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

Yimin Chen[1, 3], Xia Li*[1, 3], Bin Ai[2, 3]

[1] School of Geography and Planning, Sun Yat-sen University, Guangzhou, P.R. China
[2] School of Marine Sciences, Sun Yat-sen University, Guangzhou, P.R. China
[3] Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, P.R. China
email corresponding author: lixia@mail.sysu.edu.cn

Abstract

The rapid growth of China's energy consumption has resulted in many problems in this country. Chinese government has realized the necessity to improve energy efficiency and reduce energy consumption. This study presents a model that integrates the support vector regression (SVR) and cellular automata (CA) to simulate the urban forms and to estimate the corresponding energy consumptions in one of the most developed regions in China, the Pearl River Delta (PRD). We simulated four scenarios to assess the impacts of different development strategies on urban forms and the related energy consumptions. The result indicates that land demand is more sensitive to the change of economic structure rather than energy consumption. The comparison of different simulated scenarios suggests that promoting low energy consuming industries is the most effective strategy to balance the economic development and energy and land consumptions.
Keywords: urban forms; energy consumption; cellular automata; support vector regression

1. Introduction

Just about ten years ago, China's energy consumption was a half 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). The rapid growth of energy consumption led to many environmental and social problems (Fang et al. 2009). In addition, the insufficient domestic energy supply has forced China to rely on fuel imports, which has raised the concerns about energy security (Crompton and Wu 2005).

Since over 70% of China's total energy is used for industrial production (Energy Information Association 2009), China's "12th Five-Year Plan" urges the country to improve the energy efficiency and reduce energy consumption by adjusting economic structure. Under this context, we selected one of the most developed regions in China, the Pearl River Delta (PRD), as the study area to explore the potential effects of the change of 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 the low labor and land costs. The processing technologies of manufacturing industries, featured by low energy efficiency, were dominant in the region (Fang et al. 2009) and has brought about many environmental problems (Shao et al. 2006). The provincial government has been considering to implement long-term measures to reduce energy consumption and improve the environment, including adjusting the economic structure and upgrading the processing technologies and equipment (Fang et al. 2009).

In addition, the control of urban form should also be considered as the means to reduce energy consumption. The built-environment can affect households' traveling behaviors which are related to energy consumption. In North America, residents prefer public transit or walking in those high density employment centers, as these areas usually have concentrated transit hubs (Chen et al. 2008). Residents were also less likely to own vehicles and tended to used 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 United States 2009). In China, the land and housing reform has broken the workplace-residence tie in the pre-reform urbanized areas since the 1980s. This resulted in increasing spatial separation between workplace and residence. Highways were then intensively constructed by local governments to improve the accessibility; and the automobile use was also encouraged (Yang and Gakenheimer 2007). Such rapid motorization, along with the lengthened trips, generated serious traffic problems in Chinese cities.

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 consumptions. The simulation should be useful for exploring the impacts of different development strategies on urban growth and the energy consumption. 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 being coupled with spatial optimization models (Li et al. 2011a; Li et al. 2011b).

In addition, we adopted the support vector regression (SVR) (Smola and Schölkopf 2004) to predict the corresponding energy consumption for different urban forms. SVR is a new technique of classification and prediction, and has been used to handle complex relationships in many fields (Hua et al. 2007; Oliveira 2006). 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.

2. Method

Figure 1 illustrates the flow of the proposed model. In this model, the CA was used to simulate urban forms with the urban size constraint, which was produced by a SVR model based on a set of socio-economic 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 socio-economic variables. Details of SVR and CA are provided in the following sections.

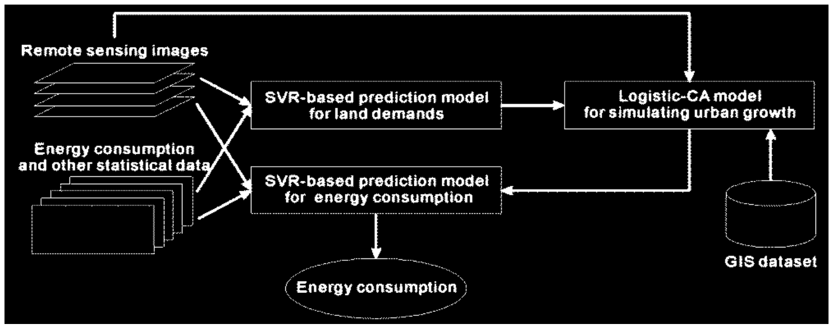


Fig. 1. An integrated model to evaluate the impacts of different development strategies on urban growth and energy consumption.

2.1. Support vector regression

In SVR, the objective is to find a function $f(x) = \langle w, x \rangle + b$ that best fits the training dataset, where w is the weight vector, b is the threshold ($w \in \chi, b \in R$), and <* , *> 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 as (Smola and Schölkopf 2004):

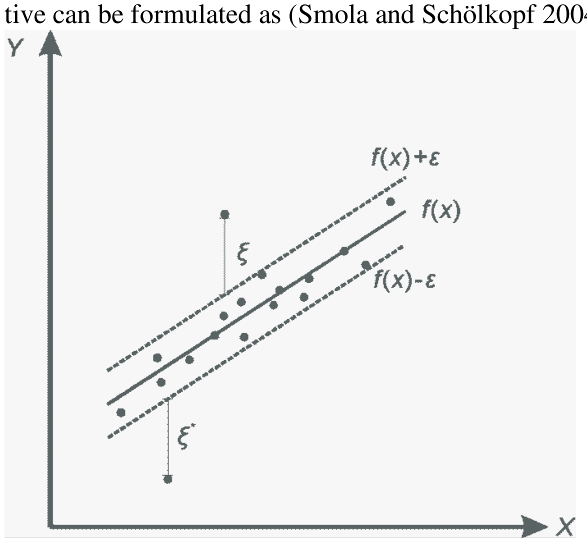


Fig. 2. The ε-insensitive loss function in SVR.

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

subject to: $\begin{cases} (< w, x_i > +b) - y_i \le \varepsilon + \xi_i, \\ (y_i - < w, x_i > +b) \le \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0. \end{cases}$ (2)

where C is a positive constant, representing the tradeoff between the flatness of f and the errors. The minimization of equation (1) is based on the Lagrange function:

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}\alpha_i\left[\varepsilon + \xi_i + < w, x_i > +b - y_i\right]$$
$$- \sum_{i=1}^{l}\alpha_i^*\left[\varepsilon + \xi_i^* + y_i - < w, x_i > -b\right] - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) \quad (3)$$

where α_i, α_i^*, η_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 the above equations into equation (3), the minimization problem becomes:

minimize:

$$-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) < x_i, x_j > -\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*)$$

(8)

subject to: $\begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \\ 0 \le \alpha_i, \alpha_i^* \le C \end{cases}$ (9)

Based on equation (4), the $f(x)$ can be re-formulated 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 (KKT) 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, whose errors 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 non-zero coefficients are called support vectors. Furthermore, the optimization process described above 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, the 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 ith observation and its estimate, respectively.

2.2. Logistic-CA

The CA model was developed largely based on Wu's (2002) method, but was enhanced by incorporating the urban size projected by the 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, $x_2 \ldots x_n$), 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; x_k is the development factors of cell_{ij}.

However, the probability $p_{g,ij}$ only addresses the influences of static physical factors. The actual urban development would be also 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:

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

where $con()$ 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 water body, mountain, or restricted areas, the cell should be excluded from development. Therefore, the development probability is revised as follows:

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

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

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

where $p_{c,\max}^t$ is the maximum value of $p_{c,ij}^t$ in space at time t; and δ is called 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.

3. Implementation and results

3.1. Study area

The study area is located in the Pearl River Delta (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 the boom of the regional economy, as well as the rapid urbanization process. At present, the PRD has the highest per capita GDP among the several most developed regions in China (Shao et al. 2006). However, it requires a vast volume of natural resources, especially fuel resources, to sustain its economic growth. According to the Guangdong Statistical Yearbook (http://www.gdstats.gov.cn/tjnj/ml_c.htm), the energy consumption of the entire province reached 226.72 million tons of the Standard Coal Equivalent (SCE) in 2008.

Fig. 3. Location of the Pearl River Delta.

3.2. 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 (kg of SCE) of city i, and P_i is the city's population. Table 1A shows the energy intensity (energy consumption per unit GDP) for all the five cities from 2005 to 2008, which were 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 1B). For the three cities, the differences between the estimates and recorded values are insignificant, which validates the use of equation (25) to estimate the energy consumptions of Foshan and Shenzhen.

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

	2005	2006	2007	2008
(A) Energy consumption per unit GDP (t. 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
(B) Total energy consumption (10^6 t. 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)

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: Guangdong Statistical Yearbook and the respective statistical yearbooks of Dongguan, Foshan, Guangzhou, Shenzhen, and Zhongshan.

While industrial production was generally the largest source of energy consumption in the five cities, the subsectors of industry can be quite different in terms using energy. To classify the subsectors based on their energy consumptions, we calculated the energy intensity for each of the 39 subsectors identified in China's statistical system for the five cities. Due to the data limitation, we had to assume that the energy intensity of a certain sub-sector is the same for all five cities, and use the provincial data to conduct the calculation. Based on the energy intensity values, we classify the subsectors into three groups: intensive energy consuming sector (avg. 2.35 t. of SCE/10^4 Yuan, including industries such as smelting and pressing of ferrous metals, nonmetal mineral products, petroleum refining, coking, nuclear fuel processing, papermaking and paper products, etc), medium energy consuming sector (avg. 0.68t. of SCE/10^4 Yuan, including industries such as manufacture of chemical fibers, manufacture of medicines, printing and record medium peproduction, manufacture of textile garments, footwear and headgear, etc), and low energy consuming sector (avg. 0.32t. of SCE/10^4 Yuan, including industries such as manufacture of general/special-purpose machinery, manufacture of communication equipment, computers and other electronic equipment, recycling and disposal of waste, etc). Table 3 lists the proportions of the industrial products of these subsectors in year 2008 for each city.

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

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: Guangdong Statistical Yearbook and the respective statistical yearbooks of Dongguan, Foshan, Guangzhou, Shenzhen, and Zhongshan.

In this study, we used multi-temporal 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. 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). The land use classes of farmland, fishpond, and bare soil were merged into non-urban area, regarded as candidates of 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) to assess the classification accuracy. This method divides the disagreements between classification and reference into quantity disagreement and allocation disagreement. 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 sample 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 and the allocation disagreement of land use class g; Q and A are the overall quantity disagreement and the allocation disagreement, respectively; D is the total disagreement.

We calculated the quantity and allocation disagreements for the binary land use data (urban and non-urban) from 2005 to 2008. 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%, whereas the quantity disagreement is only 1% to 3%. The total disagreements are less than 10% for all four years.

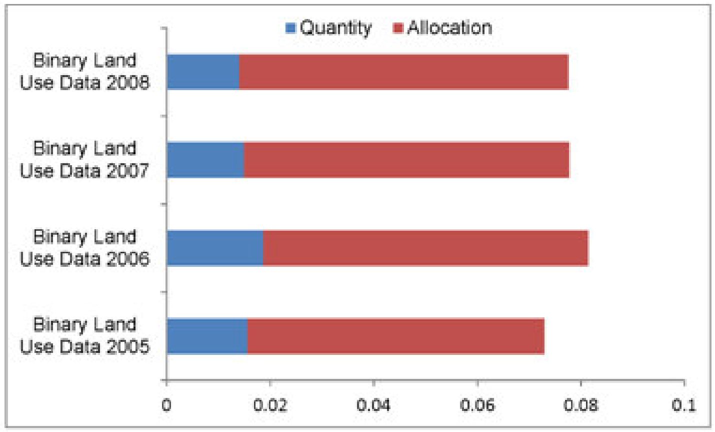
Table 5. Confusion matrices of the classification of urban areas (2005 to 2008)

		2005		2006		2007		2008	
	Reference	Urb.	Non-urb.	Urb.	Non-urb.	Urb.	Non-urb.	Urb.	Non-urb.
Classification	Urb.	397	73	447	71	487	80	520	102
	Non-urb.	142	2,229	158	2,155	182	2,082	149	2,060

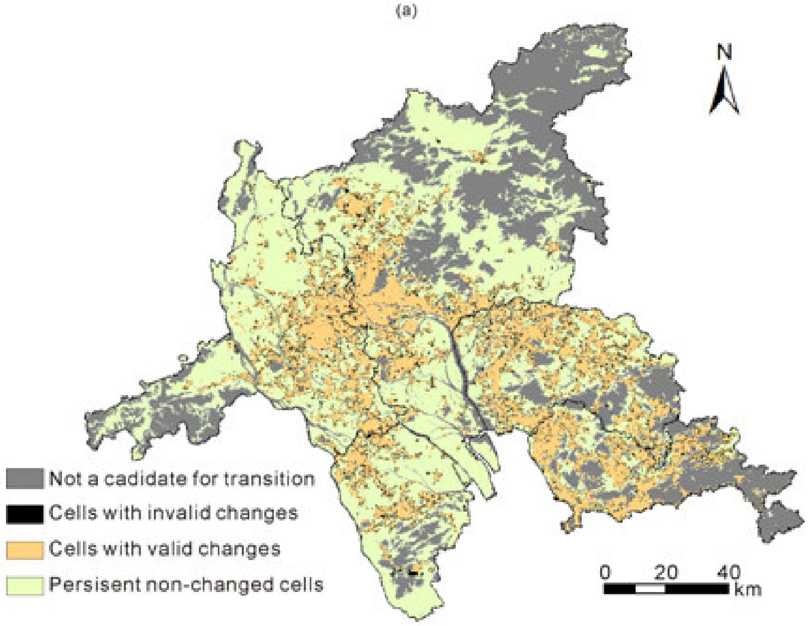
In addition, we used classification consistency to assess the over-time classification accuracy. This was carried out according to a three-step procedure: (1) detecting the consecutive land use change from 2005 to 2008 using the binary land use data (i.e., 2005 → 2006 → 2007 → 2008). As a result, there should be 16 possible changes; (2) identifying the cells that witnessed invalid (false) changes. Not all of these 16 possible changes are valid in reality. For example, the conversion from urban to non-urban is almost impossible; (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' 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 0.9249. If persistent non-urban cells are excluded from the calculation, the value of p_v becomes 0.7937.

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 et al. 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.

As the inputs to the CA model, a series of spatial variables were generated using GIS functions. The Atlas of Guangdong Province 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 the spatial data have a resolution of 30 m.



(a)



(b)

Fig. 4. The classification accuracy of theland use data. (A) Quantity and allocation disagreements. (B) Consistency of the land use classification over time.

3.3. Implementing SVR models to predict energy consumption and urban size

The energy consumptions 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 gross products of three industrial sectors: 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).

The training of SVR was implemented in WEKA, a machine learning software (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. 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 7A. It can be seen that the polynomial function (exponent = 1) has the highest modeling accuracy, with the mean relative errors of 8.93% for training and 12.63% for, respectively.

Table 6. Statistical description of the training and testing data sets (mean and standard deviation).

-	Training (10 instances)	Testing (10 instances)
E (10^6 t. 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 (1031.49)	754.59 (199.05)
V_{tert} (10^8 yuan)	2,307.68 (1779.15)	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: Standard deviations of variables are shown in the parentheses.

Another SVR model was employed to project urban size. We take into account three categories of socio-economic 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. The performances of the polynomial and radial-basis functions were compared and the results are shown in Table 7B. The respective mean relative errors of the polynomial function (exponent = 1) for training and testing are 12.87% and 16.02%, 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.

Table 7. The errors of the SVR-based models for predicting energy consumption and urban size.

	Training	Testing
(A) SVR-based energy consumption model (10^6 t. 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%
(B) 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%	- 8.71%

3.4. 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). The calibrated parameters are shown in Table 8**.**

Table 8. The calibration results of the Logistic-CA.

	Dongguan	Foshan	Guangzhou	Shezhen	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

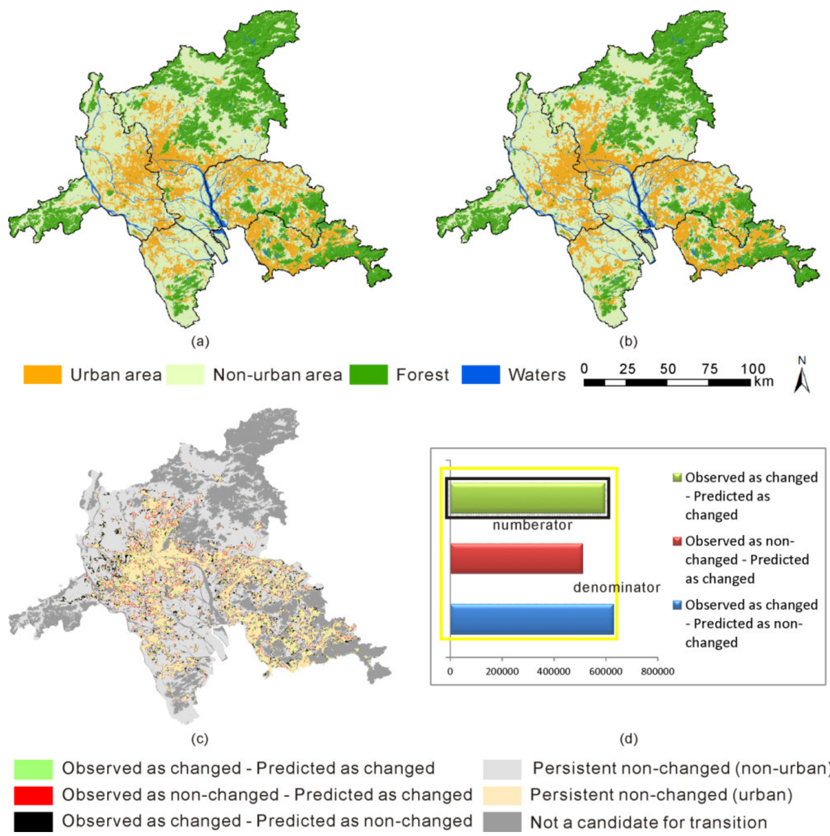


Fig. 5. (A) and (B) the actual and simulated urban land use patterns, respectively; (C) overlap of the actual and simulated land use; (D) quantity of the three grouped cells.

Figures 5A and B 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 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 correctly predicted as changed, while the denominator is the total number of instances excluding persistently non-changing instances. Based on this ratio, Pontius et al. (2008) conducted a comparison of 13 land change modeling applications, in which they found that the value of "figure of merit" ranges 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 non-changed and predicted as changed; (3) observed as changed and predicted as non-changed; and (4) persistent non-changed. 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." The value of "figure of merit" is 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 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; $a_{i,s}$ and $a_{i,o}$ are values of metrics derived from the simulated pattern and the observed pattern, respectively.

Table 9. Validating the simulated patterns using landscape metrics.

	NP	PARA	ENN
(A) 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
(B) 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
(C) Overall similarity (A)			
Dongguan	77.26%		
Foshan	71.33%	-	
Guangzhou	69.34%	-	
Shenzhen	73.79%	-	
Zhongshan	75.62%	-	

Tables 9A and B list the values of landscape metrics for the observed and simulated patterns. Table 9C shows the results of the similarity index A. The values of A are highest in the simulations of Dongguan and

Zhongshan while the lowest in the simulation of Guangzhou. The reason is that Guangzhou has a much larger territory than the other four cities. Nevertheless, the average value of the *A* is over 70% for all five cities. This indicates that the model is accurate enough for further applications.

3.5. Evaluating the impacts under different development strategies

To explore the potential impacts of the proposed development strategy on urban growth and energy consumption, four scenarios of the development in 2011 were created based on the development plans mentioned above. Scenario 1 assumes that the region will continue the 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 socio-economic 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 socio-economic 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 below:

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%, see Table 3). Such 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 re-scaled 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 outputs exclusively came from the low energy consuming sector in 2011. Recently, Shenzhen had approximately 74% 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 of 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 re-scaled 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 to the entire region.

Scenario 4: Preferring industries in the tertiary sector. The study area has recently witnessed a rapid growth in the tertiary industry (including information transmission, computer services and software, wholesale and retail trades, hotels and catering services, financial intermediation, etc). For instance, the proportion of tertiary industry was as high as 59.0% 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 of 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 of were replaced by the ones of Guangzhou.

Figure 6 shows the predicted values of V_{tert}, M_1, M_2, and M_3 in 2011 for each scenario. Table 10A lists the projected urban size of each city for these scenarios. The total urban size of the five cities is 4,564.18 km^2 in the baseline scenario, in which the region follows the current development strategy. The total urban size increases to 5,131.67 km^2 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 km^2 if the region's major industrial outputs come from the low energy consuming sector (Scenario 3). In case the tertiary industry becomes the dominant sector of the regional economy (Scenario 4), the total urban size (4,285.69 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 of promoting the tertiary industry should be lower than that of promoting industrial production. A possible reason is that the recent boom of real estate development requires a large amount of land for the construction of residential buildings and various kinds of villas.

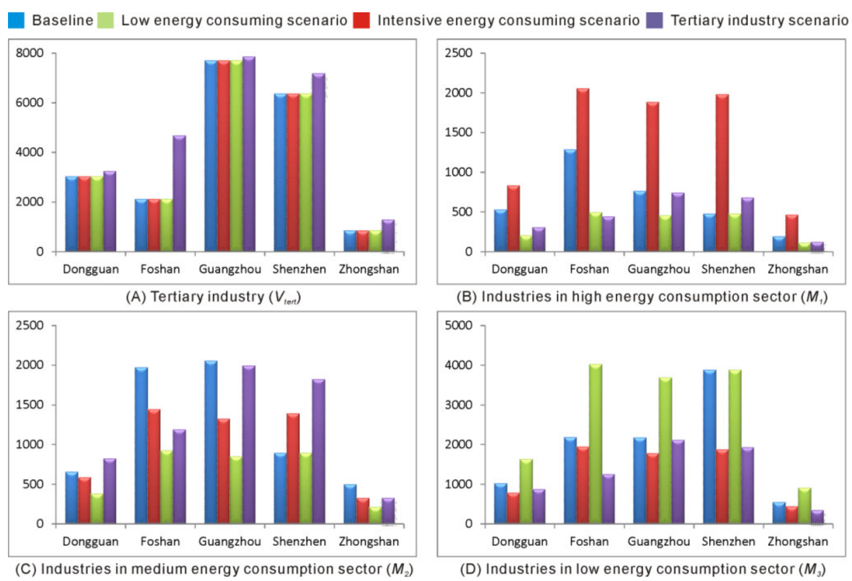


Fig. 6. Predicted gross production values of V_{tert}, M_1, M_2, and M_3 (10^8 yuan) in the four scenarios.

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 into urban land use and, in particular, the non-urban 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.

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

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
(A) Projected urban size (km^2)				
Dongguan	979.26	1058.99	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
(B) Predicted energy consumption (10^6 t. 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: Scenario 1 = Baseline; Scenario 2 = Preferring industries in the intensive energy consuming sector; Scenario 3 = Preferring industries in the low energy consuming sector; Scenario 4 = Preferring industries in the tertiary sector.

Table 10B lists the predicted energy consumptions of the five cities in 2011. In the baseline scenario, the energy consumption is 222.39 million tons of SCE (Scenario 1). The highest energy consumption (247.35 million tons of SCE) is witnessed in Scenario 2 that assumes 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 consumptions can be seen in Scenario 4 (developing the tertiary industry).

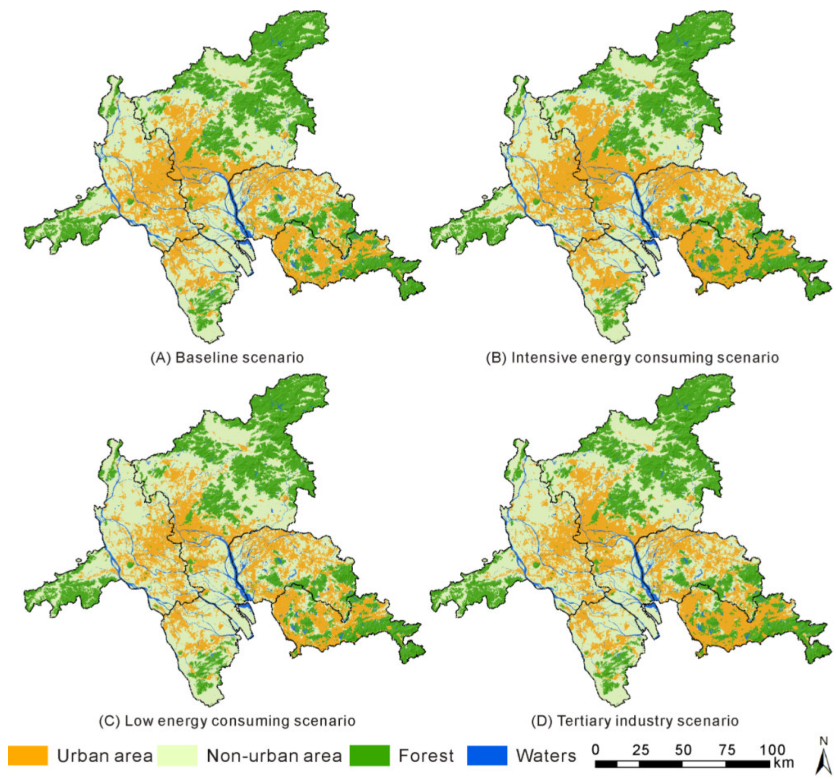


Fig. 7. Simulated urban land use patterns in the four scenarios.

Further comparison of the results reveals an interesting finding. The relative differences of projected urban size are larger than those of the predicted energy consumption among the four scenarios. For instance, the comparison between Scenarios 1 and 2 shows that the percentages increase of the urban size (12.43%) is slightly higher than that of the energy consumption (11.22%). Such differences are more obvious in the comparison between Scenarios 1 and 3; the percentage reduction of urban size is 10.59%, while the percentage reduction of energy consumption is only 4.85%. A similar result is observed in the comparison between Scenarios 1 and 4, in which the respective percentage changes are 6.10% and 3.35%. Such results indicate that compared with energy consumption, urban size is more sensitive to the adjusted economic structure, perhaps because the land requirement varies 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.

In summary, the scenario simulations above 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, whose developable land has already become scarce (see Figure 7B). On the contrary, the demands of 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 economy, increasing the share of industries in the low energy consuming sector is more helpful to balance the economic development and energy and land consumptions. Promoting the tertiary industry is another alternative for future development. Generally, a shift from the primary and secondary industry to the tertiary industry is happening in industrialized regions. The analysis in this study reveals that the strategy of promoting the tertiary industry can, to some extent, reduce both land and energy consumptions.

4. 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 the landscape metrics and other socio-economic 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 of land resources and energy. In contrast, the development strategy of Shenzhen requires far less land and energy resources for future development. Promoting the tertiary industry (Guangzhou’s strategy), to some extent, can also reduce the demands of both land and energy.

References

Badoe DA, Miller EJ (2000) Transportation-land-use interaction: empirical findings in North America, and their implications for modeling. Transportation Research Part D: Transport and Environment 5(4):235-263

Chen C, Gong H, Paaswell R (2008) Role of the built environment on mode choice decisions: additional evidence on the impact of density. Transportation 35(3):285-299

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

Crompton P, Wu Y (2005) Energy consumption in China: past trends and future directions. Energy Economics 27(1):195-208

Definiens Developer 7.0 (2003) Definiens Developer Version 7. Available from http://www.ecognition.com/document/definiens-developer-version-7 accessed Access Date Access Year)

Dietzel C, Oguz H, Hemphill J, Clarke K, Gazulis N (2005) Diffusion and coalescence of the Houston Metropolitan Area: evidence supporting a new urban theory. Environment and Planning B: Planning and Design 32(2):231-246

Energy Information Association (2009) International Energy Outlook: 2009.

Ewing R, Cervero R (2001) Travel and the built environment: a synthesis. Transportation Research Record: Journal of the Transportation Research Board 1780:87-114

Fang M, Chan C, Yao X (2009) Managing air quality in a rapidly developing nation: China. Atmos Environ 43(1):79-86

Frank E, Hall M, Holmes G et al (2010) Weka-a machine learning workbench for data mining. Data Mining and Knowledge Discovery Handbook:1269-1277

Herold M, Couclelis H, Clarke K (2005) The role of spatial metrics in the analysis and modeling of urban land use change. Computers, Environment and Urban Systems 29(4):369-399

Hua X, Ni Y, Ko J, Wong K (2007) Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. J Comput Civil Eng 21(2):122-135

Li X, Chen Y, Liu X, Li D, He J (2011a) Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. Int J Geogr Inf Sci 25(4):633-655

Li X, Shi X, He J, Liu X (2011b) Coupling Simulation and Optimization to Solve Planning Problems in a Fast-Developing Area. Ann Assoc Am Geogr 101(5):1032-1048

National Academy of Sciences of United States (2009) Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use, and CO_2 Emissions. Transportation Research Board of the National Academies, Washington, DC

Oliveira A (2006) Estimation of software project effort with support vector regression. Neurocomputing 69(13-15):1749-1753

Pontius R, Boersma W, Castella J 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

Pontius R, Walker R, Yao-Kumah R et al (2007) Accuracy assessment for a simulation model of Amazonian deforestation. Ann Assoc Am Geogr 97(4):677-695

Pontius RG, Millones M (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32(15):4407-4429

Seto K, Fragkias M (2005) Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landscape Ecol 20(7):871-888

Shao M, Tang X, Zhang Y, Li W (2006) City clusters in China: air and surface water pollution. Frontiers in Ecology and the Environment 4(7):353-361

Silva E, Clarke K (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, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199-222

Yang J, Gakenheimer R (2007) Assessing the transportation consequences of land use transformation in urban China. Habitat International 31(3-4):345-353

Zellner M, Theis T, Karunanithi A, Garmestani A, Cabezas H (2008) A new framework for urban sustainability assessments: Linking complexity, information and policy. Computers, Environment and Urban Systems 32(6):474-488