Article

Dynamic simulation of urban expansion based on cellular automata and logistic regression model: Case study of Hyrcanian region of Iran

Meisam Jafari [1], Hamid Majedi [2,*], Seyed Masoud Monavari [3], Ali Asghar Alesheikh [4],

Mirmasoud Kheirkhah Zarkesh [5]

1. Department of Environmental Science, Science and Research Branch, Islamic Azad University, Tehran, Iran; jafari.m@pci.iaun.ac.ir

2. Department of Art and Architecture, Science and Research Branch, Islamic Azad University, Tehran, Iran

3. Department of Land use Planning and Assessment, Science and Research Branch, Islamic Azad University, Tehran, Iran; monavarism@yahoo.com

4. Department of GIS, Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran, Iran; alesheikh@kntu.ac.ir

5. Department of Remote Sensing and GIS, Science and Research Branch, Islamic Azad University, Tehran, Iran; kheirkhah@alumni.itc.nl

* Correspondence: majedi.ha@yahoo.com

Abstract: A model that combines logistic regression model, Markov chain, together with cellular automata based modeling is introduced here to simulate future urban growth and development in Gilan Province, Iran. The model calibrated based on data, beginning 1989 and ending 2013, and is applied in making predictions for the years 2025 and 2037, through 12 urban development criteria. The relative operating characteristic (ROC) is validated with a very high rate of urban development. The analyze results indicate that the area of urban land is increased by more than 1.7 % that is, from 36012.5 ha in 1989 to 59754.8 ha in 2013, and the area of Caspian Hyrcanian forestland is reduced by 31628 ha. The simulation results with respect to prediction indicate an alarming increase in the rate of urban development in the province by the years 2025 and 2037 that is, 0.82 % and 1.3 %, respectively. The development pattern is expected to be uneven and scattered, without following any particular direction. The development will occur close to the existing or newly-formed urban infrastructure and around major roads and commercial areas. This

development trend, if not controlled, will lead to the loss of 25101ha of Hyrcanian forests and if the continued, 21774 ha of barren/open lands are expected to be destroyed by the year 2037. These, results demonstrate the capacity of the integrated model in establishing comparisons with urban plans, and their utility to explain both the volume and constraints of urban growth. It is very beneficial to apply the integrated approach in, urban dynamic assessment through land use modeling with respect to spatio-temporal representation in, distinct urban development formats.

Keywords: Land use change; urban sprawl; Logistic regression; Markov chain; Cellular automata; Gilan Province

1. Introduction

Globalization could not have occurred if the world community had not advanced to the point where it is. This fact is accompanied with population growth in developing countries in specific which leads to land use expansion. Urban sprawl is a term used in describing low-density contemporary settlement patterns with slight comprehensive public planning that ends with emerging economically inefficient, socially inequitable, environmentally damaging and non-aesthetic cities [1]. Increasing awareness on the negative aspects of urban sprawl is a major topic of debate [2]. Urban changes should be managed in a manner where the environmental and social sustainability for ever-growing urban population of the world is ensured [3] which by 2050 is expected to account for 60% of the world population according 10 UN predictions [4].

According to the importance of the issue, knowledge-based management of urban growth is focus many of the relevant literature on urban planning [5]. Since 1960s, this modeling is used an manner [6] in describing the previous, current, and future patterns in urban development [7]. The logistic regression-based Cellular Automata (CA) model was first introduced by Wu [8]. This combined model, which was applied by [9-11]. By comparison, the logistic regression-based is one of the most frequently applied in the past two decades in land use modeling prediction many inductive modeling are applied [12–14]; but it should be noted that, this model lacks quantification of change and temporal analysis [13-15]. On the contrary, the MC models can predict the volume of land use change and identify the structural utility, since MCs only compute the probabilities of land use transitions and the amount of change are spatially non-explicit [18]. The

spatial CA models overcome this limitation of MC based on predefined site-specific rules which mimic land use transitions [19], while the CA models are not able to determine the actual amount of change. Since every modeling technique has its unique limitations [11], proposed a hybrid approach based on logistic regression and CA transition rules, which resulted in an improved model quality; nevertheless, this model is not able to quantify the amount of land use change. Puertas and Meza [17] came up with some promising results after integrating the logistic regression, MC CA models in simulating urban expansion; but hybrid model failed to consider a spatial auto-regression in determining the driving forces of LUCCs was missing in this. CA-based models are simple and allow for dynamic spatial simulation [20,21]. CA, when combined with logistic regression models, as done by Wu [8] and Puertas and Meza [17], can predict a complex urban reality, thus the potential in being applied in this study. This hybrid framework can form the basic structure of this presented model; but, when such models are applied in the context of Gilan Province, several due considerations will have to be made. The first important consideration is to adopt the CA framework which would suit the realities in the city in a sense that the outputs of the model would provide information at an appropriate scale for urban policy making. The second consideration is to insert the designated variables in to the logistic regression models. Given the availability of data, it is important to identify the methods that can be adopted in quantifying the inputs fed into the logistic regression model. The third consideration is to build a link with the framework, that is, the CA-based framework, and to implement the model by using inputs from the logistic regression model and urban policies in a manner that the model would able to establish forward and backward linkages with urban development and control policies. This consideration should ultimately be reflected in the simulation of urban development and growth. Therefore, this study furthers the work conducted on modeling the urban development and growth of Gilan Province Jat et al. [22], which are mostly examples of modeling urban sprawl and are not able to model the development of different land uses. How could some recent applications, introduced in the empirical literature on urban growth modeling, be adopted in this study?

Due to its geo-logistic stance, Gilan has experienced a significant growth with respect to urban land use area in the recent decades. Annual rural influx expands urban population and promotes the urbanization culture. Unsustainable growth of the city has caused many socioeconomic and

environmental problems like reshaping the landscape, emissions, heavy traffic jam, deforestation, and land use change. All in all, the pleasant climate, picturesque landscape of Hyrcanian forests, the Caspian Sea have made this part of country as one of the most densely populated parts. It is usually the first attractive choice of tourism for domestic and Middle Eastern tourists with low accommodation-cost. This fact can pave the way to forecast sustainable growth in future development by applying the necessary administrative-planned measures and avoid uncontrolled construction and shapeless growth. Accordingly, this study is run with the objective of presenting a dynamic hybrid model based on LR, MC, and CA to forecast the auto-spreading orientation of by the years 2025 and 2037. The outputs here can equip the decision makers and relevant authorities with future orientations of the restricted urban growth. This article is structure as follows:

The material and method with all their relevant subdivisions are presented in section 2; the result and discussion with all their relevant subdivisions are presented in section 3; the conclusion is presented in section 4.

2. Materials and methods

2.1. Study area

Gilan is a tourist province in the north of Iran located over an area of 14042 km^2 within the latitudes of 36^0 34' - 38^0 27' N and the longitudes of 48^0 53 '- 50^0 34' E covering Eastern third of the Caspian Sea coast with MSL of -27 m on the plain to 3897 m on the mountain top, (Figure 1). This area consists of two mountainous and plain features. The major urban area covers the plain with 32 % of the total area with a gradient less than 5 %.

As for the climate, annual precipitation is 1506 mm, 38% of which is confined to the autumn season, October in specific. This region has predominant naturally resources, like fertile and productive soil types and rich Hyrcanian forests and agricultural lands, covered with 52 cities, 13 of which are in coastal line, and 2916 villages [23]. Domestic immigration and tourism are the two important factors in significant population growth recorded in the recent decades, that is, any increase from 2081037 1986 to 2480874 in 2011 [23]. This is followed by urban sprawl and extensive land degradation in Gilan. One of the most striking aspects of land degradation here is the loss of great parts of
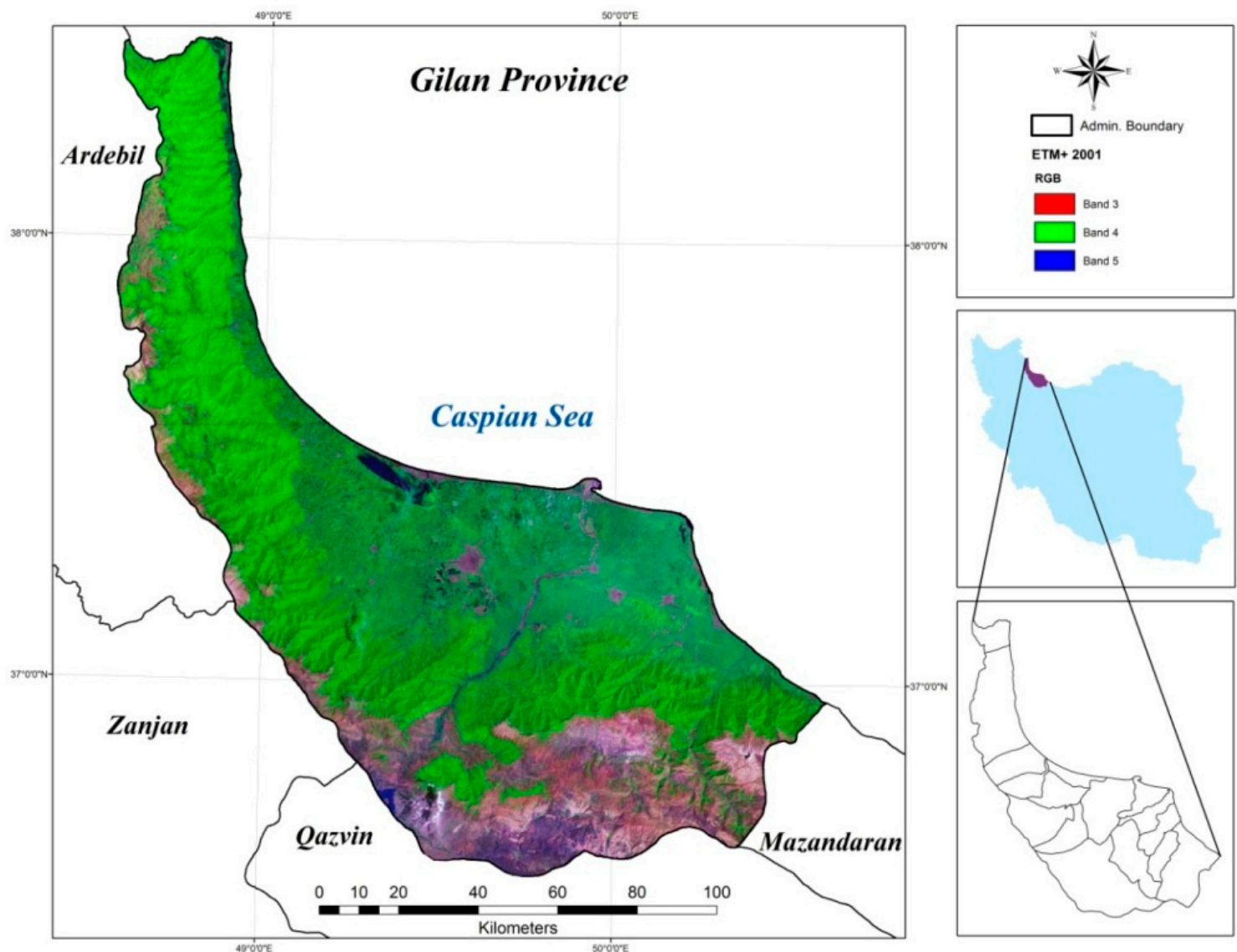
Hyrcanian forests and rangeland areas.

Figure 1. Situation of Gilan Province in Iran

2.2. Research procedure

The urban growth dynamics of the study area are found where techniques are integrated CA, MC and logistic regression modeling. This model is designed through the data from multiple time scales, obtained from remote sensing and census data. Calibration of the model is based on LULC data compiled from 1989 to 2013, while simulations are conducted for the years 2025 and 2037. The data and their implementation processes are presented in Figure 2.

2.3. Land Use/Land Cover (LULC): Change Detection

To begin with, the changes in different land use are detected in order to quantify variations in the city boundaries over the years 1989 - 2013. The land use maps are extracted from Landsat TM, ETM+, and OLI satellite images within two 12-year time intervals of 1989, 2001 and 2013. The coordinate system of the images is set as Universal Transverse Mercator (UTM), at Zone 39, and North WGS 84. The images are geo-referenced by Nearest Neighborhood algorithm, where, using a total of 220 training points, are well-distributed throughout the study area. The geo-referenced images are corrected radio-metrically and geometrically. In order to enhance the contrast of the features and facilitate classification of different land uses, Optimum Index Factor (OIF) is used in finding the best color composite. After pre-processing the images, they are classified through the Maximum Likelihood supervised classification method. Accordingly, the land use maps of the study area are prepared in five land use classes (Table 1), including: built-up areas, agricultural lands, water bodies, forest coverage, and open land (rangeland and bare land). At the end, the changes in different land use types were detected by cross tabulation matrix in IDRISI Selva software.

2.4. The integrated land use change model

The integrated model of logistic regression, Markov chain, and cellular automata begins with the logistic regression of the two dataset of urban and non-urban explanatory factors, which determine the potential for change from one class of land use to another, in each cell. Next, using MC and reference maps of land use, the third step, is estimated, that is, the proportion of the area that will likely change from one land use to another. Subsequently, the location of this estimated area of conversion is calculated through the CA. This allows the researchers to assign, the land use class for a given year of prediction to each pixel. In order to calibrate and validate the adjustment of the model a reference map from 2013 is compared to this estimated area map of land use/cover for the

same year. Finally, the calibrated model was used to simulate future land covers for the target years 2025 and 2037.

2.5. Simulation of future land use change by LR

In this study, the future land use maps of the target years 2025 and 2037 are simulated once again logistic regression. This is done to achieve more accurate estimate of future land use changes. According to relevant literature [14, 21] and expert knowledge, a total of 12 independent variables are recognized as the factors which determine the urban sprawl, the: distance to and from airport, distance to and from central business district (CBD), distance from major roads, distance to the nearest cities, distance to the seashore, Digital Elevation Model (DEM), distance to the rivers, slope (in percent), forest areas, northing and easting coordination's and agricultural lands. These were separately introduced to the model as the input data as to predict any changes in the extent of the built-up areas over the periods of 12 years: 1989 – 2001 and 2001 – 2013. The dependent variable in this model is the developed urban areas cells (i.e. conversion of other land use types into the built-up areas) as a binary raster map layer, where, the cells with 0 value represent areas without land use change while the cells with the value of 1 indicate that the areas underwent the land use change over a given time period (e.g. 1989 – 2001). It is worth mentioning that the cell size of all the map layers is 30 m and the analysis is run in IDRISI Selva software. The logistic regression was estimated for the 12 variables in a manner to making achieving the maximum relative operating characteristic (ROC) possible and adjust the odd ratio values. These values represent the dependent and independent nature of the variables between two binary data as degree of significance level. Regression is one of the well-known statistical approaches applied when the dependent variable is dichotomous [25]. This approach model can reveal the correlation between a binary dependent and several independent categorical and continuous variables [26]. Systematic sampling and stratified random sampling are the two approved approaches applied in logistic regression model; the first one is applied to reduce spatial dependency when homogeneous nature of the population is not the second is applicable for running regression [16]. Here, the second method is selected and to conduct regression [27]. As a basic assumption in LR, where, the probability of a dependent

variable taking the value of 1 is based on the logistic curve, the probability value is calculated through the following equation:

$$P = E(Y) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + c_p x_p)} \tag{1}$$

where **p** is the probability of forest loss in the cell, **E(Y)** is the expected value of the binary dependent variable $\boldsymbol{\beta_0}$ is a constant value and $\boldsymbol{\beta_1}$ is the coefficient of the independent variable X_i

The logistic function of Eq. (1) is transformed into a linear response as follows:

$$P' = \log\left(\frac{P}{1-P}\right) \tag{2}$$

Furthermore, the binary dependent variable value is either 1 or 0 and the standard linear regression model is expressed as:

$$P' = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \beta_3 X_3 \tag{3}$$

where **P′** is the probability of a grid cell for the occurrence of land use type i. **X** is the related driving factors; $\boldsymbol{\beta}$ represents the regression coefficients between land use types and driving factors.

2.6. Simulation of future land use change through MC

In this study, the Markov chain model is applied to simulate the spatio-temporal changes of land uses indifferent time future series. A finite set of states as discrete values are after specified in Markov Chain. The possibility in making a transition in the states of this specified value in the system is recorded in the form of transition matrix. In simple terms, the transition matrix records the probability of moving from one state (land use class) to another (land use class) [28]. In this study the probability of changing each land use class to another is determined by Markov transition probability matrices for the periods between 1989 and 2013. In order to verify the CA-Markov model for the target years of 2025 and 2037, it is assumed that the target year is 2013; accordingly, based on the transition probability matrix of 1989 - 2001, a land use map is simulated for the year 2013. The model is validated through the Kappa coefficient by comparing the simulated land use map of 2013 with the actual land use map, derived from the satellite image of 2013. MC

model is a discrete random process with the cellular lattice where the transition probability matrix based on neighborhood effects is applied in a spatial influence logarithm [29]. MC models is lately being applied in a big scale in modeling LULC changes in future predictions including both urban and non-urban areas [27,28]. this model is key-descriptive and is implemented in a convenient manner [17]. This method is developed into an essential prediction approach in geographic research and spatio-temporal simulations, due to its features like the of descriptive power, simple trend projection of land use and land cover transition [29,30]. A chain is defined as a stochastic process having the property where the value of the process at   times t, and X_t, depend only on its value at time $t-1$, X_{t-1}, not on the sequence of values X_{t-2}, X_{t-3};……; X_0 that the process passed through in arriving at X_{t-1} [34]. It can be expressed as:

$$P(x_t = \frac{j}{x_0} = \dot{I}_0, X_1 = i_1, \ \dots, X_{t-1} = i_{t-1})$$

$$= P(x_t = J / \quad X_{t-1} = i_{t-1}) \tag{4}$$

If a Markov sequence of random variant X_n takes the a_n discrete value a_1,………, then:

$$P(x_n = a_{in}/x_{n-1} = a_{i_{n-1}} ...., X_1 = a_{i_1}) \tag{5}$$

$$= P(x_n = a_{i_n}/X_{n-1} = a_{i_{n-1}})$$

The Chi-square (X^2-test) is run to test the change matrix to justify Markov chain, which is expressed as follows:

$$\Sigma_{i=1}^{r}\Sigma_{j=1}^{c} \frac{(Qij - Eij)^2}{Eij} \tag{6}$$

where, $X^2 = $ r = number of row,   c = number of columns, Q*ij* = Actual number in cell *ij*,

E*ij* = Expected number in cell *ij* and   (r-1) * (c-1) = the degrees of freedom

2.7. Simulation of future land use change by CA

Models based on cellular automata (CA) are in extensive use in simulating urban expansion [20]. This is due to the fact that they have demonstrated the capacity to represent complex spatial processes, such as the representation of a variety of local behaviors together with global patterns [14,15]. The (CA) component in the Land use/cover model determines the assignment of land use in

each cell. The (CA) component in the Land use/cover model determines the assignment of land use in each cell.
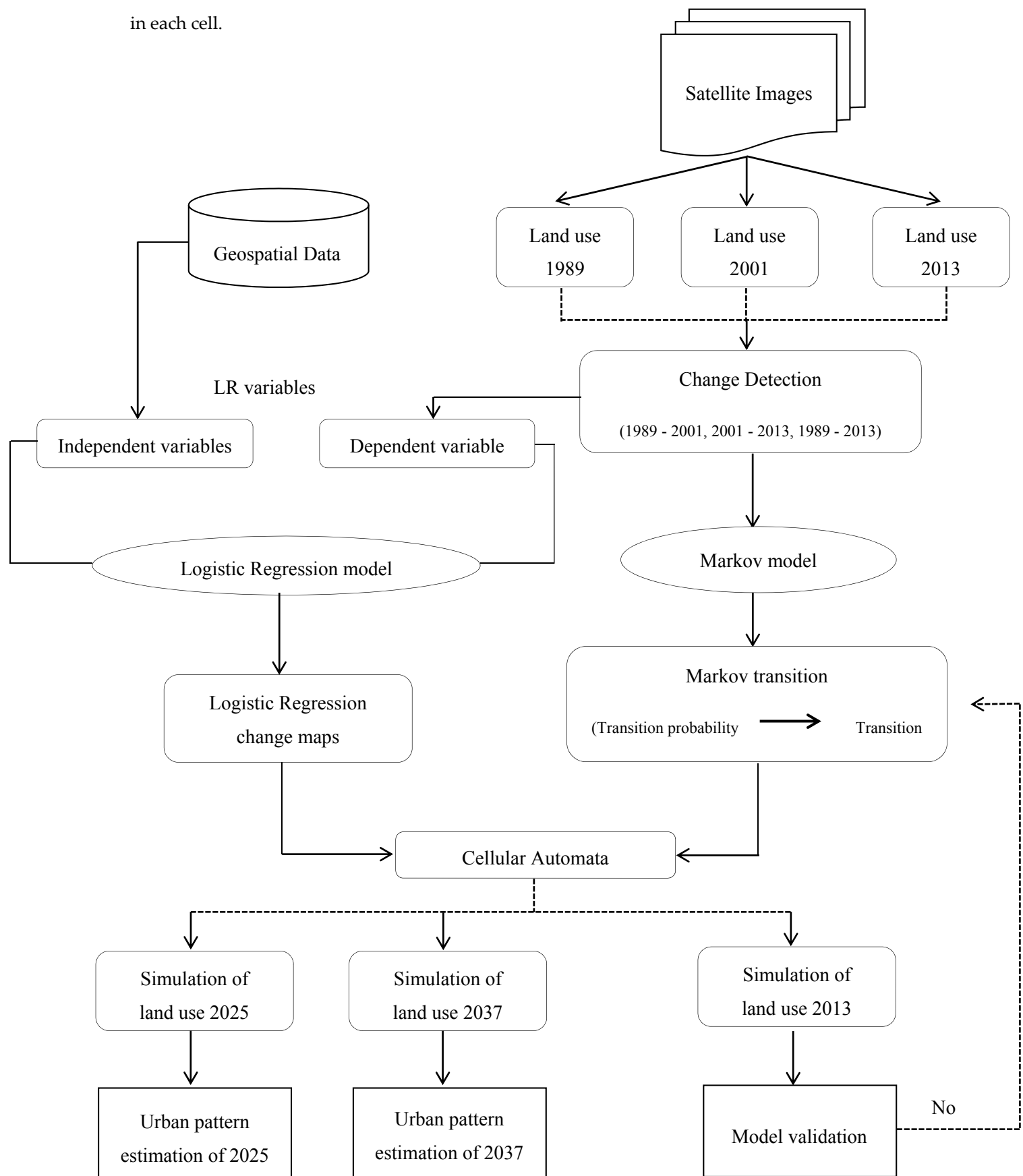

Figure 2. Schematic presentation of the logistic and CA-Markov methods

The (CA) component in the Land use/cover model determines the assignment of land use in each cell. CA models define a new state of land use for each cell for a particular period and as the previously stated function of the cell and cell neighbors it is, in agreement with a group of defined transition rules [35]. In this article, model "CA-Markov" is applied in the IDRISI software for modeling LULC. In this model, the Markov Chain controls the temporal dynamics (gains and losses), while the spatial dynamics is controlled by cellular automata, according to the characteristics of a spatial filter and the potential of land use change [36].

3. Results and discussion

3.1. LULC analysis

The LULC maps of this study area in the two 12 year periods are prepared through Landsat TM, ETM+, and OLI satellite images. Five major land uses are determined in the study area as: built-up areas, agricultural lands, water bodies, forest coverage, and open land (rangeland and bare land). The overall accuracy of the generated maps is estimated through the error matrices. The kappa coefficients of 89.23%, 90.84%, and 91.08% confirm the accuracy of the classified maps of each one of the periods, respectively. The LULC maps of the study area in different periods under study are shown in Figure 3. The LULC map of the province in 2013 shows that the built-up areas are spread mainly in the central parts, over an area of 59754.8 ha.

Table 1. Temporal land use areas in Gilan Province

Class name	Land use 1989		Land use 2001		Land use 2013	
	ha	%	Ha	%	ha	%
Built - up Area	36012.5	2.56	54853	3.91	59754.8	4.26
Forest Coverage	782870	55.76	768541	54.74	751242	53.5
Barren / Open Land	318986	22.72	281880	20.08	266963	19.01
Water Bodies	8632.54	0.61	12237.6	0.87	7213.15	0.51
Agricultural Land	257559	18.34	286548	20.41	318887	22.71

The built-up areas are surrounded by fertile agricultural lands, mostly paddy fields covering 318887 ha and a strip of Hyrcanian forests in the south in an area of 751242 ha. The position of the built-up areas reveals the potential loss of fertile agricultural lands and the unique Caspian Hyrcanian forest ecosystems in its vicinity. In addition, barren/open lands making 266963 ha as a relatively wide strip in the south is narrowing down as the scattered spots in the West and Northwest increase.
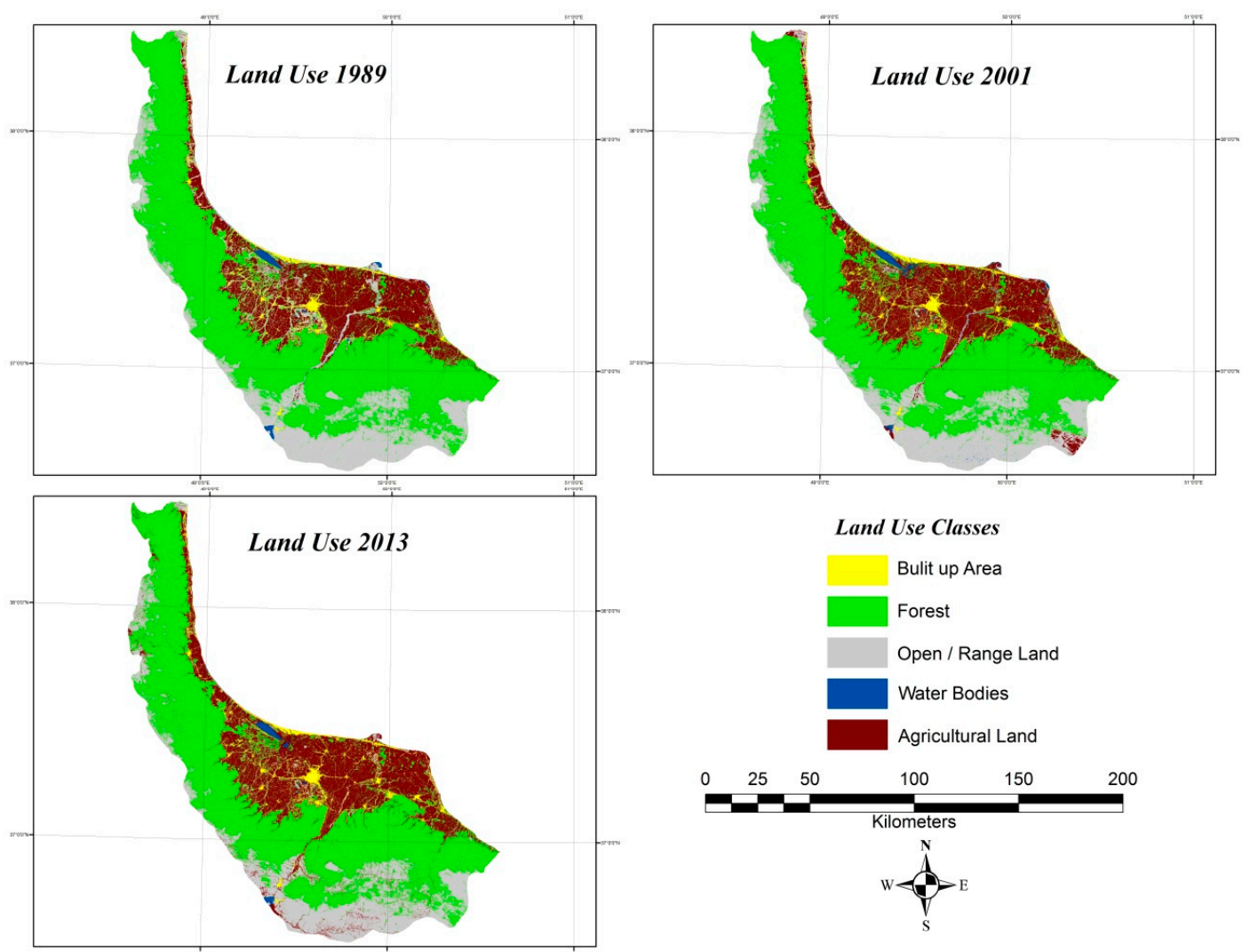

**Figure 3.** LULC maps of Gilan Province in 1989, 2001, and 2013

3.2. LULC change detection

After checking the accuracy of the land use maps, they are cross-tabulated to quantify the variations in different land use classes over time. The results tabulated in Table 2 in the three periods of 1989 - 2001, 2001 - 2013, and 1989 - 2013. As observed in this Table, changes in built-up areas follow

an upward trend over the study period from 1989 to 2013. The expansion is mainly due to conversion of 8644.41 ha of barren lands and 12623.01 ha of agricultural lands to the urban land use (Figure 4). In addition, deforestation of 2076.3 ha of Hyrcanian forests is another reason for urban sprawl in the province. Conversion of 12637.46 ha of forest coverage and 6744 ha of agricultural lands into the barren land could also be a clear indicator for extensive land
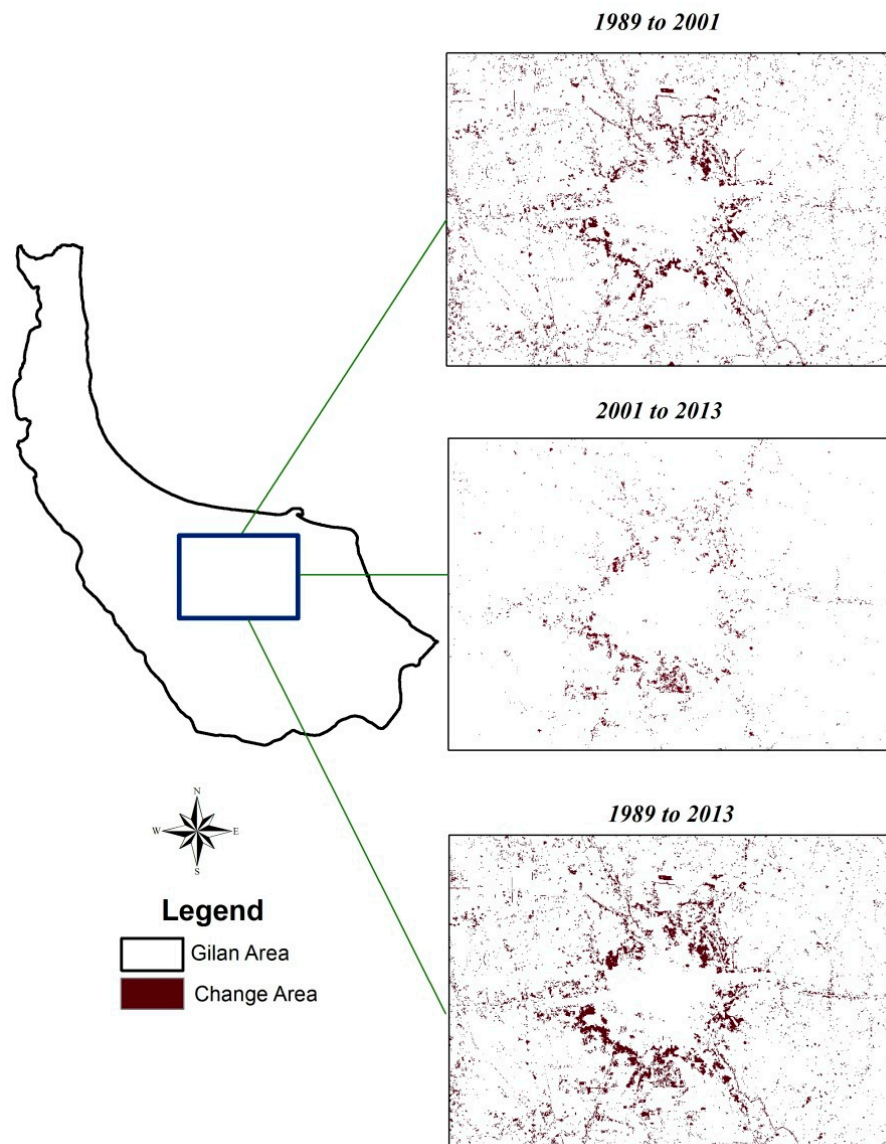
degradation in the study area.

Figure 4. Changes in urban land use in 1989, 2001, and 2013 periods

3.3. CA-Markov projection model

The transitional probability matrix of CA-Markov model in Table 3 reveals the probability of changing from one land use to another which would make it possible to predict changes in future in different land use types in a quantitative sense. Before running the forecast model for the target years 2025 and 2037, the model validity is pre-tested for the year 2013 based on the changes over the period of 1989 – 2001, the probability of the land use changes predicted by the year 2013. According to the obtained results, the greatest land use changes could be observed in the conversion of barren lands into urban areas (probability rate = 0.1367) followed by agricultural lands (probability rate = 0.1147). This is well-approved by the results of the change detection analysis.

Table 2. Change detection results of Gilan Province in two 12 year periods

2001					
	Class	Built - up Area	Forest Coverage	Barren / Open Land	Agricultural Land	Total
1989 Built - up Area	36012.5	0	0	0	36012.5
Forest Coverage	1835.46	774539	2407.86	9017.68	787800
Barren / Open Land	6785.91	2138	313619.1	30395.5	352938.51
Agricultural Land	10075.2	6193	2959	218145.2	237372.43
Total	54709.07	782870	318986	257558.4	1414123.5

2013					
	Class	Built - up Area	Forest Coverage	Barren / Open Land	Agricultural Land	Total
2001 Built - up Area	54853	0	0	0	54853
Forest Coverage	240.84	741270.7	10229.6	16799.9	768541.04
Barren / Open Land	1858.5	1180.98	247402.2	31438.3	281880
Agricultural Land	2547.81	7788.96	3785	272426.2	286548
Total	59500.15	750240.6	261416.8	320664.4	1391822

2013					
	Class	Built - up Area	Forest Coverage	Barren / Open Land	Agricultural Land	Total
1989 Built - up Area	36411.08	0	0	0	36411.08
Forest Coverage	2076.3	733941	12637.46	25818.17	774472.93
Barren / Open Land	8644.41	3319	247581.54	61833.8	321378.75
Agricultural Land	12623.01	13982	6744	231235.03	264584.04
Total	59754.8	751242	266963	318887	1396846.8

As mentioned earlier, the area of barren lands was changed by 52023 ha over the years 1989 to 2013, confirming the fact that barren lands were subject to the greatest changes in the given period. By comparing the simulated and real land use changes for the year 2013, it can be deduced that this developed model is capable to well estimate the future land use changes. The obtained Kappa coefficient of 0.89 indicates a good fitness between the actual and the simulated maps. This fitness indicates the accuracy of the simulation. Therefore, it was run to simulate changes by the years 2025

and 2037. According to the probability matrix, the probability of the barren lands to be converted into urban areas by the year 2025 is very high. The probability rate is estimated at 0.1235.

Table 3. Markov transition probability matrixes for 1989 - 2001, 2001 - 2013, and 1989 - 2013

	CA - Markov 2013	Built - up Area	Forest Coverage	Barren / Open Land	Water Body	Agricultural Land
1989 – 2001	**Built - up Area**	0.9697	0.0098	0.0084	0.0053	0.0068
	Forest Coverage	0.0548	0.8397	0.0915	0.0003	0.0137
	Barren / Open Land	0.1367	0.0069	0.7349	0.0068	0.1147
	Water Body	0.0000	0.0014	0.0089	0.9858	0.0039
	Agricultural Land	0.0557	0.0239	0.0489	0.0000	0.8715
	CA - Markov 2025	Built - up Area	Forest Coverage	Barren / Open Land	Water Body	Agricultural Land
2001 – 2013	**Built - up Area**	0.9859	0.0084	0.0031	0.0002	0.0024
	Forest Coverage	0.0431	0.8195	0.1289	0.0002	0.0083
	Barren / Open Land	0.1235	0.0043	0.7456	0.0049	0.1217
	Water Body	0.0000	0.0013	0.0106	0.9872	0.0009
	Agricultural Land	0.0521	0.0102	0.0445	0.0000	0.8932
	CA - Markov 2037	Built - up Area	Forest Coverage	Barren / Open Land	Water Body	Agricultural Land
1989 – 2013	**Built - up Area**	0.9881	0.0055	0.0033	0.0002	0.0029
	Forest Coverage	0.0691	0.8164	0.1023	0.0006	0.0116
	Barren / Open Land	0.1449	0.0074	0.7095	0.0059	0.1323
	Water Body	0.0000	0.0007	0.0085	0.9892	0.0016
	Agricultural Land	0.0579	0.0312	0.0504	0.0000	0.8605

The agricultural and forest areas with probability rates of 0.0521 and 0.0431 respectively are the other two land uses with high potential in being changed into urban areas by 2025. As expected, conversion of water bodies into the built-up areas is almost impossible. The probability rate is zero, in simple terms; water bodies are the only land use that is likely to remain unchanged in the known future. The remarkable point is the high probability for conversion of forests to barren lands (probability rate is 0.1289). This indicates accelerated land degradation in the study area, which is apparent in the form of loss of forest areas possibly due to mismanagement of Hyrcanian Mixed Forests. Likewise, the agricultural and barren lands are highly subject to be converted to urban

areas with probability rate of 0.0521 and 0.1235 respectively. This clearly clarifies the need for a reform in the current agricultural practices. The transitional probability matrix for the target year 2037 predicts the same trend of conversion in different land use. Among various land use classes, barren lands account for the greatest conversion into the urban areas (0.1449) followed by agricultural lands (0.1323). Furthermore, forest areas are highly prone to change into barren lands (0.1023). Further details are tabulated in Table 3. The findings in Table 4 reveal that the built-up areas will be expanded up to 71263 ha (5.08%) by the end of the first target year 2025. The urban land use increasing trend in urban development will continue up to 78073 ha (5.56%) until the second target year 2037. The greatest change is expected to occur in the forest coverage area so that 17951 ha of the Hyrcanian forests will be disappeared by 2025 and 25101 ha by the year 2037. Changes in the area of agricultural land will be significant, that is, an increase of 16509 ha (1.18 %) by 2025 and 28461 ha (2.03 %) by 2037. It is also likely to loss 9931 ha (0.7 %) and 21774 ha (1.55 %) of the barren lands by 2025 and 2037, respectively. The future expansion of urban areas by the target years 2025 and 2037 are shown in Figure 5.

Table 4. Predicted land use changes in Gilan Province, available: 2013 and predicted: 2025 and 2037

	CA - Markov 2013		CA - Markov 2025		CA - Markov 2037	
Class Name	ha	%	Ha	%	ha	%
Built - up Area	58749	4.18	71263	5.08	78073	5.56
Forest Coverage	747154	53.21	733291	52.23	726141	51.72
Barren / Open Land	265397	18.90	257032	18.31	245189	17.46
Water Bodies	6302	0.45	7078	0.5	7309	0.52
Agricultural Land	326458	23.25	335396	23.89	347348	24.74

As observed, the urban development would go through its gradual evolution primarily on the periphery of the current population centers with an orientation towards surrounding agricultural lands and forested areas. Most of the to be established built-up areas will be observed with a central concentration surrounded with scattered networks of low-density areas.
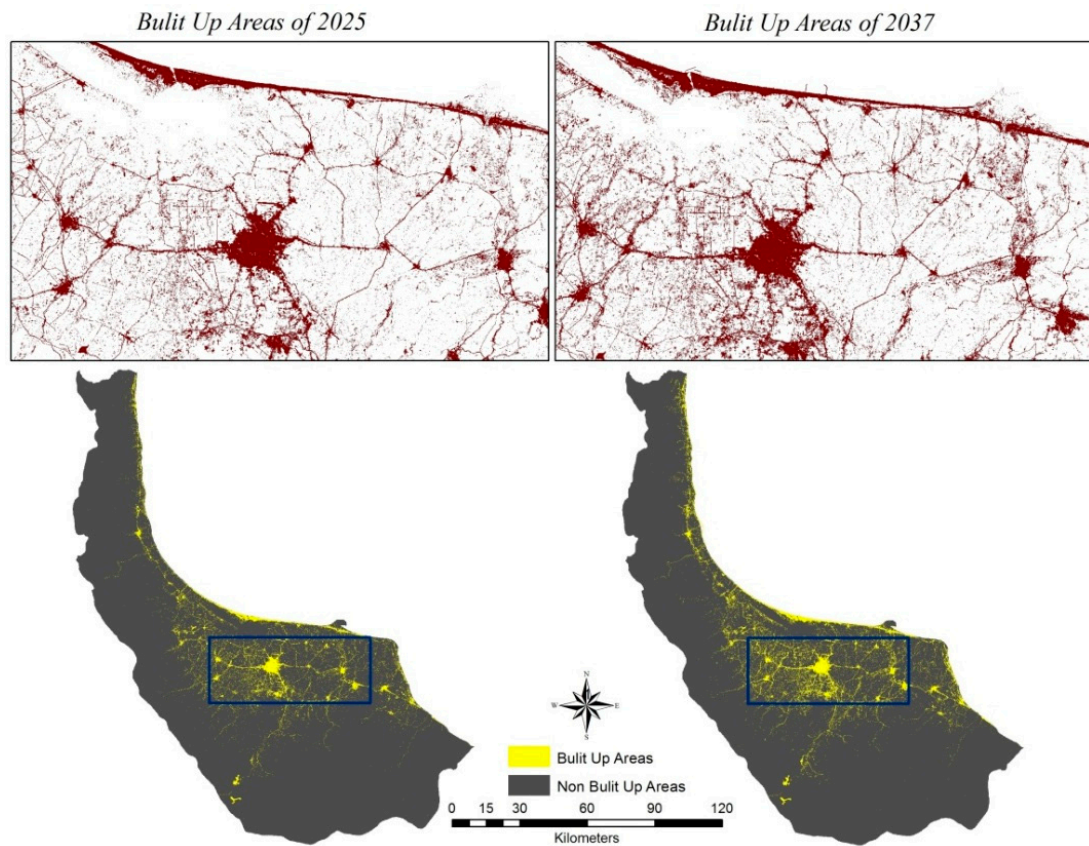
Figure 5. Simulated urban expansion based on CA-Markov model by the target years 2025 and 2037

3.4. CA-logistic regression model

The logistic regression model is developed by 12 independent variables, determined by literature reviews and expertise opinion (Figure 6). This model is implemented separately for each of these variables as the input data for the period 1989 to 2013. The dependent variable in this implementation is the developed cells, that is, change from no built-up areas to built-up areas which is prepared as a binary raster map where the value 1 indicates the changed and zero the unchanged areas during the study period. Figure 6 represents the pattern of dependent variable. The output of the regression model determines the positive or negative role of any variable on the probability of any change regarding land use to be converted to urban. The influence of the independent variables in urban growth development in over the study area in the 1989 to and 2013 period is presented in Table 5. According to this table, agricultural and forest lands, with the coefficient values of 17.901376 and 8.363241, respectively have the highest potential to be converted into urban areas.
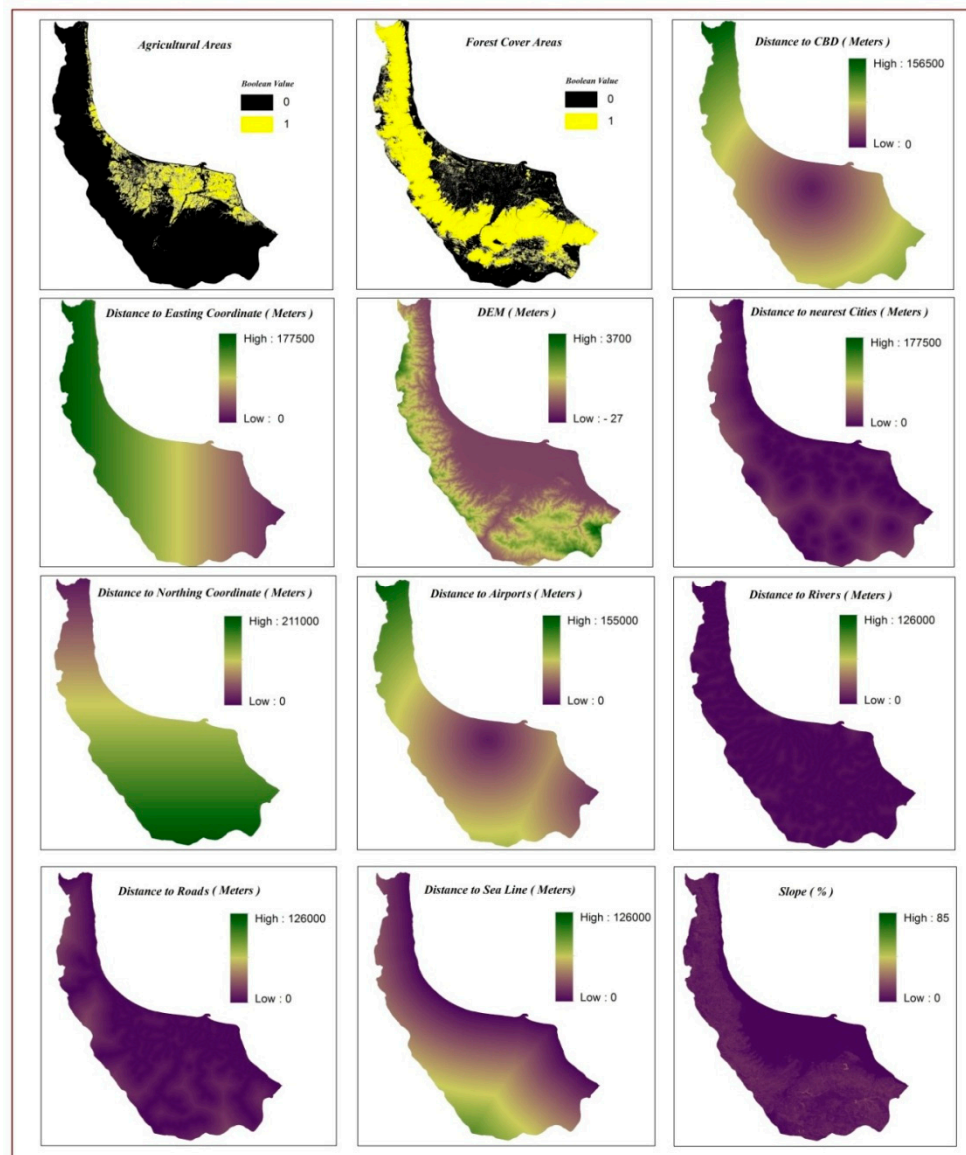
Figure 6. Raster layers of independent variables as the continuous and binary values

The variables of distance to nearest cities (with a coefficient of -0.000498), northing coordinates (with a coefficient of -0.000835), easting coordinates (with a coefficient of -0.000068), and slope (with a coefficient of -0.000319) have negative impact on development of cities in the province. Here, it is assumed that the extent and direction of urban development are dependent on the Northing and Easting geographical coordinates. Since the coefficients of the northern and eastern aspects in LR model are negative, there exists the possibility that land uses in southern and western sections would be converted into the built-up areas. All in all, the intensity of urban development in the south is more than the west and this fact is approved by the higher coefficient value of the north direction in relation to that of the east (-0.000068). The area map subject to urban sprawl of the
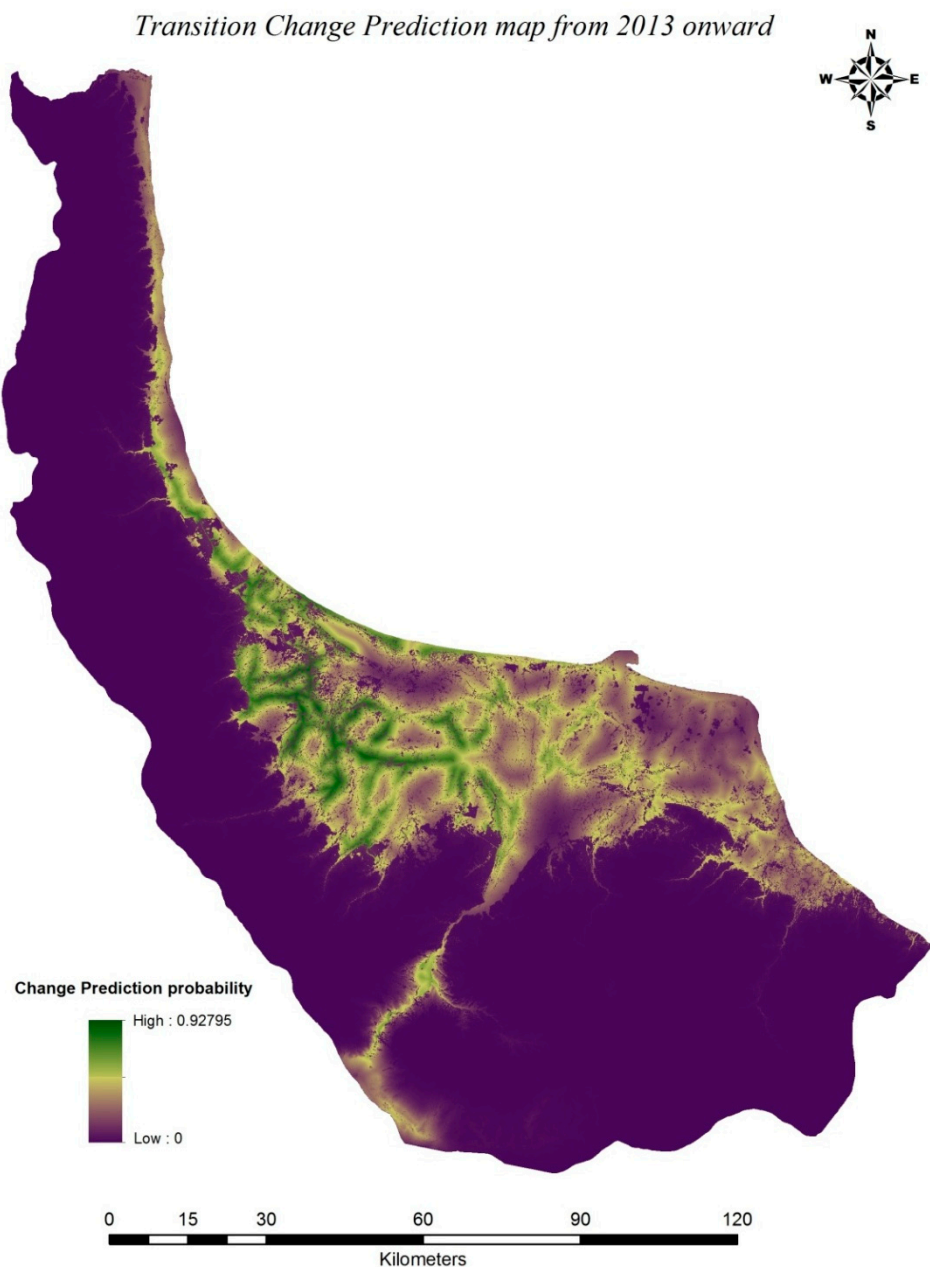
study area is based on logistic regression model, (Figure 7), where the faded areas indicate higher

probabilities for urban expansion.

Table 5. Influence of the independent variables in urban development in 1989 to 2013 period

	Variables	Coefficient	Standard error
X 1	Distance to CBD	0.034231	0.014891
X 2	Distance to airports	0.001698	0.001291
X 3	Distance to rivers	0.014287	0.012732
X 4	Distance to major roads	0.035783	0.023195
X 5	Distance to sea line	0.000031	0.000012
X 6	Distance to nearest cities	-0.000498	0.000184
X 7	Forest lands	8.363241	3.984701
X 8	Agricultural lands	17.901376	11.390572
X 9	Northing coordination	-0.000835	0.000364
X 10	Easting coordination	-0.000068	0.000046
X 11	DEM	-0.001973	0.001394
X 12	Slope in percent	-0.000319	0.000185
	Intercept	-19.13703	10.380974

The expanding pattern in future is distinct and clear in this map, indicating that areas near the major roads and existing urban clusters have the high potential in becoming converted into urban areas in future. The sensitivity analysis is run through ROC method to identify the factors with greatest impact on growth and expansion of built-up areas. To do this, the LR model is run for 12 times in a deductive manner, that is, the next run contains -1 independent variable. Comparing the ROC values of the whole model with the individual ROCs obtained in each run the effect of each of the independent variables is determined in Table 6, where ROC value of 1 shows an excellent spatial

adaption between the simulated and actual urban development maps and ROC value of 0.5 indicates the existence of a chance of agreement between the projected probability and actual map assigned to random locations. The best independent variables are determined by the ROC values

and the adjusted odds' ratios for each set of the variables.

Figure 7. Probable areas of urban sprawl in Gilan Province based on logistic regression model

The validity and fitness of the selected collection is confirmed through the adjusted odds' ratios. As observed in Table 6, the ROC of the LR model run by excluding the variable of distance from residential areas is 0.7351. It can deduced that distance to the nearest cities will have the greatest impact on future urban development, followed by distance to agricultural lands, slope, and distance to main roads.

As observed in Figure 7, the pixels closer to the major roads have a higher potential to become converted into urban areas are of low probability of change due to the cells with a high elevation (DEM). Moreover, the contribution of other input data like: distance to industrial centers, distance to educational centers, and barren/open lands is assessed as being negligible due to their low ROC values, (Table 6). In general, the obtained results on capability of Markov chain model and cellular automata for simulation of future changes in urban land are in accordance with those of Alsharif and Pradhan [26] and Deep and Saklani [37].

Table 6. Estimated statistical values for the logistic regression model consisting of 12 independent variables

Independent Variables	Odds ratio	ROC	Standard error
Collection 1	2.397853	0.9017	0.014891
Collection 2	1.852981	0.9164	0.023195
Collection 3	2.461983	0.8735	0.012732
Collection 4	1.679826	0.8436	0.001291
Collection 5	2.789532	0.9216	0.000012
Collection 6	0.000085	0.7351	0.000364
Collection 7	9.873901	0.9431	3.984701
Collection 8	0.000381	0.7539	0.000185
Collection 9	2.287643	0.9142	0.000184
Collection 10	0.090367	0.8974	0.000046
Collection 11	7.679031	0.9673	11.390572
Collection 12	0.000238	0.7976	0.001394

4. Conclusion

Understanding the dynamics of urban expansion would assist the managers and urban planner to draw rational decisions when the control of urban sprawl and its consequences in future are of

concern. Simulation and forecast of urban expansion is a must for urban planners and land conservationists in formulating sustainable development strategy(s). In this assessment, the LR model is combined with MC and CA models in order to develop an efficient hybrid geospatial explicit approach. The logistic regression model explores correlation between land conversion and the causative factors in a quantitative sense, which enables the researchers to distinguish the essential driving forces. The customary LR models in spatial studies have the restrictions of their own like: mere focus on natural factors and disregarding the spatial-interacted impact on key industries; accordingly, the integrated approach is introduced and implemented to compensate the existing constraints. How has the unsustainable urban development delivered its adverse effect on the natural landscape at the cost of major loss in the Hyrcanian forests and rangelands is revealed in this article. It is predicted that this deteriorating trend would continue in future. According to a close estimation by CA imagines the built-up areas will be developed in 71263 ha by the year 2025 and 78073 ha by the year 2037. According to the estimates, the urbanization in Gilan Province follows an upward trend, that is, an expansion of about 5.65 % over the period between 1989 and 2037. Moreover, the results of land use change detection demonstrated a continuous decrease in agricultural and forest lands in the study area for the same period. The situated land use maps of 2025 and 2037 reveal an alarming descending acceleration in the non-urban areas. Based on the predicted results, the expansion of the urban areas would follow a scattered pattern, while indicating extensive loss of forests and natural rangelands which would in turn become an obstacle for sustainable development. This fact justifies the need for adopting of comprehensive planning approaches to develop an up-dated urban sprawl management system to control expansion of the cities in Gilan Province.

Acknowledgments

References

[1]   G. Caruso, M. Rounsevell, and G. Cojocaru, "Exploring a spatio-dynamic neighbourhood-based model of residential behaviour in the Brussels periurban area," *Int. J.*

Geogr. Inf. Sci., vol. 19, no. 2, pp. 103–123, 2005.

[2]   N. Nazarnia, C. Schwick, and J. A. G. Jaeger, "Accelerated urban sprawl in Montreal,

Quebec City, and Zurich: Investigating the differences using time series 1951–2011," *Ecol.*

Indic., vol. 60, pp. 1229–1251, Jan. 2016.

[3]   G. Engelen, C. Lavalle, J. I. Barredo, M. der Meulen, and R. White, "The MOLAND

modelling framework for urban and regional land-use dynamics," in *Modelling land-use*

change, Springer, 2007, pp. 297–320.

[4]   U. N. UN, "World urbanization prospects: The 2007 revision," *New York UN*, 2008.

[5]   P. N. Dadhich and S. Hanaoka, "Remote sensing, GIS and Markov's method for land use

change detection and prediction of Jaipur district," *J. Geomatics*, vol. 4, no. 1, pp. 9–15, 2010.

[6]   A. G. Wilson, "Urban and regional models in geography and planning," 1974.

[7]   N. C. Goldstein, J. T. Candau, and K. C. Clarke, "Approaches to simulating the 'March of

Bricks and Mortar,'" *Comput. Environ. Urban Syst.*, vol. 28, no. 1, pp. 125–147, 2004.

[8]   F. Wu, "Calibration of stochastic cellular automata: the application to rural-urban land

conversions," *Int. J. Geogr. Inf. Sci.*, vol. 16, no. 8, pp. 795–818, 2002.

[9]   O. L. Puertas, A. Brenning, and F. J. Meza, "Balancing misclassification errors of land cover

classification maps using support vector machines and Landsat imagery in the Maipo river

basin (Central Chile, 1975--2010)," *Remote Sens. Environ.*, vol. 137, pp. 112–123, 2013.

[10]   T. Liu and X. Yang, "Monitoring land changes in an urban area using satellite imagery, GIS

and landscape metrics," *Appl. Geogr.*, vol. 56, pp. 42–54, 2015.

[11]   L. Poelmans and A. Van Rompaey, "Complexity and performance of urban expansion

models," *Comput. Environ. Urban Syst.*, vol. 34, no. 1, pp. 17–27, Jan. 2010.

[12]   P. Verhagen and T. G. Whitley, "Integrating archaeological theory and predictive modeling:

a live report from the scene," *J. Archaeol. Method Theory*, vol. 19, no. 1, pp. 49–100, 2012.

[13]   Y. Li, X. Zhu, X. Sun, and F. Wang, "Landscape effects of environmental impact on bay-area wetlands under rapid urban expansion and development policy: A case study of Lianyungang, China," *Landsc. Urban Plan.*, vol. 94, no. 3–4, pp. 218–227, Mar. 2010.

[14]   J. Liao, L. Tang, G. Shao, X. Su, D. Chen, and T. Xu, "Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations," *Environ. Model. Softw.*, vol. 75, pp. 163–175, 2016.

[15]   Z. Hu and C. P. Lo, "Modeling urban growth in Atlanta using logistic regression," *Comput. Environ. Urban Syst.*, vol. 31, no. 6, pp. 667–688, 2007.

[16]   J. J. Arsanjani, M. Helbich, W. Kainz, and A. D. Boloorani, "Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 21, pp. 265–275, 2013.

[17]   O. L. Puertas, C. Henríquez, and F. J. Meza, "Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago Metropolitan Area, 2010--2045," *Land use policy*, vol. 38, pp. 415–425, 2014.

[18]   A. J. Mousivand, A. Alimohammadi Sarab, and S. Shayan, "A new approach of predicting land use and land cover changes by satellite imagery and Markov chain model (Case study: Tehran)," MSc Thesis. Tehran: Tarbiat Modares University, 2007.

[19]   J. Han, Y. Hayashi, X. Cao, and H. Imura, "Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China," *Landsc. Urban Plan.*, vol. 91, no. 3, pp. 133–141, 2009.

[20]   P. M. Torrens and D. O'Sullivan, "Cellular automata and urban simulation: where do we go from here?," *Environ. Plan. B Plan. Des.*, vol. 28, no. 2, pp. 163–168, 2001.

[21]   T. Munshi, M. Zuidgeest, M. Brussel, and M. van Maarseveen, "Logistic regression and cellular automata-based modelling of retail, commercial and residential development in the city of Ahmedabad, India," *Cities*, vol. 39, pp. 68–86, Aug. 2014.

[22]   M. K. Jat, P. K. Garg, and D. Khare, "Monitoring and modelling of urban sprawl using remote sensing and GIS techniques," *Int. J. Appl. earth Obs. Geoinf.*, vol. 10, no. 1, pp. 26–43, 2008.

[23]   Iranian Statistic Center, "No Title," *Census data. Tehran,Iran.*, 2011.

[24]   O. Dubovyk, R. Sliuzas, and J. Flacke, "Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 2, pp. 235–246, 2011.

[25]   C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, 2002.

[26]   A. A. A. Alsharif and B. Pradhan, "Urban sprawl analysis of Tripoli Metropolitan city (Libya) using remote sensing data and multivariate logistic regression model," *J. Indian Soc. Remote Sens.*, vol. 42, no. 1, pp. 149–163, 2014.

[27]   C. Xie, B. Huang, C. Claramunt, and C. Chandramouli, "Spatial logistic regression and GIS to model rural-urban land conversion," in *Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications*, 2005, pp. 12–15.

[28]   W. A. V Clark, "Markov chain analysis in geography: an application to the movement of rental housing areas," *Ann. Assoc. Am. Geogr.*, vol. 55, no. 2, pp. 351–359, 1965.

[29]   I. Kim, G. Jeong, S. Park, and J. Tenhunen, "Predicted land use change in the Soyang river basin, South Korea," in *2011 TERRECO Science Conference, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany*, 2011, pp. 2–7.

[30]   J. R. Drewett, "A stochastic model of the land conversion process: An interim report," *Reg. Stud.*, vol. 3, no. 3, pp. 269–280, 1969.

[31]   Q. Wu, H. Li, R. Wang, J. Paulussen, Y. He, M. Wang, B. Wang, and Z. Wang, "Monitoring and predicting land use change in Beijing using remote sensing and GIS," *Landsc. Urban*

Plan., vol. 78, no. 4, pp. 322–333, 2006.

[32] X. Yang, X.-Q. Zheng, and L.-N. Lv, "A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata," *Ecol. Modell.*, vol. 233, pp. 11–19, 2012.

[33] Q. Zhang, Y. Ban, J. Liu, and Y. Hu, "Simulation and analysis of urban growth scenarios for the Greater Shanghai Area, China," *Comput. Environ. Urban Syst.*, vol. 35, no. 2, pp. 126–139, 2011.

[34] F. Fan, Y. Wang, and Z. Wang, "Temporal and spatial change detecting (1998--2003) and predicting of land use and land cover in Core corridor of Pearl River Delta (China) by using TM and ETM+ images," *Environ. Monit. Assess.*, vol. 137, no. 1–3, pp. 127–147, 2008.

[35] A. Rabbani, H. Aghababaee, and M. A. Rajabi, "Modeling dynamic urban growth using hybrid cellular automata and particle swarm optimization," *J. Appl. Remote Sens.*, vol. 6, no. 1, pp. 63581–63582, 2012.

[36] D. J. Maguire, M. Batty, and M. F. Goodchild, *GIS, spatial analysis, and modeling*. Esri Press, 2005.

[37] S. Deep and A. Saklani, "Urban sprawl modeling using cellular automata," *Egypt. J. Remote Sens. Sp. Sci.*, vol. 17, no. 2, pp. 179–187, 2014.