Predicting Future State of Land cover by Accounting for Urban Activities and Local Demographic Indicators

Toshihiro Osaragi, and Takayuki Mori

Department of Mechanical and Environmental Informatics,
Graduate School of Information Science and Engineering,
Tokyo Institute of Technology
osaragi@mei.titech.ac.jp

INTRODUCTION

Remotely sensed images provide basic representations of land cover variations, and they are often useful aids for further analysis (Aplin 2004). For instance, land cover change analysis can be adopted for investigating the effects of human development (Herold et all. 2002, Steinmeier and Mueller 2004). Also, land cover information is often used for assisting management decisions of urban planners (Civco et al. 2002, Epstein et al. 2002). The most popular form of land cover analysis in the field of remote sensing is land cover classification (Briem et al. 2002, Franklin and Wulder 2002, O'Hara et al. 2003, Smith et al. 2003, Aplin 2004). On the other hand, monitoring of land cover changes is used in environmental management since knowledge of land cover dynamics can help indicate where natural resources require protection (Aplin 2004). A fairly straightforward means of monitoring land cover change is to compare images of a given area acquired at different times (Chen P et al. 2002, Aplin 2004).

Thus, remote sensing technology makes data available immediately, so that the current land cover can be quickly viewed (or monitored), however, with a limit in predicting future land cover. Even though there are some approaches to simulate land cover changes (Soares-Filho et al. 2002, Pontius et al. 2003, 2005), a method to predict future land cover is needed to extend the application of land cover analysis. This study proposes a statistical model for predicting future land cover. With the fact that land cover is closely connected to urban activities, such as land use and local demographic indicators, a model is constructed to employ a wide variety of information, including land use in particular (figure.1). Actual data are also employed to generate estimates of future values for land use and local demographic indicators, which are subsequently used to forecast future land cover.

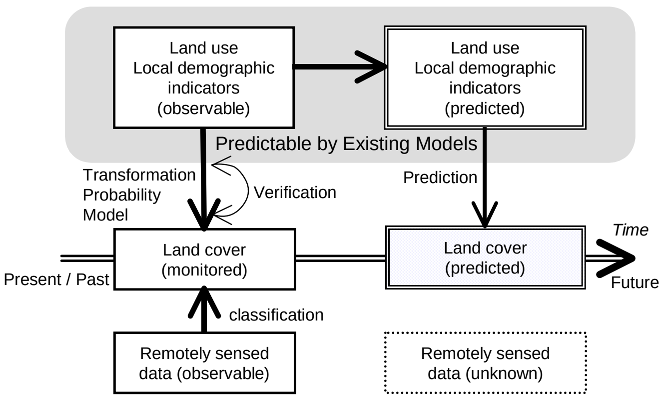


Figure 1: Method for predicting future land cover.

CONSTRUCTING A MODEL FOR ESTIMATING LAND COVER

Basic model for estimating land cover

Types of land cover in certain location are closely related to the uses of such locations – or zones. Accordingly, the first step in modeling land cover is to apply the equation given below, which transforms land use data directly to land cover data:

$$y_i(k) = \sum_{j=1}^{J} P_{ij} x_j(k)\,, \tag{1}$$

where $y_i(k)$: area of i-th land cover class in zone k
$x_j(k)$: area of j-th land use class in zone k
P_{ij}: transformation probability relating land use j with land cover i.

This is the basic model that expresses the area of any one kind of land cover $y_i(k)$ in a selected space (zone k) as the linear sum of areas of land use $x_j(k)$.

Formulation of model for estimating land cover

The basic model described above is very simple: it states that if the land usage situations in different locations are the same, then they will have the same state of land cover. Actually, though, differences in details of activities in different locations may result in large disparities in the state of land cover. For example, two locations may be used for residence, but if one contains rows of single-family dwellings with front and back yards, while the other contains high-rise multiple-family dwellings, their states of land cover would be far different. Thus, land cover depends not only on the land use class, but also on the activities particular in each location – or zone – and such zone's characteristic local demographic indicators. Let us substitute the non-uniform transformation probability matrix for P_{ij} for converting data about local land use to data about land cover. The elements of this matrix express some of the unique features of each zone. Thus, transformation probability P_{ij} for zone k is $P_{ij}(k)$ and the basic model is rewritten as:

$$y_i(k) = \sum_{j=1}^{J} P_{ij}(k) x_j(k)\,, \tag{2}$$

where $P_{ij}(k)$ is a transformation probability for land use and land cover in zone k.

Still, it is difficult to believe that transformation probabilities of all zones would differ to any great degree. Zones with similar local demographic indicators are likely to contain similar structures for $P_{ij}(k)$, and the likelihood of such assumptions also helps to keep the model "simple," which is a desirable attribute. Consequently, zones are classified into some groups on the basis of the characteristics of their local demographic indicators, denoted by l, and each group is assigned its own transformation probability P_{ijl} on the basis of similarities in the local demographic indicators. Here, it shall be noted that we consider groups of zones can differ according to land use class. The reason for this is as follows: the local demographic indicators which determine land cover in residential districts can differ from those local demographic indicators which determine land cover in commercial districts. This reasoning suggests that Eq.(2) should be re-written. The set of zones designated for land use j and local demographic indicator l are expressed as $G_j(l)$ (l=1,…, L_j), and the members of this group $G_j(l)$ share a transformation probability P_{ijl}:

$$P_{ij}(k) = \sum_{l=1}^{L_j} P_{ijl} z_{jl}(k) \quad \textit{for all i, j, k}\,, \tag{3}$$

where P_{ijl}: transformation probability for designated land use j and local demographic indicator l.

$z_{jl}(k)$: indicator variable (l = 1, …, L_j) expressing the local demographic indicators connected with land use j for each zone k,

where $z_{jl}(k) = \begin{cases} 1 & : \quad k \in G_j(l) \\ 0 & : \quad k \notin G_j(l) \end{cases}$ *for all j, l.*

The purpose of the indicator variable $z_{jl}(k)$ here is to state which group each zone k belongs to for each land use classification j.

Method for estimating probability of transformation P_{ijl}

First, the number of locations in zone k with land use j and land cover i is written as $n_{ij}(k)$. Next, it is assumed that the data $n_{ij}(k)$ are observed in the stochastic process based on transformation probability P_{ijl}. The simultaneous probability is assumed to be a function solely of P_{ijl}, allowing us to define the following likelihood function:

$$L = \prod_i \prod_j \prod_k \left(\sum_l P_{ijl} z_{jl}(k)\right)^{n_{ij}(k)} . \tag{4}$$

We can estimate the values of transformation probability P_{ijl} by maximizing the value of L for each land use j. Thus we can obtain the following maximum likelihood estimator $\hat{P}_{ijl}$:

$$\hat{P}_{ijl} = \frac{\sum_k (n_{ij}(k) \sum_l z_{jl}(k))}{\sum_i \sum_k (n_{ij}(k) \sum_l z_{jl}(k))} . \tag{5}$$

The indicator variable $z_{jl}(k)$ has not yet been determined, so Eq.(5) does not yet give a maximum likelihood estimator $\hat{P}_{ijl}$. The determination of $z_{jl}(k)$ is deferred to a later discussion.

It is obvious that the more groups each zone is classified into, the greater the value of the likelihood function. However, classifying zones into a large number of groups results in a complicated model. Therefore, we employ the Akaike Information Criterion (AIC) to assess the fitness and simplicity of the model simultaneously. A model, which shows a lower value of the AIC, can be considered of higher quality. The logarithmic likelihoods to each land use class are mutually independent, so each class can be assessed using the following expression:

$$\mathrm{AIC}_j = -\sum_i \sum_k n_{ij}(k) \ln \sum_{l=1}^{L_j} \hat{P}_{ijl} z_{jl}(k) + 2(I-1)L_j , \tag{6}$$

where $\hat{P}_{ijl}$ is a maximum likelihood estimator, I is the number of classes of land cover, and L_j is the number of local demographic indicators.

Setting the values of indicator variable $z_{jl}(k)$

The value of $z_{jl}(k)$ should be set so that it minimizes AIC_j in Eq.(6). Indicator variable $z_{jl}(k)$ can take on a very large number of values, however, and it would be impossible to compile a list of all these values within any practical calculation time. Therefore, a Neural Network is employed to perform an efficient search for the value of $z_{jl}(k)$. Aoki et al. (1993, 1996) proposed a method to divide a total region into more appropriate sub-areas using a simulated Neural Network, and reported that land use prediction error could be minimized whenever the area-dividing method was used.

In this portion of the analysis, the indicator variable $z_{jl}(k)$ corresponds to a neuron of the network, and AIC corresponds to a potential energy of the network. The states of neurons (active/inactive) are varied so as to reduce the potential energy. The states of neurons when the Neural Network reaches equilibrium conditions are taken as values for the indicator variable $z_{jl}(k)$ that provides the lowest value of AIC.

ANALYSIS OF RELATION BETWEEN LOCAL DEMOGRAPHIC INDICATORS AND LAND COVER

Outline of analyzed data

We use the land cover estimation model described in the previous section with the data observed in the region of Metro Manila, the capital of the Philippines. Figure 2 and tables 1 – 3 show the region employed for this study and the details of the data.

Greater Manila region, the Philippines

The basic unit for this analysis is the zone (shown in white on map), which is the spatial unit in the person-trip survey. All the data collected for land cover, land use and activities from all 233 zones are examined for this analysis.

Figure 2: Study area.

Table 1: Classification of land cover data.

<Land cover classes>	<Remarks>
Paved	Represents any pavements in urban region
Buildings (dense)	Low-rise buildings in dense concentration
Buildings (sparse)	Areas with green coverage and building-to-land ratio that are clearly different from those in Buildings (dense)
Soil B (no vegetation)	Little vegetation; under development, so soil is dry
Soil A (vegetation)	Far more likely to have grass than land under development
Lawn	Much vegetation
Aquatic plants	Plants growing densely in rivers or lake shores
Forest C (sparse veg.)	Forest with sparser vegetation than in Forest A and Forest B
Forest B	Forest midway between Forest A and Forest C
Forest A (copious veg.)	Untouched forest with copious vegetation

Table 2: Classification of land use data.

<Land use classes>	<Land use categories in NCTS*>
Residence	Residential
Industry	Commercial/Business
	Industrial
Agriculture	Agricultural
Forest	Forest
Vacant	Open
	Other
Grass	Grassland
	Water Related
Park	Park Recreational
Public	Government Quasi Public
	Educational Cultural
	Sports Athletics
	Health Welfare
	Religious Cemetery
	Transport Service Facility
	Military
Riverside	River
Road	Road

* NCTS: National Center for Transportation Studies, University of the Philippines

Table 3: Local demographic indicators.

<Demographic Features >	<Indicators>
Accessibility	Distance from main road
Population	Population density
	Labor population fraction
	Housewife population fraction
	Unemployment rate
	Average household size
Industrial	Labor population fraction in primary industries
	Labor population fraction in secondary industries
	Labor population fraction in tertiary industries
Income	Fraction of population with low income
	Fraction of population with medium income
	Fraction of population with high income
Ownership	Fraction of population owning their own home
	Fraction of population owning land
	Fraction of population owning car(s)
	Average number of cars per household

Extracting estimates for the model and local demographic indicators

In the case study described here, the data for land use and land cover are transformed into raster data with a consistent format; then, these are employed as observation data. For convenience, the location size, or zone size, is set by the resolution of the LandSat thematic mapper (TM) data, about 30 m × 30 m. The chief purpose of this portion of the study is to investigate the validity of all the steps of the method described above. The number of groups is fixed, and the method for defining groups is discussed. The zones are grouped using local demographic indicators. In this scheme, the following four standards must be considered:

(1) Which local demographic indicator is appropriate?

(2) How should the indicator be classified?

(3) How many local demographic indicators should be used?

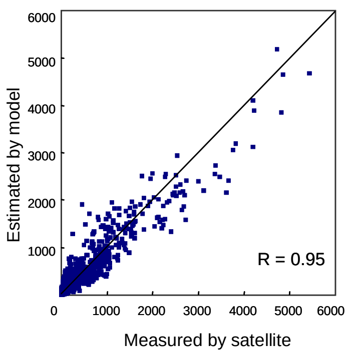
(4) How should local demographic indicators be combined?

If all these items are to be considered simultaneously in all possible combinations, it would consume excessive calculation time to identify the optimal combinations. Therefore, from the viewpoint of (3) above, we consider just two local demographic indicators, and from the viewpoint of (2), we establish three classes for each local demographic indicator. Each locality is thus assigned into one of nine (3 x 3) groups, and the transformation probability is extracted. Specifically, all combinations of land use classes with local demographic indicators are examined, and the combination, which has the lowest value for AIC, is used.

Reliability of model and local demographic indicators selected

Figure 3 presents the results of a verification of reliability of the model, which was provided at the end of the search for the optimal combinations of land use classes and local demographic indicators. The estimated area $\hat{y}_i(k)$ of land cover i shows a correlation of 0.95 with the actual measured area $y_i(k)$, indicating high reliability of the model.

Table 4 lists the local demographic indicators used in this model. The population density and labor population fraction are the indices with the highest explanatory power in the commercial/business and industrial districts, i.e., commercial/urban land. Meanwhile, in the non-commercial land-use classifications of residential, road and governmental quasi-public, activity indicators such as accessibility (average distance in pixels from the nearest main roadway), structure of employment (fraction of the employed who are engaged in primary industries), and income structure (fraction of population with medium income) relate closely to land cover. Another close connection is found between the differences in local demographic indicators such as industry (fraction of the employed who are engaged in primary industries), income structure and ownership structure (fraction of households owning their home) in natural areas used for agriculture and forests.



Land use data
NCTS data (polygon data) were re-classified from their 18 original categories into 10 categories for display in the table.

Land cover data
The data set were taken from Landsat TM data (taken May 15, 1997) and classified into 10 land cover categories.
Each pixel from the land cover data and land use data on the 30 m grid was classified and zones of collected pixels were employed for the analysis.

Local demographic indicators
The indicators were created using the collected data for each zone from the person trip data set.

Figure 3: Reliability of the model.

Table 4: Selection of local demographic indicators.

Land use categories / Local demographic indicators		Commercial/ urban		Non-commercial			Natural				
		Industry	Riverside	Residence	Road	Public	Grass	Park	Agric	Forest	Vacant
Accessibility	Distance from main roadway			×	×	×					
Population	Population density	×	×								
	Labor population fraction	×	×								
	Housewife population fraction						×				
	Unemployment rate										
	Average household size										
Industry	Labor population in primary ind.			×	×			×	×	×	×
	Labor population in secondary ind.										
	Labor population in tertiary ind.							×			
Income level	Pop. fraction with low income								×		
	Pop. fraction with medium income					×	×			×	
	Pop. fraction with high income										
Ownership	Pop. fraction owning house										×
	Pop. fraction owning land										
	Pop. fraction owning car(s)										
	Average number of cars per household										

Local demographic indicators and their connection with land cover

Transformation probabilities adjusted for the unique local demographic indicators of each group are graphed and the relation between local demographic indicators and land cover is observed in detail.

Residential area: The product (average distance in pixels from the nearest main roadway) × (fraction of the employed who are engaged in primary industries) is selected to assess the cases where land is designated for residential use (figure 4). When a distance from the nearest main roadway is far and a fraction of employment in primary industries is high (Group 3), there is a high probability for the land cover to be classified into Buildings (sparse); conversely, locations where the opposite trends (Group 1) are found to have a high probability for the land cover to be classified into Buildings (dense). Geographically, Group 1 consists of areas that are comparatively developed, and Group 3 zones are in the periphery of Group 1. In summary, even when two zones have the same classification as Residential, inhabitants engaging in differing local demographic indicators affect the probability of whether the zones' land cover will be Buildings (dense) or Buildings (sparse).

Agricultural land: The product (fraction of the employed who are engaged in primary industries) × (labor population fraction) is selected to assess the cases where land is designated for agricultural use (figure 5). Group 2 designates areas where both the fraction of people working in primary industries and the fraction of the population working as laborers are high. Here, Soil A (vegetation) shows a somewhat high fraction, assumingly for its use of cultivation activities. In Group 3, however, there is a low fraction of laborers engaging in primary industries, though the fraction of laborers is high; this indicates large numbers of laborers in secondary and tertiary industries. This group has a lower probability than the other groups to contain the classification Buildings (sparse) but a higher probability for Lawn. The main observed difference from Group 2 is the low count of agricultural activities.

Riverside: The product (population density) × (labor population fraction) is selected to assess the cases where land is designated for riverside use (figure 6). Locales with the combination of low population densities and high population fractions of labor (Group 3) have a high probability of land cover with soil or natural areas. Still, it is also highly possible that the locale has a dense population and is paved. This indicates that the neighborhoods of rivers tend to be either well-developed regions covered with artificial coverings or relatively undeveloped regions in their natural state.

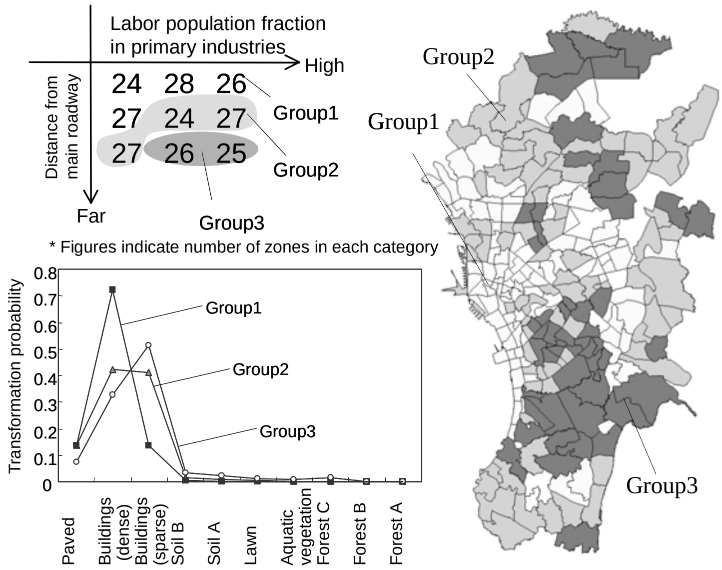


Figure 4: Local demographic indicators for residential.

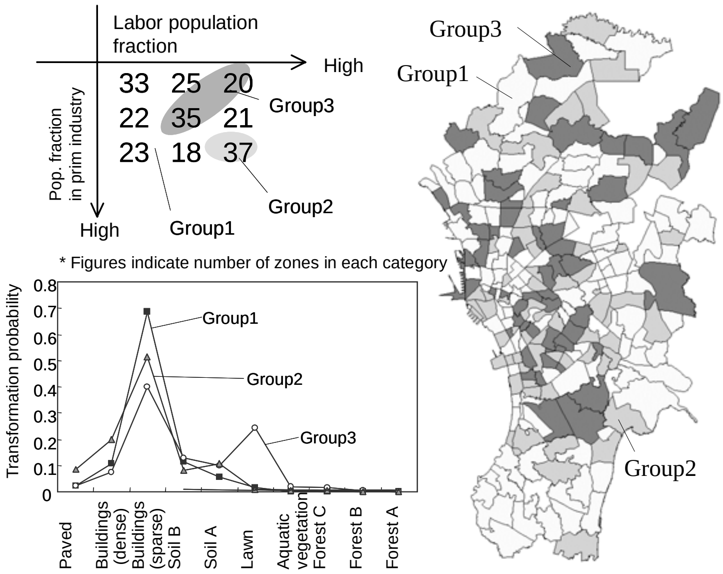


Figure 5: Local demographic indicators for agricultural land.

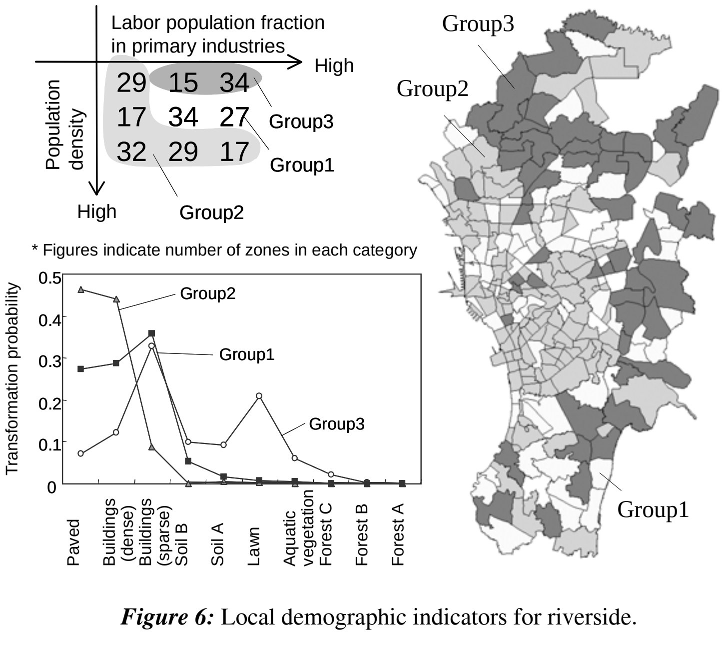


Figure 6: Local demographic indicators for riverside.

PREDICTION OF FEATURE LAND COVER

Method for predictions

Forecasts of land use and local demographic indicators are employed to simulate the land cover in the greater Manila region in the same way as above. The details of the data are provided in figure 7. Two kinds of simulations were prepared:

(1) Simulations employing forecasts of land use and local demographic indicators.

(2) Simulations employing only forecasts of land use (to examine the influence of changes in land use only, and also to provide a counter-reference for examining the influence of local demographic indicators).

Land use data
Predictions of land use in the year 2015 were used (Doi and Kurokawa, 2001). Data were transformed to 30 m grid, collected into zone units.

Local demographic indicators
(1) Population, number of households
A logistic curve was fitted to data that had been obtained for 1980, 1990, and 1995, and the data were gathered in zone units.
(2) Number of households by income class
The obtained forecasts for land use in 2015 were classified into 3 categories; residences were classified into low-, medium- and high-income residences. The area of each category was summed into zone totals and the respective area fraction was multiplied by the number of residences found in (1).

Local demographic indicators		Data on conditions in 2015
Accessibility	Distance from main roadway	Constant
Population	Population density	Estimated from population (1)
	Labor population fraction	Constant
	Housewife population fraction	
	Unemployment rate	Estimated from population and number of households (1)
	Average household size	
Industry	Labor pop. in primary ind.	Constant
	Labor pop. in secondary ind.	
	Labor pop. in tertiary ind.	
Income level	Pop. fraction w/ low income	Estimated according to (2)
	Pop. fraction w/ medium income	
	Pop. fraction w/ high income	
Ownership	Pop. fraction owning house	Constant
	Pop. fraction owning land	
	Pop. fraction owning car(s)	
	Av. num. of cars per household	

Figure 7: Relations between local demographic indicators affecting land cover.

Simulation results and observations

First, we discuss the changes in land cover due to land use and local demographic indicators. Predicting the land cover in the greater Manila region in 2015 was attempted, using the predicted data for land use and local demographic indicators; figure 8 shows the results. There is a tendency toward an increase in artificial land covers such as Paved and Buildings, and a general decrease in natural covers such as Soil, Lawn and Forest. Locations classified as Buildings (dense), of which there were many in the city center, will decrease while showing increase in the outskirts of the city. On the contrary, locations classified as Buildings (sparse), of which there are now many in the outskirts of Manila, will decrease in the city center as they decrease in the periphery. The growing number of locations classified Paved and Buildings (dense) reflect the construction of residences in the outskirts of the region as the process of urban sprawl continues. In contrast, some of the locations in downtown Manila classified as Buildings (dense) will revert to Buildings (sparse) and the number of locations classified as Soil B (no vegetation) will increase slightly. There will be some laudable advances in land use in downtown Manila, but insufficient attention to creating a green environment may be a concern.

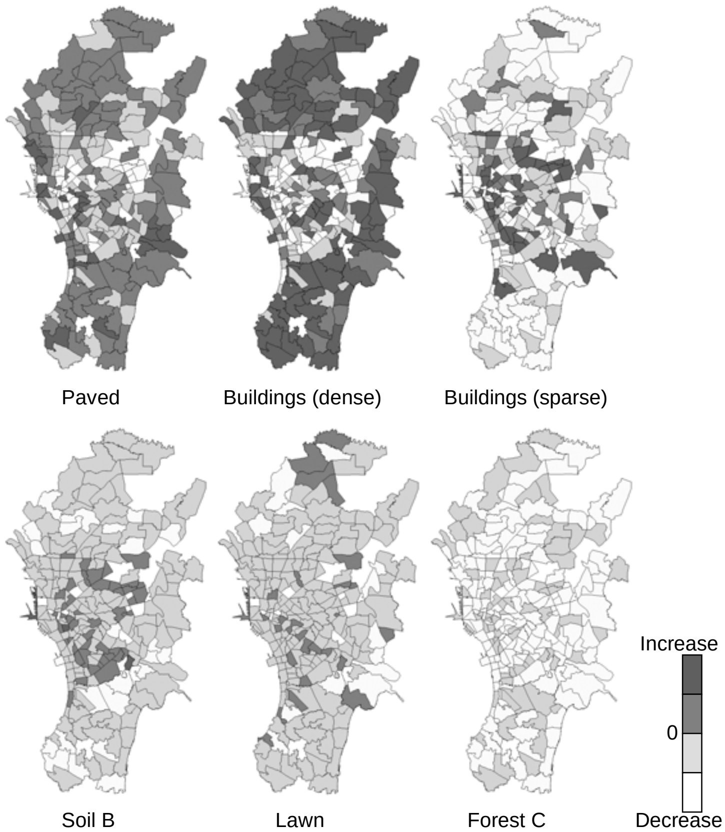


Figure 8: Predictions based on changes in both land use and local demographic indicators.

Next, we discuss the changes in land cover due only to land use. In this portion of the study, the local demographic indicators are assumed to remain constant and only the predictions of land use data in 2015 are employed to forecast the land cover. Those results are shown in figure 9. Comparing these with the results described in the previous section (figure 8), the locations classified into artificial cover such as Paved or Buildings generally show identical trends, but larger changes are predicted by the model that takes the local demographic indicators into account. In other words, changes in land cover that cannot be explained just by changes in land use are explained by consideration of local demographic indicators.

Comparing the results for Soil and Lawn between these models, a large number of locations are found where the predictions show increases in one but decreases in the other. Specifically, when only changes in land use are considered, Soil and Lawn are predicted to decrease, but when changes in local demographic indicators are also taken into account, some zones show tendencies to increase. This discrepancy is particularly noticeable in the city center. These results show that future intensive use of land in downtown cannot be well anticipated by examining only past changes in land use. Turning to a comparison of Forest locations, the prediction based only on land use changes brings somewhat more changes in these locations than does the other model. This indicates that predictions based partially on local demographic indicators tend to underestimate the effect of land use on land cover. In other words, even when there are changes in land use in a location covered with Forest, these changes are not significant, from the viewpoint of local demographic indicators. Thus, changes in land use do not necessarily bring changes in land cover.

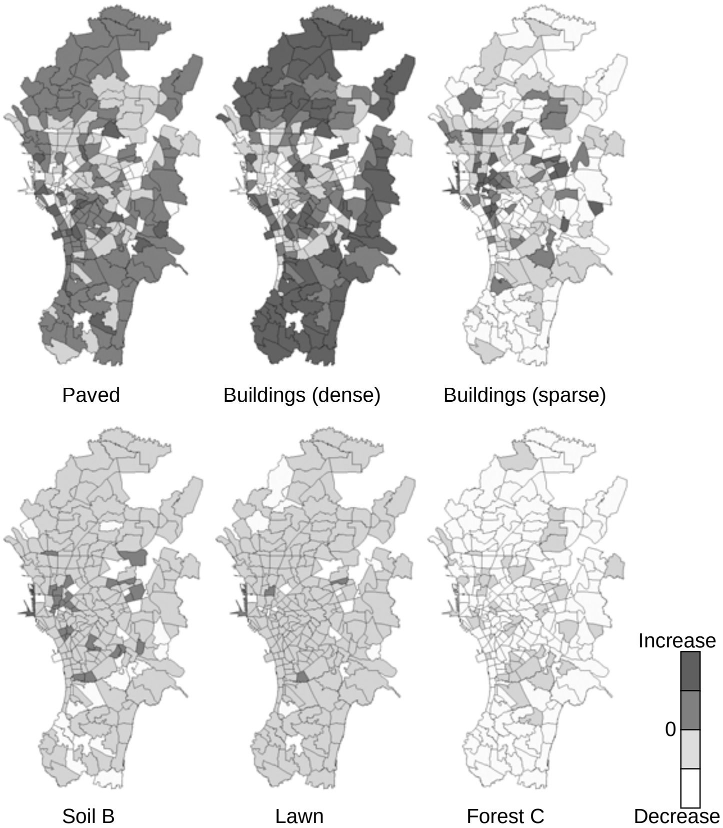


Figure 9: Predictions based only on changes in land use.

SUMMARY AND CONCLUSIONS

This study presents the construction of a model for estimating land cover using data for land use and for local demographic indicators. The proposed model is checked against actual land cover data and is verified as reliable. Then, a simulation of land cover in the future is conducted, using a forecast of future data on land use and local demographic indicators. It is shown that the combination of the factors of land use and local demographic indicators is quite significant for predicting land cover.

It is expected that there will be growing demands to predict and evaluate environmental conditions such as green coverage, thermal environment, etc., in future urban environments. The use of the land cover prediction model proposed in this paper may provide basic data useful in this research field.

Due to the limitation of available data, the validation of predicted states of future land cover was unable to be produced. In further study, we would be able to confirm the accuracy of our proposed model, by predicting the states of land covers, which can be verified from the historical data accessible at such point of time.

Acknowledgements

The authors would like to acknowledge the variable comments from anonymous referees. This work has been funded by "Japan Society for the Promotion of Science" as part of the project, "Impact Analysis of Metropolitan Policies for Development & Conservation in the Philippines."

BIBLIOGRAPHY

Aoki Y, Osaragi T, Nagai A: Area Dividing Method by Neural Network. Third International Conference on Computers in Urban Planning and Urban Management 2: 379-392, 1993.

Aoki Y, Osaragi T, Nagai A: Use of the area-dividing method to minimize expected error in land-use forecasts. Environment and Planning B, Planning and Design 23:655-666, 1996.

Aplin P: Remote sensing: land cover. Progress in Physical Geography 28-2:283-293, 2004.

Brien G J, Benediktsson J A, Sveinsson J R: Multiple classifiers applied to multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing 40-10:2291-2299, 2002.

Chen P, Lu X, Liew S C, Kwoh L K: Quantification of land cover change and its impact on hydro-geomorphic processes in the upper Yangtze using multi-temporal Landsat imagery: an example of the Minjiang area. Proceedings international geoscience and remote sensing symposium (IGARSS) 2:1216-1218,2002.

Civco D L, Hurd J D, Wilson E H, Arnold C L, Prisloe J M P: Quantifying and Describing Urbanizing Landscapes in the Northeast United States. Photogrammetric Engineering & Remote Sensing 68-10:1083-1090, 2002.

Doi K, Kurokawa K: Integrated design of LUCC policies and environmental simulation in metro manila. Proc. of LUCC Symposium 1:11,2001.

Epstein J, Payne K, Kramer E: Techniques for mapping suburban sprawl. Photogrammetric Engineering & Remote Sensing 63-9:913-918. 2002.

Franklin S E, Wulder M A: Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. Progress in Physical Geography 26-2:173-205, 2002.

Herold M, Scepan J, Clarke K C: The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environment and Planning A 34:1443-1458, 2002.

O'Hara C G, King, J S, Cartwright J H, King R L: Multitemporal land use and land cover classification of urbanized areas within sensitive coastal environments. Transactions on Geoscience & Remote Sensing 41-9-1:2005-2014, 2003.

Pontius R G J, Agrawal A, Huffaker D: Estimating the uncertainty of land-cover extrapolations while constructing a raster map from tabular data, Journal of Geographical Systems 5:253-273, 2003.

Pontius R G J, Spencer J: Uncertainty in extrapolations of predictive land-change models, Environment and Planning B: Planning and Design 32-2:211-230, 2005.

Smith J H, Stehman S V, Wickham J D, Yang L: Effects of landscape characteristics on land cover class accuracy. Remote Sensing of Environment 84:342-349, 2003.

Soares-Filho B S, Cerqueira G C, Pennachin C L: DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecological Modelling 154:217-235, 2002.

Steinmeier C, Mueller K: A monitoring concept for detection of land cover change along main traffic routes in Switzerland. Proceedings international geoscience and remote sensing symposium (IGARSS) 5:3409-3412, 2004.