# A CA based two-stage model of land use dynamics in urban fringe areas: An application to the Tokyo metropolitan region

**Takeshi Arai and Tetsuya Akiyama**
Department of Industrial Administration
Faculty of Science and Technology
Tokyo University of Science, Japan
tarai@rs.noda.tus.ac.jp

## 1. Introduction

From the late 1980s, Cellular automata (CA) has been applied to many modeling efforts as a tool for modeling spatial dynamics [O'Sullivan and Torrens 2000]. However, only a few CA based land use models were developed based on a rigorous statistical analysis of detailed land use transition data, for example, the multinomial logit model by Wu and Webster [Wu and Webster 1998] and the multiple regression model by Wu and Yeh [Wu and Yeh 1997]. In Japan, "Detailed Digital Information (10m Grid Land Use) Metropolitan Area" [Geographical Survey Institute 1998] was published recently. This data set includes land use data of each 10-m cell in the Tokyo Metropolitan Region, surveyed in 1974, 1979, 1984, 1989 and 1994, and the number of land use categories is fifteen.

Some studies on the CA modeling were carried out in Japan by using this data source. For instance, Arai and Akiyama [Arai and Akiyama 2001] constructed a short term land use dynamics model which simulates land use transition of each 100-m square cell in a suburban area of Tokyo, the size of which is a 10 km square. In the model, land use transition potentials of a cell are defined by similar linear equations to the formula introduced by White et al.[White, Engelen and Uljee 1997] . As the determinants of the transition potentials of a cell, the number of cells each land use within its neighborhood, which consists of 24 surrounding cells, and the proximity of it to the transport services are picked out. The process of land use transition of a cell is modeled as multiple stages of "dichotomies" based on the existing main types of land use conversion in the study area. Discriminant analysis is applied to the estimation of the parameters in the model. Nevertheless, according to comparison of real land use transitions between 1979 and 1989 with predicted ones by the model, the ratios of the number of correctly predicted cells to the total number of cells of individual land uses are at most 70%.

A reason for the lower performance of the model may be that the state of a 100-m square cell is represented by a single land use category, the number of unit data cells (10-m square cells) used by which is the largest in the 100-m square cell. Accordingly there are a wide diversity of mixed land use situations of the 100-m square cells, the states of which are the same category of land use. Since the difference between the case where one land use occupies the greater part of the cell and the case where some land uses have almost equal shares in the cell are neglected in the calibration process of the model, the errors of prediction are inevitably large. One way to address this problem is to construct a model which simulates the land use transition of a smaller cell, namely 10-m square cell, than in the preceding model. However, since the change in land use of a 10-m square cell for five or ten years may be affected by a variety of particular factors, it is not easy to make a model which can predict the land use transition of each 10-m square cell correctly.

Therefore, in this study the state of a 100-m square cell is represented by a multi-dimensional vector, the components of which are the numbers of 10-m square cells of individual land use categories within the cell. The purpose of this study is to construct a modified model which consists of the linear equations to predict the changes in values of the state vectors of individual 100-m square cells. The model has a two-stage structure, that is, a classification stage of cells into several groups according to their land use transition patterns using the discriminant equations and an estimation stage of the vectors of cells by the multiple regression equations. It was applied to the test area, which is located 30 kilometres northeast of the centre of Tokyo.

## 2. Two-stage model

### 2.1. The basic concept of the model

In this study, the state of a block, which we call a 100-m square cell here, is defined by the following multidimensional vector **S** instead of a single land use category :

$$\mathrm{S}^{t}_{xy} = \begin{pmatrix} s^{t}_{1,xy} \\ \vdots \\ s^{t}_{j,xy} \\ \vdots \\ s^{t}_{J,xy} \end{pmatrix} , \tag{1}$$

where $\mathbf{S}^{t}_{xy}$ is the state vector of a block located xy at the time t, the dimension of which is J; $s^{t}_{j,xy}$ is the j th components of the state vector S and it means the area of the j th land use category in the block; J is the total number of land use categories. The purpose of the model presented here is to predict the dynamic changes in the $s^{t}_{j,xy}$'s. The relationship between $s_{j,xy}$ at the time t, $s^{t}_{j,xy}$, and the one at the next time (t+1), namely $s^{t+1}_{j,xy}$, can be defined as follows:

$$s^{t+1}_{j,xy} = s^{t}_{j,xy} + ss^{t,t+1}_{j,xy} , \tag{2}$$

$$ss^{t,t+1}_{j,xy} = \sum_{i}^{J} z_{ij,xy} , \tag{3}$$

where $ss^{t,t+1}_{j,xy}$ is the increase or decrease in $s^{t}_{j,xy}$ between the time t and t+1; $z_{ij,xy}$ is the amount of the land use conversion from the i th category into the j th category; and xy is the location of the block. Moreover, we assumed the following relationships between $z_{ij,xy}$ and the transition potentials $p_{i,xy}$ and $p_{j,xy}$:

$$pp_{ij,xy} = p_{j,xy} - p_{i,xy} \tag{4.1}$$

$$z_{ij,xy} = f\,(pp_{ij,xy}, s^{t}_{i,xy}, s^{t}_{j,xy}) , \tag{4.2}$$

$$\text{(i) } \partial f / \partial pp_{ij} \geq 0, \partial f / \partial s_i \geq 0, \partial f / \partial s_j \leq 0 , \tag{4.3}$$

$$\text{(ii) } if\ \ pp_{ij} \geq 0\ \ then\ z_{ij} \geq 0\ and\ \ if\ \ pp_{ij} \leq 0\ \ then\ z_{ij} \leq 0 , \tag{4.4}$$

$$and\ \text{(iii) } z_{ij} = -z_{ji} , \tag{4.5}$$

where $pp_{ij,xy}$ is the difference between $p_{i,xy}$ and $p_{j,xy}$; $p_{i,xy}$ is the transition potentials to the i th land-use and $f$ is the function which has the characteristics (i), (ii) and (iii). (i) means that the bigger the difference between the transition potentials $p_{i,xy}$ and $p_{j,xy}$ is, the larger the amount of land use conversion from the i th category into the j th category is. Similarly, the bigger the area of the i th land-use is, the larger $z_{ij}$ is. And, the smaller the area of the j th land-use is, the larger $z_{ij}$ is. (ii) means that if $p_{j,xy}$ is larger than or equal to $p_{i,xy}$, $z_{ij}$ is positive or equal to zero. (iii) means that the amount of land use conversion from the i th category into the j th category plus the one from the j th category into the i th category equal to zero.

An example of the function which satisfy the above conditions is as follows:

$$g(pp_{ij}, s_i, s_j) = w_0 \cdot pp_{ij}^{\ w_1} \cdot s_i^{\ w_2} \cdot (s_j^{\ *} - s_j)^{w_3} \quad , \tag{5.1}$$

$$if \ \ pp_{ij} \geq 0 \ then \ \ f(pp_{ij}, s_i, s_j) = g(pp_{ij}, s_i, s_j) \tag{5.2}$$

$$otherwise \qquad f(pp_{ij}, s_i, s_j) = -g(pp_{ji}, s_j, s_i) \tag{5.3}$$

where $g$ is a name of function; $w_0$, $w_1$, $w_2$, $w_3$ are nonnegative parameters; and $s_j^*$ is the upper limit of $s_j$.

Another simpler example of the function $f$ is as follows:

$$\begin{aligned} h(pp_{ij}, s_i, s_j) &= w_0^* \cdot pp_{ij} + w_1^* \cdot s_i + w_2^* \cdot (s_j^* - s_j) + w_3^* \quad if \ \ pp_{ij} \neq 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad if \ \ pp_{ij} = 0 \end{aligned} \tag{6.1}$$

$$if \ \ pp_{ij} \geq 0 \ then \ \ f(pp_{ij}, s_i, s_j) = h(pp_{ij}, s_i, s_j) \tag{6.2}$$

$$otherwise \qquad f(pp_{ij}, s_i, s_j) = -h(pp_{ji}, s_j, s_i) \tag{6.3}$$

where $h$ a name of function; and $w^*_0$, $w^*_1$, $w^*_2$, $w^*_3$ are nonnegative parameters.

In this study, we assumed the function $f$ in the equation (4.2) to be the latter one of the above examples, because it is a simple quasi linear function and estimation of its parameters is easier than in the former example.

In addition, transition potentials of a block (namely, a 100-m square cell) can be defined by the following linear function based on the precedent study [Arai and Akiyama 2001]:

$$p_{j,xy} = \sum_m a_{jm} \cdot n_{m,xy} + \sum_k b_{jk} \cdot q_{k,xy} + c_j \quad , \tag{7}$$

where $p_j$ is the transition potential to the state j; $n_m$ is the number of blocks of land-use m within the neighborhood of the block; $q_k$ is the distance from the block to the nearest station (k=1) or main road (k=2); $xy$ is the location of the block; $a_{jm}$ and $b_{jk}$ are the weighting parameters; and $c_j$ is a constant parameter. Consequently, we can derive the following equation (8) from (3), (4.1), (4.2), (6) and (7):

$$ss^{t,t+1}_{j,xy} = \sum_{\mathrm{m}} v^{(1)}_{jm} \cdot n_{m,xy} + \sum_{\mathrm{k}} v^{(2)}_{jk} \cdot q_{k,xy} + \sum_{\mathrm{l}} v^{(3)}_{jl} \cdot s_{l,xy} + v^{(0)}_j \quad , \tag{8}$$

where $v^{(1)}_{jm}$, $v^{(2)}_{jk}$, $v^{(3)}_{jl}$ and $v^{(0)}_j$ are weighting parameters, which can be estimated by using the regression analysis.

### 2.2. Consideration of particularities in urban fringe areas

In urban fringe areas in Japan, the main type of land use conversion is that of non-urban land into residential land and industrial land, occasionally by way of vacant land for future construction. Based on the above fact, we can simplify the problem of estimating the parameters of equation (8). We picked out only three land use categories, namely non-urban land, vacant land for future building development and land for urban activities which includes residential land, industrial land and land for public facilities. Then, the following relationships can be assumed:

$$ss_{NU,xy} + ss_{V,xy} + ss_{U,xy} = 0 \quad , \tag{9}$$

$$and \quad ss_{NU,xy} \leq 0 \;\; and \;\; ss_{U,xy} \geq 0 \quad , \tag{10}$$

$$and \quad ss_{V,xy} \cdot ss_{U,xy} \leq 0 \quad , \tag{11}$$

*therefore*

$$if \;\; ss_{NU,xy} = 0 \;\; then \quad ss_{V,xy} = 0 \;\; and \;\; ss_{U,xy} = 0 \;\; , \tag{12.1}$$

$$or \quad ss_{V,xy} < 0 \;\; and \;\; ss_{U,xy} > 0 \;\; , \tag{12.2}$$

$$if \;\; ss_{NU,xy} < 0 \;\; then \quad ss_{V,xy} > 0 \;\; and \;\; ss_{U,xy} = 0 \;\; , \tag{12.3}$$

$$or \quad ss_{V,xy} = 0 \;\; and \;\; ss_{U,xy} > 0 \;\; , \tag{12.4}$$

where $ss_{NU,xy}$ is the increase (or decrease) in the area of non-urban land in a block in a unit period; $ss_{V,xy}$ is the increase in the area of vacant land; and $ss_{U,xy}$ is the increase in the area of land for urban activities. (9) means that total area of a block (xy) is unchanged, and (10) means the assumption that the area of non-urban land in the block decreases or is unchanged and that of land for urban activities increases or unchanged in urban fringe areas. In addition, in (11) we assume that the area of land for urban activities and that of vacant land do not increase simultaneously. From the above three premises, four possible situations of land use dynamics in the block, namely (12.1-12.4), can be derived. On the basis of the above framework, we constructed a model which has a two-stage structure, that is, a classification stage of blocks into several groups according to their land use transition patterns using the discriminant equations and an estimation stage of the state vectors of the blocks by the multiple regression equations. Figure 1 shows the outline of the model.

## 3. Application of the model

### 3.1. Outline of the study area

We applied the model presented here to the Kashiwa city area, which is located in the northeast of the Tokyo metropolitan region and the shape of which is a rectangle two km by three. The distance from the center of Tokyo to the area is about 30 km and residential areas have increased by about 50% during the last three decades in this area.

Although the number of land-use categories of the original data are 15, we categorized them again into following seven categories: Non-urban, Vacant, Industry, Houses, Commercial, Road, and Public uses.

In addition, we use a supplementary category, Urban, which includes Industry, Houses and Commercial.

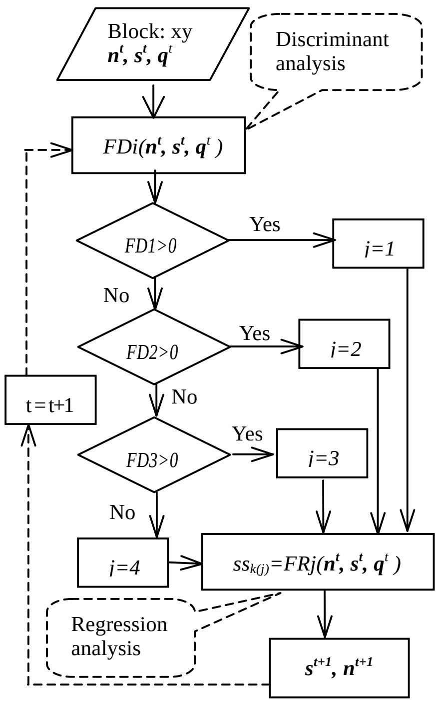



**Figure 1.** Two-stage model

Practically, since there were very few blocks used for Industry and Commercial, we analyzed land-use transitions among the following land-uses: Non-urban, Vacant, and Houses. ‘Neighborhood’ of a block means the eight blocks located in the Moore neighborhood of the block and ‘Enlarged neighborhood’ of a block means the 24 blocks, which include eight blocks in the neighborhood of the block and 16 blocks that abut on them. In order to analyze land-use dynamics at 10-year intervals, we used the data surveyed in 1974, 1984 and 1994.

### 3.2 Discriminant analysis

We classified the change in the state vector **S**, i.e. **SS**=($ss_{NU}$, $ss_V$, $ss_H$)’, of each block into 27 types according to the following rules:

For $i$= *NU, V,* and *H,* if $ss_i$>10 the extent of change in land-use i is ‘Plus’,
if $ss_i$<-10 the extent of change in land-use i is ‘Minus’,
otherwise the extent of change in land-use i is ‘O’

where *NU* means Non-urban land-use, *V* means Vacant land and *H* means Houses.
In reality, types of change at most of the blocks in the study area fall into only seven groups as shown in Table 1. The percentage of blocks which belong to the four groups that correspond to the situations (12.1)-(12.4) is 77.8%, and this fact supports the assumption (9)-(11) . Next, we estimated the discriminant equations that determine which group each block belongs to among the seven groups.

**Table 1.** Groups of land-use changing patterns existent in the study area

| | $SS_{NU}$ | $SS_V$ | $SS_H$ | # of blocks | Percentage |
|---|---|---|---|---|---|
| ***Group*** | | | | | |
| **(1)** | **O** | **O** | **O** | 329 | 54.8% |
| **(2)** | **O** | **Minus** | **Plus** | 55 | 9.2% |
| **(3)** | **Minus** | **Plus** | **O** | 55 | 9.2% |
| **(4)** | **Minus** | **O** | **Plus** | 28 | 4.7% |
| **(1)+(2)+(3)+(4)** | | | | 467 | 77.8% |
| **(5)** | **O** | **O** | **Plus** | 41 | 6.8% |
| **(6)** | **Minus** | **O** | **O** | 41 | 6.8% |
| **(7)** | **O** | **Minus** | **O** | 17 | 2.8% |
| **(8)** | | **All the rest** | | 34 | 5.7% |
| **Total** | | | | 600 | 100.0% |

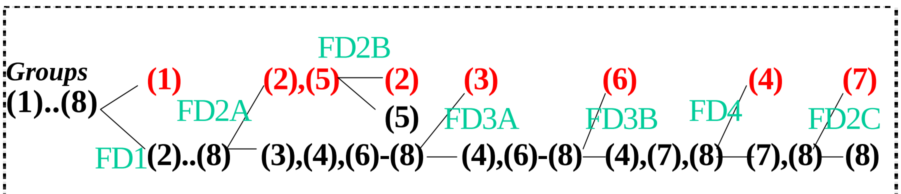



**Figure 2.** Multi-stage estimation of discriminant functions.

We worked out multi-stage discriminant analysis based on the model algorithm stated in chapter 2. Figure 2 shows the multi-stage estimation procedure of discriminant functions. The results are shown in Table 2. Since correct prediction rates of discriminant functions are between 67.5% and 87.5%, we are able to use them in our model.

**Table 2.** Discriminant equations

| | FD*i*: Discriminant functions | | | | | | |
|---|---|---|---|---|---|---|---|
| | ***i*:** 1 | 2A | 2B | 3A | 3B | 4 | 2C |
| Groups (FD>0) | (1) | (2), (5) | (2) | (3) | (6) | (4) | (7) |
| Groups (FD<0) | (2)-(8) | (3),(4) (6)-(8) | (5) | (4),(6) (7),(8) | (4),(7), (8) | (7),(8) | (8) |
| *Number of cells* | | | | | | | |
| Groups (FD>0) | **329** | **96** | **55** | **55** | **41** | **28** | **17** |
| Groups (FD<0) | 271 | 175 | 41 | 120 | 79 | 51 | **34** |
| *Factors(Indipendent variables)* | | | | | | | |
| Distance to the nearest station | (+) | | | (-) | | | |
| The number of 10-m cells of non-urban uses in a cell | (-) | | | (+) | | (+) | |
| The number of 10-m cells of vacant land in a cell | (-) | (+) | (+) | | (-) | | (+) |
| The number of 10-m cells of public uses in a cell | (-) | (+) | (-)* | | | | |
| The number of 10-m cells of houses in the neighborhood | | | (-)* | | | | (+) |
| The number of 10-m cells of commercial uses in the neighborhood | (+) | | | (-)* | | | |
| The number of 10-m cells of public uses in the neighborhood | | | | | (-) | | |
| The number of 10-m cells of urban uses in the neighborhood | | (+)* | | | | (+) | |
| The number of 10-m cells of non-urban uses in the enlarged neighborhood | | | | (+) | | | |
| The number of 10-m cells of public uses in the enlarged neighborhood | | | | (-) | | | |
| The number of 10-m cells of | | | | | (+) | | |
| Constant | (+) | (-) | (-) | (-) | (-) | (-) | (-) |
| **Correct prediction rate (%)** | 67.5 | 79.3 | 87.5 | 74.9 | 73.3 | 75.9 | 78.4 |

The significant parameters of discriminant equation 1(FD1) suggest that if a block has larger area of Non-urban land-use, Vacant land and Public uses within the block, and it is located closer to the nearest station and commercial facilities, the land-use of the block is not stable. FD2A and FD4 show that the primary factor which promotes land-use transition to Houses may be the existence of Urban land-use blocks in its neighborhood and that the existence of the resource for development, i.e. Non-urban land and Vacant land, within the block is another important factor. In addition, we could not find irrational relationships in all the discriminant equations.

### 3.3 Regression analysis

In order to estimate the equation (8) on each $ss_i$, we carried out regression analysis. The results are shown in Table 3.

**Table 3.** Regression equations

| | FRj: Regression equations | | | | |
|---|---|---|---|---|---|
| | Group(2) | Group(3) | Group(6) | Group(4) | Group(7) |
| *Dependent variables* | Increase in 10-m cells of houses in a cell | Increase in 10-m cells of vacant land in a cell | Decrease in 10-m cells of non-urban uses in a cell | Increase in 10-m cells of houses in a cell | Decrease in 10-m cells of vacant land in a cell |
| *Independent variables* | | | | | |
| Distance to the nearest station | -0.00467* | -0.0323** | | -0.0145* | |
| The number of 10-m cells of non-urban uses in a cell | | 0.448** | -0.557** | | |
| The number of 10-m cells of vacant land in a cell | 0.529** | | | | -0.885** |
| The number of 10-m cells of non-urban uses in the enlarged neighborhood | | | | -0.0171* | |
| The number of 10-m cells of urban uses in the enlarged neighborhood | | -0.0386** | | | |
| The number of 10-m cells of vacant land in the enlarged neighborhood | | | | 0.0779* | |
| Constant | 8.305 | 37.07 | -1.89 | 9.80 | 9.61 |
| Coefficient of determination($R^2$) | 0.53 | 0.52 | 0.34 | 0.43 | 0.85 |

Since the coefficients of determination ($R^2$) are between 0.34 and o.85, they are not so high. Moreover, the number of significant explanatory variables in each regression equation is less then four. However, all the signs, i.e. positive or negative, of the coefficients of independent variables are rational. Therefore, we used them in our model. The results suggest that proximity to the nearest station promotes development of housing areas, and over a half of areas of Non-urban land-use and Vacant land are converted their land-uses to residential land-use. As for Group (5), any significant explanatory variables couldn't be identified. So, we used the average value of $ss_H$'s at the blocks in the group as the estimate.

## 4. Validation of the model

In order to test the validity of our model, we carried out the simulation which reproduced the land-use situation in the study area in 1984 and 1994, starting from the situation in 1974 by using the model. The comparison of predicted results and real states are shown in Table 4. and 5. Although the average error percentages of $ss_{NU}$ and $ss_H$ at each block are not so high, the average error percentage of $ss_V$ is rather high as shown in Table 3.

Next we calculated the representative land-use of each block and compared the predicted one and the real one. Table 5 shows that correctly predicted percentages of the block states used for Non-urban and Houses are very high.

**Table 4.** Average error and error rates

| | Non-urban | Vacant | Houses |
|---|---|---|---|
| Average error | | | |
| 1984 | 12.3 | 11.0 | 10.9 |
| 1994 | 16.9 | 12.8 | 17.7 |
| Average error rate1984 | 16.1% | 70.9% | 26.4% |
| 1994 | 29.4% | 74.5% | 41.2% |

**Table 5.** Correctly predicted rates of representative land-use of a block

| | 1984 | | | 1994 | | |
|---|---|---|---|---|---|---|
| | Correctly predicted | Number of cells | Correctly predicted rate | Correctly predicted | Number of cells | Correctly predicted rate |
| *Land use categories* | | | | | | |
| Non-urban | 284 | 297 | 96% | 224 | 252 | 89% |
| Vacant | 11 | 38 | 29% | 8 | 26 | 31% |
| Houses | 150 | 192 | 78% | 166 | 236 | 70% |

## 5. Conclusions

We created a two-stage CA based land use model which describe the transition of land-use of each 100-m square cell. As the factors which may affect the land-use transition potential, we picked out aggregate land-use state in the neighborhood and accessibility to transport services. We tried to apply this model to a suburban city in the Tokyo metropolitan region. We calibrated the model by using discriminant analysis and regression analysis. The comparison of predicted results and real states shows performance of the model was improved comparing with the precedent model by the authors.[Arai and Akiyama 2001]

## References

Arai, T., Akiyama T. (2001), "A Statistical Analysis for Making a Block-based Land Use Model Used for Planning in Local Governments", Proceedings of CUPUM 2001 (in CD-ROM), 1-20.

Geographical Survey Institute. (1998), "Detailed Digital Information (10 m Grid Land Use) Metropolitan Area 1979, 1984, 1989 and 1994", Japan Map Center (Tokyo). (in CD-ROM).

O'Sullivan, D., Torrens, P. M. (2000), "Cellular Models of Urban Systems", CASA(Centre for Advanced Spatial Analysis) Working Paper, 22, 1-11.

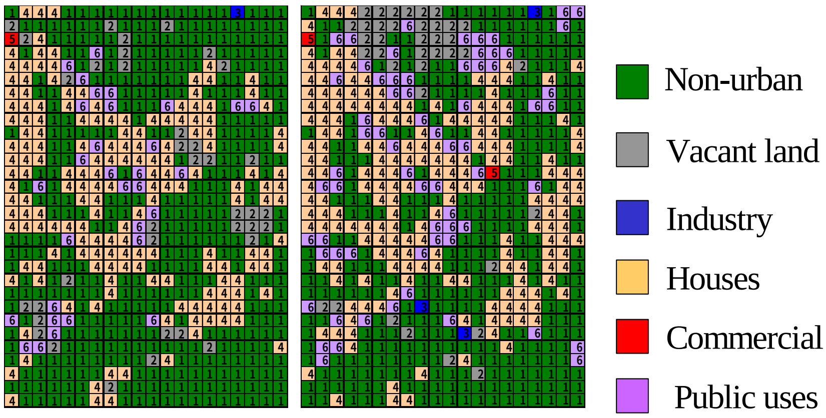
White, R., Engelen, G., Uljee, I. (1997), "The use of constrained automata for high-resolution modelling of urban land-use dynamics", Environment and Planning B, 24, 323-343.

Wu, F., Webster C. J. (1998), "Simulation of land development through the integration of cellular automata and multicriteria evaluation", Environment and Planning B, 25, 103-126.

Wu, F. L., Yeh, A. G. O. (1997), "Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: a case study of Guangzhou", Urban Studies, 34 (11), 1851-1879.

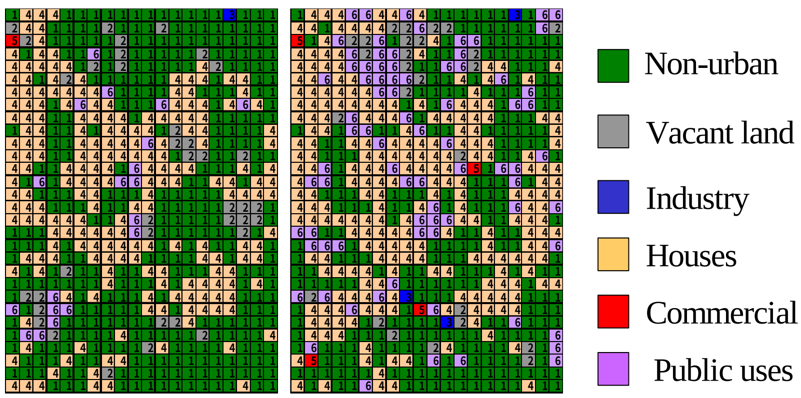
**Appendix:**
**Comparison between predicted landuse and real landuse**

## Land-use of Kashiwa area in 1984




Predicted (**1984**) Real state (**1984**)

## Land-use of Kashiwa area in 1994




Predicted (**1994**) Real state (**1994**)