{15}	Distance to the Secondary Centers (SC) (km)	Continuous
X_{16}	Agriculture areas in distance more than 5km of Urban area	Design
X_{17}	Distance to the Industry Activities (IA)	Continuous
X_{18}	Distance to CBD and Secondary centers	Continuous

5. LOGISTIC REGRESSION MODEL

The logistic regression model assumes that observations are independent of each other and the residuals are mutually independent. But this assumption may be contravened due to the spatial autocorrelation. Spatial autocorrelation is the trend for data values to be similar to neighboring data values.

To test the logistic regression residual for spatial autocorrelation, Moran’s I for the King’s case was calculated under a normality assumption that the cell values display independent drawings from a single normally distributed population, hence a null hypothesis that there is no spatial autocorrelation. For area within the residual image at the resolution of 350 m, the value of Moran’s I is 0.768, indicating positive spatial autocorrelation. The Z test statistic value is 1891.23 with the p value of 0.0002. The p value is much less than 5%, which leads to the conclusion that the null hypothesis of no spatial autocorrelation in the residuals can be rejected. In other words, spatial autocorrelation is present among the residual values. Alternatively, there are two methods for reduce dependency: (1) pixel thinning that reducing the number of rows and columns while simultaneously decreasing the cell resolution, (2) sampling that leads to a smaller sample size that loses certain information base on maximum likelihood method. Nevertheless, it is a more sensible approach to remove spatial auto-correlation and a reasonable design of spatial sampling scheme will make a perfect balance between the two sides [3, 20, and 21]. In this study applied spatial sampling to correct for the spatial autocorrelation.

Systematic sampling and random sampling are two sampling comportments in logistic regression. Modeling via systematic sampling reduces spatial dependence, while important information like relatively isolated sites are lost when population sample does not spatially resemble. On the other hand, random sampling is efficient in representing population, but does not efficiently reduce spatial dependence, local spatial dependence in specific. Stratified random sampling is thought to perform well when it is necessary to make sure that small, but important, areas are represented in the sample [35]. So, a sampling scheme both of systematic and random sampling is employed to overcome the conflict of sample size and spatial dependence.

At the resolution of 350 m, the number of cells within the counties is 1,560,552, of which 15,659 cells have been sampled. Within the area, the number of cells that have changed from non-urban to urban, i.e., the number of 1s for variable Y (1 = urban growth), is 150,801 accounting for 9.66% of the total number of cells. Of the 15,659 sample points, there are 1,510 points whose cell values are 1 in the Y variable layer. The percentage of 1s in the sample is 9.64%, matching very well with the percentage of 9.66% for the full data set, which demonstrates the representativeness of the stratified random sampling.

A maximum likelihood estimator [14] was used to fit the model. The results of fitting the logistic regression model with the full 14 independent variables are given in Table 2. At the level, all of variables are significant. A probability map was derived using the refined model and a residual map calculated to evaluate the extent to which autocorrelation has been reduced. The value of Moran's I becomes 0.001 (p = 0.084), indicating very weak spatial autocorrelation. McFadden's pseudo R-square was used to test the goodness-of-fit of the model. Pseudo R square values between 0.2 and 0.4 are considered a good fit [29, 36].

The pseudo R2 value of 0.3648 indicates a good fit of the model .Statisticians suggest that we must be careful in our use of the Wald statistics to assess the significance of the coefficients and that whenever a categorically scaled independent variable is included from a model (both of continuous and design variables) [14]. Strict adherence to the level of significance would justify excluding the three land use-cover types from the model. However, the probability of urbanization of a land lot should be influenced by its initial land use-cover status and initial land use-cover should be considered important in land cover change dynamics in a biophysical and cultural sense. Thus the two Continuous variables contained Distance to the regional Parks and Distance to the Sporty Facilities were eliminated and variables Distance of CBD and Distance to the Secondary Centers were synthesized in the reduced model. The results of fitting the educed logistic regression model are shown in Table 2 and equation 4.

A logistic regression model is used to associate urban growth with driving force factors and to generate urban growth probability data. Herein, the urban growth equation with 14 driving force factors based on stepwise regression analysis of data in 1991 and 2002 was used as the initial equation for the logistic regression model to predict urban growth in 2006 as shown in the following equation.

$$\begin{aligned} UG &= Y_{1991-2006} \\ &= -13.8572 + 4.8344\,X_1 - 2.7554\,X_2 - 2.3803\,X_3 \\ &+ 1.9624\,X_5 + 4.4812\,X_6 + 3.2018\,X_7 \\ &+ 0.1968\,X_9 \\ &+ 0.3054\,X_{10} - 1.4347\,X_{11} - 0.5012\,X_{12} \\ &+ 6.1464\,X_{14} + 5.3135\,X_{16} - 1.8788 X_{17} + 2.1889 X_{18} \end{aligned} \tag{4}$$

This result indicates that the urban growth between 1991 and 2006 was driven spontaneously and naturally by socio economic development such as land use change. In addition, predicted urban growth in 2006 based on spatial data in 1991 and 2002 was an output from the LOGISTIC REG module of IDRISI software, having probability values between 0.00 and 0.92 (Fig. 7). These probability values for urban growth (0.00-0.92) were firstly used to extract the predictive urban and built-up area. After that we compared this predictive result with the actual urban and built-up area in 2006 using Relative Operating Characteristic (ROC) and Kappa Index of Agreement (KIA). These 2 accuracy assessment methods and the probability values for urban growth are calculated. The best threshold value coefficient of agreement in kappa analysis is 0.913. The KIA offers one comprehensive statistical analysis that answers simultaneously two important questions. How well do a pair of maps agree in terms of the quantity of cells in each category? How well do a pair of maps agree in terms of the location of cells in each category? This method of validation calculates various Kappa Indices of Agreement and related statistics to answer these questions (For details, see 21&39).

6. MODEL EXPLANATION

In a logistical regression, the odds of Y being 1 is calculated using the equation $\varphi = e^{\alpha+\sum_{i=1}^{k}\beta_i X_i}$ (38.54), and the odds ratio for the dichotomous variable Xi is demonstrated relation between each of the variables with urban growth (Y)[1].The parameter a can be explanted as the logarithm of the background odds that would result for the logistic model without any X's at all.

A land lot with more neighboring areas that are urban is more likely to be developed for urban use. The variable number of urban cells within a neighborhood of 5 *5 cell size (X1) has an odds ratio equal to 125.76. With an increase of 1 urban cell within the neighborhood, the odds of development will increase extremely. The use of a land lot is often influenced by the land use/cover status of the adjacent area. Land managers and real estate developers have some propensity of imitating the land use/cover behaviors in the neighborhood.

Secondary centers are developed in near to the CBD in Ahvaz metropolitan, so spatial correlation between these variables is positive (0.91). So these variables (X_4&X_{15}) were synthesis (X_{18}). The decentralized, polycentric urban growth trend in the metropolitan Ahvaz area is evidenced by the interpretation of the odds ratios for the two predictors: distance to the CBD and Secondary centers (X_{18}) and distance to High road (X_7). The odds of urban growth in an area 1 cell closer to from the X18 is estimated as 8.9254 as large as that in area further away this variable. The odds ratio for distance to High road is 24.5767, which means that the odds of urban growth in area close to CBD and SC is estimated 1.087629 times as large as that in area 1 cell further away from CBD and SC. The closer it is to High road, rather than to the CBD, the more probable a land lot will be developed for urban use.

Urban areas tend to grow close to the nearest Main road (X_6). Distance to the nearest Main road (X_6) has a coefficient of 3.2018. The odds ratio is equal to 24.57. The probability of urban development in an area is estimated at 0.0406 times as large as the probability of urban growth in an area 1 cell further away from the nearest urban area. This demonstrates that urban growth has been controlled by road accessibility.

The probability of transition in vacant land (X_{14}) is larger than the probability of transition in areas covered with Agriculture areas (X_{16}).This can be seen from the odds ratio values of 467.03 and 203.05 in a decreasing order for vacant land and Agriculture areas, respectively. All values are greater than one, illustrating a higher probability of urban growth [3].

Constraints (X_2) as Design variables and Distance to the bridge (X_3), rural c (X_{11}), planned towns (X_{12}) and industry activities (X_{17}) as continuous variables have negative coefficients (odds ratio < 1).

The probability of change from constraint lands (X_2) to urban use is smallest for urban land use (odds ratio values=0.00). Values are greater than 1, having a higher likelihood of urban growth in those areas. New urban growth has occurred mainly in spaces far from bridges (X_3), rural c (X_{11}), planned towns (X_{12}) and industry activities (X_{17}). It should be noted that urban growth in Ahvaz not tends to infill development. Also, variables of sprawl agents indicate more power than to compact agents.

7. MODEL VALIDATION USING ROC METHOD

Model variables were selected by entry testing based on the significance of the score statistic, and removal testing was based on the probability of the Wald statistic. Probability for entry and removal was respectively set to 0.05 and 0.10. Collinear was accounted for by eliminating the variable with the least significant Wald statistic contribution to the model [37]. The performance of the resulting regression models was evaluated by the relative operating characteristic (ROC) [15].

ROC was used to validate the logistic regression model. Recently the ROC method was brought to the field of land transformation change and urban growth modeling to measure the relationship between simulated change and real change [15, 38]. ROC method is an excellent method to evaluate the validity of a model that predicts the occurrence of an event by comparing a probability image depicting the probability of that event occurring and a binary image showing where that class actually exists. In this study, the ROC method offers a statistical analysis that solves one important question: ‘‘how well is urban growth concentrated at the locations of relatively high suitability for urban growth?’’ ROC evaluated how well the pair of maps agrees in terms of the location of cells being urbanized generally. Model validation using ROC reported a summary ROC value on base area under the curve as well as the peculiarities of the points on the curve that was used to calculate the ROC value. A ROC value of 0.5 is the agreement that would be expected due to random locations [20, 21]. Also, a ROC value of 1 indicates that there is a best spatial agreement between the actual urban growth map and the predicted probability map.

Table 2 Estimated coefficients and odds ratios for the logistic Regression model containing the 14 independent variables (M14)

Variables		Coefficient	Odds ratio	Standard Error
X_1	Number of UC	4.8344	125.7629	0.2155
X_2	Constraints	-26.4778	0.0000	0.3368
X_3	Distance to the B	-2.3803	0.0925	0.2491
X_5	Distance to the H	1.9624	7.1164	0.2468
X_6	Distance to the MR	4.4812	88.3405	0.2661
X_7	Distance to the HR	3.2018	24.5767	0.2254
X_9	Distance to the RL	0.1968	1.2175	0.2889
X_{10}	Distance to the R	0.3054	1.3572	0.2899
X_{11}	Distance to the RC	-1.4347	0.2382	0.2888
X_{12}	Distance to the PT	-0.5128	0.5988	0.2893
X_{14}	Vacant lands	6.1464	467.0322	0.4956
X_{16}	Agriculture areas	5.3135	203.0594	0.4660
X_{17}	Distance to the IA	-1.8788	0.1528	0.1376
X_{18}	Distance to CBD and SC	2.1889	8.9254	0.2929

To conduct model validation, the image map of urban growth probability predicted from the logistic regression model was decency against that of actual urban growth obtained by comparison of the actual urban growth map with the predicated urban growth map in 2006 year. First the ranked image of probability of urbanization was sliced at a series of threshold levels. A threshold accounts to the percentage of cells in the predicated image to be re-classed as 1 in preparation for comparison with the actual image. ROC began with the cell ranked the highest for probability, reclassified it as 1 and Calculations in equal parts (5%-10%, …100%) continues until of the cells had been reclassified as 1.. at each stages, sliced image was then compared with the 2006 actual image. Then ROC continued for the successive threshold.

In Table 3, A represents the number of true positive cells which are predicted as urban growth and are actually urban growth in the actual image. From each contingency table for each threshold, one data point (x, y) was generated where x is the rate of false positives (false positive %) and y is the rate of true positives (true positive %):

$$true\ positive\ \% = \frac{A}{A+C} = 97.52\% \tag{5}$$

$$false\ positive\ \% = \frac{B}{B+D} = 4.82\% \tag{6}$$

Table 3 Contingency table showing the comparison of the slice image of predicted urban growth probability with the reference image

		Reference image	
		Urban growth (1)	Urban growth (0)
Simulated Image	Urban growth (1)	A (true positive)	B (false positive)
	Urban growth (0)	C (false negative)	D (true negative)

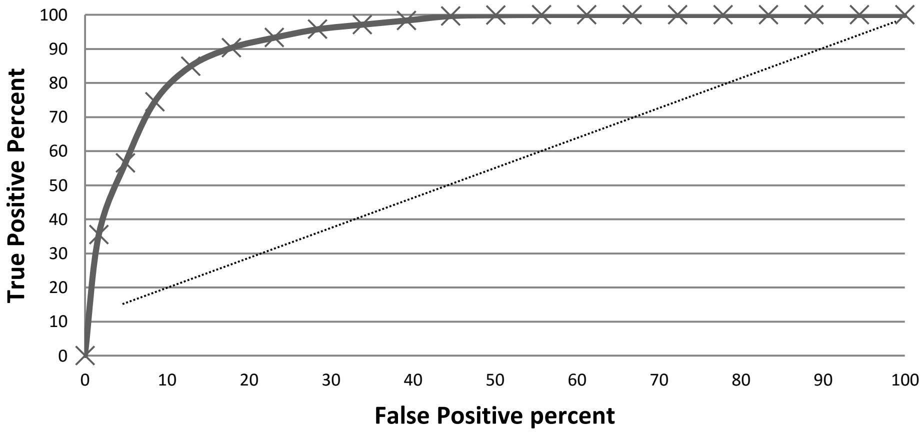

Fig. 6 Relative Operating Characteristic (ROC) curve

These data points were connected to create a ROC curve from which the ROC value was calculated. The ROC statistic is the area under the curve that connects the plotted points. The ROC statistic is the area under the curve that connects the plotted points that uses the trapezoidal rule from integral calculus to compute the area, where X_i is the rate of false positives for threshold i, Y_i is the rate of true positives for threshold i, and n+1 is the number of thresholds. The ROC curve (Area under Curve) is shown in Fig. 6. The ROC value in this study is 0.906 that calculated as:

$$(AUC) = \sum_{i=1}^{n} [x_{i+1} - x_i] \times [y_i + (y_{i+1} - y_i)/2] \tag{7}$$

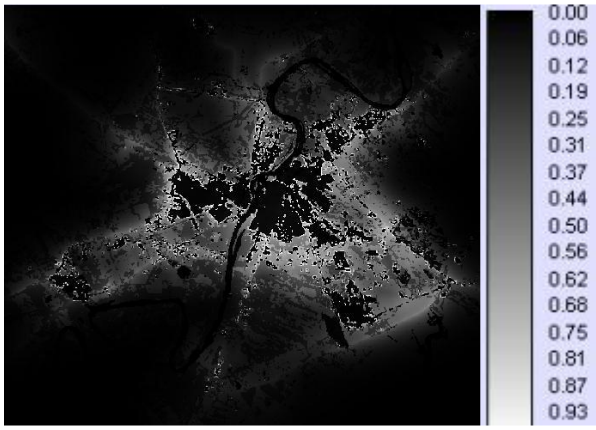

Fig. 7 Urbanization probability maps of Ahvaz, Khuzestan. Lighter tones indicate higher probabilities of urban growth

8. PREDICTION OF SPATIAL PATTERNS OF URBAN GROWTH

The probability map can be used for recognizing probability of future urban growth location and quantity. In the study area, urban area accounted for 53% in 2006 toward 1991 urban area, while same ration in relation with population growth is 37%. To produce the spatial pattern of urban distribution given a certain amount of urban area, the increase of the number of urban cells compared to the 1991 base urban map was calculated. Then the number of urbanized cells was represented base on 1 equation to the probability map (Fig. 7). This generated a growth map. The residuals are not standardized and are bounded between -1 and 1. The residuals are calculated as (Fig. 8):

$$Residual\ [i] = bserved\ (Y_i) - Predicated\ probability\ (i) \tag{8}$$

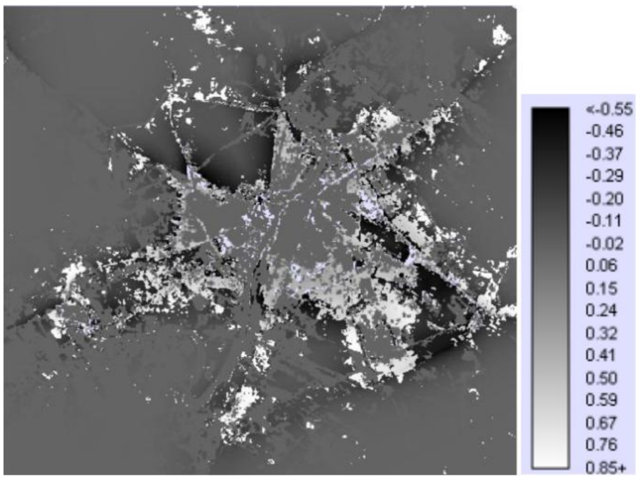

Fig. 8 Residual of regression

9. DISCUSSION AND CONCLUSION

Examine the relationship between urban growth and growth variables showed that:

Urban growth in Ahvaz has been affected severely by accessing to the main roads, the city center and sub-centers. Urban growth during the period under review occurred at distances far from the river, so there is weak relation between the growth of city and factor of distance to the bridges. River has always been one of the most important elements of the original structure. Indeed, the access to the east and west of city is declined, so urban growth has become more independent than to the River gradually. Therefore, as a strategy recommended that linking between the northern and southern parts of the city through the construction of bridge would be increased. Usage of this strategy will affect significantly on the agglomeration economies and will reduce infrastructure costs. Also access to the CBD and sub-centers through which it currently is on the wane, are improved and therefore urban growth will support endogenously.

Urban growth over any variable is affected of built up areas. Accordingly, we can say that urban management can play an important role in orientation of urban growth via controlling and prioritization the construction. Examine the relationship between urban growth and agricultural and vacant land show that the results and reflections of agricultural land protection laws (203 odds ratio) have protected with only 43% of power equivalent compared to the terms of Normal condition (Vacant land, with the 467 odds ratio).

Effect of spatial distribution of hospitals on urban growth is roughly effect equivalent of the town center and sub-centers, therefore it should be paying more attention the choosing the location of hospitals.

In modeling level, logistic regression modeling was used to identify and improve our understanding of the demographic, econometric and biophysical forces that have driven the urban growth and to find the most probable of locations of urban growth in Ahvaz. The following two groups of factors were found to affect urban growth in different degrees as indicated by odd ratios:

(1) Constraints (X_2) , Distance to the bridge (X_3), rural c (X_{11}), planned towns (X_{12}) and industry activities (X_{17}) (all with odds ratios < 1); and (2) distance to the CBD and SD, high road, main road, rail line, hospital, river, number of urban cells within a 5·5 cell window, vacant land, agriculture land, forest, and UTM northing coordinate (all with odds ratios > 1).

Urban growth will mainly be around existing urban areas and close to major roads, while some new clusters located at far from the existing urban areas. Despite the, this study has shown logistic regression model's strengths and limitations of the model. First, Logistic regression modeling is suitable for Understanding and measuring of driving forces effect on urban growth. Second, unlike the CA model, the logistic regression model is not temporally explicit. Its output probability map can only indicate where urban development will occur, but not when this will take place. Therefore, it is suggested in urban growth modeling, CA model and logistic regression method to be applied together.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

NOTE

1. odds ratio is calculated based on the equation $\frac{\varphi_1}{\varphi_0} = e^{\beta_i}$

REFERENCES

[1] Braimoh AK, Onishi T. Geostatistical techniques for incorporating spatial correlation into land use change models, International Journal of Applied Earth Observation and Geoinformation, 2007, Vol. 9, pp. 438-446.

[2] Allen J, Lu K. Modeling and prediction of future urban growth in the Charleston Region of South Carolina: a GIS-based integrated approach, Conservation Ecology, 2003, Vol. 8, No. 2.

[3] Hu Z, Lo CP. Modeling urban growth in Atlanta using logistic regression, Computers, Environment and Urban Systems, 2007, Vol. 31, pp. 667-688.

[4] Landis J, Zhang M. The second generation of the California urban futures models, Environment and Planning B, 1998, Vol. 25, pp. 795-824.

[5] Pijanowski B, Brown D, Shellito B, Manik G. Using neural networks and GIS to forecast land use changes: a land transformation model, Computers, Environment and Urban Systems, 2002, Vol. 26, pp. 553-575.

[6] Ou X, Zhang Y, Ren J, Yao D. A neural network model for urban traffic volumes compression, Journal of Systemic Cybernetics and Informatics, 2003, Vol. 1, No. 4, pp. 1-7.

[7] Maithani S, Jain RK, Arora MK. An artificial neural network based approach for modeling urban spatial growth, ITPI Journal, 2007, Vol. 4, No. 2, pp. 43-51.

[8] Tsekeris T, Stathopoulos A. Gravity models for dynamic transport planning: development and implementation in urban network, Journal of Transport Geography, 2006, Vol. 14, pp. 152-160.

[9] Batty M. Urban modeling. In R. Kitchin, & N. Thrift (Eds.), International Encyclopedia of Human Geography, Oxford: Elsevier, 2009, Vol. 12, pp. 51-58.

[10] He C, Okada N, Zhang Q, Shi P, Zhang J. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China, Applied Geography, 2006, 323-345.

[11] Zhang H, Zeng Y, Bian L, Yu X. Modelling urban expansion using agent based model in the city of Changsha, Journal of Geographical Systems, 2010, Vol. 20, No. 4, pp. 540-556.

[12] Serneels S, Lambin EF. Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model, Agriculture, Ecosystems and Environment, 2001, Vol. 85, p. 65-81.

[13] Luo J, Kanala NK. Modeling urban growth with geographically weighted multinomial logistic regression, Proceedings of SPIE, 2008, Vol. 71, No. 44.

[14] Hosmer DW, Lemeshow S. Applied logistic regression, New York, Wiley, 1989.

[15] Schneider L, Pontius RG. Modeling land use change in the Ipswich watershed, Massachusetts, USA, Agriculture, Ecosystems and Environment, 2001, Vol. 85, pp. 83-94.

[16] Geoghegan J, Villar SV, Klepeis P, Mendoza PM, Ogneva-Himmelberger Y, Chowdhury RR, Turner II BL, Vance C. Modeling tropical deforestation in the southern Yucatán peninsular region: comparing survey and satellite data, Agric, Ecosyst, Environ, 2001, Vol. 85, pp. 25-46.
[17] Walsh SJ, Crawford TW, Welsh WF, Crews-Meyer KA. A multiscale analysis of LULC and NDVI variation in nang rong district, Northeast Thailand, Agriculture Ecosystems and Environment, 2001, Vol. 85, Nos. 1-3, pp. 47-64.
[18] Wu F, Yeh AG. Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally Planned Economy to a Socialist Market Economy: a case study of Guangzhou, Urban Studies, 1997, Vol. 34, No. 11, pp. 1851-1879.
[19] Triantakonstantis D, Mountrakis G. Urban growth prediction: a review of computational models and human perceptions, Journal of Geographic Information System, 2012, Vol. 4, pp. 555-587.
[20] Khajeh Borjsefidi A. Simulation and analysis of urban growth using cellular automata, Case study: Ahvaz metropolitan area, Msc thesis, Art and Architecture department, Shiraz University, Shiraz, Iran, 2009.
[21] Eastman JR. IDRISI Kilimanjaro, Guide to GIS and Image Processing, Clark Labs, Clark University, Worcester, USA, 2003, 360 p.
[22] Clarke KC, Hoppen S, Gaydos L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area, Environment and Planning B: Planning and Design, 1997, Vol. 24, pp. 247-261.
[23] Dietzel C, Clarke K. The effect of disaggregating land use categories in cellular automata during model calibration and forecasting, Computers, Environment and Urban Systems, 2006, Vol. 30, pp. 78-101.
[24] Yang X, Lo CP. Modeling urban growth and landscape changes in the Atlanta metropolitan area, International Journal of Geographical Information Science, 2003, Vol. 17, No. 5, pp. 463-488.
[25] Moellering H, Tobler W. Geographic variances, Geographical Analysis, 1972, Vol. 4, pp. 34-50.
[26] Meentemeyer V. Geographical perspectives of space, time, and scale, Landscape Ecology, Vol. 3, pp. 163-173.
[27] Bian L, Walsh SJ. Scale dependencies of vegetation and topography in a mountainous environment of Montana, Professional Geographers, 1993, Vol. 45, pp. 1–11.
[28] Kok K, Veldcamp A. Evaluating impact of spatial scales on land use pattern analysis in Central America, Agriculture, Ecosystems and Environment, 2001, Vol. 85, pp. 205-221.
[29] Clarke KC. Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method, Computers and Geosciences, 1986, Vol. 12, No. 5, pp. 713-722.
[30] Jaggie S, Quattrochi A, Lam S. Implementation and operation of three fractal measurement algorithms for analysis of remote-sensing data, Computers and Geosciences, 1993, Vol. 19, No. 6, pp. 745-767.
[31] Lam NS, Quattrochi DA. On the issues of scale, resolution, and fractal analysis in the mapping sciences, Professional Geographers, 1992, Vol. 44, pp. 88-98.
[32] Mandelbrot BB. Fractal geometry of nature, San Francisco, Freeman, 1983.
[33] Goodchild MF, Mark DM. The fractal nature of geographical phenomena, Annals of the Association of American Geographers, 1987, Vol. 77, No. 2, pp. 265-278.
[34] Cao C, Lam NS. Understanding the scale and resolution effects in remote sensing and GIS, In D. A, 1996.
[35] Congalton RG. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, Photogrammetric Engineering and Remote Sensing, 1988, Vol. 54, No. 5, pp. 593-600.
[36] Domencich TA, McFadden D. Urban travel demand: Behavioral analysis, Amsterdam: North-Holland, 1975.
[37] Wassenaar T, Gerber P, Verburg PH, Rosales M, Ibrahim M, Steinfeld H. Projecting land use changes in the Neotropics: the geography of pasture expansion in forest, Global Environmental Change, 2006, Vol. 17, pp. 86-104.
[38] Pontius RG. Quantification error versus location error in the comparison of categorical maps, Photogrammetric Engineering and Remote Sensing, 2000, Vol. 66, No. 8, pp. 1011–1016.

AUTHOR (S) BIOSKETCHES

Khajeh Borj Sefidi, A., *PhD Candidate in Urban Planning, Urban Planning Group, Art, Architecture & Urban Planning Department, Najafabad Branch, Islamic Azad University, Isfahan, Iran*
Email*: armankbs@gmail.com*

Ghalehnoee, M., *PhD in Urban Design, Department of Urban Studies, Art University of Isfahan, Iran*
Email: *m.ghalehnoee@aui.ac.ir*

COPYRIGHTS

HOW TO CITE THIS ARTICLE

Khajeh Borj Sefidi, A., Ghalehnoee, M., (2016). Analysis of urban growth pattern using logistic regression modeling, spatial autocorrelation and fractal analysis, case study: Ahvaz city. Int. J. Architect. Eng. Urban Plan, 26(2): 183-194, December 2016.

URL: http://ijaup.iust.ac.ir/article-1-207-en.html