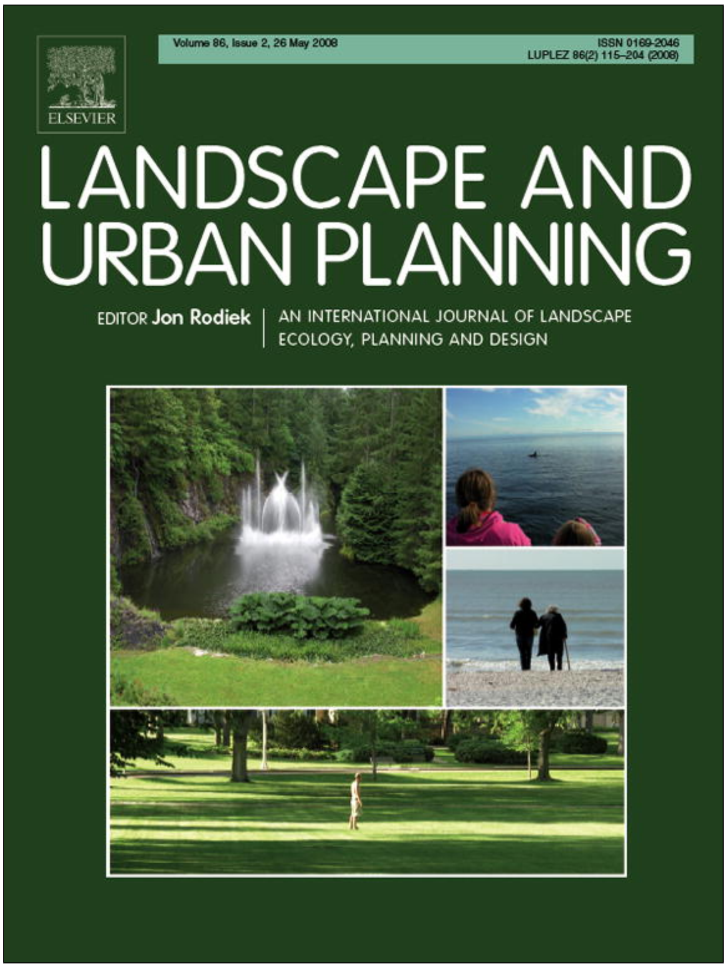
Volume 86, Issue 2, 26 May 2008
ISSN 0169-2046
LUPLEZ 86(2) 115–204 (2008)
ELSEVIER
LANDSCAPE AND URBAN PLANNING
EDITOR Jon Rodiek
AN INTERNATIONAL JOURNAL OF LANDSCAPE ECOLOGY, PLANNING AND DESIGN

Landscape and Urban Planning 86 (2008) 177–186

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

# Discovering and evaluating urban signatures for simulating compact development using cellular automata

Xia Li*, Qingsheng Yang, Xiaoping Liu

*School of Geography and Planning, Sun Yat-sen University, 135 West Xingang Road, Guangzhou 510275, PR China*

ARTICLE INFO

*Article history:*
Received 22 May 2007
Received in revised form 11 December 2007
Accepted 22 February 2008
Available online 11 April 2008

*Keywords:*
Cellular automata
Urban signatures
GIS
Compact cities

ABSTRACT

This paper provides a new method for retrieving, evaluating and modifying urban signatures for simulating compact development using urban cellular automata (CA). Urban CA usually adopt fixed transition rules for simulating urban dynamics in large complex regions. However, these regions can be segmented into sub-regions so that separate transition rules can be retrieved for generating better simulation results. Moreover, urban signatures or "genes" can be extracted from GIS data to assist the understanding of urban evolution for each sub-region. Good "genes" from a sub-region can be cloned to other sub-regions for producing better urban forms. A heuristic swapping technique is developed to modify existing "genes" so that compact development patterns can be generated for planning purposes. The proposed method has been applied to the simulation of compact development in the Pearl River Delta. The analysis indicates that this model can help improve the compactness of urban development in this fast growing region.



## 1. Introduction

The simulation and prediction of urban growth can provide key inputs to many environmental and planning models. In the last two decades, a set of urban models based on cellular automata (CA) techniques were developed for better understanding of urban evolution (Batty and Xie, 1994; Couclelis, 1997; Wu and Webster, 1998; Ward et al., 2000; Li and Yeh, 2004a; Straatman et al., 2004; Herold et al., 2005). Actually, CA were developed by Ulam in the 1940s and soon used by Von Neumann to investigate the logical nature of self-reproducible systems (White and Engelen, 1993). CA have powerful modeling capabilities and high degree of reality when integrated with GIS (Li and Yeh, 2000). These models have inherent advantages in simulating various spatio-temporal processes by using local interaction rules. They can model a variety of natural phenomena in a way that is conceptually clearer, more accurate, and more complete than conventional mathematical systems (Itami, 1994; Weber and Puissant, 2003). Cities become computable in various ways within the generic framework of CA models (Batty and Xie, 1994).

The core of CA is how to define transition rules, which can be represented in many forms. For example, transition rules can be represented by using weight matrices (White and Engelen, 1993), the SLEUTH model (slope, landuse, exclusion, urban extent, transportation and hillshade) (Clarke and Gaydos, 1998), multicriterion evaluation (MCE) (Wu and Webster, 1998), logistic regression (Wu, 2002), neural networks (Li and Yeh, 2002) and decision trees (Li and Yeh, 2004a). In these methods, many variables are involved for defining transition rules. Each variable is usually associated with a parameter that indicates its importance in simulation. These parameters significantly affect the outcomes of urban simulation (Wu, 2002; Li and Yeh, 2002).

The calibration of CA is essential for simulating urban dynamics of realistic cities (Silva and Clarke, 2002; Li and Yeh, 2002; Straatman et al., 2004). However, existing methods are based on a uniform set of parameters for simulating urban dynamics in a whole region. This assumes that the relationships are fixed in the spatio-temporal dimension. In reality, the relationships may be complex between the state conversion (e.g. converted to urban land or not) and its geographical variables, and discrepancy can be created for the simulation based on a uniform set of transition rules. Heterogeneous development patterns can be observed in large complex regions (Li and Yeh, 2004b). The irregularities of development patterns can be attributed to localized physical, social and political factors which cannot be captured by models. For example, special land use policies can be adopted by local governments, resulting in the changes in development patterns. A way to capture these localized features is to divide the study region into some relative homogenous sub-regions based on segmentation methods. Administrative boundaries may be used to divide a large region into sub-regions since localized policies are usually implemented within the jurisdiction of each sub-region.

* Corresponding author. Tel.: +86 20 33047500; fax: +86 20 84115833.
*E-mail addresses:* lixia@mail.sysu.edu.cn, lixia@graduate.hku.hk (X. Li).

0169-2046/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.landurbplan.2008.02.005

Recently, some primitive efforts have been made to define the signatures of a region according to the calibration of CA. Silva (2004) perhaps is the first to propose the concept of "the DNA of our regions" through the use of urban CA models. He suggests that CA can be calibrated to produce realistic simulation and the use of the calibration values of CA can have another function by identifying DNAs of each region. This opportunity arises from the possibility of defining a "signature" that identifies the uniqueness of a region without compromising its universality. His experiments are based on the SLEUTH CA which has five control parameters for the growth components of diffusion, breed, spread, slope resistance, and road gravity. A key of numbers resulting from the calibration process can be used to reflect the local characteristics of a region. Cloning a DNA to another region can have very different results. This also indicates that the shift of existing policies can lead to unexpected results.

Urban morphology is an important element in urban planning. Compact development should be promoted so that less land will be consumed by development (Jenks et al., 1996). There are evidences indicating a strong link between urban form and sustainable development, although it is not simple and straightforward. Significant relationships have been found between energy use in transport and physical characteristics of cities, such as density, size, and amount of open space (Banister et al., 1997).

Dispersed development patterns have been witnessed in some fast growing cities in China. For example, land development is in a rather fragmented form in the Pearl River Delta (Li and Yeh, 2004b). A series of land use problems are associated with sprawl development, including increases in development costs, energy consumption and wasteful use of land resources (Yeh and Li, 2001). Simulation of compact urban forms is the first step toward the implementation of this compact development initiative. Urban signatures play a key role in determining urban morphology in urban simulation.

The objective of this study is to explore the relationships between urban signatures and compact development. Methodologies are developed for retrieving, cloning and modifying urban signatures by using urban CA. A large complex region is first segmented into a number of sub-regions according to administrative boundaries. A set of "genes" is retrieved for each sub-region by using GIS data. Genetic algorithms are used to assist the search for these "genes" since they are difficult to obtain for non-linear equations. These "genes" are then evaluated according to some spatial metrics. The modification of "genes" is finally carried out by using a heuristic swapping technique. The proposed method should be useful for simulating compact urban development under various assumptions for planning purposes.

## 2. The study area

The study region constitutes the core of the Pearl River Delta with an area of about 41,157 km$^2$, situated in the central part of Guangdong, south China. The fertile delta is very suitable for agricultural production. However, a large amount of land use changes, associated with the loss of a significant amount of agricultural land, has been observed in this fast growing region since the economic reform in 1978 (Li and Yeh, 2004b).

Localized features of growth patterns can be observed in this large complex area. For example, two significantly different growth patterns can be visually identified for two neighboring cities in the Pearl River Delta (Fig. 1). This region can be segmented into a number of sub-regions for capturing the complexity of urban dynamics. In this research, the study area is divided into six major sub-regions based on the administrative boundaries of cities. They are the cities of Guangzhou city proper, Zengcheng, Conghua, Shenzhen, Dongguan and Zhongshan (Fig. 2). A major land use problem in the study area is related to sprawl development along transport networks (Yeh and Li, 2001). It can be easily identified from the classification of remote sensing images (Fig. 3).

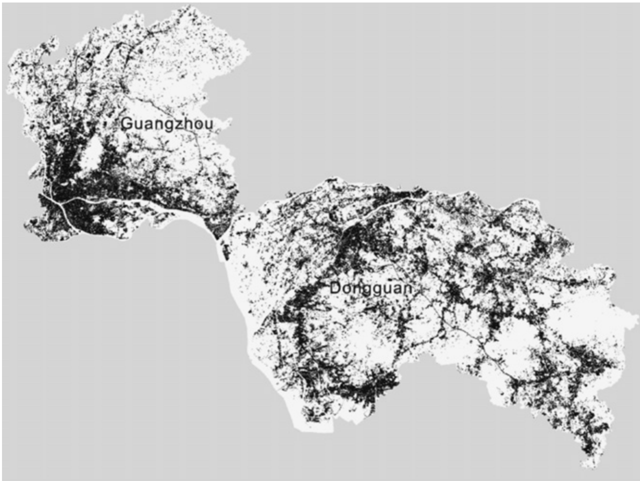


**Fig. 1.** Two significant growth patterns identified for two neighboring cities in the Pearl River Delta.

Spatial variations of land use patterns and development processes can be identified among these cities (Li and Yeh, 2004b). This can be attributed to the differences in geographical locations and localized land use policies. Fig. 4 illustrates the validity of the segmentation which is based on the administrative boundaries. There is a strong conformity between the transition of growth patterns and the administrative boundaries. This indicates that administrative boundaries can be used to differentiate the growth patterns for this aggregate region.

## 3. Capturing and modification of urban signatures using cellular automata

### 3.1. Capturing urban signatures

Cellular automata consist of four elements–cells, states (e.g. urbanized or non-urbanized), neighborhoods, and transition rules. The definition of transition rules is essential for implementing CA. In an urban CA, transition rules are usually represented by using a probability function which determines if land use conversion will take place (Wu and Webster, 1998). This involves a number of spatial variables that represent various forces in urban evolution. The combined effects of these forces can be addressed by incorporating multicriteria evaluation (MCE) into cellular automata (Wu and Webster, 1998).

The development probability is determined by a combined evaluation score $r_{ij}$, of which nonlinear transformation is used to discriminate the simulation patterns (Wu and Webster, 1998):

$$p_{ij}^t = \phi(r_{ij}^t) = \exp\left[\alpha\left(\frac{r_{ij}^t}{r^{\max}} - 1\right)\right] \quad (1)$$

where $\alpha$ is a dispersion parameter ranging from 0 to 1; $r_{ij}^t$ is the combined evaluation score at location $ij$; $r^{\max}$ is the maximum value of $r_{ij}^t$.

The composite evaluation score ($r_{ij}^t$) is calculated by using the following linear equation:

$$r_{ij}^t = a_0 + a_1x_1^t + a_2x_2^t + \cdots + a_mx_m^t + \cdots + a_Mx_M^t \quad (2)$$

where $a_0$ is the constant; $x_1^t, x_2^t, \ldots, x_m^t, \ldots, x_M^t$ are spatial variables representing the driving forces for land development;

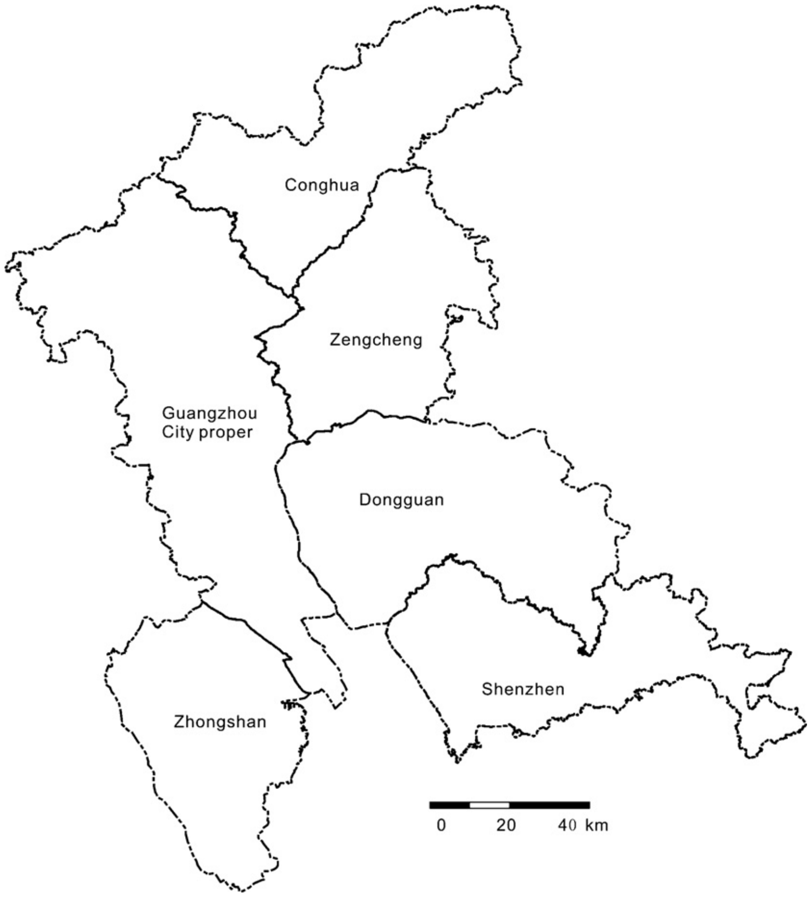


**Fig. 2.** A complex region with a hierarchy of cities.

$a_1$, $a_2$, ..., $a_m$, ..., $a_M$, are the parameters (weights) of these variables.

It is straightforward to understand the meanings of the weights in the MCE expression. A larger weight indicates that the associated variable makes a greater contribution to the development probability. However, this MCE-CA model cannot be calibrated for simulating realistic cities. A modification of this model is to transform it into a logistic form so that the calibration is possible (Wu, 2002):

$$p_{ij}^t = \frac{\exp(-r_{ij}^t)}{1+\exp(-r_{ij}^t)} = \frac{1}{1+\exp(-r_{ij}^t)} \tag{3}$$

Urban development is subject to a series of physical constraints and some uncertainties. By incorporating a series of constraints plus a stochastic factor, the above equation can be further revised as follows:

$$p_{ij}^t = (1+(-\ln \gamma)^{\alpha}) \times \frac{1}{1+\exp(-r_{ij}^t)} \times \mathrm{con}(s_{ij}^t) \times \Omega_{ij}^t \tag{4}$$

where $\gamma$ is a stochastic factor ranging from 0 to 1, $\Omega_{ij}^t$ is the development intensity in the neighborhood, and $\mathrm{con}(s_{ij}^t)$ is the total constraint score ranging from 0 to 1.

The terms of $r_{ij}^t$, $\Omega_{ij}^t$ and $\mathrm{con}(s_{ij}^t)$ are dynamically updated during CA simulation. At each iteration, $p_{ij}^t$ is compared with a threshold value to determine if a non-urbanized cell will be converted into urbanized cell:

IF $p_{ij}^t >$ Threshold and cell *ij* is undeveloped,
THEN The state of the cell will be converted into urban land.

The parameters associated with these spatial variables in the CA model have crucial effects on determining state conversion (urban dynamics) in the simulation. They are analogous to the "genes"

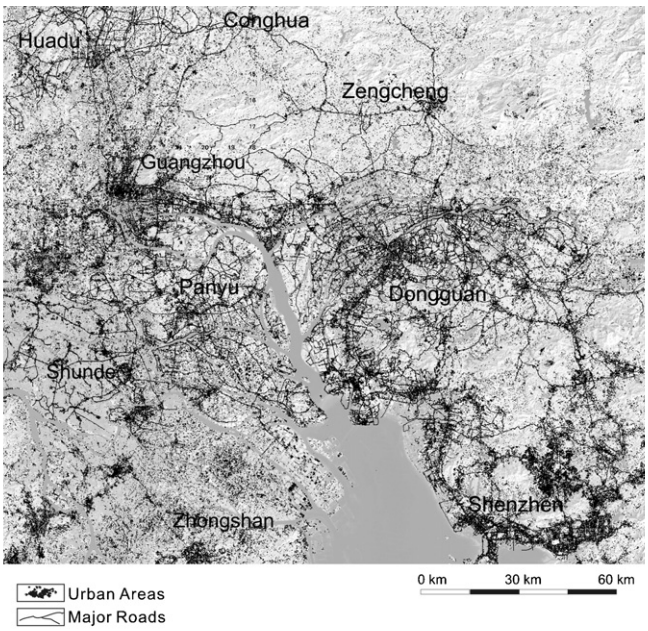


**Fig. 3.** Urban sprawl along roads in the Pearl River Delta according to classified TM images.

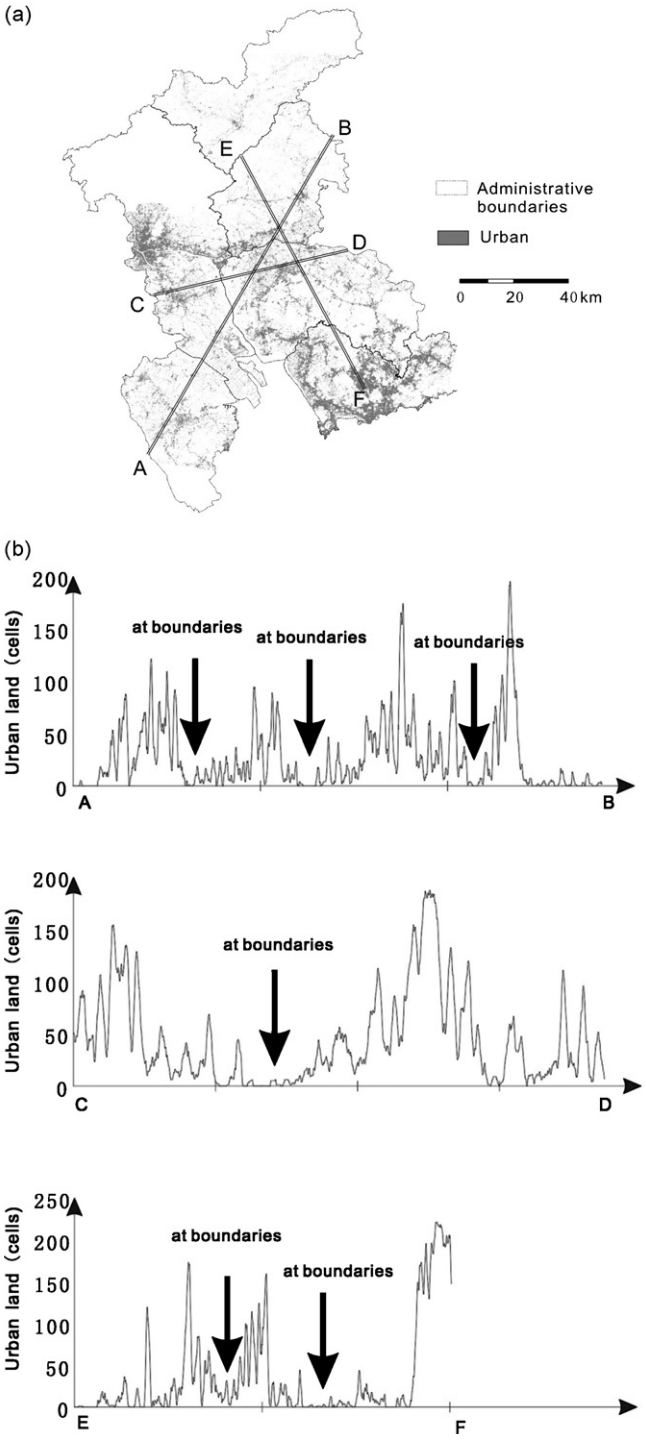


**Fig. 4.** Conformity between the transition of growth patterns and the administrative boundaries (a). (b) Profile of land development

in biology because they have a unique role in representing urban evolution during simulation. Since there are significant spatial variations of growth patterns in a large region (Li and Yeh, 2004b), the use of a single set of "genes" or urban signatures may not be suitable for simulating land use dynamics. An aggregated region usually consists of many sub-regions, such as a number of cities and towns. It is preferable to partition this region into a number of sub-regions. Each sub-region can be associated with a unique set of "genes" to capture local characteristics in urban development. When a region is divided into a number of sub-regions, the transition rules in Eq. (3) can be revised as follows:

$$p^t_{ij,k} = \frac{\exp(-r^t_{ij,k})}{1+\exp(-r^t_{ij,k})} = \frac{1}{1+\exp(-r^t_{ij,k})} \quad (5)$$

where $r^t_{ij,k} = a_{0,k} + a_{1,k}x^t_{1,k} + a_{2,k}x^t_{2,k} + \cdots + a_{m,k}x^t_{m,k} + \cdots + a_{M,K}x^t_{M,k}$ $(k = 1, 2, \cdots, K)$, $K$ is the total number of sub-regions. $x^t_{m,k}$ is the $m$th variable for sub-region $k$ and $a_{m,k}$ is the weight of the variable.

Therefore, the set of urban signatures (CM) for sub-region $k$ can be represented as follows:

$$\mathrm{CM} = [a_{0,k}, a_{1k}, \ldots, a_{m,k}, p_{\mathrm{threshold},k}] \quad (6)$$

Conventional calibration procedures have difficulties in determining the above signatures because they are not in a linear form. In this study, a genetic algorithm (GA) is used to find the optimal set of urban signatures for each sub-region, which is crucial for producing realistic simulation results. GA has advantages in solving a lot of complex optimization problems in many disciplines because specific programs are not required (Goldberg, 1989). The optimization procedure is based on the concept of natural selection which controls the evolution process in biology. GA is excellent for quickly finding an approximate global maximum or minimum value. Moreover, the form of GA is generally applicable to a variety of complex optimization problems.

Fitness functions should be defined for finding the optimal parameters of the CA model. Fitness functions are used to indicate the performance of each solution or individual (chromosome) in solving an optimal problem. The definition of fitness functions is domain-dependent.

In this study, the fitness function is defined by calculating the difference between the actual state (e.g. urbanized or not) and the predicted state. The optimal set of parameters should produce the minimum value (the least error) of the fitness function. Therefore, the fitness function is represented as follows:

$$f(x) = \sum_{i=1}^{n} (\widehat{f}_i - f_i)^2 \quad (7)$$

$$\widehat{f}_i = \begin{cases} 1 & \text{if } \widehat{f}_i \geq p_{\mathrm{threshold}} \\ 0 & \text{if } \widehat{f}_i < p_{\mathrm{threshold}} \end{cases}$$

where

$$\widehat{f}_i(x_1, x_2, \cdots, x_m) = \frac{1}{1+\exp(-(a_{0,k} + a_{1,k}x_{1,k} + a_{2,k}x_{2,k} + \cdots + a_{m,k}x_{m,k} + \cdots + a_{M,K}x_{M,K}))}.$$

$f_i$ is the actual state ($f_i = 1$ for urbanized cells; $f_i = 0$ for non-urbanized cells).

The actual state is obtained from the classification of remote sensing imagery. The predicted state is calculated by using the logistic model. The whole region is divided into a number of sub-regions (e.g. cities) based on the administrative boundaries. The GA program is used to find the optimal set of urban signatures for each sub-region. These retrieved urban signatures will control urban evolution for each city in the simulation.

### *3.2. Modifying urban signatures for simulating compact development*

Urban forms should be 'more compact and humane', instead of the increasing sprawl of metropolitan development (Bourne, 1992). It is expected that some "genes" can have better performances in terms of compact development. The evaluation of urban morphology can help identify these good "genes". The assessment is carried out by using some common landscape metrics, which can provide a detailed description of the accuracy of the model's historical simulations (Herold et al., 2003). These spatial metrics include mean patch shape index (MPSI), mean patch fractal dimension (MPFD), mean Euclidean nearest-neighbor (MNN) distance, and aggregation index (AI). These metrics are selected because they are related to the compactness of urban forms. They are calculated at the landscape level for each city by using a landscape analysis package, FRAGSTATS 3.3 (McGarigal and Marks, 1995).

Mean patch shape index is given as follows (McGarigal and Marks, 1995):

$$\mathrm{MPSI} = \frac{0.25\sum_{i=1}^{n} P_i}{\sqrt{\sum_{i=1}^{n} A_i}} \tag{8}$$

where MPSI is mean patch shape index, $P_i$ is the perimeter of patch $i$, $A_i$ is the area of patch $i$ in terms of number of cells, $n$ is the total number of patches. MPSI increases as patch shape becomes more irregular.

Mean patch fractal dimension is calculated as follows (McGarigal and Marks, 1995):

$$\mathrm{MPFD} = \frac{\sum_{i=1}^{n}[2\,\ln(0.25P_i)/\ln(A_i)]}{n} \tag{9}$$

where MPFD is mean patch fractal dimension. MPFD approaches 1 for shapes with very simple perimeters such as squares, and approaches 2 for shapes with highly convoluted, plane-filling perimeters.

Mean Euclidean nearest-neighbor distance is represented by (McGarigal and Marks, 1995):

$$\mathrm{MNN} = \frac{\sum_{i=1}^{n} h_i}{n} \tag{10}$$

where MNN is mean Euclidean nearest-neighbor distance, $h_i$ is the distance from patch $i$ to nearest neighboring patch of the same type (class) $i$, based on patch edge-to-edge distance, computed from cell centre to cell centre. MNN decreases as patches become more compact.

Aggregation index is expressed by the following equation (McGarigal and Marks, 1995):

$$\mathrm{AI} = \left[\frac{g_{ii}}{\max g_{ii}}\right] \times 100 \tag{11}$$

where AI is aggregation index, $g_{ii}$ is the number of like adjacencies (joins) between pixels of patch type (class) $i$ based on the single-count method. max $g_{ii}$ is the maximum number of like adjacencies (joins) between pixels of patch type (class) $i$ based on the single-count method. max $g_{ii}$ is expressed as follows (McGarigal and Marks, 1995):

$$\max g_{ii} = \begin{cases} 2n(n-1), & m = 0 \\ 2n(n-1) + 2m - 1, & m \le n \\ 2n(n-1) + 2m - 2, & m > n \end{cases} \tag{12}$$

where $m = a_i - n^2$, $a_i$ is the area of class $i$ (in terms of number of cells) and $n$ is the side of a largest integer square smaller than $a_i$. AI equals 0 when the focal patch type is maximally disaggregated (i.e., when there are no like adjacencies); AI increases as the focal patch type is increasingly aggregated and equals 100 when the patch type is maximally aggregated into a single, compact patch.

Since MPSI, MPFD, MNN, AI are measured at different scales. These metrics should be normalized so that they can be compared. The following equation is used for the normalization:

$$x_i' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

These normalized metrics can be combined to form a final utility function ($U$) by representing all these morphological effects. This utility function is defined as follows:

$$U = \frac{1}{4}((1 - \mathrm{NMPSI}) + (1 - \mathrm{NMPFD}) + (1 - \mathrm{NMNN}) + \mathrm{NAI}) \tag{14}$$

where NMPSI, NMPFD, NMNN, and NAI are the normalized MPSI, MPFD, MNN and AI, respectively. $U$ is the combined utility function. The higher the utility value, the better the urban morphology becomes in terms of compact development.

The spatial data are stored in ARCGIS GRID data format, which can be imported to FRAGSTATS 3.3 for the calculation of these metrics. The best set of "genes" can be identified according to this utility function. This set of "genes" can be cloned to other sub-regions to produce better urban forms. For example, the "genes" of the city proper can be used to replace those of other cities.

Directly cloning these "genes" may not be the best option since land use problems in a city can be propagated to other cities. A solution is to modify existing "genes" based on the assessment of their performance. Different parts of "genes" will play specific roles in controlling urban morphology. For example, some "genes" will result in road-based development, but others will produce town-centre-based development. "Genes' can be modified to produce more compact growth scenarios under various planning objectives.

The study area consists of a hierarchy of centres, a number of major centres (city centres) and many sub-centres (town centres). Therefore, compact development can be implemented in a mono-centric form (around city centres) or a polycentric form (around town centres). Two options are then available for modifying the existing best "genes" (e.g. the "genes" of the city proper) before they are used for the whole region. These two options of modification include: (1) "city centers transport" concentrated development; (2) "town centers transport" concentrated development. The first option is to address the trade-off between the attractions from city centres, plus transport networks. Most of the land development is attracted by city centres, but some by transport networks, such as roads, railways, and expressways. The second option is to address the trade-off between the attractions from town centres, plus transport networks. Most of the land development is attracted by town centres, but some by transport networks.

The "genes" from the most compact city (e.g. the city proper) are used as the starting point for the modification. A heuristic swapping technique is proposed for the search for better "genes" compared to the existing ones. The search is constrained by the total amount of land use conversion, which is obtained from the classification of remote sensing images. This is to ensure that the total amount of land use conversion is the same between the simulated and the actual (expected). The new "genes" are obtained by interactive modification of the weights between city centres, town centres and roads.

The detailed procedure of modification for "city centres transport" concentrated development is as follows:

(1) The initial weights ($a_{M,K}$) are set to the minimum absolute value (−0.0001) for all the variables, such as city centres, roads, railways, and expressways (Fig. 5a). This is to guarantee the minimum attraction to these factors.
(2) The absolute weight for the variable of urban centres will then be increased, constrained by the total amount of land conversion. The constraint is to guarantee that the amount of the simulated land conversion is equal to that of the actual

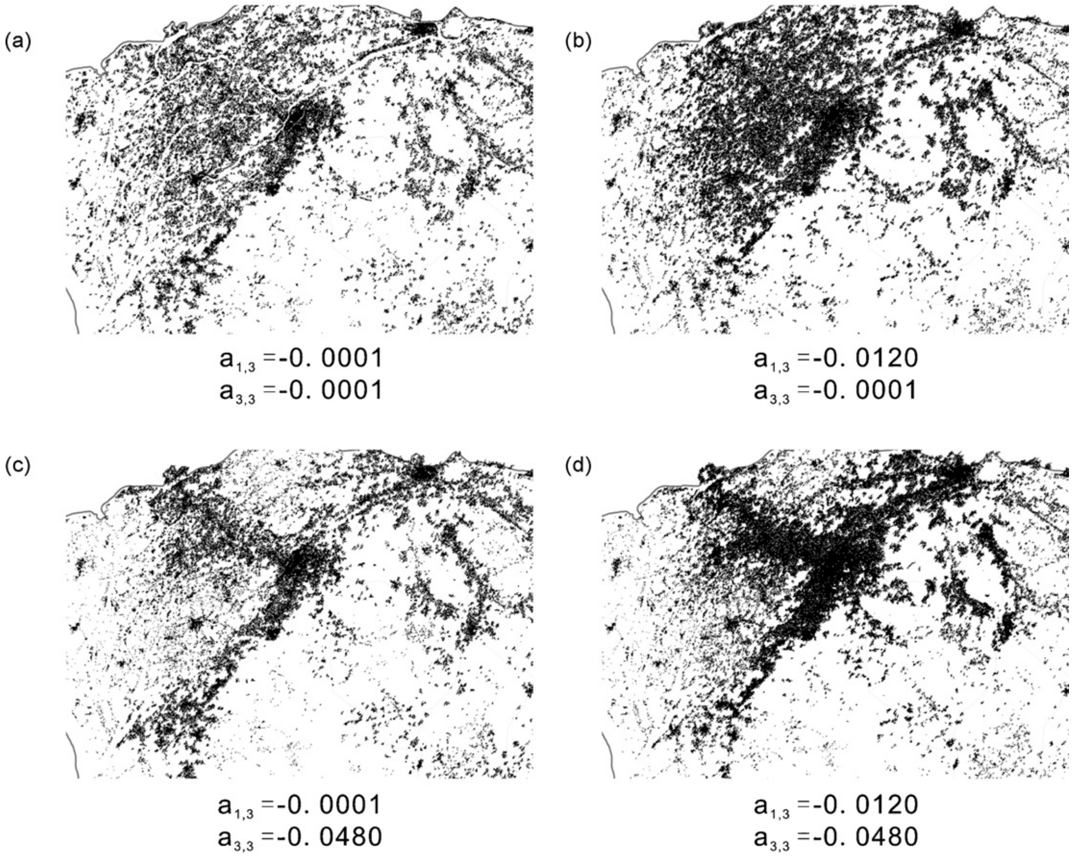


**Fig. 5.** Obtaining new "genes" by interactive modification of the attractions between city centres and roads.

(expected). This increase will result in a higher amount of land development around urban centres, and a lower amount of land development around transport networks. It will thus create a polarized effect of land development around urban centres (Fig. 5b).

(3) The modification is also applied to the absolute weight for transport networks. This change will result in a higher amount of land development around transport networks, and a lower amount of land development around urban centres. It will thus create a polarized effect of land development around transport networks (Fig. 5c).

(4) Repeat steps (2) and (3) again until the urban form cannot be further improved significantly in terms of compact development (Fig. 5d).

The same procedure can be used to modify these "genes" for "town centres-transport" concentrated development. This method provides a transparent tool for formulating alternative development scenarios, which links the patterns to processes by using the concept of "genes".

## 4. Implementation and results

### *4.1. Mining urban signatures for simulating realistic development patterns*

Urban "genes" are defined to characterize a region for predicting and regulating urban dynamics. The "genes" for simulating realistic urban growth are obtained by using empirical data from remote sensing and GIS. The dependent variable, land use conversion, was obtained by the classification of the Landsat TM images dated 10 December 1988 and 24 December 1993, respectively. The accuracy assessment for land use classification was carried out with reference to available land use maps, air photographs and field investigation. The total accuracy is 0.87 and the kappa coefficient is 0.83 according to the accuracy assessment (Li and Yeh, 2004b). The independent variables, a series of spatial variables (e.g. proximity to centres), were retrieved by using GIS functions.

In this study, the set of "genes" for each sub-region was obtained by using a genetic algorithm (GA). Stratified random sampling (Congalton, 1991) was first employed to collect the samples from remote sensing and GIS. 10% of the original data were selected in

**Table 1**
Retrieved urban "genes" for the cities in the Pearl River Delta

| | $a_{0,k}$ | $a_{1,k}$ | $a_{2,k}$ | $a_{3,k}$ | $a_{4,k}$ | $a_{5,k}$ | $P_{\text{threshold}}$ |
|---|---|---|---|---|---|---|---|
| Guangzhou | 1.476 | −0.00079 | −0.00794 | −0.02519 | −0.00245 | −0.00402 | 0.445901 |
| Zengcheng | 1.500 | −0.00089 | −0.00010 | −0.02048 | −0.00010 | −0.00032 | 0.512726 |
| Conghua | 1.500 | −0.00088 | −0.00872 | −0.02832 | −0.00010 | −0.00010 | 0.542195 |
| Shenzhen | 1.500 | −0.00010 | −0.00480 | −0.01656 | −0.00794 | −0.00010 | 0.512844 |
| Dongguan | 0.978 | −0.00010 | −0.00559 | −0.01421 | −0.00167 | −0.00010 | 0.729893 |
| Zhongshan | 1.034 | −0.00089 | −0.00010 | −0.02205 | −0.00167 | −0.00010 | 0.559767 |

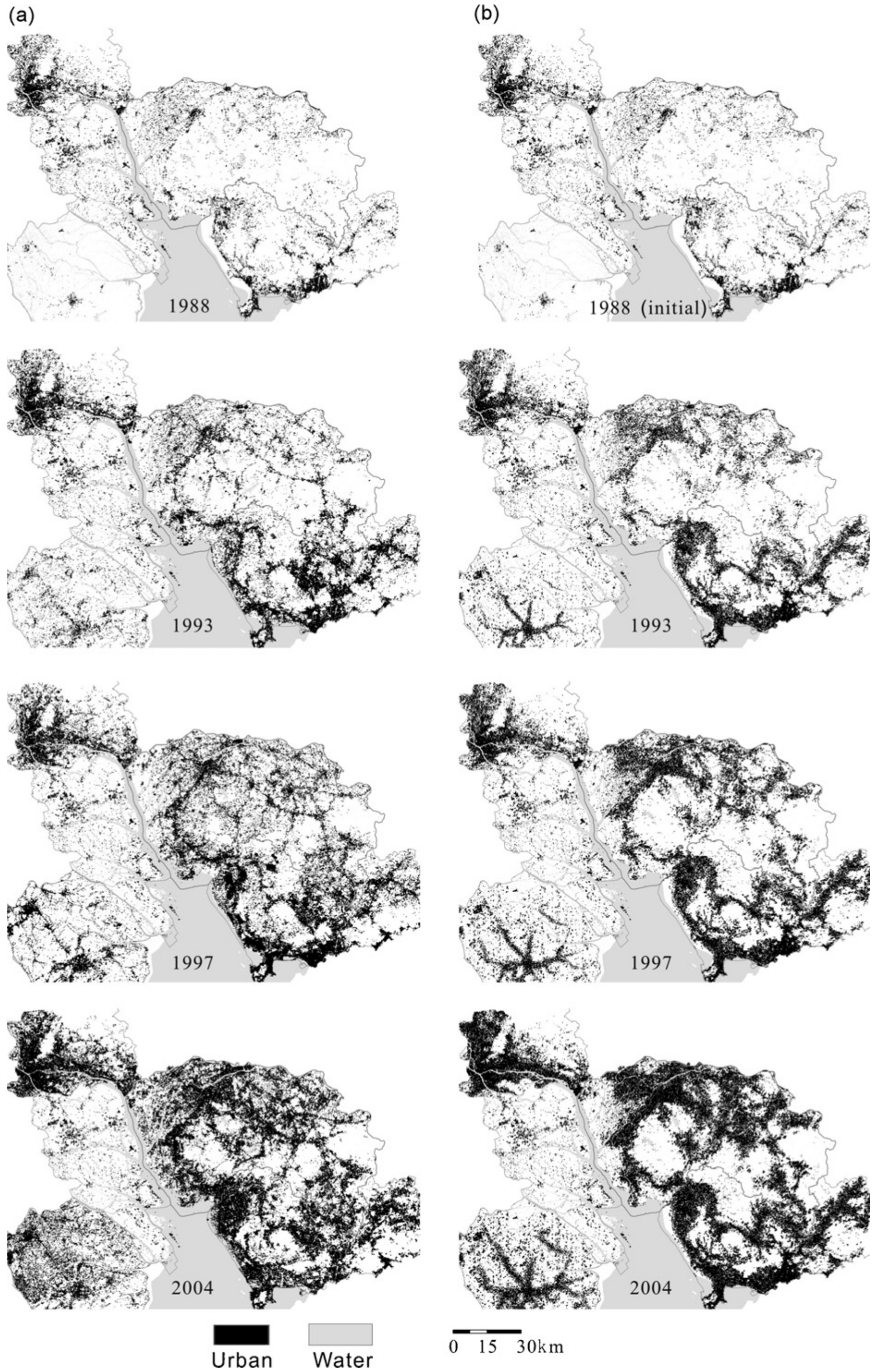


**Fig. 6.** Realistic simulation of urban growth for the cities in the Pearl River Delta in 1988–2004. (a) Actual and (b) simulated.

the classified remote sensing imagery as the training data. These data were then used to calculate the fitness function for the GA programming.

In this study, the population size was set to 100. The initial value of $a_{0,k}$ was 0.5, and all the initial values of $a_{1,k}, \ldots, a_{m,k}, \ldots, a_{M,K}$, were $-0.01$. The crossover rate and the mutation rate were 0.90 and 0.01, respectively. The strategies of elitist selection and diversity operation were also adopted to facilitate the search for the optimal parameters.

Table 1 shows the retrieved "genes" for simulating urban evolution in this region. Distinct sets of "genes" are obtained for different cities in this region. Each set of "genes" will control the unique evolution of urban morphology for a city. Traditional methods have difficulties in determining these parameters because of the complexities.

These "genes" can be applied for generating realistic urban growth without any modifications (Fig. 6b). They can be used to simulate urban development in the same period (1988–1993) from which the empirical data were obtained, and predict urban development in the "future" period (1993–2004) based on the growth trajectory. Very plausible results have been obtained by using these "genes" to simulate urban development in 1988–2004, although

**Table 2**
Comparison of simulation accuracies between using separate transition rules and using unified transition rules

| | Simulated non-urban | Simulated urban | Accuracy (%) |
|---|---|---|---|
| (A) Separate transition rules | | | |
| 1988–1993 (cells) | | | |
| Actual non-urban | 906188 | 80750 | 91.82 |
| Actual urban | 236621 | 505090 | 68.10 |
| Total accuracy (%) | 81.64 | | |
| Kappa coefficient | 0.62 | | |
| 1993–1997 (cells) | | | |
| Actual non-urban | 667065 | 88293 | 88.31 |
| Actual urban | 256176 | 717115 | 73.68 |
| Total accuracy (%) | 80.07 | | |
| Kappa coefficient | 0.60 | | |
| 1997–2004 (cells) | | | |
| Actual non-urban | 397942 | 95836 | 80.59 |
| Actual urban | 275731 | 899140 | 76.53 |
| Total accuracy (%) | 77.73 | | |
| Kappa coefficient | 0.52 | | |
| (B) Unified transition rules | | | |
| 1988–1993 (cells) | | | |
| Actual non-urban | 856626 | 137255 | 86.19 |
| Actual urban | 273166 | 475380 | 63.51 |
| Total accuracy (%) | 76.45 | | |
| Kappa coefficient | 0.51 | | |
| 1993–1997 (cells) | | | |
| Actual non-urban | 617503 | 144798 | 81.01 |
| Actual urban | 292721 | 687405 | 70.13 |
| Total accuracy (%) | 74.89 | | |
| Kappa coefficient | 0.50 | | |
| 1997–2004 (cells) | | | |
| Actual non-urban | 408380 | 152341 | 72.83 |
| Actual urban | 322276 | 859430 | 72.72 |
| Total accuracy (%) | 72.76 | | |
| Kappa coefficient | 0.42 | | |

these "genes" are retrieved by using empirical data in 1988–1993. This is conformed by comparing the simulated patterns (Fig. 6b) with the actual patterns (Fig. 6a) which is obtained by classifying remote sensing data. This indicates that CA have a strong capability of predicting urban development if they have been calibrated by using empirical data.

This proposed model is based on separate transition rules obtained by dividing the whole region into six sub-regions according to the administrative boundaries. Comparison indicates that the simulation accuracies of using separate transition rules are much better than those of using unified transition rules (Table 2). The total accuracy and kappa coefficients of using separate transition rules are 81.64% and 0.62 for 1988–1993, 80.07% and 0.60 for 1993–1997, and 77.73% and 0.52 for 1997–2004, respectively. However, the total accuracy and kappa coefficients of using unified transition rules are 76.45% and 0.51 for 1988–1993, 74.89% and 0.50 for 1993–1997, and 72.76% and 0.42 for 1997–2004, respectively.

### *4.2. Modifying urban signatures for simulating compact development patterns*

It is obvious that the simulation based on the "genes" mined from the empirical data will inherit the past sprawl patterns. Modification of urban "genes" is required to produce possible development alternatives to avoid these land use problems. Table 3 lists the results of assessing the actual urban forms of various cities in the study area. It is found that Guangzhou city proper has the largest value of the combined utility function. This provincial capital has the most compact form because of its implementation of strict development control. The whole region can have a better urban form if other cities can follow the behavior of the city proper. This can be realized by cloning these "genes" from the city proper to other cities.

**Table 3**
Assessment of the urban forms for the cities in the Pearl River Delta using spatial metrics

| | MPSI | MPFD | MNN | AI | *U* |
|---|---|---|---|---|---|
| Guangzhou | 1.3789 | 1.0507 | 143.2747 | 69.7479 | 1.0000 |
| Zengcheng | 1.4712 | 1.0712 | 171.8017 | 58.2507 | 0.4349 |
| Conghua | 1.4224 | 1.0748 | 197.2318 | 39.6987 | 0.2444 |
| Shenzhen | 1.4983 | 1.0559 | 161.6018 | 68.7973 | 0.6676 |
| Dongguan | 1.5213 | 1.0603 | 172.5922 | 55.7620 | 0.4367 |
| Zhongshan | 1.4920 | 1.0647 | 218.5303 | 55.1756 | 0.2850 |

**Table 4**
Comparison of the simulated urban forms between using the original "genes" and using the cloning "genes"

| | MPSI | MPFD | MNN | AI | *U* |
|---|---|---|---|---|---|
| The original genes | 1.4988 | 1.0657 | 258.7611 | 78.5389 | 0.0000 |
| Cloning the "genes" of the city proper | 1.4409 | 1.0586 | 249.6434 | 79.4347 | 0.4350 |

**Table 5**
Modified "genes" for "city centres transport" concentrated development

| $a_{0,k}$ | $a_{1,k}$ | $a_{2,k}$ | $a_{3,k}$ | $a_{4,k}$ | $a_{5,k}$ |
|---|---|---|---|---|---|
| 1.2 | −0.00183 | −0.0001 | −0.018 | −0.0001 | −0.0001 |

Fig. 7b shows the results of simulating the development patterns of the whole region by using the "genes" of the city proper. Cloning these genes to the whole region has resulted in a significant increase of compactness for the whole region. This fact is supported by the significant increase of the combined utility value (Table 4). Therefore, this proposed method can produce not only a compact but also a practical form by cloning the realistic good "genes". It is possible to generate a complete compact form, but this form may not be practical. This proposed method is useful for creating a more acceptable urban form, assuming that the mechanism of urban development in the city proper is applicable to other cities.

A further improvement is to modify the "genes" of the city proper before the cloning. This is because the sprawl pattern is also obvious in the city proper. Therefore, the "genes" controlling land development should be modified to reduce this sprawl pattern for the whole region. The swapping technique was used to modify the "genes" for producing the scenario of "city centres transport" concentrated development. The modified "genes" are shown in Table 5. Fig. 7c is the simulation results based on this set of modified "genes".

The same procedure was applied to the derivation of the "genes" for "town centres transport" concentrated development. Table 6 shows the retrieved "genes" according to this modification. The simulation outcome based on this set of modified "genes" is shown in Fig. 7d.

The performances of the above development options were assessed in terms of compact development. Table 7 lists the improvement of the combined utility value for these options, compared to the realistic development. There is an improvement of the utility by directly using the "genes" of the city proper. More

**Table 6**
Modified "genes" for "town centres transport" concentrated development

| $a_{0,k}$ | $a_{1,k}$ | $a_{2,k}$ | $a_{3,k}$ | $a_{4,k}$ | $a_{5,k}$ |
|---|---|---|---|---|---|
| 1.2 | −0.0023 | −0.0023 | −0.0023 | −0.0001 | −0.0001 |

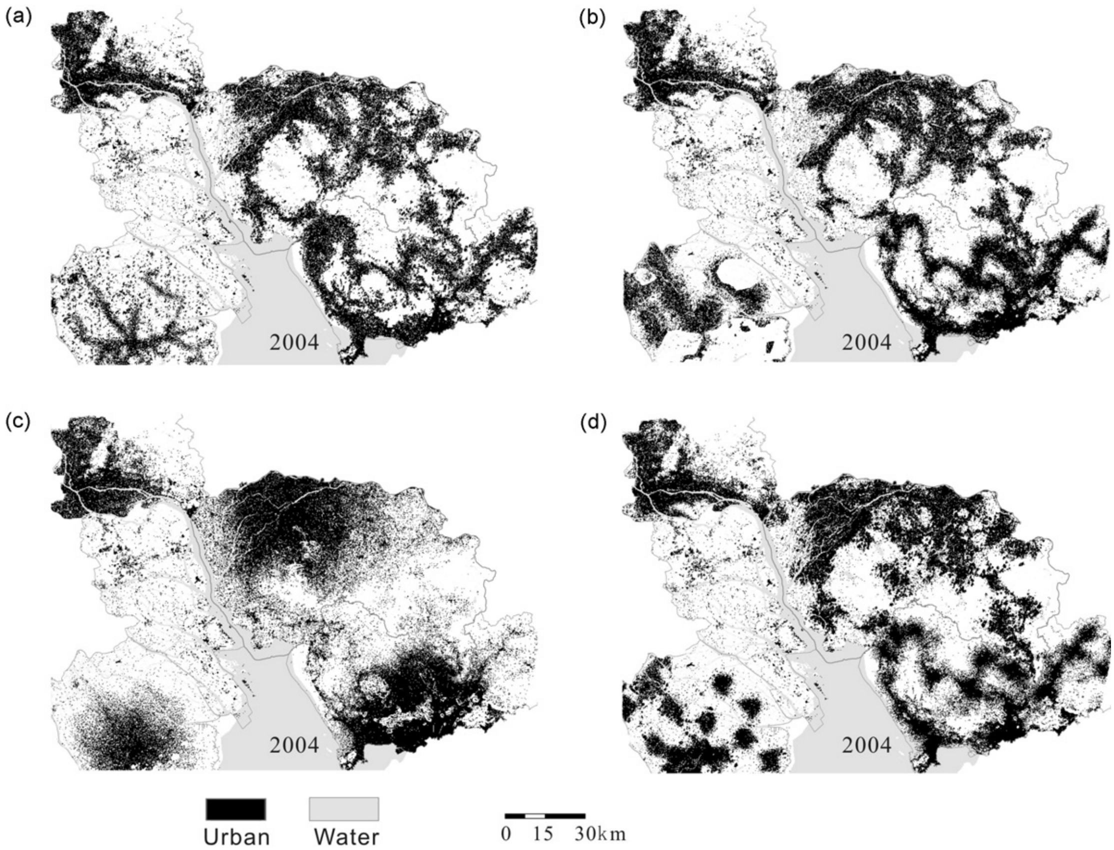


**Fig. 7.** Simulation of compact cities in the Pearl River Delta in 2004: (a) realistic simulation (original retrieved "genes"); (b) using the "genes" of the city proper (Guangzhou's "genes"); (c) modified genes for "urban centre road" development; (d) modified genes for "town centre road" development.

**Table 7**
Improvement of the combined utility value for various development alternatives

| | MPSI | MPFD | MNN | AI | *U* |
|---|---|---|---|---|---|
| The original genes | 1.4988 | 1.0657 | 258.7611 | 78.5389 | 0.0000 |
| Cloning the genes of the city proper | 1.4409 | 1.0586 | 249.6434 | 79.4347 | 0.4350 |
| "City centres transport" concentrated | 1.3529 | 1.0521 | 249.2032 | 84.2656 | 0.9242 |
| "Town centres transport" concentrated | 1.3965 | 1.0566 | 245.0450 | 81.4199 | 0.7183 |

improvement of the utility is obtained by modifying these existing "genes" according to this heuristic method.

## 5. Conclusion

This study indicates that cellular automata can become a useful exploratory tool for formulating compact development. Rapid urban development has resulted in intensive land use conflicts in many fast growing countries. Compact development can be formulated to alleviate land use problems in these regions. The simulation, prediction, and optimization of urban development are essential for promoting compact cities. CA can be used to simulate the evolution of cites by using local rules. These models can be also used to assist land use planning by incorporating planning objectives in the simulation. In urban CA, the contribution of each spatial variable to the simulation outcomes is quantified by its parameter or weight. Each set of parameters can be analogous to "genes" in biology because they control the process of urban evolution in the simulation.

In this paper, the "genes" related to urban morphology are determined by a calibration procedure. Genetic algorithms are used to find suitable "genes" of each sub-region by using empirical data from remote sensing and GIS. Better simulation results can be obtained by using separate transition rules instead of unified ones. There are spatial variations of urban dynamics in a large complex region due to localized land use policies. It is better to divide a complex region into sub-regions for producing more consistent simulation results.

Some good "genes" can be identified according to the assessment of urban morphology using some spatial metrics. The morphological utility can be conveniently calculated after the classification of remote sensing images. It is possible to produce better urban forms by replacing existing "genes" with better "genes" based on the assessment. The "genes" of good performance can be cloned from a city to other cities to improve urban morphology.

The existing "genes" can be further modified according to a heuristic swapping method. This provides an operational method to create more compact patterns around urban centres, town centres, and transport networks. The modification is accomplished by the interactive increase of the polarized effect of land development around urban centres, town centres and transport networks. Experiments indicate that significant improvement of the utility in terms of urban morphology is obtained by modifying existing "genes".

## Acknowledgements

This study was supported by the National Outstanding Youth Foundation of China (Grant No. 40525002), the National Natural Science Foundation of China (Grant No. 40471105), and the Hi-tech

Research and Development Program of China (863 Program) (Grant No. 2006AA12Z206).

## References


Banister, D., Watson, S., Wood, C., 1997. Sustainable cities: transport, energy, and urban form. Environ. Plann. B 24, 125–143.

Batty, M., Xie, Y.C., 1994. From cells to cities. Environ. Plann. B 21, 531–548.

Bourne, L.S., 1992. Self-fulfilling prophecies? Decentralization, inner city decline, and the quality of urban life. J. Am. Plann. Assoc. 58 (4), 509–513.

Clarke, K.C., Gaydos, L.J., 1998. Loose-coupling a cellular automata model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. Int. J. Geogr. Inf. Sci. 12 (7), 699–714.

Congalton, R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data. Remote. Sens. Environ. 37, 35–46.

Couclelis, H., 1997. From cellular automata to urban models: new principles for model development and implementation. Environ. Plann. B 24, 165–174.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimisation and Machine Learning. Addison-Wesley, Reading, MA, p. 412.

Herold, M., Couclelis, H., Clarke, K.C., 2005. The role of spatial metrics in the analysis and modeling of urban land use change. Comput. Environ. Urban. 29, 369–399.

Herold, M., Goldstein, N.C., Clarke, K.C., 2003. The spatiotemporal form of urban growth: measurement, analysis and modeling. Remote Sens. Environ. 86, 286–302.

Itami, R.M., 1994. Simulating spatial dynamics: cellular automata theory. Landsc. Urban. Plan. 30, 24–47.

Jenks, M., Burton, E., Williams, K., 1996. Compact cities and sustainability: an introduction. In: Jenks, M., Burton, E., Williams, K. (Eds.), The Compact City: A Sustainable Urban Form? E&FN SPON, London, pp. 11–12.

Li, X., Yeh, A.G.O., 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. Int. J. Geogr. Inf. Sci. 14 (2), 131–152.

Li, X., Yeh, A.G.O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. Int. J. Geogr. Inf. Sci. 16 (4), 323–343.

Li, X., Yeh, A.G.O., 2004a. Data mining of cellular automata's transition rules. Int. J. Geogr. Inf. Sci. 18 (8), 723–744.

Li, X., Yeh, A.G.O., 2004b. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landsc. Urban. Plan. 69 (4), 335–354.

McGarigal, K., Marks, B.J., 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. General Technical Report PNW-GTR-351. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Silva, E.A., 2004. The DNA of our region: artificial intelligence in regional planning. Futures 36 (10), 1077–1094.

Silva, E.A., Clarke, K.C., 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. Comput. Environ. Urban. 26, 525–552.

Straatman, B., White, R., Engelen, G., 2004. Towards an automatic calibration procedure for constrained cellular automata. Comput. Environ. Urban. 28, 149–170.

Ward, D.P., Murray, A.T., Phinn, S.R., 2000. A stochastically constrained cellular model of urban growth. Comput. Environ. Urban. 24, 539–558.

Weber, C., Puissant, A., 2003. Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. Remote. Sens. Environ. 86, 341–352.

White, R., Engelen, G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environ. Plann. A 25, 1175–1199.

Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. Int. J. Geogr. Inf. Sci. 16 (8), 795–818.

Wu, F., Webster, C.J., 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environ. Plann. B 25, 103–126.

Yeh, A.G.O., Li, X., 2001. The need and challenges for compact development in the fast growing areas in China—the Pearl River Delta. In: Jenks, M., Burgess, R. (Eds.), Compact City: Sustainable Urban Form for Developing Countries. SPON Press, London, pp. 73–90.