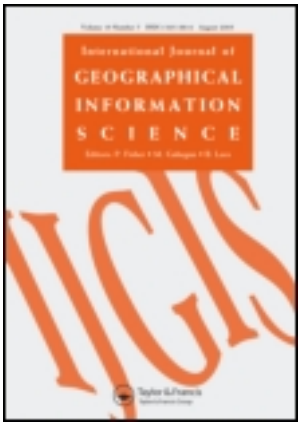

International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

Knowledge transfer and adaptation for land-use simulation with a logistic cellular automaton

Xia Li[a], Yilun Liu[a], Xiaoping Liu[a], Yimin Chen[a] & Bin Ai[a]
[a] School of Geography and Planning, and Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, 510275, PR, China
Published online: 12 Sep 2013.

To cite this article: Xia Li, Yilun Liu, Xiaoping Liu, Yimin Chen & Bin Ai (2013) Knowledge transfer and adaptation for land-use simulation with a logistic cellular automaton, International Journal of Geographical Information Science, 27:10, 1829-1848, DOI: 10.1080/13658816.2013.825264

To link to this article: http://dx.doi.org/10.1080/13658816.2013.825264

International Journal of Geographical Information Science, 2013
Vol. 27, No. 10, 1829–1848, http://dx.doi.org/10.1080/13658816.2013.825264

Knowledge transfer and adaptation for land-use simulation with a logistic cellular automaton

Xia Li*, Yilun Liu, Xiaoping Liu, Yimin Chen and Bin Ai

School of Geography and Planning, and Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou 510275, PR China

(*Received 8 June 2012; accepted 10 July 2013*)

Few studies have been conducted into the use of knowledge transfer for tackling geo-simulation problems. Cellular automata (CA) have proven to be an effective and convenient means of simulating urban dynamics and land-use changes. Gathering the knowledge required to build the CA may be difficult when these models are applied to large areas or long periods. In this paper, we will explore the possibility that the knowledge from previously collected data can be transferred spatially (a different region) and/or temporally (a different period) for implementing urban CA. The domain adaptation of CA is demonstrated by integrating logistic-CA with a knowledge-transfer technique, the *TrAdaBoost* algorithm. A modification has been made to the *TrAdaBoost* algorithm by incorporating a dynamicweight-trimming technique. This proposed model, CA_{trans}, is tested by choosing different periods and study areas in the Pearl River Delta. The 'Figure of Merit' measurements in the experiments indicate that CA_{trans} can yield better simulation results. The variance of traditional logistic-CA is about 2–5 times the variance of CA_{trans} until the number of new data reaches 30. The experiments have demonstrated that the proposed method can alleviate the sparse data problem using knowledge transfer.

Keywords: knowledge transfer; cellular automata; urban simulation; model adaptation

1. Introduction

The sharing and reuse of existing data and knowledge for geographical applications among different regions and domains has proven to be necessary and useful (Fonseca *et al*. 2000). Some studies have been carried out for sharing data and knowledge in geographical information systems (GIS) using the method of spatial ontologies (Fonseca and Egenhofer 1999). However, this assumes that designers should define and conceptualize the knowledge in a domain by providing standardized vocabulary, semantic terminology, and methodologies during the design stage.

Cellular automata (CA) are bottom-up simulation tools that rely on transition rules for modeling the behavior of complex systems (Wolfram 1986, 2006, Toffoli and Margolus1987). Over the last three decades, CA for geographical simulation have proliferated because of their simplicity, flexibility, and intuitiveness (Santé *et al*. 2010). Although there are other types of bottom-up model (e.g., agent-based models), CA have been widely used for simulating a variety of geographical phenomena, such as urban development

*Corresponding author. Email: lixia@mail.sysu.edu.cn; lixia@graduate.hku.hk

(Batty and Xie 1994, Wu 1998, Li 2011), land-use changes (White and Engelen 1993), landscape evolution (Soares-Filho *et al*. 2002), wildfire spread (Clarke *et al*. 1997), and population dynamics (Couclelis 1988).

The reusability of CA in geographical applications is appealing for a number of reasons: (1) inexperienced users do not want to build a brandnew CA model; (2) the collection of a new set of training data is expensive and time-consuming; and (3) past experiences or old data are useful for capturing long-term trends. However, the reuse of CA for solving real problems does pose certain challenges. Studies have shown that the use of fixed transition rules will result in a large amount of simulation error because of spatiotemporal heterogeneity at a regional scale or over a long period (Li *et al*. 2008). CA can be implemented by rebuilding the models from scratch using newly collected training data. However, obtaining labeled training data about land cover from remote-sensing imagery is still a tedious job in most situations (e.g., relatively unfamiliar environment and inaccessible locations) (Rajan *et al*. 2008). A solution to this dilemma is to transfer the transition rules of CA from past applications to new applications.

Transfer learning techniques have attracted increasing attention in computer sciences in recent years (Dai *et al*. 2007). These techniques have mainly been developed in the field of data mining and machine learning. The goal of transfer learning is to improve learning by transferring knowledge obtained from previous tasks to new tasks in the target domain. Transfer learning was primarily motivated by the need to learn efficiently (Schmidhuber 1995).

Traditional machine-learning methods often make predictions using statistical models that are trained on labeled data. Many applications often contain a lot (e.g., hundreds or thousands) of previously (old) collected labeled data. It would be useful if these old data, plus a tiny set (e.g., 5–10) of new, collected data, could be used for building models. Most traditional methods assume that the sample distribution of previously collected data is the same as that of new data. However, this assumption may not be true because old data usually have a different sample distribution than the data from the target domain. Such distribution variations (diff-distribution) are responsible for poor simulation performances using all these training data for large areas which may consist of a number of cities (Li *et al*. 2008).

Transfer learning can effectively deal with the situations in which the domains, tasks, and distributions vary between training and testing data.A well-known method for transfer learning is that of boosting algorithms. One of the first simple boosting procedures in computational learning theory was developed by Schapire (1990).Freund and Schapire (1995) later proposed a ‘boost by majority’ algorithm which uses many weak (not accurate) learners simultaneously to improve the performance of the simple boosting algorithm (Friedman *et al*. 2000).

Transfer learning techniques for solving geographical problems are quite unique because of the inherent spatiotemporal characteristics. This paper attempts to develop a new method of domain adaption across spatiotemporal dimensions that can facilitate urban and land-use simulation. We have used transfer learning techniques because previously collected data are useful for defining transition rules of CA. However, these previous experiences cannot be used directly without knowledge transfer. We can get round this problem using an instance-based approach, which is based on the revised *TrAdaBoost*. We will test this method for the urban simulation of different cities in the Pearl River Delta, China.

2. Knowledge transfer of CA based on the revised *TrAdaBoost* algorithm

This paper will present a domain adaptation method for using CA according to knowledge-transfer techniques. This method is based on an instance-transfer approach, by which some parts of the old instances are reused together with a few new labeled instances (e.g., new samples with known land-use types) for a new application (target domain). The knowledge transfer of CA is implemented by revising the *TransferAdaBoost*(*TrAdaBoost*) learning algorithm. This algorithm was originally proposed by Dai *et al*. (2007), although some modifications must be made in order to adapt it to the knowledge transfer of CA.

The procedure for the modifications includes: (1) defining logistic-CA as the basic learner; and (2) revising the *TrAdaBoost* algorithm for constructing the ensemble of weak logistic-CAs. Two types of labeled land-use data are used in this method, including the abundant supply of auxiliary (old) labeled data from previous tasks (e.g., previous periods or other regions) and a tiny set of the base (new) labeled data from the target domain. The detailed methodology of this proposed model, CA_{trans}, is described as follows:

(1) Logistic-CA as the basic learner

The first step for building CA_{trans} is to define the basic learner, which is dependent on the application domain. In this study, a typical urban cellular automaton, the logistic-CA, is selected as the basic learner for urban simulation. The logistic-CA is quite easy to define and convenient to calibrate using training data (Wu 2002, Li *et al*. 2011). The methodology will be the same if the basic learner is replaced by other CA, such as ANN-CA (Li *et al*. 2011) and genetic-CA (Li *et al*. 2008). The logistical model can be used to represent the conversion probability from the non-urban to the urban land for simulating land-use dynamics (Wu 2002). The conversion probabilities of CA are subject to change according to a series of factors. This is quite different from the fundamental Markov assumption of time-invariant transition probabilities (Rabiner and Juang 1986). The variant conversion probability of logistic-CA is estimated as follows:

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

where p_{ij}^{t} is the conversion probability for cell ij at time t, z_{ij}^{t} is the combined assessment score for conversion suitability ($z_{ij}^{t} = a_0 + a_1 x_{1} + a_2 x_{2} + \cdots + a_m x_{m} + \cdots + a_M x_{M}$), a_0 is the constant, x_{m} is a spatial (physical) variable representing a driving force for land-use conversion, and $a_m$is the parameter (weight) associated with variable x_{m}.

The above combined score, p_{ij}^{t}, only addresses the global factors in terms of various proximity variables. However, urban development is influenced by the interactions at local as well as global levels. Moreover, some spatial constraints can be incorporated to reflect the site conditions that also affect land-use conversion. By integrating all these geographical factors, the development probability for urban simulation is further revised as follows (Li *et al*. 2011):

$$p_{ij}^{t} = (1+(-\ln\gamma)^{\alpha})\frac{1}{1+\exp(-z_{ij}^{t})} \times f(\Omega_{ij}^{t}) \times con_{ij} \tag{2}$$

where γ is a stochastic factor ranging from 0 to 1, α is the parameter to control the stochastic degree, $f(\Omega_{ij}^t)$ is the development intensity in the neighborhood of Ω_{ij}, and con_{ij} is the total constraint score ranging from 0 to 1.

Finally, p_{ij}^t is compared with a threshold value to determine if a non-urbanized cell will be converted into an urbanized cell at each iteration of simulation:

$$S_{ij}^{t+1} = \begin{cases} Converted, p_{ij}^t \geq Q_{land} \\ NonConverted, p_{ij}^t < Q_{land} \end{cases} \quad (3)$$

where S_{ij}^{t+1} is the state (land-use type) of cell ij at next time $(t + 1)$, and Q_{land} is the threshold value which is related to the amount of land conversion.

The threshold (Q_{land}) is calculated using observation data or an exogenous growth model by predicting land demand. For example, this value can be determined in such a way that the total number of converted cells will be equal to the real number estimated from the observed remote-sensing data.

(2) Revised *TrAdaBoost* algorithm for logistic-CA

The second step for building CA_{trans} is to solicit transition rules of CA from old data plus a tiny set of new data. The so-called boost algorithm will be used to find a set of weak rules which are combined together to form the final prediction model. This algorithm is implemented by adjusting the weights of these training data for learning a (weak) rule accordingly from these weighted samples. This process allows the knowledge (transition rules) from an old domain to be adapted into a new domain by minimizing the prediction errors.

The *AdaBoost* algorithm, which was proposed by Freund and Schapire in 1995, has been widely used for machine learning (Schapire 2001). However, *AdaBoost* requires that the distributions of the training and test data must be identical. Dai *et al*. (2007) further developed the so-called *TrAdaBoost* algorithm, which is based on *AdaBoost*, to tackle the problem of different distributions. In our study, a revised *TrAdaBoost* algorithm is proposed to use the labeled data collected in different regions or different periods for the domain adaptation of CA.

Figure 1 illustrates how *TrAdaBoost* works for predicting land conversionbased on independent variables. This figure shows an example of establishing a logistic regression model between converted probability (urbanized or not) and development suitability. It is difficult to obtain an accurate regression model if only a small amount of new training data (blue points) is available in the target domain. The model built on these sparse data will be inaccurate or wrong (dash line in Figure 1a). If there are a lot of old data (orange points) with the same distribution as the new ones, these old data can be used to build a much more accurate model (Figure 1b). However, it will be problematic to build the model if these old data have a different (probability) distribution (Figure 1c). Figure 1d shows that these old data can only be used appropriately with a weight-adjusted scheme according to the assessment of these data. This scheme is to allow the old data that fit the target domain better (yellow points in the dotted circles) to have larger weights for the prediction.

In our study, the modifications of *TrAdaBoost* include two aspects: (1) designing a dynamicweight-trimming technique to facilitate the domain adaptation of CA; and (2) embedding logistic-CA into *TrAdaBoost* for urban simulation. The detailed methodology for revising *TrAdaBoost* for urban simulation is described as follows:

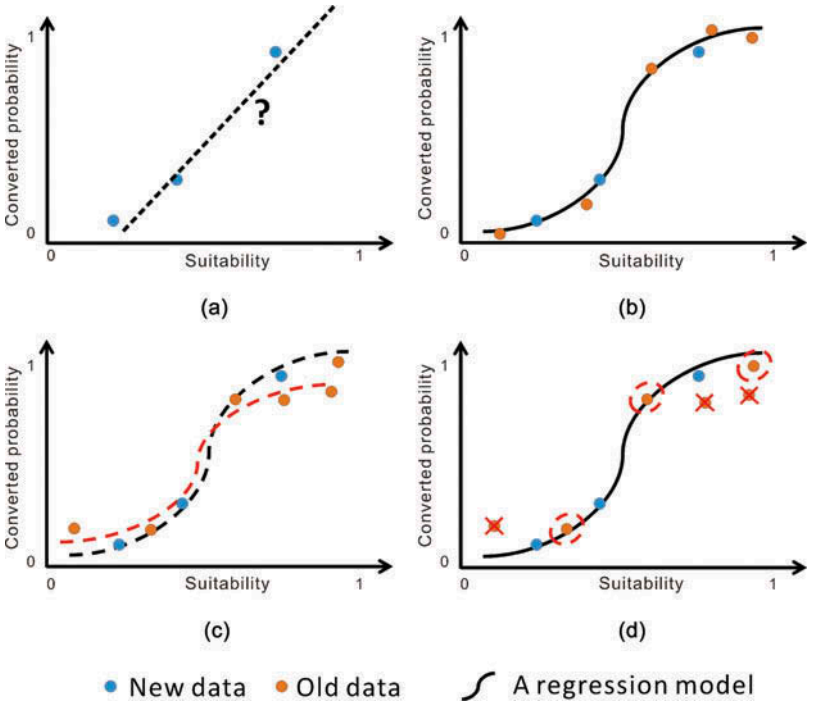

Figure 1. An example of *TrAdaBoost* algorithm for dealing with training data of different distributions. (a) Obtaining a satisfactory regression model is difficult without enough new data. (b) Old data can be directly used to train a satisfactory regression model if their distributions are the same. (c) There is a problem to use old data directly because the distributions of new data and old data are different. (d) The TrAdaBoost algorithm provides a way to use old data for training a better regression model.

Step 1: Preparing the inputs.

The first step is to prepare the labeled land-use data and define the maximum number of iterations (T) for the *TrAdaBoost* algorithm. The labeled land-use data provide the empirical information for creating weak learners. The maximum number of iterations which defines the set (ensemble) of weak learners is determined according to the iteration curve of error rate (Dai *et al*. 2007). It is assumed that the improvement of prediction accuracy will be stabilized after a certain number of iterations.

The labeled-land use data (instances) are usually obtained by classifying remote-sensing images or carrying out field investigations. These instances include an auxiliary (old) data set $D_a\{(x_1, y_1), \cdots, (x_N, y_N)\}$ (N is the number of old data) collected from a previous task, and a base (new) data set $D_b\{(x_{N+1}, y_{N+1}), \cdots, (x_{N+L}, y_{N+L})\}$ (L is the number of new data) collected from a new task ($x_i \in X = R^p$ and $y_i \in Y = \{1, 0\}$; R^p refers to all the instances or the instance space in which each instance is assumed to be represented by a set of attribute–value pairs). In each instance (sample), the variables of x and y represent the site attributes at a location and its category label (e.g., land-use type) respectively. A Boolean function is used to map X to Y. The base learning algorithm is the above logistic-CA.

Step 2: Initializing the weights for auxiliary data and base data.

These two labeled data sets (samples) usually have different distributions because of spatiotemporal variations (Dai *et al*. 2007). A weight can be used to represent the importance of a sample for the prediction. The first step is to initialize the weights for both auxiliary data and base data before calculating the weight decay factor of auxiliary data. It is assumed that all these data can be used for regression or prediction, but the contribution of each sample to the prediction varies according to its weight.

At the beginning, all these weights are initialized as equalfor both auxiliary data and base data. These initial equal weights are defined as follows:

$$w_1(i) = \begin{cases} \frac{1}{N} & \text{If } 1 \le i \le N \\ \frac{1}{L} & \text{If } \mathrm{N}+1 \le i \le N+L \end{cases} \tag{4}$$

where N and L are the total numbers of the auxiliary data and base data, respectively.

Step 3: For u = 1, . . . , U, running the base learner (logistic-CA) and adjusting the weight of each sample according to its prediction performance.

CA_{trans} is calibrated from a weighted set of auxiliary data (D_a) and base data (D_b). This CA consists of a number of weak learners which are generated using different combinations of these data. At each iteration, a part of auxiliary data and base data are randomly selected to create a weak learner.

First, the weight of each sample (a pair of labeled data) is normalized using the following equation:

$$w(i) = w(i) \Big/ \sum_{i=1}^{N+L} w(i),\ 1 \le i \le N+L \tag{5}$$

A dynamicweight-trimming technique is proposed so that *TrAdaBoost*can be applied to the knowledge transfer of CA. Only those samples whose weights are greater than a dynamic threshold will be selected to generate a weak rule. This dynamic threshold is defined as follows:

$$\beta_u = mean\left(w_u(1), \cdots, w_u(N)\right) \cdot \gamma,\ 0 \le \gamma \le 1 \tag{6}$$

where γ is a random variable, and u is the current number of iterations ($u = 1, \ldots, U$).

The above procedure will find a set of suitable samples for building a weak learner (logistic regression model). This learner will map the land-use conversion according to the following equation:

$$f_u : X \to \{1, 0\} \tag{7}$$

where X is all the set of x (attributes).

The base data (D_b) from the target domain is then used to estimate the model error of this weak learner (f_u) (Dai *et al*. 2007):

$$\varepsilon_u = \sum_{i=N+1}^{N+L} w_u(i)\, Abs\left(f_u(x_i) - y_i\right) \Big/ \sum_{i=N+1}^{N+L} w_u(i) \tag{8}$$

where ε_u is the model error, $w_u(i)$ is the weight of the ith sample, $Abs(f_u(x_i)\text{-}y_i)$ is the difference between the predicted state (land-use type) from the logistic model and the true state from the observation (the training data), and ε_u is required to be less than 0.5.

The weight decay factors for D_a and D_b are defined as follows (Dai *et al*. 2007):

$$\alpha = 1/\left(1+\sqrt{2\ln N/U}\right) \text{ and } \alpha_u = \varepsilon_u/(1-\varepsilon_u) \tag{9}$$

The weights are dynamically updated according to the following equation (Dai *et al*. 2007):

$$w_{u+1}(i) = \begin{cases} w_u(i)\,\alpha^{Abs(f_u(x_u)-y_u)}, & 1 \le i \le N \\ w_u(i)\,\alpha_u^{-Abs(f_u(x_u)-y_u)}, & \mathrm{N}+1 \le i \le N+L \end{cases} \tag{10}$$

Step 4: Generating the hypothesis according to the ensembles of weak learners.

The above procedure will create U number of weak learners. The model error will decrease as the weights are updated according to Equation (10). As a result, the weak learners built at the later stages will be better than their precedents in terms of classification accuracy. In machine-learning literature, the hypothesis is that a given set of instances (or samples) can be used to produce a learner, sometimes also called a classification rule (Schapire *et al*. 1998). The final output of the model is thus based on the ensembles of the last $U/2$ weak learners (Dai *et al*. 2007):

$$\text{argmax}\left(\sum_{u=U/2}^{U}\alpha_u I\left(f_u(x_i)=y_i\right)\right) \tag{11}$$

where I is a sign function in which, if $f(x)=y$, then $I(f(x))=1$, else $I(f(x))=0$.

The innovation of this proposed method is the integration of the revised *TrAdaBoost* algorithm with logistic-CA. The final simulation is based on the ensembles of the last $U/2$ weak logistic-CA. This proposed method can allow *TrAdaBoost* to be extended to the knowledge transfer of urban simulation by dealing with different distributions of empirical data.

3. Model implementation

3.1. Study areas and spatial data

The proposed method was tested in the Pearl River Delta, which has an area of about 41,157 km^2. Situated in the mid-south of Guangdong, this study area consists of a number of administrative cities and districts, such as Guangzhou, Dongguan, Shenzhen, Zhongshan, and Foshan (Figure 2). Guangzhou is the largest city in southern China, and includes the city proper, Huadu district, Conghua district, Zengcheng district, and Panyu district. Before the economic reform in 1978, a major part of this region was engaged in intensive agricultural activities (e.g., growing rice, sugar cane, or banana) and fisheries (Seto *et al*. 2002). As a result of fast urbanization, this region has witnessed a large amount of land-use conversion and agricultural land loss. There is rich literature on the development of change detection and urban simulation methods for revealing urbanization and land-use problems in this fast-growing region (Seto *et al*. 2000, 2002, Li *et al*. 2008).

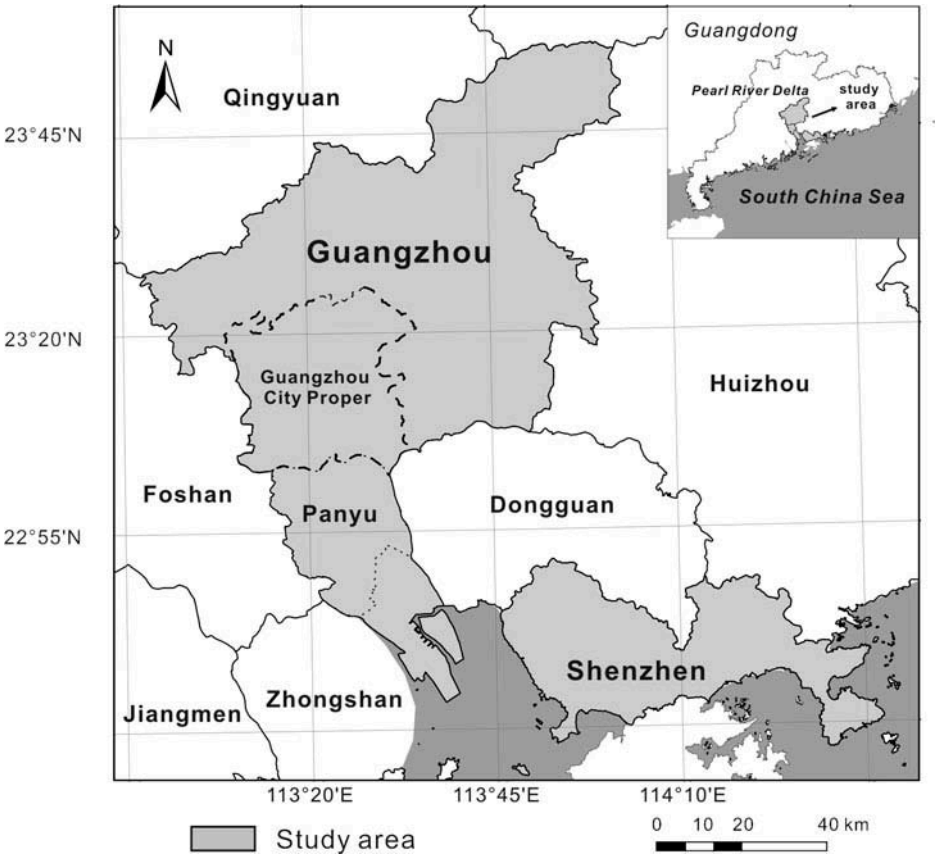

Figure 2. Location of the study area in the Pearl River Delta.

The labeled land-use data which were used to calibrate CA models were obtained from the classified Landsat TM images. Land-use classification was applied to the Landsat TM images of Guangzhou scene (Scene No. 122–44 in China Remote Sensing Ground Station reference system) dated 31 July 1986, 24 October 1994, 28 July 2000, and 4 March 2008, respectively. These images were radiometrically and geometrically corrected before the classification. First, the dark object subtraction (DOS) method was used to minimize the influences of different weather and light conditions on land-use classification (Chavez 1988). This procedure was implemented using the dark subtract tool of ENVI. Second, geometric corrections of these images were performed according to ground-control points. The total Root Mean Square (RMS) error of the geometric correction was less than 0.5 pixels. These corrected images were then classified using a series of techniques, such as object-based classification, manual editing, and intensive field labeling with GPS.

The classification accuracies for urban land uses are about 86–89% according to field checking (Chen *et al*. 2011). This means that the classification has an error of 11–14%.We used the method proposed by Pontius and Millones (2011) for comparing this error with the land-use changes across time. This method divides the disagreements between classification and reference into two parameters: quantity disagreement and allocation disagreement. We calculated the quantity and allocation disagreements with random sampling method for

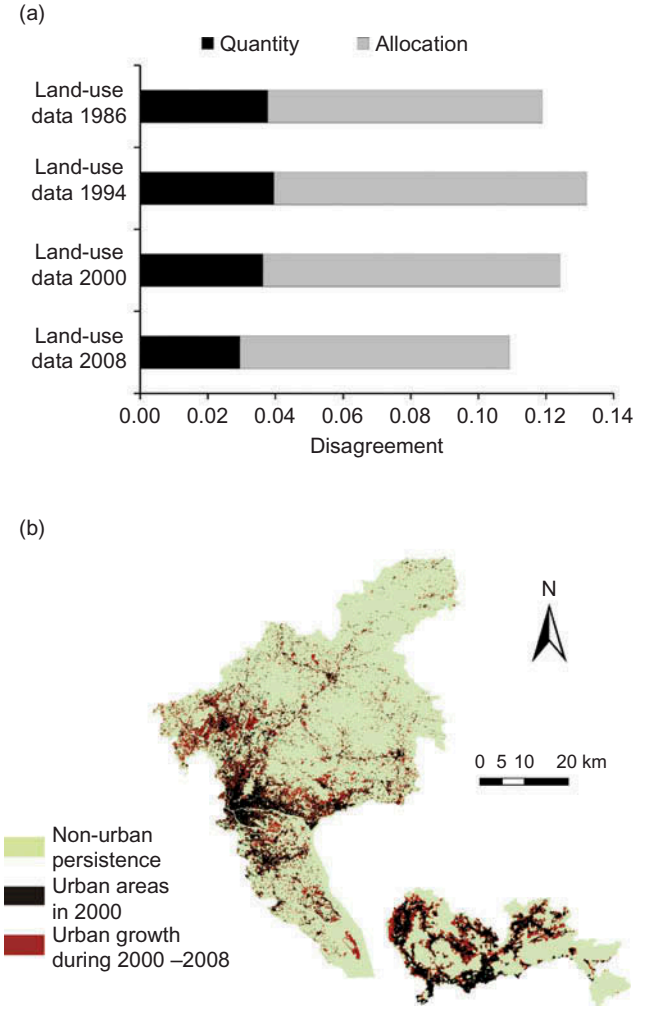

Figure 3. The classification accuracy of the land-use data. (a) Quantity and allocation disagreements and (b) consistency of the land-use classification over time.

the land-use data from 1986 to 2008. Figure 3a shows the two components of disagreement, which are stacked to show the total disagreement for these four years. The majority of disagreement comes from allocation disagreement, ranging from 8% to 10%, whereas the quantity disagreement is only 3–4%. Figure 3b just displays the urban growth for a part of the study area during 2000–2008, with the urban growth ratio ranging from 12.2% to 21.3%. The average classification disagreement is 11.7% for years 2000 and 2008, which is less than the difference (related to urban growth) between each pair of sequential maps.

We selected three cities/districts, Guangzhou, Panyu, and Shenzhen, to examine the effects of spatiotemporal knowledge transfer for urban simulation. Three sets of labeled data were collected for Guangzhou (7 districts), Panyu (1 district), and Shenzhen (Figure 4). Each set consists of training data and test data. The sites for collecting these data

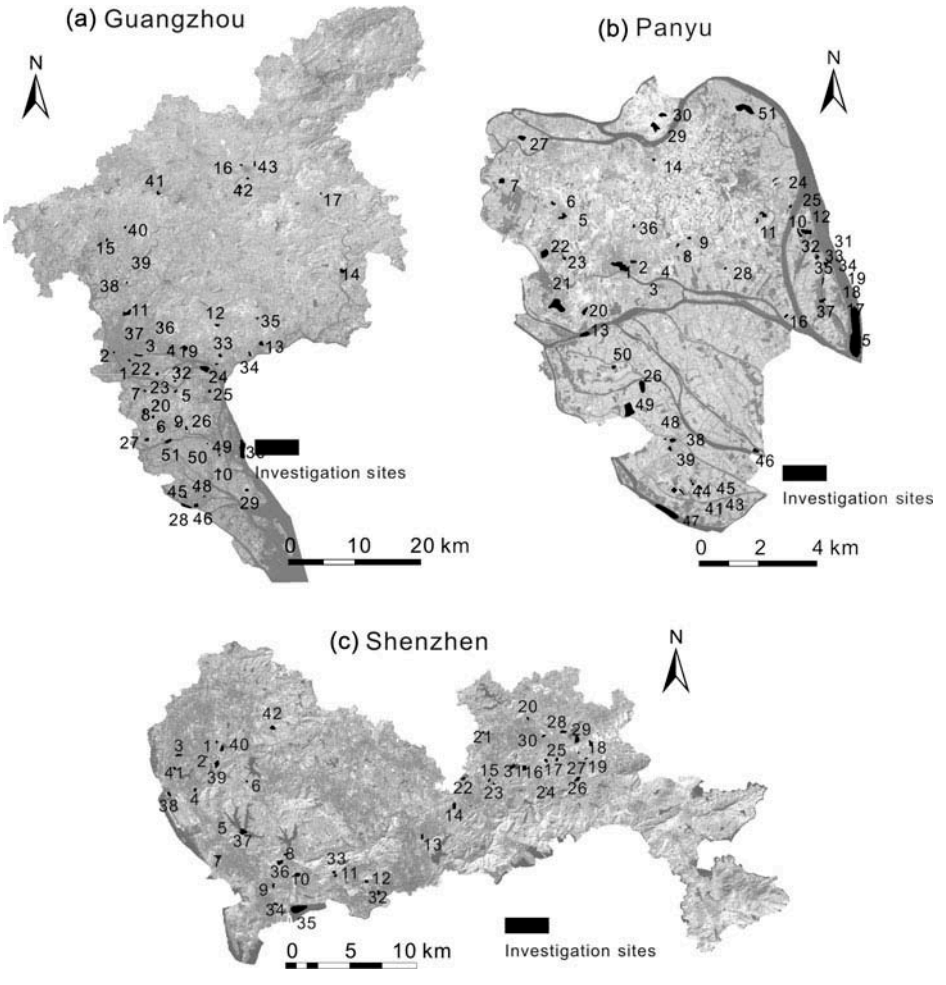

Figure 4. Obtaining three sets of labeled data for (a) Guangzhou, (b) Panyu, and (c) Shenzhen from classified remote-sensing data.

were determined based on the criteria of covering broad geographical locations and diverse land types. The detailed information about these labeled data is described as follows:

(1) Guangzhou labeled data

Guangzhou labeled data over two periods, 1986–1994 and 2000–2008, were used to build the temporal knowledge-transfer model for urban simulation. The Thematic Mapper (TM) images of 1986, 1994, 2000, and 2008 were classified to obtain the labeled data for land-use classes in 1986–1994 and 2000–2008, respectively. A total of 42 sites (patches) were identified for the period of 1986–1994 (Figure 4a). Within these sites, we randomly extracted a total of 2723 samples. These data were divided into two sets, 1000 as the training samples and 1723 as the test samples. A total of 51 sites (patches) were also investigated for the period of 2000–2008. These sites randomly yielded a total of 2109 pixels as the labeled data, which were divided into 1000 training samples and 1109 test samples.

(2) Panyu labeled data

Spatio-knowledge transfer for building the transition rules of CA was implemented using two sets of labeled data from Guangzhou and Panyu in 2000–2008. Besides the above Guangzhou's data, Panyu's labeled data of land-use classes were obtained from the classified 2000 and 2008 TM images. We selected 53 sites (patches) for collecting Panyu's training and test data (Figure 4b). A total of 2390 samples were randomly extracted from these sites. These samples were further split into 1000 training samples and 1390 test samples for building and validating the proposed model, respectively.

(3) Shenzhen labeled data

Shenzhen's labeled data in 2000–2008 were also acquired to test the effects of spatio-knowledge transfer for urban simulation. A total of 42 sites (patches) were selected for obtaining Shenzhen's training and test data (Figure 4c). A random selection from these sites created a total of 2620 samples, which were further divided into 1000 training samples and 1620 test samples, respectively.

The basic learner (logistic-CA) consists of two components of interactions, global interaction and local interaction, for addressing urban and land-use dynamics. The global interaction is represented by a logistic function of various proximity factors (e.g., urban centers, highways, and railways). The local interaction is represented by a neighborhood function of various land-use types (e.g., the amount of a land-use class in the neighborhood). The importance of these proximity factors and the neighborhood effects of land use for urban simulation have been extensively discussed by previous studies (White and Engelen 1993, Clarke *et al*. 1997, Verburg *et al*. 2002, Wu 2002, Li *et al*. 2008).

3.2. Experiments and model parameters

Knowledge transfer for urban and land-use simulation can be carried out in two folds, spatio transfer and temporal transfer. Three experiments were designed to examine the effects of spatiotemporal knowledge transfer with the use of the three labeled data sets above (Table 1).

Table 1. Three experiments of spatiotemporal knowledge transfer for urban simulation using various sets of labeled data.

Data	Region	Period
Experiment 1		
Base data set (D_b)	Panyu	2000–2008
Auxiliary data set (D_a)	Guangzhou	2000–2008
Experiment 2		
Base data set (D_b)	Shenzhen	2000–2008
Auxiliary data set (D_a)	Guangzhou	2000–2008
Experiment 3		
Base data set (D_b)	Guangzhou	2000–2008
Auxiliary data set (D_a)	Guangzhou	1986–1994

Experiment 1 (spatio-knowledge transfer from Guangzhou to Panyu)

The aim of the first experiment is to study the effects of spatio-knowledge transfer between the places at a closer distance (from Guangzhou district to Panyu district). The labeled data were previously collected in a larger area (Guangzhou). They will be used as the auxiliary data to facilitate urban simulation for a small, nearby area (Panyu). Although there is plenty of previously collected data (auxiliary data) in Guangzhou, some of the new data (base data) is still required for simulation at this new region (Panyu).

Experiment 2 (spatio-knowledge transfer from Guangzhou to Shenzhen)

The aim of this experiment is to examine the possibility of spatio-knowledge transfer between places that are further apart (from Guangzhou to Shenzhen). The labeled data previously collected in Guangzhou will be used for the simulation of another large city (Shenzhen), which is about 150 km away. Although sharing some similarities, these two cities have experienced quite different growth patterns during the study periods (Li *et al*. 2008).

Experiment 3 (temporal knowledge transfer from 1986–1994 to 2000–2008)

This experiment was just designed for the temporal knowledge transfer between different periods (from 1986–1994 to 2000–2008), but at the same place. In this experiment, the data previously collected in Guangzhou during the period1986–1994 were used for the simulation of the same city during the period 2000–2008.

The objective of these experiments is to explore the potential of using a large amount (e.g., 500 samples) of previously collected auxiliary (old) data set (D_a) with a tiny amount (e.g., 10–15 samples) of currently collected base (new) data (D_b) to construct CA models. There is usually a large amount of empirical data accumulated from past applications. It will be attractive if these data can serve as the empirical information for building a new simulation model. In the experiments, different combinations (ratios) of D_b/D_a are tested to identify the minimal amount of new data which is required for building the model. The amount of D_a is fixed to 500 samples while the amount of D_b is changed from 10 to 50 samples for the exploration. The data of D_b and D_a are randomly drawn from the labeled data described in section 3.1. In Experiment 1, for example, 10 samples of D_b were randomly drawn from the 1000 training samples of Panyu and 500 samples were randomly drawn from the 1000 training samples of Guangzhou. The ratio of D_b/D_a is then equal to 2%. A total of 21 combinations with the variations of D_b/D_a were obtained using this random sampling procedure. Each combination was repeated 10 times to reduce the uncertainties by randomly selecting these samples. Then, the final simulation was obtained from the average of the 10 repeated simulations.

The first step is to determine the ensemble size (the total number) of basic learners for producing satisfactory prediction results. A larger size of the ensemble can reduce the model error, but this will be at the expense of longer computation time. Actually, the ensemble size is decided according to the convergence curve of model error. We found that the decrease of the model error will be stabilized after the iteration (u) reaches 20 for most of the combinations (Figure 5). The maximum number of iterations (U) is determined based on the convergence trend. The value of U is set to 50, as some combinations may need to run 40–60 iterations in order to reach convergence. The total number of weak learners which are used for the simulation is, thus, equal to 25 ($U/2$) according to Equation 11.

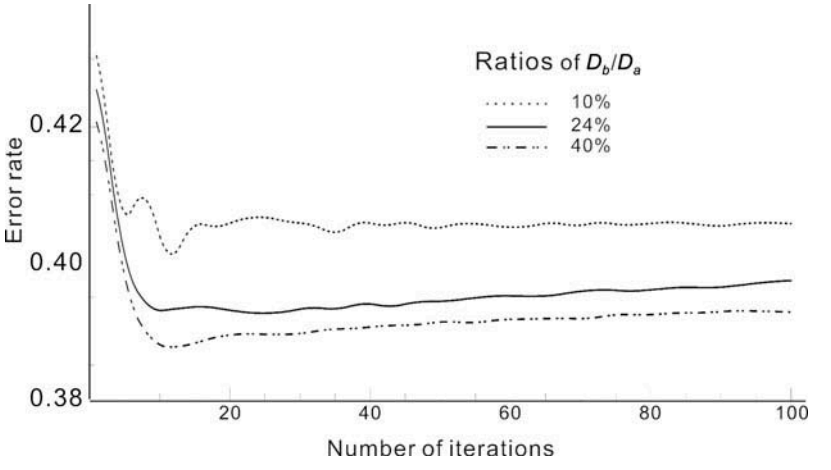

Figure 5. The error rate of the basic learner with the increase of the iteration (T) for the Shenzhen data set.

3.3. *Knowledge transfer of transition rules in spatiotemporal dimensions*

There are at least four major methods for implementing the knowledge transfer of CA using *TrAdaBoost* and *LogisticCA*. These methods are: (1) the proposed model based on the integration of *TrAdaBoost* and *LogisticCA* (CA_{trans}); (2) traditional *LogisticCA* based on the new data ($CA_{new\ data}$); (3) traditional *LogisticCA* based on the whole set of old data and new data ($CA_{all\ data}$); and (4) traditional *LogisticCA* just based on the old data ($CA_{old\ data}$).

These four models adopt different strategies of handling old and new labeled data. First, the proposed model (CA_{trans}) treats these two different sets of data according to the weights defined from the revised *TrAdaBoost* algorithm. It is expected that this method can well tackle the diff-distributionproblem with the use of these weights. Second, the traditional method of $CA_{new\ data}$ is to construct the model from scratch. This method is not efficient because it abandons all the previously collected labeled data. Third, the method of $CA_{all\ data}$ simply utilizes all these data for calibrating CA without considering their potential diff-distribution. This diff-distribution can cause the poor performances of simulation if the distribution bias of two sets of training data is large. Fourth, the method of $CA_{old\ data}$ also has drawbacks because this previously built CA is outdated without domain adaptation.

We use 'figure of merit' (FoM) to assess the simulation results of these models (Pontius *et al*. 2008). The FoM measurements in these experiments were derived from overlays of the reference map of the initial time, the reference map of the subsequent time, and the prediction map for the subsequent time. This indicator focuses on change instead of giving credit to correctly predicted persistence. Actually, FoM is a ratio, where the numerator is the intersection of the observed change and predicted change, while the denominator is the union of the observed change and predicted change (Pontius *et al*. 2008).

4. Results and validation

Experiment 1

The aim of this experiment is to test if the samples collected in Guangzhou can be reused for implementing urban simulation in Panyu. Figure 6a displays the relationship between

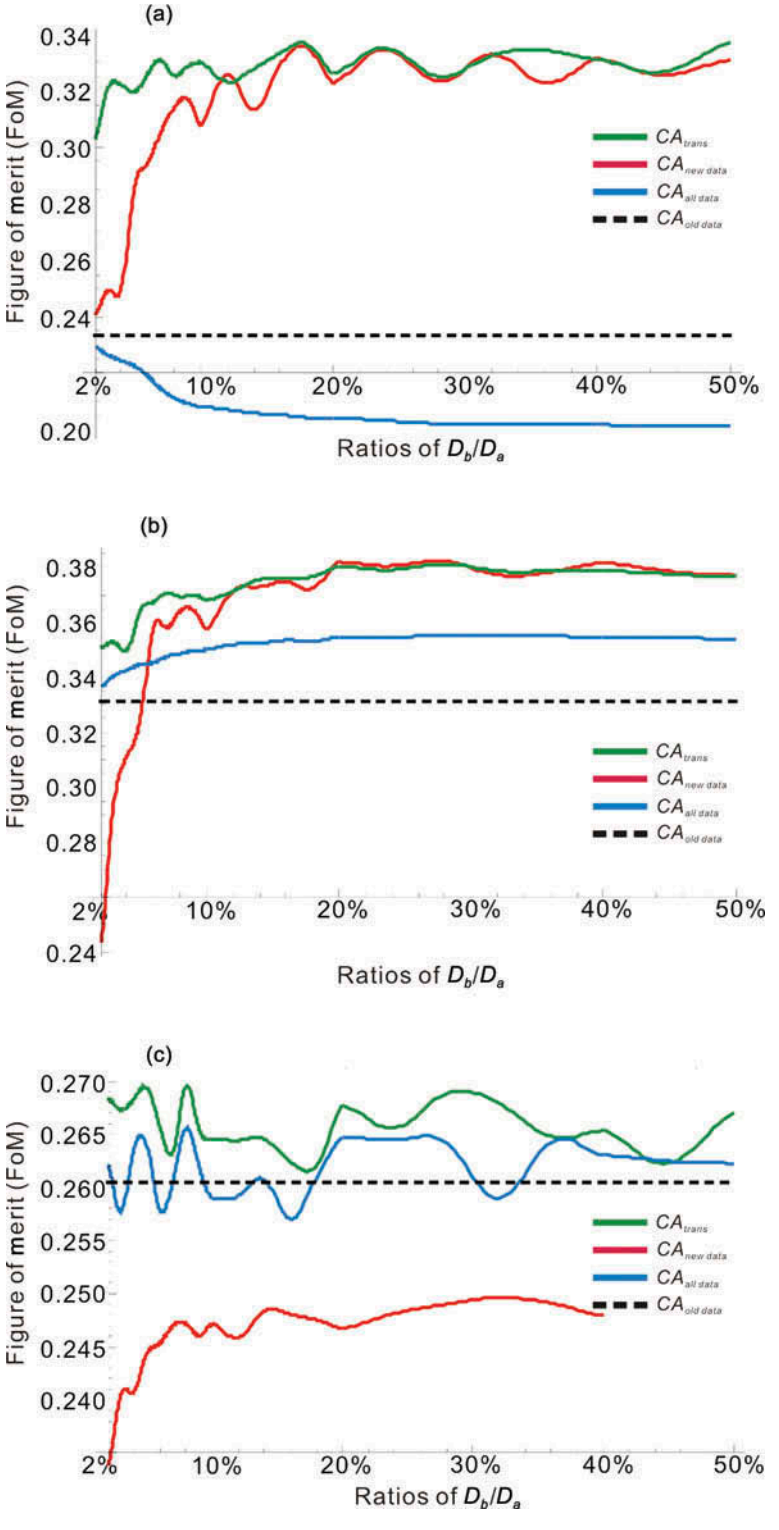

Figure 6. The figure of merit (FoM) of the simulation from the four methods. (a) Experiment 1 (from Guangzhou to Panyu), (b) Experiment 2 (from Guangzhou to Shenzhen), and (c) Experiment 3 (from Guangzhou to Shenzhen).

FoM and the ratio of D_b/D_a of these four models. This figure clearly shows that the proposed method (CA_{trans}), which uses the combination of D_b and D_a data, has a higher FoM value than $CA_{new\ data}$ if the amount of D_b is small. The fewer the new data, the more effects the proposed method will have for improving the simulation accuracy in terms of FoM. If there is a large amount of new data, however, the performance is almost the same as that of $CA_{new\ data}$ (the traditional method of building CA from scratch).

Compared with $CA_{new\ data}$, CA_{trans}can increase the FoM value by 25.96% if there are only 10 samples (Ratio = 2%) of new data. Using 30 samples (Ratio = 6%) of new data, CA_{trans}can produce 98.52% of the FoM value that $CA_{new\ data}$ can do with 250 samples (Ratio = 50%) of new data.

It is interesting to find that the blue curve of $CA_{all\ data}$ (using all these data with an equal weight) decreases with an increase in the D_b/D_a ratio for Experiment 1. This means that by adding more new data to the study area, the model yields even worse results. The explanation is that $CA_{all\ data}$ cannot handle these data well because they have different distributions.

Experiment 2

This experiment is to carry out the knowledge transfer between two places at a greater distance apart (from Guangzhou to Shenzhen). Figure 6b clearly confirms that the proposed method (CA_{trans}) has much better performance than the other three methods if the ratio of D_b/D_a is less than 6%. CA_{trans}can increase the FoM value by 44.14% if there are only 10 samples (Ratio = 2%) of new data, compared with the $CA_{new\ data}$. This proposed method can produce 96.58% of the FoM value that traditional CA ($CA_{new\ data}$) produces with 250 samples (Ratio = 50%) of new data. However, the performance of this proposed model is close to that of the traditional method ($CA_{new\ data}$) if a large amount of new data is available.

Experiment 3

This experiment is to compare the effect of temporal knowledge transfer using different periods of collected data at the same location. The labeled data include the auxiliary data set (D_a) and the base data set (D_b) for Guangzhou during the periods1986–1994 and 2000–2008, respectively.

This experiment also confirms that a tiny set of new data can be used to implement temporal knowledge transfer using the proposed method. Figure 6c shows that CA_{trans} produces the highest simulation accuracy for all the sampling ratios. This method improves the FoM value by 14.82%, compared with $CA_{new\ data}$. The effect is more obvious if the amount of new data is small (Ratio <10%).

These two methods, $CA_{old\ data}$(the back dash line)and $CA_{all\ data}$, also yield almost the same simulation results as CA_{trans}. This means that even the old data can be directly used for producing very good simulation results. The reason is that the growth patterns of this study area have not changed much between 1986–1994 and 2000–2008. This characteristic is the main reason why CA models can have the predictive capability if they are applied to the same study area.

The final simulated patterns of urban development for these regions were obtained using these four models. Figure 7a–d only shows the comparison of the simulated results of CA_{trans} and $CA_{new\ data}$ for Experiment 1 and Experiment 2. These simulated patterns are

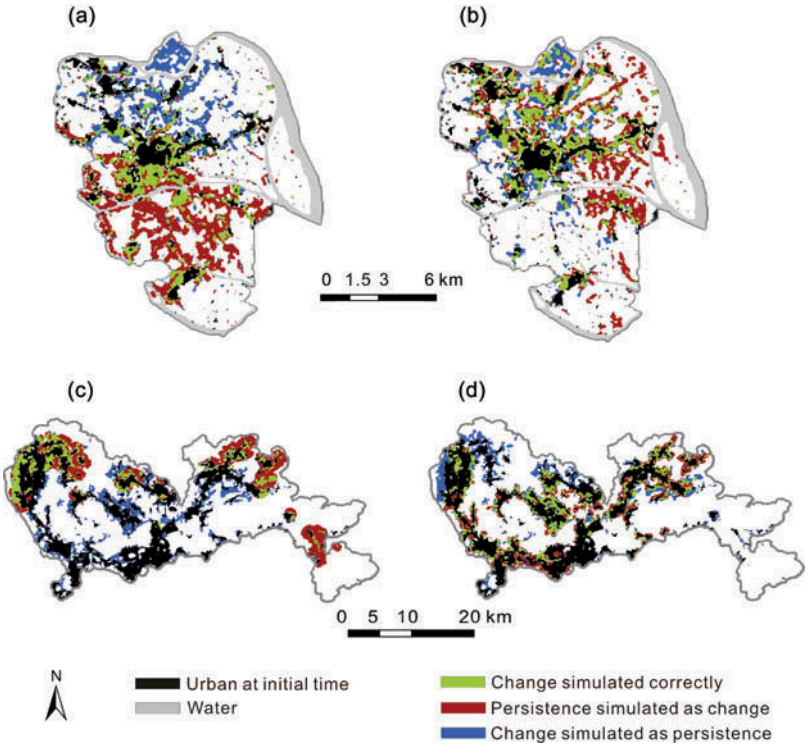

Figure 7. Simulation of urban growth using 10 base training samples: (a) Experiment 1 using $CA_{new\ data}$; (b) Experiment 1 using CA_{trans}; (c) Experiment 2 using $CA_{new\ data}$; and (d) Experiment 2 using CA_{trans}.

compared with the reference map for the subsequent time. The areas in green are the correctly simulated while those in red (persistence simulated as change) and blue (change simulated as persistence) are the falsely simulated. We found that $CA_{new\ data}$ performs much worse than the proposed method with more areas of red and blue (Figure 7a and c). However, a more realistic pattern can be simulated using the proposed method, CA_{trans} (Figure 7b and d). These figures reveal an interesting fact that the falsely simulated patches are usually situated in remoter areas away from urban centers. Future efforts are required to improve the simulation accuracies in these remoter areas.

The above experiments are based on a fixed amount of old data (500 samples) and variable amounts of new data. In most situations, the FoM value stabilizes if the ratio of D_b/D_a is larger than 15%. The dependency of the simulation accuracy on the ratio is just a matter of uncontrolled variation of the data after the threshold. This indicates that the knowledge-transfer method is only useful when new data for calibrating CA are sparse. Experiments 1 and 2 show that the red curves of traditional $CA_{new\ data}$ can catch up very quickly with the green curves of CA_{trans} and yield almost the same result at about 6% (or 35 new samples) of the D_b/D_aratio. This means that the traditional CA can do a very good job with about 30–40 new samples. However, we argue that field investigations for collecting these additional new data may not be easy for a number of reasons, such as labor costs, inexperienced users, and limited knowledge about the study area. Previous studies have also indicated that labeling data is expensive and sample size must be kept

to a minimum (Congalton 1991). Interpolation is often carried out to obtain the attribute values for each spatial element because sampling is expensive and data points are sparse (Heuvelink *et al*. 1989). Our above experiments have shown that knowledge transfer is a good option to solve such a data problem for calibrating simulation models.

It is appealing that the amount of new required data can be reduced using available (old) data as much as possible. A further experiment was designed to examine how these new data can be reduced if old data are available. In this experiment, the amount of old data will be changed to examine this possibility. Figure 8a shows the FoM value using different amounts of new data with respect to 20 and 100 old samples. As mentioned before,

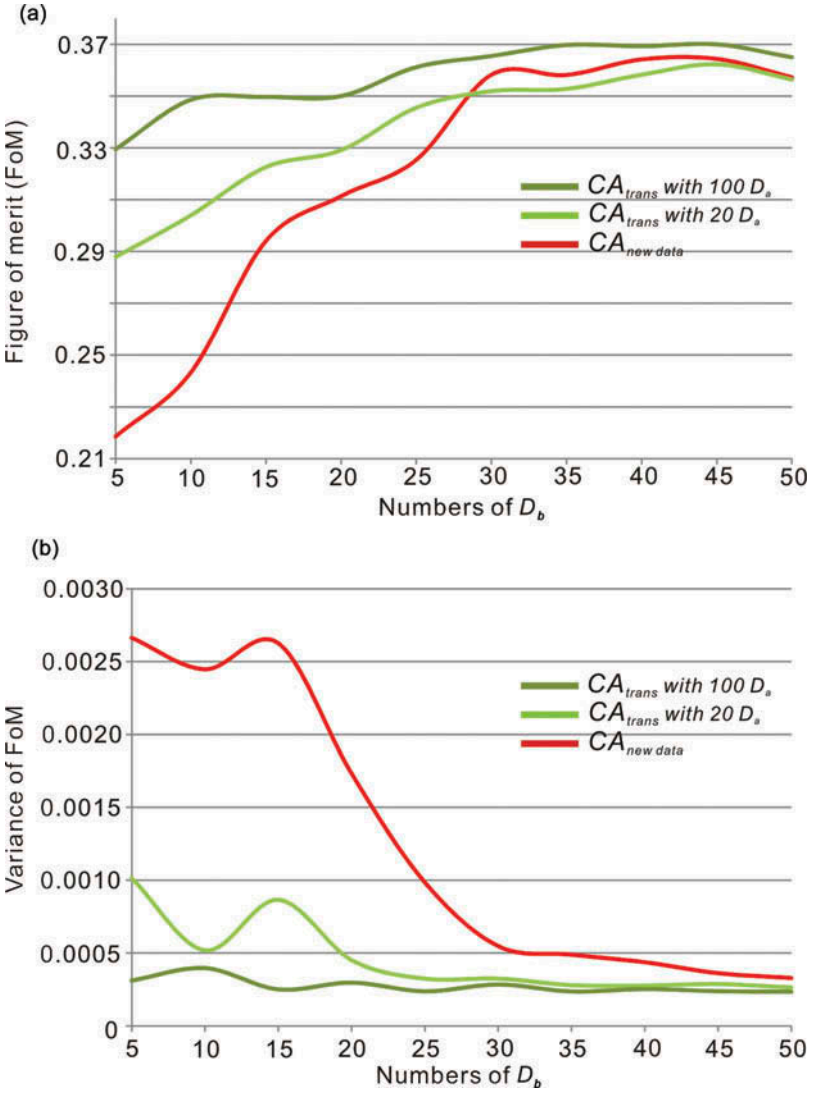

Figure 8. The figure of merit (FoM) and its variances using 20 and 100 old samples. (a) Figure of merit (FoM) (from Guangzhou to Shenzhen) and (b) variances of FoM (from Guangzhou to Shenzhen).

all these accuracies were obtained by randomly drawing the samples 10 times and yielding 10 simulations. Figure 8b is the variance of FoM for the transfer between Guangzhou and Shenzhen. This figure clearly indicates that the variance of traditional CA is large (2–5 times the variance of the proposed method) until the number of new data reaches 30. This means that traditional CA will yield much larger errors than the proposed methodif there are not enough new data. However, the proposed method can obviously reduce the uncertainties even if the amount of new data is small (within 5–30). Moreover, the FoM value of the proposed method is 6–50% higher than that of the traditional method before the threshold (30 new samples). These figures also indicate that 100 old samples will produce better performances than 20 old samples under the same condition. In terms of the FoM value, the use of 25 new samples for the proposed method is almost equivalent to the use of 50 new samples for the traditional method. The efficiency is quite clear because 50% of the labor costs will be saved if field investigations are required to label the data (obtaining land-use classes in the field).

5. Conclusions

Calibration of CA models is often faced with the bottleneck of collecting field information in large complex areas. Serious field investigations rely on experience and knowledge of the study area. Model adaptation is necessary if existing models are reused for solving new simulation problems. Transfer learning techniques can be employed to reduce the costs of building new models.

This study has demonstrated that the knowledge transfer for urban simulation can be based on the integration of logistic-CA and *TrAdaBoost*. A number of experiments were carried out to examine the applicability of this proposed method under different situations. Experiment 1 tested the spatio-knowledge transfer for urban simulation between two close regions. With the increase of accuracy up to 25.96%, this proposed model, CA_{trans}, can produce better simulation results than traditional methods. This experiment utilizes a tiny amount of new samples (e.g., 10–15 samples) and a large amount of old samples. Experiment 2, which is to test knowledge transfer between two more distant regions, also yields much better results for the spatio-knowledge transfer, with the increase of accuracy up to 44.14%.

Our analysis indicates that knowledge transfer can alleviate the problems of uncertainties if there are a few new data for calibrating CA models. Instead, traditional methods have limitations to produce reliable simulation results without enough new data (e.g., less than 15 new samples). However, our proposed method can handle this problem of sparse data well. Generally speaking, this method can save 50% of the labor costs using existing knowledge from the empirical data.

Experiment 3 indicates that the proposed model is able to implement the temporal knowledge transfer of transition rules. However, itdoesnothave obvious advantages over traditional methods. The analysis shows that the performances of $CA_{old\ data}$ and $CA_{all\ data}$, which use old data for calibrating model directly,are as good as the performances of the proposed model. This confirms the fact that CA models can be used to simulate future land-use dynamics if the historical trend of urban dynamics continues.

We find that more simulation errors happen in remoter areas away from urban centers. Stronger reinforced calibration will be taken in these areas to improve simulation accuracies in our future studies. There is also a need to consider the implementation of knowledge transfer using multi-sources of previously collected data. It would also be interesting to develop the methods of knowledge transfer for constructing other bottom-up models, such

as agent-based models. Moreover, the simulation model can incorporate other exogenous forces, such as land-use policy, which will affect land-use dynamics. In the future studies, we also need to examine how the effect of spatio-knowledge transfer is related to the similarity between two cities, and we must consider the measurement of similarity. Perhaps, some indicators can be developed so that urban structures and growth patterns can be quantified to give users the information as to whether or not the knowledge transfer is appropriate.

Acknowledgments

This study was supported by the National Basic Research Program of China (973 Program) (Grant No. 2011CB707103) and National Natural Science Foundation of China (Grant No. 41171308).

References

Batty, M. and Xie, Y., 1994. From cells to cities. *Environment and Planning B*, 21, 31.

Chavez, P.S., Jr, 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote sensing of environment*, 24(3), 459–479.

Chen, Y.M., *et al*., 2011. Estimating the relationship between urban forms and energy consumption: a case study in the Pearl River Delta, 2005–2008. *Landscape and Urban Planning*, 102(1), 33–42.

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*, 24, 247–262.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1), 35–46.

Couclelis, H., 1988. Of mice and men: what rodent populations can teach us about complex spatial dynamics. *Environment and Planning A*, 20(1), 99–109.

Dai, W., *et al*., 2007. Boosting for transfer learning. In: *Proceedings of the 24th International Conference on Machine Learning*.Corvallis, OR: ACM, 193–200.

Fonseca, F.T., *et al*., 2000. Ontologies and knowledge sharing in urban GIS. *Computers, Environment and Urban Systems*, 24(3), 251–272.

Fonseca, F.T. and Egenhofer, M.J.,1999. Ontology-driven geographic information systems. In: *7th ACM Symposium on Advances in Geographic Information Systems*. Kansas City, MO: ACM, 14–19.

Freund, Y. and Schapire, R., 1995. A desicion-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 23–37.

Friedman, J., Hastie, T., and Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. *Annals of statistics*, 28, 337–374.

Heuvelink, G.B.M., Peter, A.B., and Stein, A., 1989. Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information System*, 3(4), 303–322.

Li, X., 2011. Emergence of bottom-up models as a tool for landscape simulation and planning. *Landscape and Urban Planning*, 100(4), 393–395.

Li, X., Yang, Q., and Liu, X., 2008. Discovering and evaluating urban signatures for simulating compact development using cellular automata. *Landscape and Urban Planning*, 86(2), 177–186.

Li, X., *et al*., 2011. Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. *International Journal of Geographical Information Science*, 25(4), 633–655.

Pontius, G.R. and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429.

Pontius, G.R., *et al*., 2008. Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42(1), 11–37.

Rabiner, L. and Juang, B., 1986. An introduction to hidden Markov models. *IEEEASSP Magazine*, 3(1), 4–16.

Rajan, S., Ghosh, J., and Crawford, M.M., 2008. An active learning approach to hyperspectral data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4), 1231–1242.

Santé, I., *et al*., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape and Urban Planning*, 96(2), 108–122.
Schapire, R.E., 1990. The strength of weak learnability. *Machine learning*, 5(2), 197–227.
Schapire, R.E., *et al*., 1998. Boosting the margin: a new explanation for the effectiveness of voting methods. *The annals of statistics*, 26(5), 1651–1686.
Schapire, R.E., 2001. *The boosting approach to machine learning: an overview*. Lecture Notesin Statistics. New York: Springer Verlag, 149–172.
Schmidhuber, J.,1995. *On learning how to learn learning strategies*. Technical Report FKI-198-94. Munich, Germany: Fakultät für Universität München.
Seto, K.C., *et al*., 2002. Monitoring land-use change in the Pearl River Delta using Landsat TM. *International Journal of Remote Sensing*, 23(10), 1985–2004.
Seto, K.C., Kaufmann, R.K., and Woodcock, C.E., 2000. Landsat reveals China's farmland reserves, but they're vanishing fast. *Nature*, 406(6792), 121–121.
Soares-Filho, B.S., Coutinho Cerqueira, G., and Lopes Pennachin, C., 2002. A stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154(3), 217–235.
Toffoli, T. and Margolus, N., 1987. *Cellular automata machines: a new environment for modeling*. Cambridge, USA: The MIT Press.
Verburg, P.H., *et al*., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management*, 30(3), 391–405.
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, 1175–1175.
Wolfram, S., 1986. *Theory and applications of cellular automata.* Advanced Series on Complex Systems. Singapore: World Scientific Publication.
Wolfram, S., 2006. *Cellular automata and complexity*. Boston, USA: Addison-Wesley.
Wu, F., 1998. An experiment on the generic polycentricity of urban growth in a cellular automatic city. *Environment and Planning B: Planning and Design*, 25(5), 731–752.
Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science*, 16(8), 795–818.