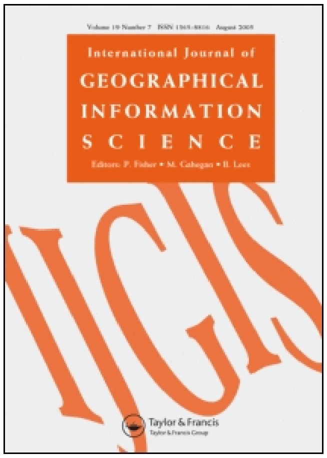

International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713599799

An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region

Xia Li [a]; Xiaoping Liu [a]
[a] School of Geography and Planning, Sun Yat-sen University. Guangzhou, 510275. PR China

To link to this article: DOI: 10.1080/13658810600816870
URL: http://dx.doi.org/10.1080/13658810600816870

International Journal of Geographical Information Science
Vol. 20, No. 10, November 2006, 1109–1136

Research Article

An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region

XIA LI* and XIAOPING LIU
School of Geography and Planning, Sun Yat-sen University, Guangzhou, 510275, PR China

(*Received 17 October 2005; in final form 12 May 2006*)

Rule-based cellular automata (CA) have been increasingly applied to the simulation of geographical phenomena, such as urban evolution and land-use changes. However, these models have difficulties and uncertainties in soliciting transition rules for a large complex region. This paper presents an extended cellular automaton in which transition rules are represented by using case-based reasoning (CBR) techniques. The common k-NN algorithm of CBR has been modified to incorporate the location factor to reflect the spatial variation of transition rules. Multi-temporal remote-sensing images are used to obtain the adaptation knowledge in the temporal dimension. This model has been applied to the simulation of urban development in the Pearl River Delta which has a hierarchy of cities. Comparison indicates that this model can produce more plausible results than rule-based CA in simulating this large complex region in 1988–2002.

Keywords: Cellular automata; Case-based reasoning; GIS; Dynamic transition rules

1. Introduction

Cellular automata (CA) were first proposed by Ulam and Von Neumann to investigate the logical nature of self-reproducible systems in the 1940s (White and Engelen 1993). They were primarily used to solve computation problems in the design of digital computers in the early days. However, they were soon applied to the simulation of either physical or artificial complex systems in many disciplines, such as lattice gases and propagation phenomena in physics, and crystal growth in chemistry (Wolfram 1984, Binder 1989, Goles 1989). There is a growing body of research of using CA to simulate and predict complex geographical phenomena. Many interesting studies have been documented in the simulation of residential development (Deadman *et al.* 1993), spatial dynamics of animals' population (Couclelis 1985, 1988), wildfire propagation (Clarke *et al.* 1994), changes of desert landscape (Jenerette and Wu 2001), forest evolution (Favier and Dubois 2004), landscape dynamics (Soares *et al.* 2002), evolution of ecological systems (Wu and David 2002), urban expansion (Batty and Xie 1994, Wu and Webster 1998, Clarke *et al.* 1997), and land-use changes (White *et al.* 1997, Li and Yeh 2002).

CA are discrete models in the spatio-temporal dimension that can be easily implemented in computer programs. These models can be conveniently incorporated

*Corresponding author. Email: lixia@mail.sysu.edu.cn

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810600816870

into GIS when the raster programming environment is provided. For example, CA can be implemented by using the AML language of a GIS package, ARC/INFO (Wu and Webster 1998, Li and Yeh 2000). CA have become a powerful tool for simulating many geographical phenomena in which global patterns can be generated by just using some local rules. This 'bottom-up' approach has more advantages than the traditional 'top-down' approach which is based on strict mathematical equations (White and Engelen 1993, Batty and Xie 1994).

It is flexible and transparent to set up CA for simulating geographical phenomena. CA are usually implemented through a sequence of processes which are bounded by transition rules. These rules can be heuristically defined according to the intuitive understanding of the process of geographical phenomena. Various types of transition rules have been proposed according to experts' preferences and knowledge (Batty and Xie 1994, Clarke *et al.* 1997, Wu and Webster 1998). 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), and neural networks (Li and Yeh 2002).

There is no agreement on how to define model structures and parameter values of CA. Transition rules are usually provided in a quite relaxed way. However, it is more useful to generate realistic patterns by elaborately providing transition rules. The outcomes of CA are very sensitive to transition rules and their parameter values (Wu and Webster 1998, Li and Yeh 2002, Wu 2002). Some calibration procedures are especially useful when CA are applied to the simulation of real cities. Fine tuning of transition rules requires considerable effort, which is greatly dependent on calibration techniques and available data. A number of methods have been proposed to calibrate CA. Trial and error is a useful way to calibrate CA (Clarke *et al.* 1997, Wu and Webster 1998). A simple procedure is to use visual testing to establish parameter ranges and make rough estimates of the values of the parameters for transition rules.

Calibration of this type of dynamic spatial model is not a trivial problem (Straatman *et al.* 2004). More elaborate methods have been proposed with regard to different model structures of CA. For example, self-modification techniques have been used to calibrate the SLEUTH-CA (Clarke *et al.* 1997, Clarke and Gaydos 1998, Silva and Clarke 2002). In the SLEUTH model, diffusion, breed, spread, slope, and road coefficients control growth behaviour. The calibration allows the model to "learn" its local setting over time. This learning is quantified by the variation during calibration of the five control parameters (Silva and Clarke 2002).

Wu (2002) provides a method to estimate the global development probability by using a logistic regression model. The initial global probability is calibrated according to historical land-use data. It seems to be easy to understand the coefficients in the logistic regression equation. However, the regression model assumes that spatial variables should be independent of each other.

The calibration of CA can also be implemented by using a neural network (Li and Yeh 2002). This can solve the problem of uncertainties in defining parameters. Moreover, neural networks are able to deal with complex interactions among variables. Variables are not required to be independent of each other.

The rule induction procedure using data mining has been proposed for reconstructing explicit transition rules automatically (Li and Yeh 2004a). Remote sensing and GIS provide the basic information for the mining procedure. This procedure can minimize the uncertainties and time consumed in defining and testing transition rules because they are automatically reconstructed by machine-learning. The benefits of this method include faster speed of rule base construction, convenient calibration, and transparent rule structures.

CA can also be calibrated to find a set of weights such that the highest potential for each cell and its neighbourhood corresponds to the desired state for that cell (Straatman *et al.* 2004). The coupling of a neighbourhood and its desired cell is established by a series of maps of the past. Some automatic search methods, such as the width search, are used to find possible improvements for reducing total error. The iterative procedure is applied for deriving the optimal weights when a minimum of two high-quality land-use maps are available.

Although the above calibration methods are useful, concrete knowledge about the application domain must be available for extracting explicit transition rules. The rule-based techniques assume that knowledge is well bounded and can be clearly expressed. However, there are difficulties in soliciting rules which deal with complicated elements and relations (Watson 1997). The constructing of rule bases is very time-consuming because a large set of rules is usually required for solving a real application (McAvoy and Krakowski 1989, Li and Yeh 2004c).

Most of the existing methods assume that transition rules are invariant in the spatio-temporal dimension. The same set of transition rules will be applied to any location and time, regardless of the possible changes to the simulation environment. In reality, the relationships between the state conversion and its geographical variables may be complex for real geographical phenomena. There are difficulties in providing invariant rules to describe the relationships which may vary significantly in the spatio-temporal dimension. It is unreasonable to use the same set of rules when a study area is large, and the simulation period is long. The use of 'fixed' rules will produce poor simulation results for a large region consisting of a hierarchy of cities. This is because each of these cities may experience a different growth stage associated with a unique growth pattern (Li and Yeh 2004b). The transition rules should be further relaxed to accommodate the variations in control factors and parameter values so that better simulation results can be produced.

Case-based reasoning (CBR) techniques can help to solve these difficult knowledge solicitation problems in implementing CA. CBR can be traced back to the work of Roger Schank and his students at Yale University in the early 1980s (Kolodner 1993, Watson 1997). It is developed to overcome the problems of rule-based systems (KBS) but has the advantages inherited from KBS, such as artificial intelligence, reduction of repetitive tasks, and highly automated capability. It can save much of the time in applications because eliciting rules from past experiences are not required. The case-based reasoning (CBR) approach has more appealing features than the traditional rule-based approach. CBR can also allow users to find the solutions in domains that are not completely understood by them (Watson 1997). Moreover, CBR can well handle the domains where problems have many exceptions to rules (Holt and Benwell 1999).

Recently, CBR has been used to solve the problems in environmental sciences, urban planning, and geography (Branting and Hastings 1994, Lekkas *et al.* 1994, Yeh and Shi 1999). For example, CBR methods are used to classify soil types (Holt and Benwell 1999) and predict air pollution (Lekkas *et al.* 1994). CBR can also be a useful tool for urban planning and management (Yeh and Shi 1999).

This paper presents a new method to establish CA using case-based reasoning (CBR) techniques. The use of discrete cases can provide a powerful tool to capture the complexities exhibited in many geographical phenomena which may have exceptions, irregularities, and variations. The advantages of CBR should allow case-based CA to have much better potential for self-learning and adaptation than traditional rule-based CA. Moreover, this paper presents an improved k-NN algorithm to capture the spatial variations of transition rules so that the proposed CA can have a better performance.

2. Methodology

2.1 *Rule-based cellular automata*

In traditional CA, transition rules are used to determine the state conversion of each cell at time t. Strict transition rules have been relaxed for simulating complex geographical phenomena by using various forms of neighbourhood functions. For example, the probability of land-use conversion can be expressed as a function of a series of spatial variables (Wu and Webster 1998):

$$P(i) = \sum_{l=1}^{N} w_l a_l(i) \tag{1}$$

where $P(i)$ is the probability of land-use conversion (e.g. from agricultural land use to urban land use) at location i, $a_l(i)$ is the lth site attribute (dependent variable) at location i, w_l is the weight for the site attribute, and N is the total number of variables.

It is essential to determine the weight for each variable. A common method to determine these weights is based on the multicriteria evaluation (MCE) techniques (Wu and Webster 1998). However, this MCE-CA may have uncertainties, since its weights are determined according to experts' knowledge. A modified version is to employ logistic regression to calibrate CA by using empirical data (Wu 2002):

$$P(i) = \frac{\exp(z(i))}{1+\exp(z(i))} = \frac{1}{1+\exp(-z(i))} \tag{2}$$

where $z(i) = w_0 + w_1 a_1(i) + w_2 a_2(i) + \cdots + w_N a_N(i)$.

More methods have been proposed to extract appropriate CA transition rules, including computer exhaustive searches (Clarke and Gaydos 1998), neural networks (Li and Yeh 2002), and data mining (Li and Yeh 2004a). Complex global patterns can be generated by using various types of transition rules. However, these strict transition rules or equations may have problems in representing complex relationships between the probability of state conversion and its independent variables. Standard CA use 'fixed' rules in the spatio-temporal dimension for simulating complex systems. The relationships are independent of locations and time, since the parameter values will not change during simulation. The strict transition rules may have difficulties in representing realistic spatio-temporal processes in geography, which are often more complex than those in other disciplines. The simulation of geographical processes will encounter at least the following problems:

- Neighbourhood influences are very complex with various kinds of distance decay functions. The calibration is difficult because of involving a huge combination of parameters.

- There are hidden and external factors which are beyond human knowledge.
- Environmental factors or constraints are rapidly changing under accelerating economic growth and environmental degradation.
- A region to be simulated is usually very complex and may consist of a hierarchy of cities with various growth patterns and processes.

Standard CA are based on 'fixed' neighbourhood windows for representing transition rules. A larger window size may be preferable for reflecting complex relationships, but calibration of such models is very difficult. Moreover, explicit transition rules may have difficulties in capturing complex relationships which have exceptions and irregularities. These exceptions and irregularities usually come from the unexplained variables that cannot be covered by these rules. Therefore, the rule-based approach has difficulties in balancing the trade-off between neighbourhood window sizes and complexities for simulating geographical phenomena. It may create significant discrepancy between the simulated and the actual patterns, especially when the study area is large, and the simulation period is long. Extended cellular automata should be useful for dealing with these problems by relaxing transition rules.

2.2 *Extended cellular automata using case-based reasoning*

The use of case-based reasoning (CBR) techniques can avoid knowledge-soliciting problems in CA simulation. A CBR system for solving geographical problems usually consists of: (1) construction of case library using geographical data; (2) case matching; (3) problem solving; and (4) verification and adaptation according to training data (figure 1). Cases are the basic units for reasoning process in a CBR system. Case-based learning is often termed 'lazy learning', as there is typically no 'transformation' of training cases into more general 'statements' or rules. Instead, the presented training data are simply stored, and when a new queried case is encountered, a set of similar, related cases are retrieved from memory and used to provide solutions to the new queried case.

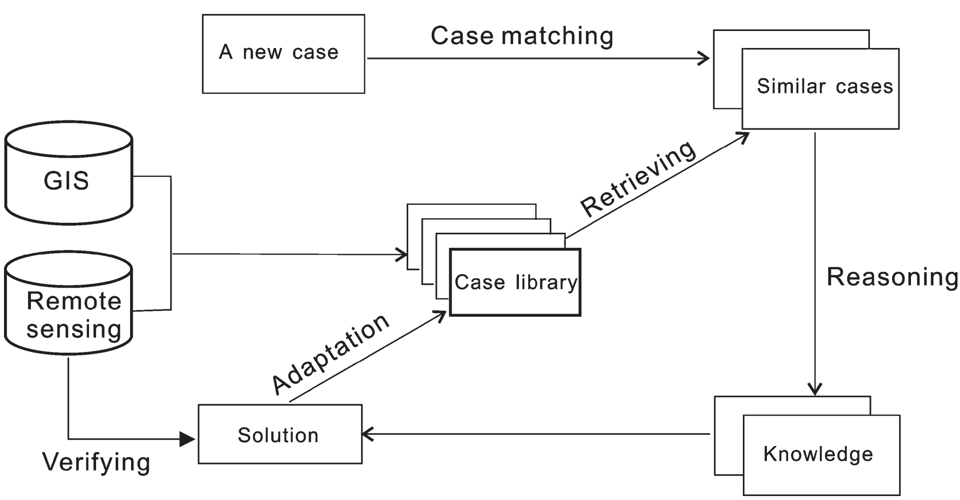

Figure 1. CBR system for solving geographical problems.

In a case library, each case is represented by a description of the problem, plus a solution and/or the outcome. The knowledge for solving a problem is not recorded but is implicit among the cases. Similar cases will be matched and retrieved in the case library to solve a current problem. The retrieved cases are used to suggest a possible solution which can be reused and tested, or even adapted to a current problem. This can allow users to formulate solutions to problems quickly and avoid the time necessary to derive those answers from scratch. CBR can also allow users to find the solutions in domains that are not completely understood by them (Watson 1997). CBR can easily handle the domains where problems have many exceptions to rules (Holt and Benwell 1999).

Discrete cases can be conveniently obtained to represent spatio-temporal variations of transition rules for CA. Figure 2 shows the methodology of using case-based reasoning for establishing CA. Detailed procedures are provided in the following sections.

2.2.1 Establishing case library. The first step of this proposed method is to establish the case library which will replace explicit transition functions or rules for traditional CA. A case is a contextualized piece of knowledge representing an experience that can help a reasoner to achieve his goals (Kolodner 1993). In this proposed urban CA, each case is represented by two parts: (1) attributes (features) of each cell; (2) state conversion (solution). The first part includes proximity (distance) variables and land-use types. The second part is the solution, which determines if the state will be converted or not (e.g. urbanized or not).

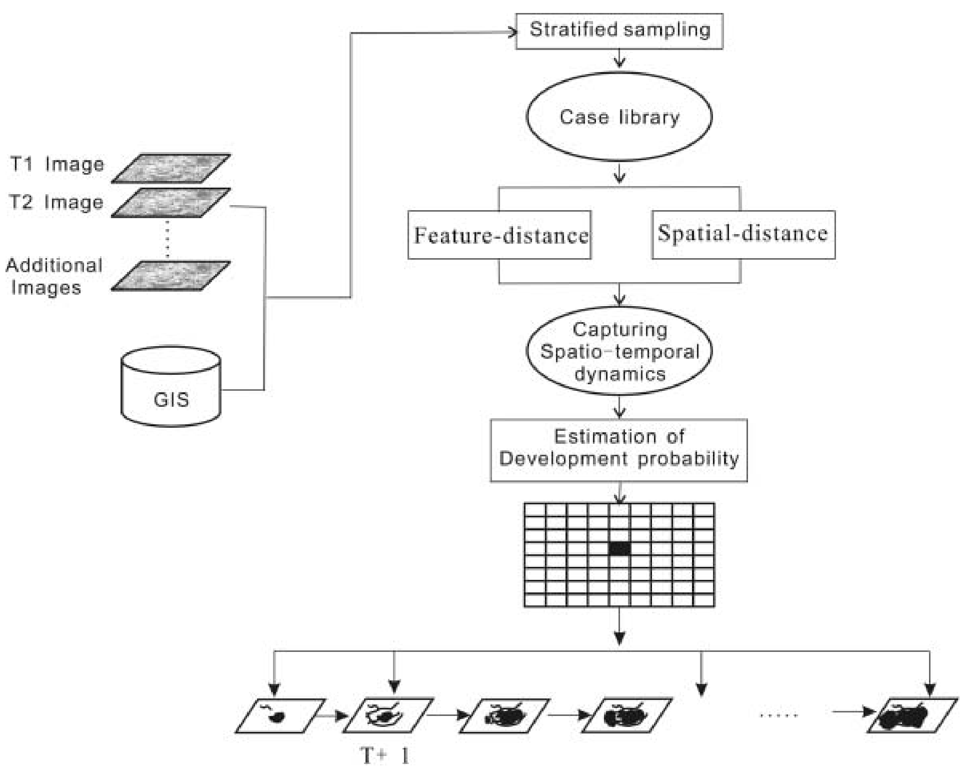

Figure 2. Case-based approach for cellular automata.

A case can then be represented as follows:

$$I = (a_1(i), a_2(i), \cdots, a_N(i); s) \tag{3}$$

where $a_1(i), a_2(i), \cdots, a_N(i)$ are the attributes (features) of case i in terms of a series of spatial variables—proximity variables and existing land-use types. s is a Boolean variable in which the urbanized is 1, and the non-urbanized is 0. N is the total number of features (variables).

Empirical data from remote sensing and GIS will be used to establish the case library. First, multi-temporal satellite images will be acquired and classified so that the data about land use and land-use conversion (state conversion) are obtained for each cell. Second, GIS analyses will be carried out to calculate the site attributes of each cell. They include the proximity variables—from a location to a series of 'attraction' centres (e.g. city centres, town centres, roads, expressways, and railways). A closer distance to these centres will have a higher probability of land-use conversion (Wu and Webster 1998). Stratified randomly sampling methods (Li and Yeh 2002) can then be employed to retrieve only a portion of original data as the cases.

2.2.2 Improved k-NN algorithm for reflecting spatio-temporal dynamics. The case library stores the experience (knowledge) that can determine if the state of a cell will be converted in the simulation. Case matching should be carried out to retrieve the experience for making the decision. The matching is usually based on the similarity between an input (questioned) case (i) and a known case (j) of the samples in the case library. The similarity can be calculated by using the following Euclidean distance function:

$$d(i,j) = \sqrt{\sum_{l=1}^{N} (a_l(i) - a_l(j))^2}, \tag{4}$$

where $a_l(i)$ is the lth feature of a case.

Two cases will be more similar if they have a closer Euclidean distance in the feature space. The similarity is calculated according to a number of features, which can be treated equally for their importance. In some situations, the importance of each feature may be different in the calculation. Weights can be assigned to address the contribution of each feature in calculating the similarity. The above equation is then revised as follows:

$$d(i,j) = \sqrt{\sum_{l=1}^{N} w_l^2 (a_l(i) - a_l(j))^2} \tag{5}$$

where w_l is the weight for the lth feature.

In this study, the weights are determined by an entropy method, since there is no prior knowledge. First, the variables are standardized into the range of [0, 1]. The entropy is used to represent the content of information (Theil 1967, Thomas 1981). It can be calculated as follows:

$$H_n = \sum_{i=1}^{n} p_i \log(1/p_i)/\log(n) \tag{6}$$

where $p_i = a(i) \Big/ \sum_{i=1}^{n} a(i)$, and n is the total number of observation.

The entropy falls within the range of [0, 1]. The smallest value of 0 represents the maximum amount of the information exhibited in the variable. The largest value of 1 indicates the minimum amount of the information. Therefore, the amount of the information is directly proportional to this form: $1 - H_n$.

A feature that has a larger amount of information is expected to have a larger weight. Then, the entropy weight for the lth feature can be represented as follows:

$$\phi_l = \frac{1 - H_{nl}}{n - \sum_{i=1}^{n} H_{nl}} \tag{7}$$

The entropy weights can then be integrated with the initial (subjective) weights to produce the integrated weights (Xu 2004). These final adjusted (integrated) weights can be computed as follows:

$$w_l = w_l^0 \cdot \phi_l \Big/ \sum_{i=1}^{N} w_i^0 \cdot \phi_i \tag{8}$$

where w_l^0 is the initial weight defined by experts, and w_l is the final adjusted weight for the lth feature.

The reasoning first locates the queried case i to its nearest known case j in the library. The known case has a target function of $f(j)$ (e.g. indicating converted or not). The reasoning assumes that the case closest to j tends to have a target function close to $f(j)$. Actually, the matching is often carried out by comparing the queried case with a number of known cases, its k-nearest neighbours. This is the most popular k-nearest-neighbour (k-NN) algorithm, which works well on many practical problems and is fairly noise-tolerant in CBR applications (Dasarathy 1991).

Intuitively, the k-NN algorithm assigns to each new queried case the majority class (state) among its k nearest neighbours. For a discrete value of $f(j)$, it can be estimated by using the following expression (Dasarathy 1991, Houben *et al.* 1995):

$$\hat{f}(i) \leftarrow \arg\max_{s \in S} \sum_{j=1}^{k} \delta(s, f(j)) \begin{cases} \delta(s, f(j)) = 1, \text{if } s = f(j) \\ \delta(s, f(j)) = 0, \text{if } s \neq f(j) \end{cases} \tag{9}$$

where k is the total number of the nearest neighbours, and s is the finite set of target class value. In this study, it represents the state of a cell (e.g. 1 for urbanized and 0 for non-urbanized).

It is possible that the use of different k values may have different enquiry results. For example, the queried case is classified as A by using five nearest neighbours, but as B by using 10 nearest neighbours (figure 3). The appropriate value of k is considered to be problem-dependent (Houben *et al.* 1995). Experiments can be carried out to find out the appropriate value for an application.

The above equation assumes that each neighbour has the same vote in determining the target class value (state). However, it is reasonable that a closer neighbour should have more influences than others in making the decision. A distance-weighted function can be used to treat these neighbours differently. The contribution of each neighbour is decided according to its distance to the queried

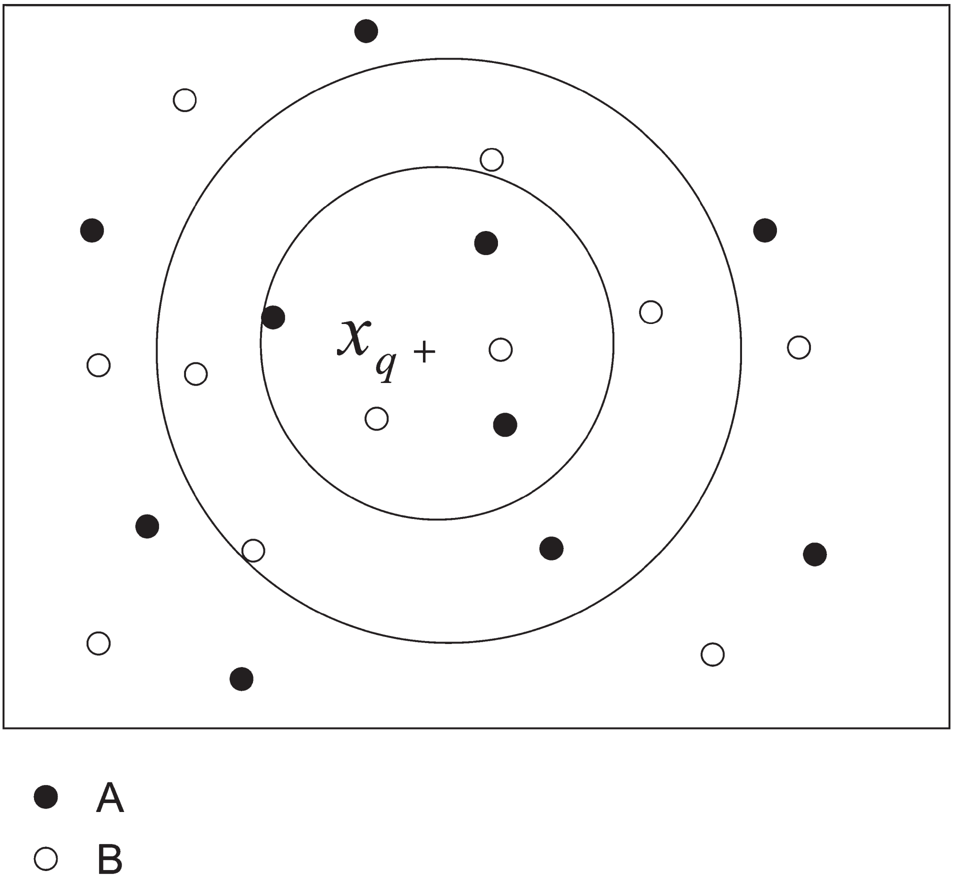

Figure 3. Impacts of k values on reasoning results—classified as A using five neighbours, but as B using 10 neighbours.

case. Equation (9) is then revised by adding the weights to the inference equation (Jia and Richards 2005):

$$\hat{f}(i) \leftarrow \underset{s \in S}{\arg\max} \sum_{j=1}^{k} w_{fj} \cdot \delta(s, f(j)) \tag{10}$$

These feature-distance weights (w_{fj}) are proportional to the inversed distance function:

$$w_{fj} = \frac{1}{d(i,j)^2} \tag{11}$$

where $\hat{f}(i) = f(j)$ if $d(i,j) = 0$.

The above k-NN algorithm only deals with the proximity influences in the feature space. The distance-weighted function, which is calculated in the feature space, cannot reflect the spatial variations of transition rules of CA. It is obvious that the influence of a case may not be the same at different spatial locations in the reasoning process. The location information should be included as a part of the attribute for a case. This can be done by adding the coordinates of a case as additional attributes in equation (3).

Therefore, the distance-weighted function should have two parts: (1) feature-distance-weighted; and (2) spatial-distance-weighted. The later element is essential in dealing with spatial variations of transition rules. The spatial-distance weights (w_{sj}) can be defined to count the proximity influences related to case locations in original space. w_{sj} is written as follows:

$$w_{sj} = \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \tag{12}$$

where x and y are the coordinates of a case.

Finally, equation (10) is revised as follows:

$$\hat{f}(i) \leftarrow \arg\max_{s \in S} \sum_{\xi = j}^{k} w_{fj} w_{sj} \cdot \delta(s, f(j)) \tag{13}$$

or

$$\hat{f}(i) \leftarrow \arg\max_{s \in S} \sum_{j=1}^{k} W_j \cdot \delta(s, f(j)) \tag{14}$$

where $W_j = w_{fj} w_{sj}$

This method adopts a slightly different approach to address the neighbourhood influences of CA by using the modified k-NN algorithm. The neighbourhood window is the most basic element of CA. Standard CA usually adopt 'fixed' neighbourhood windows to represent transition rules (White and Engelen 1993). However, this proposed method has only the 'fixed' neighbours, but its neighbourhood window varies in size in either the original space or the feature space. The modified k-NN algorithm uses fixed k nearest neighbours to represent the transition rules of CA. The neighbourhood influences are quite complex, since these neighbours are determined in both the original and feature spaces. General calibration procedures for rule-based CA are not suitable for addressing these complex neighbourhood influences. The lazy-learning techniques from CBR have advantages in soliciting inexplicit transition rules for CA. The simplicity of CA can still be maintained by using these convenient techniques.

2.2.3 Estimating development probability using case-based reasoning. The state conversion of a cell can be determined according to its k-nearest neighbours by using CBR. However, this can only yield a Boolean value—converted or not. The conversion probability is often used to produce more plausible simulation results (Wu 2002). In an urban CA, there are two states—1 (urbanized) and 0 (non-urbanized). The conversion probability can thus be estimated by using the following equation

$$P_{\text{proximity}}(i) = \frac{\sum_{j=1}^{k} W_j \cdot \delta(1, f(j))}{\sum_{j=1}^{k} W_j \cdot \delta(1, f(j)) + \sum_{j=1}^{k} W_j \cdot \delta(0, f(j))} \tag{15}$$

where $P_{\text{proximity}}(i)$ is the development probability for cell i.

There are other factors that play roles in determining the conversion probability. The amount of urbanized cells in the neighbourhood can significantly influence urban development. For example, more developed cells in the neighbourhood will increase conversion probability (Batty and Xie 1994). The conversion probability related to the neighbourhood function is as follows:

$$P_{\text{neigh}}(i) = \kappa \sum_{\Omega} N(i) \tag{16}$$

where $N(i)=1$ for urbanized cells and $N(i)=0$ for others; κ is a scaling coefficient.

The final conversion probability is the joint of these two probabilities, plus some constraint factors (e.g. topography and land use) which are used to adjust the conversion probability. For example, it is unlikely that urban development takes place at the locations with steep slopes and at the locations of water. The final probability is presented by using the following equation:

$$\begin{aligned} P(i) &= P_{\text{proximity}}(i) \times P_{\text{neigh}}(i) \times \left(1 - \sum_{\gamma} \delta_{\gamma}(i)\right) \\ &= \kappa \frac{\sum_{j=1}^{k} W_j \cdot \delta(1, f(j))}{\sum_{j=1}^{k} W_j \cdot \delta(1, f(j)) + \sum_{j=1}^{k} W_j \cdot \delta(0, f(j))} \times \sum_{\Omega} N(i) \times \left(1 - \sum_{\gamma} \delta_{\gamma}(i)\right) \end{aligned} \tag{17}$$

where $\delta_{\gamma}(i)$ is the γth constraint factor with the range of [0, 1]. A larger value of $\delta_{\gamma}(i)$ indicates a higher degree of constraint influence for the factor.

Urban dynamics has manifested some degree of uncertainties, which cannot be reflected by using deterministic models (White and Engelen 1993, Batty and Xie 1994). Stochastic CA can reproduce some complex features of urban systems by properly adding perturbation in urban simulation (White and Engelen 1993). These models are usually implemented by using the Monte Carlo method. In this study, the final land-use conversion is determined by comparing the development probability with a random variable (Wu and Webster 1998):

$$S_{t+1}(i) = \begin{cases} \text{Converted}, P(i) > \text{Rand()} \\ \text{Non} - \text{converted}, \text{Others} \end{cases} \tag{18}$$

where Rand() is a random variable ranging from 0 to 1.

The initial case library is produced at the beginning of the simulation by using t_1 and t_2 satellite images. The case library provides the knowledge for determining land-use conversion in the simulation. The library can be updated by using the same procedure if additional t_3 and t_4 satellite images are available for reflecting the possible changes of relationship. New cases are obtained from these remote-sensing data and stored in the case library. This is useful for reflecting the possible changes to the simulation environment. The dynamic transition rules in the temporal dimension can then be easily implemented by using this updated library.

There are questions if this type of case-based models still belongs to CA, which assume the use of 'fixed' transition rules. One may argue that the same rules should be applied to anywhere and anytime for standard CA. However, the assumption is still valid for this extended CA under the case-based framework by using the same k-NN algorithm instead of the same rules. This can provide more flexibility of using

CA to simulate complex geographical phenomena. Actually, the strict transition rules of traditional CA have already been relaxed when they are applied to the simulation of urban systems by using various model structures (White and Engelen 1993, Batty and Xie 1994, Clarke and Gaydos 1998, Li and Yeh 2002, Wu 2002). The case-based approach is another way to relax transition rules so that the spatio-temporal complexity exhibited in geographical phenomena can be appropriately addressed.

3. Implementation and results

3.1 *Study area and data*

The proposed model is applied to the simulation of fast urban expansion in the Pearl River Delta which has an area of about 41 157 km^2 (figure 4). It has led the nation in economic growth and urbanization with unprecedented land-use changes since the economic reform in 1978. Fast urban expansion has triggered the loss of a large amount of agricultural land in this region (Li and Yeh 1998).

Instead of a single city, this simulation involves a hierarchy of cities which have different scales of population and economic development (figure 4). Internal variations of land-use change patterns and development processes can be identified in this agglomeration region (Li and Yeh 2004b). Traditional CA which are based on static transition rules may have difficulties in simulating urban development in this large area because of its inhomogeneous geographical features.

In this study, the initial training data about land-use conversion were obtained by the classification of the 1988 and 1993 TM images. The classification accuracy will definitely affect the simulation accuracy. 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 (e.g. proximity to centres) for determining land-use conversion were retrieved by using GIS functions. Stratified random sampling (Congalton 1991) was then employed to construct the case library which consists of two major groups of cases—urbanized and non-urbanized. A total of 4000 cases were obtained for representing the complex relationships between land-use conversion and its independent variables.

The case library was updated to reflect the possible change in relationships by using additional satellite images, the 1995 and 1997 TM images. Studies have indicated that the land-development patterns of the study area have changed significantly in different periods because of the variations in economic growth and land-use policies (Li and Yeh 2004b).

The weight for each feature in calculating the similarity was determined by using the entropy method in equation (8). These weights will vary during simulation when the case library has been updated. There are only very limited studies on the retrieval of the adaptation knowledge of CA so that future land-use changes can be more reasonably forecast from the past trends. For example, self-modification rules have been obtained by the SLEUTH model (Clarke *et al.* 1997). The fluctuation of global constraints can also be obtained by using data-mining techniques (Li and Yeh 2004a). However, CBR is more straightforward and convenient for acquiring adaptation knowledge through the learning from case libraries (Kolodner 1993, Wilke and Bergmann 1998).

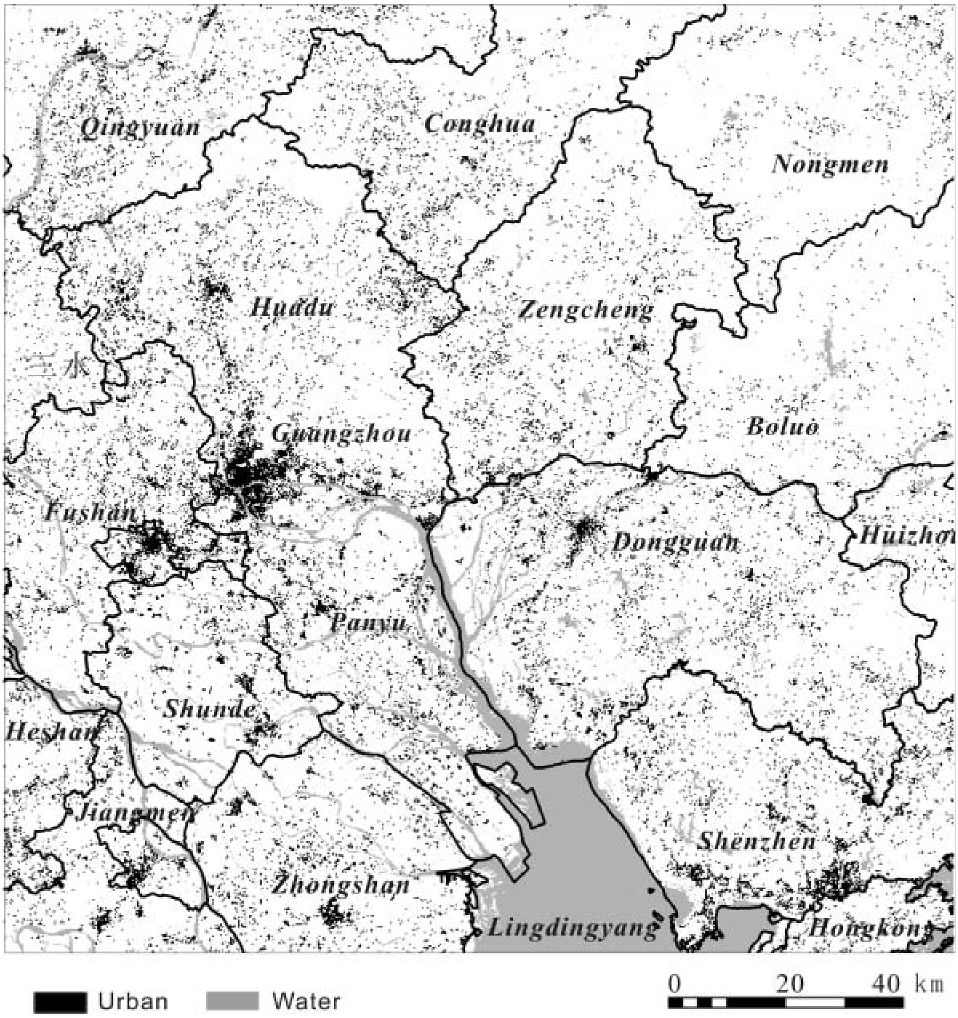

Figure 4. Simulation of urban expansion in a large region—the Pearl River Delta.

3.2 *Simulation results*

In this study, the *k*-NN algorithm is modified by using spatial-distance weights. This is useful for capturing spatial variations of urban dynamics and formulating dynamic transition rules of CA. Moreover, the transitional rules are not frozen in the temporal dimension by using multi-temporal satellite images to update the case library. New cases are collected from additional satellite images in 1995 and 1997. These new cases can be used to reflect the new relationships between land-use conversion and its independent variables. Transition rules, which are represented by cases, can thus become dynamic in the temporal dimension by updating the case library. This case-based method can thus provide a more flexible approach to capture complex relationships for urban simulation.

This CA is mainly implemented by using the modified *k*-NN algorithm. The first step was to determine the proper *k* value for the simulation. Experiments were carried out to test the influences of various *k* values on the classification results. Two hundred cases from the case library were used to examine the relationships. The experiments indicate that the increase in k values can improve the classification accuracy but significantly increase the computation time. Figure 5 is the relationship between the number of k and prediction accuracy based on the training data. The accuracies are 69.1% (k=1), 73.2% (k=10), and 73.4% (k=20), respectively. It is obvious that the accuracy improvement becomes stabilized after k is greater than 10. Therefore, 10 neighbours are used to avoid too much computation time in the simulation.

It is necessary for the simulated and actual land-use patterns to have the same amount of land-use conversion. The actual amount of land-use conversion can be obtained by classification of remote-sensing data. The amounts of future land-use conversion can also be estimated based on the growth trends. One way to capture the growth trend is to use more than two 2 of satellite images (Li and Yeh 2004a). These extra remote-sensing data can be used to provide aggregated information about the development trajectory for simulating future urban development.

There are many iterations of simulation before the outcome is obtained. A shorter interval between t and $t+1$ means that a larger number of iterations are required. The relationship between the number of iterations (K), the iteration interval (Δt) and the observation interval (ΔT) between two satellite images is as follows (Li and Yeh 2004a):

$$K = \Delta T / \Delta t \tag{19}$$

where ΔT is the observation interval for the two remote-sensing images, Δt is the iteration interval between t and $t+1$, and K is the number of iterations.

Land consumption at each period ΔQ_t is obtained from remote-sensing data. It is possible to estimate the amount of land-use conversion between t and $t+1$. The amount is calculated using the following equation:

$$\Delta q_t = \Delta Q_t / K \tag{20}$$

where ΔQ_t is the amount of land-use conversion for the observation interval, and Δq_t is the amount of land-use conversion for the iteration interval.

The final land-use conversion is usually determined by comparing the development probability with a random variable (Batty and Xie 1994, Wu and Webster 1998, Wu 2002). The random variable may introduce the 'noise' in the simulation process by reflecting the uncertainties of urban development.

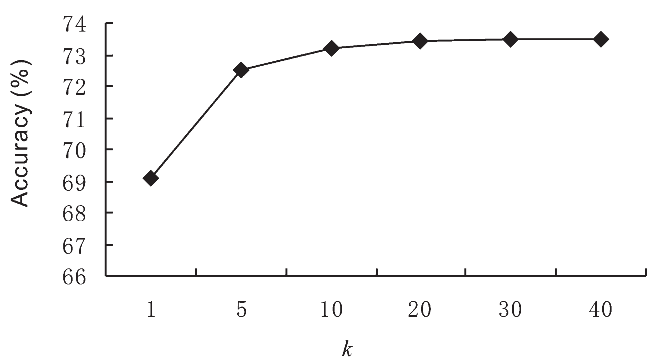

Figure 5. Relationship between the number of k and prediction accuracy.

In this study, the land-use conversion is determined using the following rules:

$$S_{t+1}(i) = \begin{cases} \text{Converted}, P(i) > \text{Rand}() \text{ and } \text{Rand}() < \Delta q_t) \\ \text{Non} - \text{converted}, \text{Others} \end{cases} \tag{21}$$

where Rand() is a random variable with the range of [0, 1].

This case-based CA was used to simulate the urban expansion of the whole Pearl River Delta in 1988–2002. The urban land use at the initial stage was classified from the 1988 TM image. The actual urban land use in 1993 and 2002, which was used as the reference data for calibration and verification, was also classified from the TM images of these two years. Figure 6 shows a comparison of the actual and simulated patterns of urban development in the Pearl River Delta in 1988, 1993, and 2002.

There are inhomogeneous growth patterns in this large study area which has a hierarchy of cities. A uniform set of transition rules has difficulties in simulating the whole complex region simultaneously for traditional rule-based CA. However, the experiment indicates that the proposed case-based CA can generate very plausible simulation results by using a single case library. Consistent simulation results are obtained, although the growth patterns vary significantly between different cities. For example, various growth patterns can be effectively simulated in the cities of Shenzhen (figure 7), Dongguan (figure 8), Guangzhou (figure 9), and Zhengchen (figure 10) using this model. A visual comparison can reveal that the simulated and actual growth patterns are very close at a local level.

3.3 *Model validation*

Validation of urban CA is required when they are applied to the simulation of real cities. It is preferable for a model to produce a high goodness of fit. It is impossible to reproduce the exact patterns of land use because of the existence of unexplained variables. A simple method to assess the goodness of fit is to compare the simulated patterns with the actual patterns visually. Figures 6–10 compare these two patterns of urban development for the whole region and several selected cities. A visual comparison can provide a rough estimation of the accuracy of this proposed model in simulating urban development.

A further quantitative analysis is to produce a confusion matrix about the concordance between the simulated and the actual development patterns. It is based on the spatial overlay of these two patterns cell by cell. Table 1 compares these two patterns in 1993 and 2002. The total accuracies are 0.86 and 0.82, and the kappa coefficients are 0.53 and 0.51 for the simulation of urban development in 1993 and 2002, respectively. The simulation in 1993 has a better accuracy because it uses the cases closer in time.

The cell-by-cell comparison has difficulties in providing the morphology information about land-use patterns. Urban studies are usually concerned about the morphological and structural features of urban land use, such as connectivity, fractals, and compactness. A convenient way to measure these features is based on several aggregated indicators, such as compactness indices (Yeh and Li 2001a), fractal dimensions (White and Engelen 1993), and Moran's I (Wu 2002).

In this study, the simple indicator of Moran's I is applied to the measurement of land-use patterns. It is quite easy to calculate Moran's I values in a GIS package, ARC/INFO GRID. Moran I is a useful spatial indicator that can reveal the degree of spatial autocorrelation (Goodchild 1986). The indicator is able to estimate how close the simulated land-use pattern is to the actual pattern (Wu 2002). The

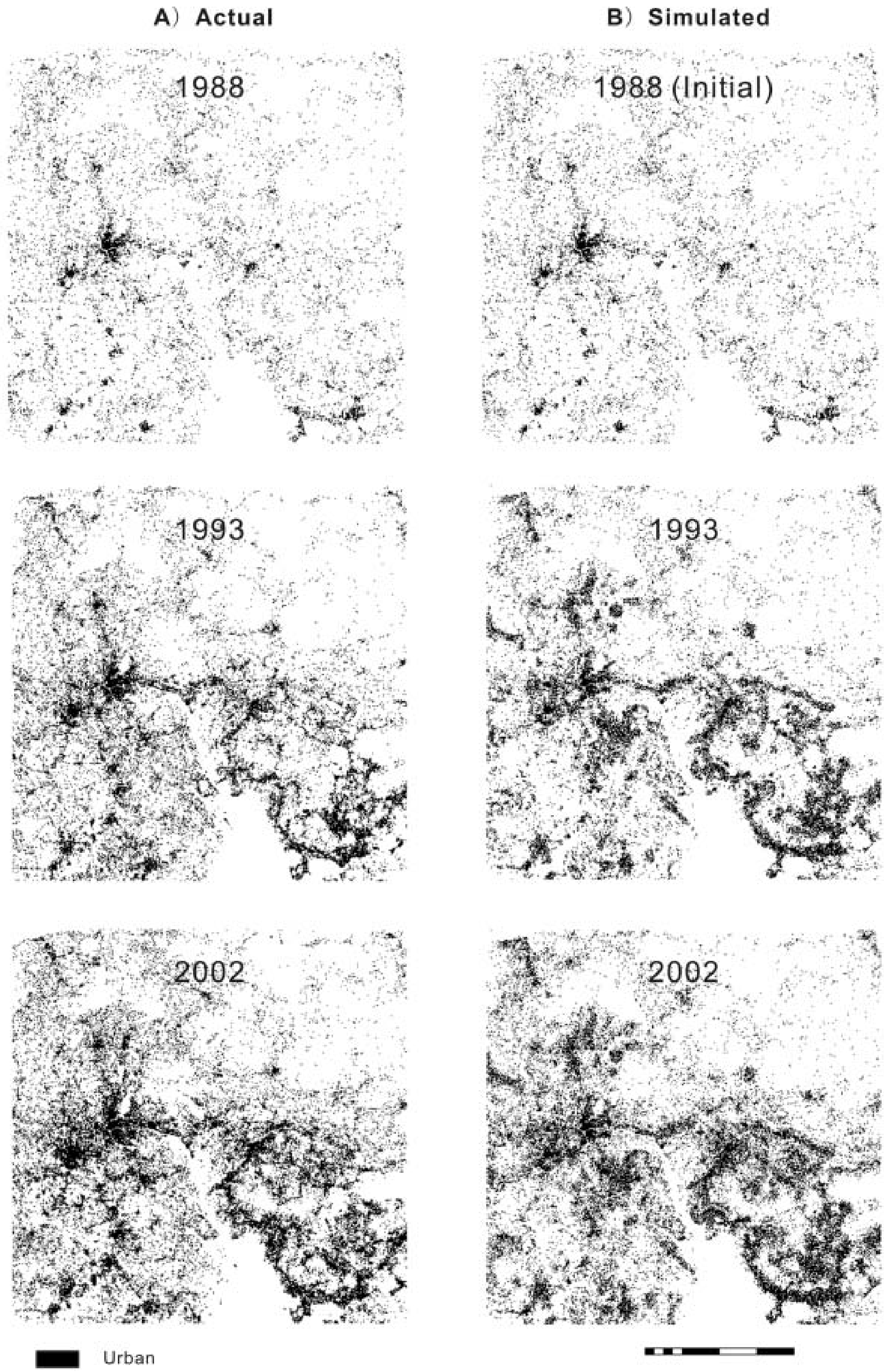

Figure 6. Actual and simulated patterns or urban development in the Pearl River Delta in 1988, 1993, and 2002.

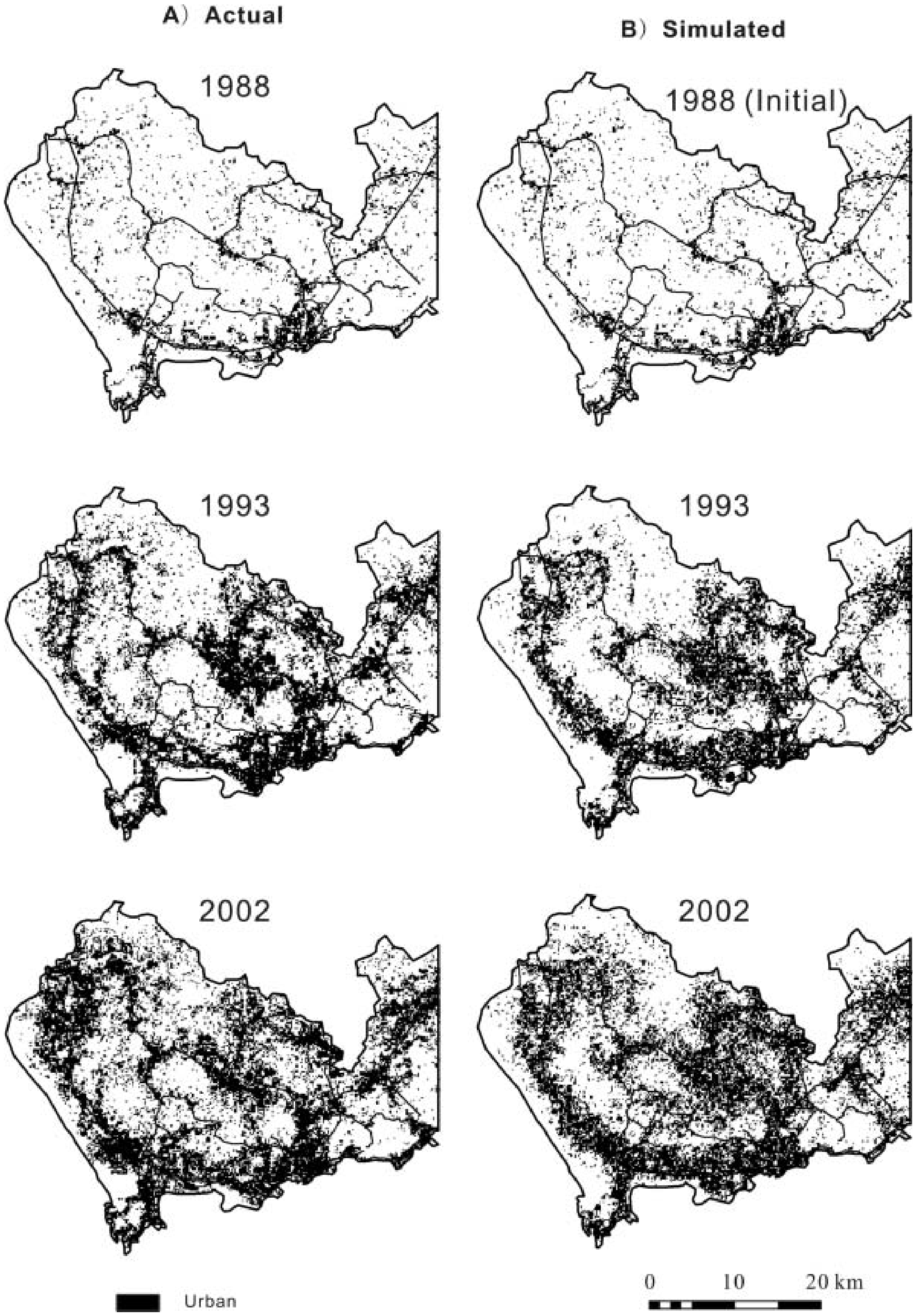

Figure 7. Actual and simulated patterns or urban development in Shenzhen in 1988, 1993, and 2002.

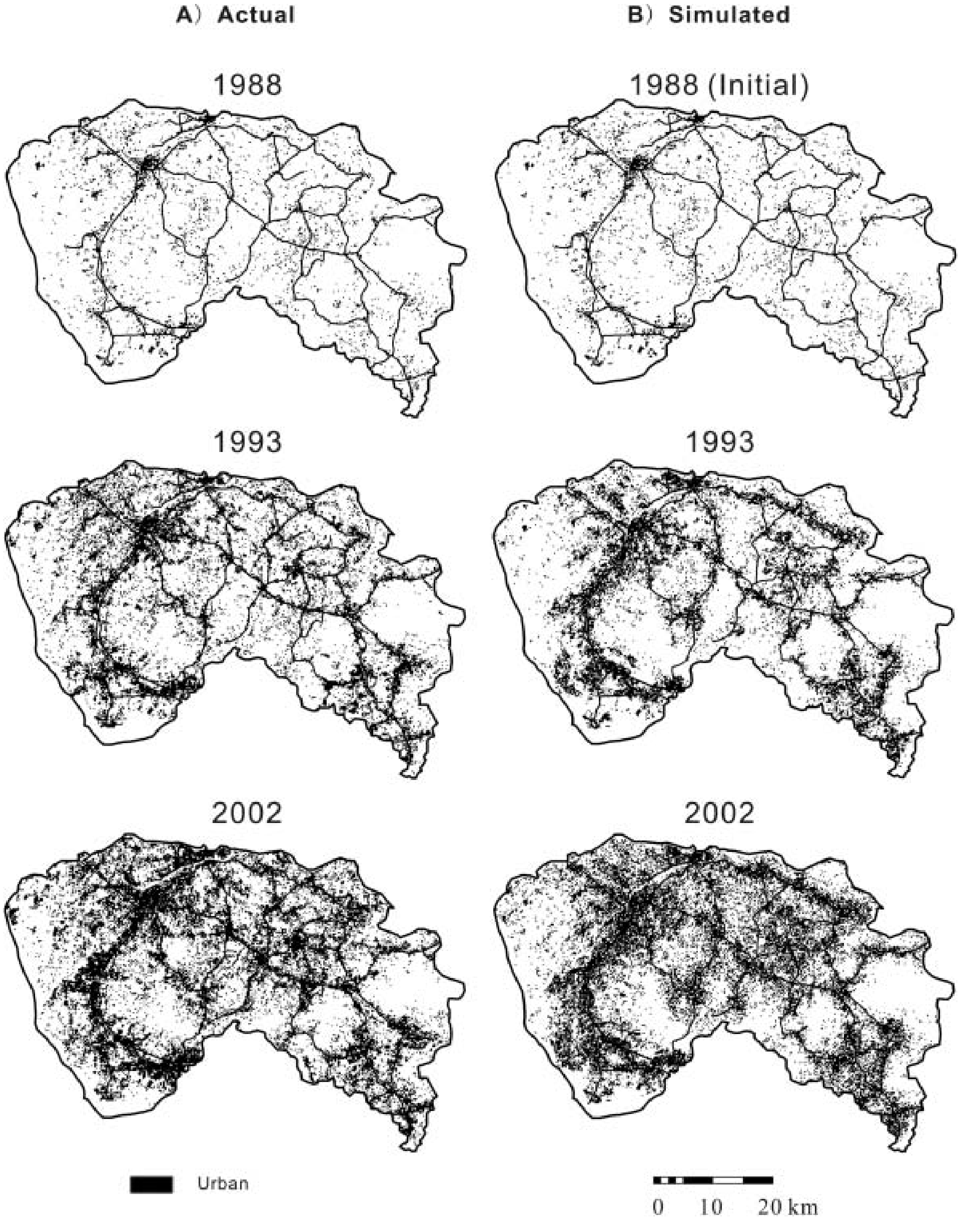

Figure 8. Actual and simulated patterns or urban development in Dongguan in 1988, 1993, and 2002.

maximum value is one which indicates the absolute concentration of a land-use type. A smaller value, which can be below zero, indicates a more even distribution of the land-use type.

The analysis indicates that the simulated patterns are very close to the actual patterns based on the Moran I indicator. The Moran I values are 0.376 and 0.382 for the actual and simulated patterns in 1993, respectively. They become 0.385 and 0.393 for the actual and simulated patterns in 2002, respectively.

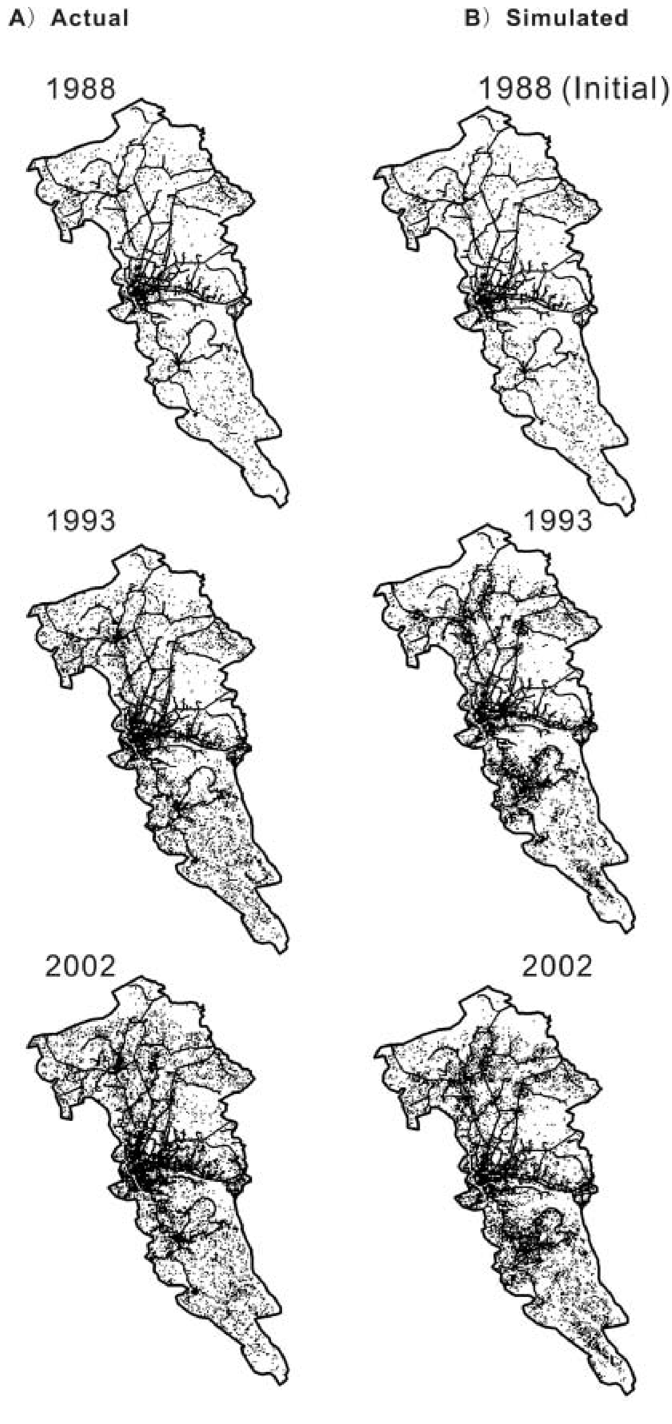

Figure 9. Actual and simulated patterns or urban development in Guangzhou in 1988, 1993, and 2002.

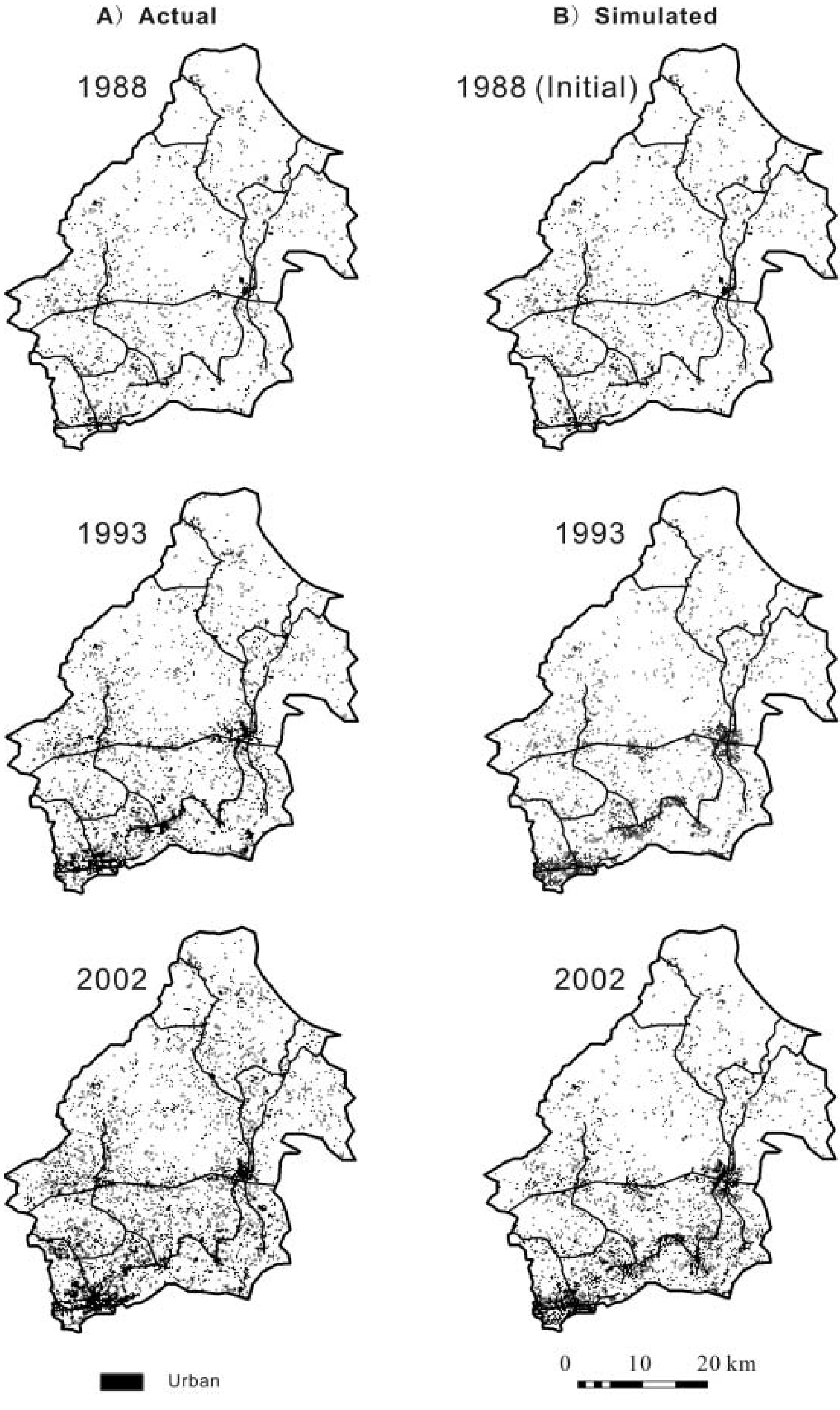

Figure 10. Actual and simulated patterns or urban development in Zengcheng in 1988, 1993, and 2002.

Table 1. Simulation accuracies of the case-based CA for the Pearl River Delta.

	1988~1993 (cells)		
	Simulated 1993 non-urban	Simulated 1993 urban	Accuracy
Actual 1993 non-urban	476 678	58 315	0.89
Actual 1993 urban	31 953	71 454	0.69
Total accuracy			0.86
Kappa coefficient			0.53

	1993~2002 (cells)		
	Simulated 2002 non-urban	Simulated 2002 urban	Accuracy
Actual 2002 non-urban	432 997	65 273	0.87
Actual 2002 urban	48 513	91 616	0.65
Total accuracy			0.82
Kappa coefficient			0.51

Another experiment is to compare the performances of this proposed model with those of traditional rule-based CA. A common logistic rule-based CA (Wu 2002) was applied to the simulation of urban development in the study area as the comparison. This simulation utilizes static transition rules in which the parameter of each variable is calibrated using a logistic regression. Table 2 shows the simulation accuracies of this logistic rule-based CA. The total accuracy is 0.81 and 0.75, and the kappa coefficients are 0.34 and 0.28 for the simulation of urban development in 1993 and 2002, respectively. It has much lower accuracies than the proposed model because the 'fixed' rules are used.

Table 2. Simulation accuracies of the logistic rule-based CA for the Pearl River Delta.

	1988~1993 (cells)		
	Simulated 1993 non-urban	Simulated 1993 urban	Accuracy
Actual 1993 non-urban	461 699	73 294	0.86
Actual 1993 urban	50 462	52 944	0.51
Total accuracy			0.81
Kappa coefficient			0.34

	1993~2002 (cells)		
	Simulated 2002 non-urban	Simulated 2002 urban	Accuracy
Actual 2002 non-urban	416 056	82 214	0.84
Actual 2002 urban	76 651	63 478	0.45
Total accuracy			0.75
Kappa coefficient			0.28

The rule-based CA also has a poorer performance according to the aggregated indicator. The Moran I values are 0.376 and 0.489 for the actual and simulated patterns in 1993, respectively. They become 0.385 and 0.526 for the actual and simulated patterns in 2002. The differences are larger than those of the proposed model.

There are many successful examples of using standard rule-based CA for simulating complex urban systems (White *et al.* 1997, Clarke and Gaydos 1998). Satisfactory results in terms of simulation accuracies have been obtained for these studies. However, most of these studies mainly focus on the simulation of a single city or a portion of a city using 'fixed' rules. For example, White *et al.* (1997) use a rule-based CA to simulate the central area of a small city, Cincinnati in Ohio. The kappa coefficient is 0.69. Wu (2002) has applied the logistic-CA to simulate the dynamics of urbanized and non-urbanized land use in the Tianhe District, a suburban district of Guangzhou. The simulation has an overall accuracy of 0.72~0.79. However, it has an area of only 304 km^2, which is much smaller than that of our study area. Ward *et al.* (2000) use a constrained rule-based CA to simulate the residential development of a single satellite city in the Gold Coast of eastern Australia. The model performs reasonably well with agreement between the simulated and actual growth of 63%. Silva and Clarke (2002) also utilize the SLEUTH urban growth model to simulate two metropolitan areas separately, the Lisbon Metropolitan Area (312 km^2) and the Porto Metropolitan Area (817 km^2). In both case studies, the model accurately reflects the evolution of urbanization in both metropolitan areas (a score of 0.90 for Lisbon and 0.97 for Porto).

Our experiments indicate that the case-based CA has much better simulation performances than the rule-based CA for this complex region. The poor performance of using rule-based CA is due to the complexities exhibited in this large region. It has an area of about 41 157 km^2, consisting of a hierarchy of fast-growing cities, such as Shenzhen, Dongguan, Guangzhou, and Zhengchen. It is inappropriate to use the same set of rules to simulate these cities simultaneously, as they have quite different growth patterns and processes (Li and Yeh 2004b). Moreover, many chaotic patterns have been witnessed in this fast-growing region (Yeh and Li 2001b). It is difficult for standard CA to obtain good simulation results under these complex situations. The case-based reasoning is more suitable for capturing unexplained variables and revealing complex relationships in this large area.

Another issue related to the simulation accuracy is the error propagation for source data. There are classification errors when remote-sensing data are used to obtain the empirical information about land-use changes. Additionally, a large amount of GIS data are usually required as the main inputs to these urban models. There are many possibilities of creating errors in spatial data because these errors can come from source maps and even map digitizing. New errors can also be generated in GIS operations. All these errors will propagate through CA simulation and have an impact on the results. Most of these errors are difficult to assess because of unavailable source information. Moreover, error propagation is a fairly complex process for these nonlinear dynamic models.

Studies indicate that errors inherited from the source data can propagate through CA simulation, but the errors are much reduced in the simulation because of the averaging effects of neighbourhood functions and the use of iterations in CA (Yeh and Li 2006). It is also found that there are spatial patterns for the uncertainties related to data errors. In this study, most of the simulated patterns have good conformity with the actual patterns (figures 6–10). Experiments indicate that the areas of disagreement are mainly situated in the peripheries of urban areas (Yeh and Li 2006). This fact can also be observed in the above figures. Many more studies are still required to address the error propagation in urban simulation when a variety of spatial data are involved.

Recently, Pontius and Malanson (2005) suggested that a predictive model should be compared with a null model of pure persistence (no change) for model validation. The baseline is that a predictive model should have better performances than a null model. For example, the typical amount of change on the landscape was about 10% in many situations, so a null model of pure persistence would be 90% correct based on the standard overall accuracy. However, this overall accuracy has a bias because of the difference between the actual agreement and chance agreement (Congalton 1991). This accuracy is extremely popular and frequently misinterpreted (Pontius and Malanson 2005). The Kappa coefficient can be used to avoid this problem. Table 3 lists the accuracies of the null model of pure persistence for the Pearl River Delta. The total accuracies have problems related to distinction of the performances of these models. However, the Kappa coefficient indicates that the proposed CBR-CA has a much better performance.

Figure 11 further shows the spatial distribution of agreement and disagreement of the simulated patterns of urban development in the major part of Guangzhou in 1993 and 2002. The first three classes are predicted correctly, whereas the last three classes are errors, by the proposed simulation model. The null model of pure persistence (no change) would correctly predict the first two classes and the green class, but would wrongly predict the others.

4. Conclusion

The essential part of geographical cellular automata (CA) is to provide appropriate transition rules so that realistic patterns can be simulated. Transition rules can be

Table 3. Accuracies of the null model of pure persistence for the Pearl River Delta.

	1988~1993 (cells)		
	1988 non-urban (no change)	1988 urban (no change)	Accuracy
Actual 1993 non-urban	522 010	12 984	0.98
Actual 1993 urban	72 424	30 982	0.30
Total accuracy			0.87
Kappa coefficient			0.36

	1993~2002 (cells)		
	1988 non-urban (no change)	1988 urban (no change)	Accuracy
Actual 2002 non-urban	484 379	14 152	0.97
Actual 2002 urban	109 448	30 451	0.22
Total accuracy			0.81
Kappa coefficient			0.25

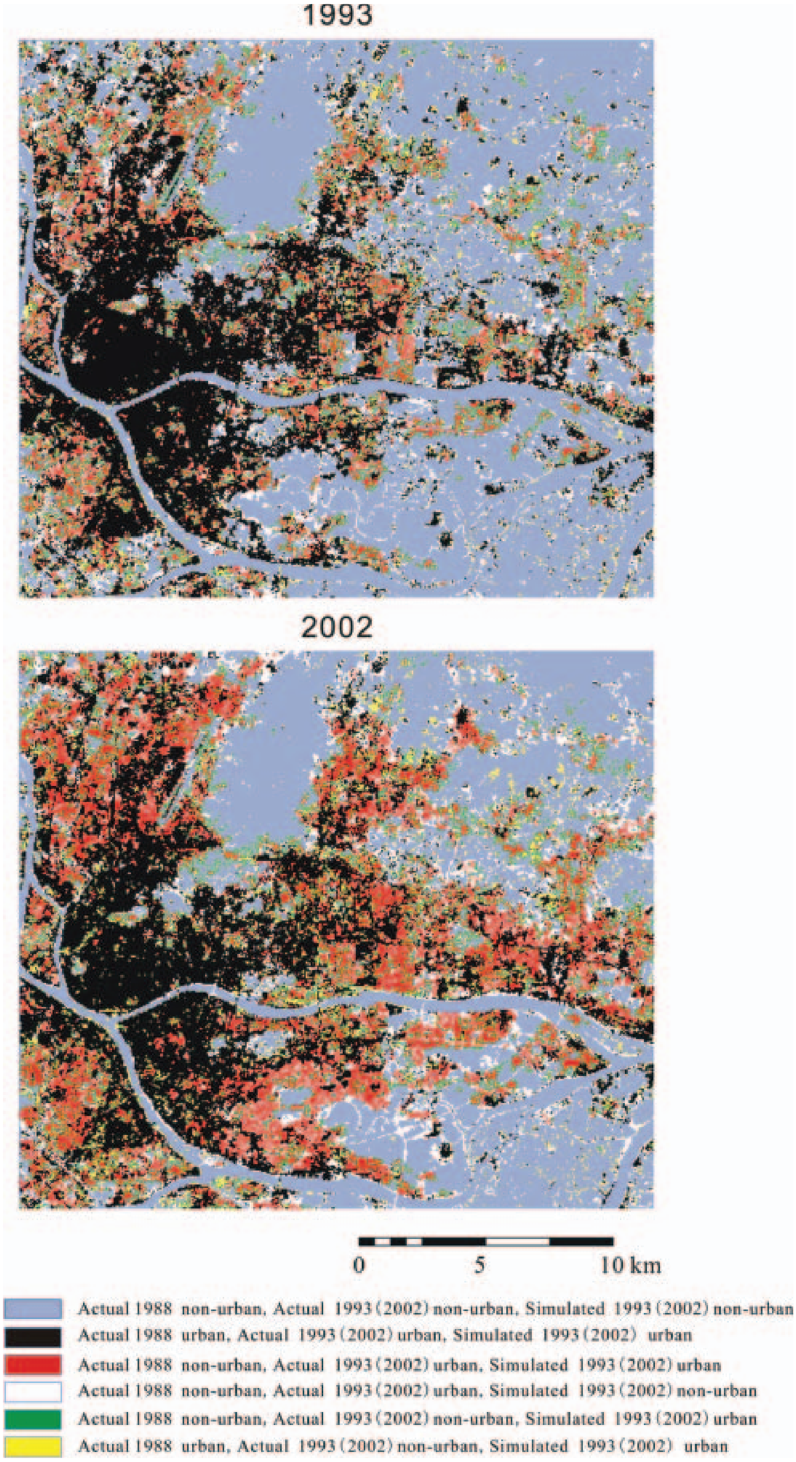

Figure 11. Distribution of agreement and disagreement of the simulated patterns or urban development in the major part of Guangzhou in 1993 and 2002.

defined by a variety of methods, such as multicriteria evaluation (MCE), logistic regression, neural networks, and data mining. CA usually assume that transition rules are constant in the spatio-temporal dimension. However, the uniform transition rules may not be suitable for simulating complex geographical phenomena (e.g. the evolution of urban systems) at a large regional scale. The validity of CA is affected by some unexplained variables because of nature's complexities. Static transition rules cannot capture the influences of these unexplained variables and thus reflect complex relationships. Discrete cases are more adapted to representing dynamic transition rules and are more suitable for the simulation of a large area and a long period.

The solicitation of concrete knowledge (transition rules) is often difficult for many applications. There are problems in representing complex relationships by using detailed rules. This study demonstrates that the case-based approach can avoid the problems of the rule-based approach in defining CA. The proposed method is based on the case-based reasoning techniques, which do not require the procedure of soliciting explicit transition rules. The knowledge for determining the state conversion of CA is inexplicitly embedded in discrete cases. The lazy-learning technology can be used to represent complex relationships more effectively than detailed equations or explicit transition rules.

In this study, the common k-NN algorithm has been modified to incorporate the location factor in the reasoning process. Spatial-distance weights are defined to reflect spatially dynamic transition rules. The dynamic characteristics in the temporal dimension are also obtained by updating the case library with remote-sensing data. The use of CBR can help CA more adapted to the simulation of complex cities with possible changes to simulation environment, such as population, economy, land-use policy and resources

The proposed model has been applied to the simulation of urban development in the whole Pearl River Delta in 1988–2002. This large area consists of a hierarchy of cities, among which there are significant variations in population, economic growth, urbanization, and land-use patterns. However, the case-based CA can produce very plausible simulation results for this large region. The consistency between the actual and simulated patterns is verified by using classified remote-sensing data. Satisfactory results can be achieved not only at a regional level but also at a city level.

The model validation is carried out by using the indicators of cell-by-cell overlay and Moran I. A standard rule-based CA is also used as a comparison. The experiment indicates that the case-based CA has better simulation performances than the rule-based model for this large complex region.

Mining the trajectory of future from the past is still a difficult problem for any forecasting models. Some studies have been carried out to address this issue by using the methods of self-modification (Clarke *et al.* 1997) and data-mining techniques (Li and Yeh 2004a). However, CBR has been well recognized for its adaptation potential through the learning from case libraries (Kolodner 1993, Wilke and Bergmann 1998). This learning capability is useful for tackling the complexities in urban simulation.

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 '985 Project' of GIS and Remote Sensing for Geosciences from the Ministry of Education of China (Grant No. 105203200400006).

References

BATTY, M. and XIE, Y., 1994, From cells to cities. *Environment and Planning B: Planning and Design*, **21**, pp. 531–548.

BINDER, P., 1989, Evidence of Lagrangian tails in a lattice gas. In *Cellular Automata and Modeling of Complex Physical Systems*, P. Manneville, N. Boccara, G.Y. Vichniac and R. Bidaux (Eds), pp. 155–160 (Berlin: Springer).

BRANTING, K.L. and HASTINGS, J.D., 1994, An Empirical Evaluation of Model-based Cased Matching and Adapation. In *American Association for Artificial Intelligence, Case-based Reasoning Workshop*, Seattle, WA, pp. 72–78.

CLARKE, K.C. and GAYDOS, L.J., 1998, Loose-coupling a cellular automata model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, **12**, pp. 699–714.

CLARKE, K.C., BRASS, J.A. and RIGGAN, P.J., 1994, A cellular automata model of wildfire propagation and extinction. *Photogrammetric Engineering & Remote Sensing*, **60**, pp. 1355–1367.

CLARKE, K.C., HOPPEN, S. and GAYDOS, L., 1997, A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, **24**, pp. 247–261.

CONGALTON, R.G., 1991, A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, **37**, pp. 35–46.

COUCLELIS, H., 1985, Cellular worlds: a framework for modeling micro-macro dynamics. *Environment and Planning A*, **17**, pp. 585–596.

COUCLELIS, H., 1988, Of mice and men: what rodent populations can teach us about complex spatial dynamics. *Environment and Planning A*, **20**, pp. 99–109.

DASARATHY, B.V., 1991, *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*, p. 447 (Los Alamitos, CA: IEEE Computer Society Press).

DEADMAN, P.D., BROWN, R.D. and GIMBLETT, H.R., 1993, Modelling rural residential settlement patterns with cellular automata. *Journal of Environmental Management*, **37**, pp. 147–160.

FAVIER, C. and DUBOIS, M.A., 2004, Reconstructing forest savanna dynamics in Africa using a cellular automata model, FORSAT. *Lecture Notes in Computer Science*, **3305**, pp. 484–491.

GOLES, E., 1989, Cellular automata, dynamics and complexity. In *Cellular Automata and Modeling of Complex Physical Systems*, P. Manneville, N. Boccara, G.Y. Vichniac and R. Bidaux (Eds), pp. 10–20 (Berlin: Springer).

GOODCHILD, M.F., 1986, *Spatial Autocorrelation: Concepts and Techniques in Modern Geography*, p. 47 (Norwich, UK: Geo Books).

HOLT, A. and BENWELL, G.L., 1999, Applying case-based reasoning techniques in GIS. *International Journal of Geographical Information Science*, **13**, pp. 9–25.

HOUBEN, I., WEHENKEL, L. and PAVELLA, M., 1995, Coupling of K-NN with decision trees for power system transient stability assessment. In *Proceedings of the 4th IEEE Conference*, pp. 825–832.

JENERETTE, G.D. and WU, J.G., 2001, Analysis and simulation of land-use change in the central Arizona-Phoenix region. *Landscape Ecology*, **16**, pp. 611–626.

JIA, X.P. and RICHARDS, J.A., 2005, Fast k-NN classification using the cluster-space approach. *IEEE Geoscience and Remote Sensing Letters*, **2**, pp. 225–228.

KOLODNER, J., 1993, *Case-Based Reasoning* (San Mateo, CA: Morgan Kaufmann).

LEKKAS, G.P., AVOURIS, N.M. and VIRAS, L.G., 1994, Case-based reasoning in environmental monitoring applications. *Applied Artificial Intelligence An International Journal*, **8**, pp. 359–376.

LI, X. and YEH, A.G.O., 1998, Principal component analysis of stacked multi-temporal images for monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing*, **19**, pp. 1501–1518.

LI, X. and YEH, A.G.O., 2000, Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, **14**, pp. 131–152.

LI, X. and YEH, A.G.O., 2002, Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, **16**, pp. 323–343.

LI, X. and YEH, A.G.O., 2004a, Data mining of cellular automata's transition rules. *International Journal of Geographical Information Science*, **18**, pp. 723–744.

LI, X. and YEH, A.G.O., 2004b, Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. *Landscape and Urban Planning*, **69**, pp. 335–354.

LI, X. and YEH, A.G.O., 2004c, Multitemporal SAR images for monitoring cultivation systems using case-based reasoning. *Remote Sensing of Environment*, **90**, pp. 524–534.

MCAVOY, J.G. and KRAKOWSKI, E.M., 1989, A knowledge based system for the interpretation of SAR images of sea ice. *International Geoscience and Remote Sensing Symposium (IGARSS'89)*, **2**, pp. 844–847.

PONTIUS, G.R. and MALANSON, J., 2005, Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science*, **19**, pp. 243–265.

SILVA, E.A. and CLARKE, K.C., 2002, Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, **26**, pp. 525–552.

SOARES, B.S., CERQUEIRA, G.C. and PENNACHIN, C.L., 2002, DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, **154**, pp. 217–235.

STRAATMAN, B., WHITE, R. and ENGELEN, G., 2004, Towards an automatic calibration procedure for constrained cellular automata. *Computers, Environment and Urban Systems*, **28**, pp. 149–170.

THEIL, H., 1967, *Economics and Information Theory* (Amsterdam: North-Holland).

THOMAS, R.W., 1981, *Information Statistics in Geography* (Norwich, UK: Geo Abstracts, University of East Anglia).

WARD, D.P., MURRAY, A.T. and PHINN, S.R., 2000, A stochastically constrained cellular model of urban growth. *Computers, Environment and Urban Systems*, **24**, pp. 539–558.

WATSON, I., 1997, *Applying Case-Based Reasoning: Techniques for Enterprise Systems* (San Mateo, CA: Morgan Kaufmann).

WHITE, R. and ENGELEN, G., 1993, Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land use patterns. *Environment and Planning A*, **25**, pp. 1175–1199.

WHITE, R., ENGELEN, G. and UIJEE, I., 1997, The use of constrained cellular automata for high-resolution modelling of urban land use dynamics. *Environment and Planning B: Planning and Design*, **24**, pp. 323–343.

WILKE, W. and BERGMANN, R., 1998, Techniques and knowledge used for adaptation during case based problem solving. In *Lecture Notes in Artificial Intelligence*, 1416, pp. 497–505 (Berlin: Springer).

WOLFRAM, S., 1984, Cellular automata: a model of complexity. *Nature*, **31**, pp. 419–424.

WU, F., 2002, Calibration of stochastic cellular automata: the application to rural–urban land conversions. *International Journal of Geographical Information Science*, **16**, pp. 795–818.

WU, F. and WEBSTER, C.J., 1998, Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B*, **25**, pp. 103–126.

Wu, J.G. and David, J.L., 2002, A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling*, **153**, pp. 7–26.

Xu, X.Z., 2004, A note on the subjective and objective integrated approach to determine attribute weights. *European Journal of Operation Research*, **156**, pp. 530–532.

Yeh, A.G.O. and Li, X., 2001a, A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. *Environment and Planning B: Planning and Design*, **28**, pp. 733–753.

Yeh, A.G.O. and Li, X., 2001b, Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogrammetric Engineering & Remote Sensing*, **67**, pp. 83–90.

Yeh, A.G.O. and Li, X., 2006, Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems*, **30**, pp. 10–28.

Yeh, A.G.O. and Shi, X., 1999, Applying case-based reasoning to urban planning: a new planning-support system tool. *Environment and Planning B*, **26**, pp. 101–115.