

Article

# Simulation of Dynamic Urban Growth with Partial Least Squares Regression-Based Cellular Automata in a GIS Environment

Yongjiu Feng <sup>1,2,\*</sup>, Miaolong Liu <sup>3</sup>, Lijun Chen <sup>4,5</sup> and Yu Liu <sup>1,2</sup>

<sup>1</sup> College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; liuy@shou.edu.cn

<sup>2</sup> The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources (Ministry of Education), Shanghai Ocean University, Shanghai 201306, China

<sup>3</sup> College of Surveying and Geo-informatics, Tongji University, Shanghai 200092, China; liuml@tongji.edu.cn

<sup>4</sup> Key Laboratory of Ecohydrology of Inland River Basin, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China; chenlj2001@126.com

<sup>5</sup> School of Geography, Planning and Environmental Management, University of Queensland, Brisbane 4072, Australia; lijuan.chen@uq.edu.au

\* Correspondence: yjfeng@shou.edu.cn; Tel.: +86-21-6190-0304

Academic Editors: Qiming Zhou, Zhilin Li and Wolfgang Kainz

Received: 25 September 2016; Accepted: 30 November 2016; Published: 16 December 2016

**Abstract:** We developed a geographic cellular automata (CA) model based on partial least squares (PLS) regression (termed PLS-CA) to simulate dynamic urban growth in a geographical information systems (GIS) environment. The PLS method extends multiple linear regression models that are used to define the unique factors driving urban growth by eliminating multicollinearity among the candidate drivers. The key factors (the spatial variables) extracted are uncorrelated, resulting in effective transition rules for urban growth modeling. The PLS-CA model was applied to simulate the rapid urban growth of Songjiang District, an outer suburb in the Shanghai Municipality of China from 1992 to 2008. Among the three components acquired by PLS, the first two explained more than 95% of the total variance. The results showed that the PLS-CA simulated pattern of urban growth matched the observed pattern with an overall accuracy of 85.8%, as compared with 83.5% of a logistic-regression-based CA model for the same area. The PLS-CA model is readily applicable to simulations of urban growth in other rapidly urbanizing areas to generate realistic land use patterns and project future scenarios.

**Keywords:** urban growth; dynamic simulation; cellular automata; partial least squares (PLS) regression; geographical information systems (GIS); accuracy analysis

## 1. Introduction

Cellular automata (CA) method is a discrete dynamic modeling technique that has been widely applied in fields related to spatiotemporal distributions [1–4]. Classical CA formalism has been extended to accommodate the complexity of many systems [5,6]. Geographical information systems (GIS) based CA models have attracted extensive attention because of their ability to simulate urban growth and land use change [7–9], following the pioneering work of Tobler [10].

Over the past two decades, remarkable achievements have been made in geographical CA-based dynamic urban growth and land use change modeling, particularly in rapidly urbanizing areas [11–16]. Substantial progress has also been made in CA methodology, including transition rules retrieval, neighborhood configuration, scale effects, and results assessment [17–20]. One important issue in CA modeling is the quantification of the impacts of the factors that drive urban growth and land use change at both global and local scales. Many approaches have been developed to define CA transition rules

and each is aimed at improving the overall accuracy and reducing errors of simulation [21–24]. These approaches vary widely in theoretical assumptions, underlying methodologies, and spatio-temporal resolutions and extents [25]. For example, a CA model based on artificial neural networks (ANN) was developed to calculate land conversion probabilities and model dynamic land use in a GIS environment [21]. This model was used to simulate the multiple land use changes in a rapidly growing area of Guangdong Province, China. A heuristic CA model of urban land use change was proposed based on a simulated annealing (SA) algorithm and was successfully applied to simulate the urban growth in one of Shanghai's outer suburbs [22]. This model was built around a function that minimizes the difference (residual) between observed and simulated land use patterns, resulting in improved locational accuracy when compared to a logistic-regression-based CA model (named logistic-CA). Other heuristic optimization algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) have been used to optimize CA parameters from logistic regression and calibrate CA models [22,26–28]. A landscape expansion index was incorporated into CA (LEI-CA) to simulate both the adjacent and outlying urban growth of Dongguan City in southern China [15]. This approach demonstrated an improvement when compared to the logistic-CA model in terms of urban simulation accuracy. A random forest based CA model was used to simulate urban growth in Harare Metropolitan Province, Zimbabwe from 1984 to 2013 [24]. This model outperformed CA models based on support vector machine (SVM) and logistic regression in the study area. Markov chain integrated CA (CA-Markov) models are another class of methods developed in the last decade to simulate multiple land use changes [29,30]. The CA-Markov has become increasingly popular in geographical modeling since it was included in IDRISI. Most of these proposed new models perform better than earlier models, substantially advancing CA-based modeling of urban growth/expansion and land use change across the world. Current trends of CA model calibration, such as ANN, SVM, GA, SA and PSO, have become more complex [2,27,29]. Therefore, reconsideration of the statistical approaches is necessary for CA-based urban modeling.

Statistical approaches such as logistic regression and principal components analysis (PCA) are relatively simple and easy to implement using modern software packages. As a classical method, logistic regression has proved to be reliable in CA modeling [11,18,31,32]; however, most of the studies were conducted without consideration of correlation among variables. Moreover, the logistic regression method is incapable of eliminating the negative effects of the multicollinearity among variables [21,28]. By adding an auto-covariate term, logistic regression can be used to reduce the effect of correlation and, hence, increase its predictive accuracy in modeling land use change. A case study of the Paochiao watershed region in Taiwan shows that auto-logistic regression performs better than logistic regression [33]. PCA was used to reduce the effect of multicollinearity among spatial variables and obtain more reasonable CA parameters [34], yielding an improvement in performance when compared to the logistic-CA model. Statisticians have pointed out that the PCA method produces principal components that reflect only the covariance structure between the independent variables [35], and, as a consequence, the extracted components may only weakly explain the variance of the independent variable corresponding to the dependent variable in the regression.

The issue of variable multicollinearity, therefore, has continuously pushed researchers to develop more accurate, justifiable, and defensible models for simulating urban growth. Partial least squares (PLS) regression appears to be useful in addressing correlation because it integrates and generalizes features from PCA and multiple regression methods [36,37]. The method offers three advantages: (1) it removes data redundancies and extracts components from highly correlated spatial variables that better represent and explain the dependent variables (land conversion); (2) it avoids the detrimental effects in modeling due to multicollinearity and can regress when the number of observations is less than the number of variables; and (3) it integrates the basic functions of regression models, PCA, and canonical correlation analysis. In summary, PLS searches for the principal components that explain as much as possible of the covariance between the independent and dependent variables. The parameters

obtained using the PLS method might then better explain the dependent variables, i.e., the conversion probability of urban growth.

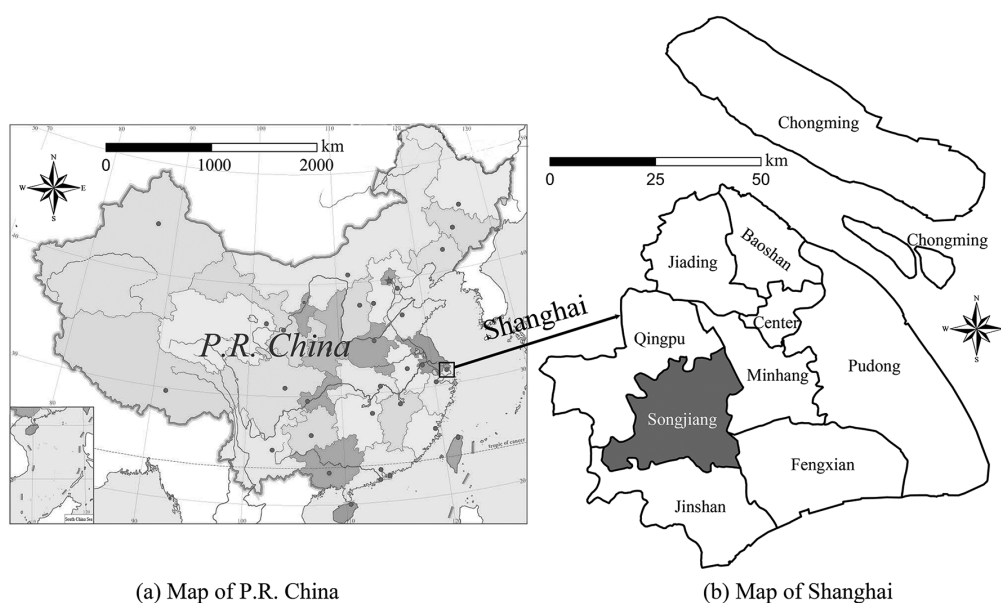
This paper presents a novel CA model based on the PLS approach that we call PLS-CA. This approach was used to derive principal components of the spatial variables for regressing CA parameters. Compared to logistic-CA, PLS can extract variables that are uncorrelated amongst the explanatory variables, and also between the explanatory and response variables. The result is the discovery of important transition rules from a number of driving factors that may be highly correlated. Our PLS-CA model was applied to simulate urban growth in the Songjiang district, an outer suburb of Shanghai Municipality, from 1992 to 2008. For comparison, a logistic-CA model was also applied to simulate the urban growth in the same study area.

## 2. Material

### 2.1. Study Area and Data

Songjiang is an outer suburb in the southwest part of Shanghai Municipality that is centered at  $121^{\circ}45'$  E and  $31^{\circ}00'$  N. Songjiang has a total area of  $598.5 \text{ km}^2$ ,  $15.5 \text{ km}^2$  of which is water (Figure 1). Over the past two decades, the urban area of Songjiang has grown rapidly with a significant increase in economic activity and concomitant dramatic land use change. According to the local government census, the total registered population has increased from 498,600 in 1995 to 1,074,200 in 2008. Rapid population growth has resulted in an explosive expansion of the urban area [38]. Such large-scale land use change and rapid urban growth have led to the degradation of the landscape, environment, and ecosystem [39,40].

Two Landsat-5 TM/ETM+ images acquired on 18 July 1992 and 24 March 2008 were collected to derive the changes in patterns in the study area. Other, essential ancillary datasets including 1:5000 administrative, topographic, and transportation maps were also collected from the local government. A total of 21 ground control points (GCPs) were identified on the remote sensing images using the topographic map as a reference source. A polynomial method was adopted for geometric rectification and the resulting accuracy obtained was 0.34 and 0.28 pixels for 1992 and 2008, respectively. Finally, the areal extent of Songjiang was clipped from the rectified Landsat images using the administrative map as the boundary.



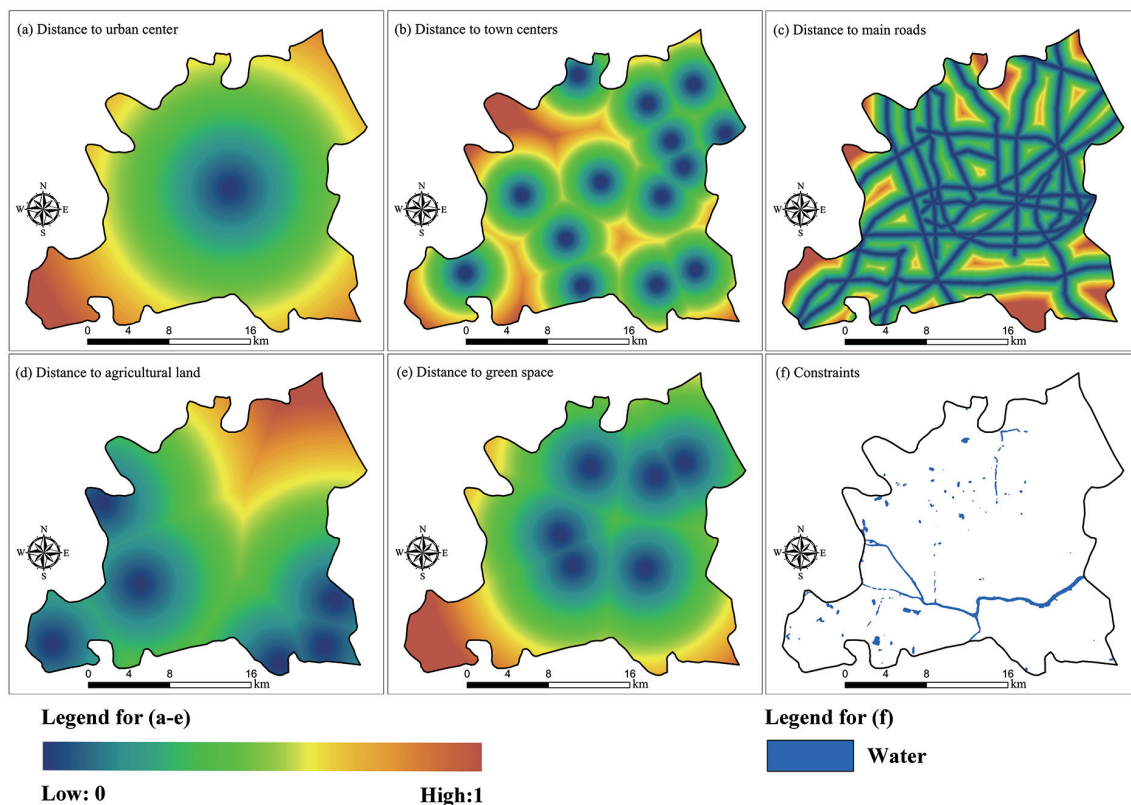
**Figure 1.** The study area of Songjiang district in Shanghai, China. (a) Map of P. R. China and (b) Map of Shanghai.

### 2.2. Input Variables

Nine factors affecting land use change were chosen to model urban growth in Songjiang from 1992 to 2008. These factors were distance-based variables, neighborhood, constraints, and a stochastic factor (Table 1); all are closely related to urban development and land use changes [2,41,42]. We then visualized the spatial variables and constraints in ArcGIS and produced them as input layers for the PLS-CA model (Figure 2).

**Table 1.** The spatial variables used to simulate urban growth in the partial least squares-based cellular automata (PLS-CA) model.

Variable	Meaning	Type	Acquisition Method
$y$	Conversion probability	Criterion variable	Remote sensing classification
$D_{urban}$ $D_{town}$ $D_{mrd}$ $D_{agri}$ $D_{gs}$	Distance to urban center Distance to town centers Distance to main roads Distance to agricultural land Distance to green space	Spatial variable	Euclidean Distance tool in ArcGIS
Neighborhood	3 × 3 neighborhood	Local variable	Retrieved dynamically during simulation
Constraints	Local constraints Global constraints		
Stochastic	Stochastic factors	Global variable	Assigned randomly



**Figure 2.** Visualization of spatial variables used in the PLS-CA model. (a)  $D_{urban}$ ; (b)  $D_{town}$ ; (c)  $D_{mrd}$ ; (d)  $D_{agri}$ ; (e)  $D_{gs}$ ; and (f) Constraint.

Topographic data play an important role in generating spatial variables for CA models. As an example, it is sometimes difficult to convert rural land on a steep slope into urban use.

As a result, a slope factor should be included in any credible model. However, the Songjiang study area lies on a very flat land in the Yangtze River Delta [43], and, therefore, the impact of slope can be omitted in the modeling. Distance-based variables and neighborhood reflect the agglomerative effect of urban development and the attractive power of infrastructure [44]. Spatial variables used in the PLS-CA model can be categorized as positive and negative distances. Positive distances include distances to urban center, town centers, and main roads; these factors are significant “push” forces to urban growth. Conversely, the negative distances, such as distances to agricultural land and green space, yield a “repellent” effect on urban development.

Apart from the aforementioned quantifiable factors, there are still many uncertainties and errors in modeling urban growth, resulting in the departure of actual urban growth from some well-known trajectories. Some of these uncertainties are intangible and can be difficult to identify and/or quantify. To represent these uncertainties, a stochastic factor was introduced into our CA model (Table 1). The real values of these spatial variables were acquired from both remotely sensed imagery and vector maps. The conversion probability ( $y$ ) was calculated by detecting land use change using the thematic mapper (TM) images from 1992 to 2008.

### 3. The PLS-CA Model

#### 3.1. A Generic CA Model

The global conversion probability of land conversion from non-urban to urban can be calculated as the combined effect of the static probability, neighborhood effect, constraints, and random impact [9,45]. A general form of the global conversion probabilities for  $u \times v$  cells (in a lattice) is:

$$\left[ P_{ij}^t \right]_{u \times v} = \left[ P_d \times \text{con}(S_{ij}^t = \text{suitable}) \times P_{\Omega,ij}^t \times \left( 1 + (-\ln(Rnd))^{\beta} \right) \right]_{u \times v} \quad (1)$$

where  $P_{ij}^t$  is the global probability of rural-to-urban conversion for cell  $ij$  at time  $t$ ;  $P_d$  is the static probability determined by spatial distances [11,34];  $\text{con}()$  is a constraint function which returns either 0 or 1 [46];  $P_{\Omega,ij}^t$  is the effect for cell  $ij$  at time  $t$  within  $\Omega_l \times l$  neighborhood and it is calculated by  $P_{\Omega,ij}^t = \frac{\sum_l \times l \text{con}(S_{ij} = \text{urban})}{l \times l - 1}$  where  $\text{con}(S_{ij}^t = \text{suitable})$  returns 1 if the state of the cell  $ij$  is urban, otherwise, it returns 0;  $\left( 1 + (-\ln(Rnd))^{\beta} \right)$  is the stochastic factor [47], where  $Rnd$  is a random real number ranging from 0 to 1, and  $\beta$  is a parameter ranging from 0 to 10 that adjusts the influence of the stochastic factor.

The global conversion probability, therefore, consists of: (1) the conversion probability based on spatial variables, (2) cell conversion constraints including planning regulation, protected farmland, and water bodies, (3) neighborhood effects, and (4) a stochastic factor. The first component is the observed conversion probability  $P_d$  [18,41]:

$$\left[ P_d \right]_{u \times v} = \left[ \frac{1}{1 + \exp(-(\alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p))} \right]_{u \times v} \quad (2)$$

where  $\alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p$  represents the comprehensive impacts of distance-based variables on cell  $ij$ ,  $x_i (i = 1, \dots, p)$  are the distances from the cell  $ij$  to a key point such as the urban center, town centers, main roads, etc.; and  $a_i (i = 0, 1, \dots, p)$  are their corresponding parameters. These distances are also defined as spatial or independent variables in our CA modeling.

#### 3.2. The PLS Method

We assume that  $\mathbf{y} = (y_1, \dots, y_q)_{n \times q}$  is a set of dependent variables (i.e., the observed rural-to-urban conversion), where  $n$  is the size of samples and  $q$  is the number of dependent variables,  $\mathbf{x} = (x_1, \dots, x_p)_{n \times p}$  is a set of independent variables with  $p$  as the number of independent

variables. We also assume that  $E = (E_{01}, \dots, E_{0p})_{n \times p}$  and  $F = (F_{01}, \dots, F_{0p})_{n \times p}$  are the normalized (mean-centered and variance-scaled) matrix forms of  $x$  and  $y$ , respectively,  $t_1$  is the first principal component vector of  $E_0$ , i.e.,  $t_1 = E_0 w_1$ ,  $w_1$  is the corresponding unit weight vector of  $E_0$  and  $\|w_1\| = 1$ , and that  $u_1$  is the first principal component vector of  $F_0$ , i.e.,  $u_1 = F_0 c_1$ ,  $c_1$  is the corresponding unit weight vector of  $F_0$  and  $\|c_1\| = 1$ .

In PLS regression, the goal is to obtain a first pair of vectors  $t_1 = E_0 w_1$  and  $u_1 = F_0 c_1$  under the condition that  $\|w_1\| = 1$  and  $\|c_1\| = 1$ , and maximizing  $t_1^T u_1$ . The objective can be re-written as an optimization problem [36,37]:

$$\begin{cases} \max_{w_1, c_1} \langle E_0 w_1, F_0 c_1 \rangle \\ \text{subject to} \begin{cases} w_1^T w_1 = 1 \\ c_1^T c_1 = 1 \end{cases} \end{cases} \quad (3)$$

By applying the Lagrange algorithm, we obtained eigenvalue equations resolving a first pair of weight vectors  $w_1$  and  $c_1$  as follows:

$$\begin{cases} E_0^T F_0 F_0^T E_0 w_1 = \theta_1^2 w_1 \\ F_0^T E_0 E_0^T F_0 c_1 = \theta_1^2 c_1 \end{cases} \quad (4)$$

where  $w_1$  and  $c_1$  are the unit eigenvectors of the matrices  $E_0^T F_0 F_0^T E_0$  and  $F_0^T E_0 E_0^T F_0$ , respectively,  $\theta_1^2$  is the corresponding eigenvalue, and  $\theta_1 = F_0^T E_0^T F_0 c_1$ . According to Equation (1),  $\theta_1$  is supposed to be maximal in the sense of PLS regression.

We compute the first pair of component vectors  $t_1 = E_0 w_1$  and  $u_1 = F_0 c_1$ , and run the regression of  $E_0$  and  $F_0$  with respect to  $t_1$  and  $u_1$ , respectively. The equation is:

$$\begin{cases} E_0 = t_1 p_1^T + E_1 \\ F_0 = t_1 r_1^T + F_1 \end{cases} \quad (5)$$

where  $E_1$  and  $F_1$  are the residual matrices, and  $p_1$  and  $r_1$  are the coefficient vectors that can be given by:

$$\begin{cases} p_1 = \frac{E_0^T t_1}{\|t_1\|^2} \\ r_1 = \frac{F_0^T t_1}{\|t_1\|^2} \end{cases} \quad (6)$$

Substituting the residual matrices  $E_1$  and  $F_1$  for  $E_0$  and  $F_0$  and repeating the above method, we obtained the second component vectors  $t_2$  and  $u_2$  as:

$$\begin{cases} t_2 = E_1 w_2 \\ u_2 = F_1 c_2 \\ \theta_2 = \langle t_2, u_2 \rangle = w_2^T E_1^T F_1 c_2 \end{cases} \quad (7)$$

where  $w_2$  and  $c_2$  are the unit eigenvectors of matrices  $E_1^T F_1 F_1^T E_1$  and  $F_1^T E_1 E_1^T F_1$ , respectively, corresponding to the maximum eigenvalue  $\theta_2^2$ .

Running the regression of  $E_1$  and  $F_1$  with respect to  $t_2$  and  $u_2$ , respectively, we have:

$$\begin{cases} E_1 = t_2 p_2^T + E_2 \\ F_1 = t_2 r_2^T + F_2 \end{cases} \quad (8)$$

where the coefficient vectors  $p_2$  and  $r_2$  are calculated from:

$$\begin{cases} p_2 = \frac{E_1^T t_2}{\|t_2\|^2} \\ r_2 = \frac{F_1^T t_2}{\|t_2\|^2} \end{cases} \quad (9)$$

The procedure is iterated until  $E_0$  becomes a null matrix, and the final components  $t_i (i = 1, \dots, m)$  are determined by cross-validation. Therefore, we have the following equations:

$$\begin{cases} E_0 = t_1 p_1^T + \dots + t_m p_m^T \\ F_0 = t_1 r_1^T + \dots + t_m r_m^T + F_m \end{cases} \quad (10)$$

Since  $t_1, \dots, t_m$  can be represented as the linear combination of the original variables  $E_{01}, \dots, E_{0p}$ , and  $F_0$  in Equation (10) is recovered by the regression equation of  $y_j^* = F_{0k} (k = 1, \dots, q)$  with respect to  $x_j^* = E_{0j} (j = 1, \dots, p)$  as follows:

$$y_k^* = \alpha_{k_1} x_1^* + \dots + \alpha_{k_p} x_p^* + F_{mk} (k = 1, \dots, q) \quad (11)$$

where  $\alpha_{k_1}, \dots, \alpha_{k_p}$  are the corresponding coefficients and  $F_{mk}$  is the  $k^{\text{th}}$  column of residual matrix  $F_m$ .

Cross-validation checks the contributions of the extracted principal components to determine how well the regression model predicts the data. The cross-validation for the component  $t_n$  is:

$$Q_h^2 = 1 - \frac{PRESS_h}{SS_{h-1}} \quad (12)$$

where  $PRESS_h$  is the sum of squares of prediction error with a total of  $h$  components ( $t_1, \dots, t_h$ ), and  $SS_{h-1}$  is the sum of squares of combination error of  $y$  with the first  $(h-1)$  components ( $t_1, \dots, t_{h-1}$ ).

If  $\frac{PRESS_h}{SS_{h-1}} \leq 0.95^2$ , the contribution margin of the newly added component  $t_n$  is significant, and as a result, iteration stops when  $Q_h^2 \geq 0.00975$  [36,37].

### 3.3. PLS-Based CA Model

Since the conversion probability of each cell in CA is a single decimal variable, Equation (11) can be re-written as [36,37]:

$$\hat{y} = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p \quad (13)$$

where  $\alpha_i (i = 0, 1, \dots, p)$  is the  $i^{\text{th}}$  regression estimator.

The form of Equation (13) retrieved by PLS method is similar to that from PCA method but the regression estimator  $\alpha_p$  obtained from the PLS method contains information about the dependent variable  $y$  as shown in Equation (7), while  $\alpha_p$  obtained by PCA does not contain any contribution of the responsive variable  $y$  [34,36,37]. Although data redundancy can be eliminated by PCA, the regression estimators obtained are not related to the independent variables and, thus, have less strong ability to interpret the independent variable  $y$ . PLS is more robust than PCA at explaining the responsive variable  $y$ .

Integrating Equations (1), (2) and (13), we derived the global conversion probability in the PLS-CA model:

$$\left[ P_{ij}^t \right]_{u \times v} = \left[ \frac{1}{1 + \exp[-(\alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p)]} \times \frac{\sum_{l \times l} \text{con}(S_{ij} = \text{urban})}{l \times l - 1} \times \text{con}(cell_{ij}^t = \text{suitable}) \times (1 + (-\ln(\text{Rnd}))^\beta) \right]_{u \times v} \quad (14)$$

If the calculated global probability  $P_{ij}^t$  exceeds the predefined threshold ranging from 0 to 1, the cell  $ij$  at time  $t$  will be converted to urban land use at time  $t + 1$ . Otherwise, it will retain its current state at next time  $t + 1$  [18,41].

### 3.4. Structure of the PLS-CA Model

The PLS-CA model workflow consists of five steps: raw data collection, data processing, CA rule discovery with PLS, determination of other CA factors, and model implementation and results assessment (Figure 3). Each step of the model plays a distinct role in the modeling as follows:

(1) Raw data collection: Data used in the model include historical raster images such as remotely sensed images, an administrative vector map, a topographic map, and a transportation map.

(2) Data processing: Spatial variables were extracted from raw data using the ArcGIS Spatial Analyst tool. These spatial variables included the distance to the urban center ( $D_{urban}$ ), town centers ( $D_{town}$ ), main roads ( $D_{mrd}$ ), agricultural land ( $D_{agri}$ ), and green space ( $D_{gs}$ ). The five spatial variables were normalized by:

$$D_{norm} = \frac{D_{ori}}{D_{max}} \tag{15}$$

where  $D_{max}$  is the maximum value of the spatial variable,  $D_{ori}$  is the original distance value from the raw data, and  $D_{norm}$  is the normalized value in the range (0, 1). Normalization enables a precise interpretation of the geographic meaning of the parameters. For instance, if a cell is situated at the urban center, its normalized  $D_{urban}$  value will be 0, where if the cell is situated far from the urban center, its normalized  $D_{urban}$  will approach 1.

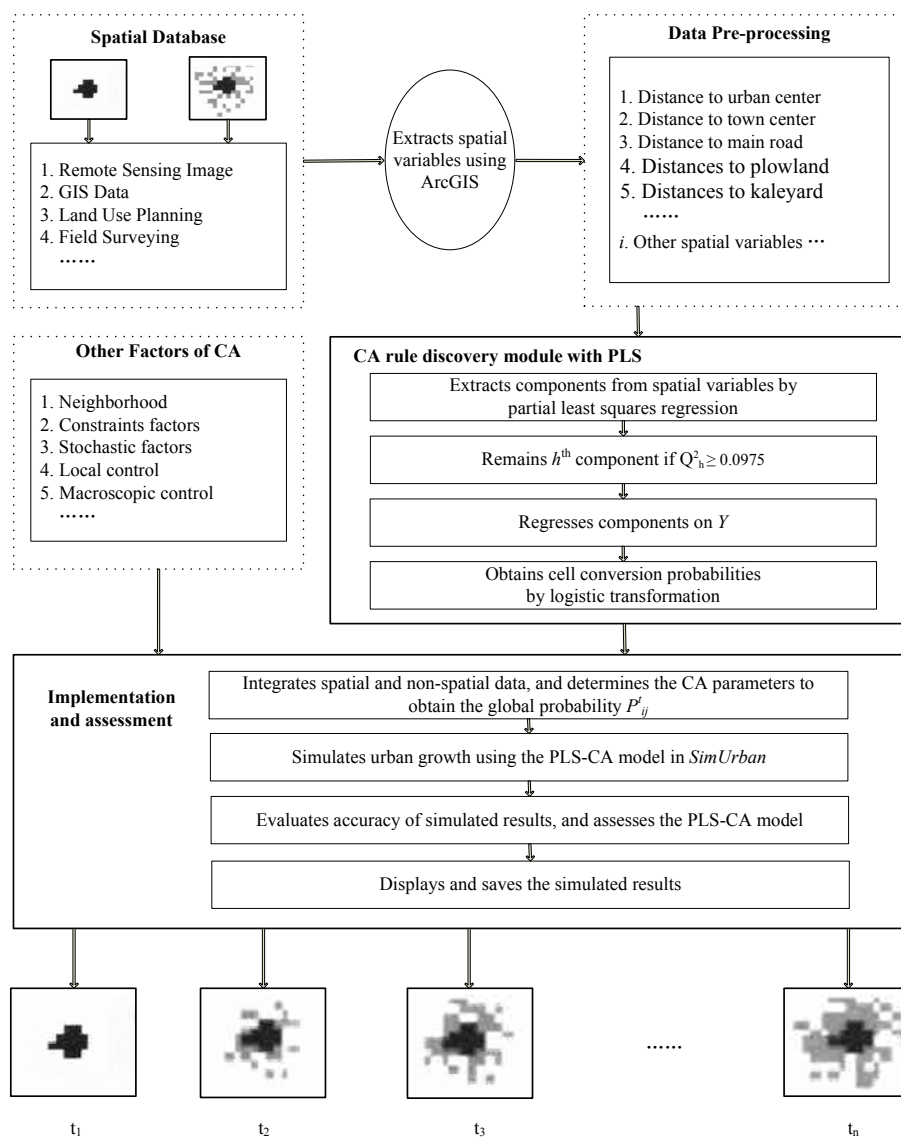


Figure 3. Structure of the PLS-CA model.

(3) CA rule discovery: This module derives uncorrelated spatial components using PLS. It determines whether the derived spatial variables satisfy the cross-validation of  $Q^2_h \geq 0.0975$  and,



hence, it is used to define CA parameters (i.e., weights of spatial variables) by which the land conversion probability  $P_d$  under variables can be obtained. The PLS regression was conducted using the “PLSR” package of R-language [48].

(4) Other CA factors: These include non-spatial factors such as neighborhood effect, constraints of basic farmland, and a stochastic factor.

(5) PLS-CA implementation and assessment: This module enables the simulation of the PLS-CA model and incorporates simulation accuracy assessment by generating overall accuracy, producer’s accuracy, user’s accuracy, Kappa coefficient, and the compared urban growth rate (*CUGR*). The module also displays and exports simulation outcomes.

The simulated area of each category from the CA modeling was not exactly equal to the actual area. Therefore, an indicator termed the compared urban growth rate (*CUGR*) is calculated to assess the accuracy of the PLS-CA model by comparing the observed and simulated urban growth rates. The *CUGR* indicator was computed as:

$$CUGR = \frac{S_{sim2008}}{S_{obs2008}} \times 100\% \quad (16)$$

where *CUGR* is the difference between the observed and simulated areas of each category in terms of growth rate,  $S_{sim2008}$  is the simulation area of the urban or non-urban category at 2008, and  $S_{obs2008}$  is the statistical areas of observed urban growth in 2008 or non-urban loss in 1992, respectively.

## 4. Results and Discussion

### 4.1. Assessment of Correlation

A total of 5000 samples were randomly selected from spatial variables and the classified land use patterns in 1992 and 2008 to determine the CA transition rules. The correlation matrix of spatial variables was calculated using the samples (Table 2), showing significant correlations among these spatial variables. Traditional methods, such as multi-criteria evaluation (MCE) technique and logistic regression, are not able to avoid the negative effects of multicollinearity and are relatively weak in providing correct weights for the variables. We, therefore, applied PLS to extract the uncorrelated principal components from spatial variables to achieve more reasonable CA transition rules and improve the performances of the CA model.

**Table 2.** Correlation matrix of spatial variables.

Variable	$D_{town}$	$D_{mrd}$	$D_{agri}$	$D_{gs}$	$y$
$D_{urban}$	0.6462	0.6109	−0.5372	−0.5280	0.2264
$D_{town}$		0.7058	−0.4496	−0.7754	0.5635
$D_{mrd}$			−0.4714	−0.4971	0.2518
$D_{agri}$				0.8537	−0.1572
$D_{gs}$					−0.1893

### 4.2. CA Transition Rules

Among the three components acquired by PLS regression, only the first two satisfied the cross-validation requirement but explained more than 95% of the total variance (Table 3). The first component is mainly related to urban center, and the second component is principally related to main roads. For the third component, its  $Q_h^2$  is less than the critical value and, therefore, it is not a valid component. By comparison, PCA can reduce data redundancy but it extracts exactly five components for the same samples used in this research. Its first three explained 84.268% of the variance, lower than that of the PLS regression. Based on PLS regression, suitable weights for CA models can be easily defined, since the principal components are independent, avoiding the repeated counting that may occur in general MCE [21]. The CA parameters acquired by logistic regression are quite

different with  $D_{mrd}$  (1.7590) being very large and  $D_{town}$  (0.5846) relatively small (Table 4) as compared with those in the PLS regression. This indicates that the logistic-CA model over-weights  $D_{mrd}$  but undervalues  $D_{urban}$ . In contrast, PLS regression generated CA parameters that more reasonably reflect the actual urban growth in Songjiang. In the PLS approach, the negative weights for the distance factors are:  $D_{urban}$  (−1.1063),  $D_{mrd}$  (−0.8274),  $D_{town}$  (−0.5841), followed by the positive weights of  $D_{agri}$  and  $D_{gs}$  which reflect the factors that tend to prevent non-urban land from being developed. The land conversion potential was produced as map layers based on the calibrated logistic regression and PLS methods (Figure 4), which varied from 0.38 to 0.72 for logistic-CA and from 0.34 to 0.70 for PLS-CA.

Table 3. Principal components derived from PLS.

Component	Cross-Validation			Spatial Variables				
	R	$Q^2_h$	Critical Value	Urban Center	Town Center	Main Road	Agricultural Land	Green Space
1	0.8625	0.8517	−0.0975	−0.8156	−0.4744	0.3012	−0.2290	−0.2967
2	0.3240	0.1054	−0.0975	−0.3694	−0.2536	−0.5515	0.6557	0.3553
3	0.1572	−0.0092	−0.0975	−0.0916	0.1228	0.2031	0.1232	0.0802

Table 4. Comparison of CA parameters generated by PLS and logistic regression.

Models	Variable				
	$D_{urban}$	$D_{town}$	$D_{mrd}$	$D_{agri}$	$D_{gs}$
PLS-CA	−1.1063	−0.5841	−0.8274	0.1924	0.1513
logistic-CA	−0.5846	−0.2837	−1.7590	2.0263	1.5978

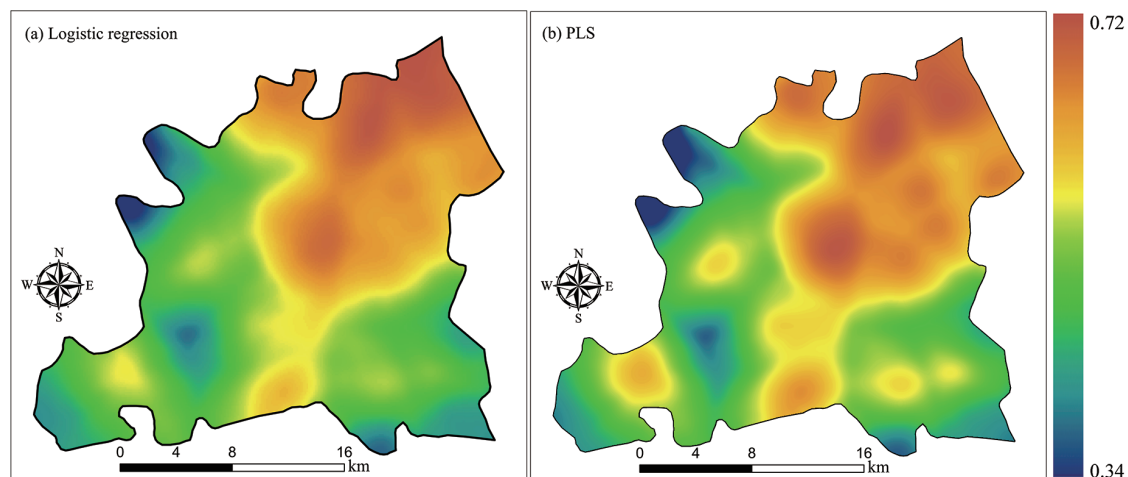


Figure 4. Land conversion potential based on spatial variables. (a) Logistic regression and (b) PLS.

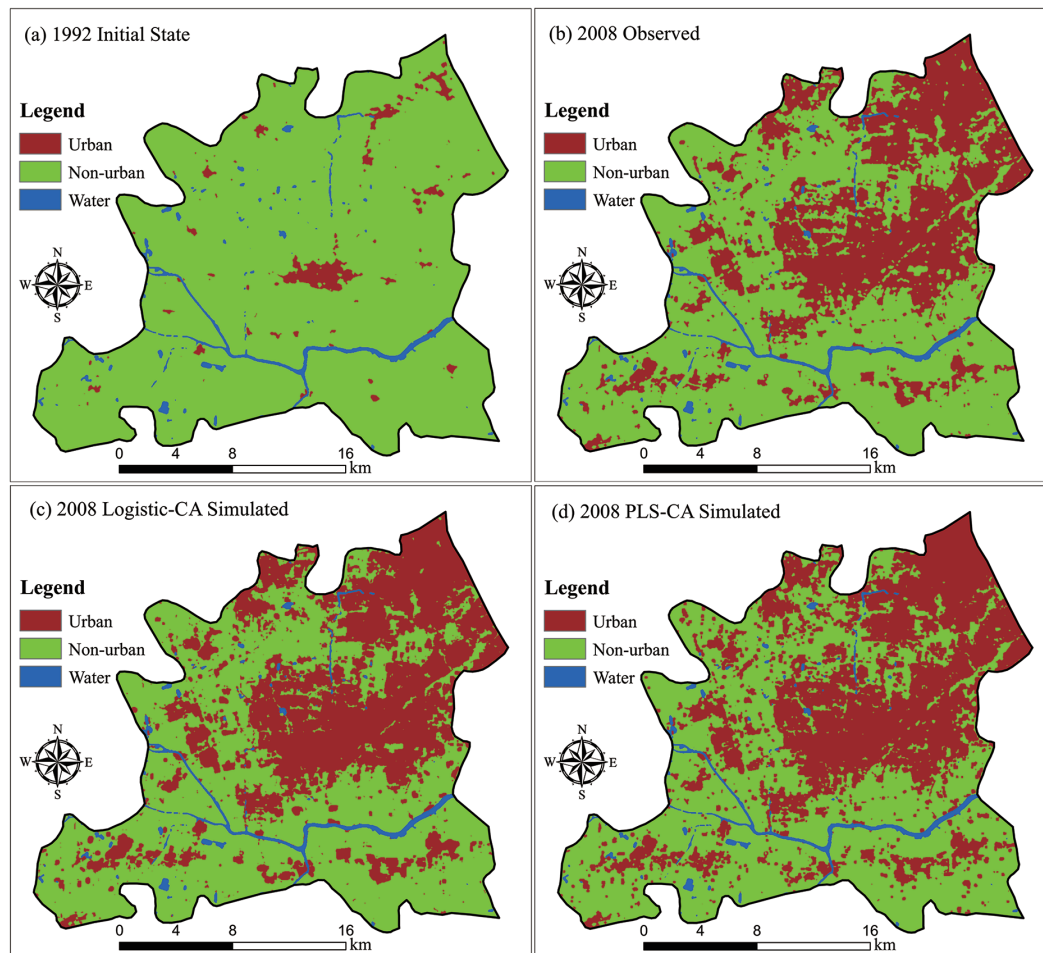
#### 4.3. Simulation Results

The PLS-CA model was applied to simulate urban growth of Songjiang from 1992 to 2008 (Figure 5). In the simulation, land use types were generalized as urban, non-urban, and water body.

Before running the model, the best combination of threshold value  $P_{thd}$  and the number of iterations was determined for the calibration of the PLS-CA model. The meaning of each iteration should also be defined. As an initial trial, a  $P_{thd}$  value of 0.40 was used to test if a non-urban cell can be converted to an urban cell. By running the model, simulated results that approximated the actual urban growth were realized within a certain number of iterations. For the next trial,  $P_{thd}$  was increased by 0.02 and a simulation result with the highest overall accuracy was acquired with another number of iterations.  $P_{thd}$  increased by 0.02 from 0.40 to 0.80, indicating that there were

21 trials for this model. After comparing the results of all trials with different  $P_{thd}$  values, the PLS-CA model generated the highest overall simulation accuracy of 85.8% at  $P_{thd} = 0.68$  and Iteration = 16. By comparison, a logistic-CA model was also calibrated with the best overall simulation accuracy of 83.5% at  $P_{thd} = 0.66$  and Iteration = 16.

Visual inspection demonstrates a good match between the observed and simulated patterns, for both the logistic-CA and PLS-CA models (Figure 5). Further, all three observed and simulated urban patterns in 2008 show that urban growth of Songjiang occurred around the urban centers and northeastern areas in the late 1990s to early 2000s.



**Figure 5.** The observed and simulated patterns in Songjiang. (a) The 1992 initial state; (b) The 2008 observed pattern; (c) The 2008 simulated pattern by logistic-CA; and (d) The 2008 simulated pattern by PLS-CA.

#### 4.4. Accuracy Analysis

To quantitatively evaluate the simulation accuracy and the performance of the PLS-CA model, a pixel-by-pixel comparison was used to calculate a confusion matrix about the concordance between the simulated results and the observed pattern [11,21,49,50]. The reference land use map illustrating the observed urban growth was the classified result using a supervised minimum distance classifier in ENVI 5.2. The confusion matrix derived from the simulated results and the observed reference map was produced for comparison (Table 5). Kappa coefficients were also calculated to quantify their actual degree of agreement [51,52].

User's accuracy was 72.9% for non-urban areas and 96.8% for urban areas in 2008, while the producer's accuracies for non-urban and urban categories were 95.2% and 80.7%, respectively (Table 5).

The user's accuracy of urban informs that, of all the observed cells considered as urban in the classified patterns, 96.8% were correctly classified in the simulated pattern, and the probability of identification of urban mislabeled as non-urban (i.e., commission error) was 3.2% [53]. For the same urban category in Table 5, of all the urban cells in the simulation result, 80.7% actually correspond to urban in the classified pattern. In other words, the probability of identification of a cell erroneously labeled as urban category (i.e., an omission error) was 19.3%. The overall accuracy shows that 85.8% of all the cells under assessment were correctly categorized in the simulation result. The Kappa coefficient means that the simulation achieved an accuracy that was 70.9% better than what would be expected from the chance assignment of cells to categories.

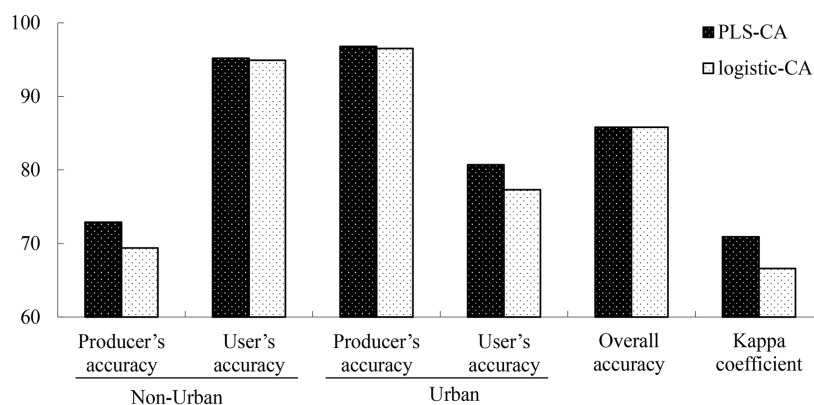
**Table 5.** Confusion matrix between remote sensing-based classification and simulated urban pattern using the PLS-CA model for Songjiang in 2008.

Item	Observed (%)			
	Urban	Non-Urban	Total	
Simulated (%)	Urban	33.6	12.5	46.1
	Non-urban	1.7	52.2	53.9
	Total	35.3	64.7	100
User's Accuracy			Commission error	
Non-urban	= 33.6/46.1 = 72.9%		27.1%	
Urban	= 52.2/53.9 = 96.8%		3.2%	
Producer's Accuracy			Omission error	
Non-urban	= 33.6/35.3 = 95.2%		4.8%	
Urban	= 52.2/64.7 = 80.7%		19.3%	
Overall accuracy		85.8%		
Kappa coefficient		70.9%		

#### 4.5. Discussion

The detailed simulation accuracies were calculated for Songjiang in 2008 for the logistic-CA and PLS-CA models for Songjiang (Figure 6). The user's accuracy for non-urban generated by the PLS-CA model was 3.5% greater than that of the logistic-CA model, and the producer's accuracy for non-urban from the PLS-CA model was nearly equal to that of the logistic-CA model (94.9%). For the urban category, the user's and producer's accuracies of the logistic-CA model were 96.5% and 77.3%, respectively, lower than those of the PLS-CA model. The overall accuracy of the simulation results in 2008 was 83.5% for the logistic-CA model, 2.3% less than the new PLS-CA model. The Kappa coefficient of the PLS-CA model was 70.9%, which outperforms the logistic-CA model (66.6%). The comparison suggests that the PLS-CA model generated more accurate results compared to the logistic-CA model.

The *CUGR* indicator illustrates the growth rate of the simulated urban area compared with the actual urban development. A *CUGR* for urban areas larger 100% indicates that the simulated growth exceeds the observed growth; otherwise, the simulation growth is slower than the observed growth. A *CUGR* value approaching 100% suggests that the CA model performs well in terms of the area control. Statistics for observed urban growth (excluding water bodies) were retrieved from remote sensing, corrected by the data from the local government of Songjiang. Logistic-CA and PLS-CA models both simulated more urban growth than actually occurred (Table 6). The *CUGR* of the urban category was 114.2% for the PLS-CA model, lower than that of the logistic-CA model (121.1%). The *CUGR* of the non-urban was 92.3% for the PLS-CA model, which is closer to 100% than that for the logistic-CA model (88.6%). This result suggests that the overall area control performance of the PLS-CA model was better than that of the logistic-CA model, while still overestimating urban growth and underestimating the persistence of non-urban.



**Figure 6.** Simulation accuracy (%) of the two CA models in 2008.

**Table 6.** Observed and simulated urban growth rates from 1992–2008 (CUGR stands for compared urban growth rate).

Urban Growth		Urban	Non-Urban
Observed	Area 1992 (km <sup>2</sup> )	17.9	565.1
	Area 2008 (km <sup>2</sup> )	205.2	377.8
logistic-CA	Area 2008 (km <sup>2</sup> )	248.4	334.6
	CUGR (%)	121.1	88.6
PLS-CA	Area 2008 (km <sup>2</sup> )	234.3	348.7
	CUGR (%)	114.2	92.3

Urban development is a complex open system whose trajectory is affected by drivers that may be significantly, spatially correlated. Integrated with the analytical functions of GIS, logistic regression can be used to evaluate the impact of these factors on urban growth. However, logistic regression cannot eliminate the correlation of spatial variables, while PCA can eliminate spatial correlation only to a certain degree, and the principal components found in the independent variables may not adequately explain the dependent variables [35]. The proposed PLS method can extract variables that are uncorrelated from amongst the explanatory variables, and also between the explanatory and response factors [36,37], resulting in the discovery of transition rules from a number of driving factors that are usually highly correlated. This relationship could explain why PLS-CA modeling outperformed the traditional logistic-CA model, at least from a theoretical point of view. Our results show that the simulated patterns of urban growth accord well with the actual urban pattern of Songjiang. Compared with the logistic-CA model, the PLS-CA model achieved better simulation accuracies in modeling the urban growth of Songjiang through time. We speculate that our model is better than the PCA-based CA model as inferred by the principal components, but not necessarily better than auto-logistic regression which performs nearly as well as ANN [33]. It still needs to be tested whether the PLS-CA model is more or less accurate than other CA models based on artificial intelligence and machine learning. However, our new model poses the advantages that it is simpler than these models and generates parameters having clear physical meanings.

In addition, CA models contain various types of uncertainty caused by several other factors such as sampling, neighborhood configuration, constraints, stochastic perturbation, and spatial scale [19,47,54,55]. We took only one group of samples for training the PLS-CA model in this study. Like any other CA models, our PLS-CA could be sensitive to samples that are determined by both sampling method and sample grouping [9]. The effect of sampling on the simulation results is reflected by the CA parameters of drivers. Such an effect is relatively greater for statistically significant drivers, whereas it is much smaller for statistically non-significant drivers which can even be excluded in modeling [32,56]. Neighborhood configuration such as the shape and number of neighbors influences the CA models by local interactions [19,57]. A stochastic factor in CA transition rules is to simulate less tangible uncertainties and perturbations that may affect the simulation results [11,42,47]. Moreover,

raster-based CA models are also sensitive to cell size (grain size) in terms of simulation accuracy and landscape structure [39,55,58,59]. The proposed PLS-CA model is no exception because it depends on a rasterized space.

## 5. Conclusions

This paper demonstrates that CA models can accurately simulate urban growth using global and local constraints that reflect various environmental concerns. The advantages of urban growth simulation by GIS-based CA modeling include the identification of the driving factors of land use change and the identification of spatial patterns across space and over time. The most important part of developing these new models is to discover mature CA transition rules. Our PLS-CA model is capable of extracting uncorrelated factors from the candidate explanatory variables. Thus, PLS-CA can well eliminate redundancy of the input data and, thus, allow for the discovery of better and more reasonable transition rules. The PLS-CA model was successfully applied to simulate the urban growth in Songjiang, demonstrating better simulation accuracy than a conventional logistic-CA model.

Further improvements could be made by testing the response and robustness of the PLS-CA model on sampling, neighborhood configuration, constraints, and spatial scale. In addition, advanced CA models could be packaged with simple, robust, and easily implementable modules such as real-time and dynamic display of simulation results.

**Acknowledgments:** We thank Professor Xiaohua Tong at Tongji University for his assistance in writing the manuscript. This research was supported by the National Natural Science Foundation of China (41406146), Natural Science Foundation of Shanghai Municipality (13ZR1419300), and Shanghai Universities First-class Disciplines Project-Fisheries (A).

**Author Contributions:** Yongjiu Feng conceived and developed the model and wrote the paper. All authors analyzed the data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mondal, B.; Das, D.N.; Bhatta, B. Integrating cellular automata and Markov techniques to generate urban development potential surface: A study on Kolkata agglomeration. *Geocarto. Int.* **2016**. [[CrossRef](#)]
2. Feng, Y.; Liu, Y.; Batty, M. Modeling urban growth with GIS based cellular automata and least squares SVM rules: A case study in Qingpu–Songjiang area of Shanghai, China. *Stoch. Env. Res. Risk Assess.* **2016**, *30*, 1387–1400. [[CrossRef](#)]
3. Barredo, J.I.; Kasanko, M.; McCormick, N.; Lavalle, C. Modelling dynamic spatial processes: Simulation of urban future scenarios through cellular automata. *Landsc. Urban Plan.* **2003**, *64*, 145–160. [[CrossRef](#)]
4. Jantz, C.A.; Goetz, S.J.; Shelley, M.K. Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore-Washington metropolitan area. *Environ. Plan. B* **2004**, *31*, 251–271. [[CrossRef](#)]
5. Tobler, W. Cellular geography. In *Philosophy in Geography*; Springer: Berlin, Germany, 1979; pp. 379–386.
6. Batty, M.; Xie, Y.; Sun, Z. Modeling urban dynamics through GIS-based cellular automata. *Comput. Environ. Urban* **1999**, *23*, 205–233. [[CrossRef](#)]
7. Verburg, P.H.; Schot, P.P.; Dijst, M.J.; Veldkamp, A. Land use change modelling: Current practice and research priorities. *GeoJournal* **2004**, *61*, 309–324. [[CrossRef](#)]
8. Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals*; The MIT Press: Cambridge, MA, USA, 2007.
9. Clarke, K.C.; Gaydos, L.J. Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *Int. J. Geogr. Inf. Sci.* **1998**, *12*, 699–714. [[CrossRef](#)] [[PubMed](#)]
10. Li, X.; Yeh, A.G.-O. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 131–152. [[CrossRef](#)]
11. Wu, F. Calibration of stochastic cellular automata: The application to rural-urban land conversions. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 795–818. [[CrossRef](#)]

12. Cao, K.; Batty, M.; Huang, B.; Liu, Y.; Yu, L.; Chen, J. Spatial multi-objective land use optimization: Extensions to the non-dominated sorting genetic algorithm-II. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1949–1969. [[CrossRef](#)]
13. Cao, M.; Bennett, S.J.; Shen, Q.; Xu, R. A bat-inspired approach to define transition rules for a cellular automaton model used to simulate urban expansion. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1–19. [[CrossRef](#)]
14. Feng, Y.; Liu, Y. Scenario prediction of emerging coastal city using CA modeling under different environmental conditions: A case study of Lingang New City, China. *Environ. Monit. Assess.* **2016**, *188*, 540. [[CrossRef](#)] [[PubMed](#)]
15. Liu, X.; Ma, L.; Li, X.; Ai, B.; Li, S.; He, Z. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 148–163. [[CrossRef](#)]
16. Liu, Y.; Feng, Y. Simulating the impact of economic and environmental strategies on future urban growth scenarios in Ningbo, China. *Sustainability* **2016**, *8*, 1045. [[CrossRef](#)]
17. Liu, X.; Li, X.; Shi, X.; Zhang, X.; Chen, Y. Simulating land-use dynamics under planning policies by integrating artificial immune systems with cellular automata. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 783–802. [[CrossRef](#)]
18. Liu, Y.; Feng, Y. A logistic based cellular automata model for continuous urban growth simulation: A case study of the Gold Coast City, Australia. In *Agent-based Models of Geographical Systems*; Springer: Berlin, Germany, 2012; pp. 643–662.
19. Liao, J.; Tang, L.; Shao, G.; Qiu, Q.; Wang, C.; Zheng, S.; Su, X. A neighbor decay cellular automata approach for simulating urban expansion based on particle swarm intelligence. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 720–738. [[CrossRef](#)]
20. Verstegen, J.A.; Karssenber, D.; Van Der Hilst, F.; Faaij, A.P. Identifying a land use change cellular automaton by Bayesian data assimilation. *Environ. Model. Softw.* **2014**, *53*, 121–136. [[CrossRef](#)]
21. Li, X.; Yeh, A.G.-O. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 323–343. [[CrossRef](#)]
22. Feng, Y.; Liu, Y. A heuristic cellular automata approach for modelling urban land-use change based on simulated annealing. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 449–466. [[CrossRef](#)]
23. Liu, Y.; Tang, W.; He, J.; Liu, Y.; Ai, T.; Liu, D. A land-use spatial optimization model based on genetic optimization and game theory. *Comput. Environ. Urban* **2015**, *49*, 1–14. [[CrossRef](#)]
24. Kamusoko, C.; Gamba, J. Simulating urban growth using a Random Forest-Cellular Automata (RF-CA) model. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 447–470. [[CrossRef](#)]
25. Triantakou, D.; Mountrakis, G. Urban growth prediction: A review of computational models and human perceptions. *J. Geogr. Inf. Syst.* **2012**, *4*, 263–273. [[CrossRef](#)]
26. Cao, K.; Huang, B.; Li, M.; Li, W. Calibrating a cellular automata model for understanding rural–urban land conversion: A Pareto front-based multi-objective optimization approach. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1028–1046. [[CrossRef](#)]
27. Feng, Y.; Liu, Y. An optimised cellular automata model based on adaptive genetic algorithm for urban growth simulation. In *Advances in Spatial Data Handling and GIS*; Springer: Berlin, Germany, 2012; pp. 27–38.
28. Feng, Y.; Liu, Y.; Tong, X.; Liu, M.; Deng, S. Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landsc. Urban Plan.* **2011**, *102*, 188–196. [[CrossRef](#)]
29. Guan, D.; Li, H.; Inohae, T.; Su, W.; Nagaie, T.; Hokao, K. Modeling urban land use change by the integration of cellular automaton and Markov model. *Ecol. Model.* **2011**, *222*, 3761–3772. [[CrossRef](#)]
30. Yang, X.; Zheng, X.-Q.; Lv, L.-N. A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecol. Model.* **2012**, *233*, 11–19. [[CrossRef](#)]
31. Munshi, T.; Zuidgeest, M.; Brussel, M.; van Maarseveen, M. Logistic regression and cellular automata-based modelling of retail, commercial and residential development in the city of Ahmedabad, India. *Cities* **2014**, *39*, 68–86. [[CrossRef](#)]
32. Alqurashi, A.F.; Kumar, L.; Al-Ghamdi, K.A. Spatiotemporal modeling of urban growth predictions based on driving force factors in five Saudi Arabian cities. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 139. [[CrossRef](#)]
33. Lin, Y.-P.; Chu, H.-J.; Wu, C.-F.; Verburg, P.H. Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling—A case study. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 65–87. [[CrossRef](#)]
34. Li, X.; Yeh, A.G.-O. Urban simulation using principal components analysis and cellular automata for land-use planning. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 341–352.

35. Dunn, W.; Scott, D.; Glen, W. Principal components analysis and partial least squares regression. *Tetrahedron Comput. Method* **1989**, *2*, 349–376. [[CrossRef](#)]
36. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [[CrossRef](#)]
37. Abdi, H. Partial least square regression (PLS regression). *Encycl. Res. Methods Soc. Sci.* **2003**, *2003*, 792–795.
38. Deng, X.; Huang, J.; Rozelle, S.; Uchida, E. Growth, population and industrialization, and urban land expansion of China. *J. Urban Econ.* **2008**, *63*, 96–115. [[CrossRef](#)]
39. Feng, Y.; Liu, Y. Fractal dimension as an indicator for quantifying the effects of changing spatial scales on landscape metrics. *Ecol. Indic.* **2015**, *53*, 18–27. [[CrossRef](#)]
40. Feng, Y.; Liu, Y.; Liu, Y. Spatially explicit assessment of land ecological security with spatial variables and logistic regression modeling in Shanghai, China. *Stoch. Environ. Res. Risk Assess.* **2016**. [[CrossRef](#)]
41. Wu, F.; Webster, C.J. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environ. Plan. B* **1998**, *25*, 103–126. [[CrossRef](#)]
42. White, R.; Engelen, G. Cellular automata as the basis of integrated dynamic regional modelling. *Environ. Plan. B* **1997**, *24*, 235–246. [[CrossRef](#)]
43. Feng, Y.; Liu, Y.; Liu, D. Shoreline mapping with cellular automata and the shoreline progradation analysis in Shanghai, China from 1979 to 2008. *Arab. J. Geosci.* **2015**, *8*, 4337–4351. [[CrossRef](#)]
44. He, C.; Okada, N.; Zhang, Q.; Shi, P.; Zhang, J. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Appl. Geogr.* **2006**, *26*, 323–345. [[CrossRef](#)]
45. Blečić, I.; Cecchini, A.; Trunfio, G.A. How much past to see the future: A computational study in calibrating urban cellular automata. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 349–374. [[CrossRef](#)]
46. Feng, Y.; Liu, Y. A cellular automata model based on nonlinear kernel principal component analysis for urban growth simulation. *Environ. Plan. B* **2013**, *40*, 117–134. [[CrossRef](#)]
47. García, A.M.; Santé, I.; Crecente, R.; Miranda, D. An analysis of the effect of the stochastic component of urban cellular automata models. *Comput. Environ. Urban* **2011**, *35*, 289–296. [[CrossRef](#)]
48. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
49. Clarke, K.C.; Hoppen, S.; Gaydos, L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environ. Plan. B* **1997**, *24*, 247–261. [[CrossRef](#)]
50. Almeida, C.; Gleriani, J.; Castejon, E.F.; Soares-Filho, B. Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 943–963. [[CrossRef](#)]
51. Liu, X.; Li, X.; Liu, L.; He, J.; Ai, B. A bottom-up approach to discover transition rules of cellular automata using ant intelligence. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 1247–1269. [[CrossRef](#)]
52. Campbell, J.B.; Wynne, R.H. *Introduction to Remote Sensing*; Guilford Press: New York, NY, USA, 2011.
53. Liu, Y. *Modelling Urban Development with Geographical Information Systems and Cellular Automata*; CRC Press: New York, NY, USA, 2008.
54. Yeh, A.G.-O.; Li, X. Errors and uncertainties in urban cellular automata. *Comput. Environ. Urban* **2006**, *30*, 10–28. [[CrossRef](#)]
55. Ménard, A.; Marceau, D.J. Exploration of spatial scale sensitivity in geographic cellular automata. *Environ. Plan. B* **2005**, *32*, 693–714. [[CrossRef](#)]
56. Wang, F.; Hasbani, J.-G.; Wang, X.; Marceau, D.J. Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set Theory. *Comput. Environ. Urban* **2011**, *35*, 116–125. [[CrossRef](#)]
57. Verburg, P.H.; de Nijs, T.C.; van Eck, J.R.; Visser, H.; de Jong, K. A method to analyse neighbourhood characteristics of land use patterns. *Comput. Environ. Urban* **2004**, *28*, 667–690. [[CrossRef](#)]
58. Pan, Y.; Roth, A.; Yu, Z.; Doluschitz, R. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Comput. Environ. Urban* **2010**, *34*, 400–408. [[CrossRef](#)]
59. Feng, Y.; Yang, Q.; Hong, Z.; Cui, L. Modelling coastal land use change by incorporating spatial autocorrelation into cellular automata models. *Geocarto. Int.* **2016**, 1–44. [[CrossRef](#)]

