FUZZY CELLULAR AUTOMATA APPROACH FOR URBAN GROWTH MODELING

Sharaf Alkheder Jun Wang Jie Shan
Geomatics Engineering, School of Civil Engineering,
Purdue University, 550 Stadium Mall Drive,
West Lafayette, IN 47907, USA, Phone: +1-765-494-2168, Fax: +1-765-496-1105
{salkhede, wang31, jshan}@ecn.purdue.edu

ABSTRACT

This work focuses on adapting artificial intelligence techniques for urban growth modeling using multitemporal imagery. Fuzzy set theory and cellular automata are used for this purpose. Fuzzy set theory preserves the spatial continuity of the growth process through allowing a test pixel to be partially developed unlike the binary crisp system (developed/undeveloped). The development level identified by the fuzzy system is incorporated in cellular automata transition rules definition to identify the urban neighborhood threshold, beyond which a test pixel is allowed to develop. The fuzzy system parameters are defined based on the effect each input variable has on the urban growth potential. Synthetic data are used to present and test the principles of using fuzzy cellular automata for artificial city growth modeling. Using the developed fuzzy cellular automata algorithm, the growth of Indianapolis over three decades is studied. In addition to satellite images, digital elevation model, road networks and population data are included in the modeling. The transition rules of the cellular automata are calibrated spatially and over time using multitemporal images. Modeling results are evaluated on a township basis for accuracy assessment. Results indicate the role of fuzzy system in preserving the spatial continuity of the growth process. Partial development concept introduced by the fuzzy system takes into account the contribution of the test pixel in identifying the required urban neighborhood for its development. The results indicate also the importance of spatial calibration based on township specific features in improving the prediction results. Temporal calibration of the model adapts the dynamic growth pattern over time to improve the modelling accuracy. Careful definition of fuzzy parameters is crucial for accurate modeling. Stable accuracy results for prediction are achieved over time. Average accuracies of 95.81% for short term prediction (5 years) at 1992 and 95.49% for long term prediction (11 years) at 2003 are obtained.

INTRODUCTION

Urban growth is a complex process resulting from an interaction between numbers of growth variables. The complexity nature of this process makes it extremely hard to identify it using a mathematical model. As one of the artificial intelligence techniques, Cellular Automata (CA) gains recently a lot of popularity in modeling the growth process due to the many success stories achieved (Batty and Xie, 1994a; White and Engelen, 1997; Wu and Webster, 1998; Li and Yeh, 2000). Most of the existing CA models are based on crisp set theory, where each pixel is tested using a binary system (developed or undeveloped). However, the urban growth process is continuous in space and a pixel can be fully developed or partially developed (undeveloped). Many researchers have argued that the traditional crisp approach is not appropriate for representing geographic data and entities in GIS databases (Leung, 1987; Burrough and Frank, 1996). The process of urban development resembles a fuzzy process both spatially and temporally (Liu and Phinn, 2003).

Fuzzy set theory is first introduced by Zadeh (1965; 1971) as an extension of the binary crisp set theory to cover the continuous classifications case. The classes of elements are represented as gradual transitions where the fuzzy set provides a way to express the degree of membership to a particular class or set. Fuzzy set theory is useful to represent geographic data, entities and the boundaries in GIS framework (Wang and Hall, 1996). Wu (1996; 1998), define transition rules based on fuzzy logic concept, where the state of cells (pixels) is defined as non-urban or urban under the crisp set theory without addressing the multiple or fuzzy characteristics of non-urban, partly-urban and urban states in the process of urban development (Liu and Phinn, 2003). Liu and Phinn (2003) delimit urban areas using a fuzzy membership function. Different transitions are defined where each pixel is tested to check to which transition it belongs and hence associated rules for development are used. This approach needs to define the growth patterns within the study area, which in most cases is not possible. Also dividing the growth patterns to finite number of transitions restricts the continuity that urban growth process has over space. So the need arises to define a continuous pattern of spatial urban development where the pixel can be classified based on its development level without the need of assuming such transition patterns. Manetos and Photis (2003) used the fuzzy logic in the data preparation stage for multivariate object clustering, such that each object can belong to more than one cluster.

Our work focuses on urban growth modeling based on partial development concept by designing a fuzzy cellular automata algorithm. The growth pattern is defined as a continuous function based on which the test pixel's development level can be classified according to its input values. The development level of each test pixel is identified as a function of the activated fuzzy rules of the input variables. The development level is transformed through the defuzzification process to the required number of urban pixels in the neighborhood for any test pixel to develop. The fuzzy system output is used as an input to the cellular automata to identify the best urban neighborhood needed for accurate modeling. The approach is evaluated by a set of Thematic Mapper (TM) satellite imagery covering Indianapolis area, Indiana, USA over 30 years. Model evaluation and calibration is performed spatially and temporally on a township basis.

PRINCIPLES OF FUZZY CELLULAR AUTOMATA

This section describes the principles of designing the fuzzy CA algorithm for urban growth modeling. It covers the development phase using a synthetic framework to simulate the growth of artificial city. This stage includes the definition of the fuzzy membership functions for the selected input and output data, designing the fuzzy rules based on the defined membership functions, selection of the fuzzy model parameters, and finally CA transition rules definition for dynamic growth modeling.

For the purpose of algorithm design, a synthetic framework is used. Three inputs shown in Figure 1 are used to feed the fuzzy system part. A 200x200 image is used as the first input representing a land use structure of an artificial city containing 6 classes namely: lake, river, road, urban, non-urban and pollution sources. Normalized digital elevation model (DEM) is used as another input. The third input represents the normalized distance from the city center. The next step is the definition of the fuzzy membership functions for each of the inputs and the output. Membership functions simulate the continuous fuzzy degrees of effect for each variable. For example, distance variable can be classified as close, medium or far according to the location of the pixel from the city center. Figure 2 describes these membership functions for all the inputs and output data. For the normalized distance input, 3 membership functions are used: close, medium and far. Similarly, the normalized DEM input is also categorized to 3 membership functions according to height level: low, medium and high elevations. The output variable represents the development level of the test pixel based on which the required number of neighborhood urban pixels for development is identified as an input for the CA model. The output membership functions are defined based on the development level as high, medium and low. As an example, high membership function category for the output indicates that the test pixel has high development level so that the number of urban pixels needed in the neighborhood for it to develop is small as compared to other pixels with lower development potential

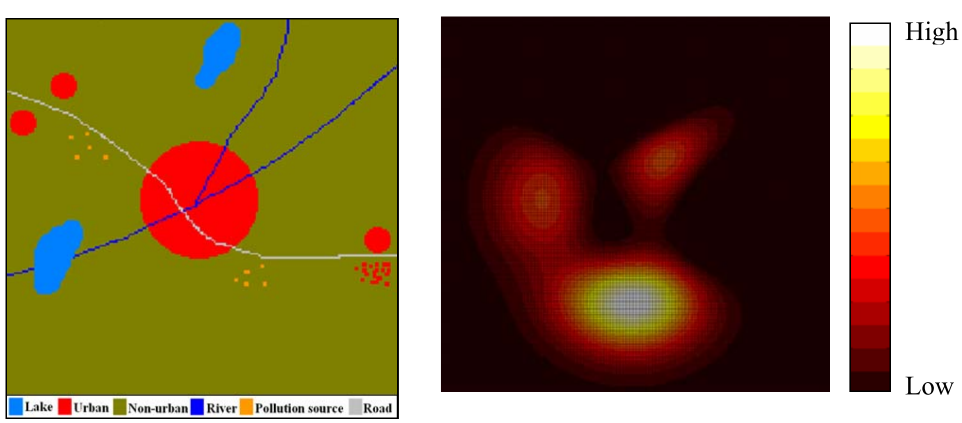


Figure 1. Synthetic city (left) and its elevation (right)

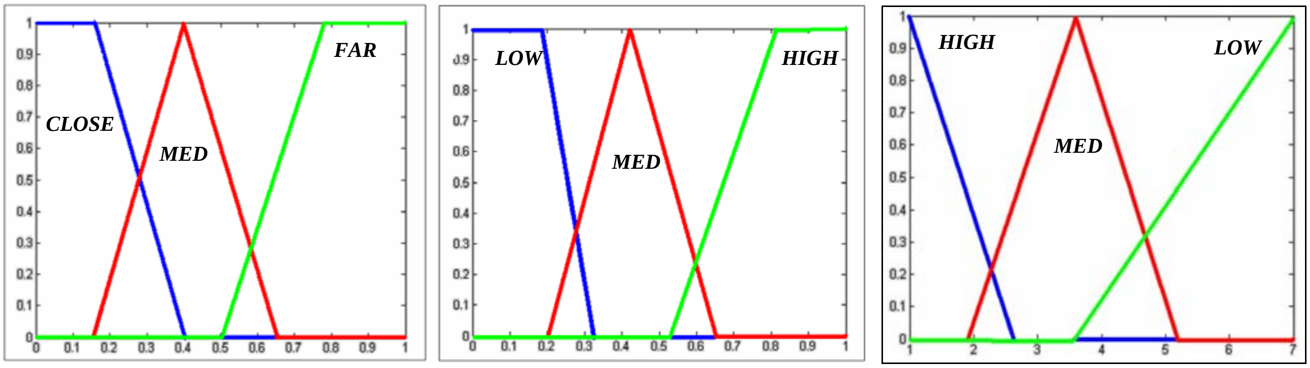


Figure 2. Membership functions definition for distance (left), elevation (middle) and output (right)

Once the membership functions are defined, the fuzzy rules are designed based on the expectation that each input variable has on the output. Fuzzy rules for the synthetic fuzzy model are summarized as a rule table in Table 1. As shown, seven rules are defined based on the two input variables (DISTANCE & DEM) influence on the output.

Table 1. Output of the membership functions

DEM / DIST	LOW	MED	HIGH
CLOSE	**1**	**1**	**2**
MEDIUM	**0**	**2**	**3**
FAR	**0**	**3**	**3**

Numbers in the table refer to the membership functions of the output variable namely: High (1), Medium (2), Low (3) and No rule (0). For example, if a test pixel has low DEM height and close distance to city center, then it has high development level and hence needs small number of urban neighborhood pixels to develop. Minimum-maximum (Mamdani method, Tsoukalas and Uhrig, 1997) method is used for the fuzzification-defuzzification process to compute the output variable based on the input variables' activated membership functions as shown in Equation 1. The fuzzy output is identified as a function of the input data, fuzzy membership functions and defined fuzzy rules. The fuzzifcaition-defiuzzification process is described below. For each test pixel, the DEM and DISTANCE values are calculated and entered to the corresponding fuzzy rules' membership functions to check which rules are activated. The fuzzy output $\mu_{out}(i, j)$ for the test pixel is calculated as in Equation 2 where the maximum of the minimum values for the activated rules is used to represent a single point in the development level fuzzy graph of the test pixel. The test pixel is checked over all the 7 rules (k=7) where for each rule the membership functions' values of DEM $\mu_{DEM}(h_{i,j})$ and DISTANCE $\mu_{DIST}(x_{i,j})$ are calculated among which the minimum value is used to activate the related membership function in the output variable. The last step is to defuzzify the fuzzy output graph to a single value representing the required number of urban pixels in the neighborhood to develop the test pixel. Center of Area (COA) method (Tsoukalas and Uhrig, 1997) is used for this purpose as can been seen in Equation 3 where u_i represents the x-coordinate for the fuzzy output function, u^* is the output value and N is all the fuzzy graph points.

$$Fuzzy\_output = f(input\_data, memebership\_functions, fuzzy\_rules) \quad (1)$$

$$\mu_{out}(i, j) = \max_{k=1....7}\{\min_k[\mu_{DEM}(h_{i,j}), \mu_{DIST}(x_{i,j})]\} \quad (2)$$

$$u^* = \frac{\sum_{i=1}^{N} u_i^* \mu_{out}(u_i)}{\sum_{i=1}^{N} \mu_{out}(u_i)} \quad (3)$$

The defuzzified output represents the required urban neighborhood pixels' number for a test pixel to develop to urban. It is used as an input to the cellular automata transition rules:

- IF the test pixel state is river, road, lake or pollution source THEN no growth.
- IF the state of test pixel is urban, THEN keep it urban.
- IF the test pixel is non-urban then there are 5 possible cases:
    - IF it is non-urban AND one or more of neighborhood pixels are pollution source THEN keep non-urban.
    - IF it is non-urban AND # of urban pixels in the neighborhood is $\geq$ than fuzzy output AND no pollution pixel THEN change it to urban.
    - IF it is non-urban AND one or more of the neighborhood pixels are road AND # of urban pixels in the neighborhood is $\geq$ two pixels less than the fuzzy output AND no pollution pixel, THEN change it to urban.
    - IF it is non-urban AND one or more of the neighborhood pixels are lake AND # of urban pixels in the neighborhood is $\geq$ two pixels less than the fuzzy output AND no pollution pixel, THEN change it to urban.
    - ELSE keep non-urban.

The cellular automata model is run based on the above defined rules and the urban growth is simulated at growth steps of 0, 25, 50 and 60. Results are shown in Figure 3.

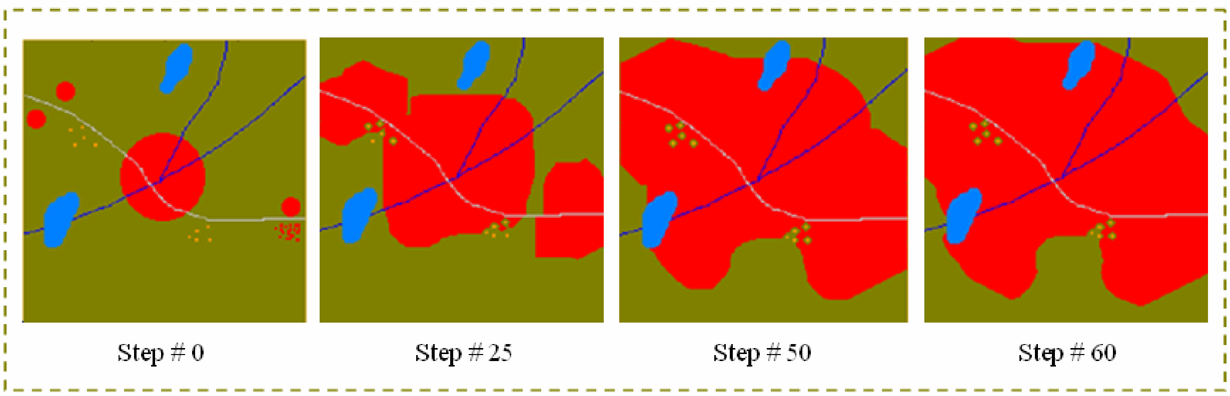


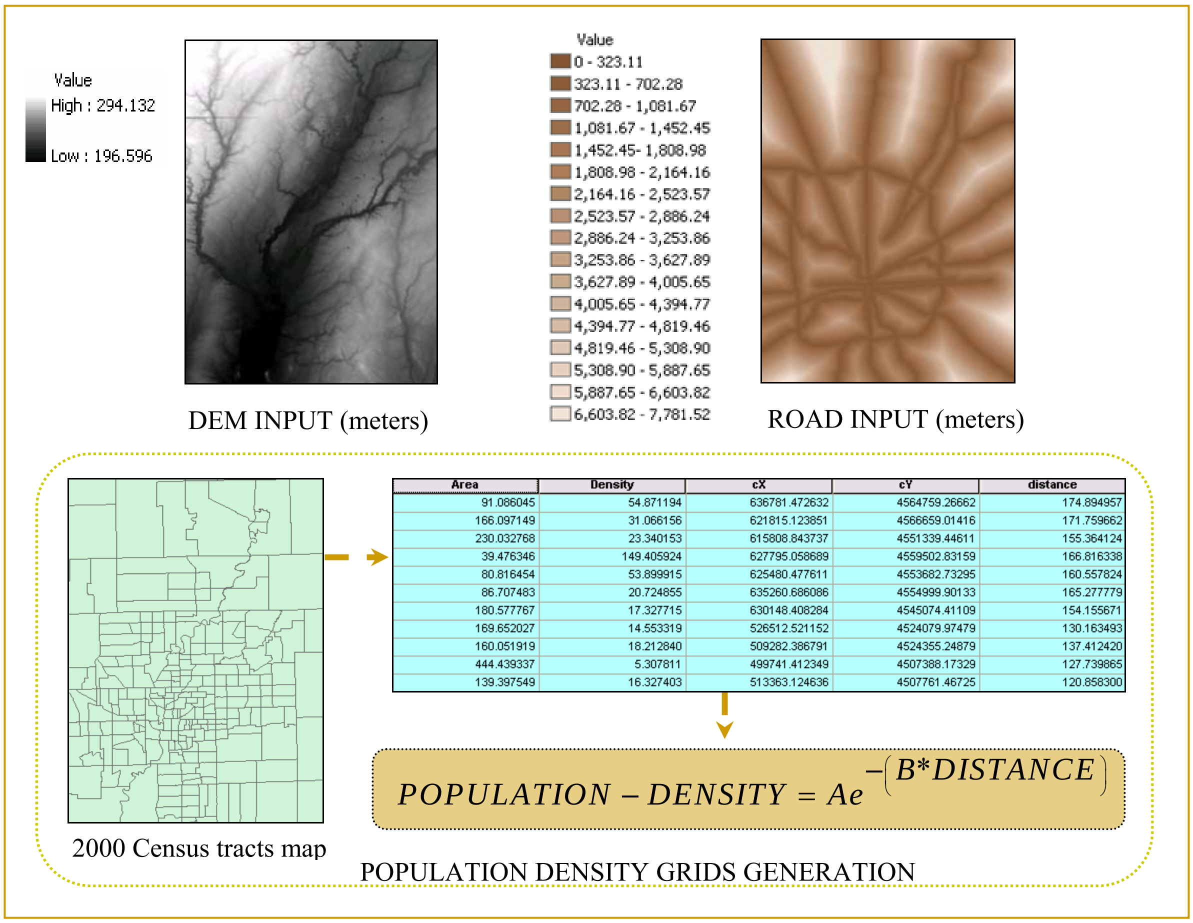
Figure 3. Fuzzy cellular automata synthetic modeling results

The results clearly show the effect of the fuzzy CA transition rules definition on the simulated urban growth. As can be seen in the lower part of the images in Figure 3, those regions of the city characterized by high DEM heights where the growth is very difficult - as defined in the fuzzy rules - turn out to be undeveloped. This indicates that previous knowledge about the nature of the problem we are trying to solve is necessary to select the best fuzzy rules defining the effect of input variables on the output. By this step the fuzzy cellular automata design is completed and the next step is to implement it to more complex growth environment represented by the real city.

URBAN GROWTH MODELING OF INDIANAPOLIS

In the next stage, the developed algorithm is used to model the growth over Indianapolis, IN, USA. For the real city modeling, a number of modifications are implemented on the model to fit the real data specific features. Selection of the input variables, fuzzy rules and CA transition rules definition, fuzzy parameters identification, designing evaluation methodology for modeling results and finally spatial and temporal modification of the CA rules all are tested and calibrated for accurate growth modeling and prediction.

Multitemporal Thematic Mapper (TM) satellite imagery represents the space over which the model works. Five of these images are collected over Marion County covering a period of 30 years with the aim to model the growth process spatially and temporally. Images are spatially registered to the same reference system of Universal Transverse Mercator (UTM) NAD 1983. Seven classes of interest are specified in the images namely, water, road, residential, commercial, forest, pasture and row crops. These classes are identified using high resolution orthophotographs and Indiana Geological Survey land classification maps as ground reference. Besides the imagery, three data sets are used for this study: DEM, road network and census tract population maps as shown in Figure 4. The DEM is normalized to scale the real heights between 0 and 1. An image showing the distances to the closest road is created from the road input and then normalized. The centroids for the whole census tracts map and for each tract are calculated. The distance is computed between each census tract centroid and the overall centroid. Population density is calculated for each census tracts and an exponential function is fitted between the population density and the distance variables. The population density is calculated for each pixel in the input satellite imagery using the exponential function as a function of its location from the downtown. The population density model parameters are modified over time and a population density grid is produced for each year of the simulation time period. A threshold value of the population density is selected based on the spatial extension of the city on a yearly basis to be used in transition rules definition.



Area	Density	cX	cY	distance
91.086045	54.871194	636781.472632	4564759.26662	174.894957
166.097149	31.066156	621815.123851	4566659.01416	171.759662
230.032768	23.340153	615808.843737	4551339.44611	155.364124
39.476346	149.405924	627795.058689	4559502.83159	166.816338
80.816454	53.899915	625480.477611	4553682.73295	160.557824
86.707483	20.724855	635260.686086	4554999.90133	165.277779
180.577767	17.327715	630148.408284	4545074.41109	154.155671
169.652027	14.553319	526512.521152	4524079.97479	130.163493
160.051919	18.212840	509282.386791	4524355.24879	137.412420
444.439337	5.307811	499741.412349	4507388.17329	127.739865
139.397549	16.327403	513363.124636	4507761.46725	120.858300

$$POPULATION - DENSITY = Ae^{-\left(B*DISTANCE\right)}$$

Figure 4. Input data for Indianapolis growth modeling

The membership functions are defined in the same way presented for the artificial city case for each of the three input variables. The output of the fuzzy analysis representing the development level of a test pixel is converted to the needed neighborhood urban pixels' number for the pixel to develop. Three membership functions are defined for each variable based on its degrees of fuzzy effects as shown in Figure 5.

The definition of the fuzzy rules is done based on the effect of input data on identifying the urban neighborhood. The following list is a summary of these fuzzy rules:

- IF test pixel has high population density (PD) THEN it has high development level (DL).
- IF test pixel is close to road THEN it has medium DL.
- IF test pixel has high PD AND medium distance to road THEN output DL is medium.

- IF test pixel has high PD AND far distance from road THEN medium DL.
- IF PD is medium AND road is medium THEN DL is medium.
- IF PD is low AND road is medium THEN DL is medium.
- IF PD is low AND road is far THEN DL is low.
- IF PD is high AND DEM is low THEN DL is high.
- IF PD is medium AND DEM is medium THEN DL is high.
- IF PD is low AND DEM is high THEN DL is low.
- IF PD is med AND DEM is high THEN DL is medium.
- IF PD is high AND DEM is high THEN DL is high.
- IF PD is medium AND DEM is low AND road is far THEN DL is medium.

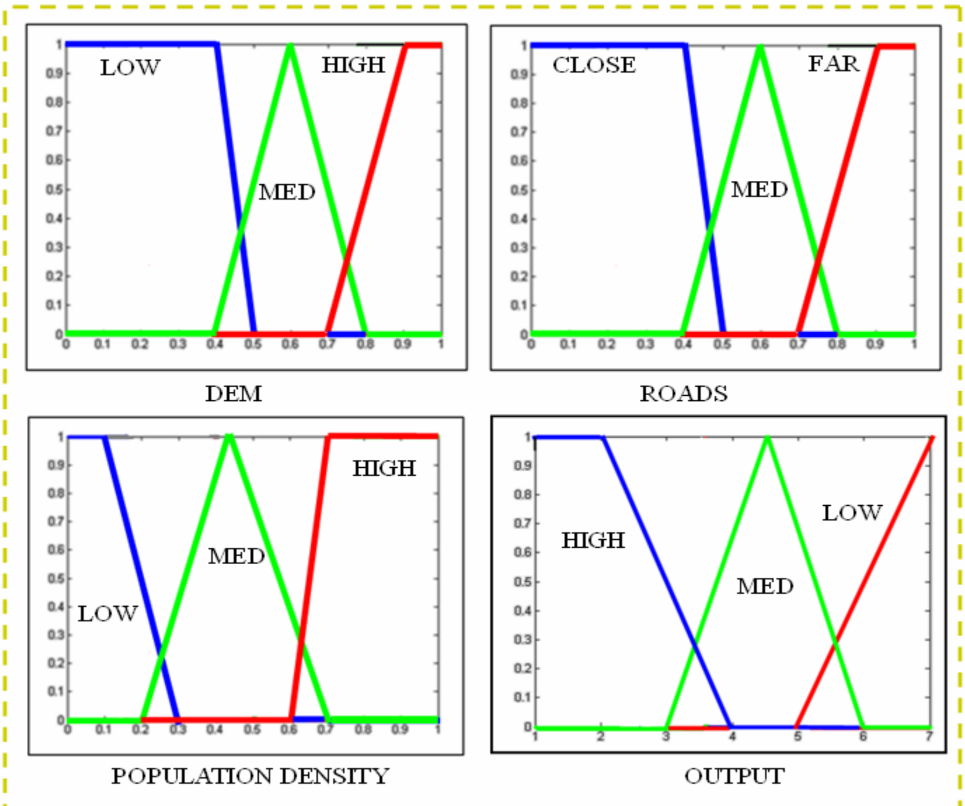


Figure 5. Membership functions for Indianapolis study

Fuzzy development level output is converted after defuzzification process to crisp value, which represents how many urban pixels in the neighborhood are needed for a test pixel to develop to urban. For example, a pixel with high development level required lower number of pixels to develop compared to other pixels that have low development level. This crisp development level value is used through the CA transition rules as a threshold beyond which the pixel is guaranteed development, otherwise no development is allowed. The fuzzy system output and the multitemporal imagery will be used as an input for the cellular automata calculation.

The fuzzy cellular automata algorithm starts from the oldest image of 1973 and continues the prediction till the first ground truth image of 1982 used for rules calibration. The calibration is performed on a township basis, i.e., the predicted image and ground truth image are compared in the corresponding areas defined by a township polygon. Accordingly, the transition rules are modified spatially for each township. For example, the rules for a specific township that underestimate the urban growth are tuned to increase the growth for this region. Once township based spatial rules are calibrated, the simulation is started again form 1973 till 1982 as second iteration. This process of calibration continues till the simulated image of 1982 is sufficiently close to the 1982 ground truth. Once this happens, another calibration is performed at the next ground truth for 1987 to consider the temporal variation of the growth process. The same spatial calibration method is repeated at 1987 till the best calibrated rules are obtained. The final calibrated rules at 1987 are used to predict the growth at 1992 (short term, 5 years). Growth at 2003 (long term, 11 years) is also predicted based on recalibrated rules at 1992. The predicted results are compared with the ground truth at 1992 and 2003 respectively. Figure 6 shows the ground truth and predicated images, while Table 2 list the numerical accuracy assessment for prediction results.

DISCUSSIONS

The developed fuzzy cellular automata model uses the advantage of fuzzy set theory regarding continuous classifications in defining the partial development principle of test pixel for urban growth modeling to preserve the continuous spatial nature of the growth process. Unlike the traditional crisp CA approaches where the test pixel state is defined using a binary system (developed/undeveloped), fuzzy CA approach takes the pixel development level into account. Our model distinction from other models appears in the algorithm structure. The model uses a number of input variables through predefined fuzzy parameters (membership functions, fuzzy rules and defuzzification method) to identify the fuzzy development level of the test pixel. The development level is defuzzified to a crisp output value of the needed urban neighborhood pixels' number for the pixel to develop. It designs the threshold for pixel development in its future state based on the input data. Pixels that meet this threshold based on their development status as approved by the CA transition rules are allowed to develop. The structure of our model doesn't assume any previous knowledge about the urban growth patterns within the study area.

The output of the fuzzy system is used as one of the input variables to the CA model part to drive the growth process over time. The CA rules define the best neighborhood structure based on the fuzzy output, land use data and constraints for accurate growth modeling. New features that our algorithm considers to improve the prediction accuracy are: spatial and temporal calibration. Transition rules are calibrated spatially on township basis to consider the spatial difference in the urban growth as a function of site specific features. Spatial calibration proves to be efficient in improving the prediction accuracy and reducing the variability of spatial growth results where transition rules are tuned to adapt the growth pattern at specified township area. Temporal calibration over time contributes significantly in obtaining a good prediction results. Temporal calibration helps the designed algorithm to learn the dynamic growth pattern over time so that the rules calibration module can be adapted to such changes. This is important since some growth periods experienced excessive or slower growth rates which can be detected and hence the model can modify itself to take them into consideration. Calibration results at 1982 and 1987 show that it takes 7 iterations till the simulated township results predict accurately the corresponding ones in the ground truth data. After each iteration, that represents new spatial tuning of the CA rules based on township specific site features, the improvement in the modeling accuracy can be easily noticed till the design criteria of convergence (±10%) is achieved.

The prediction results for short term prediction of 5 years time interval at 1992 show good accuracy on township basis where the predicted growth match closely the real one with average accuracy of 95.81%. Long term prediction results for 11 years interval at 2003 shows also good results with average accuracy of 95.49%. This indicates the stability of model prediction capability over time interpreted by the role of the temporal calibration element in adapting the changes in growth dynamics. The multitemporal imagery is an excellent medium that fits the CA grid nature over which the growth process is simulated besides its richness of spatial information that is used as an input for fuzzy CA system design. All the work of fuzzy CA modeling is done in ArcGIS VBA programming environment where the analytical results are coupled with visual evaluation.

It should be noted that prediction results of township 1 are highly overestimated. Although the exact cause is unclear, the ground truth from image classification suggests that there is de-urbanization over the years in this township. Since our model is designed for growth only and is unable to take deurbanization into account, the prediction results for this region are then overestimated. One of our future targets is to extend the model capabilities to consider zero or negative growth. The other possible cause is misclassification. Misclassification at this region, where non-urban pixels with similar spectral response as urban pixels can be classified as urban, can also result in this problem. Since the simulated image is evaluated with reference to the ground truth images; any misclassification in the ground truth at this region can results in such overestimation. Misclassification issue will be one of our future efforts through studying the effect of misclassification uncertainty on cellular automata growth modeling results. Error propagation and data adjustment principles can be used for qualitative assessment for uncertainty issues.

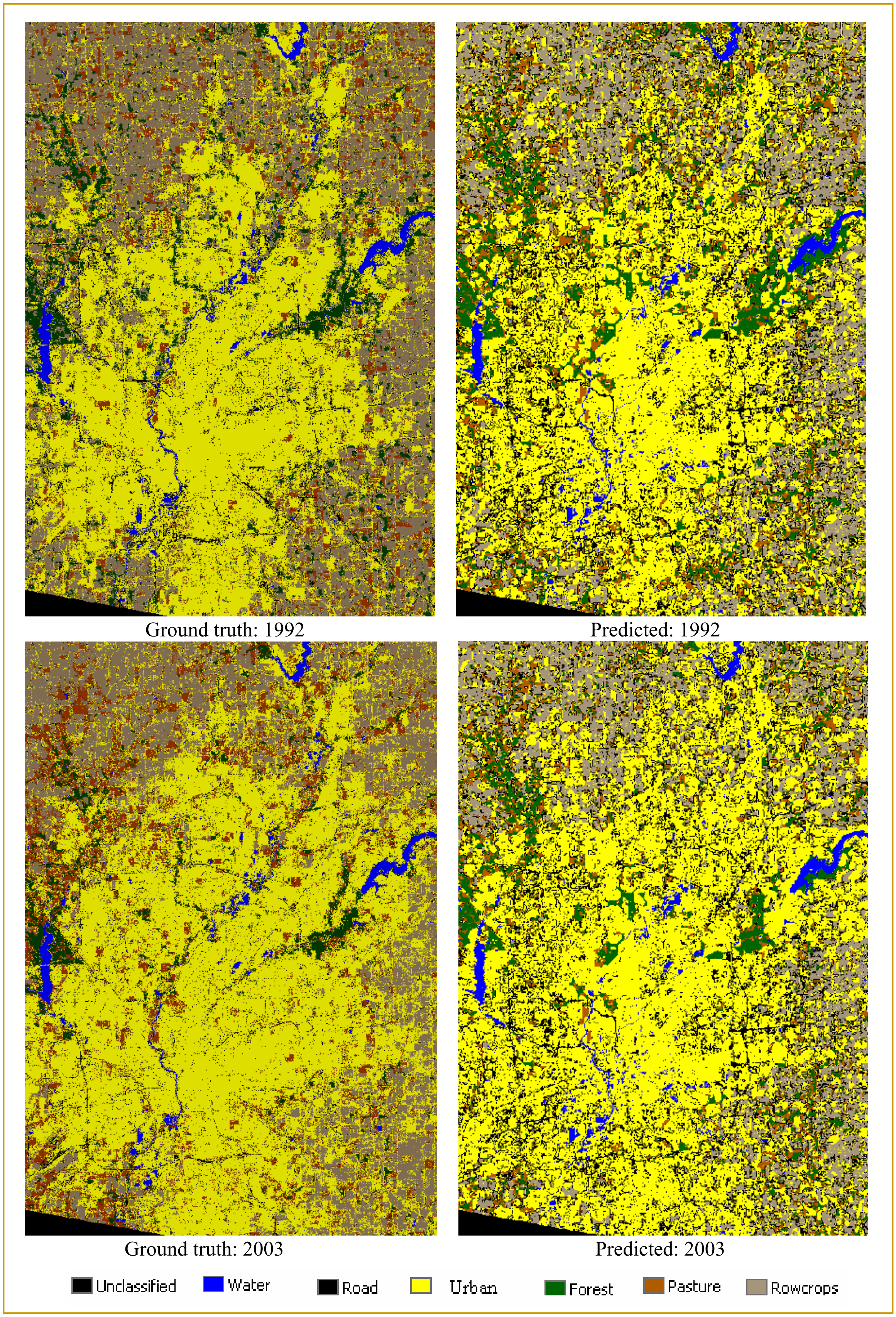


Figure 6. Indianapolis modeling results

Table 2. Fuzzy CA prediction results evaluation for 1992 (5 year prediction) and 2003 (11 year prediction)

Township	1992 Ground Truth (cells)	1992 Prediction (%)	2003 Ground Truth (cells)	2003 Prediction (%)
1	1791	(+)165.66	1550	(+)191.42
2	2999	(+)128.24	3671	(+)104.82
3	5717	(-)85.25	6705	(+)133.04
4	4572	(+)102.67	5685	(-)82.69
5	4058	100.00	4579	(-)90.00
6	7846	(+)111.69	12580	(+)103.88
7	11578	(+)104.73	14860	(+)110.23
8	6578	(-)99.01	11262	(-)96.08
9	7666	(+)103.73	9099	(+)107.01
10	17335	(-)99.20	20183	(-)96.02
11	18576	(-)88.23	19953	(-)90.19
12	10014	(+)101.70	14473	(+)113.82
13	10057	(-)96.95	11175	(+)104.21
14	20222	(-)91.37	20693	(-)94.23
15	21986	(-)90.93	22584	(-)89.59
16	12272	(-)94.51	16801	(-)69.88
17	12133	(-)84.33	12616	(-)98.42
18	20847	(-)89.30	20924	(-)89.32
19	19491	(-)91.51	21005	(-)85.57
20	8542	(+)106.71	12502	(-)73.52
21	2814	(-)89.98	2891	(-)95.71
22	7920	(-)85.25	9249	(+)110.10
23	9946	(-)86.16	12540	(-)97.66
24	4509	(+)132.80	6675	(-)89.71
Average accuracy		**(-) 95.81**		**(-)95.49**

REFERENCES

Batty, M., and Y. Xie (1994a). From cells to cities. *Environment and Planning B: Planning and Design*, *21*: 531-548.

Burrough, P. A., and A. Frank (1996). Geographic objects with intermediate boundaries. *Innovations in GIS 2*. London: Taylor and Francis.

Leung, Y. (1987). On the imprecision of boundaries. *Geographical Analysis, 19*:125-151.

Li, X., and A. Yeh (2000). Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science, 14*: 131-152.

Liu, Y., and S. R. Phinn (2003). Modelling urban development with cellular automata incorporating fuzzy-set approaches. *Computers, Environment and Urban Systems, 27*:637-658.

Gardner, M. (1970). Mathematical Games: the fantastic combinations of John Conway's new solitaire game 'life'. Scientific American, 120-123.

Manetos, P. and Y. N. Photis (2003). Simulation of urban system evolution in a synergetic modelling framework. The case of Attica, Greece. Proceeding from the 43rd European Congress of the RSA, 2003, Jyvaskyla, Finland.

Tsoukalas, L. H., and R. E. Uhrig (1997). *Fuzzy and Neural Approaches in Engineering*. John Wiley and Sons, NewYork., pp. 13-48.

von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press, Illinois. Edited and completed by A. W. Burks.

Wang, F., and G. B. Hall (1996). Fuzzy representation of geographical boundaries in GIS. *International Journal of Geographical Information Systems, 10*:573-590.

White, R., and G. Engelen (1997). Cellular automata as the basis of integrated dynamic regional modelling.

Environment and Planning B: Planning and Design, 24: 235-246.
Wu, F. (1996). A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. Computers, Environment, and Urban Systems, 20:367-387.
Wu, F., and C. J. Webster (1998). Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B, 25*: 103-126.
Wu, F. (1998). Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Journal of Environmental Management, 53*: 293-308.
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*: 335-353.
Zadeh, L. A. (1971). Towards a theory of fuzzy systems, in Aspects of Network and Systems Theory. Eds R. E. Kalman & N. DeClaris, Holt Rinehart, Winston.