

Article

Spatially-Explicit Simulation of Urban Growth through Self-Adaptive Genetic Algorithm and Cellular Automata Modelling

Yan Liu ^{1,*}, Yongjiu Feng ² and Robert Gilmore Pontius Jr. ³

¹ School of Geography, Planning and Environmental Management, The University of Queensland, St Lucia, QLS 4072, Australia

² College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China;
E-Mail: yjfeng@shou.edu.cn

³ Graduate School of Geography, Clark University, Worcester, MA 10610, USA;
E-Mail: rpontius@clarku.edu

* Author to whom correspondence should be addressed; E-Mail: yan.liu@uq.edu.au;
Tel.: +61-7-3365-6483; Fax: +61-7-3365-6899.

Received: 21 April 2014; in revised form: 3 July 2014 / Accepted: 8 July 2014 /

Published: 18 July 2014

Abstract: This paper presents a method to optimise the calibration of parameters and land use transition rules of a cellular automata (CA) urban growth model using a self-adaptive genetic algorithm (SAGA). Optimal calibration is achieved through an algorithm that minimises the difference between the simulated and observed urban growth. The model was applied to simulate land use change from non-urban to urban in South East Queensland's Logan City, Australia, from 1991 to 2001. The performance of the calibrated model was evaluated by comparing the empirical land use change maps from the Landsat imagery to the simulated land use change produced by the calibrated model. The simulation accuracies of the model show that the calibrated model generated 86.3% correctness, mostly due to observed persistence being simulated as persistence and some due to observed change being simulated as change. The 13.7% simulation error was due to nearly equal amounts of observed persistence being simulated as change (7.5%) and observed change being simulated as persistence (6.2%). Both the SAGA-CA model and a logistic-based CA model without SAGA optimisation have simulated more change than the amount of observed change over the simulation period; however, the overestimation is slightly more severe for the logistic-CA model. The SAGA-CA model also outperforms the logistic-CA model with fewer quantity and allocation errors and slightly more hits. For Logan City, the most

important factors driving urban growth are the spatial proximity to existing urban centres, roads and railway stations. However, the probability of a place being urbanised is lower when people are attracted to work in other regions.

Keywords: self-adaptive genetic algorithm (SAGA); cellular automata (CA); urban land transition; simulation accuracies; Logan City

1. Introduction

Cellular automata (CA) models have been increasingly applied to simulate the systematic spatio-temporal processes and the stochastic behaviour of land use change and urban growth [1–18]. Central to a CA-based urban model is the definition of the model's transition rules, which determine how the state of a cell changes over time [9–11,15]. CA models are intuitive and capable of incorporating both the spatial and temporal dimensions of land evolution into the simulation process. According to [19], the transition rules of urban CA models can be classified into five categories, including (1) strictly orthodox transition rules [9,20]; (2) rules based on key drivers [5,21]; (3) rules based on artificial intelligence [22]; (4) fuzzy logic transition rules [6,11,12,16,23]; and (5) other types of transition rules [24–26]. Of these five categories, rules based on key drivers have been the most widely applied, which require strict definitions of a number of variables and parameters representing various spatial and non-spatial factors driving urban growth.

A diverse range of statistical methods have been developed to select such variables and parameters; these include logistic regression [5], spatial logistic regression [27], multi-criteria evaluation (MCE) [28] and hybrid models [29]. While the integration of spatial statistics and cellular automata has laid the foundation of modelling land use dynamics, it can become a challenge to define suitable transition rules or the relevant parameter values and to construct the architecture of the models [11]. For instance, the MCE and logistic regression methods do not remove the effects caused by the multi-collinearity amongst the spatial variables [11,15]. Principal components analysis (PCA) can extract independent components of information for the set of independent variables, but PCA does not guarantee that the principal components are relevant for the dependent variables [30]. Consequently, results generated by CA models incorporating spatial statistics may not mimic the actual land use patterns, making them ineffective in capturing the spatial dynamics of urban land use change [6,15].

A CA model can be developed using a hybrid method due to the advantages of its spatial explicitness implied by the key drivers and the powerful computation capacity inherited from artificial intelligence [31]. The development of genetic algorithms (GA) has provided researchers with new ways to identify and to search for suitable transition rules and their defining parameters in urban modelling [31–36]. A GA can be used to search for an optimal solution to a problem based on the mechanics of natural genetics and natural selection [37]. Compared with other evolutionary methods, such as particle swarm optimisation [15], the unique feature of the GA exists in its operators, including selection, crossover and mutation. Substantively, a GA is a randomised method rather than a simple random operation, because historical information is used to speculate on new candidate solutions [37].

In practice, the GA technique has been applied to address various optimisation problems of geographical systems [38–43]. In CA-based urban research, the general method of integrating cellular automata with GA was initially proposed in [44] and subsequently adopted to parameterise the Markov cellular automata model for land use change simulation [45]. The physical meanings of the “genes” in the CA model were defined in [33], which were then applied to retrieve the geographical CA transition rules. A modified GA was developed to search for optimal parameters and neighbourhood rules for a Markov chain model to simulate the spatio-temporal urban landscape change processes in China’s Daqing City [45]. Their results show that the combination of the GA and Markov model is capable of capturing the spatio-temporal trend in the landscape pattern associated with urbanization for their region. More recently, GA was applied to calibrate the SLEUTH model of urban growth, which not only reduced the computation time in model calibration, but also improved the simulation accuracy of the model [34]. In another application, a pattern-calibrated and GA-optimized CA model was developed, which incorporates the percentage of landscape, patch metric and landscape division into the fitness function of the GA model [46]. The GA optimisation technique has also been combined with statistical techniques to calibrate an urban CA for a small urban settlement of northwest Spain; their work shows that the model can be adapted to urban areas with various characteristics and dynamics [47].

This paper presents a method for optimising the land use transition rules and parameters of a cellular automata urban growth model using a self-adaptive genetic algorithm (SAGA). The model was applied to simulate the spatio-temporal pattern of non-urban to urban land conversion in Southeast Queensland’s Logan City, Australia. Section 2 presents the study area and the data collected and processed for the modelling practice. Section 3 introduces the modelling approach, which includes a CA urban model, followed by the SAGA used to optimise the spatial parameters of the CA transaction rules. The optimisation process is guided through a fitness function to minimize the differences between the simulated and the observed urban growth. Section 4 presents model outcomes followed by discussions and conclusion in the last two sections.

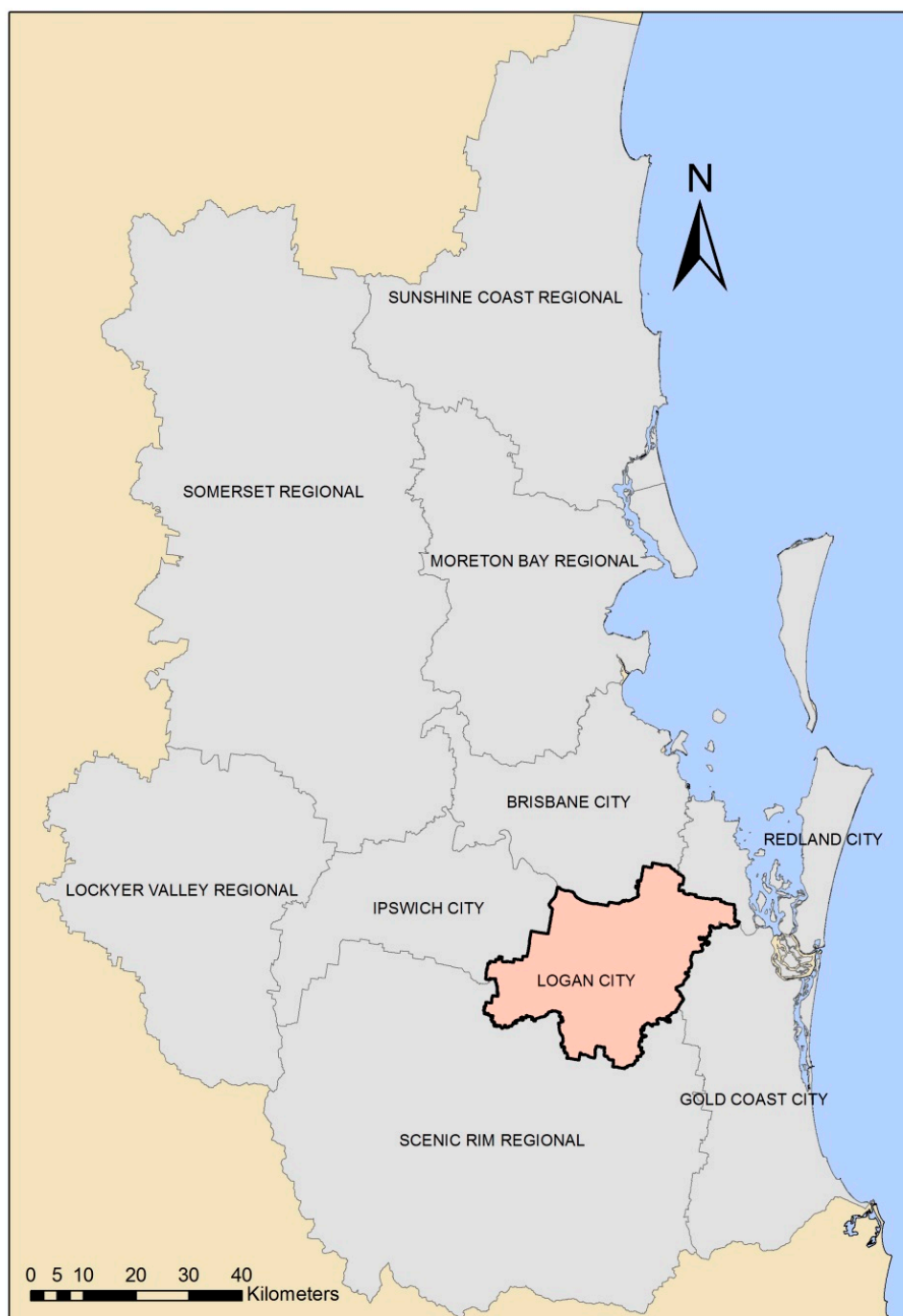
2. Study Area and Data

Logan City in South East Queensland, Australia, was selected as the case study site to apply the SAGA-CA model to simulate its land use evolution from non-urban to urban during the 1991–2001 period. Logan City is situated between the state capital city of Brisbane to the north and Gold Coast City to the south. It also borders Redland City to the northeast, the City of Ipswich to the northwest and the Scenic Rim to the southwest (Figure 1). Originally established in 1978 as a local government area, Logan City’s land size tripled in 2008, due to major changes to the local government in Queensland [48]. The current land size is 913 km² with a total population of 278,000 in 2011 [49].

Two land cover maps were collected from the Department of Natural Resources and Mines of the State of Queensland and used in this study, one representing the land cover in 1991 and one in 2001. These land cover maps were the product of the Statewide Landcover and Trees Study (SLATS) program funded by Queensland Government [50]. Landsat TM imageries acquired between June and September 1991 were used for the baseline land cover data, with a partial update of the baseline land cover dataset for 2001. The raw imagery has a nominal spatial resolution of 30 m. These raw imageries

were pro-processed through several stages of automated and semi-automated image, classification together with visual interpretation, field calibration and validation [51,52]. The final product was resampled to a spatial resolution of 25 m for release to the public [50,51].

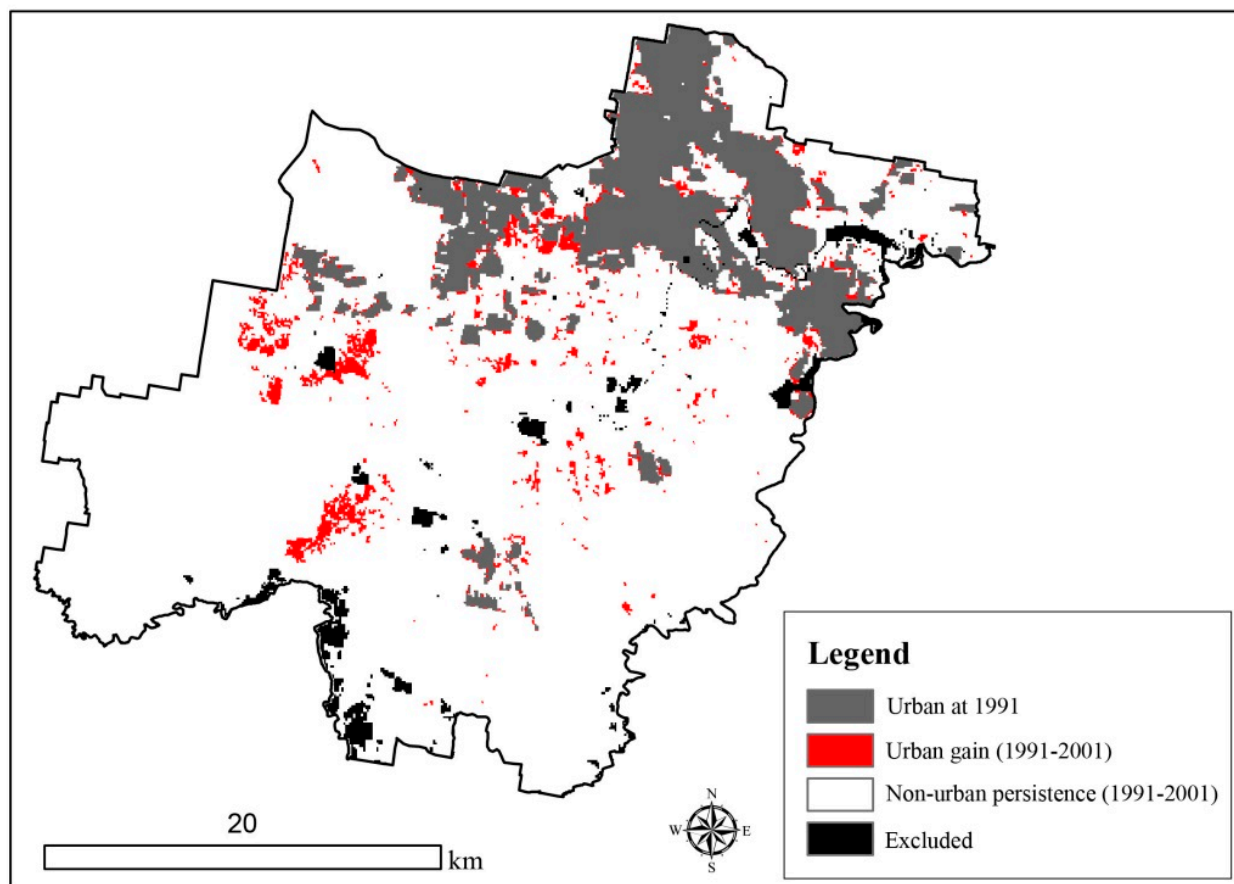
Figure 1. Logan City in South East Queensland, Australia.



Three land categories were identified from the SLATS land cover data in 1991 and 2001: urban (consisting of settlement areas with more than a population of 200, delineated using an analysis of digital cadastral boundaries based on lot size and population statistics), non-urban (consisting mostly of pasture land, woody vegetation and bare lands), and excluded (consisting of crop land and water) (Figure 2). As the focus of this research is on land use change from non-urban to urban states, the

presence of water bodies and crop land were considered as a hard constraint to urban growth; these areas were excluded from being urbanised. However, spatial proximities to water and crop land were taken as factors impacting urban growth. Table 1 lists the topographical and spatial factors, as well as the neighbourhood and stochastic factors used in the CA urban model.

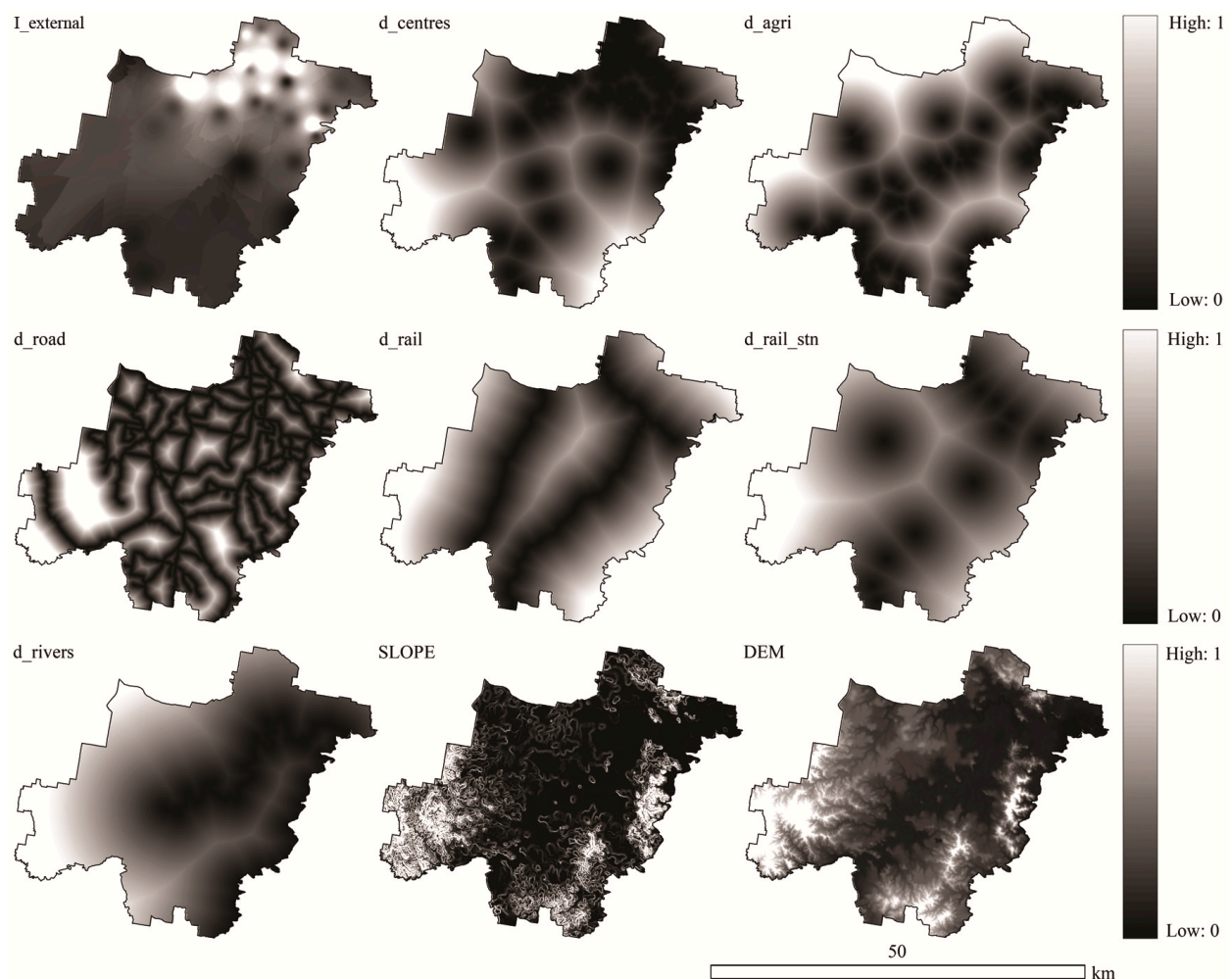
Figure 2. Logan City land use change from 1991 to 2001.



Other data collected include a 1-second DEM from the national Shuttle Radar Topographic Mission (SRTM) data and data on urban centres, main roads, railway lines and stations collected from the relevant local government authorities. A land slope data layer was generated from the DEM. The spatial proximity factors to urban centres ($d_{centres}$), to main roads (d_{road}), to railway lines (d_{rail}), to railway stations ($d_{railstn}$), to agricultural land (d_{agri}) and to water bodies (d_{rivers}) were generated using the Spatial Analyst Tools in ArcGIS from the relevant data sources. In addition, an external impact factor ($I_{external}$) measuring the impact of other urban centres in the neighbouring regions (including Brisbane, Ipswich, Gold Coast and Redland cities) on Logan City's urban growth was also considered in the model. This factor was quantified using the Journey to Work data from Logan to the surrounding regions; the earliest data available is in the 2006 census [53]; hence, this data was used as a proxy to quantify the external impact to urban growth of the city. A large external impact factor value indicates that more people have travelled to work outside the region rather than within the region. All data were processed to raster grids at 25-m spatial resolution to match with the spatial resolution of the SLATS land cover data; the values of the spatial variables were normalised to a range between 0 and 1 inclusive (Figure 3).

Table 1. Variables used to compute land conversion probabilities.

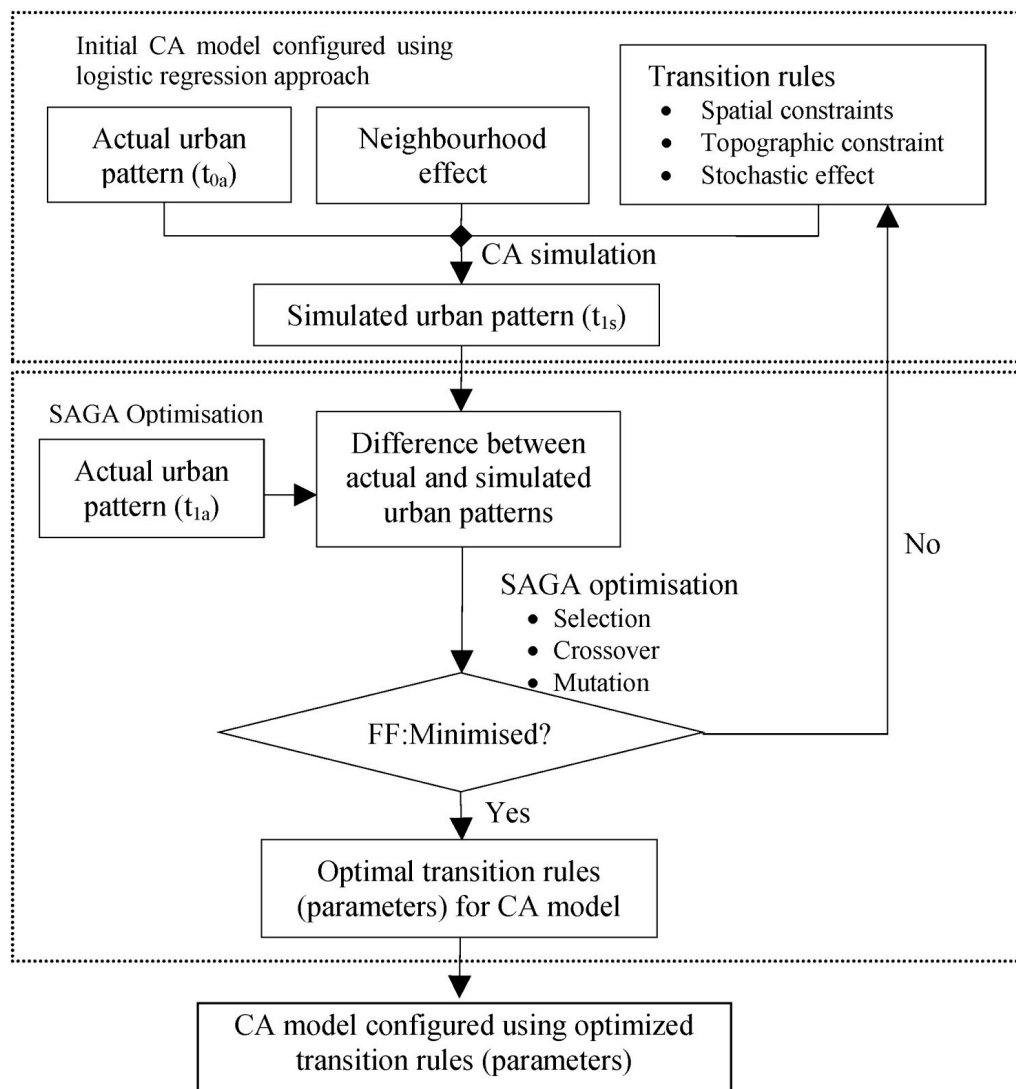
Variable	Meaning	Data Extraction
I_{external}	Impact of urban centres surrounding Logan City (including Brisbane, Ipswich, Gold Coast and Redland cities) on the urban growth within the region	Calculated using <i>Journey to Work</i> data from Logan to the neighbouring regions from the 2006 Census
d_{centres}	Distance to urban centres within the region	Measured in GIS from the urban centres data layer
d_{agri}	Distance to agricultural land	Measured in GIS from the land use data layer
d_{road}	Distance to main roads	Measured in GIS from the transportation network data layer
d_{rail}	Distance to railway lines	
$d_{\text{rail_stn}}$	Distance to railway stations	
d_{rivers}	Distance to rivers	Measured in GIS from the land use data layer
SLOPE	Land slope	Extracted from the DEM
DEM	Land elevation	
N_i^t	Probability of a cell changing its state from time t to the next time within a square 5×5 -cell neighbourhood	Calculated using the focal function in GIS according to Equation (3) presented in Section 3.1
R	Stochastic disturbance of unknown errors	Generated randomly using Equation (4) presented in Section 3.1

Figure 3. Spatial variables used in the SAGA-CA model in Logan City, Australia.

3. Method

The modelling framework consists of a generic CA urban growth model, which is linked to an optimisation module using SAGA to search for an optimal set of transition rules and parameters. The optimal set of rules and parameters will then be used by the CA model to simulate the dynamic process of urban growth (Figure 4). Each iteration of the model represents one year.

Figure 4. The SAGA process searching for the optimal CA transition rules.



Note: FF = Fitness Function

3.1. Cellular Automata Model

The generic CA urban growth model was initially configured using the logistic regression approach [54], where the land use conversion probability of a cell at location i from time t to the next time $(t + 1)$ is represented as:

$$P_i^t = F_i^t \times N_i^t \times Con \times R \quad (1)$$

where:

P_i^t represents the overall land use conversion probability of cell i from time t to the next time;

F_i^t is a land use conversion probability at location i from time t to the next time that is determined by the topographical features of the area and its spatial proximity to facilities and services. The topographical features impacting urban land conversion include slope and elevation, while the spatial proximity factors include distances to town centres, roads, railways, rivers and prime agricultural land. Therefore, F_i^t can be written as:

$$F_i^t(\mathbf{a}) = \frac{1}{1 + e^{-(a_0 + \sum_{j=1}^k a_j x_{ij}^t)}} \quad (2)$$

where: $\mathbf{a} = a_j$ ($j = 0, 1, \dots, k$), a_0 is a constant; a_j ($j = 0, 1, \dots, k$) are parameters representing the impact of the j -th factor x_{ij}^t (including both topographical and spatial factors) on land use conversion probability at location i at time t . The values for a_j were initially generated using the logistic regression method [54]; they were subsequently optimised using the SAGA optimisation approach.

N_i^t is a land use conversion probability at location i from time t to the next time due to neighbourhood support. In this research, a square neighbourhood with $m \times m$ cells was adopted; the probability that a cell develops from one state to another was defined as a proportion of the accumulative state of urban cells within the $m \times m$ neighbourhood. For the case study applied in this research, the neighbourhood size is 5×5 cells.

$$N_i^t = \frac{\sum_{j=1}^{m \times m} (S_j^t = \text{Urban}) (j \neq i)}{m \times m - 1} \quad (3)$$

where the processing cell i is not considered as a neighbouring cell, i.e., $j \neq i$.

Con represents a condition where simulated urban growth cannot occur at certain locations, such as large water bodies and land used primarily for farmland, which is prevented from undergoing urban development by land use planning regulations. Con takes on Boolean values of either 0 or 1, with 0 indicating that the cell is excluded from being urbanised and 1 indicating that the cell can change from non-urban to urban during a time step.

R is a stochastic disturbance factor on urban development, which is defined as [1,25,26]:

$$R = 1 + (-\ln r)^a \quad (4)$$

where r is a random real number that ranges from 0 to 1, while a is a real number that ranges from 0 to 10 and controls the effect of the stochastic factor on urban development.

3.2. Optimisation through Self-Adaptive Genetic Algorithm

3.2.1. Encoding the Chromosomes

Chromosomes are the abstract representations of candidate solutions. A chromosome is a set of parameters that defines a proposed solution to the problem the GA is trying to solve. In the CA urban modelling practice, all possible CA transition rules impacting on urban growth are considered as chromosomes. Each chromosome is coded by a simple string as:

$$\mathbf{C} = [\beta_0, \beta_1, \dots, \beta_k] \quad (5)$$

where \mathbf{C} represents a candidate solution; k is the total number of topographic and spatial variables (as in Equation (2)); β_0 is a constant, that is, a “gene” in a candidate solution; β_1 to β_k represent the evaluation score of each variable in a candidate solution. The optimal candidate solution values of β_0 to β_k are the parameter values of a_j ($j = 0, 1, \dots, k$) used in Equation (2).

Initially, a number of chromosomes are randomly generated to form the possible solutions for the SAGA to begin its searching process. After many generations of selection, crossover and mutation operations, only those chromosomes that acquire better fitness values remain, resulting in the emergence of an optimal chromosome structure.

3.2.2. The Fitness Function

A fitness function is used to evaluate and quantify the optimality of a solution during the optimisation process [55]. The fitness function is an objective function to quantify the difference between simulated and observed urban growth patterns. An optimal set of solutions is achieved when the fitness function value, that is, the difference between the simulated and observed urban growth pattern, is minimised. It is created by selecting sample cells, within the cellular urban space, to minimize the differences between the simulation results produced by the model and the observed urban growth patterns identified from remotely-sensed imagery. The optimisation objective is written as:

$$\text{Minimise } D(\mathbf{C}) = \sqrt{\sum_{i=1}^M (P_i^t - U_i)^2} \quad (6)$$

Where: $D(\mathbf{C})$ is the fitness function for the candidate solution \mathbf{C} ; M is the number of samples that were selected from the cellular urban space and used to retrieve the CA transition rules; P_i^t is the global conversion probability of the state of cell i during time t to $t + 1$, as defined in Equation (1) for candidate solution \mathbf{C} ; and U_i is the reference conversion of cell I , which can only take one of the two values, 0 or 1, with $U_i = 0$ meaning that the state of the cell i persists as non-urban and $U_i = 1$ meaning the state of the cell changes from non-urban to urban for candidate solution \mathbf{C} .

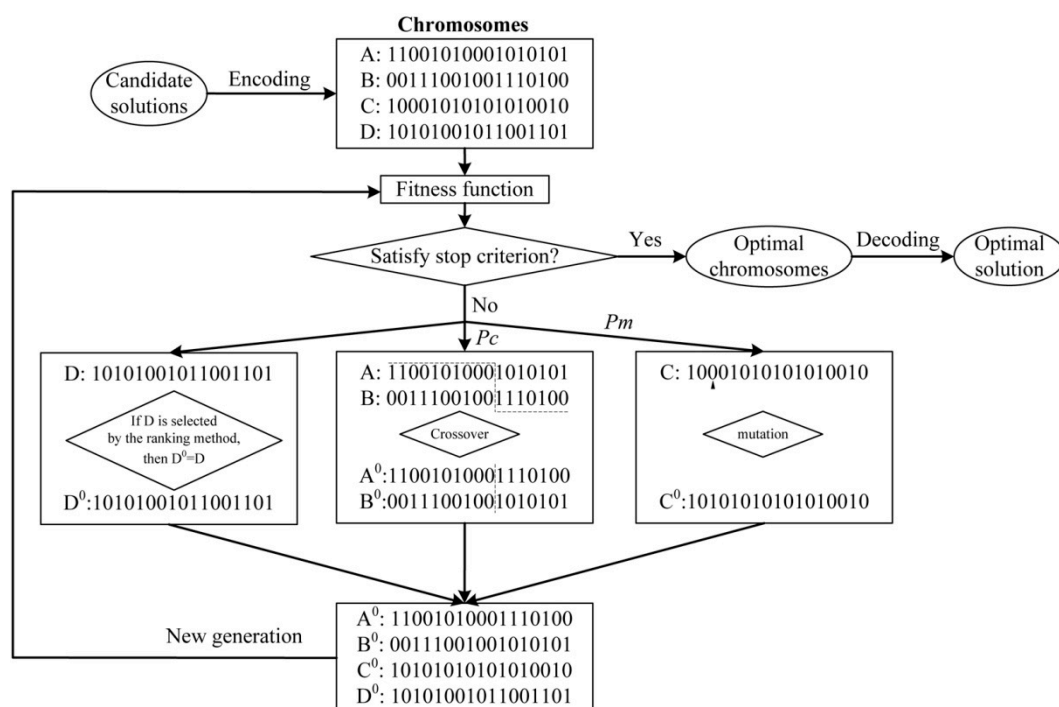
The simulation process of urban growth can be calibrated by dynamically updating the various parameters of the transition rules to minimize the value of the fitness function, so that the simulated urban patterns match the observed patterns of urban growth. The calibration process of the model is completed once the fitness function reaches a stable value over generations. At that point, the model's transition rules and their defining parameters can be considered optimised for operation in the CA model.

3.2.3. Selection, Crossover and Mutation

GA uses natural selection rules, including selection, crossover and mutation, to search for and optimise a solution from a set of chromosomes. Selection is the key operation in which individual genomes are chosen from a population of candidate solutions for later breeding, including recombination and crossover. Individual solutions are selected through a fitness-based process during each successive generation, where solutions with better fitness values are more likely to be selected. The smaller the difference between the simulated and observed results, the better the fitness value. Crossover is an exchange of genetic material between homologous individuals for final genetic recombination. Mutation is a genetic operator used to maintain genetic diversity from one generation of a population of chromosomes to the next.

However, standard GA commonly uses fixed and non-interactive selection, crossover and mutation rates, which can be problematic, because such operators cannot be modified during the search and optimisation process [37,56,57]. A SAGA can overcome this problem, because SAGA keeps population diversity and ensures the existence of all possible solutions in the solution domain; SAGA also ensures the identification of an optimal solution and improves the performance of the local and premature convergences [58,59]. In addition, the SAGA improves the search speed and precision of the standard GA and, hence, accelerates the search and optimisation process for the problem solutions. The three operations used in the SAGA are illustrated as follows (Figure 5).

Figure 5. The selection, crossover and mutation operations. Here, A, B, C and D are candidate solutions; P_c and P_m are the crossover and mutation probability, respectively.



Specifically, the selection operator used in this research was a ranking method that retains the best set of individuals that remain unchanged in the next generation of the selection operation [60]. The crossover and mutation operators are defined through a probability measure, which changes in accordance with the fitness values [59]. The crossover probability P_c and mutation probability P_m are defined as:

$$P_c = \begin{cases} P_{c1} - \frac{(P_{c1} - P_{c2})(f_{min} - f_{avg})}{(f_{max} - f_{avg})} & \dots\dots\dots f_{min} \geq f_{avg} \\ P_{c1} & \dots\dots\dots f_{min} < f_{avg} \end{cases} \quad (7)$$

$$P_m = \begin{cases} P_{m1} - \frac{(P_{m1} - P_{m2})(f_{min} - f_{avg})}{(f_{max} - f_{avg})} & \dots\dots\dots f_{min} \geq f_{avg} \\ P_{m1} & \dots\dots\dots f_{min} < f_{avg} \end{cases}$$

where: P_{c1} and P_{c2} are the maximum and minimum crossover probabilities, the values of which were set to 0.95 and 0.45, respectively; P_{m1} and P_{m2} are the maximum and minimum mutation

probabilities, the value of which were set to 0.1 and 0.001, respectively; f_{min} and f_{max} are the minimum and maximum fitness values, respectively; and f_{avg} is the average fitness value.

3.3. Model Implementation

A total of 25,000 sample pixels were selected randomly from the candidate region, meaning the entire study area minus the existing urban areas, agricultural land and water bodies in 1991. Therefore, the candidate region is areas that potentially can be developed into urban land use from 1991 onwards. Each sample includes a land use change variable from 1991 to 2001, as well as the external impact factor, distance factors, slope and DEM, as listed in Table 1 and shown in Figure 4. The land use change variable is a bi-value (with 0 or 1 value) indicating whether the state of a land cell has changed from non-urban to urban during 1991–2001 (in which case, the value is 1) or not (in which case, the value is 0). These sample data were used to build the logistic regression model (termed logistic-CA hereafter), which generates the initial values of a_j ($j = 0, 1, \dots, k$). This initial set of parameters was then used to form the possible solutions for the SAGA to construct the fitness function and begin its searching and optimisation process.

4. Results

4.1. The Optimal Chromosome/CA Transition Rules

Figure 6 shows the fitness track in the evolutionary computation of the SAGA module, which initially converges rapidly to the best fitness line. After 200 generations, the fitness value stabilized with the best fitness value of 47.1 and a mean fitness value of 47.2 (Figure 6).

Figure 6. Fitness track of the SAGA model.

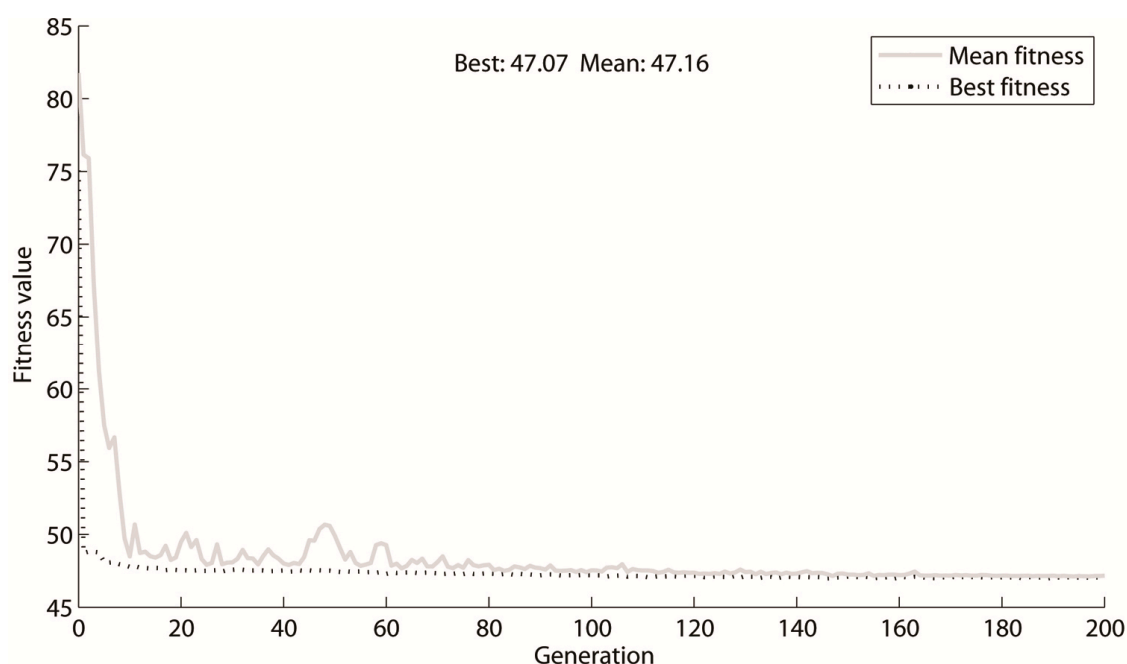


Table 2 lists the optimal solution by the SAGA in comparison with the initial setting of parameters generating the logistic-CA model. This optimal solution determines the initial input data for

the SAGA-CA model to compute the conversion probability of cells from non-urban to urban, as designated in Equation (2).

Table 2. Optimized CA parameters by the SAGA in comparison with parameters generated by the logistic regression model.

Variables	Parameters	
	Logistic	SAGA
a_0 (Constant)	−0.77	−0.82
a_1 (I_external)	1.02	1.19
a_2 (d_centres)	−0.98	−1.13
a_3 (d_agri)	0.93	0.98
a_4 (d_road)	−1.09	−1.05
a_5 (d_rail)	0.41	0.58
a_6 (d_railstn)	−1.15	−1.11
a_7 (d_rivers)	0.75	0.68
a_8 (SLOPE)	−0.34	−0.34
a_9 (DEM)	−0.91	−1.12

The optimal chromosome (*i.e.*, the optimal set of CA parameters) demonstrated the different impacts of the various factors on urban land use conversion in Logan City. According to Equation (2), a_j ($j = 0, 1, \dots, 9$) is a string of parameters representing the constant, I_external, d_centres, d_agri, d_road, d_rail, d_railstn, d_rivers, slope and DEM, respectively. A negative parameter of a_j leads to a larger F_i value, that is, a higher probability for a cell to convert from a non-urban to an urban state. Likewise, a positive a_j value results in a smaller F_i ; hence, a lower probability for the cell to convert into an urban state the next time. The parameter values optimised by the SAGA show that the external impact factor has the largest positive value, with a_1 (I_external) = 1.19, which is larger than the 1.02 generated by the logistic regression model. This positive value indicates that a place with a large number of people commuting to work out of the place is associated with a smaller probability of urban growth at that place. This is an opportunity cost for a place in Logan to be urbanised, as more people commute to work outside the region. On the other hand, the distance from existing urban centres correlates negatively with the probability of urban growth. This is reflected by the lowest negative value of a_2 (d_centres) = −1.13; such an effect is stronger in the SAGA-CA model than the logistic regression-based CA model, given the value of −0.98 generated by the logistic regression approach. Likewise, elevation, land slope and distances to roads and railway stations are also associated negatively with the probabilities of simulated urban growth, with a_9 (DEM) = −1.12, a_8 (SLOPE) = −0.34, a_4 (d_road) = −1.05 and a_6 (d_railstn) = −1.11, respectively. However, close proximity to agricultural land, rivers or railway lines tends to decrease the probability of simulated urban growth with a_3 (d_agri) = 0.98, a_7 (d_rivers) = 0.68 and a_5 (d_rail) = 0.58 after optimisation by the SAGA. Hence, the closer a cell is to rivers or railway lines, the less likely the cell is to be developed as an urban state.

4.2. Simulation Accuracies of the SAGA-CA Model

Using the 1991 land use map as the initial input data, the SAGA-CA model was operated with a set of land use transition rules, optimised by the SAGA, to generate a series of land use patterns. Each

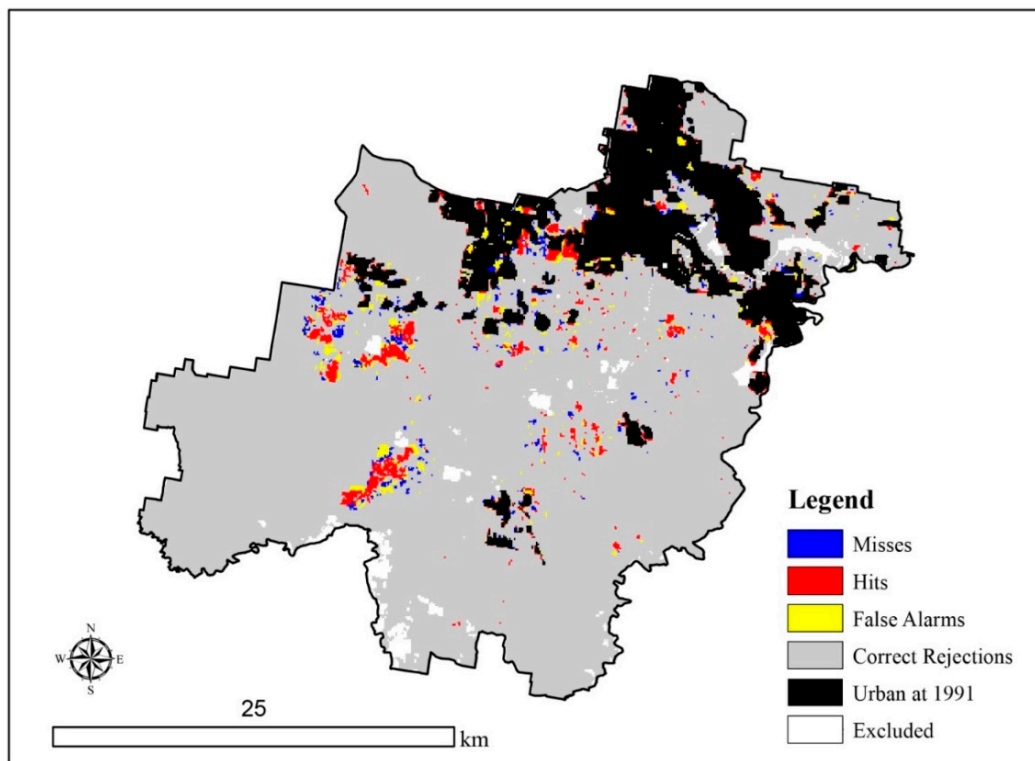
iteration of the model represented one year. Hence, the result representing urban land use pattern in Logan City in 2001 was generated after 10 iterations of the model.

The simulated urban growth map during 1991–2001 produced by the SAGA-CA model was compared with the reference map of urban growth during 1991–2001 generated from the SLATS land cover data [50]; this was realised through the comparison of three maps using the approach introduced in [61]. This produces a set of measures to evaluate the correctness and error of the simulation output (Figure 7a). Here, hits (H) indicate that growth areas from 1991 to 2001 shown on the reference maps were simulated as growth; misses (M) indicate that growth areas shown on the reference maps were simulated as persistence (*i.e.*, no change in simulated state over the two time points); false alarms (FA) indicate that the persistence shown on the reference maps were simulated as growth; correct rejections (CR) indicate that persistence shown on the reference maps was simulated as persistence; and excluded indicates the union of water and agriculture at the base year in 1991. Figure 7b illustrates the simulation correctness and error by the logistic-CA model without SAGA optimisation. To assess the simulation accuracies of the model, existing urban areas, agricultural land and water bodies in 1991 were excluded from the calculations. According to the SAGA-CA model, many of the false alarms are near the misses; this is different from the result generated by the Logistic-CA model, where many false alarms are rings around the excluded region. Figure 8 shows the size of the misses, hits, false alarms and correct rejections for the candidate region by the SAGA-CA model in comparison with the result by the logistic-CA without SAGA optimisation.

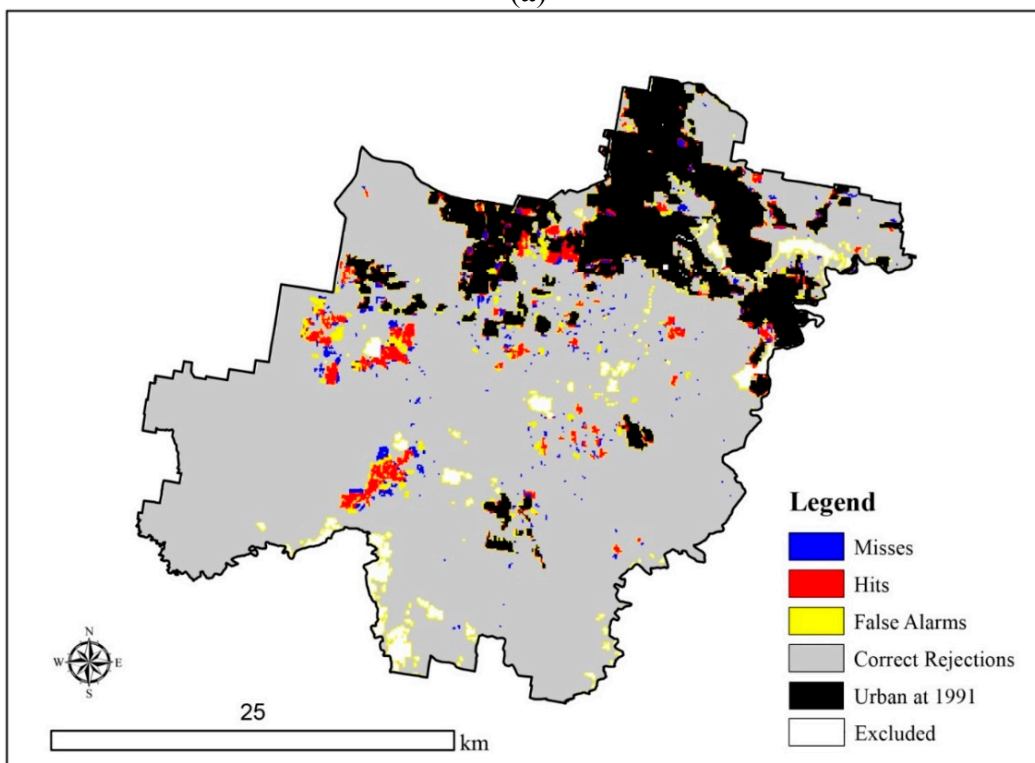
Overall, the SAGA-CA model generated 86.3% correctness, mostly due to observed persistence being simulated as persistence (79.2%) and some due to observed change being simulated as change (7.1%). The 13.7% simulation error was due to either observed persistence being simulated as change (7.5%) or observed change being simulated as persistence (6.2%). The observed change (OC) accounts for 13.3% of the candidate region, whereas the simulated change (SC) by the SAGA-CA model accounts for 14.6% of the region. Compared to the logistic-CA model, the SAGA-CA has resulted in slightly less overall error, 13.7% *versus* 15.1%. Figure 8 shows that both the logistic-CA and SAGA-CA models have simulated more change than the amount of observed change, while the overestimation is slightly more severe for the logistic-CA model.

The simulation errors generated by the model can also be presented through the measures of error due to allocation (A) and error due to quantity (Q) [61–65]. The error due to quantity measures how much less than perfect is the match between the observed and simulated quantity of change. The error due to allocation measures how much less than maximum is the match in the spatial allocation of the changes, given the specification of the quantities of the changes in the observed and simulated change maps [65]. Amongst 13.7% of the total errors between the SAGA-CA simulated and the actual land use change for the 1991–2001 interval, 1.3% was due to quantity disagreement; the other 12.4% were due to allocation disagreement. The quantity disagreement occurred due to the SAGA-CA model simulating slightly more growth than the reference growth. The results also show that the SAGA-CA model outperforms the logistic-CA model with fewer quantity and allocation errors and slightly more hits (Figure 9).

Figure 7. Simulation correctness and error based on the reference growth *versus* the simulated growth during 1991–2001. (a) The simulation correctness and error by the SAGA-CA model; (b) the simulation correctness and error by the logistic-CA model without SAGA optimisation.



(a)



(b)

Figure 8. Misses, hits and false alarms in the non-urban area of 1991, where SC denotes simulated change and OC denotes observed change from 1991 to 2001.

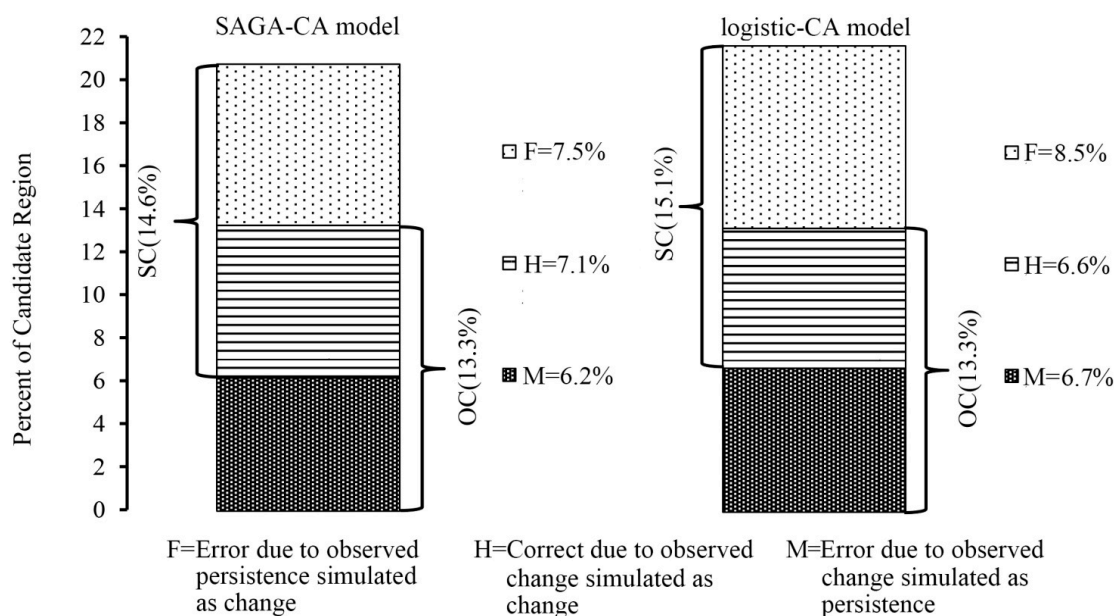
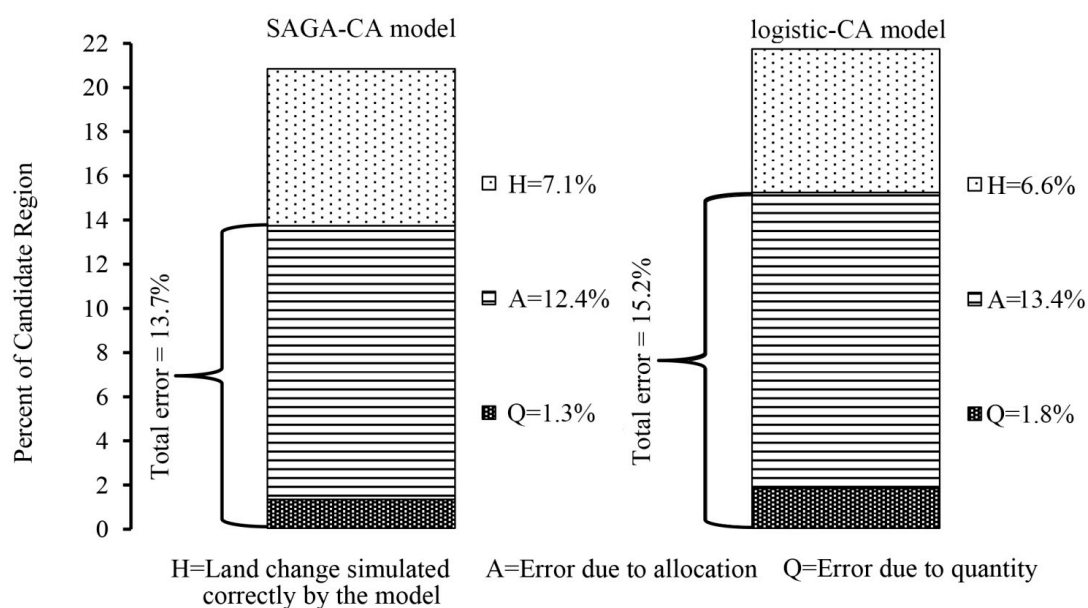


Figure 9. Quantity error, allocation error and hits in the non-urban area of 1991.



5. Discussion

Cellular automata techniques have been developed to explore urban land use evolution over the last two decades, and the reliability of the models has been studied [16,35,62]. The simulation accuracies of an urban CA model can be affected by the methodologies used in retrieving the transition rules, the spatial and thematic resolution of the model, as well as the physical, socio-environmental and institutional situations of the areas under study.

Firstly, the simulation accuracy of the model is subject to the characteristics of the region under study. In this research, three thematic categories (*i.e.*, urban, non-urban and excluded) were extracted

from a secondary data source (*i.e.*, the SLATS land cover data) to simulate the process of land use change from non-urban to urban. These three land use categories are highly generalised from the more complex land use types on the ground.

The simulation accuracy of the CA-based urban models is also sensitive to the resolution of cells [15,66]. Regardless of other factors impacting on the simulation accuracies of the CA model, a model configured at 250-m resolution can generate lower simulation accuracies compared to models configured at 30-m cell resolution [15]. The low simulation accuracy of the model at a coarse scale is usually due to the isolation of urban cells at such a scale, where only a small number of isolated urban cells can be identified in the initial input data [15].

In addition, the methodology adopted in retrieving the CA's transition rules is crucial for plausible simulation results. A number of new methods have been developed to capture land use dynamics and to improve simulation accuracy. By integrating the SAGA method into a typical logistic regression-based CA model, the integrated model is capable of taking into account feedback from individual "genes" during the modelling process, leading to the identification of a set of optimised transition rules.

6. Conclusions

Spatially-explicit simulation of urban land use change has attracted widespread interest in recent years with the focus on the spatio-temporal dynamics of the urban system and its land use evolution. Many CA-based urban models have been developed and applied in various situations to simulate the dynamic change of urban land use over time. However, it remains challenging for urban modellers to identify suitable transition rules reflecting the driving factors on land use change in the modelling practice.

This paper contributes to the field by developing an urban CA model with its transition rules optimised by a SAGA. It builds on the evolutionary computation technique [58,59] by implementing a set of essential driving forces to urban land use change and using SAGA to optimize the design configuration, similar to evolutionary processes. Consequently, a set of optimised transition rules and their defining parameters were identified and used to simulate the spatio-temporal process of land use evolutions. For the application of the model in Logan City in Queensland, Australia, most of the areas changed from non-urban to urban occurred around the existing urban centres, due to the effect of existing urban centres attracting more growth within their proximity. Likewise, the spatial proximity to roads and railway stations are also important factors attracting urban growth. On the other hand, the probability of a place being urbanised is lower when more people are attracted to work in other regions, mostly in Brisbane, Ipswich and the Gold Coast cities. In addition, spatial proximities to agricultural land, to rivers and railway lines also discourage an area from developing into urban land use. The application of the SAGA-CA model to Logan City demonstrates that the SAGA technique can increase simulation accuracy compared to a conventional logistic method, because the SAGA technique optimizes the transition rules of an urban CA model for simulating urban growth.

Author Contributions

All authors contributed to the writing of the paper. Data processing, analysis and modelling were performed by Yan Liu and Yongjiu Feng. Reviewers' comments were responded to by all authors.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. White, R.W.; Engelen, G. Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land use patterns. *Environ. Plan. A Environ. Plan.* **1993**, *25*, 1175–1193.
2. Clarke, K.C.; Gaydos, L. 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.
3. Wu, F. SimLand: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *Int. J. Geogr. Inf. Sci.* **1998**, *12*, 63–82.
4. Batty, M.; Xie, Y.; Sun, Z. Modelling urban dynamics through GIS-based cellular automata. *Comput. Environ. Urban Syst.* **1999**, *23*, 205–233.
5. Wu, F. Calibration of stochastic cellular automata: The application to rural-urban land conversions. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 795–818.
6. Liu, Y.; Phinn, S.R. Modelling urban development with cellular automata incorporating fuzzy-set approaches. *Comput. Environ. Urban Syst.* **2003**, *27*, 637–658.
7. He, C.Y.; Okada, N.; Zhang, Q.F.; Shi, P.J.; Zhang, J.S. Modelling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Appl. Geogr.* **2006**, *26*, 323–345.
8. Stevens, D.; Dragičević, S. A GIS-based irregular cellular automata model of land-use change. *Environ. Plan. B Plan. Des.* **2007**, *34*, 708–724.
9. Stevens, D.; Dragičević, S.; Rothley, K. iCity: A GIS-CA modelling tool for urban planning and decision making. *Environ. Model. Softw.* **2007**, *22*, 761–773.
10. Dietzel, C.; Clarke, K.C. Toward optimal calibration of the SLEUTH land use change model. *Trans. GIS* **2007**, *11*, 29–45.
11. Al-kheder, S.; Wang, J.; Shan, J. Fuzzy inference guided cellular automata urban-growth modelling using multi-temporal satellite images. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 1271–1293.
12. Al-Ahmadi, K.; See, L.; Heppenstall, A.; Hogg, J. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecol. Complex.* **2009**, *6*, 80–101.
13. Mondal, P.; Southworth, J. Evaluation of conservation interventions using a cellular automata-Markov model. *For. Ecol. Manag.* **2010**, *260*, 1716–1725.
14. Heppenstall, A.; See, L.; Al-Ahmadi, K.; Kim, B. CA City: Simulating urban growth through the application of Cellular Automata. In *Cellular Automata—Simplicity behind Complexity*; Salcido, A., Ed.; InTech: Winchester, UK, 2011; pp. 87–104.
15. 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.
16. Liu, Y. Modelling sustainable urban growth in a rapidly urbanising region using a fuzzy-constrained cellular automata approach. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 151–167.

17. Vaz, E.; Nijkamp, P.; Painho, M.; Caetano, M. A multi-scenario forecast of urban change: A study on urban growth in the Algarve. *Landsc. Urban Plan.* **2012**, *104*, 201–211.
18. Chaudhuri, G.; Clarke, K.C. The SLEUTH land use change model: A review. *Int. J. Environ. Resour. Res.* **2013**, *1*, 88–104.
19. Santé, I.; García, A.M.; Miranda, D.; Crecente, R. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landsc. Urban Plan.* **2010**, *96*, 108–122.
20. Ward, D.P.; Murray, A.T.; Phinn, S.R. A stochastically constrained cellular model of urban growth. *Comput. Environ. Urban Syst.* **2000**, *24*, 539–558.
21. White, R.; Engelen, G.; Uljee, I. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environ. Plan. B Plan. Des.* **1997**, *24*, 323–343.
22. Almeida, C.M.; Gleriani, J.M.; Castejon, E.F.; Soares-Filho, B.S. Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 943–963.
23. Wu, F. Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *J. Environ. Manag.* **1998**, *53*, 293–308.
24. Li, L.; Sato, Y.; Zhu, H. Simulating spatial urban expansion based on a physical process. *Landsc. Urban Plan.* **2003**, *64*, 67–76.
25. 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.
26. Feng, Y.; Liu, Y. A cellular automata model based on nonlinear kernel principal component analysis for urban growth simulation. *Environ. Plan. B* **2013**, *40*, 116–134.
27. Tayyebi, A.; Delavar, M.R.; Pijanowski, B.C.; Yazdanpanah, M.J.; Saeedi, S.; Tayyebi, A.H. A spatial logistic regression model for simulating land use patterns, a case study of the Shiraz metropolitan area of Iran. In *Advances in Earth Observation of Global Change*; Chuvieco, E., Li, J., Yang, X., Eds.; Springer: Berlin, Germany, 2010; pp. 27–42.
28. Wu, F.; Webster, C.J. Simulation of land development through the integration of cellular automata and multi-criteria evaluation. *Environ. Plan. B Plan. Des.* **1998**, *25*, 103–126.
29. Arsanjani, J.J.; Helbich, M.; Kainz, W.; Boloorani, A.D. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 265–275.
30. Glen, W.G.; Dunn, W.J.; Scott, D.R. Principal components analysis and partial least squares regression. *Tetrahedron Comput. Methodol.* **1989**, *2*, 349–376.
31. Bies, R.R.; Muldoon, M.F.; Pollock, B.G.; Manuck, S.; Smith, G.; Sale, M.E. A genetic algorithm-based, hybrid machine learning approach to model selection. *J. Pharmacokinet. Pharmacodyn.* **2006**, *33*, 196–221.
32. Xu, X.; Zhang, J.; Zhou, X. Integrating GIS, cellular automata, and genetic algorithm in urban spatial optimization: A case study of Lanzhou. *Proc. SPIE* **2006**, *6420*, doi:10.1117/12.712954.
33. Li, X.; Yang, Q.S.; Liu, X.P. Genetic algorithms for determining the parameters of cellular automata in urban simulation. *Sci. China Ser. D Earth Sci.* **2007**, *50*, 1857–1866.
34. Clarke-Lauer, M.D.; Clarke, K.C. Evolving simulation modeling: Calibrating SLEUTH using a genetic algorithm. In Proceedings of the 11th International Conference on GeoComputation, London, UK, 20–22 July 2011.

35. Feng, Y.; Liu, Y.; Han, Z. Land use simulation and landscape assessment using genetic algorithm based cellular automata under different sampling schemes. *Chin. J. Appl. Ecol.* **2011**, *22*, 957–963.
36. Feng, Y.; Liu, Y. An optimized cellular automata based on adaptive genetic algorithm for urban growth simulation. In *Advances in Spatial Data Handling and GIS: Lecture Notes in Geoinformatics and Cartography*; Yeh, A.G.O., Shi, W., Leung, Y., Zhou, C., Eds.; Springer-Verlag: Berlin/Heidelberg, Germany, 2012; Part 2, pp. 27–38.
37. Schmitt, L.M.; Nehaniv, C.L.; Fujii, R.H. Linear analysis of genetic algorithms. *Theor. Comput. Sci.* **1998**, *208*, 111–148.
38. Jenerette, G.D.; Wu, J. Analysis and simulation of land-use change in the central Arizona—Phoenix region, USA. *Landsc. Ecol.* **2001**, *16*, 611–626.
39. Stewart, T.J.; Janssen, R.; Herwijnen, M.V. A genetic algorithm approach to multi-objective land use planning. *Comput. Oper. Res.* **2004**, *31*, 2293–2313.
40. Matthews, K.B.; Buchan, K.; Sibbald, A.R.; Craw, S. Combining deliberative and computer-based methods for multi-objective land-use planning. *Agric. Syst.* **2006**, *87*, 18–37.
41. Cao, K.; Batty, M.; Huang, B.; Liu, Y.; Yu, L.; Chen, J.F. Spatial multi-objective land use optimization: Extensions to the non-dominated sorting genetic algorithm-II. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1949–1969.
42. Shan, J.; Alkheder, S.; Wang, J. Genetic algorithms for the calibration of cellular automata urban growth modeling. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 1267–1277.
43. Seppelt, R.; Voinov, A. Optimization methodology for land use patterns using spatially explicit landscape models. *Ecol. Model.* **2002**, *151*, 125–142.
44. Mitchell, M.; Crutchfield, J.P.; Das, R. Evolving cellular automata with genetic algorithms: A review of recent work. In *Proceedings of the First International Conference on Evolutionary Computation and Its Applications (EvCA'96)*, Moscow, Russia, 1–4 June 1996.
45. Tang, J.; Wang, L.; Yao, Z. Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm. *Int. J. Remote Sens.* **2007**, *28*, 3255–3271.
46. Li, X.; Lin, J.; Chen, Y.; Liu, X.; Ai, B. Calibrating cellular automata based on landscape metrics by using genetic algorithms. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 594–613.
47. García, A.M.; Santé, I.; Boullón, M.; Crecente, R. Calibration of an urban cellular automaton model by using statistical techniques and a genetic algorithm. Application to a small urban settlement of NW Spain. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1593–1611.
48. Queensland Local Government Reform Commission. *Report of the Local Government Reform Commission*; Queensland Government: Brisbane, QLD, Australia, 2007; Volume 2.
49. Australian Bureau of Statistics. Census Product 2011. Available online: <http://www.abs.gov.au/> (accessed on 1 October 2013).
50. Queensland Government. Statewide Landcover and Trees Study (SLATS). Available online: <http://www.qld.gov.au/environment/land/vegetation/mapping/slats/> (accessed on 30 June 2014).
51. Danaher, T.; Scarth, P.; Armston, J.; Collett, L.; Kitchen, J.; Gillingham, S. Remote sensing of tree-grass systems: The Eastern Australian woodlands. In *Ecosystem Function in Savannas: Measurement and Modeling at Landscape to Global Scales*; Hill, M.J., Hanan, N.P., Eds.; CRC Press: Boca Raton, FL, USA, 2010; pp. 175–194.

52. Collett, L.; Goulevitch, B.M.; Danaher, T.J. SLATS radiometric correction: A semi-automated multi stage process for the standardisation of temporal and spatial radiometric differences. In Proceedings of the 9th Australasian Remote Sensing and Photogrammetry Conference, Sydney, NSW, Australia, 24 July 1998.
53. Australian Bureau of Statistics. Census Product 2006. Available online: <http://www.abs.gov.au/> (accessed on 1 October 2013).
54. Liu, Y.; Feng, Y.J. 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*; Heppenstall, A.J., Crooks, A.Y., See, L.M., Batty, M., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 643–662.
55. Muzy, A.; Nutaro, J.J.; Zeigler, B.P.; Coquillard, P. Modelling landscape dynamics in an Atlantic rainforest region: Implications for conservation. *Ecol. Model.* **2008**, *219*, 212–225.
56. Goldberg, D.E.; Holland, J.H. Genetic algorithms and machine learning. *Mach. Learn.* **1988**, *3*, 95–99.
57. Mardle, S.; Pascoe, S. An overview of genetic algorithms for the solution of optimisation problems. *Comput. High. Educ. Econ. Rev.* **1999**, *13*, 16–20.
58. Angeline, P.J. Adaptive and self-adaptive evolutionary computation. In *Computational Intelligence: A Dynamic System Perspective*; IEEE Press: New York, NY, USA, 1995; pp. 152–161.
59. Srinivas, M.; Patnaik, L.M. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Trans. Syst. Man Cybern.* **1994**, *24*, 656–667.
60. Houck, C.R.; Joines, J.A.; Kay, M.G. *A Genetic Algorithm for Function Optimization: A Matlab Implementation*; Technical Report NCSU-IE-TR-95-09; North Carolina State University: Raleigh, NC, USA, 1995; pp. 1–14.
61. Pontius, R.G.; Peethambaram, S.; Castella, J. Comparison of three maps at multiple resolutions: A case study of land change simulation in Cho Don District, Vietnam. *Ann. Assoc. Am. Geogr.* **2011**, *101*, 45–62.
62. Pontius, R.G.; Shusas, E.; McEachern, M. Detecting important categorical land changes while accounting for persistence. *Agric. Ecosyst. Environ.* **2004**, *101*, 251–268.
63. Pontius, R.G.; Malanson, J. Comparison of the structure and accuracy of two land change models. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 243–265.
64. Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429.
65. Chen, H.; Pontius, R.G. Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. *Landsc. Ecol.* **2010**, *25*, 1319–1331.
66. Kocabas, V.; Dragičević, S. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Comput. Environ. Urban Syst.* **2006**, *30*, 921–953.