

## Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm

J. TANG, L. WANG\* and Z. YAO

Texas Center for Geographic Information Science, Department of Geography, Texas State University, San Marcos, TX 78666, USA

(Received 4 February 2005; in final form 10 July 2005)

The landscape pattern of Daqing City, China, has undergone a significant change over the past 20 years, as a result of the rapid urbanization process. To understand how urbanization has influenced the landscape in Daqing City, the largest base of the petrochemical industry in China, we conducted a series of spatial analyses with landscape pattern maps obtained from Landsat images in 1979, 1990 and 2000. Results indicate that a substantial urban area has been extended during the past two decades, along with the shrinking of wetland and woodland.

Spatio-temporal optimization is not a trivial task in developing landscape models. In previous studies, the optimization of spatial and temporal factors was achieved separately, because of the difficulty in formulating them together in a single model. In this study, we adapted the traditional Markov model by obtaining model parameters and neighbourhood rules from a modified genetic algorithm (GA). Model performance was evaluated between the empirical landscape map from the Landsat image and the simulated landscape map from the models. Over three simulation runs, the global deviation (GD) for the three models was 1.37, 1.10 and 1.15, respectively. This result shows that the Markov model and the GA together are able to effectively capture the spatio-temporal trend in the landscape pattern associated with urbanization for this region. The future landscape distribution in 2010, 2030 and 2050 was derived using a spatial Markov model (SMM) for further urban change and planning research.

### 1. Introduction

Increasing awareness concerning the importance of sustainable urban development is stimulating the improvement of current methods to better understand urban landscape evolution, which is the result of complex interactions between physical, biological and social forces in time and space (Turner 1987). A fundamental problem confronted by researchers is the difficulty in finding an effective model that can incorporate both, spatial and temporal knowledge into predicting future patterns. Such a spatial dynamics model is crucial for the analysis, understanding, representation and modelling of city dynamics.

Remote sensing data, in conjunction with geographic information systems (GIS), have been recognized as an effective tool in quantitatively measuring urban area and modelling urban growth at a relatively large spatial scale (Yeh and Li 1997, Weng 2001, Herold *et al.* 2003, Yagoub 2004). Given the advantage of repeatedly

---

\*Corresponding author. Email: lewang@txstate.edu

measuring a large spatial area, satellite remote sensing has been effectively applied to better understand and monitor landscape development and processes, as well as estimate biophysical characteristics of land surfaces (Roy and Tomar 2001, Stow and Chen 2002). GIS technology provides a seamless environment for integrating, visualizing and analysing digital data to facilitate change detection and database development (Abed and Kaysi 2003, Stewart *et al.* 2004). Significant progress in the acquisition of remotely sensed data in a finer spatio-temporal resolution, compounded with the development of geographic and environment process models, has greatly extended research capability to examine the chronology, causes and impacts of the urbanization process.

Motivation to model urban landscape dynamics arises from the process of examining what, where and to what extents landscape change has occurred, and furthermore, to understand how and why the changes can occur (Weng 2002, Yang and Lo 2002). Basically, landscape process models can be categorized into two groups: statistical description models (Baker 1989, Griffith *et al.* 2003, Herold *et al.* 2003) and spatial transition models (Turner 1987, Muller and Middleton 1994, Weng 2002). The spatial transition models make better use of spatial information by including the location or state configuration of the landscape. In terms of their application, the majority of the statistical description models reveal dynamics of landscape structure and pattern through a set of statistic variables, for example landscape configuration (Baker 1989) or spatial metrics (Griffith *et al.* 2003, Herold *et al.* 2003). Early Markovian analysis is used as a descriptive tool to predict land use change on a local or regional scale (Bell 1974, Bourne 1976). Spatial transition models, however, attempt to derive transition probabilities from a series of temporal landscape maps and apply them to the landscape prediction model, such as the Markov chain model (Muller and Middleton 1994, Weng 2002) or the cellular automata model (Jenerette and Wu 2001). A spatially explicit result can be expected from the spatial transition models, although they may be much more difficult to develop than the descriptive models.

To model urban land use change at the regional level, the spatial dependency has to be considered at an enhanced level of detail with respect to the complex spatial assemblage of land use types in the urban area (Turner 1987, Barnsley and Barr 1996, Clarke and Hoppen 1997, Ridd and Liu 1998). Turner (1987) incorporated the spatial neighbourhood relationship in developing the transition models of landscape changes in Georgia. Clarke and Hoppen (1997) developed a self-modifying cellular automaton model, based on the circumstances in neighbouring cells for studying the historical urbanization in the San Francisco Bay area. Pontius and Malanson (2005) quantified spatial contiguity for the cellular automata Markov model to predict land change in central Massachusetts. Nevertheless, there is still a lack of methods that can streamline the process from acquisition of the spatial information from remote sensing data all the way to application of such information in the spatial transition models.

Stochastic Markov models (Muller and Middleton 1994) provide one way of simulating and exploring the process of dynamic systems. Adopted from stochastic mathematical statistics, the Markov chain model has been applied to model the changes in land use/cover at a variety of spatial scales (Bell 1974, Muller and Middleton 1994, Weng 2002) with the use of remote sensing or GIS data. One obstacle to building a robust spatial transition model using Markovian analysis lies in the difficulty in concurrently achieving the spatial and temporal optimization of a given model (Turner 1987, Shibasaki and Huang 2001). Interpolation of model

parameters and neighbourhood rules using a genetic algorithm (GA) has been previously tested in a few studies (Jenerette and Wu 2001, Shibasaki and Huang 2001). Although these studies demonstrate that incorporation of spatial parameters will offer improved description and representation of landscape structure and process, we need a standardized method to properly apply the GA in estimating transition probability from a set of satellite images.

In this study, we integrate remote sensing images and GIS into landscape changes analysis for the urban region of Daqing City between 1979 and 2000 and predict its pattern in the future. To best predict the future urban landscape change, three different Markov process models were developed, based on different decision rules for deriving the Markov transition probability. The first model derives its transition probability solely from the changed area, without giving consideration to the spatial information around it. Here, we have named it the Markov model (MM). The other two models do consider the spatial neighbourhood information when coming up with the transition probability, but with different strategies: one takes account of the four immediate neighbours for all the pixels, while the other one applies the four-nearest-neighbour consideration only to the pixels under boundary condition. These are called the spatial Markov model (SMM) and the boundary Markov model (BMM), respectively. A modified GA was developed to combine the spatial information into the formation of the transition matrix. Finally, model performance was validated by comparing the empirical landscape map classified from a 2001 Landsat image and the simulated landscape map from the model in 2001.

## **2. Study site and data preparation**

### **2.1 Study site**

The study was carried out in the central part of Daqing City. As the energy capital of China, Daqing City maintains a variety of landscape types due to its unique geology and climate. Centred at 124°15' E longitude and 46°20' N latitude, the study area covers four major urban areas, Shaertu district, Ranghulu district, Longfeng district, and Honggang district, in Daqing City (figure 1). The terrain consists of a relatively flat plain with a mean elevation of 126–165 m and an elevation difference of 10–39 m.

Daqing City, once a rural area, has become the largest oil production base since the oil was explored in 1959. Although Daqing City is now diversifying its energy-oriented economy, the petroleum and petrochemical industries still comprise the backbone of its economy. The continual construction of oil fields has spoiled the original landscape pattern over the past 50 years. Reduction of swamp, grassland and forest resulted in deterioration and desertification of the environment, potentially affecting the future landscape pattern, regional environment and climate. Moreover, with rapid economic development, the population in the Daqing region has grown dramatically over the past 50 years, increased from one hundred thousand in 1945 to two and a half million in 2000 (Statistics Bureau of Daqing 2001). As a result, our study area, which contains mostly the urban area in the Daqing region, is subject to rapid changes in the urban landscape pattern, such as the addition of human constructions and the loss of natural swamps.

### **2.2 Data preparation**

Four 1500 × 1500-pixel Landsat Thematic Mapper (TM) and Multispectral Scanner (MSS) satellite images were chosen in this study for the change detection and model

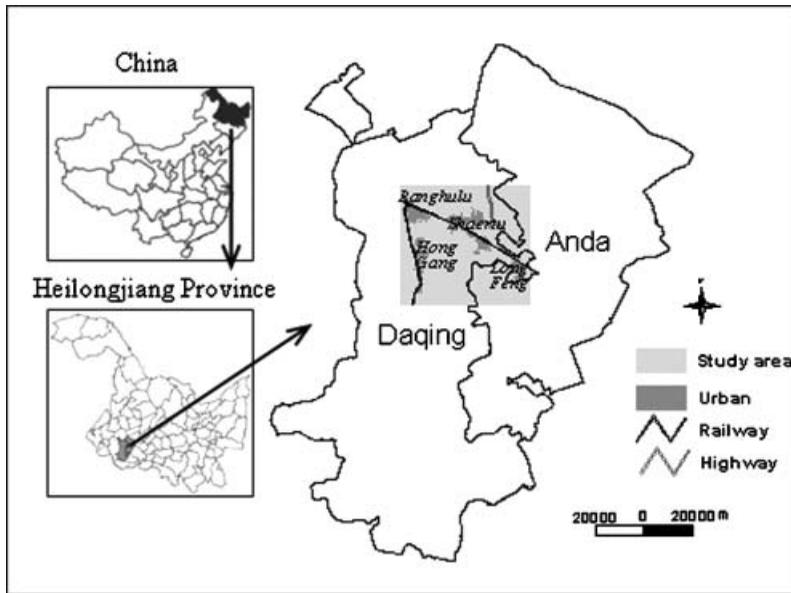


Figure 1. Study area, urban area of Daqing City, Heilongjiang Province, China.

construction of urban districts in Daqing during a span of 20 years from 1979 to 2000. One scene of a Landsat Enhanced Thematic Mapper (ETM) image acquired in 2001 was used as empirical data for the model validation. All the images were acquired between late June to September, which falls in the growing season of vegetation. One scene of Landsat MSS was acquired in 23 August 1979 and three scenes of Landsat TMs were obtained on 20 July 1990, 22 June 2000 and 11 August 2001. All images were radiometrically and geometrically corrected on a SUN workstation using ERDAS<sup>TM</sup> software. The MSS was resampled to 30 m spatial resolution, as was the TM image. The conventional supervised classification, Maximum Likelihood Classification (MLC), was conducted to obtain four classification maps in our study area (table 1). Seven classes were mapped: agriculture, urban or built-up, grass, saline or barren land, water, wetland, and woodland. Figure 2 presents the original images and sequential maps classified from the images.

Considering the requirement of traditional MLC and the size of our study area, we chose a set of training samples of 300 pixels for the imagery in each year. The accuracy of the resultant landscape maps was assessed with an independent set of test samples on the study area. Initially, an error matrix was generated. The producer accuracy, user accuracy and overall accuracy, as well as the kappa coefficient, were derived and are reported in table 2. Given the nature of the broad land-use classes defined, the classification results were thus able to provide input data for the spatio-temporal model.

### 3. Method

#### 3.1 Markov chain models in Landscape changes

Markov chain models have been used to model landscape changes in understanding and predicting the behaviour of complex systems (Baker 1989, Weng 2002, Fortin *et al.* 2003) using discrete state spaces. All landscape spatial transition models can be

Table 1. Accuracy assessment of 1979, 1990, 2000 and 2001 landscape maps from Landsat images by Maximum Likelihood Classification.

	Training sample				Test sample				User accuracy (%)				Producer accuracy (%)			
	1979	1990	2000	2001	1979	1990	2000	2001	1979	1990	2000	2001	1979	1990	2000	2001
Agriculture	324	332	335	334	316	324	329	312	80.66	76.38	66.94	75.44	64.92	83.97	91.04	75.44
Urban or built-up	313	324	322	314	312	308	313	311	94.08	98.78	87.92	96.97	92.64	78.90	92.09	96.97
Grass	315	314	307	317	300	315	314	306	94.86	80.86	76.69	76.68	89.64	83.17	61.78	76.68
Saline or barren land	317	305	302	321	300	309	322	313	89.40	79.73	83.24	75.54	99.99	95.41	92.60	75.54
Water	312	305	306	312	317	303	333	331	96.98	99.44	99.53	98.78	96.98	99.15	99.05	98.78
Wetland	311	310	312	301	314	310	309	303	80.73	98.18	80.31	95.54	96.27	84.33	97.84	95.54
Woodland	303	327	308	312	303	320	308	307	69.58	81.56	96.77	71.23	84.03	83.28	67.80	71.23
Overall accuracy (%): 86.23 (1979); 87.05 (1990); 83.86 (2000); 83.16 (2001)																
Kappa (%): 83.93 (1979); 84.89 (1990); 81.12 (2000); 80.35 (2001)																

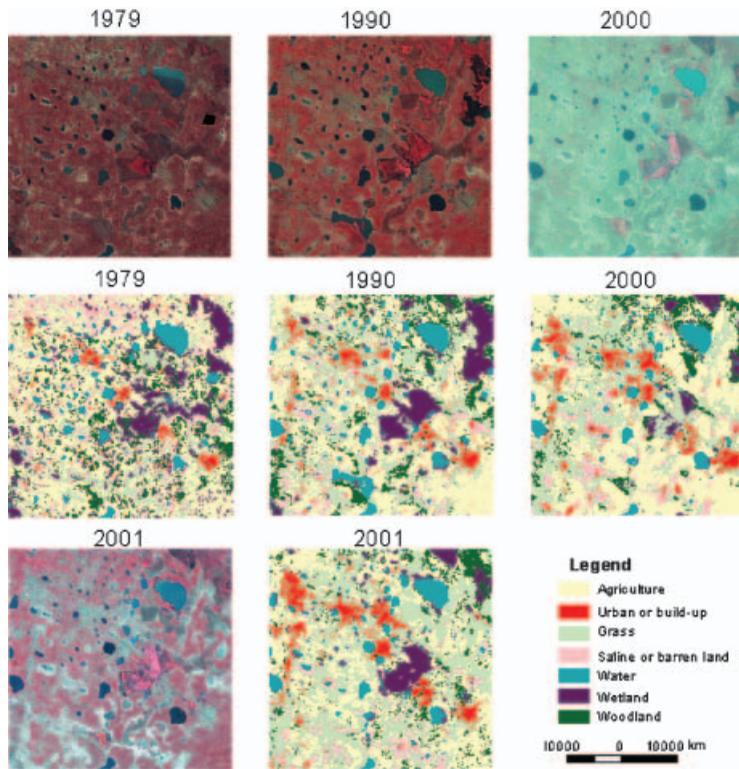


Figure 2. The original images and land use maps of the Daqing region from 1979 to 2000. Note: the classified map of 2001 was used as an empirical map to validate the model result.

expressed in a simple matrix equation as follows:

$$N_{t+1} = N_t \times P \quad (1)$$

where  $N_{t+1}$  and  $N_t$  are vectors composed of the fractions of each landscape type at time  $t+1$  and time  $t$ , respectively;  $P$  is a square matrix, whose cell  $P_{ij}$  is the transition probability from landscape  $i$  to  $j$  during times  $t$  and  $t+1$ .

The transition probabilities  $P$  are derived from the landscape transitions occurring during some time interval. In this study, we chose maximum likelihood

Table 2. Transition matrix of land use area ( $\text{km}^2$ ) between 1979 and 1990.

	Agriculture	Urban or built-up	Grass	Saline or barren land	Water	Wetland	Woodland
Agriculture	315.84	19.78	240.26	32.31	9.95	12.36	54.33
Urban or built-up	5.23	29.23	12.98	2.28	1.09	0.59	0.80
Grass	126.95	16.38	236.75	49.23	20.15	31.02	64.08
Saline or barren land	31.40	10.69	70.20	41.18	12.43	4.57	11.91
Water	1.55	2.54	10.64	8.92	79.07	13.66	1.81
Wetland	40.33	4.08	41.52	6.44	17.31	90.84	29.66
Woodland	32.92	3.99	37.86	4.52	2.62	12.83	109.84

(Anderson and Goodman 1957) to estimate the transition probabilities in  $k$  steps, as:

$$P_m(i, j) = N(i, j) / \sum_{i=1}^n n_{ij} \quad (2)$$

$$P(i, j) = \sum_{k=1}^k \frac{P_m(i, j)}{kn_m}$$

Where  $N(i, j)$  are the observed data during the transition from state  $i$  to  $j$ , and  $n_{ij}$  is the number of years between time step  $i$  and step  $j$ , and the total number of years is  $m$ ;  $P(i, j)$  is the yearly transition probability after normalizing the transition probability in multiyears.

There are several assumptions (Baker 1989, Stewart 1994, Weng 2002) in the first-order homogeneous Markov chains. Basically, we assume that the landscape change is stochastic, as opposed to deterministic; and the landscape distribution at a given time is the independent state of the Markov chain. Thus, in the Markov chain model, the next landscape distribution  $N_{t+1}$ , at time  $t+1$ , only depends on the current distribution  $N_t$  at time  $t$  without considering the other historical values,  $N_1, N_2, N_3, \dots$ . Moreover, as the cells in the transition matrix are probabilities, it follows that:

$$\sum_{j=1}^m P(i, j) = 1 \quad (3)$$

In this first-order homogeneous Markov chains model, the behaviour of any selection time can be completely predicted if we know the transition probability matrix [ $P_{ij}$ ] and the initial distribution vector  $N$  (Chakraborty *et al.* 1995). The  $s$ -step transition probabilities  $P_{ij}^s$  are obtained as the elements of  $P^s$ .

Theoretically, the Markov chain model assumes that the transition probability is spatially independent (Brown *et al.* 2000). However, the future trend of a pixel to change is not a simple function of its current state, but is often affected by its neighbouring cells. As a result, a large amount of spatial information is ignored in the stand-alone Markov chain model. Therefore, additional steps are needed to incorporate both spatial and temporal information.

### 3.2 The genetic algorithm (GA) and its simplification in estimating Markov transition probability

The original GA was designed to search and optimize solutions based on natural selection and natural genetics (Hood 1975, Goldberg 1989). In general, GA operates on a set of coded individuals and each one receives a fitness value using the coding of their genes (Mertens *et al.* 2003) to produce the new population through a set of genetic operators: reproduction, crossover and mutation. Reproduction is a process, in which individual values are copied to the new population, according to their objective function values or the fitness values, crossover mates and swaps the individual within the neighbours, while mutation alters the value in an individual position. The individuals with a higher fitness are more likely to be selected over others in the evolution process and the new population is most likely to have a higher average fitness than the old one (Bornholdt 1998).

In this study, we defined a simple version of this algorithm. The GA was modelled as a stochastic system with an optimization scheme. The algorithm starts from the initial landscape pattern in our study area. The code of their genes is the landscape

class assigned to each pixel. In each step, the landscape class will be transformed according to probabilities defined on the basis of their fitness values, which create a new generation. That is, the generations with a higher fitness are more likely to 'survive' than those with lower fitness. Therefore, the new population comes with a higher average fitness than the old one (Bornholdt 1998). The fitness function  $f(s)$  to be optimized depends on three control parameters, reproduction probability (denoted as  $P_r$ ), crossover probability (denoted as  $P_c$ ) and mutation probability (denoted as  $P_m$ ). To incorporate this algorithm into the Markov chain, we combined  $P_r$  and  $P_c$  into calculating a spatial probability  $P_{rc}$ . Let  $x$  be an element from an  $X$  population within the class space  $S = \{s_1, s_2, s_3, \dots, s_N\}$ , so that the initial probability distribution is given as

$$D_i^{(0)} = \sum^X P_r \{x^{(0)} = s_i\}, \quad i = 1, 2, 3, \dots, N$$

$$\sum_{i=1}^N D_i^{(0)} = 1$$
(4)

The class value of the target pixel at centre  $x_c$  (0) is compared to that of its four neighbours  $x_d$  (1, 2, 3, 4), respectively, to produce the spatial probabilities for the whole map (equation (5)). Through this equation, we can build a square matrix to denote the spatial relationship between the two land use classes.

$$\begin{aligned} &\text{While } (x \in X); \\ & \quad i = s(x); \quad j = s(x); \quad d = 1, 2, 3, 4; \\ & \quad \text{Matrix } (i, j) + 1; \\ & \text{End} \end{aligned}$$
(5)

As the ecotone between two classes usually demonstrates a greater spatial heterogeneity and exhibits a greater extent of landscape changes, the class located in the ecotone usually has a higher spatial influence on its neighbours. Applying the same spatial transition probability for the entire landscape will ignore this difference between the pixels inside the landscape patch and the pixels under the boundary condition (figure 3). To optimize the spatial parameter set in this landscape process model, we developed a contrasting algorithm using the four adjacent neighbours only if they are lying on the boundary. This is called the boundary Markov predict. The comparison between the spatial Markov predict result and the boundary Markov predict result could provide an additional method for considering spatial change in the spatial model.

The spatial probability  $P_{rc}(i, j)$  is then defined as

$$P_{rc}(i, j) = \frac{P_r \{X^{(0)} = s_i\}}{P_r \{X^{(0)} = s_j\}}$$

$$= \frac{\text{Matrix}(i, j)}{\sum_j \text{Matrix}(i, j)}$$
(6)

where  $P_{rc}(i, j)$  is the conditional probability calculated from the probability  $P_r \{X^{(0)} = s_j\}$  and  $P_r \{X^{(0)} = s_i\}$  when class  $i$  was found next to class  $j$  randomly. The matrix  $(i, j)$  will be acquired from equation (5).

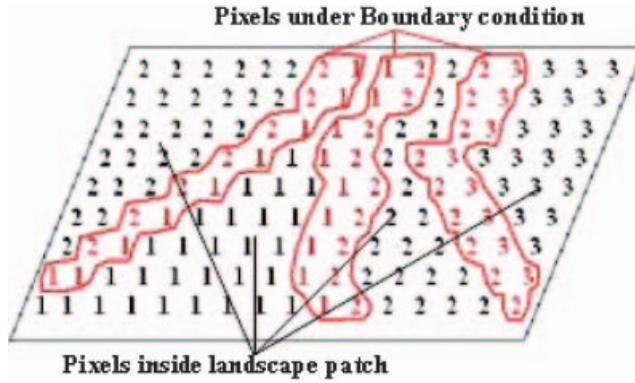


Figure 3. Pixels under boundary condition: codes denote land use class.

The mutation probability  $P_m(i,j)$  is the probability of class  $j$  being replaced by class  $i$  between two successive states,  $t$  and  $t+1$ , whose  $(i,j)$  element is:

$$P_{m(i,j)}^{(t,t+1)} = P\{x^{(t+1)} = s_j | x^{(t)} = s_i\} \tag{7}$$

As shown in figure 4, we consider the  $P_{rc}$  as spatial continuity and  $P_m$  as temporal continuity. For example, if one agriculture pixel is surrounded by several grassland pixels, it will have a greater potential to change into grassland. In other words, the agriculture pixel has a higher crossover probability from grassland to agriculture. If the agriculture pixel is situated among other agriculture pixels, it will have a higher reproduction probability to itself. The mutation probability only depends on the exchange among the classes, which can be derived from multitemporal data. Obviously, the spatial probability  $P_{rc}(i,j)$  and temporal probability  $P_m(i,j)$  are independent. Then the transition probability  $P(i,j)$  is calculated in the form of the

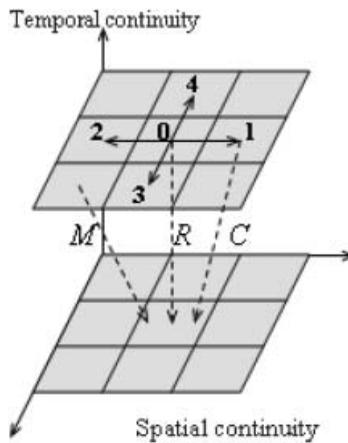


Figure 4. Spatio-temporal relationship of variable data. M, mutation; R, reproduction; C, crossover.

fitness value, which is further calculated as a multinomial distribution during  $n$  processes:

$$\begin{aligned}
 P(i, j) &= \prod_1^k \left\{ \frac{\prod_1^{k_1} P_T(i, j) P_S(i, j)}{2k_1} + \frac{\prod_1^{k_2} P_T(i, j) P_S(i, j)}{2k_2} + \dots + \frac{\prod_1^{k_n} P_T(i, j) P_S(i, j)}{2k_n} \right\} \\
 &= \prod_1^k \left\{ \frac{\prod_1^{k_1} P_m(i, j) P_{rc}(i, j)}{2k_1} + \frac{\prod_1^{k_2} P_m(i, j) P_{rc}(i, j)}{2k_2} + \dots + \frac{\prod_1^{k_n} P_m(i, j) P_{rc}(i, j)}{2k_n} \right\} \quad (8)
 \end{aligned}$$

where  $k_i$  is the temporal step number between the state, and  $k = k_1 + k_2 + \dots + k_n$ .  $P_T$  is the probability derived from multitemporal data, which corresponds to the mutation probability  $P_m$ ;  $P_S$  is the spatial probability, which corresponds to the reproduction and crossover probability  $P_{rc}$ .

### 3.3 Model validation

Validation of a landscape dynamics model is usually carried out by comparing the predicted result to the empirical map, pixel by pixel, to determine the prediction ability of the model. With regard to landscape processes, the absolute location of landscape elements is likely to be less important than the overall pattern in the landscape (Jenerette and Wu 2001). As the algorithm used in this study to simulate the spatial dynamics of landscape elements is patch-based instead of pixel-based, the per-pixel comparison is not suitable for this study.

For validation, the model's simulated output was compared to the empirical map from the same year (Pontius *et al.* 2001). The validation process runs across two steps. First, the land-use maps classified from 1979, 1990 and 2000 remote sensing images were used to build a prediction model, from which a simulated land-use map in a different year was derived. Then the simulated map in 2001 was compared to the empirical map from the same year (figure 2) to validate the model result. Two indexes, the class-specific individual deviation (ID) and the overall global deviation (GD), were adopted to assess the discrepancy between the simulated and empirical map (Pontius *et al.* 2001):

$$\begin{aligned}
 ID &= \frac{P_o - P_e}{P_e} \\
 GD &= \sum_{i=1}^n ID
 \end{aligned} \quad (9)$$

where  $P_o$  is the percentage of each class of the model's simulated output,  $P_e$  is the percentage of each class of empirical map, and  $n$  is the number of classes. A 'successful' simulation occurs when the model's simulated output best matches the empirical land-use map with the lowest ID and GD.

## 4. Results and discussion

### 4.1 Pattern change

A tremendous change in the urbanized area within Daqing City is evident over the past 20 years. The total change areas between 1979 and 1990 and between 1990 and 2000 are 1191.06 km<sup>2</sup> and 887.42 km<sup>2</sup> and the percentage changes are 59.05% and 44.00%, respectively. As indicated in figure 5, the most significant changes are the

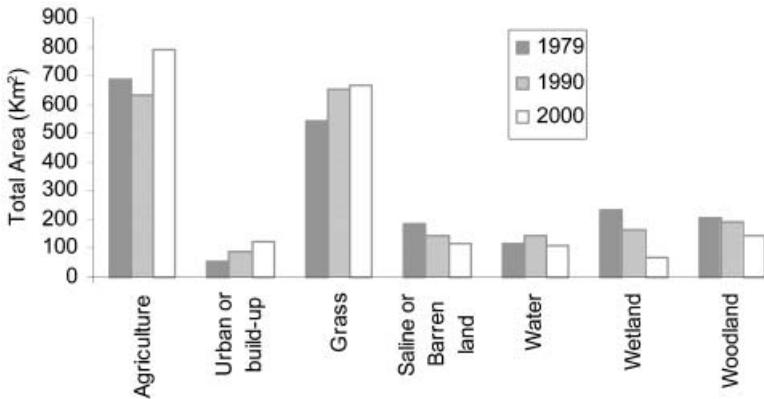


Figure 5. The total area of each land use class between August 1979 and June 2000.

spread of urban or built-up landscape, which increased from 2.59% in 1979 to 6.23% in 2000 and the loss of wetland from 11.41% in 1979 to 3.51% in 2000. The dominant landscape, agriculture, decreased from 33.95% in 1979 to 31.29% in 1990 and increased back to 39.13% in 2000.

The number of patches in each landscape class was calculated through the STATISTIC function in ARC/INFO. The pattern of changes in the number of patches is class-specific (figure 6). Over the 20 years studied, the number of urban patches increased initially and then decreased. This is primarily due to the ‘nibble’ of other land use by human disturbance. With the increased urbanization, smaller urban patches merged into a larger and continuous patch when human disturbances increased in this region. Figure 2 shows that the spatial expansion of urban area is along the major transportation routes, such as the Bingzhou and Rangtong railways. Wetland and woodland have decreased in both, the patch area and number of patches during the 20-year period, because of its conversion to grass or cultivated lands. Shrinkage of small wetlands and fragmentation at the edge of the large ones are mostly found in the northeast of the study area. On the contrary, the total area kept decreasing during 1979–1990 and increasing during 1990–2000, while the number of patches in agriculture kept increasing during 1979–1990 and decreasing

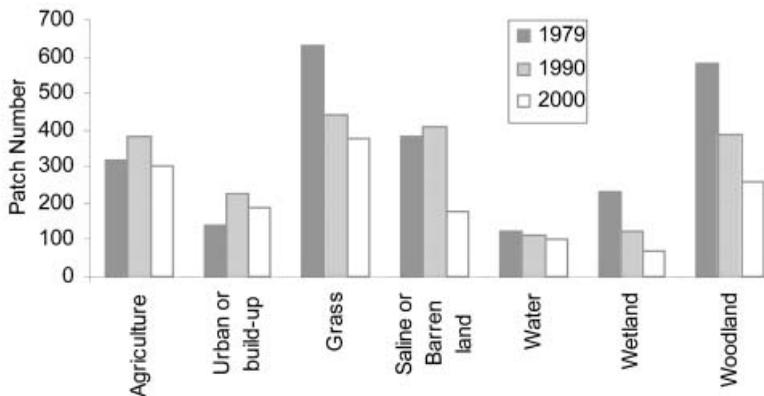


Figure 6. The number of patches for each land use class between August 1979 and June 2000.

Table 3. Transition matrix of land use area (km<sup>2</sup>) between 1990 and 2000.

	Agriculture	Urban or built-up	Grass	Saline or barren land	Water	Wetland	Woodland
Agriculture	505.57	12.58	81.58	5.30	0.66	0.96	24.50
Urban or built-up	4.30	49.56	29.53	1.59	1.27	0.09	0.36
Grass	180.08	40.42	339.11	51.75	3.68	4.46	30.93
Saline or barren land	16.65	13.62	60.05	49.25	2.29	0.23	2.79
Water	3.17	6.94	28.39	4.79	89.85	3.33	6.15
Wetland	10.96	0.72	61.95	0.60	8.50	50.03	33.11
Woodland	46.12	2.10	64.29	1.25	1.45	11.63	68.65

during 1990–2000. This result indicates a strengthened fragmentation and spread process of agriculture during this period. By examining the detailed landscape change data (tables 2 and 3), it is clear that the exchange between agriculture and grassland was the most frequent transition. During the first decade, a considerable cultivated area was converted to grass because of construction of an oil field in the 1980s. As this construction spoiled the original fertile soil, the regional government endeavoured to protect the cultivated area, which resumed agriculture in the 1990s (Statistics Bureau of Daqing 2001).

#### 4.2 Modelling results and validation

Before we apply the modified Markov model to time series data, we use a simple  $3 \times 3$  contingency table with the  $\chi^2$  statistic to test that the landscape change in the study region was not random, which is required for further Markovian analysis. The test statistic ( $P < 0.01$  with degree of freedom = 7) was significant for each comparison, 1979 with 1990, and 1990 with 2000.

Our algorithm was implemented in Matlab for the landscape maps from 1979, 1990 and 2000, respectively. The computation is based on the actual observations and time span during the landscape change, regardless of the way that the change process occurred. The temporal transitional probability matrix is calculated using equation (2), by accumulating the periods 1979–1990 and 1990–2000, and the spatial probability matrix is calculated using equation (6), based on the initial landscape map in 1979. Table 4(a) is the yearly transitional probability matrix without spatial simulation, which is based solely on Markov transition probabilities. To further test our modified GA, we applied a contrasting approach to examine the spatial factors. Table 4(b) is the yearly transitional probability matrix, based on the four nearest neighbours of all pixels, while table 4(c) is based only on the pixels under boundary conditions (figure 4).

Using three different transition probability matrices in table 4, we predicted and compared the simulated landscape maps to the empirical landscape map from the 2001 satellite image (table 5). From table 5, we can see that the GA is valid for incorporating the complex spatio-temporal information into the landscape dynamical model, but easily reaches its limit due to its inherent inability to consider spatial arrangement of pixels depending upon different classes. For the landscape type with large patch areas, such as agriculture and grass, the spatial Markov predict (38.53% and 31.51%) is better than both the boundary Markov predict (37.02% and 32.56%) and the Markov predict (36.67% and 31.97%). For the small-area landscape or landscape with small patches, such as urban and wetland, the boundary Markov

Table 4. Yearly transition probability (%) matrix for (a) the Markov model (MM), (b) the spatial Markov model (SMM) and (c) the boundary Markov model (BMM) from 1979 to 2000.

	Agriculture	Urban or built-up	Grass	Saline or barren land	Water	Wetland	Woodland
<i>(a) MM</i>							
Agriculture	96.56	0.23	2.24	0.26	0.07	0.09	0.55
Urban or built-up	0.70	95.86	2.83	0.29	0.17	0.06	0.09
Grass	2.44	0.45	95.04	0.81	0.20	0.29	0.77
Saline or barren land	1.36	0.74	3.82	93.18	0.39	0.12	0.39
Water	0.17	0.34	1.40	0.51	96.65	0.64	0.29
Wetland	1.13	0.10	2.69	0.15	0.60	93.76	1.58
Woodland	4.20	0.14	2.49	0.13	0.10	0.58	92.37
<i>(b) SMM</i>							
Agriculture	96.37	0.27	2.33	0.29	0.07	0.09	0.58
Urban or built-up	0.89	95.21	3.33	0.36	0.16	0.01	0.03
Grass	2.82	0.47	94.52	0.89	0.21	0.30	0.80
Saline or barren land	1.71	0.77	4.24	92.41	0.44	0.09	0.34
Water	0.11	0.37	1.71	0.53	96.34	0.60	0.34
Wetland	1.19	0.13	2.95	0.12	0.63	93.28	1.68
Woodland	4.96	0.16	2.49	0.08	0.06	0.59	91.66
<i>(c) BMM</i>							
Agriculture	96.39	0.27	2.35	0.28	0.06	0.08	0.57
Urban or built-up	0.75	95.82	3.00	0.26	0.14	0.00	0.02
Grass	2.59	0.44	95.00	0.79	0.19	0.26	0.73
Saline or barren land	1.53	0.73	3.88	93.08	0.39	0.07	0.32
Water	0.08	0.36	1.63	0.46	96.60	0.53	0.33
Wetland	1.05	0.13	2.88	0.09	0.59	93.64	1.62
Woodland	4.59	0.16	2.37	0.07	0.06	0.51	92.25

predict (5.81% and 4.94%) is better than the spatial Markov predict (5.57% and 4.98%) and Markov predict (5.58% and 5.45%). It can also be seen that a majority of boundary Markov predict values based on the boundary condition pixels fall between the Markov predict and the spatial predict. Using the modified GA, the outcome can be attributed to the dominance of the spatial probabilities  $P_{rc}$  over the temporal probability  $P_m$  for the large-patch landscape types, while changes will be more prevalent on the boundary area for the small-patch landscape types. Therefore, the spatial factor for the small patches has a larger influence than that for the large patches. Standardizing the weight factor of patch size and balancing it between the spatial factor and temporal factor will be difficult.

Table 6 shows the ID and GD calculated from the simulated map and empirical map in 2001 using equation (9). The results from the spatial-based models are superior to those from the Markov prediction. Both, the smallest ID (0.002) and the smallest GD (1.103) were found in the spatial Markov predict. These results indicate that inclusion of spatial pattern information in a transition probability matrix by means of GA can effectively improve the performance of the Markov prediction model. Specifically, among the seven classes, saline or barren land and water, which

Table 5. Comparison of simulated results (%) using the Markov model (MM), the spatial Markov model (SMM) and the boundary Markov model (BMM) with the empirical classification map in 2001.

	Agriculture	Urban or built-up	Grass	Saline or barren land	Water	Wetland	Woodland
Empirical landscape	39.60	6.51	31.59	6.35	6.22	3.11	6.63
MM	36.67	5.58	31.97	6.59	5.52	5.45	8.23
SMM	38.53	5.57	31.51	6.36	5.27	4.98	7.78
BMM	37.02	5.81	32.56	6.41	5.30	4.94	7.96

Table 6. The class-specific individual deviation (ID) and the overall global deviation (GD) between the empirical map and simulated map in 2001.

Model	Agriculture	Urban or built-up	Grass	Saline or barren land	Water	Wetland	Woodland	GD
MM	0.07	0.14	0.01	0.04	0.11	0.75	0.24	1.37
SMM	0.03	0.14	0.00	0.00	0.15	0.60	0.17	1.10
BMM	0.07	0.11	0.03	0.01	0.15	0.59	0.20	1.15

MM, the Markov model; SMM, the spatial Markov model; BMM, the boundary Markov model.

belong to natural landscape cover types, present a similar order of individual deviation between the spatially-based prediction and the Markov prediction, while for the most common human-disturbed landscape, agriculture, the spatially-based prediction results performed better than the sole Markov prediction. It seems that the spatially-based prediction using GA has a better capability to capture the spatial characterization of human-disturbed landscape than the natural landscape as the human-disturbed landscape is less likely to change randomly.

As a demonstration of the utility of our approach, we have generated future landscapes in 2010, 2030 and 2050 by the SMM (table 7). A notable trend is discerned; the sprawl of city and oil field will increase cultivation of the grass and lead to further loss of the natural landscape. An increasing amount of wetland and woodland will be fragmented into grass, resulting in a more fragmented landscape. Agriculture will be degraded to grass or even barren land because of damage to the soil structure after the construction of the oil field.

Table 7. The predicted results (%) from the spatial Markov model (SMM) in 2010, 2030 and 2050.

	Agriculture	Urban or built-up	Grass	Saline or barren land	Water	Wetland	Woodland
2010	37.53	6.47	33.11	6.20	5.06	4.04	7.59
2030	38.03	7.27	33.54	6.08	4.68	3.19	7.21
2050	38.18	7.60	33.66	6.07	4.46	2.95	7.08

## 5. Conclusions

The spatio-temporal model of landscape patterns using multitemporal Landsat TM and MSS imagery enabled us to identify the patch distribution in our study area and monitor the landscape dynamics for Daqing City. This study explored the potential of satellite remote sensing images and Markov chain models to predict future landscape change. Moreover, utilization of the GA for representing spatial information in the spatio-temporal model has proved to be a practical and effective method.

The Markov chain model, coupled with the GA, has indicated the descriptive capability of trend projection. This spatio-temporal model provides not only a quantitative description of change in the past but also the direction and magnitude of change in the future. However, based on the experimental results and exploratory analysis, several limitations are present in the current study:

- As the modelling process involves the use of data from multiple sources, the accuracy of prediction results will be closely related to the individual accuracy of each type of data. Developing a solid method for incorporating data from different sources, different data structures, as well as different spatial resolutions, is still a challenge.
- The transition probability in the Markov chain model is assumed to be uniform. Therefore, it is still difficult to accommodate the unpredictable influence of variables, such as the climate, governmental policy or human disturbance. In addition, the pace of landscape change is not usually steady over a whole period.
- Although the modified GA in this study achieved a considerable improvement over the Markov chain model, the exact location of classes still cannot be simulated. An examination of the relationship between landscape change and its location remains an interesting research topic.

At this time, it is not fully conclusive that the transition probability based only on boundary pixels is inferior to that based on all pixels. It is necessary to apply the developed methods to a series of varied landscapes that may present different spatial arrangements. However, it can be concluded that by incorporating more spatial algorithms into the prediction of landscape change, more accurate long-term forecasts can be made in the future.

## References

- ABED, J. and KAYSI, I., 2003, Identifying urban boundaries: application of remote sensing and geographic information system technologies. *Canadian Journal of Civil Engineering*, **30**, pp. 992–1000.
- ANDERSON, T.W. and GOODMAN, L.A., 1957, Statistical inference about Markov chains. *Annals of Mathematical Statistics*, **28**, pp. 89–110.
- BAKER, W.L., 1989, A review of models of landscape change. *Landscape Ecology*, **2**, pp. 111–133.
- BARNESLEY, M.J. and BARR, S.L., 1996, Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric Engineering and Remote Sensing*, **62**, pp. 949–958.
- BELL, E.J., 1974, Markov analysis of land use change: an application of stochastic processes to remotely sensed data. *Socio-Economic Planning Sciences*, **8**, pp. 311–316.
- BORNHOLDT, S., 1998, Genetic algorithm dynamics on a rugged landscape. *Physical Review E*, **57**, pp. 3853–3860.

- BOURNE, L.S., 1976, Monitoring change and evaluating the impact of planning policy on urban structure: a Markov chain experiment. *Plan Canada*, **16**, pp. 5–14.
- BROWN, D.G., PIJANOWSKI, B.C. and DUH, J.D., 2000, Modelling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management*, **59**, pp. 247–263.
- CHAKRABORTY, U.K., DEB, K. and CHAKRABORTY, M., 1995, Analysis of selection algorithms: a Markov chain approach. *Evolutionary Computation*, **4**, pp. 133–167.
- CLARKE, K.C. and HOPPEN, S., 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.
- FORTIN, M.J., BOOTS, B., CSILLAG, F. and REMMEL, T.K., 2003, On the role of spatial stochastic models in understanding landscape indices in ecology. *OIKOS*, **102**, pp. 203–212.
- GOLDBERG, D.E. (Ed.), 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning*, pp. 15–23 (Reading, MA: Addison-Wesley Longman).
- GRIFFITH, J.A., STEHMAN, S.V., SOHL, T.L. and LOVELAND, T.R., 2003, Detecting trends in landscape pattern metrics over a 20-year period using a sampling-based monitoring programme. *International Journal of Remote Sensing*, **24**, pp. 175–181.
- HEROLD, M., GOLDSTEIN, N.C. and CLARKE, K.C., 2003, The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote Sensing of Environment*, **86**, pp. 286–302.
- HOOD, J.H. (Ed.), 1975, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, pp. 30–62 (Ann Arbor: University of Michigan Press).
- JENERETTE, G.D. and WU, J., 2001, Analysis and simulation of land-use change in the central Arizona–Phoenix region, USA. *Landscape Ecology*, **16**, pp. 611–626.
- MERTENS, K.C., VERBERKE, L.P.C., DUCHEYNE, E.I. and WULF, R.R.D., 2003, Using genetic algorithms in sub-pixel mapping *International Journal of Remote Sensing*, **24**, pp. 4241–4247.
- MULLER, M.R. and MIDDLETON, J., 1994, A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecology*, **9**, pp. 151–157.
- PONTIUS, R.G., JR. 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.
- PONTIUS, R.G., JR., CORNELL, J.D. and HALL, C.A.S., 2001, Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Agriculture, Ecosystems and Environment*, **1775**, pp. 1–13.
- RIDD, M.K. and LIU, J., 1998, A comparison of four algorithms for change detection in an urban environment. *Remote Sensing of Environment*, **63**, pp. 95–100.
- ROY, P.S. and TOMAR, S., 2001, Landscape cover dynamics pattern in Meghalaya. *International Journal of Remote Sensing*, **22**, pp. 3813–3825.
- SHIBASAKI, R. and HUANG, S., 2001, Integration of observational data and behavioral models for spatio-temporal interpolation: application to reconstructing long-term land use and land cover changes. In *Present and Future of Modeling Global Environmental Change: Toward Integrated Modeling*, T. Matsuno and H. Kida (Eds), pp. 293–309 (Tokyo: Terra Scientific Publishing Company).
- STATISTICS BUREAU OF DAQING, 2001, *Daqing Statistical Yearbook in 2001* (Heilongjiang: Heilongjiang Statistics Bureau).
- STEWART, D.J., YIN, Z., BULLARD, S.M. and MACLACHLAN, J.T., 2004, Assessing the spatial structure of urban and population growth in the Greater Cairo area, Egypt: a GIS and imagery analysis approach. *Urban Studies*, **41**, pp. 95–116.
- STEWART, W.J. (Ed.), 1994, *Introduction to the Numerical Solution of Markov Chains*, pp. 12–22 (Princeton, NJ: Princeton University Press).

- STOW, D.A. and CHEN, D.M., 2002, Sensitivity of multitemporal NOAA AVHRR data of an urbanizing region to land-use/land-cover changes and misregistration. *Remote Sensing of Environment*, **80**, pp. 297–307.
- TURNER, M.G., 1987, Spatial simulation of landscape changes in Georgia: a comparison of 3 transition models. *Landscape Ecology*, **1**, pp. 29–36.
- YAGOUB, M.M., 2004, Monitoring of urban growth of a desert city through remote sensing: Al-Ain, UAE, between 1976 and 2000. *International Journal of Remote Sensing*, **25**, pp. 1063–1076.
- YANG, X. and LO, C.P., 2002, Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing*, **23**, pp. 1775–1798.
- YEH, A.G.-O. and LI, X., 1997, An integrated remote sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the pearl river delta, China. *International Planning Studies*, **2**, pp. 193–211.
- WENG, Q., 2001, A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. *International Journal of Remote Sensing*, **22**, pp. 1999–2014.
- WENG, Q., 2002, Land use change analysis in the Zhujiang delta of China using satellite remote sensing, GIS and stochastic modeling. *Journal of Environmental Management*, **64**, pp. 273–284.