Predicted Land Use Change in the Soyang River Basin, South Korea

Kim, Ilkwon (1); Jeong, GwanYong (1); Park, SooJin (1); Tenhunen, John (2)

(1) Department of Geography, Seoul National University, Seoul, South Korea, scruble@snu.ac.kr
(2) Department of Plant Ecology, University of Bayreuth, 95440 Bayreuth, Germany, john.tenhunen@uni-bayreuth.de

Abstract: In the case of Soyang River basin, water quality has deteriorated due to land use change within the watershed. This study aims to extract variables which influence land use change, and predict future land use changes in the Soyang River basin using 1995 and 2000 land use maps. The prediction map is validated with the map of actual land use for 2006 to test the applicability of the land use prediction model, using relative operating characteristic (ROC) and variations of the Kappa index of agreements (Kno, Klocation, Kstandard) for major land use classes (urban, forest, agriculture). The forest class has a high accuracy for both distribution ratio and spatial agreement, while urban and agriculture classes have a high accuracy for their distribution ratio and a low accuracy for spatial agreement. These results demonstrate that the model can predict overall distribution ratio without predicting exactly where land use changes occur, because urban and agricultural changes are more closely related to socio-economic factors than environmental factors used in this research.

Keywords: *land use change, land use simulation, logistic regression, CA-Markov analysis, Soyang River Basin*

1. Introduction

Land use and land cover changes are the result of human activities and natural environmental changes (Oh et al., 2010). In the case of the Soyang River basin, water pollution of the Soyang River has been aggravated by land use changes in the watershed, particularly highland farming which is considered a main cause of water pollution. In this circumstance, a prediction of land use change can provide an opportunity to deal with future environmental problems and regional changes.

Modeling is an important technique for the projection of alternative pathways into the future, for conducting experiments that test our understanding of key processes, and for describing the future in quantitative terms (Lambin et al., 2000, Veldkamp and Lambin, 2001). Land use change modeling for future prediction is conducted by relating spatial patterns and driving factors on land use change using statistical methods to examine spatial and temporal changes on socio-economic changes. Although land use models are complex systems that reflect the interface of multiple social and ecological systems, various modeling approaches such as IMAGE (Integrated Model to Assess the Global Environment), CLUE (Conversion of Land Use change and its Effects), SALU (SAhelian Land-Use), and LEAN (Land use Evolution and impact Assessment Model) have been developed to consistently understand future land use with respect to growth management (Verburg et al., 2002). Land use change modeling research is based on past land use change patterns and future scenarios to predict land use changes.

This research uses the CA-Markov analysis statistical technique combined with Markov chain and Cellular Automata (CA) theory frameworks. Although the Markov chain model is easy to calculate using grid based GIS data and current land use change patterns, it does not accurately reflect actual land use changes because of the difficulty in processing of spatial data. This method uses fixed transition probabilities and is applied equally to all locations despite of temporal changes.

To solve this problem, CA-Markov changes the states of adjacent grids consistently by applying common change patterns of temporal and spatial data to adjacent grids. Status of changed adjacent grids that are repeatedly practiced can simulate complex attributes and forms, and change the rules of adjacent grids so local characteristics are applied equally to local grids (Lee and Kim, 2007).

The first aim of this study is to understand land use change patterns and their driving factors on the Soyang River basin. Next, we predict future land use change using the CA-Markov model and validate the results to show the probability of prediction of land use changes.

2. Materials and Methods

2.1 Site Description

The Soyang River basin is located in the northeastern part of South Korea and contains the largest branch area of the North Han River. Because it is located near the Taebaek mountain range, the overall physical geographical feature is an incised meander that has a narrow river valley.

2.2 Basic Data

Land use maps consisting of 7 land cover classes (water, urban, bare land, wetland, grassland, forest, and agriculture) were obtained from the Environmental Ministry. Parameters include natural-environmental (altitude, slope, aspect, relief, and curvature, from the DEM) and socio-economic (distance to road, river, stream, urban area, and administrative centers, zoning of reserve forest).

2.3 Methods

2.3.1 Logistic Regression

Logistic regression is a form of regression which is used when the dependent variable is a dichotomy and the independent variables are continuous and categorical variables (Kim, 2002). In land use studies, logistic regression is easier to use because land use studies use a mixture of continuous and categorical variables (Hinde, 2001, Kim, 2002). Moreover, a logistic regression gives the probability of each land use class increasement as a function of the explanatory variables (Schneider and Pontius, 2001).

The function is a monotonic curvilinear response bounded between 0 and 1, given by a logistic function of the form:

$$P = E(Y) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_1 X_2 + \ \ldots}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots}$$

where p is the probability of land use increase in the cell, E(Y) the expected value of the binary dependent variable Y, $ß_0$ is a constant to be estimated, $ß_1$ a coefficient to be estimated for each independent variable X_i (Schneider and Pontius, 2001).

2.3.2 CA-Markov Chain

The Markov chain analysis is a widely used technique for predictive land use change modeling in landscape studies. Markov analysis is a statistical tool using transition probability matrix based on neighborhood effects in a spatial influence algorithm (Cole, 2006). One inherent problem with Markov is that it provides no sense of geography. The transition probabilities may be accurate on per category basis, but there is no knowledge of the spatial distribution of occurrences within each land use category (Ye and Bai, 2008). To solve this problem, CA (Cellular Automata)-Markov chain was developed to add a spatial dimension to the model using cellular automata. A cellular automaton is an agent or object that has the ability to change its state based upon the application of a rule that relates the new state to its previous state and its neighbor (Eastman, 2009). In this research, we used CA-Markov chain to predict expected land use changes using IDRISI Andes (15.0) software.

2.3.3 Validation

For validation, the model's output is compared to a map of land use prediction with actual land use map of 2006. The relative operating characteristic (ROC) is an index of discrimination accuracy that can validate land use prediction maps independently of any specified quantity of land use change (Schneider and Pontius, 2001). ROC method evaluates the predicted probabilities by comparing them with the observed values over the whole domain of predicted probabilities instead of only evaluating the percentage of correctly classified observations at a fixed cut-off value (Verburg et al., 2002). ROC equals to 1 when a prediction map has perfect suitability, while ROC equals to 0.5 when a prediction map has random suitability values (Pontius et al., 2000).

With ROC method, we use Kappa index of agreement (KIA) that has several variations, each of which measures different characteristics of agreement (Pontius, 2000). Kappa equals 1 when agreement is perfect, while kappa equals 0 when agreement is as expected by chance (Pontius et al., 2000). In this research we use three Kappa index of agreement (Kno, Klocation, Kstandard). Using the combination of Kno, Klocation, and Kstandard for the evaluation allows for a determination of an overall success rate while providing an understanding of the factors that contribute to the strength or weakness of the results (Geri et al., 2011).

Table 1. Definitions of the Kappa index of agreement

Kno	Measure of the overall proportion correctly classified versus the expected proportion correctly classified.
Klocation	Measure of the spatial accuracy due to correct assignment of values.
Kstandard	The proportion assigned correctly versus the proportion that is correct by chance.

3. Results

3.1 Land Use Change Patterns

Temporal and spatial land use change patterns (1970’s~2000’s) of the Soyang River Basin have different patterns for each land use class. Figure 1 shows a temporal land use for major land use classes in the research area. Urban land uses have increased steadily and urban growth has been occurring in Chuncheon city and centers of counties near the river. Forest covers have increased under the influence of zoning of national protection areas. Agricultural lands have decreased steadily despite the expansion of high income earning high land farming.

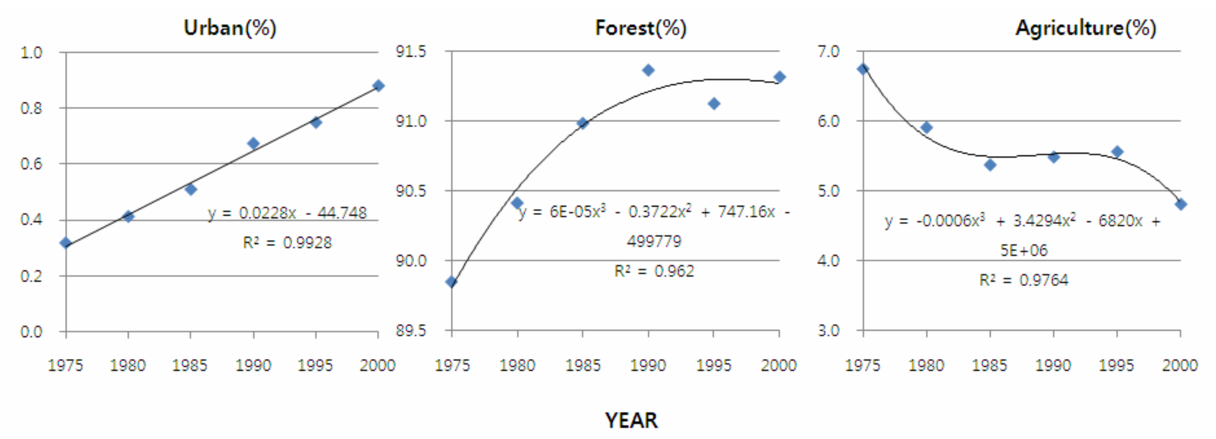


Figure 1. Temporal land use changes of major land use classes in the Soyang River Basin

3.2 Land Use Change Factors

Logistic regression parameters were estimated from the actual land use maps between 1995 and 2000 and used to predict land use changes of 2005. From the results of the logistic regression, we found that these variables for major land use classes had different explanatory powers and p-values. Most explanatory environmental variables that have affected the growth in urban land use are relief, which means urban growth has occurred in flat areas which are easy to convert to urban land use. However, the regression results show urban growth is more related to socio-economic factors than environmental factors which have low explanatory powers and no significances except relief. Forest growth is more influenced by natural-environmental factors rather than urban growth. Factors related to forest growth are relief, upslope area index and curvature that have negative correlations, while wetness index has a positive correlation. Forest growth mainly occurred in valley areas. Agricultural growth is more sensitive than other classes to variables that reflect shapes of geographical features such as slope.

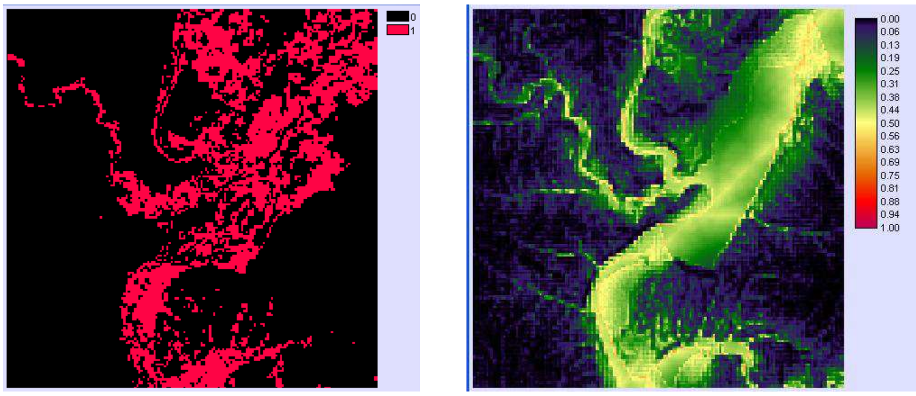
Among the artificial variables, distance variables have no explanatory power for determining land use change. The zoning of a reserve forest reflects policy variables, where major land use change is occurring in the non-reserve forest zones, especially urban land use growth. This demonstrates that zoning of reserve forests has an influence on land use change in the research area via general suppression of change. The results indicate that factors affecting land use change are different for each land use class. Table 2 shows the results of regression analysis of land use changes for major land use classes.

Table 2. Results of regression analysis of land use change variables for major land use classes

variables	Urban		Forest		Agriculture	
	B	p-value	B	p-value	B	p-value
Elevation	-.001	.006	-.002	.000	-.003	.000
Slope	-	-	-	-	-.016	.025
Aspect	.001	.011	.001	.000	.001	.000
Relief	-.103	.000	-.036	.000	-.026	.000
Wetness index	-	-	.035	.000	.020	.000
Upslope area	-	-	-.132	.000	-	-
curvature	-	-	-.016	.044	-	-
Distance_office	-	-	.000	.001	.000	.000
Distance_stream	.000	.007	-.001	.000	-.001	.000
Distance_road	.000	.000	.000	.000	.000	.000
Distance_urban	-.002	.000	.000	.000	.000	.000
Reserve forest[1]	.745	.000	.391	.000	.534	.000
Constant term	-2.670	.000	-.824	.000	-1.188	.000

[1] In the SPSS program, areas on zoning of reserve forest are coded 0 as a dummy coding.

Based on logistic regression, we made probability maps to predict major land use changes. The predicted land uses were then compared with actual land uses as illustrated in Figure 2. Areas that have high probability values for occupying agriculture class are in fact similar to actual observed locations of agricultural lands.



a) b)

Figure 2. (a) Observed agricultural land use 2007, and (b) probability of agricultural land use based on the logistic regression model

Figure 3 extends this illustration to include major land use class probability maps for Soyang River Basin. Urban growth is centered around Chuncheon city and the Soyang River, and agricultural growth is centered around the river and its streams, whereas the forest class indicates a reverse pattern. Although the comparison of predicted land use maps (a) to actual land use maps (b) shows that probabilities for urban and agricultural areas are overestimated, whereas forest is underestimated, we can neverthelss establish a predictability for land use.

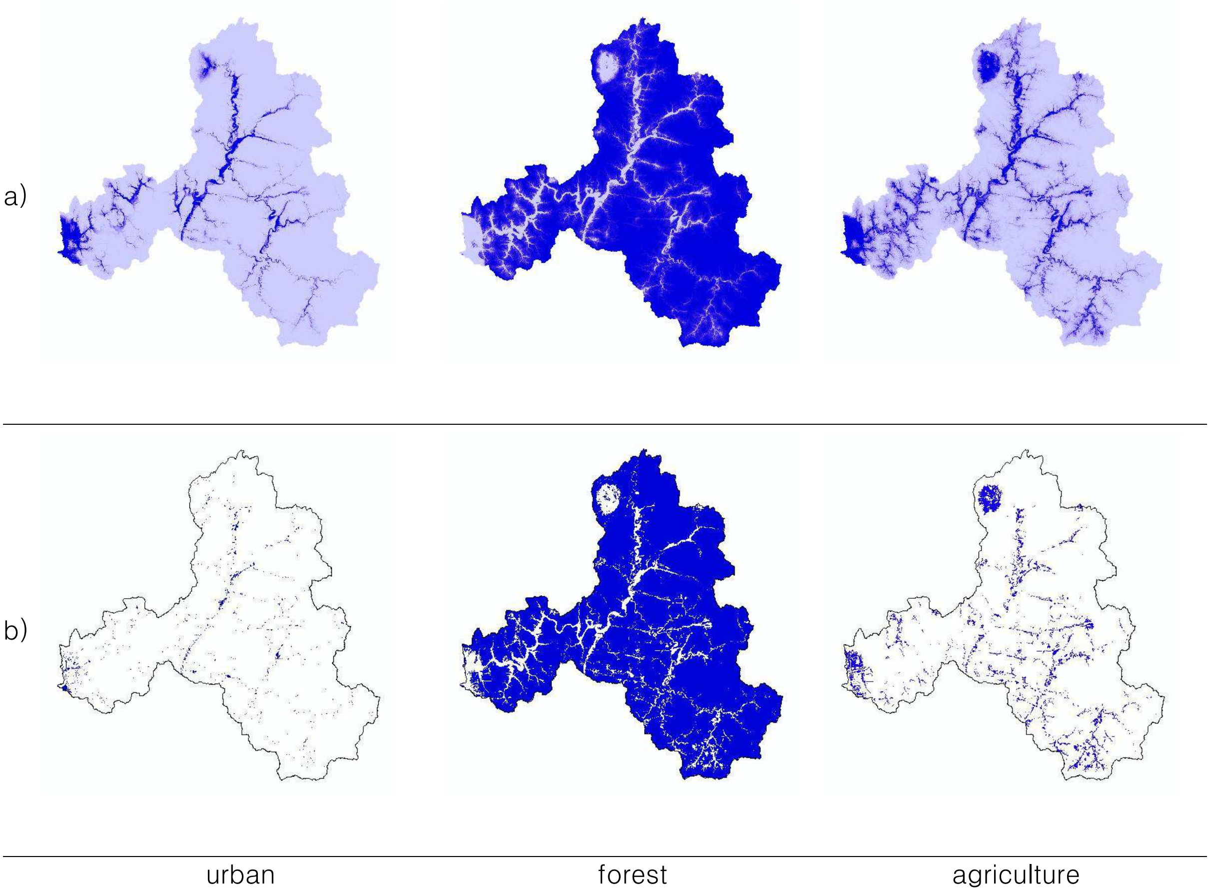
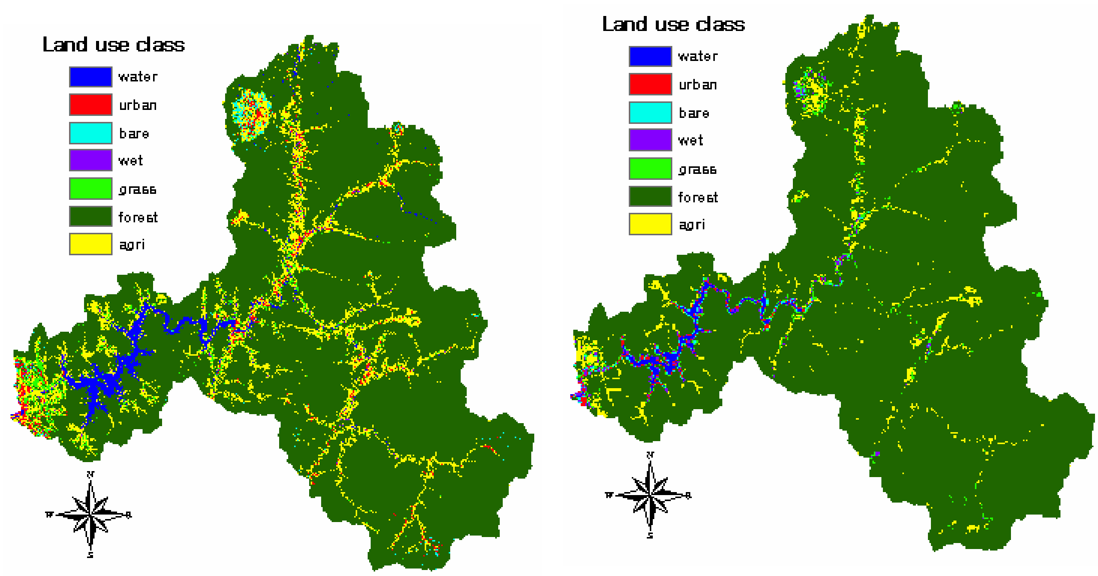


Figure 3. Comparison of (a) predicted land use map, with (b) actual land use maps for major land use classes

3.3 Land Use Prediction using CA-Markov Chain

We produced predicted land use change maps using probability maps from a logistic regression and CA-Markov chain to compare a land use prediction map with an actual land use map of 2006. In the prediction map of 2006, agriculture and urban class are overestimated, whereas forest is underestimated. These patterns are founded areas which are adjacent to the river.



a) b)

Figure 4. (a) predicted land use 2006 from CA_Markov model, and (b) land use map 2006

3.4 Validation

We compared the predicted land use maps from CA-Markov chain with the actual land use maps of 2006, by using relative operating characteristic (ROC) and variations of the Kappa index of agreements (KIA). Forest class has the highest accuracy on predicted land use maps based on ROC, while urban class shows the lowest accuracy.

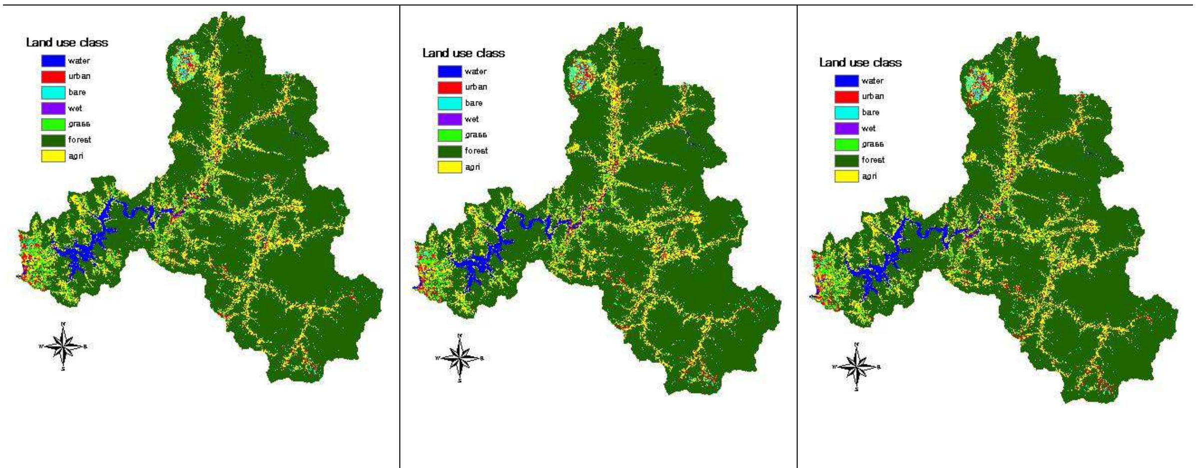
In the case of forest, ROC and Kappa Index both have high values because the whole extent of forest class is very large, although the extent of forest class adjacent to river networks is estimated less than the real extent. In the case of agriculture, the accuracy of the class has a low level at the aspect of spatial accuracy. Urban class has the lowest accuracy at the level of random distributions due to a low accuracy of spatial distributions, whereas the prediction of the extent is the most similar to the actual extent. It means scales of urban growth is appeared at the predictable levels, while it is hard to predict locations of urban growth, because urban growth is more related to socio-economic factors rather than natural-environmental factors used in this research, which make predicting urban growth difficult. Overall validation results are shown in Table 3.

Table 3. Validation results using ROC and Kappa index of major land use changes

	urban	forest	agriculture
ROC	0.515	0.890	0.643
Kno	0.9787	0.8669	0.8808
Klocation	0.2021	0.9520	0.4538
Kstandard	0.1678	0.8583	0.2622

3.5 Land Use Simulation

Although there are several limits for land use prediction in the process of validations, we predict future land use change simulations using the prediction model per every 5 years until 2020 year (Figure 5) and areas that can have the possibility of land use changes. The results of overlay analysis on prediction maps indicate that the urban class is predicted to increase in areas of former agricultural lands and the growth of the forest class is predicted to occurred primarily in small scale agriculture lands that are located far away from the river, which are causing a decline in the agriculture class.



2010	2015	2020

Figure 5. Predicted land use changes until 2020

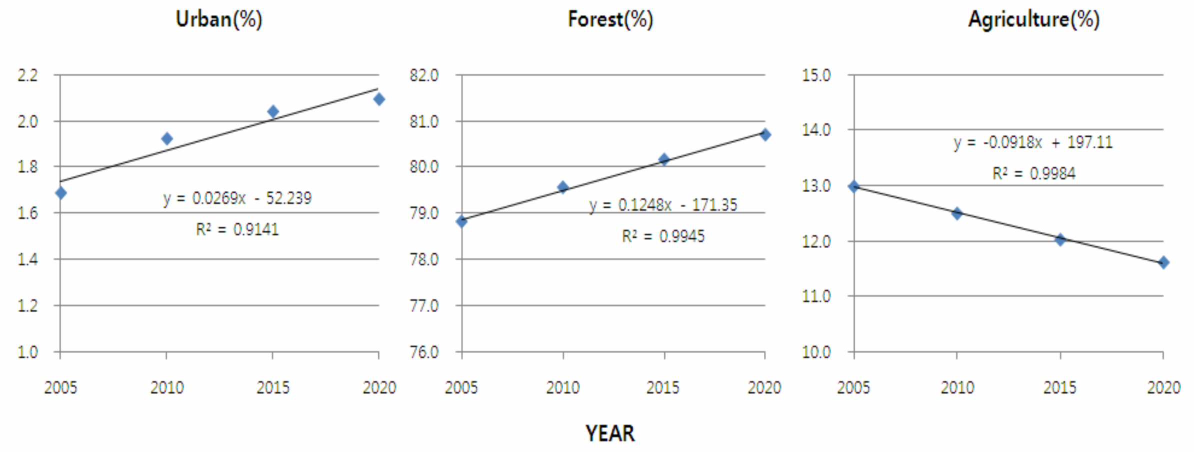


Figure 6. Temporal patterns of predicted land use change simulations for major land use classes

4. Summary

We established a land use change model using logistic regression and CA-Markov chain and compared our model with the actual land use map of 2007 to validate the accuracy of the model. Then, we predict future land use changes using the model to compare future land use change patterns with past land use change patterns.

Although the results of the land use simulation model have an explanatory accuracy compared with other land use simulation models based on the logistic regressions, one of the problems of our model is that its accuracy is based on the accuracy of forest class coverage which is approximately 90% of the research area. When we consider the accuracy of each land use class, only three land use classes have significant explanation for land use changes (water, forest, agriculture).

In the model, we use independent variables for land use changes that mainly reflect physical attributes. However, in the Soyang River Basin, artificial factors are an important factor influencing land use change, particularly in the case of urban class that is increasing steadily in the area. In order to develop the model more accurately, we have to consider artificial aspects by extracting values from the socio-economic and household survey data.

References

Kim, J, H, 2002, An analysis of land use change in the urban fringe using GIS and logistic regression in Korea, The Korea Spatial Planning Review, 33: 175-200.

Geri, F, Amici, V, and Rocchini, D, 2011, Spatially-based accuracy assessments of forestation prediction in a complex Mediterranean landscape, Applied Geography, 31, 881-890.

Lambin, E., F., Rounsevell, M., Geist, H., 2000, Are current agricultural land use models able to predict changes in land use intensity?, Agriculture, Ecosystems, & Environment, 82, 321-331.

Lee, Y, J, and Kim, S, J, 2007, A modified CA-Markov technique for prediction of future land use change, Journal of Korean Society of Civil Engineers, 27(6), 809-817.

Oh, Y, G, Choi, J, Y, Bae, S, J, Yoo, S, H, and Lee, S, H, 2010, A probability mapping for land cover change prediction using CLUE model, Journal of Korean Society of Rural Planning, 16(2), 47-55.

Pontius, R, G, 2000, Quantification error versus location error in comparison of categorical maps, Photogrammetric Engineering & Remote Sensing, 66(8), 1011-1016.

Pontius, R, G, Claessens, L, Hopkinson, C, Marzouk, A, Rastetter, E, B, Schneider, L, 2000, Scenarios of land use change and nitrogen release in the Ipswich watershed, Massachusett, USA, 4th International Conference on Integrating GIS and Environmental Modeling(GIS/EM4).

Veldkamp, A., and Lambin, E. F., Predicting land use change, Agriculture, Ecosystems and Environment, 85, 1-6.

Verburg, P, H, Soepboer, W, Veldkamp, A, Limpiada, R, Espaldon, V, and Mastura, S, S.A., 2002, Modeling the spatial dynamics of regional land use: The CLUE-S model, Environmental Management, 30(3), 391-405.

Ye, B, and Bai, Z, Simulating land use/cover changes of Nenjiang county based on CA-Markov model, Computer and Computing Technologies in Agriculture, 1:321-329.