Modeling Land Conversion in the Colombo Metropolitan Area Using Cellular Automata

Prasanna Divigalpitiya*1, Akira Ohgai2, Takeru Tani1, Kojiro Watanabe4 and Yoshimizu Gohnai5

1 Graduate Student, Department of Environment & Life Science, Toyohashi University of Technology, Japan
2 Professor, Department of Architecture & Civil Engineering, Toyohashi University of Technology, Japan
3 Assistant Professor, Department of Architecture & Civil Engineering, Toyohashi University of Technology, Japan
4 Assistant Professor, Department of Ecosystem Design, Institute of Technology and Science, University of Tokushima, Japan
5 Researcher, Research Center for Future Technology, Toyohashi University of Technology, Japan

Abstract
This paper proposes a Cellular Automata (CA) model to evaluate the urbanization patterns arising from the regulation of urban growth on paddy lands in the Colombo Metropolitan Region (CMR). Most of the historic map data available for the CMR before 1990 are temporally sporadic and spatially incomplete. As an alternative to maps, classified remote sensing data are used to analyze the urbanization process. Logistic regression is applied to derive factors of urbanization and the various relationships among them. The relation between 'urban' and 'non-urban' serves as an explanatory variable. The factors explaining that relationship are calculated by exploratory logistic regression analyses. The probability calculated from the statistical model is used for CA transition with a random number. Several growth patterns are simulated based on a range of transition thresholds to test the CA model. Status quo growth and several growth control scenarios are simulated for the period from 1987 to 2002 based on an optimum threshold. The simulation result of the status quo growth is evaluated with several evaluation methods. The level of agreement between the estimated result from the status quo model and the actual data is 62%, while the multi-scale goodness-of-fit method produces highly accurate values for the given range of resolutions.

Keywords: cellular automata; Colombo Metropolitan Area; developing countries; logistic regression

1. Introduction
The environmental impact of urban problems has been intensifying and spreading over wider areas, posing serious threats to many developing countries in Asia. The impacts of urbanization on biodiversity and environmental quality are profoundly negative (Pickett et al., 2001). Habitats and terrestrial aquatic ecosystems are lost and fragmented in cities as urbanization advances, and biodiversity continues to wane (McDonnell and Pickett, 1990). The deteriorating ecological conditions affect air quality, rainwater drainage, and urban services such as recreational activities.

The loss of agricultural lands due to urban encroachment is one of the more complex problems of rapid urbanization. The agricultural lands in urban areas and their peripheries form part of the urban ecological system and do much to encourage the greening and cleaning of the city. They turn open spaces into green zones, buffer zones, and reserve zones for future development, while improving the micro-climate within the city.

Urban agricultural lands can be improved by regulating land subdivision and taxation policies. But when growth is restricted in one location, a city will grow in a different location. Planning regulations and an ever growing urban population create a complex relationship between environmental preservation and the form of a city as it develops.

A possible starting point for understanding these problems is to analyze the urban growth patterns and to project their spatiotemporal dynamics. Cellular automata (CA) models are excellent depicters of dynamic behavior. Nowadays they are widely used for modeling urban growth and simulating metropolitan scale land-use dynamics.

Many CA models assume that development units are correlated in space, and they apply sets of rules to reflect the correlations (Batty & Xia, 1997; Clark et al., 1997). These rules are used to calculate changes in land use, and the modeled results can be evaluated by comparison with actual urban growth.

*Contact Author: Prasanna Divigalpitiya, Graduate Student, Dept. of Environment & Life Science, Toyohashi University of Technology, Hibarigaoka 1-1, Tempakuchou, Toyohashi, 441-8580 Japan
Tel: +81-532-44-1178 Fax: +81-532-44-6831
E-mail: prasanna@urban.tutrp.tut.ac.jp
(Received April 9, 2007, accepted September 6, 2007)
CA modeling applies a range of alternative rules such as deterministic rules (White & Engelen, 1997), self modification rules (Clarke et al., 1997), stochastic rules (White et al., 1997), and utility maximization rules (Wu, 1998a).

In this research we develop a CA model to evaluate the urbanization patterns arising from the regulation of urban growth on paddy lands in the Colombo Metropolitan Region (CMR) of Sri Lanka.

CA models rely on historical observations for the calculation of parameter values. This poses a challenge with our topic, as most of the historic map data on the CMR before 1990 are temporally sporadic and spatially incomplete. The labeling of the land-use classifications can be especially erroneous, as the definitions have been repeatedly changed. Recognizing that map data can lead us into error in modeling the historic urbanization, we apply a set of classified remote sensing data to derive land-use data on the CMR.

Empirical methods for analyzing data provide more interpretable modeling output. Logistic regression has been used to interpret spatial data in many of the earlier studies (Wu and Yeh 1997; Cheng and Masser 2003; Hu and Lo 2007). By interpreting statistical models we can gain better knowledge on underlying spatial patterns.

This study uses logistic regression to calculate the relationship of urbanization factors with land-use data derived from classified remote sensing images. The probability values thus obtained are applied in a CA model, and the model is calibrated by comparing the resulting simulation against a time series GIS data layer. The study has three specific objectives:

- Demonstrate a framework for a CA urban growth model capable of simulating the urban growth process of the CMR.
- Develop a method to calculate transition probabilities using logistic regression.
- Evaluate the ability of the CA model to simulate the growth patterns that form in response to the regulation of paddy lands.

2. Study Area

Colombo, the largest city in Sri Lanka, has expanded rapidly and continuously since the local economy was opened in the early 1980s. Lacking any planned framework, the city has succumbed to various problems such as urban sprawl, unmanaged ribbon developments along the main trunk roads, and fragmentation in the peripheral urban areas. And with the progressive decentralization of activities in Colombo, the suburbs are growing faster than the central areas (Deheragoda et al., 1992).

To overcome the functional and spatial shortcomings of the existing planning units in the metropolitan region, the Urban Development Authority of Sri Lanka demarcated a new planning unit by including surrounding local authorities in 1998. This unit, the 'Core Area' of the CMR, covers about 16,000 ha. In 1998, the Urban Development Authority of Sri Lanka proposed a Core Area plan as part of the Colombo Metropolitan Region Structure Plan to regulate urban development of the Core Area.

The environmental portions of the Core Area plan include various programs to regulate wetlands, water bodies, seafront areas, hillocks, agricultural lands, and nature reserves. The subdivision and filling of paddy lands is an important issue in the whole of the CMR. Paddy lands are introduced as a major land-use function and commercial activity in the Core Area plan, as a means of discouraging filling and subdivision.

In this research we analyze an extended area encompassing the Core Area and surrounding CMR (143,438 ha; see Fig.1.) in an attempt to elucidate how the urbanization patterns relate to the growth regulations on paddy lands.

3. Model Structure

3.1 Model mechanisms

The CA approach can be used to simulate land-use changes by imposing a set of rules. Wu F. (Wu, 1998a) proposed a generic cellular model where transition rules are based on the utility maximization approach.

\[ S^t = (s^t, w^t) \]  

Fig.1. Study Area
where $s'$ is a cell state, $s_{ij}^t$ is a binary variable (whether the cell is selected or not), and $w$ is the quantitative variable indicating accumulative urban activities in the location.

The binary variable in this generic model is given by

$$s_{ij}^t = \begin{cases} 1, & \text{if cell } ij \text{ is selected,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

A quantitative variable is derived from the initial population logit function gives the conditional probability $P$ of transformation from one state to another.

$$P_{ij} = \frac{\exp(U_{ij})}{\sum_x \exp(U_{x,ij})} \quad (3)$$

The utility value $U_{ij}$ can be expressed by the linear relationship shown below.

$$U_{ij} = \alpha + \beta w + \epsilon \quad (4)$$

where $\alpha$ is an unknown constant, $\beta$ is a vector of parameter coefficients, $w$ is a vector of quantitative variables, and $\epsilon$ is an error variable.

The above relationship is advantageous, in that it can be used as a training method with binomial responses such as urban or non-urban. Having only two cell states in the model proposed in this paper, we can express the probability of transition by:

$$P_{ij} = \frac{1}{1 + \exp(-U_{ij})} \quad (5)$$

The independent variables are identified through repeated experiments with logistic regression and estimation of model sensitivity to initial urbanization data. Parameter coefficients are estimated by the maximum likelihood method. The only change in the cell state in the CA model is from non-urban to urban. The state of a cell at $t+1$ is given by:

$$s_{ij}^{t+1} = P_{ij} = \frac{1}{1 + \exp(-U_{ij})} \quad (6)$$

where $s_{ij}^{t+1}$ is the state of the cell at time $t+1$, $s_{ij}^t$ is the cell state at time $t$, $P_{ij}$ is the probability derived from utility value $U_{ij}$ using Equation 5, and $z$ is a random variable between 0 and 1. During the simulation, the model calculates the probability value for each cell. The model generates a 100m x 100m cell area is set as the unit of transformation in the model. A cell of this size is thought to have a resolution fine enough to capture the details of the urban fabric. The availability of basic urban services within a walking distance is assumed to be an important factor of urbanization, and a 5 x 5 cell neighborhood is used. The interactions of neighboring cells are considered in the calculations of parameters and utility values.

### 3.2 Model implementation

We implement the model on a UNIX workstation. Parameter values and boundary conditions (initial urban extent, growth rate, and simulation period) are hard coded in the initial model. Raster grid data required by the model are prepared using ArcGIS.

Parameters and parameter coefficients are calculated by the public domain statistical language R (R Development Core Team 2006) using logistic regression. We employ R script to automate the experimental calculations and visualize the results. This procedure provides a convenient framework for model estimation.

### 4. Calculation of Locational Probability

#### 4.1 Spatial information

The CA model requires four types of data: the initial urban configuration, as seed data for the starting point of the simulation; the rate of urbanization, for setting the boundary conditions; institutional control, as constraints of growth; and topographical and environmental constraint data. Land-use data are derived from two Landsat data sets, from 1987 and 2003 (Table 1.). Each image pixel is assigned to one of 90 classes by an iterative self-organizing data analysis (ISODATA) routine. Water bodies, ocean, and open ground are identified from the images in the initial classification. Next, in the second classification, water bodies and bare ground are masked from the original image and reclassified into 50 classes. Clouds, cloud shadows, and apparently unclassifiable pixels are defined as unknown. The remaining pixels are then assigned into six land-use classes: urban areas, paddy lands, wetlands, water bodies, ocean, and green lands. Later, the 1981 agricultural base map is used to verify the accuracy of paddy lands in the final classification image. All land-use classes that encompass built-up structures are classified as urban areas, including urban parks and airports. Wetlands within the core area, large industries, refineries, and military installations are identified from the agricultural base map from 1987 and reclassified as a growth-restricted class. The low resolution of the Landsat data somewhat compromises the accuracy of low-density urban areas and mixed rural areas. The classification map from 1987 provides the initial conditions for the CA simulation and for the logistic regression model, while the simulated results are evaluated by the classification map from 2002.

#### Table 1. Data and Data Sources

| Data                           | Source                                                                 |
|--------------------------------|------------------------------------------------------------------------|
| Land cover data                | Classified Landsat TM data images 1987 and 2002                        |
| Slope and elevation data       | 90 meter digital elevation model data                                  |
| Road network                   | Agricultural base maps (1972, 1981, 1991), Survey Department of Sri Lanka |
| Soil types                     | Soil map (1981)                                                        |
Data on the road network are prepared with paper maps from 1983 and 1992. Proposed new roads are excluded from the data. Distances from major and minor roads are calculated by the Euclidian distance function in the ArcGis spatial analysis tools. All data layers are prepared as 100m x 100m raster layers and converted to ASCII data for use in the statistical calculations and the CA model.

4.2 Data sampling for the logistic regression model

The size of the sample and the spatial dependence of data points are important considerations when analyzing spatial data. The maximum likelihood method relies on large sample sizes for accurate prediction. At the same time, a reduced sampling of data resulting from an increase of distance between data points will reduce the spatial autocorrelation effect.

In this study we define the data samples for the logistic regression model by stratified random sampling. With this approach, however, the smaller sample sizes might be less sensitive for small urban areas and for new urban growth in the periphery of the study area. We therefore select 20% of the total cells from the Core Area where 10% of the total samples are urban cells. This is proportional to the original ratio of urban and non urban cells in the total data set.

4.3. Factors of urbanization

Equation 4 in the previous section can be expressed in the following logistic regression form.

\[
\logit(p_i) = \ln \left( \frac{p_i}{1-p_i} \right) = \alpha + \beta w
\]

(7)

where \( \alpha \) and \( \beta \) are unknown values and \( w \) is a vector of independent variables. Logistic regression is an approach to learning function \( P(Y|X) \) in the case where \( Y \) is discrete-valued and \( X = (X_1, \ldots, X_\rho) \) is any vector containing discrete or continuous variables. This relationship is helpful for identifying the factors of urbanization and their relative effects on urbanization. Our regression analysis employs urbanization data from 1987 as explanatory variables. Two unknown variables, \( \alpha \) and \( \beta \), can be calculated once the independent variables of the model are identified. When \( Y \) is 1 (urbanized cell), the parametric model assumed by logistic regression can be written out as shown in Equation 5.

The factors of urbanization should be significant enough to predict the process of urbanization. The chosen factors in the calculations here include existing urban growth at the initial year, distances from urban centers and roads, site conditions such as topology, soil, and existing land cover types, and links between urban areas (Table 2).

Urbanization depends on the availability of land and adjacency to existing development. The rate of development in a neighborhood can thus be applied itself as a factor of urbanization (Yeh and Li, 1998; Li and Yeh, 2000; Ward et al., 2000). The ratio of urbanization is calculated in a 5x5 cell neighborhood to derive \( ROW1 \), the variable defined as the availability of infrastructure in the neighborhood.

Constraint factors derived from existing categories of natural land cover are given as a separate parameter (COVER). This parameter assigns a weight based on the environmental sensitivity of the land class. In this model we apply flat weights for paddy fields and green lands. The weights for our initial calculations are set based on a cross-classification-table and the rate of land consumption between 1987 and 2002. Paddy fields and green lands are assigned initial weights of 0.1 and 0.8, respectively.

| Variable/key | Definition |
|--------------|------------|
| ROW1 | Ratio of urban cells within a 5x5 neighborhood |
| DCBD | Distance measured from CBD to each cell (in km). |
| DR | Distance to the closest road (in km). |
| COVER | Type of land cover paddy, not developed or urbanized. (Each class categorically weighted.) |
| SOIL | 9 soil categories. |
| SLOPE | Slope of the site (given as a percentage) |
| ELE | Height<3 or height>3 (in meters). |

Table 2. Variable Key and Definition

Our chief goal in developing this model is to evaluate how urbanization patterns relate to the regulatory control of urban growth on paddy lands. We calculate the variable \( ROW1 \) (Table 2.) by dividing the total number of urbanized cells by the total number of cells in the neighborhood. The calculation is not sensitive to the land-use class of the cells. Thus, we derive a new variable \( ROW1 \_C \) with the following relationship.

\[
ROW1\_C = ROW1 \times COVER
\]

(8)

By including the COVER variable in the above form, we improve the performance of the model and the sensitivity of the model to non-urbanized cells.

A third factor included is accessibility. Access to the CBD, to urban centers, and to work locations is an important factor in land-use modeling. Public transportation and travel to work are primary factors of location choice in developing countries. The accessibility factor is calculated as two variables: the distance from each cell to the CBD (DCBD) and the distance from each cell to the closest road (DR). To adjust for the skewness of the distribution, the DCBD and SLOPE variables are log-transformed. The width of one cell is added to both the DCBD and DR variables, in order to avoid 0 distance samples. When Pearson’s product-moment correlation test is applied to test the independence of the continuous variables, the result suggests a correlation between DCBD and \( ROW1\_C \). This relationship is given as an interaction term \( ROW1\_C:DCBD \) in the model.

Logistic regression is carried out for the total sample
set of cells. Exploratory calculations are performed using the saturated model and variables are eliminated from the model in an iterative process. ELE (Elevation) and SOIL parameters do not improve the performance of the model, so we remove them from the final model after the initial calculations.

After altering the weights assigned to paddy lands and other non-urban lands in the COVER variable, we estimate the optimum weights by repeated calculations using logistic regression and AIC. The weights calculated for paddy lands and other non-urbanized lands are 0.2 and 0.8 for the model with the lowest estimated AIC value. Table 3. shows the results from the final model, and Fig.2. compares the estimated probability values with the original data. The estimated sensitivity and specificity for the model are 0.7788 and 0.9491, respectively. Here we find a clear abundance of non-urbanized cells with comparatively high probability values. These cells are presumably in various stages of urbanization.

5. Simulated Growth and Calibration
5.1 Growth based on different thresholds of transition probability

The probability of being an urban or non-urban cell is calculated based on probabilities estimated using a logistic regression model at a threshold of 0.5. The CA model updates the urban cells discretely via a dynamic routine. The use of newly urbanized cells to calculate the probability makes it difficult to explain the urban areas using the same thresholds within the CA model. Here we simulate several growth patterns based on a range of transition thresholds, in order to obtain the threshold that can most suitably simulate the urban dynamics of CMR. We employ thresholds of 0.6, 0.4, and 0.2 for the three simulations presented in this section (Fig.3.A, B, C). The cells for transition are selected with uniformly distributed random numbers between each threshold and 1. Only cells with higher probabilities are transformed when the simulations are done at high thresholds.

By applying a random number to choose transition cells, we can attain a realistic simulation with a certain degree of uncertainty (White & Engelen, 1993). As a consequence of this method, the model produces different results for the same parameter settings. These three simulations are repeated 150 times for each threshold setting, and the results with the most frequent level of agreement are used for the analysis in this section. Table 4. compares the simulated growth and actual urban growth, cell by cell, from 1987 to 2002. As the table shows, simulation C gives the best level of agreement among the three simulations.

Simulations based on higher thresholds are highly sensitive to the transportation network and environmental constraints. More growth can be observed away from the CBD, where there tend to be larger plots of open areas. Growths along roads are clearly visible in maps A and B, where the thresholds are 0.6 and 0.4 respectively.

In contrast to the patterns found in simulations A and B, we find prominent spreading and infill growths in C, the simulation based on the threshold of 0.2. Expansions of existing agglomerations and infill-type growths are observed closer to the CBD. In contrast, only few infill-type growths are visible in simulations A and B. The most visible growth phenomenon in these

---

**Table 3. Estimated Parameters for Model Variables Using Logistic Regression**

| Parameters | Estimate | Std. Error | z value | Pr(>|z|) | 95% confidence interval |
|------------|----------|------------|---------|----------|------------------------|
| Intercept  | -0.97212 | 0.33433    | -2.908  | 0.003641 | (-1.63041, -0.319472)  |
| DCBD       | -1.22373 | 0.11547    | -10.598 | <2.00E-16| (-1.449139, -0.896309) |
| ROW1_C     | 3.70335  | 0.53679    | 6.899   | 5.23E-12 | (2.6836011, 4.7900544) |
| SLOPE      | -0.18534 | 0.04951    | -3.743  | 0.000182 | (-0.282215, -0.088104) |
| DR         | -0.35471 | 0.10751    | -3.299  | 0.00097  | (-0.566678, -0.145137) |
| DCBD:ROW1_C| 2.01033  | 0.23104    | 8.701   | <2.00E-16| (1.5471238, 2.4535923) |

---

**Table 4. Levels of Agreement among Urban Cells in Simulations A, B, and C**

| Map | Threshold | % of agreement* with actual urban growth of 2002 |
|-----|-----------|-----------------------------------------------|
| A   | 0.6       | 53                                            |
| B   | 0.4       | 58                                            |
| C   | 0.2       | 61                                            |

* The cell is urban in both the simulated map and in the actual data
simulations is the extension of existing agglomerations. Old urban areas undergo infill-type growths.

The threshold of transition is adjusted for the best possible accuracy of the model. The CA model is calibrating by the visual test (Clark et al., 1997) and the level of agreement.

We obtain the optimal model from one of the simulations performed using the threshold 0.1 (Fig. 5.-A, B). Two tests are applied to assess the model performance for this setting. The locational accuracy of the model is assessed based on the omission and commission errors and the levels of agreement.

Table 5. shows the level of agreement between simulated and observed data for 2002. Most of the omission errors are visible around the scattered urban areas in the periphery and in new urban growths in the southern part of the study area. These scattered urban patches represent the new residential developments in newly sub-divided agricultural lands, where there is no road network at the initial year of simulation. Large commission errors are observed around the peripheral urban centers and along the main roads, where the urban boundaries are less accurately classified due to the mixed urban classes and low resolution of the Land sat data.

The spatial accuracy of the simulation results is assessed by a multiple-resolution fitting method. Turner et al. (1989) suggest that comparison at a single resolution isn’t adequate to evaluate spatial models. Here we evaluate the spatial agreement of the simulated map by a multi-scale goodness-of-fit method. To do so we have to calculate the distribution of land use at different levels of resolution. A multi-scale goodness-of-fit method uses a moving window filter to calculate the influence of variation at all scales. The fit $F_w$ for window size $w$ is defined by (Turner et al., 1989):

$$F_w = \sum_{x=1}^{s} \frac{\sum_{n=1}^{P} [a_{ij} - a_{ij}]}{2w^2}$$

where $w$ is the linear dimension of a window, $a_{ij}$ is the number of cells of category $i$ in map $n$ ($n=actual$ data, $2=estimated$), $P$ is the number of categories in the sampling window, $t_w$ is the number of sampling windows for the window size $w$, and $s$ is the window sliding through the map. If the two maps are identical, $F_w$ will be 1 and remain constant. If two maps with the same proportion of land cover types with very different spatial patterns, $F_w$ will increase gradually with the window size. If the spatial patterns slightly differ, $F_w$ will increase rapidly at first and then approach 1 gradually (Turner et al., 1989).

The $F_w$ is calculated for seven window sizes (2x2, 4x4, 6x6, 8x8, 10x10, 20x20, 40x40, and 60x60) for simulations at thresholds of 0.1, 0.4 and 0.6, as shown in Equation 8 (see Fig.3.). The goodness-of-fit values are high for all simulations, and highest for the

| Agreement* (%) | Omission** (%) | Commission*** (%) | Total area of growth (ha)**** |
|---------------|---------------|------------------|-----------------------------|
| 62            | 38            | 37               | 11298                       |

* The cell is urban in both the simulated map and actual data. ** The cell is urban in the actual data and non-urban in the simulated map. *** The cell is non-urban in the actual data and urban in the simulated map. **** The area of simulated urban growth between 1987 and 2002.
thresholds 0.1 and 0.4 (curves A and B in Fig. 4).

The multi-scale goodness-of-fit results and level of agreement suggest that the CA model can simulate urban growth of the CMR with high spatial and locational accuracy.

5.2 Growth based on conserving paddy land.

The objective of the CA simulation is to explore planning scenarios and the urbanization patterns associated with them. COVER is one of the constraint variables included in the model to control the consumption of paddy land for urbanization. The COVER values for paddy lands are altered in simulations carried out from 1987 to 2002. The simulation result in Fig. 5.-A and B are produced using a reduced weight of 0.01 for paddy lands in COVER.

When we visually compare the results generated from the above simulation scenario with those generated by the status quo simulation in Fig. 5., we find a number of undeveloped patches in the periphery and in the main urban areas. Alternative growths mainly appear as road growths and expansions of existing agglomerations in the eastern suburbs. Most of them are found around agglomerations closer to the CBD or closer to transportation networks, where infrastructure conditions are good. This growth control scenario manages to prevent about 50% of urban growth on paddy lands compared to the status quo.

When assessed by direct cell counts, the levels of agreement of the controlled scenario and status quo simulation are 58% and 63%, respectively. It thus appears that control of the COVER parameter has little effect on the accuracy of the model. We can speculate, on this basis, that the urban growth on other land classes is scarcely affected and remains close to the status quo result.

The model provides no additional improvement in preventing growth on paddy lands at the extreme level of control. If the cells with high probability values quickly run out under this condition, the simulation will begin converting the paddy lands closer to the CBD, as those cells retain higher probability values because of their proximity to the CBD. The selection of cells for transition depends on the suitability for urbanization given by the calculated probability. Under an extreme level of control, the model will select less suitable cells for urbanization and convert them to urban cells due to land scarcity. One important finding from this observation is a limit to the control that can be applied.
on paddy lands closer to the CBD.

According to this simulation result, the complexity of the urban patch forms increases with the growth control. Among the patches that remain undeveloped as a consequence of the growth control, those closer to the CBD are smaller than those further away from it.

6. Discussion and Conclusion

This study uses logistic regression to explore urbanization data. The technique is effective for identifying factors of urbanization and the relationships among those factors. Urbanization is a complex social phenomenon that seldom accords with normality assumptions. The variables that compose it are both categorical and continuous. Spatial autocorrelation can always be found in spatial data. We attempt here to minimize it by sampling in stratified layers. Though complete removal of spatial autocorrelation cannot be expected, experimentation with various sampling methods is of crucial importance in coping with it.

The regulation of paddy lands has a greater effect on urbanization patterns of the CMR according to the proposed CA model. Most the new growths shift from main agglomeration of the CMR to sub-urban centers further away. Closer observation shows that most of the alternative growth locations are along the transportation network. Improved infrastructure can ease the growth pressure on paddy lands. The fragmentation of paddy lands has been a serious problem in the CMR for nearly two decades. This model does not consider the environmentally sensitive lowlands. We expect, however, that the simple framework demonstrated here can be expanded to other environmentally sensitive lands for the simulation of wide-ranging environmental changes in the CMR.

This model appears to have fairly high accuracy, as assessed by cell-by-cell comparison. For shorter simulation periods of 5 to 10 years, the level of accuracy is about 70%. The basic modeling framework yields promising results. The spatial accuracy remains reliable after a fairly large number of simulations. The multi-scale goodness-of-fit method provides high accuracy within a given range of resolution.

Measures to promote healthy urban growth rely on a clear understanding of the urban growth patterns arising from land-use control, as well as economic and environmental consequences. Growth control scenario simulations suggest that there is a threshold of growth control on the conversion of paddy lands for urban activities. To evaluate this observation, it will be important to include further economical and environmental factors in the model as it develops.

References:

1. Batty, M. & Xie, Y. (1994), 'From cells to cities', Environmental and Planning B: Planning and Design 21, pp.521-548.
2. Cheng, J. & Masser, I. (2003), 'Urban growth pattern modeling: a case study of Wuhan city, PR China', Landscape and Urban Planning 62(4), pp.199-217.
3. Clarke, K. C.; L. Gaydos & Hoppen, S. (1997), 'A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area', Environment and Planning B 24, pp.247-261.
4. Deheragoda, C. K. M.; Wanasinge, Y. A. D. S. & Karunayake, M. M. (1992), 'Urban growth in Sri Lanka: Implication for regional development', UNCRD Working Paper 92-4.
5. Department of Census and statistics (2000), 'Statistical Abstract 2000', Technical report, Department of Census and Statistics, Ministry of Finance and Planning - Sri Lanka.
6. Hu, Z., & Lo, C. P., 'Modeling urban growth in Atlanta using logistic regression', Computers, Environment and Urban Systems (2007), doi:10.1016/j.compenvurbsys.2006.11.001
7. McDonnell, M. J. & Pickett, S. T. A. (1990), 'Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology', Ecology 71, pp.1232-1237.
8. Nations, U. (2003);WORLD URBANIZATION PROSPECTS: THE 2003 REVISION, Technical report, United Nations Department of Economic and Social Affairs: Population Division.
9. R Development Core Team (2006), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
10. Pickett, S. T. A.; M. L. Cadenasso; J. M. Grove; C. H. Nilon; R. V. Pouyat; W. C. Zipperer & Costanza (2001), 'Urban ecological systems: Linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas', Annual Review of Ecology and Systematics 32, pp.127-157.
11. Urban Development Authority of Sri Lanka (1998), 'Colombo metropolitan regional structure plan - synthesis Volume 1', Technical report, Urban Development Authority of Sri Lanka.
12. Ward, D.; Murray, A. & Phinn, S. (2000), 'A stochastically constrained cellular model of urban growth', Computers, Environment and urban systems 24, pp.539-558.
13. White, R. & Engelen, G. (1997), 'Cellular automata as the basis of integrated dynamic regional modelling', Environment and Planning B: Planning and Design 24, pp.235-246.
14. White, R.; Engelen, G. & Uljee, I. (1997), 'The use of constrained cellular automata for high resolution modeling of urban land use dynamics', Environmental and Planning B 24, pp.323-343.
15. Wu, F. (1998a), 'An experiment on the generic polycentric city of urban growth in a cellular automata city', Environment and Planning B 25, pp.731-752.
16. Xie, L. & Yeh, A. G. (2000), 'Modelling Sustainable urban development by the integration of constrained cellular automata and GIS', International Journal of Geographic Information Science 14, pp.131-152.
17. Yeh, A. G. & Li, X. (1998), 'Sustainable Land development model for rapid growth areas using GIS', International Journal of Geographical Information Science 12, pp.169-189.
18. Yeh, A. G. & Li, X. (2006), 'Errors and uncertainties in urban cellular automata', Computers, Environment and Urban Systems 30, pp.10-28.