

Simulating Urban Form and Energy Consumption in the Pearl River Delta Under Different Development Strategies

Yimin Chen,* Xia Li,* Shujie Wang,* Xiaoping Liu,* and Bin Ai[†]

*School of Geography and Planning and Guangdong Key Laboratory for Urbanization and Geo-Simulation, Sun Yat-sen University, Guangzhou, China

[†]School of Marine Sciences and Guangdong Key Laboratory for Urbanization and Geo-Simulation, Sun Yat-sen University, Guangzhou, China

China is currently the world's largest energy consumer. The rapid growth in energy consumption has resulted in many problems in this country. The Chinese government has realized the necessity of improving energy efficiency and reducing energy consumption. As a useful decision-support tool, simulation models can be used to examine the potential impacts of different plans on urban development and energy consumption. This study presents a model that integrates support vector regression (SVR) and cellular automata (CA) to simulate urban forms and to estimate the corresponding energy consumption in one of the most developed regions in China, the Pearl River Delta (PRD). SVR is used to predict energy consumption and to project future urban size. The logistic CA model simulates different urban forms to evaluate their effects on energy consumption. In this study, we simulated four scenarios to assess the impacts of different development strategies on urban forms and the related energy consumption. For each scenario, we used the model to predict land demand and energy consumption. The result indicates that land demand is more sensitive to changes of economic structure than is energy consumption. The comparison of different simulated scenarios suggests that promoting low-energy-consuming industries is the most effective strategy to balance economic development and energy and land consumption. **Key Words:** *cellular automata, energy consumption, support vector regression, urban forms.*

中国是当今全世界最大的能源消耗国, 快速成长的能源消耗已导致了国内诸多问题, 而中国政府亦体认到改善能源效率和减少能源消耗的必要性。模拟模型做为有用的决策支持工具, 可以用来检视不同的计划对城市发展和能源消耗的潜在影响。本研究将引介一个整合支持向量回归 (SVR) 与细胞自动机器 (CA) 的模型, 用来模拟中国最发达的区域之一珠江三角洲 (PRD) 中的城市形态并评估相应的能源消耗。支持向量回归用来预测能源消耗并推估未来城市大小。逻辑回归CA模型则模拟不同的城市形态来评估其对能源消耗的影响。我们在本研究中模拟四个不同的情境以取得不同的发展策略对城市形态与相应之能源消耗的影响。我们在每个情境中运用模型来预测土地需求与能源消耗。研究结果显示, 土地需求对经济结构的改变较能源消耗敏感。不同模拟情境的比较指出, 推广低耗能工业是平衡经济发展和能源消耗与土地利用的最有效策略。 **关键词:** 细胞自动机, 能源消耗, 支持向量回归, 城市形态。

Actualmente China es el más grande consumidor de energía del mundo. El rápido crecimiento en consumo de energía ha dado lugar a muchos problemas en este país. El gobierno chino se ha percatado de la necesidad de mejorar la eficiencia energética y reducir el consumo de energía. Como útil herramienta en la que apoyar decisiones sobre el particular, se pueden utilizar modelos de simulación para examinar los impactos potenciales que puedan derivarse de diferentes planes sobre desarrollo urbano y consumo de energía. Este estudio introduce un modelo que integra la regresión de vectores de sostén (SVR) y el autómata celular (CA) para simular formas urbanas y calcular el correspondiente consumo de energía en una de las regiones más desarrolladas de China, el Delta del Río Perla (DRP). La regresión SVR se utiliza para para pronosticar el consumo de energía y para proyectar el tamaño urbano futuro. El modelo logístico CA simula diferentes formas urbanas para evaluar sus efectos sobre el consumo de energía. En este estudio simulamos cuatro escenarios para estimar los impactos de diferentes estrategias del desarrollo de formas urbanas y el consumo energético relacionado. Para cada escenario utilizamos el modelo para pronosticar la demanda de tierra y de consumo de energía. El resultado indica que la demanda de tierra es más sensible a los cambios de la estructura económica que el consumo de energía. La comparación de los diferentes escenarios simulados sugiere que promover industrias de bajo consumo energético

es la estrategia más efectiva para equilibrar el desarrollo económico con los consumos de energía y tierra. *Palabras clave:* autómata celular, consumo de energía, regresión de vectores de sostén, formas urbanas.

Just about ten years ago, China's energy consumption was half that of the United States. It took only a few years for China to catch up and became the world's largest energy user (Zellner et al. 2008). This rapid growth of energy consumption led to many environmental and social problems. For instance, the intensive and inefficient use of coal gave rise to serious environmental problems, such as acid rain, air pollution (Fang, Chan, and Yao 2009), and increased carbon emissions (Dhakal 2009). In addition, the insufficient domestic energy supply has forced China to rely on fuel imports, which have raised concerns about energy security (Crompton and Wu 2005). Most of China's energy is consumed by highly urbanized areas. According to Dhakal (2009), urban use accounted for up to 84 percent of China's energy consumption in year 2006, and its thirty-five largest cities, which share 18 percent of the country's population, contributed 40 percent of the total energy use.

Rapid economic growth and urbanization have stimulated the energy consumption of China (Dhakal 2009). Economic growth can directly increase energy demand, especially because the country's economy largely relies on industries with relatively low energy efficiency (Fang, Chan, and Yao 2009). In addition, the fast urbanization has given rise to increases in passenger and freight transportation demands. This has significant impacts on energy consumption and air pollution (Yang et al. 2011). Population growth is another driver of increased energy consumption. Energy demand rose partly due to the needs of the expanding population for daily living, commuting, and many other activities. Additionally, the growth of personal wealth encourages people to change to more energy-intensive lifestyles, such as the ownership of automobiles (Dieleman, Dijst, and Burghouwt 2002).

Because over 70 percent of China's total energy is used for industrial production (Energy Information Association 2009), China's 12th Five-Year Plan urges improved energy efficiency and reduced energy consumption resulting from changes in economic structure (State Council of the People's Republic of China 2011). In this context, we selected one of the most developed regions in China, the Pearl River Delta (PRD), as the study area in which to explore the potential effects of the change in economic structure on urban growth and energy consumption. As an emerging megalopolis, the PRD is a major economic

region and manufacturing base of the world. The early development in this region was mainly grounded on foreign investments and low labor and land costs. The processing technologies of manufacturing, featuring low energy efficiency, were dominant in the region (Fang, Chan, and Yao 2009) and brought about many environmental problems. For example, the ambient concentrations of SO₂ and NO₂ in the PRD have been higher than those in other parts of the province (Shao et al. 2006). The provincial government has been considering implementation of long-term measures to reduce energy consumption and improve the environment, including adjusting the economic structure and upgrading the processing technologies and equipment (Fang, Chan, and Yao 2009).

In addition, the control of urban form should be considered as a means to reduce energy consumption. Many studies have revealed that urban form plays a crucial role in urban energy consumption (Anderson 1996; Camagni, Gibelli, and Rigamonti 2002; Holden and Norland 2005; Y. Chen et al. 2011). Urban form refers to the spatial configuration of urban land use within a metropolitan area. The relationship between urban form and energy consumption can be complex. Some researchers believe that compact development can effectively reduce energy consumption (Jenks and Burgess 2000). The main idea of compact development is to promote high-density development with mixed land use types that favor the efficient use of facilities, the reduction of travel, and the development of public transit, hence achieving low energy consumption (Holden and Norland 2005). With regard to a dispersed urban development, it cannot support public transit because of the scattered demand and destinations, so dependence on automobiles increases due to both the door-to-door convenience and the declining costs of car ownership (Camagni, Gibelli, and Rigamonti 2002).

The built environment can affect households' traveling behaviors, which are related to energy consumption. In North America, residents prefer public transit or walking in high-density employment centers, as these areas usually have concentrated transit hubs (C. Chen, Gong, and Paaswell 2008). Residents were also less likely to own vehicles and tended to use transit more in high-density residential neighborhoods due to the traffic congestion and limited parking (Badoe and Miller 2000; Ewing and Cervero 2001; National Academy of Sciences of the United States 2009). In China, land and

housing reforms have broken the workplace–residence tie in the prereform urbanized areas since the 1980s, resulting in increasing spatial separation between workplace and residence. Highways were then intensively constructed by local governments to improve accessibility, and automobile use was encouraged (Yang and Gakenheimer 2007). Such rapid motorization, along with the lengthened trips, generated serious traffic problems in Chinese cities. Fortunately, there is evidence showing that residents tend to use transit, walk, or bicycle if public transit services are available or if the environments are pedestrian and cycling friendly (Pan, Shen, and Zhang 2009).

To identify the relationships between urban form and energy consumption, one needs quantitative measurements of urban form. Population density is a frequently used measurement (Mindali, Raveh, and Salomon 2004), but it cannot reflect the spatial characteristics of urban form. In this study, we used landscape metrics to measure urban form, as they can (1) improve the representation of heterogeneous urban landscapes, (2) bridge the gap between urban land use patterns and governing processes, and (3) facilitate the analysis of impacts of urban development on the surrounding environment (Herold, Couclelis, and Clarke 2005). A recent study demonstrated that the fragmentation and irregularity of urban land use patterns, measured by landscape metrics, were positively correlated with urban energy consumption (Y. Chen et al. 2011).

In this study, we propose a model that integrates support vector regression (SVR) and cellular automata (CA) to simulate the urban forms and to estimate the corresponding energy consumption levels. The simulation should be useful for exploring the impacts of different development strategies on urban growth and energy consumption. CA are a bottom-up approach (Wolfram 1984) that can generate global urban land use patterns through modeling local interactions

between geographical features and their immediate neighborhoods (environments). Previous studies demonstrated the strength of CA in simulating realistic urban growth (Clarke and Gaydos 1998; Silva and Clarke 2005) and in solving urban planning problems when coupled with spatial optimization models (Li, Chen, et al. 2011; Li, Shi, et al. 2011).

The scenario simulations of urban forms will obtain various spatial variables in terms of landscape metrics. Based on these metrics, we adopted SVR (Smola and Schölkopf 2004) to predict the corresponding energy consumption for each scenario. SVR is a new technique of classification and prediction and has been used to handle complex relationships in many fields (Oliveira 2006; Hua et al. 2007). The method employs the structural risk minimization (SRM) principle to minimize the upper bound of the generalization error instead of the error from the training set. Compared with conventional methods, SVR improves prediction accuracy through avoiding overfitting. Additionally, SVR can overcome the difficulties resulting from the normality assumption of the distribution process and the correlation among predictors (Gani, Taleb, and Limam 2010).

Method

In this study, an SVR-based model of energy consumption is integrated with a CA-based model of urban growth for two reasons. First, the impact of urban growth on energy consumption is immediate. Energy demand is expected to increase as the urban economy grows, if energy efficiency has not improved significantly. Second, the change in economic structure is likely to influence urban growth, as well as energy consumption. The overall energy demand varies among different economic structures because energy use intensity is diverse from one type of industry to another. Meanwhile, the

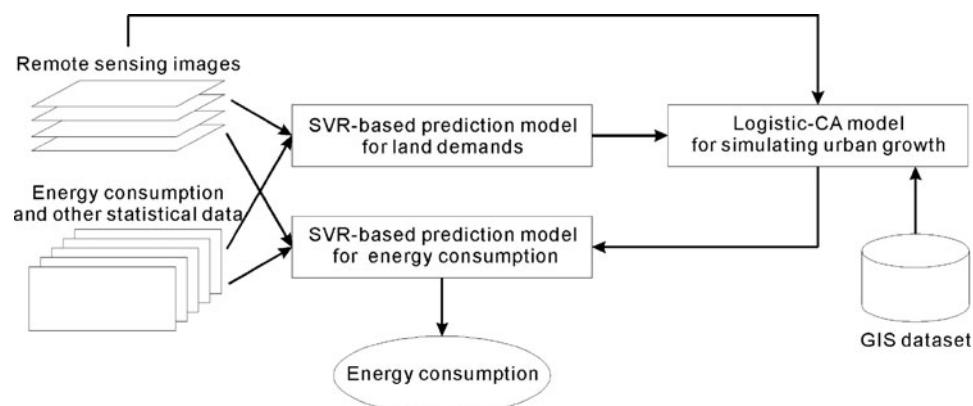


Figure 1. An integrated model to evaluate the impacts of different development strategies on urban growth and energy consumption.

urbanization process will deviate from its regular path as a result of a changed economic structure.

Figure 1 illustrates the flow of the proposed model. In this model, CA was used to simulate urban forms with the urban size constraint, which was produced by an SVR model based on a set of socioeconomic variables. After the simulation, the landscape metrics were then calculated to quantify the simulated urban forms. Another SVR model was finally applied to the prediction of energy consumption using both landscape metrics and other socioeconomic variables. Details of SVR and CA are provided in the following sections.

Support Vector Regression

An important part of this coupling model is to use SVR to predict urban size and energy consumption. In SVR, the objective is to find a function $f(x) = \langle w, x \rangle + b$ that best fits the training data set, where w is the weight vector, b is the threshold ($w \in \chi, b \in R$), and $\langle *, * \rangle$ is the dot product. An ε -insensitive loss function is further defined, where ε is the parameter representing the band of the tube around the regression function, as shown in Figure 2. Errors less than ε (inside the tube) are ignored, whereas errors larger than ε are depicted using slack variables ξ and ξ^* (Figure 2). Then, the optimization objective can be formulated (Smola and Schölkopf 2004) as:

$$\text{minimize: } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \quad (1)$$

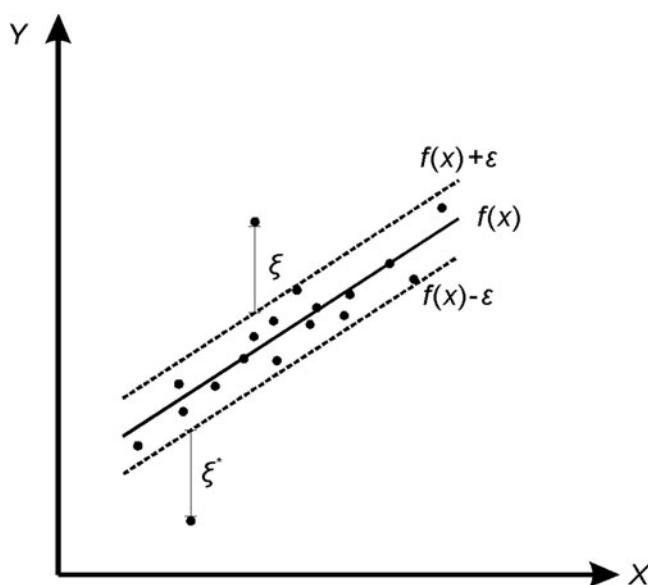


Figure 2. The ε -insensitive loss function in support vector regression.

subject to:

$$\begin{cases} (\langle w, x_i \rangle + b) - y_i \leq \varepsilon + \xi_i, \\ (y_i - \langle w, x_i \rangle - b) \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0. \end{cases} \quad (2)$$

where C is a positive constant, representing the trade-off between the flatness of f and the errors. The minimization of Equation 1 is based on the Lagrange function:

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & - \sum_{i=1}^l \alpha_i [\varepsilon + \xi_i + \langle w, x_i \rangle + b - y_i] \\ & - \sum_{i=1}^l \alpha_i^* [\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b] \\ & - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (3)$$

where $\alpha_i, \alpha_i^*, \eta_i$, and η_i^* are Lagrange multipliers. The partial derivatives of L , with respect to w, b, ξ_i , and ξ_i^* , are derived:

$$\partial_b L = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (4)$$

$$\partial_w L = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \quad (5)$$

$$\partial_{\xi_i} L = C - \alpha_i - \eta_i = 0 \quad (6)$$

$$\partial_{\xi_i^*} L = C - \alpha_i^* - \eta_i^* = 0 \quad (7)$$

Substituting these equations into Equation 3, the minimization problem becomes:

$$\begin{aligned} \text{minimize: } & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ & - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (8)$$

subject to:

$$\begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \quad (9)$$

Based on Equation 4, the $f(x)$ can be reformulated as:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) < x_i, x > + b \quad (10)$$

The threshold b is calculated using the Karush–Kuhn–Tucker conditions:

$$\alpha_i [\varepsilon + \xi_i - y_i + < w, x_i > + b] = 0 \quad (11)$$

$$\alpha_i^* [\varepsilon + \xi_i^* + y_i - < w, x_i > - b] = 0 \quad (12)$$

$$(C - \alpha_i) \xi_i = 0 \quad (13)$$

$$(C - \alpha_i^*) \xi_i^* = 0 \quad (14)$$

Therefore, the threshold b is obtained as:

$$b = y_i - < w, x_i > - \varepsilon \text{ for } \alpha_i \in (0, C) \quad (15)$$

$$b = y_i - < w, x_i > + \varepsilon \text{ for } \alpha_i^* \in (0, C) \quad (16)$$

where x_i represents the data points inside the tube, with errors that are less than ε . According to the definition of ε -insensitive loss function, the Lagrange coefficients of those data points inside the tube are zero. Those data points with nonzero coefficients are called *support vectors*. Furthermore, the optimization process described earlier can be alternatively accomplished through the kernel function $K(x_i, x_j)$:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (17)$$

The forms of kernel functions include polynomial, sigmoidal, and radial-basis functions. Details of the solution can be found in Smola and Schölkopf (2004). In this study, SVR is implemented through the machine learning software WEKA (Frank et al. 2010). The performance of SVR is assessed in terms of prediction accuracy. This can be measured by using the mean relative error (MRE), a common measurement in many applications of SVR (Oliveira 2006):

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - Y'_i|}{Y_i} \quad (18)$$

where Y_i and Y'_i are the i th observation and its estimate, respectively.

Logistic CA

In this study, the CA model was developed largely based on Wu's (2002) method but was enhanced by incorporating the urban size projected by SVR based on a set of economic variables. The CA model was formulated in a logistic form. Specifically, in a two-dimensional latticed space of this CA model, the probability of $cell_{ij}$ to be developed was estimated through a function of development factors (x_1, x_2, \dots, x_n), such as the proximity to town centers or major roads. A logistic function is used to represent the development probability:

$$p_{g,ij} = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)} \quad (19)$$

where z is the combination score of development factors of $cell_{ij}$:

$$z = b_0 + \sum_k b_k x_k \quad (20)$$

where b_0 is a constant; b_k are the coefficients of the development factors, which can be calibrated using logistic regression; and x_k is the development factors of $cell_{ij}$.

The probability $p_{g,ij}$ only addresses the influences of static physical factors, however. The actual urban development would also be subject to the influences of dynamic factors. In the CA model, they are represented in a way of neighborhood effect, denoted as Ω_{ij}^t . A simple way to obtain Ω_{ij}^t is to calculate the development density in an $n \times n$ neighborhood of $cell_{ij}$ at time t :

$$\Omega_{ij}^t = \frac{\sum_{n^2} \text{con}(s_{ij} = \text{developed})}{n^2 - 1} \quad (21)$$

where $\text{con}()$ is a conditional function that returns true if the state of a cell within the neighborhood is currently developed. Physical constraints can be incorporated into the function. For instance, if a cell belongs to a water body, mountain, or restricted areas, the cell should be excluded from development. Therefore, the development probability is revised as follows:

$$p_{c,ij}^t = p_{g,ij} \Omega_{ij}^t \text{con}(s_{ij} = \text{suitable}) \quad (22)$$



Figure 3. Location of the Pearl River Delta. (Color figure available online.)

where $\text{con}()$ is a conditional function that returns true if cell_{ij} is suitable for development. A nonlinear transformation is imposed to $p_{c,ij}^t$ to promote the probability of development in cells with higher evaluation scores:

$$p_{t,ij}^t = p_{c,ij}^t \exp \left[-\delta \left(1 - p_{c,ij}^t / p_{c,\max}^t \right) \right] \quad (23)$$

where $p_{c,\max}^t$ is the maximum value of $p_{c,ij}^t$ in space at time t ; and δ is called the dispersion parameter, ranging from 1 to 10. During the simulation process, the number of cells selected for development should meet the projected amount of urban growth. Therefore, $p_{t,ij}^t$ is further scaled as:

$$p_{s,ij}^t = q p_{t,ij}^t / \sum p_{t,i'j'}^t \quad (24)$$

where q is the expected number of cells to be converted, which can be determined through the iteration number and the projected urban size. The SVR model was used to estimate the urban size based on a set of economic variables.

The selection of cells for development is based on the Monte Carlo approach. First, the development probability for each cell in space is updated according to Equations 22, 23, and 24. Then, a cell is randomly picked and its scaled development probability $p_{s,ij}^t$ is compared with a random value γ , within 0 to 1. If $p_{s,ij}^t$ is greater than γ , the cell is converted to urban land use. Otherwise, it remains unchanged.

Implementation and Results

Study Area

The study area is located in the PRD, Guangdong Province, China (Figure 3). The five most economically important cities of this region—namely, Guangzhou, Shenzhen, Foshan, Dongguan, and Zhongshan—were selected for this study. In 1978, the economic reform of China triggered a boom in the regional economy, as well as rapid urbanization. At present, the PRD has the highest per capita gross domestic product (GDP) among the several most developed regions in China (Shao et al. 2006), but it requires a vast volume of natural resources, especially fuel resources, to sustain its economic growth. According to the Guangdong Statistical Yearbook (Statistics Bureau of Guangdong Province 2009), the energy consumption of the entire province reached 226.72 million tons of the standard coal equivalent (SCE) in 2008. The energy consumption of the industrial sector accounted for 65.68 percent of overall energy use, more than 70 percent of which was generated from coal and oil. As the most urbanized and economically developed region in Guangdong Province, the PRD accounted for 67 and 85 percent of the total provincial consumption of coal and oil, respectively (Shao et al. 2006).

Data

The primary source of the energy consumption data used in this study is the statistical yearbooks published

by the city governments. Most of these data are not available before 2005. As a result, we only collected the energy consumption data of the study area from 2005 to 2008. The total amounts of energy consumption of Guangzhou, Dongguan, and Zhongshan were found in the statistical yearbooks, but those of Foshan and Shenzhen were not available. The energy consumptions of these two cities were estimated using the following equation:

$$E = e_{GDP,i} V_i + e_{Living,i} P_i / 1,000 \quad (25)$$

where $e_{GDP,i}$ is the energy intensity (ton of SCE/ 10^4 yuan) of city i , and V_i represents the amount of GDP (10^4 yuan) of city i ; $e_{Living,i}$ is per capita energy consumption for living (kilograms of SCE) of city i , and P_i is the city's population. This method is different from Dhakal's (2009) approach, which uses statistics beyond the city level (e.g., provincial energy intensity), to estimate energy consumptions of China's cities. This is a top-down approach, which disaggregates the energy consumption at a higher level to a lower level. In this study, because the energy intensity data were available at the city scale, no disaggregation was needed when estimating energy consumption. Table 1 shows the energy intensity (energy consumption per unit GDP) for all five cities from 2005 to 2008, as retrieved from statistical yearbooks of these cities. To validate Equation 25, we used it to estimate the energy consumption of Dongguan, Guangzhou, and Zhongshan and compared the results with the values recorded in the statistical yearbooks of these three cities (Table 1). For the three cities, the differences between the estimates and recorded val-

Table 2. Percentage of gross products of industry and energy consumption for the five selected cities in 2008

	Percentage of gross products of industry (%)	Percentage of energy consumption for industrial production (%)
Dongguan	46.3	56.1
Foshan	60.8	53.3
Guangzhou	34.1	54.7
Shenzhen	43.8	55.2
Zhongshan	54.9	42.0

Sources: Statistics Bureau of Dongguan Municipality (2009); Statistics Bureau of Foshan Municipality (2009); Statistics Bureau of Guangdong Province (2009); Statistics Bureau of Guangzhou Municipality (2009); Statistics Bureau of Shenzhen Municipality (2009); Statistics Bureau of Zhongshan Municipality (2009).

ues are insignificant, which validates the use of Equation 25 to estimate the energy consumption of Foshan and Shenzhen.

For Dongguan, Foshan, Shenzhen, and Zhongshan, industrial production has been the most important economic sector, accounting for 40 to 60 percent of the cities' GDPs (Table 2). Guangzhou is an exception, with a percentage of 34.1 percent. Although industrial production was generally the largest source of energy consumption in the five cities, the subsectors of industry can be quite different in terms of energy usage. To classify the subsectors based on their energy consumption, we calculated the energy intensity for each of the thirty-nine subsectors identified in China's statistical system for the five cities. Due to the data limitations, we had to assume that the energy intensity of a certain subsector was the same for all five cities and use the provincial

Table 1. Energy consumption per unit GDP and total energy consumption of the five selected cities during 2005 to 2008 (estimated total energy consumption is shown in parentheses)

	2005	2006	2007	2008
Energy consumption per unit GDP (tons of SCE/ 10^4 yuan)				
Dongguan	0.86	0.82	0.78	0.74
Foshan	0.95	0.91	0.87	0.80
Guangzhou	0.78	0.75	0.71	0.68
Shenzhen	0.59	0.58	0.56	0.54
Zhongshan	0.78	0.74	0.70	0.67
Total energy consumption (10^6 tons of SCE)				
Dongguan	18.86 (18.76)	21.39 (21.38)	23.94 (23.92)	25.90 (25.88)
Foshan	— (22.64)	— (25.94)	— (29.53)	— (31.32)
Guangzhou	40.29 (40.20)	44.13 (44.15)	48.66 (48.49)	52.25 (51.93)
Shenzhen	— (29.21)	— (33.24)	— (37.11)	— (40.43)
Zhongshan	6.88 (6.86)	7.61 (7.62)	8.35 (8.33)	8.85 (8.84)

Note: GDP = gross domestic product; SCE = standard coal equivalent.

Table 3. The fraction of gross products of each industrial sector in terms of energy intensity in 2008 (%)

	Intensive energy consumption industries	Medium energy consumption industries	Low energy consumption industries
Dongguan	23.98	29.73	46.28
Foshan	37.76	26.52	35.72
Guangzhou	15.28	41.17	43.55
Shenzhen	9.10	16.98	73.92
Zhongshan	15.50	40.31	44.20

data to conduct the calculation. Based on the energy intensity values, we classify the subsectors into three groups: intensive energy-consuming sector (average = 2.35 tons of SCE/ 10^4 yuan), medium energy-consuming sector (average = 0.68 tons of SCE/ 10^4 yuan), and low energy-consuming sector (average = 0.32 tons of SCE/ 10^4 yuan). Table 3 lists the proportions of the industrial products of these subsectors in 2008 for each city. The intensive energy-consuming sector shares the largest proportion in Foshan (37.76 percent), whereas the low energy-consuming sector is the highest in Shenzhen (73.92 percent). We also collected other statistical data (Table 4), particularly the data on the gross products and energy consumption of tertiary industry of the five selected cities, as the tertiary industry became more important in PRD's economy.

In this study, we used multitemporal satellite images to generate urban land use data. These satellite data include four pairs of Landsat TM5 images (path 122, row 44; path 121, row 44) acquired in 2005, 2006, 2007, and 2008, with a resolution of 30 m. These images were georeferenced to the universal transverse mer-

Table 4. All statistical data used in this study

Data	Period
Total energy consumption at city level	2005–2008
Population of each city	2000–2008
Gross domestic products of each city	2000–2008
Gross products of industry at city level	2000–2008
Gross products of each industrial sector at both provincial and city level	2005–2008
Energy consumption of each industrial sector at provincial level	2005–2008
Gross products of tertiary industry at city level	2000–2008

Sources: Statistics Bureau of Dongguan Municipality (2009); Statistics Bureau of Foshan Municipality (2009); Statistics Bureau of Guangdong Province (2009); Statistics Bureau of Guangzhou Municipality (2009); Statistics Bureau of Shenzhen Municipality (2009); Statistics Bureau of Zhongshan Municipality (2009).

cator projection with a registration error of less than 15 m. The land use classification for these images was carried out using the object-oriented classification software, Definiens Developer 7.0 (Definiens Developer 7.0 2003). First, similar pixels were aggregated into objects via the image segmentation approach. Samples (objects) were then manually collected for each land use category (e.g., urban area, farmland, forest, water, fish pond, and bare soil). A set of features was selected through the feature selection tool to maximize the distance between land use classes. Finally, all objects were classified using the nearest neighbor method. The land use classes of farmland, fish pond, and bare soil were merged into nonurban area, regarded as candidates for land conversion during the urban growth simulation, whereas the forests and water areas were considered as restricted areas in which development was not permitted.

We used the method proposed by Pontius and Millones (2011), instead of the frequently used kappa indexes, to assess the classification accuracy. This method divides the disagreements between classification and reference into quantity disagreement and allocation disagreement. Therefore, this method is more helpful than kappa indexes just using a single ratio to represent the classification accuracy. The quantity disagreement and allocation disagreement can be calculated using the following equations (Pontius and Millones 2011):

$$p_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^J n_{ij}} \right) \left(\frac{N_i}{N_{ij}} \right) \quad (26)$$

$$Q = \frac{1}{2} \sum_{g=1}^J q_g = \frac{1}{2} \sum_{g=1}^J \left| \left(\sum_{i=1}^J p_{ig} \right) - \left(\sum_{j=1}^J p_{gj} \right) \right| \quad (27)$$

$$A = \frac{1}{2} \sum_{g=1}^J a_g = \frac{1}{2} \sum_{g=1}^J 2 \min \left[\left(\sum_{i=1}^J p_{ig} \right) - p_{gg}, \left(\sum_{j=1}^J p_{gj} \right) - p_{gg} \right] \quad (28)$$

$$D = Q + A \quad (29)$$

where J is the number of land use classes; n_{ij} is the number of the samples classified as i and referenced as j ; N_i is the population of land use class i ; p_{ij} is the estimated proportion of the study area classified as i and referenced as j ; q_g and a_g are the quantity disagreement

Table 5. Confusion matrices of the classification of urban areas (2005–2008)

		2005		2006		2007		2008	
	Reference	Urban	Nonurban	Urban	Nonurban	Urban	Nonurban	Urban	Nonurban
Classification	Urban	397	73	447	71	487	80	520	102
	Nonurban	142	2,229	158	2,155	182	2,082	149	2,060

and the allocation disagreement of land use class g ; Q and A are the overall quantity disagreement and the allocation disagreement, respectively; and D is the total disagreement.

We calculated the quantity and allocation disagreements for the binary land use data (urban and nonurban) from 2005 to 2008. Each year, we conducted random sampling and collected the samples' reference information. Table 5 shows the confusion matrices for each year's binary land use data, and Figure 4 demonstrates the quantity and allocation disagreements. The majority of disagreement comes from allocation disagreement, ranging from 5 to 8 percent, whereas the quantity disagreement is only 1 to 3 percent. The total disagreements are less than 10 percent for all four years, indicating that the classification is fairly accurate every year.

In addition, we used classification consistency to assess the classification accuracy over time. This was carried out according to a three-step procedure: (1) detecting the consecutive land use change from 2005 to 2008 using the binary land use data (i.e., 2005 → 2006 → 2007 → 2008), which should result in sixteen possible changes; (2) identifying the cells that witnessed invalid (false) changes; not all of these sixteen possible changes are valid in reality (e.g., the conversion from urban to nonurban is almost impossible); and (3) calculating the respective proportions of cells with valid and invalid changes, denoted as p_v and p_{iv} , respectively. The value of p_v is calculated by overlaying four years of binary land use data and counting cells with valid changes (Figure 4B). As a result, we found 11,948,717 cells with valid changes and 970,139 cells with invalid changes. Thus, the value of p_v is $[11,948,717/(11,948,717 + 970,139) = 0.9249]$. If persistent nonurban cells (i.e., no changes witnessed during the study period) are excluded from the calculation, the value of p_v becomes $[3,733,283/(3,733,283 + 970,139) = 0.7937]$. Generally, the value of p_v should be proportionate to the classification accuracy over time. The results of p_v suggest that the classification of urban area is fairly accurate to be used in subsequent urban analysis.

After the land use classification, landscape metrics were used to quantify the urban land use patterns. We selected four landscape metrics based on previous literature (Dietzel et al. 2005; Seto and Fragkias 2005), including total urban class area (UCA), the number of urban patches (NP), mean perimeter-area ratio (PARA), and mean Euclidean nearest-neighbor distance (ENN). For NP, a patch means an individual homogenous region of urban land use (Herold, Couclelis, and Clarke 2005). ENN is the average distance between a patch and its nearest neighbor. PARA is the mean value of the perimeter-area ratio of all urban patches.

These metrics were selected to reveal the characteristics of urban forms from three aspects: size, fragmentation, and regularity. UCA can be used to reflect the urban areal change. The combination of NP and ENN can be used to measure the fragmentation of land use pattern. Given the same amount of urban areas, higher values of NP and ENN_MN indicate a more fragmented or dispersed pattern. PARA represents the regularity of a pattern. A higher value of PARA suggests a more oddly shaped pattern.

As the inputs to the CA model, a series of spatial variables were generated using geographic information system (GIS) functions. The atlas of Guangdong Province for 2009 was used to obtain the distribution of city centers and town centers and transportation networks of the study area. These layers were further used to create spatial variables, such as the distance to city centers, the distance to towns, the distance to major expressways, the distance to major roads, and the distance to railways. The slope of the study area was produced using the digital elevation model (DEM). All of the spatial data have a resolution of 30 m.

Implementing SVR Models to Predict Energy Consumption and Urban Size

The energy consumption levels of the five selected cities were predicted based on SVR using two types of factors. The first type is the economic variables, including tertiary industrial output value (V_{tert}) and the

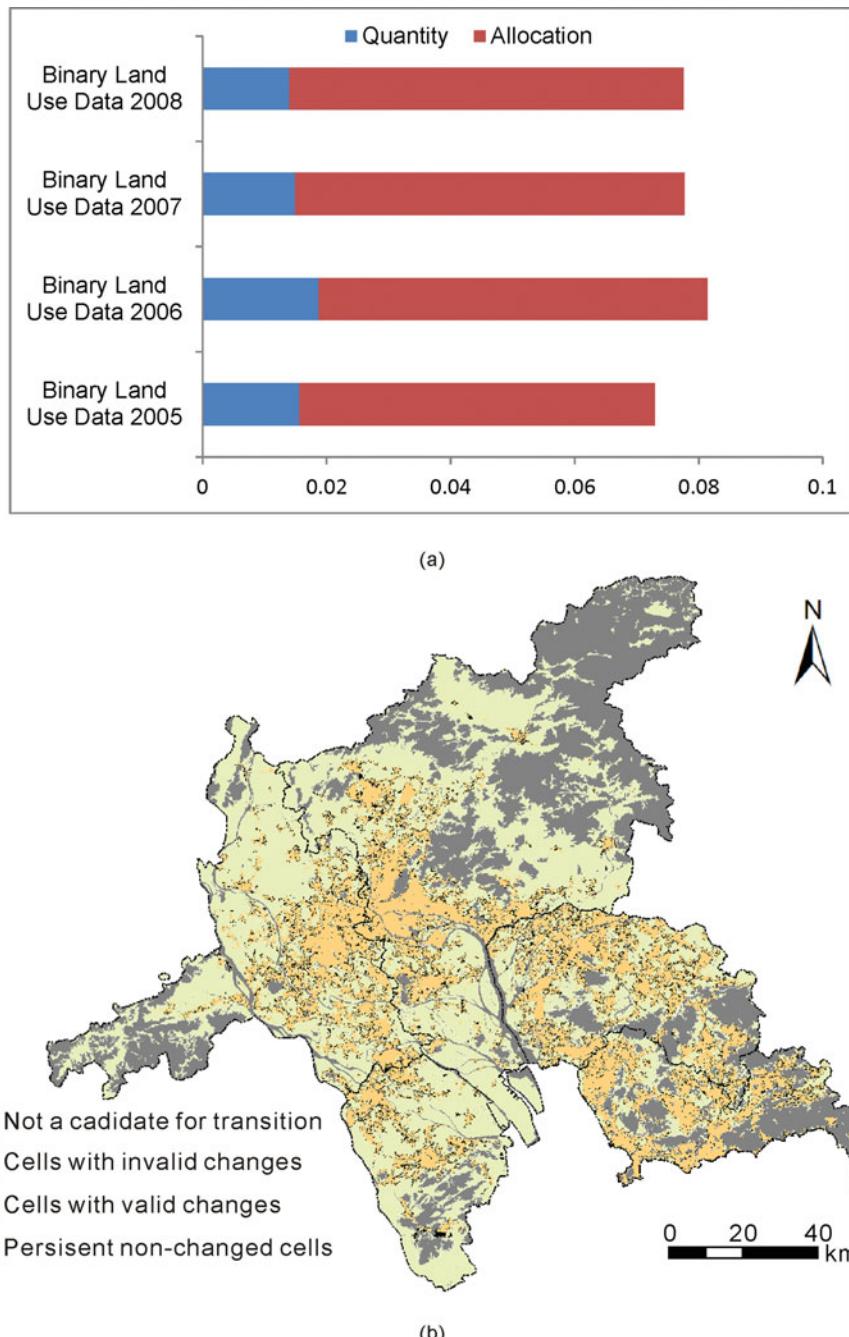


Figure 4. The classification accuracy of the land use data. (A) Quantity and allocation disagreements. (B) Consistency of the land use classification over time. (Color figure available online.)

gross products of three industrial sectors: the intensive energy-consuming sector (M_1), medium energy-consuming sector (M_2), and low energy-consuming sector (M_3). The second type includes those factors that reflect the characteristics of urban forms, represented by landscape metrics (UCA, NP, ENN, and PARA). Generally, landscape metrics are correlated, which leads to multicollinearity when using a conventional regression approach. Thanks to the principle of minimizing structural risk, however, SVR is not sensitive to intercorre-

lated variables (Gani, Taleb, and Limam 2010). Thus, the metrics can be used as input variables to predict energy consumption. The computation of these metrics was accomplished through the spatial analysis software FRAGSTATS (McGarigal et al. 2002).

The training of SVR was implemented in WEKA, a machine learning software package (Frank et al. 2010). Data were normalized and randomly split into two halves for training and testing, respectively. Table 6 shows the statistical description of these two data sets.

Table 6. Statistical description of the training and testing data sets (means and standard deviations)

	Training (10 instances)		Testing (10 instances)	
	M	SD	M	SD
E (10^6 tons of SCE)	27.26	17.92	28.39	8.19
M_1 (10^8 yuan)	279.16	139.57	393.27	140.88
M_2 (10^8 yuan)	544.84	322.91	602.31	264.51
M_3 (10^8 yuan)	1,335.72	1,031.49	754.59	199.05
V_{tert} (10^8 yuan)	2,307.68	1,779.15	1,604.45	906.41
P (10^4 persons)	643.02	345.02	699.93	145.61
UCA (km^2)	526.89	252.08	779.45	91.51
NP	220	79.56	277.70	53.49
ENN (m)	270.56	93.95	289.48	127.95
PARA	399.925	99.54	325.73	52.48

Note: SCE = standard coal equivalent; UCA = urban class area; NP = number of urban patches; ENN = Euclidean nearest neighbor differences; PARA = mean perimeter-area ratio.

The polynomial function and the radial-basis function were used to make a comparison in terms of mean relative error. The results are shown in Table 7. It can be seen that the polynomial function (exponent = 1) has the highest modeling accuracy, with the mean relative errors of 8.93 percent for training and 12.63 percent for testing, respectively.

Another SVR model was employed to project urban size. The variables used for this projection were selected based on previous studies. Deng et al. (2008) concluded that GDP was the most important driver of China's urban expansion. Han et al. (2009) regarded variables of population, GDP, and urbanization level to model the urban areal change of Shanghai. Deng et al. (2008) found that industrialization and the development of the tertiary sector were also crucial factors affecting the urban growth in China. In this study, we take into account

three categories of socioeconomic variables to predict urban size: population (P), gross products of three industrial sectors (M_1 , M_2 , and M_3), and the tertiary sector (V_{tert}). Configuration of this SVR model remains the same as the previous one: Data were normalized and split into two halves for training and testing, respectively. The performances of the polynomial and radial-basis functions were compared and the results are shown in Table 7. The respective mean relative errors of the polynomial function (exponent = 1) for training and testing are 12.87 and 16.02 percent, which are the lowest compared with the other two models. This model was used to estimate urban size during the simulation of PRD's urban growth from 2005 to 2008 and to project urban size in the scenario simulations.

Calibration of Logistic CA for Urban Growth Simulation

The logistic CA was calibrated using land use data in the years of 2005 and 2008. The input variables include distance to city centers (x_1), distance to towns (x_2), distance to expressway (x_3), distance to major roads (x_4), distance to railways (x_5), and slope (x_6). It is inappropriate to adopt the same set of calibrated parameters for all cities because the study area is large and complex. Separate calibrations were implemented for each city to avoid large simulation errors (Li, Yang, and Liu 2008). The calibrated parameters are shown in Table 8.

The performance of the logistic CA was tested through the simulation of realistic urban growth from 2005 to 2008. The number of iterations was set to 300, and the dispersion parameter δ was manually tuned using a trial-and-error approach (Table 8). Figures 5A and 5B are the observed and simulated urban land use patterns in 2008. The modeling outcome was validated at both local (pixel-by-pixel) and global (landscape metrics) axes. Pontius et al. (2007) proposed

Table 7. The errors of the support vector regression (SVR)-based models for predicting energy consumption and urban size

	Training (%)	Testing (%)
SVR-based energy consumption model (10^6 tons of SCE)		
Polynomial function (exponent = 1)	8.93	12.63
Polynomial function (exponent = 2)	11.90	15.42
Radial-basis function	33.70	53.13
SVR-based urban size model (km^2)		
Polynomial function (exponent = 1)	12.87	16.02
Polynomial function (exponent = 2)	22.75	35.44
Radial-basis function	69.52	78.71

Table 8. The calibration results of the logistic cellular automata

	Dongguan	Foshan	Guangzhou	Shenzhen	Zhongshan
b_1	-0.877	-3.853	-2.234	-0.615	-0.662
b_2	-1.417	-3.01	-4.645	-3.591	-3.520
b_3	-0.001	-0.825	-3.660	-0.228	-0.211
b_4	-1.518	-2.923	-6.686	-2.755	-1.872
b_5	-0.469	0.509	-3.284	1.037	-0.966
b_6	-14.243	-5.725	-11.899	-12.873	-2.862
b_0	1.196	1.960	3.204	0.960	2.011
δ	2.0	7.0	3.0	3.0	1.0

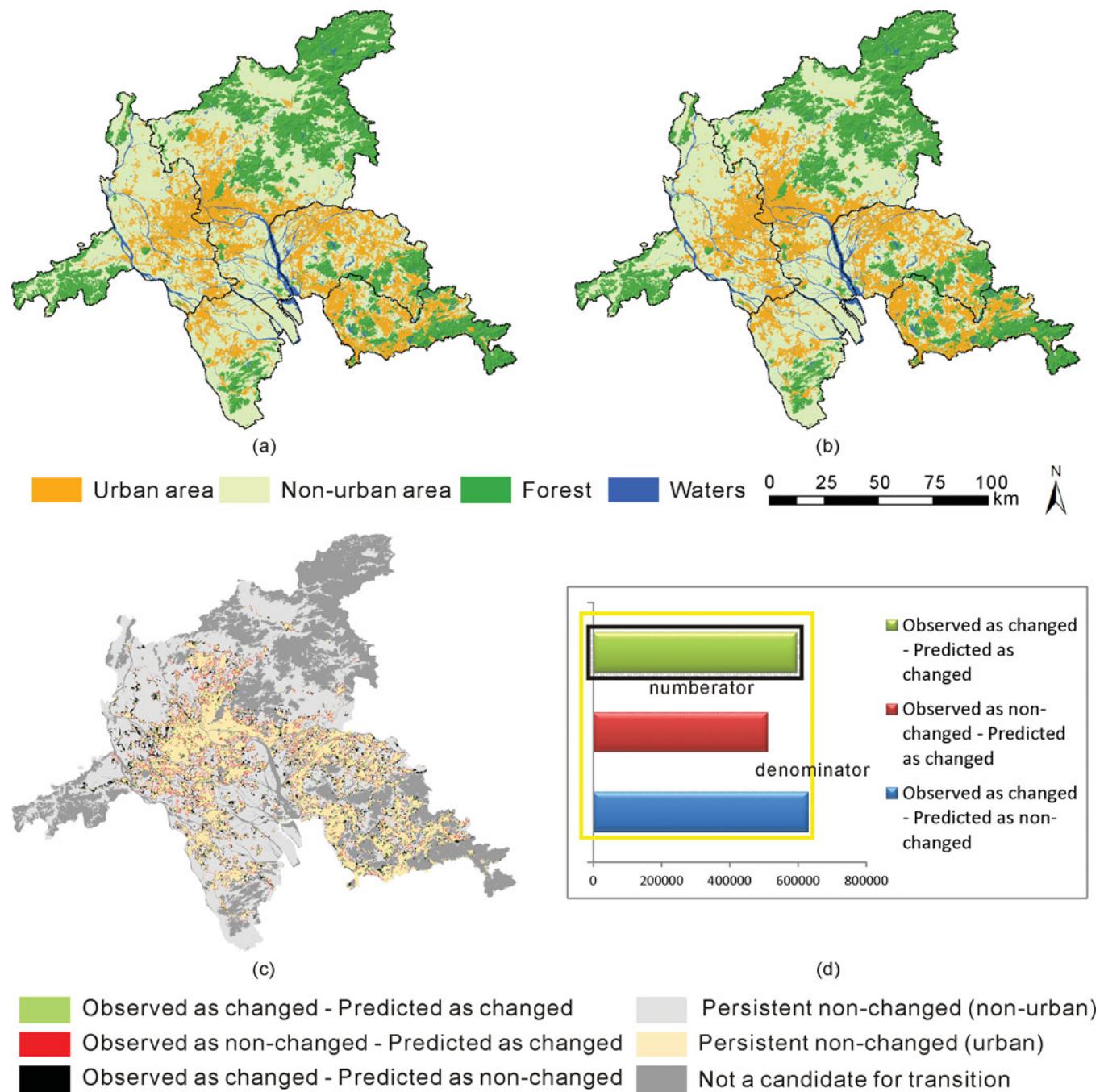


Figure 5. (A) and (B) The actual and simulated urban land use patterns, respectively. (C) Overlay of the actual and simulated land use. (D) Quantity of the three grouped cells. (Color figure available online.)

a pixel-by-pixel approach called *figure of merit* to assess the accuracy of a simulation model. Figure of merit is a ratio, where the numerator is the number of instances that changed and were correctly predicted as changed, and the denominator is the total number of instances excluding persistently nonchanging instances. Based on this ratio, Pontius et al. (2008) conducted a comparison of thirteen land change modeling applications,

in which they found that the value of the figure of merit ranged from 0.01 to 0.59.

We first overlaid the observed land use pattern with the simulated one to identify four groups of cells: (1) observed as changed and predicted as changed; (2) observed as nonchanged and predicted as changed; (3) observed as changed and predicted as nonchanged; and (4) persistent nonchanged. Figure 5C shows the

distribution of these groups of cells. Then we counted the respective number of cells for each group (Figure 5D) and calculated the figure of merit. As a result, we found 596,118 cells correctly predicted as changed and 1,141,890 cells wrongly predicted, including 511,220 cells observed as nonchanged but predicted as changed and 630,670 cells observed as changed but predicted as nonchanged. Therefore, the value of the figure of merit should be $[596,118 / (596,118 + 1,141,890)] = 0.3430$. This value is average compared with that of other land use models (Pontius et al. 2008).

The simulated land use patterns were also validated at the landscape level. This was carried out by comparing the values of landscape metrics (NP, PARA, and ENN) between the observed and simulated patterns. A similarity index was used to measure the overall accuracy:

$$A = 1 - \frac{1}{n} \sum_{i=1}^n \frac{|a_{i,s} - a_{i,o}|}{a_{i,o}} \quad (30)$$

where n is the number of metrics and $a_{i,s}$ and $a_{i,o}$ are values of metrics derived from the simulated pattern and the observed pattern, respectively.

Table 9 lists the values of landscape metrics for the observed and simulated patterns. Table 9 also shows the results of the similarity index A . The values of A are highest in the simulations of Dongguan and Zhongshan and lowest in the simulation of Guangzhou, as Guangzhou has a much larger territory than the other four cities. Nevertheless, the average value of A is over 70 percent for all five cities. This indicates that the model is accurate enough for further applications.

Evaluating the Impacts Under Different Development Strategies

Table 1 demonstrates that even though energy use intensity is declining annually for all five cities, their total energy consumption levels are growing at a much higher rate. For example, the energy use intensity of Foshan reduced by 15.8 percent from 2005 to 2008; however, its total energy consumption increased by 38.3 percent during this period. This indicates that the improvement in energy efficiency does not offset the increase in overall energy consumption. This result is similar to those of Güneralp and Seto (2012), who examined the energy consumption of building construction and operation in the PRD and discovered that the improvements in energy efficiency lagged behind the

Table 9. Validating the simulated patterns using landscape metrics

	NP	PARA	ENN
Actual land use patterns			
Dongguan	184	395.5731	196.5591
Foshan	329	267.5776	204.7529
Guangzhou	369	328.8492	435.4597
Shenzhen	181	425.139	210.2299
Zhongshan	141	303.7895	218.9249
Simulated land use patterns			
Dongguan	140	458.6587	252.2834
Foshan	243	327.4381	281.5401
Guangzhou	206	473.4667	418.7878
Shenzhen	152	557.8764	276.1894
Zhongshan	109	380.0164	274.4164
Overall similarity (A)			
Dongguan		77.26%	
Foshan		71.33%	
Guangzhou		69.34%	
Shenzhen		73.79%	
Zhongshan		75.62%	

Note: NP = number of urban patches; PARA = mean perimeter-area ratio; ENN = Euclidean nearest-neighbor differences.

growth of overall energy demand. Changing the development strategy can be another approach to reduce regional energy consumption. The proposed model can be used to evaluate the impacts of different development strategies on energy consumption through scenario simulations.

In fact, the five cities have announced development plans to guide their future economic and urban development. For example, the Guangzhou Municipal Government (2009) put forward a target of promoting the tertiary industry in the Outline of Guangzhou Program of Building a Modern Industrial System (2009–2015), and the Shenzhen Municipal Government (2009) highlighted the importance of high-tech industries in the approved Overall Plan of Shenzhen for Modern Industry System (2009–2015). Foshan argued the need for the development of petrochemical industry because it already has a basis for this industry (see Ministry of Commerce of the People's Republic of China 2010).

To explore the potential impacts of the proposed development strategy on urban growth and energy consumption, four scenarios of development in 2011 were created based on the development plans already mentioned. Scenario 1 assumes that the region will continue its current development strategy in the future. In Scenario 2, the region will prefer to develop industries in the intensive energy consuming sector, whereas in Scenario 3 the region will focus on the development of

low energy consuming industries. In Scenario 4, higher priority is given to the development of tertiary industry instead of manufacturing industries.

The quadratic model was used to extrapolate the socioeconomic variables (population and the gross products of both industrial and tertiary sectors) before the scenario simulations:

$$y = at^2 + bt + c \quad (31)$$

where y is the predicted socioeconomic variable, and t is the time variable (year). Estimation of coefficients a , b , and c was based on statistical data from 2000 to 2008, which are listed in Table 4. Other details of the four scenarios are specified next.

Scenario 1: Baseline Scenario. In this scenario, the development strategies for the five cities remain unchanged. The values of V_{tert} , M_1 , M_2 , and M_3 for each city were forecasted using Equation 31. Urban size was then projected and the urban land use patterns were simulated by the calibrated CA model. The simulated patterns were quantified using the metrics NP, ENN, and PARA. Energy consumption was then predicted based on the SVR model.

Scenario 2: Preferring Industries in the Intensive Energy Consuming Sector. Among the five cities, Foshan has the highest proportion of industries in the intensive energy consuming sector (37.76 percent; see Table 3). Such a situation will continue if the development plan of Foshan is followed. In this scenario, the development strategy of Foshan was applied to the simulation of the other four cities, using Foshan's CA model parameters. Specifically, the values of V_{tert} , M_1 , M_2 , and M_3 for Foshan in 2011 were forecasted using Equation 31, and the respective proportions of M_1 , M_2 , and M_3 can be derived, denoted as $p_{m1, FS}$, $p_{m2, FS}$, and $p_{m3, FS}$. For the other four cities, the values of V_{tert} , M_1 , M_2 , and M_3 in 2011 were extrapolated, and the generated M_1 , M_2 , and M_3 were rescaled based on $p_{m1, FS}$, $p_{m2, FS}$, and $p_{m3, FS}$. The urban land use patterns were then simulated based on the calibrated CA model, constrained by the projected urban size. The simulated patterns were quantified using the metrics NP, ENN, and PARA. Finally, the energy consumption of each city was predicted using the SVR-based energy prediction model.

Scenario 3: Preferring Industries in the Low Energy Consuming Sector. In contrast with Scenario 2, Scenario 3 assumes that the majority of industrial out-

puts exclusively came from the low energy-consuming sector in 2011. Recently, Shenzhen had approximately 74 percent of industrial outputs from industries in the low energy consuming sector (Table 3). The development plan of Shenzhen emphasizes the development of such industries in the future. In this scenario, the values of V_{tert} , M_1 , M_2 , and M_3 for Shenzhen in 2011 were first forecasted using Equation 31. Then the respective proportions of M_1 , M_2 , and M_3 can be derived, denoted as $p_{m1, SZ}$, $p_{m2, SZ}$, and $p_{m3, SZ}$. For the other four cities, the values of V_{tert} , M_1 , M_2 , and M_3 were extrapolated, and the generated M_1 , M_2 , and M_3 were rescaled based on $p_{m1, SZ}$, $p_{m2, SZ}$, and $p_{m3, SZ}$. The rest of the procedure is similar to Scenario 2, except that Shenzhen's CA model parameters were implemented for the entire region.

Scenario 4: Preferring Industries in the Tertiary Sector. The study area has recently witnessed a rapid growth in the tertiary industry. For instance, the proportion of tertiary industry was as high as 59.0 percent in Guangzhou in 2008. The development plan of Guangzhou indicates that the city will prefer to grow the tertiary industry in its future development. Thus, this scenario assumes that Guangzhou's development strategy will be implemented in the other four cities. Specifically, the values of gross domestic output V_{tert} , M_1 , M_2 , and M_3 for Guangzhou in 2011 were forecasted using Equation 31. Meanwhile, the respective proportions of the outputs of the industrial sector and tertiary sector were determined, denoted as $p_{m, GZ}$ and $p_{tert, GZ}$. For the other four cities, the values of the gross domestic output were predicted beforehand, and the values of V_{tert} , M_1 , M_2 , and M_3 were then disaggregated based on $p_{tert, GZ}$ and $p_{m, GZ}$. The rest of the procedure is similar to that of Scenario 2, except that each city's original CA model parameters were replaced by the ones for Guangzhou.

Figure 6 shows the predicted values of V_{tert} , M_1 , M_2 , and M_3 in each scenario. Table 10 lists the projected urban size of each city for these scenarios. The total urban size of the five cities is 4,564.18 km² in the baseline scenario, in which the region follows the current development strategy. The total urban size increases to 5,131.67 km² if the strategy of developing industries in the intensive energy-consuming sector is adopted (Scenario 2). On the contrary, the total urban size significantly decreases to 4,080.87 km² if the region's major industrial outputs come from the low energy-consuming sector (Scenario 3). If the tertiary industry becomes the dominant sector of the regional economy

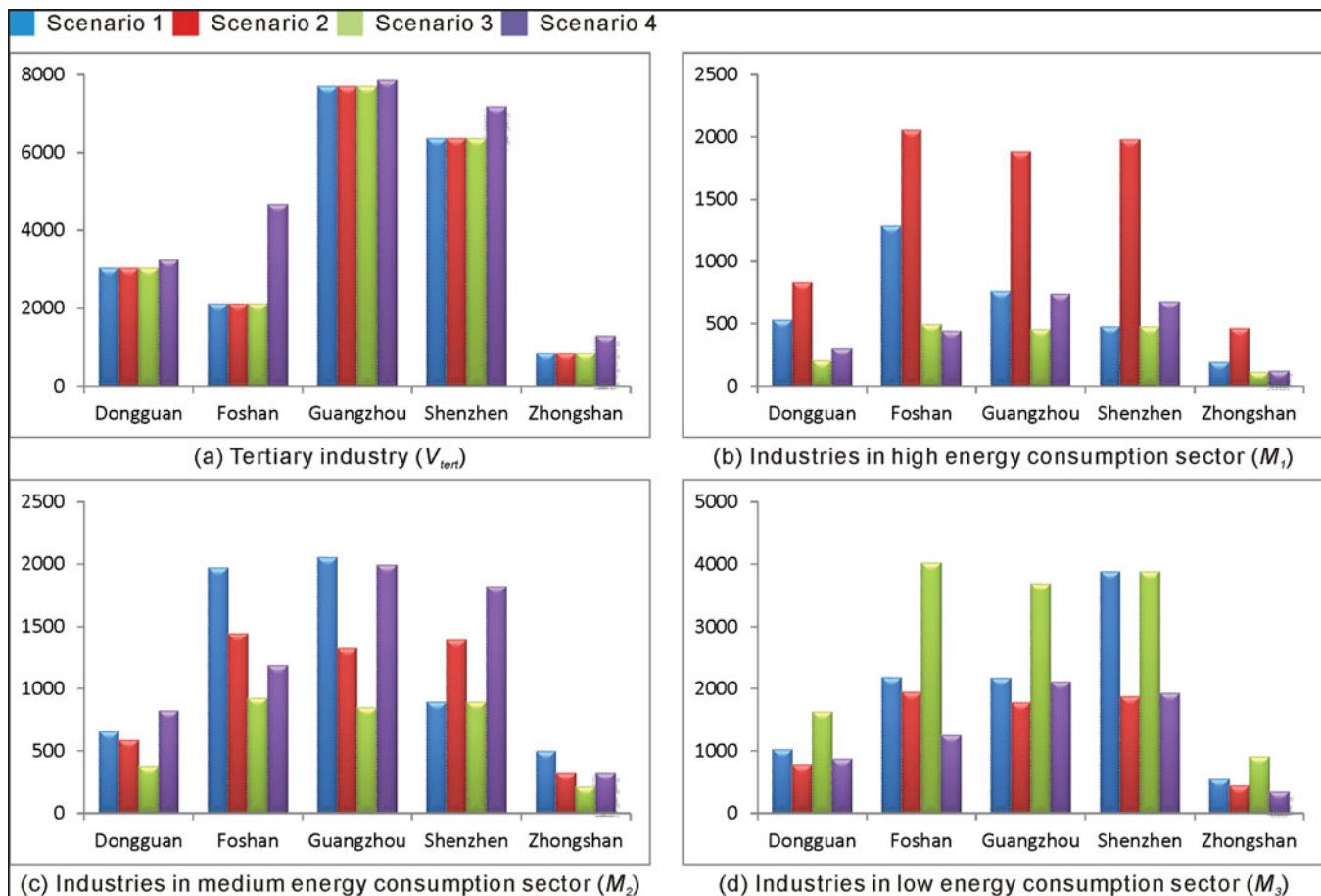


Figure 6. Predicted values of V_{tert} , M_1 , M_2 , and M_3 in the four scenarios. (Color figure available online.)

Table 10. The predicted urban size and energy consumption in the four development scenarios

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Projected urban size (km^2)				
Dongguan	979.26	1,058.99	929.17	930.47
Foshan	1,202.05	1,202.05	897.62	907.83
Guangzhou	1,211.49	1,461.57	1,139.26	1,211.49
Shenzhen	774.38	950.25	774.38	876.47
Zhongshan	397.00	458.82	350.44	359.43
Total	4,564.18	5,131.67	4,080.87	4,285.69
Predicted energy consumption (10^6 tons of SCE)				
Dongguan	34.58	38.16	33.07	32.97
Foshan	49.08	49.08	42.67	42.65
Guangzhou	71.65	81.92	69.14	71.65
Shenzhen	53.46	62.02	53.46	54.39
Zhongshan	13.62	16.16	13.27	13.27
Total	222.39	247.35	211.60	214.94

Note: SCE = standard coal equivalent.

(Scenario 4), the total urban size ($4,285.69 \text{ km}^2$) becomes less than that of the baseline scenario but much higher than that of Scenario 3. This result is unexpected because the total land demand for promoting the tertiary industry should be lower than that of promoting industrial production. A possible reason is that the recent boom in real estate development requires a large amount of land for the construction of residential buildings and various kinds of villas.

Figure 7 shows the simulated urban land use patterns for these scenarios. The simulation can help visualize the potential impacts of different development strategies. For example, Scenario 2 (Figure 7B) will cause a large quantity of land to be converted to urban use and, in particular, the nonurban area is almost depleted in Shenzhen. Scenario 3 (Figure 7C) is more reasonable because it still shows sufficient space for the city to grow in the future. All these patterns are quantified using the metrics NP, ENN, and PARA. The results, along with the predicted values of V_{tert} , M_1 , M_2 , and M_3 were used to estimate energy consumption for each city.

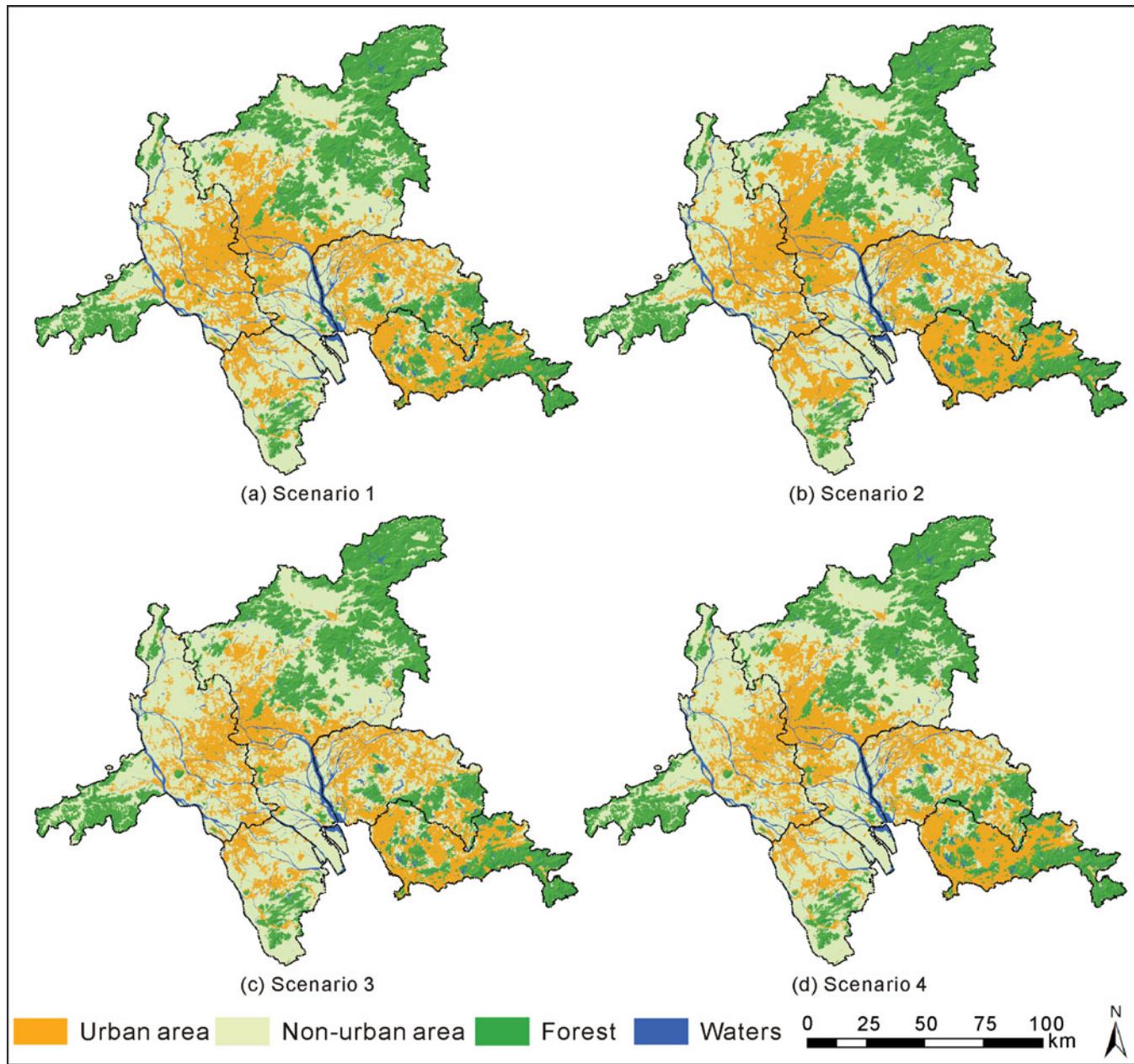


Figure 7. Simulated urban land use patterns in the four scenarios. (Color figure available online.)

Table 10 also lists the predicted energy consumption of the five cities in 2011. In the baseline scenario, energy consumption is 222.39 million tons of SCE (Scenario 1). The highest energy consumption (247.35 million tons of SCE) is shown in Scenario 2, which assumes that the region prefers to develop industries in the intensive energy-consuming sector, whereas the lowest energy consumption (211.60 million tons of SCE) is observed in Scenario 3, in which the region strongly promotes industries in the low energy-consuming sector. In addition, compared with the result of the baseline

scenario, a moderate reduction of both land and energy consumption can be seen in Scenario 4 (developing tertiary industry).

Further comparison of the results reveals an interesting finding. The relative differences of projected urban size are larger than those for predicted energy consumption among the four scenarios. For instance, the comparison between Scenarios 1 and 2 shows that the percentage increase in urban size (12.43 percent) is slightly higher than that of energy consumption (11.22 percent). Such differences are more obvious in the

comparison between Scenarios 1 and 3; the percentage reduction in urban size is 10.59 percent, and the percentage reduction in energy consumption is only 4.85 percent. A similar result is observed in the comparison between Scenarios 1 and 4, in which the respective percentage changes are 6.10 percent and 3.35 percent.

Such results indicate that compared with energy consumption, urban size is more sensitive to the adjusted economic structure, perhaps because the land requirements vary among industries in each sector. Although we never know the exact land requirements of each industry, evidence suggests that industries in the high energy-consuming sector usually occupy a larger amount of land. For example, the sector of oil refinery, coking, and nuclear fuel processing has the highest energy intensity (3.46 tons of SCE/ 10^4 yuan) across all industries, and a representative enterprise in this sector, the SINOPEC Guangzhou Company (http://english.sinopec.com/about_sinopec/subsidiaries/refineries_petrochemicals/20080326/3043.shtml), covers a vast area of 3.7 km^2 in eastern Guangzhou. By contrast, enterprises in the low energy-consuming sector require much less land resources on average. For instance, the Shenzhen High-Tech Industrial Park, with an area of 11.5 km^2 , includes hundreds of enterprises.

In summary, the scenario simulations presented represent four types of development strategies. The strategy of developing energy-intensive industries requires massive inputs of both energy and land. Hence, it is not suitable for cities like Shenzhen, where developable land has already become scarce (see Figure 7B). Meanwhile, the transportation demand might also increase because of the rapidly growing urban size. The PRD is expected to rely more on road-based transportation to meet the expanding demand for mobility, which can aggravate the energy problem and air pollution in this region (Yang et al. 2011). On the contrary, the demands for land and energy are much lower if the strategy of developing industries in the low energy consuming sector is adopted. Therefore, given the same size of the economy, increasing the share of industries in the low energy-consuming sector is more helpful in balancing economic development and energy and land consumption. Promoting the tertiary industry is another alternative for future development. Generally, a shift from primary and secondary industry to tertiary industry is happening in industrialized regions. The analysis in this study reveals that the strategy of promoting tertiary industry can, to some extent, reduce both land and energy consumption.

Conclusion

This study presents a model that integrates CA and SVR to evaluate the impacts of different development strategies on urban growth and energy consumption. The proposed model was tested in the PRD, which is a rapidly developing region in China. The logistic CA model was used to simulate the urban forms of the study area, constrained by the projected urban size. The landscape metrics were then adopted to quantify the simulated urban forms. Finally, the SVR model was employed to predict energy consumption using landscape metrics and other socioeconomic variables.

Scenario simulations were carried out based on the respective development plans of Guangzhou, Foshan, and Shenzhen to examine the effects of the modified economic structure on urban growth and energy consumption. Compared with the baseline scenario (Scenario 1), Scenario 2 (the development strategy of Foshan is implemented) will largely increase the demands for land resources and energy. In contrast, the development strategy of Shenzhen requires far less land and energy resources for future development. Promoting tertiary industry (Guangzhou's strategy), to some extent, can also reduce the demands for both land and energy.

An apparent limitation of this study is that the empirical land use data only involve urban and nonurban types due to the limited resolution of Landsat images (30 m). Land use data with more specific types (e.g., residential, commercial, and industrial uses) can be helpful to reveal the link between land use structure and energy consumption. Such land use data are also useful for analyzing the geographical placement of energy demand and providing implications for energy management and policymaking. In our future research, remote sensing images with higher resolution (e.g., IKONOS or QuickBird images) will be considered as data sources to map detailed urban land uses, although much time and funding will be needed to conduct an analysis at a regional scale using these images.

As a rapidly developing region and the world's manufacturing base, PRD sustains its economic growth at the cost of a large volume of fossil fuel use with relatively low efficiency. In addition to prediction of future energy demand, it is important to assess the impacts of energy policies on carbon emission. This field has been regarded by Horner, Zhao, and Chapin (2011) as one of the fertile research areas in synthesizing GIScience and energy issues. Therefore, in future studies, we will apply the recent scenario simulations of energy consumption

to the evaluation of their respective carbon emission volumes. Energy structure data are necessary for this kind of research, but such data are not available for all cities in the PRD. Additional efforts must be made to improve data availability.

Acknowledgments

We thank Dr. Mei-Po Kwan and the anonymous reviewers for their useful comments and suggestions. We also thank Professor Xun Shi for providing insightful comments and polishing the article's language. We are indebted to the National Basic Research Program of China (973 Program; Grant No. 2011CB707103), the Key National Natural Science Foundation of China (Grant No. 40830532), the National Natural Science Foundation of China (Grant No. 41171308), and the Sun Yat-sen Innovative Talents Cultivation Program for Excellent Tutors.

References

- Anderson, W. 1996. Urban form, energy and the environment: A review of issues, evidence and policy. *Urban Studies* 33 (1): 7–36.
- Badoe, D. A., and E. J. Miller. 2000. Transportation-land-use interaction: Empirical findings in North America, and their implications for modeling. *Transportation Research Part D: Transport and Environment* 5 (4): 235–63.
- Camagni, R., M. Gibelli, and P. Rigamonti. 2002. Urban mobility and urban form: The social and environmental costs of different patterns of urban expansion. *Ecological Economics* 40 (2): 199–216.
- Chen, C., H. Gong, and R. Paaswell. 2008. Role of the built environment on mode choice decisions: Additional evidence on the impact of density. *Transportation* 35 (3): 285–99.
- Chen, Y., X. Li, Y. Zheng, Y. Guan, and X. Liu. 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., and L. Gaydos. 1998. Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science* 12 (7): 699–714.
- Crompton, P., and Y. Wu. 2005. Energy consumption in China: Past trends and future directions. *Energy Economics* 27 (1): 195–208.
- Definiens Developer 7.0. 2012. Definiens Developer Version 7. <http://www.ecognition.com/document/definiens-developer-version-7> (last accessed 17 November 2012).
- Deng, X., J. Huang, S. Rozelle, and E. Uchida. 2008. Growth, population and industrialization, and urban land expansion of China. *Journal of Urban Economics* 63 (1): 96–115.
- Dhakal, S. 2009. Urban energy use and carbon emissions from cities in China and policy implications. *Energy Policy* 37 (11): 4208–19.
- Dieleman, F., M. Dijst, and G. Burghouwt. 2002. Urban form and travel behaviour: Micro-level household attributes and residential context. *Urban Studies* 39 (3): 507–27.
- Dietzel, C., H. Oguz, J. Hemphill, K. Clarke, and N. Gazulis. 2005. Diffusion and coalescence of the Houston Metropolitan Area: Evidence supporting a new urban theory. *Environment and Planning B: Planning and Design* 32 (2): 231–46.
- Energy Information Association. 2009. *International energy outlook: 2009*. Washington, DC: Energy Information Association.
- Ewing, R., and R. Cervero. 2001. Travel and the built environment: A synthesis. *Transportation Research Record: Journal of the Transportation Research Board* 1780:87–114.
- Fang, M., C. Chan, and X. Yao. 2009. Managing air quality in a rapidly developing nation: China. *Atmospheric Environment* 43 (1): 79–86.
- Frank, E., M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. Witten, and L. Trigg. 2010. WEKA—A machine learning workbench for data mining. In *Data mining and knowledge discovery handbook*, ed. O. Maimon and L. Rokach, 1269–77. New York: Springer.
- Gani, W., H. Taleb, and M. Limam. 2010. Support vector regression based residual control charts. *Journal of Applied Statistics* 37 (2): 309–24.
- Guangzhou Municipal Government. 2009. Outline of Guangzhou program of building a modern industrial system (2009–2015). <http://english.gz.gov.cn/publicfiles/business/htmlfiles/outline/index.html> (last accessed 10 November 2012).
- Güneralp, B., and K. C. Seto. 2012. Can gains in efficiency offset the resource demands and CO₂ emissions from constructing and operating the built environment? *Applied Geography* 32 (1): 40–50.
- Han, J., Y. Hayashi, X. Cao, and H. Imura. 2009. Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China. *Landscape and Urban Planning* 91 (3): 133–41.
- Herold, M., H. Couclelis, and K. Clarke. 2005. The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems* 29 (4): 369–99.
- Holden, E., and I. Norland. 2005. Three challenges for the compact city as a sustainable urban form: Household consumption of energy and transport in eight residential areas in the greater Oslo region. *Urban Studies* 42 (12): 2145–66.
- Horner, M. W., T. Zhao, and T. S. Chapin. 2011. Toward an integrated GIScience and energy research agenda. *Annals of the Association of American Geographers* 101 (4): 764–74.
- Hua, X., Y. Ni, J. Ko, and K. Wong. 2007. Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. *Journal of Computing in Civil Engineering* 21 (2): 122–35.
- Jenks, M., and R. Burgess, eds. 2000. *Compact cities: Sustainable urban forms for developing countries*. London: E & FN Spon.

- Li, X., Y. Chen, X. Liu, D. Li, and J. He. 2011. Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. *International Journal of Geographical Information Science* 25 (4): 633–55.
- Li, X., X. Shi, J. He, and X. Liu. 2011. Coupling simulation and optimization to solve planning problems in a fast-developing area. *Annals of the Association of American Geographers* 101 (5): 1032–48.
- Li, X., Q. S. Yang, and X. P. Liu. 2008. Discovering and evaluating urban signatures for simulating compact development using cellular automata. *Landscape and Urban Planning* 86 (2): 177–86.
- McGarigal, K., S. Cushman, M. Neel, and E. Ene. 2012. FRAGSTATS: Spatial pattern analysis program for categorical maps. <http://www.umass.edu/landeco/research/fragstats/fragstats.html> (last accessed 17 November 2012).
- Mindali, O., A. Raveh, and I. Salomon. 2004. Urban density and energy consumption: A new look at old statistics. *Transportation Research Part A: Policy and Practice* 38 (2): 143–62.
- Ministry of Commerce of the People's Republic of China. 2010. Foshan is introducing international petroleum enterprises for the development of modern petrochemical industrial areas. <http://cccmc.mofcom.gov.cn/aarticle/ztxx/201006/20100606974217.html> (last accessed 10 November 2012; in Chinese).
- National Academy of Sciences of the United States. 2009. *Driving and the built environment: The effects of compact development on motorized travel, energy use, and CO₂ emissions*. Washington, DC: Transportation Research Board of the National Academies.
- Oliveira, A. 2006. Estimation of software project effort with support vector regression. *Neurocomputing* 69 (13–15): 1749–53.
- Pan, H., Q. Shen, and M. Zhang. 2009. Influence of urban form on travel behaviour in four neighbourhoods of Shanghai. *Urban Studies* 46 (2): 275–94.
- Pontius, R., W. Boersma, J. Castella, K. Clarke, T. de Nijs, C. Dietzel, Z. Duan, E. Fotsing, N. Goldstein, and K. Kok. 2008. Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science* 42 (1): 11–37.
- Pontius, R. G., and M. Millones. 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32 (15): 4407–29.
- Pontius, R., R. Walker, R. Yao-Kumah, E. Arima, S. Aldrich, M. Caldas, and D. Vergara. 2007. Accuracy assessment for a simulation model of Amazonian deforestation. *Annals of the Association of American Geographers* 97 (4): 677–95.
- Seto, K., and M. Fragkias. 2005. Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. *Landscape Ecology* 20 (7): 871–88.
- Shao, M., X. Tang, Y. Zhang, and W. Li. 2006. City clusters in China: Air and surface water pollution. *Frontiers in Ecology and the Environment* 4 (7): 353–61.
- Shenzhen Municipal Government. 2009. Overall plan of Shenzhen for modern industry system (2009–2015). http://www.sz.gov.cn/zfgb/2009/gb660/200907/t20090723_1154710.htm (last accessed 10 November 2012; in Chinese).
- Silva, E., and K. Clarke. 2005. Complexity, emergence and cellular urban models: Lessons learned from applying SLEUTH to two Portuguese metropolitan areas. *European Planning Studies* 13 (1): 93–115.
- Smola, A., and B. Schölkopf. 2004. A tutorial on support vector regression. *Statistics and Computing* 14 (3): 199–222.
- State Council of the People's Republic of China. 2011. 12th five-year plan on national economic and social development. <http://news.sina.com.cn/c/2011-03-17/055622129864.shtml> (last accessed 10 November 2012; in Chinese).
- Statistics Bureau of Dongguan Municipality. 2009. *Dongguan statistical yearbook 2009*. Beijing, China: China Statistical Press.
- Statistics Bureau of Foshan Municipality. 2009. *Foshan statistical yearbook 2009*. Beijing, China: China Statistical Press.
- Statistics Bureau of Guangdong Province. 2009. *Guangdong statistical yearbook 2009*. Beijing, China: China Statistical Press.
- Statistics Bureau of Guangzhou Municipality. 2009. *Guangzhou statistical yearbook 2009*. Beijing, China: China Statistical Press.
- Statistics Bureau of Shenzhen Municipality. 2009. *Shenzhen statistical yearbook 2009*. Beijing, China: China Statistical Press.
- Statistics Bureau of Zhongshan Municipality. 2009. *Zhongshan statistical yearbook 2009*. Beijing, China: China Statistical Press.
- Wolfram, S. 1984. Cellular automata as models of complexity. *Nature* 311 (4): 419–24.
- Wu, F. L. 2002. Calibration of stochastic cellular automata: The application to rural-urban land conversions. *International Journal of Geographical Information Science* 16 (8): 795–818.
- Yang, J., C. Fang, C. Ross, and G. Song. 2011. Assessing China's megaregional mobility in a comparative context. *Transportation Research Record: Journal of the Transportation Research Board* 2244 (1): 61–68.
- Yang, J., and R. Gakenheimer. 2007. Assessing the transportation consequences of land use transformation in urban China. *Habitat International* 31 (3–4): 345–53.
- Zellner, M., T. Theis, A. Karunanithi, A. Garmestani, and H. Cabezas. 2008. A new framework for urban sustainability assessments: Linking complexity, information and policy. *Computers, Environment and Urban Systems* 32 (6): 474–88.