

Calibration and Assessment of Multitemporal Image-based Cellular Automata for Urban Growth Modeling

Sharaf Alkheder and Jie Shan

Abstract

This paper discusses the calibration and assessment of a cellular automata model for urban growth modeling. A number of transition rules are introduced in the cellular automata model to consider the most influential urbanization factors, such as land-cover maps obtained from satellite images and population density from the census. The transition rules are calibrated both spatially and temporally to ensure the modeling accuracy. Spatially, each township (about 6 miles \times 6 miles) in the study area is used as a calibration unit such that the spatial variability of the urban growth process can be taken into account. The temporal calibration is performed by using a sequence of remote sensing images from which the land-cover information at different years is extracted. As for the assessment, fitness (for urban level match) and two types of modeling errors (for urban pattern match) are introduced as the evaluation criteria. The study shows that the use of images reduces the need for a large number of input data. Evaluation on the rule variogram reveals that the transition rule values are correlated spatially and vary with the urbanization level. The paper reports the study outcome over the city of Indianapolis, Indiana for the past three decades using Landsat images and the population data.

Introduction

Remarkable progress has been achieved in urban dynamic modeling to understand the urban growth process (Meaille and Wald, 1990; Batty and Xie, 1994a and 1994b). Some urbanization models focus more on the physical aspects of the urban growth process (Wilson, 1978), while others on social factors (Jacobs, 1961). An example of the physical models is the land-use transition model of Alonso and Muth in landscape economics (Wilson, 1978). Social models simulate the urbanization process according to the difference between individuals' intentions and their behavior (Clarke *et al.*, 1997; Portugali *et al.*, 1997). According to Clarke *et al.* (1997), urban growth models can be designed either for a specific geographical location such as BASS II which models

the urbanization process for the San Francisco Bay area only (Landis, 1992), or as general models such as human-induced land transformations (HILT) where its growth rules are designed to be general enough to consider different city structures.

Yang and Lo (2003) classify the urban dynamic models into three categories: "cellular automata-based models" such as Clarke *et al.* (1997); "probability-based models" such as Veldkamp and Fresco (1996); and "GIS weighted models" like the Pijanowski *et al.* (1997) model. The "cellular automata-based models" are becoming popular in recent literature mainly because of their ability to model and visualize spatial complex phenomena (Takeyama and Couclelis, 1997). Urban "cellular automata models" perform better as compared to the conventional mathematical models (Batty and Xie, 1994a) and simplify the simulation of complex systems (Wolfram, 1986; Waldrop, 1992). The fact that the urban process is entirely local in nature also makes the cellular automata a preferred choice (Clarke and Gaydos, 1998).

Many urban cellular automata models are reported. The model of White and Engelen (1992a and 1992b) involves reduction of space to square grids, based on which a set of initial conditions is defined. The transition rules are implemented recursively until the reference data are matched by the modeling results. Cellular automata has been used by Batty and Xie (1994a) to model the urban growth of Cardiff of Wales, and Savannah of Georgia. Later, Batty *et al.* (1999) develop a model that tests many hypothetical urban simulations to evaluate the different model structures. Based on the previous work (von Neumann, 1966; Hagerstrand, 1967; Tobler, 1979; Wolfram, 1994), Clarke *et al.* (1997) propose the SLEUTH model, which is able to modify the parameter settings when the growth rate exceeds or drops below a critical value. Clarke and Gaydos (1998) use SLEUTH to model the urban growth in the San Francisco Bay Area and Washington D.C./Baltimore, Maryland corridor. Yang and Lo (2003) use the SLEUTH model to simulate the future urban growth in Atlanta, Georgia with different growth scenarios. Wu (2002) develops a stochastic cellular automata model to simulate rural-to-urban land conversions in the city of Guangzhou, China.

Calibration of cellular automata models is essential to achieve an accurate modeling outcome. However, it has been ignored until recent efforts were made to develop

Sharaf Alkheder is with the Queen Rania's Institute of Tourism and Heritage, The Hashemite University, Jordan, and formerly with the Geomatics Engineering, School of Civil Engineering, Purdue University, West Lafayette, IN 47907.

Jie Shan is with Geomatics Engineering, School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907 (jshan@ecn.purdue.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 12, December 2008, pp. 1539–1550.

0099-1112/08/7412-1539/\$3.00/0
© 2008 American Society for Photogrammetry
and Remote Sensing

cellular automata as a reliable procedure for urban development simulation (Wu, 2002). Calibration is meant to determine the optimal values for the parameters in the transition rules such that the modeled urban growth closely matches the actual urban growth. The difficulty of the calibration is partially due to the complexity of the urban development process (Batty *et al.*, 1999). Clarke *et al.* (1997) use visual tests to establish the ranges of the parameters and provide their initial values. About a dozen of statistical measures are calculated for certain features to check the match between the actual data and the modeling results. Such visual and statistical tests are repeated for each parameter set. Wu and Webster (1998) use multi-criteria evaluation (MCE) to identify the parameter values for their cellular automata model, whereas neural networks (NN) are used by Li and Yeh (2001). The fact that most urban cellular automata models need a large number of data input variables is not free of risk. Many uncertainties show up in the simulation output. These can result from the uncertainty in the input data, uncertainty propagation through the model, and the uncertainty of the model itself in term of to what degree the model represents the reality. Previous research shows that urban modeling is very sensitive to the errors of the input data (Li and Yeh, 2003). Therefore, it is beneficial to minimize the need for large input data to reduce the modeling uncertainty and redundancy of the input variables.

This paper is focused on two important aspects in urban cellular automata modeling: calibration and assessment. First, our cellular automata model is designed to reduce the amount of input data. For this purpose a historical set of satellite imagery is used as an alternative to cadastral maps as being used in literature. We believe that building the model over the imagery directly is more realistic as compared to cadastral maps. The imagery is a rich source of information including land-cover, urban extent, and growth constrains (e.g., water resources). This will reduce the need of having different sets of input data layers. In addition, uncertainty of urban modeling that usually rises from having multiple input data layers (and hence variable precisions) will be reduced. Other data that are not included in the imagery (such as population density) can be used as extra input layers. Secondly, most cellular automata models assume that one set of transition rules will fit the whole study area. As a matter of fact, some regions in a study area may have different urbanization behavior than others. Based on this understanding, we argue that the calibration should be carried out both spatially and temporally. The spatial calibration takes into account the spatial variability in urban process. In this study, the study area is divided into townships, each of which forms a calibration unit. The transition rules are calibrated to find the best values that fit the urban dynamics for each township. The temporal calibration is based on multitemporal imagery and allows the transition rule values to change over time to meet the variable urban pattern in time. Finally, modeling results are assessed with three quality measures, one for urban count and two for modeling errors. Calibrated rules should be able to reproduce both the same urban count and the same urban pattern as the reality. The rule values that produce urban count close to real imagery with minimum modeling errors are selected. The approach is implemented first on a synthetic city to study the effect of growth factors on urban process and then expanded to model the historical urban growth of Indianapolis, Indiana over the last three decades.

Principles of Cellular Automata

Cellular automata is originally introduced by Ulam and von Neumann in 1940s as a framework to study the behavior of complex systems (von Neumann, 1966). It is commonly defined as a dynamical discrete system in space and time that operates on a uniform grid under certain rules. It consists of four components: pixels, their states (such as land-use classes), neighborhood (square, circle, etc.), and transition rules. Cellular automata computation is iterative, with the future state of a pixel being determined based on the current pixel's state, neighborhood, and transition rules. Based on the work of Codd (1968), Sipper (1997) provides a formal definition of two dimensional (2D) cellular automata. Let I represents a set of integers, a cellular space associated with the set $I \times I$ can be defined. The neighborhood function for pixel α is:

$$g(\alpha) = \{\alpha + \delta_1, \alpha + \delta_2, \dots, \alpha + \delta_n\} \quad (1)$$

where $\delta_i (i = 1 \dots n)$ represents the index of the neighborhood pixels. Figure 1 shows an example of a 2D cellular automata grid system, where $I = 5$ represents the total space of pixels in a grid of $5 \times 5 = 25$ pixels. As an example, the state of pixel α is urban and it is surrounded by eight neighbors $\delta_i (i = 1 \dots 8)$ in a 3×3 square neighborhood. The neighborhood of pixel α can be generally represented as a city-block metric τ :

$$\tau(\alpha, \beta) = |x_\alpha - x_\beta| + |y_\alpha - y_\beta| \quad (2)$$

given that $\alpha = (x_\alpha, y_\alpha)$ and $\beta = (x_\beta, y_\beta)$. The function $\tau(\alpha, \beta)$ defines the set of pixels β around pixel α such that $\{\beta \in I \times I\}$. As an example, $x_\alpha - x_\beta = \pm 1$ in the x -direction and $y_\alpha - y_\beta = \pm 1$ in the y -direction specify a 3×3 square neighborhood for pixel α . The neighborhood states $h^t(\alpha)$ are defined as:

$$h^t(\alpha) = (v^t(\alpha), v^t(\alpha + \delta_1), \dots, v^t(\alpha + \delta_n)) \quad (3)$$

where $(v^t(\alpha), v^t(\alpha + \delta_1), \dots, v^t(\alpha + \delta_n))$ are the states of pixel α and its neighborhood pixels at time t . The selected neighborhood kernel in Figure 1 for the center pixel has states $h^t(\alpha) = [\text{water, road, urban, water, urban, urban, road, urban}]$ (in row-first order).

Finally, the relationship between the state of pixel α at time $(t + 1)$ and its neighborhood states at time t can be expressed as:

$$v^{t+1}(\alpha) = f(h^t(\alpha)) \quad (4)$$

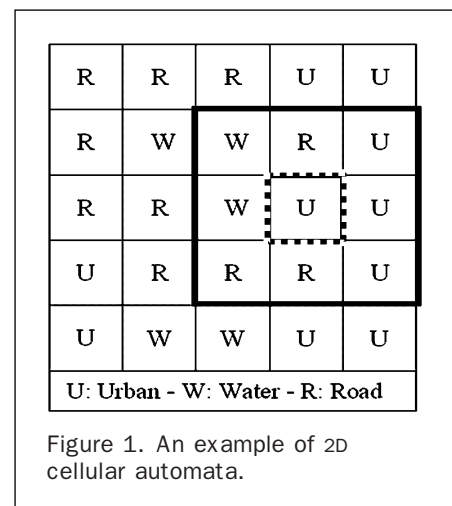


Figure 1. An example of 2D cellular automata.

where $f(h^t(\alpha))$ is the transition function that represents the cellular automata transition rules defined on α and its neighborhood. Typically, the transition function $f(h^t(\alpha))$ uses IF . . . THEN rules over $h^t(\alpha)$ to identify the future state of α at time $(t + 1)$.

Modeling of a Synthetic City

This section implements the cellular automata principle to a synthetic city to study the effect of modeling parameters on the urban growth process. It mimics the reality through introducing complex structures for an urban system. Figure 2 presents an image of 200×200 pixels used as the input to the cellular automata algorithm. Six classes are defined: road, river, lake, pollution source, urban, and non-urban (other).

The design of the cellular automata model needs to reflect the effect of the land-use on the urban growth process. Transportation systems encourage and drive the urban development. For example, commercial centers should have access to road network for customer's visit and goods delivery. Therefore, the cellular automata rules related to roads should encourage urban development for pixels near roads. River and lake pixels should be constrained such that no urban growth is allowed on these locations to conserve water resources. On the other hand, lakes are considered as one of the attractive factors for urban development especially residential and recreational types, so the corresponding rule needs to show such effect on urban development. The pollution sources are included as one of the constraints for urban development due to their effect on the degradation of ecological system. The designed cellular automata rules should prevent urban

growth in such locations. Based on the above considerations, the following rules are used

- IF a test pixel is urban, river, road, lake or pollution source, THEN no change.
- IF a test pixel is non-urban AND there is no pollution pixel in its neighborhood, then four cases are defined:
 1. IF three or more of the neighborhood pixels are urban, THEN change the test pixel to urban.
 2. IF one or more of the neighborhood pixels are road AND one or more are urban, THEN change the test pixel to urban.
 3. IF one or more of the neighborhood pixels are lake AND one or more are urban, THEN change the test pixel to urban.
 4. ELSE keep non-urban.

The above cellular automata rules first check the growth constraint to preserve certain land-cover classes (e.g., water), then test the possibility of urban development for non-urban pixels based on the urbanization level in the neighborhood. Figure 2 shows the modeled urban growth results after 0, 25, 50, and 60 growth steps with a 3×3 neighborhood. The effect of the road on driving the urban development is clear where the growth rate is higher near the road and its pattern follows the road's direction. Higher growth rate towards the lakes is also noticeable. The restriction on growth in locations close to pollution sources succeeds in decreasing the urban development rate and creates "buffered" zones around such places. Finally, the growth constraint on water pixels succeeds as well in conserving the water resources in future urban growth. The above study demonstrates that urban growth can be modeled by properly defining the transition rules.

Modeling of Indianapolis City

This section applies the cellular automata approach to a real city. The study area, transition rules, and evaluation criteria will be discussed.

Study Area and Data

Indianapolis, Indiana is selected for the study. Indianapolis is located in Marion County at latitude $39^\circ 44' N$ and longitude of $86^\circ 17' W$ as shown in Figure 3. It has experienced recognizable accelerated growth in population and urban infrastructure over the last few decades. It grows from a small part of Marion County in early 1970s to cover the entire county and parts of the neighboring counties in 2003. The necessity arises to model the urban growth over time for sustainable planning and distribution of infrastructure services.

Two types of data are used as input to the cellular automata model: land-use data (thematic imagery) and population density. The historical satellite images of 1973 (MSS, 4 bands), 1982, 1987, 1992, and 2003 (TM, 7 bands) in UTM NAD83 projection were collected over the study area. Images are classified using the Anderson *et al.* (1976) classification system to produce five land-use maps containing seven classes, namely water, road, residential, commercial, forest, pasture, and row crops. Commercial and residential classes represent urban class of interest in this study. All classified images were resampled to 60 m resolution as input to the cellular automata model.

In addition to the images, the 1990 and 2000 population census tract maps (see Figure 4, population per tract) are also used. To prepare the population density grids as the

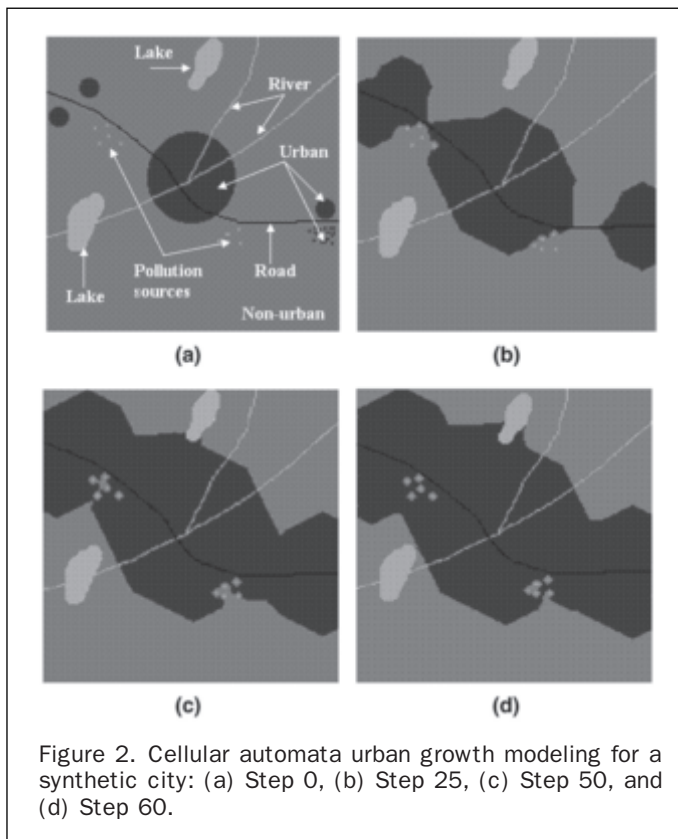


Figure 2. Cellular automata urban growth modeling for a synthetic city: (a) Step 0, (b) Step 25, (c) Step 50, and (d) Step 60.

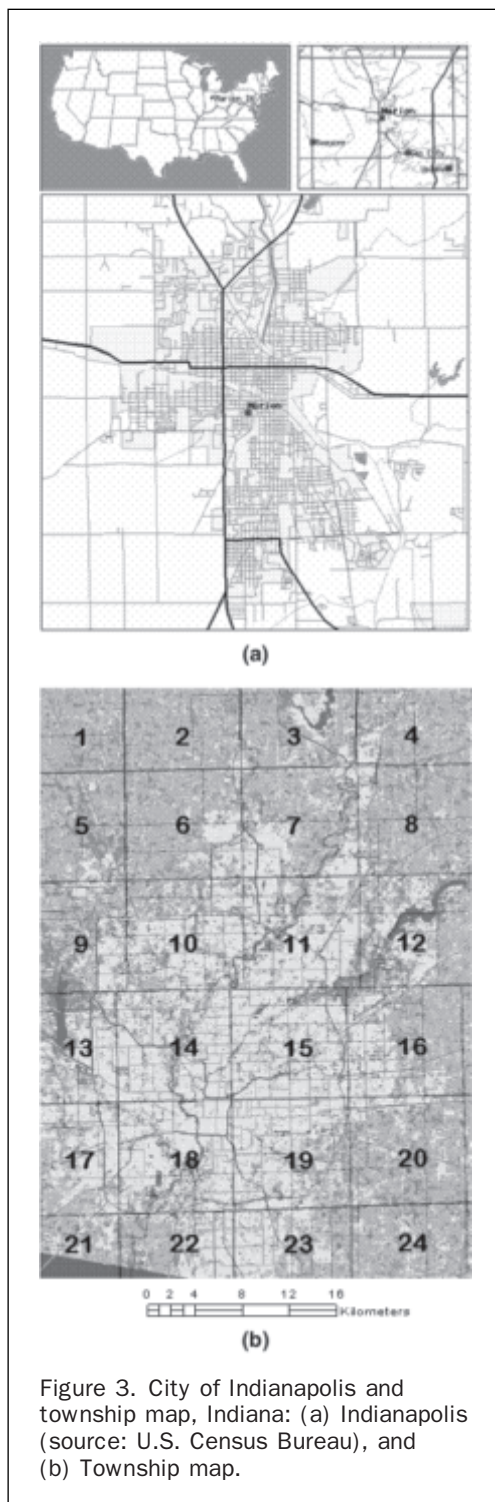


Figure 3. City of Indianapolis and township map, Indiana: (a) Indianapolis (source: U.S. Census Bureau), and (b) Township map.

input to cellular automata, the following procedure is used for both years (1990 and 2000). The area of each census tract is computed and used to produce the tract population density by dividing its population by its area. The centroid of every census tract and the overall city centroid for the study area are computed, and the distance from each census tract centroid to the city centroid is determined. Population densities for census tracts within certain distance range are averaged to reduce the variability in data. An exponential function is fitted representing population density as a

function of distance from the city center for both 1990 and 2000, separately:

$$POPULATION - DENSITY = A \cdot e^{-(B \cdot DISTANCE)}. \quad (5)$$

Parameters A and B for 1990 and 2000 are used to calculate their yearly change rates. The updated parameters (A and B values that vary year by year) are used to calculate the population density grids for each year from 1973 to 2003, where the population density for each pixel at a given year is calculated according to its distance from the city centroid using Equation 5. These population density grids are used as another input to the cellular automata.

Transition Rules

The implementation of the cellular automata consists of defining the transition rules, calibrating them, and evaluating the modeling results for prediction purpose. Cellular automata transition rules are designed as a function of land-use, growth constraints, and population density. A 3×3 neighborhood is used to minimize the number of input variables to the model. The rules identify the urban level in the neighborhood needed for a test pixel to urbanize, and take into account no growth constraints for certain land-cover classes. The effect of the closeness to urban area and infrastructure is also considered in the rule definition. The following rules are defined:

1. IF a test pixel is water, road or urban (residential or commercial), THEN no change to the test pixel.
2. IF a test pixel is non urban (forest, pasture or row crop), THEN:
 - IF its population density is equal or greater than a threshold (P_i) AND the number of neighborhood residential pixels is equal or greater than a threshold (R_i), THEN change the test pixel to residential.
 - IF its population density is equal or greater than a threshold (P_i) AND the number of neighborhood commercial pixels is equal or greater than a threshold (C_i), THEN change the test pixel to commercial.
 - ELSE keep non urban.

Evaluation Criteria and Rule Calibration

Evaluation and calibration are performed township by township. A township map is a semi-grid as shown atop the image in Figure 3. There are a total of 24 townships in the area. Dividing the study area into townships will take into consideration the effect of site specific features (spatial calibration) in each township on urban growth. The same cellular automata transition rules are defined for all townships; however, different townships may have different rule values. Spatial calibration is to find the optimal transition rule values (R, C, P) _{i} for each township. For this purpose, cellular automata is run for all possible combinations (R, C, P) _{i} in the search space. The search space for both R_i and C_i is respectively from 0 to 8 (the possible neighborhood size of 3×3 kernel) with integer increment of 1. The search space for P_i ranges from 0 to 3 with increment of 0.1. For each township, the cellular automata runs for a total of 2511 ($9 \times 9 \times 31$) combinations.

An evaluation scheme with three measures is designed for each township:

1. Fitness measure: this is the ratio of modeled urban pixel count to the ground truth count:

$$Fitness (\%) = \frac{Modeled \ urban \ count}{Ground \ truth \ urban \ count} \times 100. \quad (6)$$

2. Type I modeling error: this is used to identify the urban class modeling mistakes. It counts the pixels that are urban

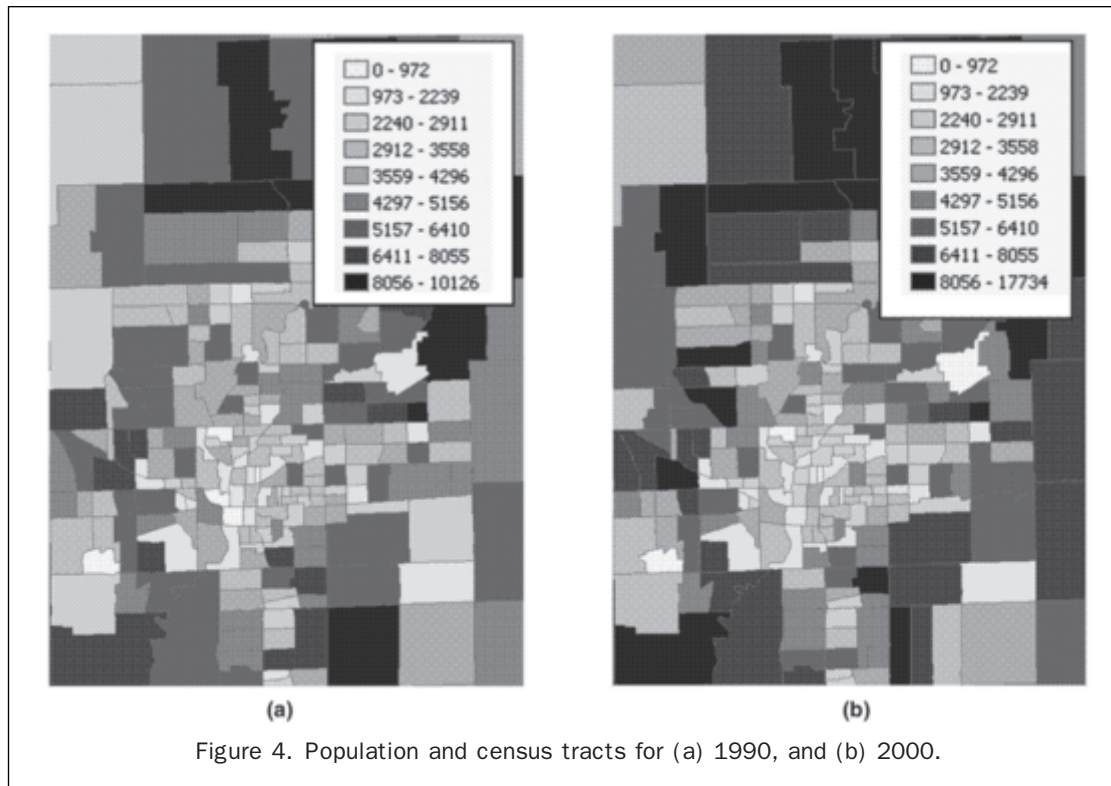


Figure 4. Population and census tracts for (a) 1990, and (b) 2000.

in the ground truth image but non-urban in the modeled image.

$$\text{Type I (\%)} = \frac{\text{Type I count}}{\text{Urban count}} \times 100. \quad (7)$$

3. Type II modeling error: this is used to identify the non-urban class modeling mistakes. Type II error counts the pixels that are non-urban in real but urban in the modeled image.

$$\text{Type II (\%)} = \frac{\text{Type II count}}{\text{Non urban count}} \times 100. \quad (8)$$

The total error ΔE (%) based on the urban and non-urban counts (total township pixels count) represents the overall modeling error:

$$\Delta E(\%) = \frac{\text{Type I count} + \text{Type II count}}{\text{Total count}} \times 100. \quad (9)$$

Fitness measure is introduced to indicate how a specific $(R,C,P)_i$ combination succeeds in reproducing the same real urban level within a township. A rule combination is said to overestimate the township urbanization level if the fitness is more than 100 percent, while a fitness less than 100 percent means underestimation of the urbanization level. Type I and Type II errors represent the pixel by pixel difference between the simulation results and ground truth. It also provides a strict measure for the mismatch between the simulated and actual urban patterns. Such errors need to be minimized for accurate modeling. Among all the rule value combinations, the one with minimum total error and with fitness value closest to 100 percent (within ± 10 percent) is selected as the best.

The cellular automata modeling starts running from 1973 to 1982, which is the first ground truth image used for

calibration. The best rule combination is selected based on the above evaluation criteria. In the next step, temporal calibration is implemented through recalibrating the cellular automata rules with the 1987 ground truth image. The same procedure is repeated in 1987 to find the best set of rule values for each township to reproduce the growth pattern in 1987. The objective of recalibration in 1987 is to take the temporal urban dynamics change into consideration in the calibration process. By doing so, the transition rules can be exposed to the changes in urban growth pattern over time and hence can be adapted to such dynamics.

The next step is to evaluate the prediction capability of the cellular automata modeling without calibration at the destination year. The set of calibrated rules for all townships that best model the ground truth in 1987 are used to predict the future growth in 1992. The prediction in 1992 represents short term prediction of five years. The predicted image in 1992 is evaluated based on the three evaluation measures. The next prediction is performed in 2003 for a long term period of 11 years starting from 1992 using the best rules after the calibration in 1992. Table 1 shows the modeling results at the calibrated years (1982 and 1987), while Table 2 shows those for the year 1992 and 2003 prediction results. The calibrated images (1982 and 1987) and predicted images (1992 and 2003) are shown in Figures 5 to 8, respectively.

Assessment and Discussion

The above results will be assessed in this section to understand the properties of the cellular automata modeling. It will start with an overall quality evaluation, followed by an evaluation on the distribution and correlation of transition rule values based on their variograms.

Quality of the Modeling

The results presented in Table 1 and 2 are a summary of the evaluation measures: fitness, Type I, Type II, and total errors

TABLE 1. 1982 AND 1987 CALIBRATION RESULTS

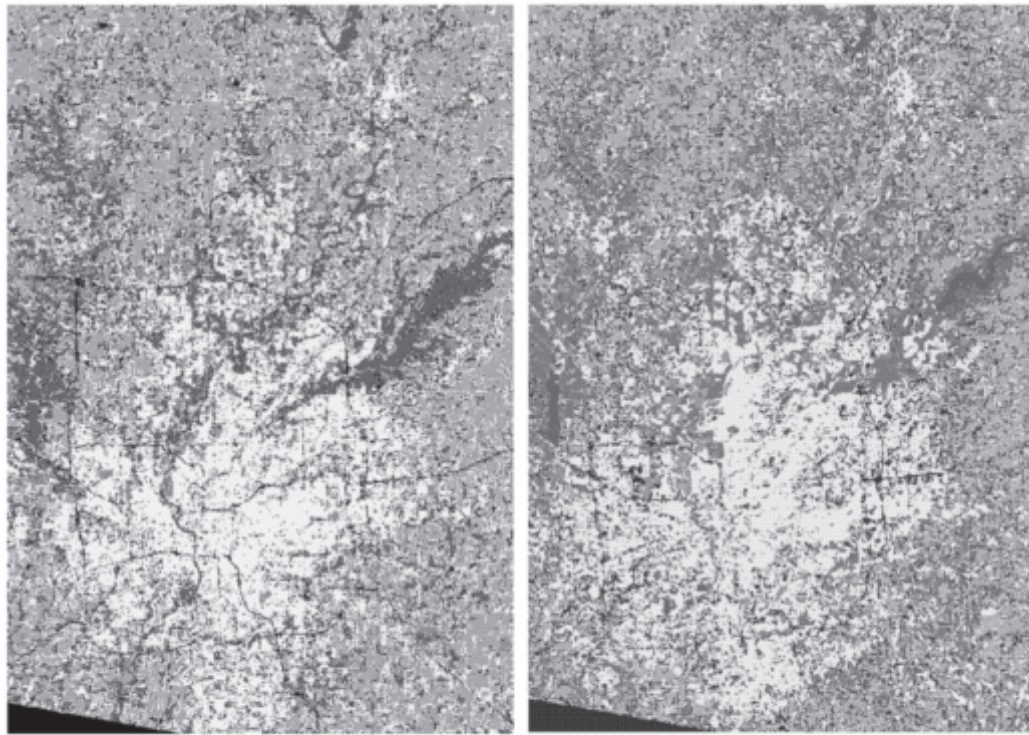
Township #	1982					1987				
	Rules (R,C,P)i	Fitness (%)	Type I Error (%)	Type II Error (%)	Total Error ΔE (%)	Rules (R,C,P)i	Fitness (%)	Type I Error (%)	Type II Error (%)	Total error ΔE (%)
1	5,4,0	91.4	64.7	16.6	26.7	5,6,0,2	148.8	56.9	17.5	22.6
2	0,0,0.1	96.9	60.5	17.7	26.8	8,5,0,3	110.0	52	16.8	23
3	4,4,0	94.9	40.1	21.8	27.3	7,8,0	97.5	40.1	20.6	26.1
4	0,0,0.1	99.0	39.7	20	25.4	6,8,0,1	99.5	43.2	19.1	25.2
5	3,2,0.2	93.5	55.8	19.2	28.2	7,5,0,8	113.8	45.3	18.3	23.6
6	3,3,0.3	91.0	50.4	19.2	27.5	6,8,0,5	99.6	43.6	18.9	25.1
7	3,6,0.3	88.3	41.1	24.3	30.3	4,8,0,6	96.5	36.6	26.1	30
8	5,4,0	87.9	43.9	17.4	24.4	7,8,0	102.6	49.5	19.7	26.8
9	4,8,0.2	91.4	50.1	17.7	27	5,3,1,8	88.7	32.8	22.5	26.3
10	4,4,0.4	90.5	26.6	30.5	28.4	4,4,2,2	96.9	16.4	36.2	24.1
11	4,2,0.5	92.2	24.7	33.7	28.8	5,3,1,5	98.5	15.2	38.6	24
12	3,2,0.5	89.2	50.9	17.7	26.1	4,4,0,8	96.4	47.4	23	30.2
13	3,7,2.6	91.3	22.8	25	23.9	4,5,2,9	97.6	17.1	34	24.9
14	5,2,2.3	90.1	14.8	30.9	19.4	5,3,3	97.4	8.4	48.4	17.3
15	4,1,2.8	100.5	2.4	57.9	11.5	4,4,0	99.6	1.9	63.8	10.2
16	3,1,1.5	92.4	24.5	24.7	24.6	6,4,3	93.6	24.3	24.7	24.5
17	4,2,2	90.7	17.3	30.2	22.8	5,5,3	96.7	17.3	36.1	25.4
18	1,0,0	99.8	0	76.4	8.8	6,4,0	96.0	3.2	58.7	10.5
19	4,4,2.8	93.0	10.1	45.6	17.6	7,5,0	99.2	7.9	42.8	16.5
20	4,7,0.4	90.2	40.2	23.4	29.5	8,7,1,8	97.2	38.4	24.3	29.1
21	3,6,1.7	90.3	34.3	24.7	28.4	8,8,2,9	118.8	27.9	25.2	26
22	3,8,3	104.0	19.8	39.2	28.9	6,8,3	96.3	21.1	31.8	26.1
23	3,2,2.9	94.2	23.7	30.5	27	8,8,0	105.0	18.4	28	23.7
24	4,7,0	90.3	47.4	26.4	33.4	8,8,0	125.6	45.4	26	30.9
Avg.		93.0	33.6	28.8	25.1		103.0	29.6	30.1	23.8

using the available ground truth in year 1982, 1987, 1992, and 2003. They present the evaluation results at a given calibrated or predicted year for each township according to its best calibrated rule set. As is shown, good fitness values (close to 100 percent) and total average errors in the range of

24 percent to 25 percent are achieved for most of the calibrated and predicted years. The images of the modeling results in Figure 5 through 8 demonstrate that an overall urban structure very close to what the city achieved in reality.

TABLE 2. 1992 AND 2003 PREDICTION RESULTS

Township #	1992				2003			
	Fitness (%)	Type I Error (%)	Type II Error (%)	Total error ΔE (%)	Fitness (%)	Type I Error (%)	Type II Error (%)	Total error ΔE (%)
1	102.4	77.6	11.4	19.9	117.1	79.6	12	19.5
2	116.7	67.6	15.6	24.1	81.7	70.6	13.5	24.7
3	86.2	55.8	20.1	32.2	97.6	42.6	25.6	32.4
4	98.5	57.3	19.4	29.4	95.6	51.8	23.4	32.7
5	95.0	64	15.5	26.1	105.0	50.4	18.2	26.1
6	82.8	55.8	16.2	28.7	99.4	33.8	34.9	34.4
7	94.2	40.5	28.2	34.1	116.1	19.9	56.3	34.2
8	91.3	61.7	18.9	30.3	131.3	24.3	46.2	36.2
9	139.4	26.8	42.3	35.5	93.4	24.7	23.6	24.1
10	119.8	8.7	55.3	21.8	93.5	12.8	41.3	18.2
11	111.1	9.9	58.2	20.5	109.7	2.8	64	13.2
12	101.6	41.5	30.4	35.4	102.4	20.3	43.3	28.6
13	105.5	17.6	35.5	23.3	112.9	4.9	57.3	18.9
14	110.4	5.5	59.6	13.2	103.4	3.9	57.7	11.9
15	111.4	0.9	68.3	5.6	100.8	1.1	61.4	5.9
16	116.1	24.9	30.4	27.8	79.1	29.2	20.4	25.9
17	91.5	23.9	32.5	26.2	117.7	4.9	64.3	18.8
18	106.0	5.4	52.1	8.7	105.6	0.6	78.5	9.0
19	100.6	12.9	39	17.6	106.4	3.2	74.5	14.3
20	107.6	43	22.6	29.3	71.1	43	16.4	29.1
21	75.7	47.7	13.9	27.6	117.3	34.3	37.3	36.1
22	101.9	24.1	28.1	25.8	95.6	20.3	36.7	25.6
23	76.7	36.1	17.1	28.6	104.0	14.1	59.3	25.4
24	103.2	59.5	20	29.8	79.2	53.5	20.5	32.3
Avg.	101.9	36.2	31.3	25.1	101.5	26.9	41.1	24.1

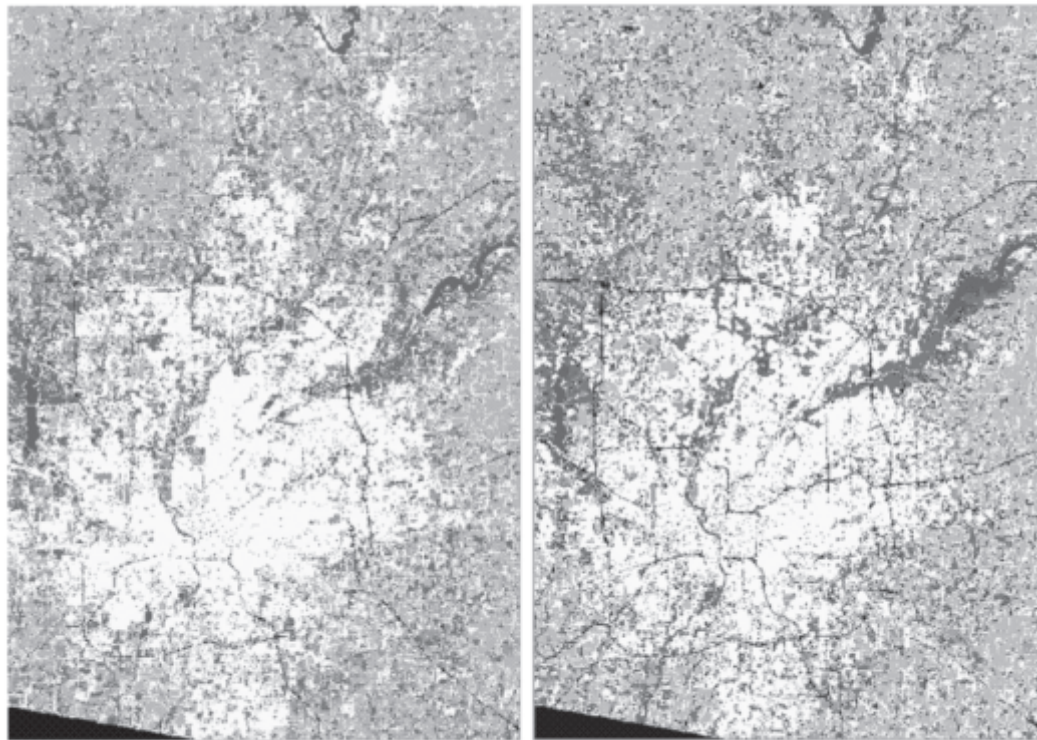


(a)

(b)

■ Unclassified ■ Water ■ Road □ Urban ■ Forest ■ Pasture ■ Rowcrops

Figure 5. (a) real, and (b) calibrated images for 1982. A color version of this paper is available at the ASPRS website: www.asprs.org.



(a)

(b)

■ Unclassified ■ Water ■ Road □ Urban ■ Forest ■ Pasture ■ Rowcrops

Figure 6. (a) real, and (b) calibrated images for 1987. A color version of this paper is available at the ASPRS website: www.asprs.org.

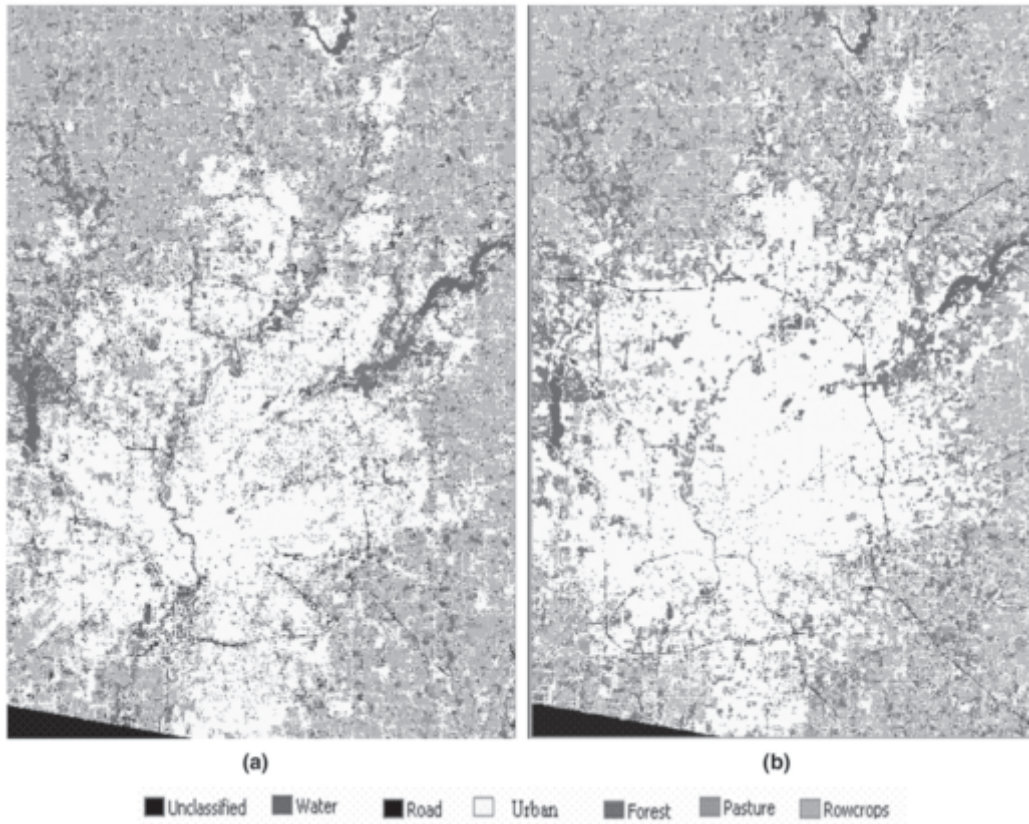


Figure 7. (a) real, and (b) predicted images for 1992 (from 1987). A color version of this paper is available at the ASPRS website: www.asprs.org.

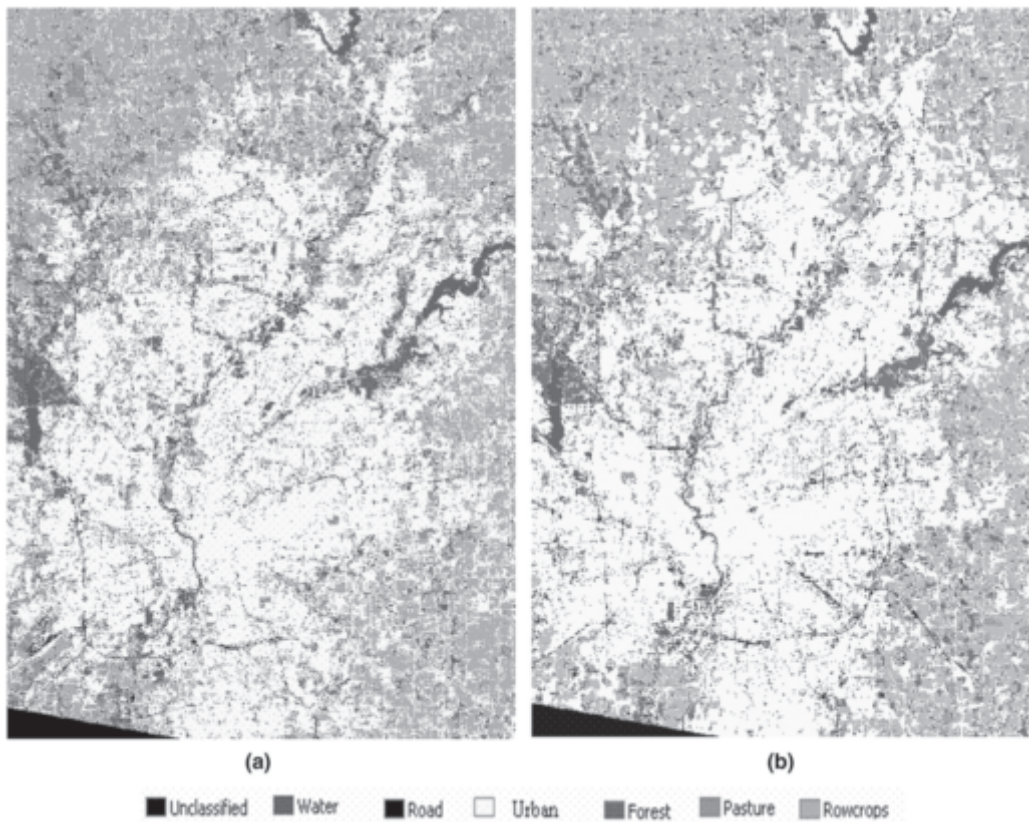


Figure 8. (a) real, and (b) predicted images for 2003 (from 1992). A color version of this paper is available at the ASPRS website: www.asprs.org.

As discussed earlier, the modeling errors represented by the total error are used as one criterion for the best rule selection. Minimizing such errors is important in cellular automata modeling to achieve urban pattern close to the ground truth. Since each error type represents the modeling quality for the class it is associated with (Type I for urban and Type II for non urban), there should be certain relationship between each class dominance level in the township and the corresponding error. Figure 9 shows the modeling errors in percentages for the calibrated year 1987 and the predicted year 1992. It is seen that the prediction modeling errors are slightly higher than the calibration errors. Also, the figure shows the corresponding class count associated with each error type. Type I error is more dominant in remote townships where the urban level is low, while Type II dominates the townships close to the city where the non urban count is small. This is related to the fact that each error type is evaluated with reference to the class it represents, thus a large class count will reduce the relative error value. The total error shows more spatial stability regardless of the township's location. This is clear in Figure 9, where the range of variation for the total error changes from 10 percent to 30 percent, with an

average of 24 percent. This is a small variation range if compared to the variation ranges of Type I or II errors that can be as low as 1 percent and as high as 80 percent. The total error takes into account the weight of each class referenced to the total count in the township. For this reason, the total error is selected as the criterion besides the fitness for rule value selection in the calibration process. Finally, spatial calibration plays a major role in matching the pattern between the modeling results and the reality. It finds the best set of rules that fits each township according to its urban level and pattern instead of using one rule set for the entire study area, which ignores the local effect of the urban process.

To further examine the compactness between the modeled and real images, Figure 10 shows a comparison using small windows (20×20 pixels) between the ground truth images in year 1987 and 1992 and their corresponding calibrated (1987) and predicted (1992) images. These test windows are distributed all over the images to cover different parts of the study area. Two remarks can be made based on Figure 10. First, the urban pixels in the modeling results are more connected compared to the high discontinuity and fragmentation in the real city data. This is the result of the

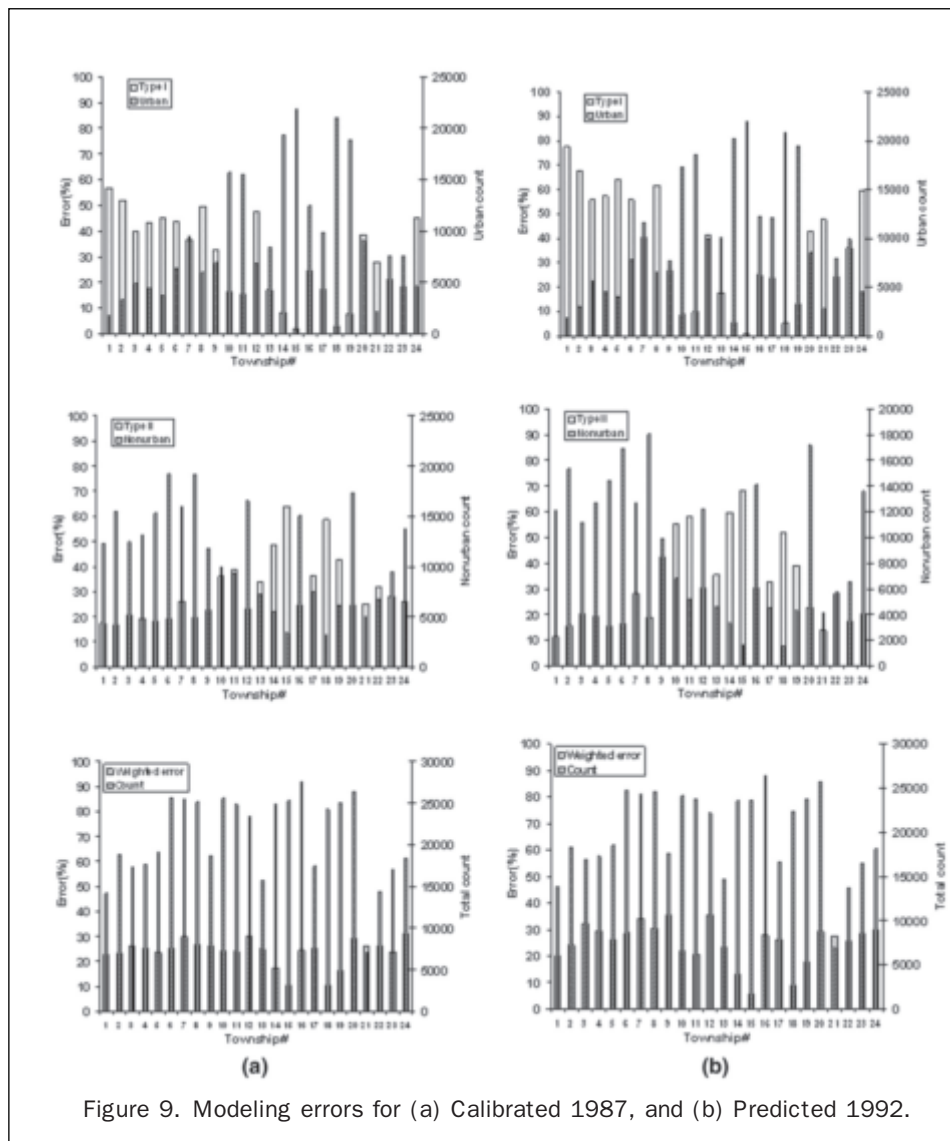
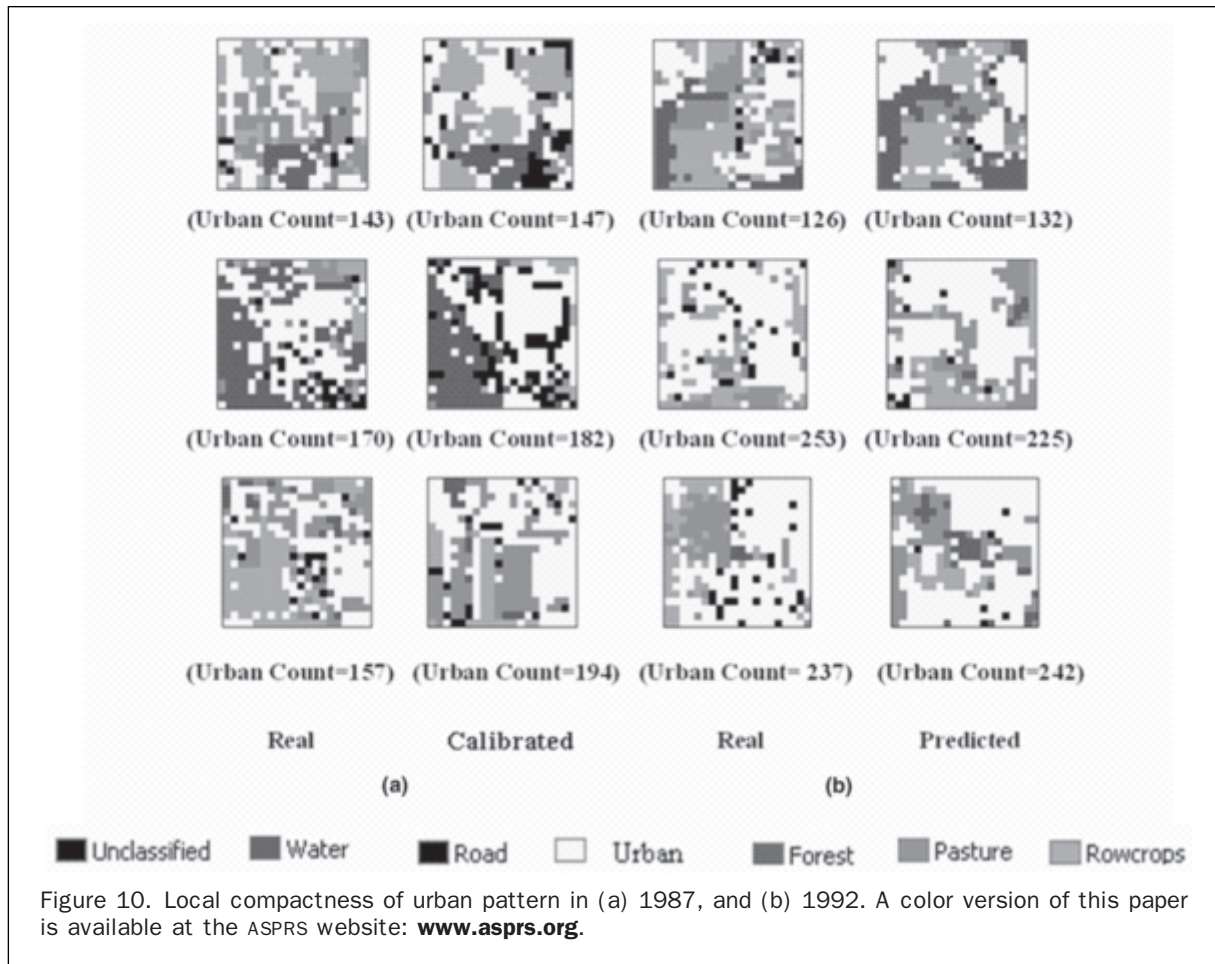


Figure 9. Modeling errors for (a) Calibrated 1987, and (b) Predicted 1992.



cellular automata modeling approach where pixels are checked in a sequential way through a connected set of neighborhood. Second, even if a very close urban count is achieved between the test windows, they may have different structures. This indicates that, in addition to the urban count match, there should be other evaluation measures that check the urban pattern match. This justifies that Type I and II measures or total error should be used as evaluation criteria.

Distribution of the Transition Rule Values

This section studies the spatial distribution of the calibrated rule values. Figure 11 shows the final calibrated rule values ((a) residential, (b) commercial, and (c) population density) for all townships for the calibration years (1982, 1987, and 1992). The residential and commercial rules have smaller values at townships close to the city center (townships 9 to 17) than other townships away from the city center. As can be seen in Figure 11, the residential rule values for townships close to the city vary between 3 to 5 while reaching high values, such as 7 and 8, for remote townships (e.g., townships 1 to 6 and 22 to 24 in year 1987 and 1992). The same observation is also clear for the commercial rules. Such results suggest that the residential and commercial rule values are inversely related to the township urbanization level. This is also evident in Figure 12, which shows the relation between rule values and urban count, where the smaller the urban count associated with each rule (e.g., commercial) the larger the rule values. Therefore, more restricted rules are needed for townships far from the city compared to the closer ones to produce realistic urban count

and pattern. As for the population density rule, there is no clear pattern as in the residential and commercial rules.

Correlation of the Transition Rule Values

This section will evaluate the correlation among the calibrated rule values. This is examined by using their variogram as a function of distance between townships. The variogram is a tool that quantifies the correlation of the dependent variable in terms of certain independent variables. For a given lag distance (h), the variogram is computed as the average squared difference of values separated approximately by h (Isaaks and Srivastava, 1989):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)} [z(u) - z(u+h)]^2 \quad (10)$$

where $\gamma(h)$ is the variogram value for a specified lag distance h ; $z(u)$ and $z(u+h)$ represent the variables for which the variogram is being computed at location u and $u+h$, respectively. $N(h)$ is the total number of pairs separated by h . In our work, the variogram is used to study the correlation in the rule values. Our evaluation is done with lag distance (h) defined as the distance between townships. At each lag distance, all the townships that have the same lag are listed, the differences between their residential rule values are identified and finally the average for the squared differences is calculated to find the variogram value. Figure 13 plots the variogram of the residential rule values for all lag distances among the townships. It is evident that there is a trend and significant degree of spatial correlation between the residential

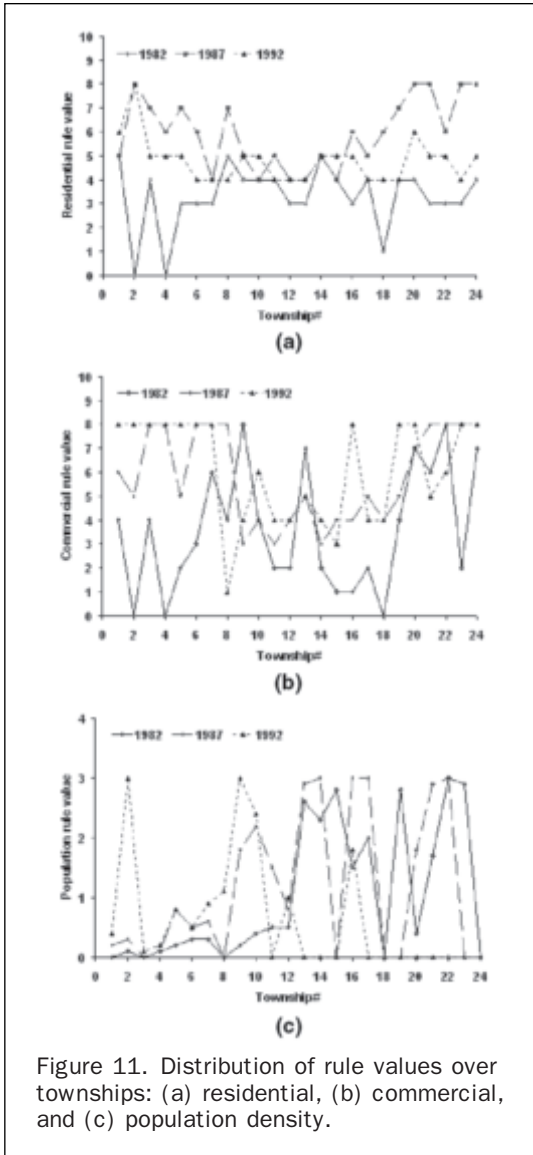


Figure 11. Distribution of rule values over townships: (a) residential, (b) commercial, and (c) population density.

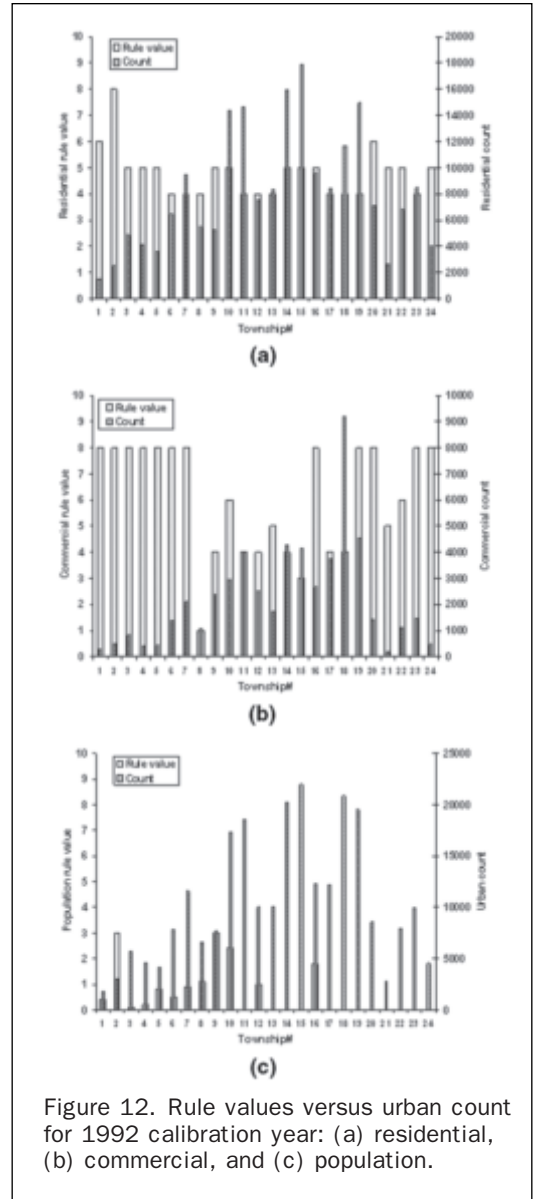


Figure 12. Rule values versus urban count for 1992 calibration year: (a) residential, (b) commercial, and (c) population.

rule values as a function of the distance between the townships. The closer the townships, the smaller the variogram values and hence the higher correlation. For townships far (> ~30km) from each other, the variogram shows an overall trend with larger values, which suggests a smaller correlation between the rule values among such townships.

Conclusions

This work presents a unique methodology for using cellular automata along with multitemporal imagery for urban growth modeling. Comparing to the exiting methods, it has the following distinctions. First, the use of multitemporal imagery can simplify the definition of the transition rules and limit the search space for the transition rule calibration. This is beneficial in reducing the need for large input data and the possible modeling uncertainty either directly from data or through error propagation. Besides, the fact that images record most of the land-cover features required for urban modeling supports such design philosophy. The availability of medium resolution satellite images at a minimal cost also makes such methodology particularly significant.

Calibration of transition rules is necessary to produce accurate modeling results in terms of urban level and pattern

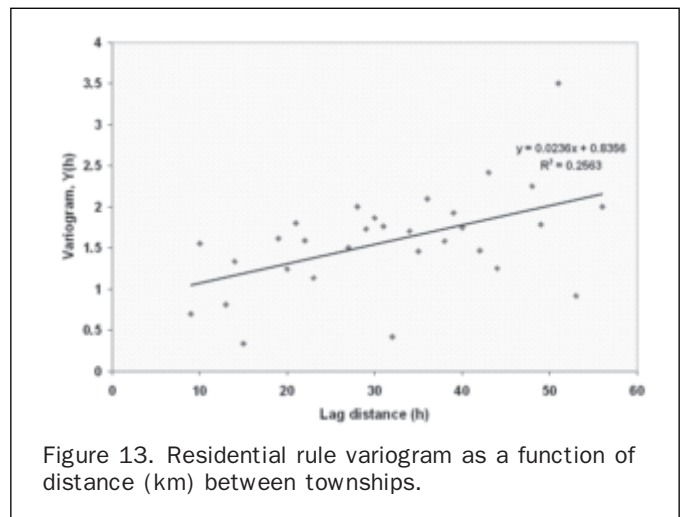


Figure 13. Residential rule variogram as a function of distance (km) between townships.

referenced to the reality. Spatiotemporal calibration introduced in this study proves to be efficient in achieving such objective. The “township-based spatial calibration” as another distinction of the proposed methodology can take into account the spatial variability in urban development. This succeeds in reducing the mismatch between the real and modeled urban patterns through assigning rule values that best fit the urbanization behavior for each township. The temporal calibration is necessary to refit the model to the urban growth variability over time, since some growth periods can experience faster or slower growth rates compared to other periods. The role of the spatiotemporal calibration in enhancing the modeling results strongly supports the argument that one set of rules for the whole study area may not be sufficient to represent the urbanization spatiotemporal variability.

Assessment methodology is essential for accurate urban modeling and model calibration. The proposed three quality measures, fitness (to check close urban count match) and two modeling errors (for urban pattern match check) prove to be very helpful in selecting the optimal rules. The total error, besides fitness, as a selection criterion shows unbiased evaluation measure that balances each error count with its corresponding class. Good fitness ensures that a close urban level be achieved, while minimizing the modeling errors can produce the realistic spatial urban structure.

The transition rule values present clear correlation or relationship with their geographical location and urbanization level. It is found that areas off the city require more restrict rules to match the lower urban development rate. Townships that are close spatially or have similar urban development level will receive similar transition rule values. This property can be used to estimate or search for the optimal rule values for spatial units (townships in this case) with similar characteristics, such as urbanization level or geometric location.

As an ongoing study, we are improving the search method for the best calibrated rule values. Instead of using the brute force method to search the entire solution space, genetic algorithm is expected to optimize the search strategy. Integrating genetic algorithm with cellular automata will likely make the current calibration process more efficient.

(Note: This work already collected in the *PE&RS* Special Issue on Artificial Intelligence in Remote Sensing; readers may refer to Shan *et al.*, 2008.)

References

- Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, USGS Professional Paper 964, Sioux Falls, South Dakota.
- Batty, M., and Y. Xie, 1994a. From cells to cities, *Environment and Planning B*, 21:531–548.
- Batty, M., and Y. Xie, 1994b. Modelling inside GIS: Part 2 – Selecting and calibrating urban models using ARC-INFO, *International Journal of Geographical Information Systems*, 8:451–470.
- Batty, M., Y. Xie, and Z. Sun, 1999. Modelling urban dynamics through GIS-based cellular automata, *Computers, Environment and Urban Systems*, 23:205–233.
- Clarke, K.C., S. Hoppen, and L. Gaydos, 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area, *Environment and Planning B*, 24:247–261.
- Clarke, K.C., and L.J. Gaydos, 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore, *International Journal of Geographical Information Sciences*, 12:699–714.
- Codd, E.F., 1968. *Cellular Automata*, Academic Press, New York.
- Hagerstrand, T., 1967. *Innovation Diffusion as a Spatial Process*, University of Chicago Press, Chicago, Illinois.
- Isaaks, E.H., and R.M. Srivastava, 1989. *Applied Geostatistics*, Oxford University Press, 592 p.
- Jacobs, J., 1961. *The Death and Life of Great American Cities*, Random House, New York.
- Landis, J., 1992. *BASS II: A New Generation of Metropolitan Simulation Models*, WP-573, Institute of Urban and Regional Development, University of California at Berkeley, Berkeley, California.
- Li, X., and A.G.O. Yeh, 2001. Calibration of cellular automata by using neural networks for the simulation of complex urban systems, *Environment and Planning A*, 33:1445–1462.
- Li, X., and A.G.O. Yeh, 2003. Error propagation and model uncertainties of cellular automata in urban simulation with GIS, *Proceedings of the 7th International Conference on GeoComputation*, 08–10 September, University of Southampton, Southampton, UK, unpaginated CD-ROM.
- Meaille, R., and L. Wald, 1990. Using geographical information systems and satellite imagery within a numerical simulation of regional urban growth, *International Journal of Geographic Information Systems*, 4:445–456.
- Pijanowski, B.C., D.T. Long, S.H. Gage, and W.E. Cooper, 1997. A land transformation model: Conceptual elements, spatial object class hierarchies, GIS command syntax and an application for Michigan’s Saginaw Bay Watershed, URL:<http://www.ncgia.ucsb.edu/conf/land-use97> (last date accessed: 18 August 2008).
- Portugali, J., I. Benenson, and I. Omer, 1997. Spatial cognitive dissonance and sociospatial emergence in a self-organizing city, *Environment and Planning B*, 24:263–285.
- Shan, J., S. Alkheder, and J. Wang, 2008. Genetic algorithms for the calibration of cellular automata urban growth modeling, *Photogrammetric Engineering & Remote Sensing*, 74(10):1267–1277.
- Sipper, M., 1997. *Evolution of Parallel Cellular Machines: The Cellular Programming Approach*, 03–08, Springer-Verlag, Heidelberg.
- Takeyama, M., and H. Couclelis, 1997. Map dynamics: Integrating cellular automata and GIS through Geo-Algebra, *International Journal of Geographical Information Sciences*, 11:73–91.
- Tobler, W., 1979. *Cellular Geography*, in *Philosophy in Geography*, (S. Gale and G. Olsson, editors), D Reidel, Dordrecht, pp. 379–386.
- Veldkamp, A., and L.O. Fresco, 1996. CLUE: A conceptual model to study the conversion of land-use and its effects, *Ecological Modelling*, 85:253–270.
- von Neumann, J., 1966. *Theory of Self-Reproducing Automata* (A.W. Burks, editor and contributor), University of Illinois Press.
- Waldrop, M.M., 1992. *Complexity: The Emerging Science at the Edge of Order and Chaos*, New York: Simon and Schuster.
- White, R., and G. Engelen, 1992a. *Cellular Dynamics and GIS: Modelling Spatial Complexity*, WP-9263, Research Institute for Knowledge Systems (RIKS), Maastricht, The Netherlands.
- White, R., and G. Engelen, 1992b. *Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land Use Patterns*, WP-9264, RIKS, Maastricht, The Netherlands.
- Wilson, A., 1978. Spatial interaction and settlement structure: Toward an explicit central place theory, *Spatial Interaction, Theory and Planning Models* (A. Karlqvist, L. Lundqvist, F. Snickars, and J.W. Weibull, editors), North-Holland, Amsterdam, 137–156.
- Wolfram, S., 1986. *Theory and Applications of Cellular Automata*, Singapore: World Scientific.
- Wolfram, S., 1994. *Cellular automata, Cellular Automata and Complexity: Collected Papers*, Addison Wesley, Steven Wolfram, Reading, Massachusetts.
- Wu, F., and C.J. Webster, 1998. Simulation of land development through the integration of cellular automata and multi-criteria evaluation, *Environment and Planning B*, 25:103–126.
- Wu, F., 2002. Calibration of stochastic cellular automata: The application to rural-urban land conversions, *International Journal of Geographical Information Science*, 16:795–818.
- Yang, X., and C.P. Lo, 2003. Modelling urban growth and landscape changes in the Atlanta metropolitan area, *International Journal of Geographical Information Science*, 17:463–488.

(Received 03 January 2007; accepted 18 April 2007; revised 28 June 2007)