

Design of a Dynamic Land-Use Change Probability Model Using Spatio-Temporal Transition Matrix

Yongjin Joo¹, Chulmin Jun², and Soohong Park³

¹ Institute of Urban Sciences, University of Seoul, Seoul, Korea
yjjoo75@uos.ac.kr

² Dept. of Geoinformatics, University of Seoul, Seoul, Korea
cmjun@uos.ac.kr

³ Dept. of Geoinformatic Engineering, Inha University, Incheon, Korea
shpark@inha.ac.kr

Abstract. This study aims to analyze land use patterns using time-series satellite images of Seoul Metropolitan Area for the past 30 years, and present a macroscopic model for predicting future land use patterns using Markov Chain based probability model, and finally examine its applicability to Korea. Several Landsat MSS and TM images were used to acquire land-use change patterns and dynamic land-use change patterns were categorized from the classified images. Finally, spatio-temporal transition matrices were constructed from the classified images and applied them into a Markov Chain based model to predict land-use changes for the study area.

Keywords: land-use change prediction, spatio-temporal transition matrix, Markov Chain, urban growth model.

1 Introduction

Urban economist and planners have consistently studied how urban areas have developed and what primary factors have affected. However, those research efforts have not sufficiently presented theoretical models. That's because the aspect of urbanization is different between countries and varies with time. In addition, the process of urbanization is so complicated that proposing theoretical validity is difficult through feasible verification [10].

Detecting an urban spatial structure and predicting changing trend is very important information in establishing the efficient urban policies. In Korea, however, there have been minimal research efforts regarding analysis and prediction of the characteristics of dynamic changes of land use. In order to predict land use change, models represented in terms of space-time is needed and a variety of variables and data supporting the model are also required [11]. The models in the previous studies, with insufficient time-series data, show limitations in incorporating the past tendencies of urbanization and explaining the past land use changes. Investigating the current state of land use and comparing with the past ones requires significant time and efforts. In

the areas as Seoul Metropolitan Area (SMA), which shows fast population growth and development, detecting the land-use variations happened in the past is very difficult. In this situation, utilizing remote sensing data is a practical alternative for monitoring visible change of urban spaces.

First of all, in this study, we examine the characteristics of land use transitions through the time-series images of Landsat in SMA. Then, we develop a prediction model for land-use change based on Markov chain methods and apply it to the simulation of the land-use transition processes. Finally, we examine the validity of prediction result using the actual data of 1984, 1992. Before the implementation of the model, we set up the prototype of the model through the consideration of the spatial characteristics of the study area, the definition of model components, establishing input variable data, and designing basic algorithm for the model. Satellite images (MSS, TM), digital maps and data of the limited development district were used to establish the prediction model. In other words, input data which are relevant to land-use changes (topography and social phenomenon) and land cover data are developed. Then, by calculating the land use conversion rate, a transition matrix was composed on two periods--1972~1984, 1984~1992. The suitability of the model was evaluated by using a validation method comparing the derived results with the actual data (1984, 1992).

2 Markov Transition Model

2.1 Model Framework

Analysis of Markov Chain, a statistical method was used for predicting how topographical and social variations affect on the land use changes in the future through examining dynamic characteristics of the past. It is based on the process of probability called Markov Chain, which assumes that present state is determined only by the immediate previous state. It is composed of the system state and transition probability. The changes of states are called transitions, and the probabilities associated with various state-changes are called transition probabilities.

When a probability analysis can be performed on matrix of random events accompanied by time, $\{X_t\}$, the row of random variables of each event is referred to as stochastic process. If random variables $X_t(t=1,2,\dots)$ change into one of state sets ($S_1, S_2, \dots S_k$) at a certain moment, transforming from state S_i to S_j is called a step. And when the transition probability from S_i to S_j (P_{ij}) is only related to the immediate previous state S_i , such probability process is Markov Process. The following formula (1) indicates the transition probability.

$$P_{ij} = P \{X_n = S_j \mid X_{n-1} = S_i \wedge P\} = P \{X_n = S_j \mid X_{n-1} = S_i\} \quad (1)$$

The transition matrix of P_{ij} is defined as follows;

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1k} \\ P_{21} & P_{21} & \dots & P_{2k} \\ P_{k1} & P_{k1} & \dots & P_{kk} \end{bmatrix} \text{ for } 0 \leq P_{ij} \leq 1 \quad \sum P_{ij} = 1 \quad (2)$$

In its simplest form, the state vector $X(t)$ can be described if types of land use are categorized into urban, water, forest, and agriculture. The formula is defined as follows;

$$X(t) = \begin{bmatrix} X1(t) \\ X2(t) \\ X3(t) \\ X4(t) \end{bmatrix} \quad (3)$$

In the above formula, vector $x(t)$ is land use/cover, transition probability P is land use and land-use change from time t to $t+1$ is defined as. $X(t+1) = P * x(t)$. Each element of transition matrix P_{ij} is the probability to move from i type of land use in time t to j type of land use in time $t+1$. For example, let's suppose there are 100 pixels of forests in t time. If, after 20 years, there still remain 78 pixels of forests, 12 pixels change into agricultural area and 10 pixels to urban area, then P_{ij} is described as;

- $P_{31} = 10 / 100 = 0.10$
- $P_{32} = 0 / 100 = 0.00$
- $P_{33} = 78 / 100 = 0.78$
- $P_{34} = 12 / 100 = 0.12$

With the transition matrix being described as;

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix} = \begin{bmatrix} \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0.10 & 0.00 & 0.78 & 0.12 \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (4)$$

2.2 Problems of Spatial Influence Algorithm

Markov model has been used to predict aspects of land-use changes by human activities and understand the changing processes of natural forms [8]. The probability of a land cover change in this model is based on spatial influence algorithm with neighborhood effects having influence on adjacent land cover [12]. On the whole, this algorithm supposes that change of a cell is carried out by transition probabilities. The range of influence that adjacent cells reach can be set up with 4 or 8 neighbor cells. The values are set in two-dimensional square cells and transition state of land use class from previous time (t_1) to next time (t_2) is calculated. After that, change of cell in the space can be simulated by the time according to calculated transition probability.

Markov model is easily computed by using digital image or raster-based GIS data and has an advantage to reflect transition tendency of current land use effectively. Even though time passes along, transition matrix is always constant and applied equally to all locations [7]. However, actual land use doesn't change exactly according to the assumption of Markov and obtaining the transition probability through independent measurement is difficult. Also, the factors for land-use transition are

more affected by political and economic factors rather than biophysical ones [1]. Therefore, in case we apply previous algorithm to our study area, the Seoul metropolitan area where urbanization has been on the rise at an extremely rapid rate for a short period in Korea, some anticipated problems are as follows;

First of all, limited development districts (also called green-belt areas) have been set in South Korea since 1972. It has been playing an important role in controlling spread of urbanization and preserving green spaces. Because the change of center pixel is affected only by the land uses of adjacent cells, it leads to a spread of habitation inside the limited development areas. Secondly, one of the most important factors relevant to land-use change is the slope. The high slope prevents the regions from being developed and populated by the development permit system in Korea. Unless the physical properties of the land are considered, habitation cells located in hill sides and low mountains will spread into neighboring cells.

In this paper, we thus improved previous model into a more practical land-use model engrossed in urban structural change, which can incorporate the concept of multi-dimensional spatial filter. In other words, political factor of land use regulation (green-belt policy) is considered to prevent urbanized cells in green area and green-belts from spreading. More importantly, we developed the methodology for dynamic probabilities of transition matrix with the help of practical multi-temporal satellite images accumulated for long periods.

3 Design of Land-Use Change Model

Land-use change model that we suggest is based on Cellular automata (CA), which are both a body of knowledge and set of techniques for solving complex dynamic-systems problems [9]. The model includes four components: a grid space, local states, neighborhoods and a transition rule. Though these components, the model evolves in discrete time steps by updating their local state according to a universal rule that is applied to each cell synchronously at each time step. The value of each cell is determined by a geometrical configuration of neighbor cells, and is specified in the transition rule. Updated values of individual cells then become the inputs for the next iteration. In this chapter, we specify each element for design of land-use change model in detail.

Grid space: The first element of this model is grid spaces, which mean regular grid of cells where interactions for urban sprawl are carried out. Theoretically, there is no restriction to the tessellation of a grid space and it could be various forms of shape. In general, however, square cells are the most common form in CA applications due to their inherent convenience of implementation in computation. Therefore, grid space in this study consists of regular square cells of 2 dimensions. Besides, because grid spaces are determined by the spatial resolution of satellite images, input data (land cover, green-belt, and slope) are determined as 60m grids according to Landsat MSS, the lowest size of the sensor resolution.

Local states: The second element of the model is the local state. The local states mean the status of each cell encoded by numerical values at a given time step. The

range of its values is defined in the transition rules and depends on the actual implementation of model. In raster GISs, these local states are directly analogous to the values of each grid cell in a layer. The local states in our study represent the land-use characteristics of cells, that is, values assigned in all cells. Table 1 shows the cell state of input data that we used in our experiment.

Table 1. The cell states of input data

Thematic Data		State	Grid Space(m)	Data type
Land cover	Urban	1	60×60	Integer
	Water	2	60×60	
	Forest	3	60×60	
	Agriculture	4	60×60	
Green-belt	Exclusive region	1	60×60	Byte
	Non-Exclusive region	0	60×60	
Gradient map	Slope	0×90	60×60	Integer

Neighborhood cell: Neighborhood is a set of cells located adjacent to focus cell. Such a neighborhood concept is very similar to the mask or moving window of spatial filters in digital image processing and GIS. In theory, there is no limitation to the size of a neighborhood and usually the configuration of a neighborhood can be extended to the temporal dimension as well as the spatial dimensions. In this study, the Neighborhood of Moor was used as the neighborhood definition. In fig 1, the cell in the 3×3 window (i.e, the neighborhood) is changed in a discrete time step according to the transition rule.

Transition rule: GIS data such as land cover data of time series built from satellite images, digital elevation models (DEM), and green-belt data are considered as input variables. The local transition rule and constraints are applied to grid spaces repeatedly resulting in state transitions from time t to time $t + 1$. This transition rule defines how each cell changes every time step, and models the process that a state of a cell changes constantly in accordance with the effects of neighboring cells. As the process of this algorithm, transition matrix is calculated by using time-periodic transition probability in the study area. Transition index is calculated through examining the state of the focus cell and the adjacent state of 8 cells representing land use followed by the computation of the transition index. The transition index is the maximum value j of $N_j \times P_{ij}$ (where N_j is the number of land use elements in the current window size and P_{ij} is the element in the transition matrix from i to j). If the returned value of transition index is urban, then model checks for such constraints as green-belt and slope. In case agricultural cell in green-belt is changed into urban, then its state is maintained. Lastly, transition index is assigned to the cell and move on to next cell. Process and algorithm for model execution are as shown in figure 1.

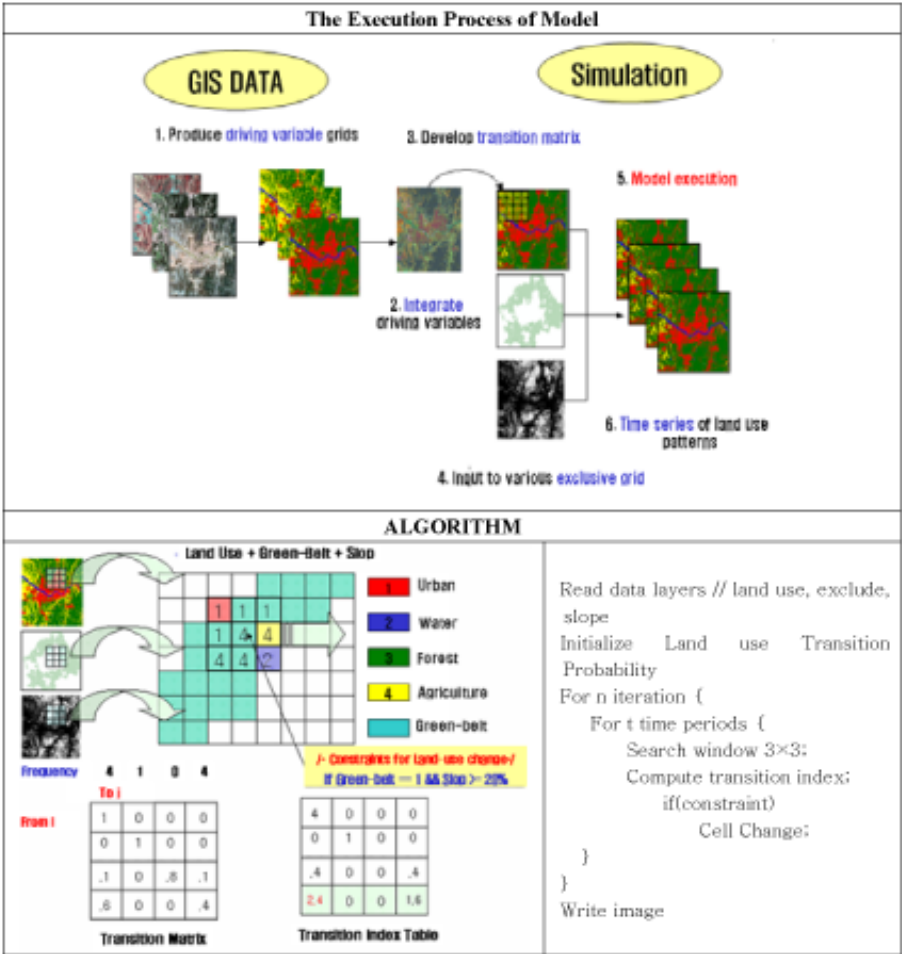


Fig. 1. Process and algorithm for model execution

4 Construction of Land-Use Change Model

4.1 Data Set

In order to apply the proposed land-use change model, we built input data on factors for the past 30 years related with land-use change (topography and social phenomenon) by utilizing the satellite images (MSS, TM), digital maps, and green-belt data. Transition of land use in the study area was analyzed using the land cover maps. We utilized MSS(1972) images, TM(1984, 1992,2000) images for land cover maps. Finally, the water bodies, greenbelt and DEM were used to limit the land-use change.

In Korea, the green-belt system has been established since 1971, and we collected green-belt data from that time to the present. Gradient maps were used by extracting contours and layers only from digital topographic maps, interpolating them in TIN (Triangulated Irregular Network), and then transforming into DEM (Digital Elevation Model) or a percent form. After creating the DEM by using a contoured digital topographical map, we used it as slope for the study. Because the elevation value was less changeable over time, we used the DEM values in the 1990s as substitutes for the 1970s, 1980s and 1990s data. The major characteristics of images and the input data are shown in figure 2.

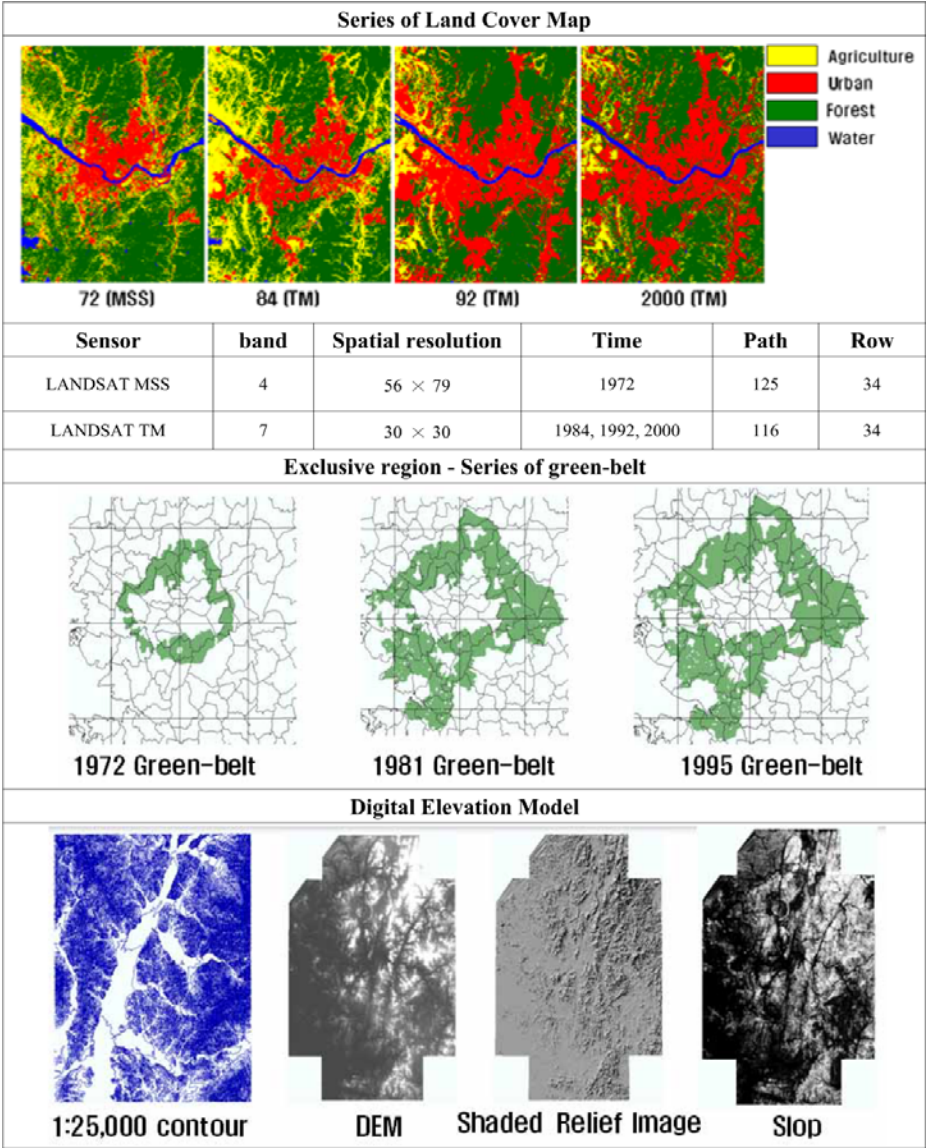


Fig. 2. GIS databases for the study

4.2 Transition Matrix Configuration

In order to configure transition matrix with change type by time, we combined each result of classification into 1972~1984, 1984~1992 respectively. So each mixing type becomes $4^2(16)$ different types for each has 4 categories of 2 periods. In other words, change type has values between 11 and 44. For example, same values mixed over 2 periods such as 11, 22, 33, and 44 means that land use doesn't change. On the contrary, different values mixed over 2 periods, let say, 41 means that the land use was changed from agriculture into urban. The elements of transition matrix are also of 16 and composed of probabilities of land-use change (Table 2).

Table 2. Configuration of transition matrix using time-series land-cover map

First time period (1972 - 1984)					
	Urban	Water	Forest	Agriculture	Total
Urban	0.939	0.011	0.014	0.037	1
Water	0.032	0.927	0.019	0.022	1
Forest	0.055	0.001	0.728	0.217	1
Agriculture	0.240	0.001	0.059	0.699	1
Second time period (1984 - 1992)					
	Urban	Water	Forest	Agriculture	Total
Urban	0.958	0.008	0.023	0.011	1
Water	0.027	0.956	0.008	0.010	1
Forest	0.090	0.001	0.870	0.038	1
Agriculture	0.452	0.003	0.048	0.496	1

4.3 Simulation and Result

Time series experiments of land-use change were performed through simulations and validation methods. Through this, we were able to evaluate the validity of the model and reliability of predicted land-use change. The range of our experiment was from the 1970s to the 1990s. We analyzed the accuracy of the results from the proposed model by using the following formula;

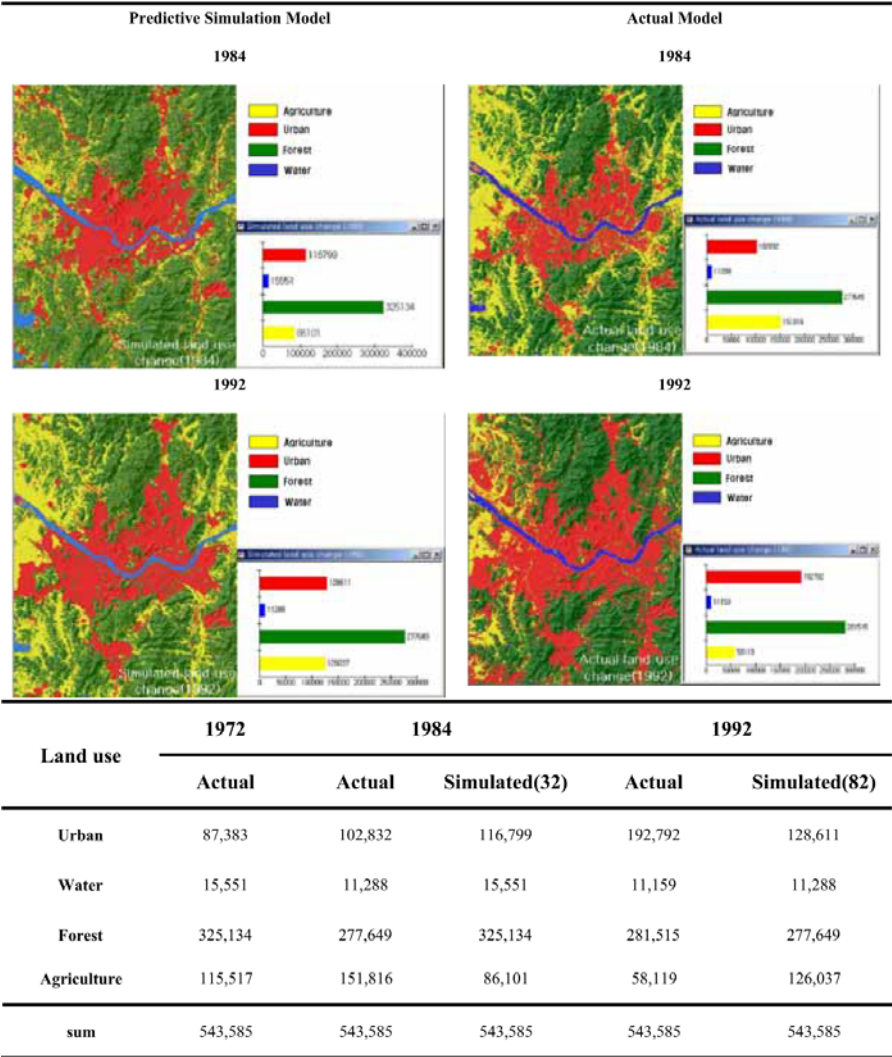
Lee-Sallee Index = No. of urban cells in simulation / No. of urban cells in
actual data

(5)

Lee-Sallee Index shows how correctly the results of modeling match the spatial shape of the actual urban area [4, 5, 6, 13]. Lee-Sallee index was used for the calibration of models, and was estimated by using the number of matching cells between images of

Table 3. The validation result using the Lee-Sallee index

Test	1984	1992
Cells of predicted urban areas	116,799	128,611
Cells of actual urban areas	102,657	115,173
No. of iteration	32	82
Lee-Sallee Index (×)	0.68647	0.61972



urbanized areas and those of from the simulation at a standard point of time. Table 3 shows the results of Lee-Sallee indexes which have the highest spatial similarity.

As a result, following figure 3 represents the results of the simulation through the predictive images calibrated by Lee-Sally index. They were compared against the actual land cover map to verify our proposed model.

(Simulated (n) indicates the simulation images obtained from n times of model processes.)

5 Conclusions

As it is shown in this paper, this study aimed to analyze land use patterns in the past using time-series satellite images of Seoul metropolitan area for the past 30 years, and present a macroscopic model for predicting future land use patterns using Markov Chain-based probability model, and finally examine its applicability to Korea. To accomplish this, we, first, selected Seoul as target area which has urbanized since 1972 at an extremely rapid rate. Second, we constructed input data with regard to constraints (slop, green-belt) and spatio-temporal land cover maps from satellite images, which help categorize dynamic land-use change patterns.

Finally, spatio-temporal transition matrices were constructed from the classified images and they were applied to a Markov Chain-based model to predict land-use changes for the study area. In addition, we evaluated our model through a validation method called Lee-Sally index and simulation experiments to predict 1984, 1992 by using 1972 and 1984 data. We expect that our proposed model, by integrating with existing urban growth models, can be effectively applied in predicting land-use changes in non-urban areas. Though this paper is shown satisfactory consequences, we need to extend our proposed model to microlevel deterministic simulation model in urban domain by using transportation usage and environmental impact so that this model is a tool for use by urban planners, policy makers, and other community stakeholders to help formulate and evaluate combinations of land use, transportation and environmental policies.

Acknowledgments. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF- 2009-413-D00001).

References

1. Alig, R.J.: Econometric analysis of the factors influencing forest acreage trends in the southeast. *For. Sci.* 32, 119–134 (1986)
2. Agarwal, C., Green, G., Grove, M., Evans, T., Schweik, C.: A Review and Assessment of Land-use change Models: dynamics of space, time, and human choice. In: 4th International Conference on Integrating GIS and Environmental Modeling (2000)
3. Acevedo, W., Foresman, T.W., Buchanan, J.T.: Origins and Philosophy of Building a Temporal Database to Examine Human Transformation Processes. In: *Proceedings, ASPRS/ACSM Annual convention and Exhibition*, Baltimore, MD, April 22-24 (1996)

4. Clark, S., Starr, J., Foresman, T.W., Prince, W., Acevedo, W.: Development of the Temporal Transportation Database for the Analysis of Urban Development in the Baltimore-Washington Region. In: Proceeding, ASPRS/ACSM Annual Convention and Exhibition, Baltimore, MD, April 22-24, vol. III, pp. 101-110 (1996)
5. Clarke, K.C., Hoppen, S., Gaydos, L.: Methods and techniques for rigorous calibration of a cellular automaton model of urban growth. In: Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-25, Santa Barbara: National Center for Geographic Information and Analysis (1996)
6. Clarke, K.C., Hoppen, S., Gaydos, L.: A Self-modifying Cellular Automata Model of Historical Urbanization in the San Francisco Bay Area. *EPB* 24, 247-261 (1997)
7. Huang, O., Cai, Y.: Simulation of land use change using GIS-based stochastic model: the case study of Shiqian County. *Stoch. Env. Res. Risk Assess.* 21, 419-426 (2007)
8. Jenerette, G.D., Wu, J.: Analysis and simulation of land-use change in the central Arizona - Phoenix region. *Landscape Ecology* 16, 611-626 (2001)
9. Jung, J.J.: Development of Cellular Automata Model for the Urban Growth. Seoul University (2001)
10. Park, S.: Design and Implementation of an Integrated CA-GIS System. *Geographic Information System Association of Korea* 5(1), 99-113 (1997)
11. Park, S., Joo, Y.-G., Shin, Y.-H.: Design and Development of a Spatio-Temporal GIS Database for Urban Growth Modeling and Prediction. *The Geographical Journal of Korea* 36(4), 313-326 (2002)
12. Turner, M.G.: Spatial simulation of landscape changes in Georgia: a comparison of three transition models. *Landscape Ecology* 1, 29-36 (1987)
13. Kang, Y., Park, S.: A Study on the Urban Growth Forecasting for the Seoul Metropolitan Area. *Journal of the Korean Geographical Society*, 621-639 (2000)