The integration of vector-based cellular automata and partitioned rules for simulation of land use change

Yi Lu*[1] and Shawn Laffan[2]

[1, 2]School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia
*Email: yi.lu@unsw.edu.au

Abstract

Cellular Automata (CA) have formed an important part of the geocomputational and spatial analysis toolbox for three decades. Some of the most important components of CA are the transition rules to determine the changes in cell states across iterations. These are applied at the micro-sale, but lead to emergent patterns at the macro-scale. An important consideration for transition rules is the spatial extent over which they will be valid. One does not expect the same rules to apply equally across an entire region, yet most CA implementations only support one set of transition rules that are applied everywhere. In this paper, a vector CA model with spatially partitioned transition rules is proposed to identify the expansion of urban residential areas across heterogeneous study area. Initial experiments using two sub-regions of Ipswich, Queensland, Australia, indicate that the spatially partitioned approach can improve the accuracy of vector CA.

Keywords: Vector-based CA, Partitioned rules, Ipswich City, Residential area, Misclassification frequency

1. Introduction

CA models have been employed in the exploration of a wide variety of urban phenomena, from traffic simulation and regional-scale urbanization to land-use dynamics, polycentricity, historical urbanization, and urban development (Torrens and O'Sullivan, 2001). For spatial scientists and urban planners, there is also an urgent need to predict future developments and land use change in an understandable way. Numerous computer-based models have been developed to address these issues, including CA (Cellular automata) (Wu and Webster, 1998; Li, Xia and Yeh, 2002; Liu et al., 2007), CLUE/CLUE-S (Conversion of Land use and its Effects at small region extent) (Verburg, P.H. et al., 1999; Verburg, P. and Overmars, 2007), MAS (Multi-agent system) (Heppenstall et al., 2011; O'Sullivan et al., 2016), SD (System dynamic) (He et al., 2006; Xu and Coors, 2012) and What If (Pettit, 2005; Pettit et al., 2015).

Among all land use change models, CA and its extension models have been widely applied due to their capability of modelling complex spatial dynamics on the basis of a set of 'simplified' transition rules (White et al., 1997). Abundant results have been achieved in the field of CA modelling, which can be classified into four groups: cell format (Flache and Hegselmann, 2001; Moreno et al., 2008), transition rules (Li, Xuecao et al., 2014; Almeida et al., 2008), neighbourhood configuration (Moreno et al., 2009), sensitivity and uncertainties (Kocabas and Dragicevic, 2006; Şalap-Ayça et al., 2018). While transition rules have attracted more attention than other parameters, there is still a key problem to be discussed and solved: Does a single set of transition rules contain enough information for all sub-regions of the study area?

Here we propose a spatially partitioned CA model to address the above question. Instead of using one set of general transition rules across the entire study region, the transformations of cells are determined by spatially local rules, partitioned by sub-regions. Using a PACA (partitioned and asynchronous cellular automata) model, Ke et al. (2016) simulated the process of urban growth during 2005 to 2013. Taking Yangtze River middle reaches megalopolis (YRMRM) as the study area, it is also indicated by Xia et al. (2019) that the development of partitioned transition rules for sub-regions can greatly improve both the overall and local accuracies of CA model. Concerning the afore-mentioned reports, it has been indicated that the spatial heterogeneity of urban growth can be better represented by using differential transition rules for partitioned zones. Additionally, these previous approaches used raster based CA models, while vector CA models have the potential to improve land use change modelling, particularly in urban environments. The aim of this research is therefore to assess the effectiveness of a spatially partitioned vector CA model for land use change modelling.

2. Methodology

2.1. Using PSO for the discovery of transition rules

In this research, particle swarm optimization (PSO) is utilized for the calibration of transition rules. PSO is a useful approach for the discovery of transition rules as it captures the complex non-linear processes of urban land use change (Feng et al., 2011) and deals well with the large number of calibration parameters (Pinto et al., 2017).

Similar to the relationship between cell and CA, particle is the smallest unit of PSO, it equals to one potential solution of the target problem, and is comprised of two parts: velocities and positions

$$Particle = (v_n, P_n)$$

Equation 1

where n is the dimension of target problem, v_n and P_n are the velocity and position of corresponding particle at a specific time point. They can be represented by n velocities and positions at time t:

$$\begin{cases} v_n = (v_1, v_2, \dots, v_n, t) \\ P_n = (P_1, P_2, \dots, P_n, t) \end{cases}$$

Equation 2

The combination of velocity and position in each particle are updated according to individual and global best positions:

$$\begin{cases} v(t+1) = w * v(t) + c1 * (P_{ib} - P(t)) + c2 * \left(P_{gb} - P(t)\right) \\ P(t+1) = P(t) + v(t+1) \end{cases}$$

Equation 3

where w is the weight of velocity, $c1$ and $c2$ are individual and global learning factors. P_{ib} is the best individual position of particle i, and P_{gb} is the best global position of all particles, namely the best one of all best individual positions. In addition, $v(t+1)$ is the velocity of a particle at time t+1, $\mathrm{P}(t)$ and $P(t+1)$ are the positions of particle at time t and t+1, accordingly.

The transfer probability $\mathrm{P}(Cell_i)$ of a single cell i can be calculated by Equation 4:

$$\mathrm{P}(Cell_i) = \frac{1}{1+\exp\left[-\left(a_0 + \sum_{i=1}^{i=n} f_i * w_i\right)\right]} * (b_0 + norNei)$$

Equation 4

Where $\mathrm{P}(Cell)$ is the probability of $Cell_i$, a_0 and b_0 are two constants, f_i and w_i are the normalized values of driving factors and corresponding weights, norNei is the normalized neighbourhood configuration. In this study, w_i are derived from PSO method, and n is the number of driving factors.

2.2. Validation method

Two indices are used here to validate the performances of PSO-CA models: Cumulative producer's spatial accuracy and misclassification frequency.

Cumulative producer's spatial accuracy (CPSA)

Producer's spatial accuracy (PSA) has been widely used for the assessment of precision in the research field of land use modelling, which can be described as:

$$\mathrm{PSA} = \frac{\mathrm{Area}_{cor}}{Area_{all}}$$

Equation 5

In Equation 5, the PSA of vector CA can be calculated by $\mathrm{Area_{cor}}$ and $\mathrm{Area_{all}}$, the correctly simulated and total area of cells.

Separate experiments are proposed in both the entire and sub-regions of the study area. Therefore, CPSA is defined as the mean value of PSA from all experiments:

$$\mathrm{CPSA} = \frac{\sum_{i=1}^{Num_{sim}} PSA}{Num_{sim}}$$

Equation 6

where Num_{sim} refers to the number of simulation experiments that conducted under each combination of CA parameters.

Misclassification frequency (MF)

Except for CPSA, the misclassification frequencies of all simulations, is used to assess the frequency a cell has been misclassified across the set of simulations:

$$MF = \frac{N_{incor}}{N_{all}}$$

Equation 7

where N_{incor} and N_{all} records the number of incorrectly classified and total simulation counts.

2.3. Model implementation

The implementation of both general and partitioned PSO-CAs can be summarized with three steps (Figure 1). At the beginning of simulation, the entire study area is divided into two sub-regions (administrative boundaries for this research). Sample data, which are required for PSO training, are then randomly selected from the spatial datasets. Afterwards, two separate PSOs, which represent the sets of general and partitioned transition rules, will be used for training, with the output of weights for different driving factors. After the completion of training, the simulated distribution of residential cells will be produced by the CA models with general and partitioned rules (general and partitioned PSO-CA). Finally, taking the land use map at the end of simulation period as reference, we evaluate the accuracy of both general and partitioned CAs.

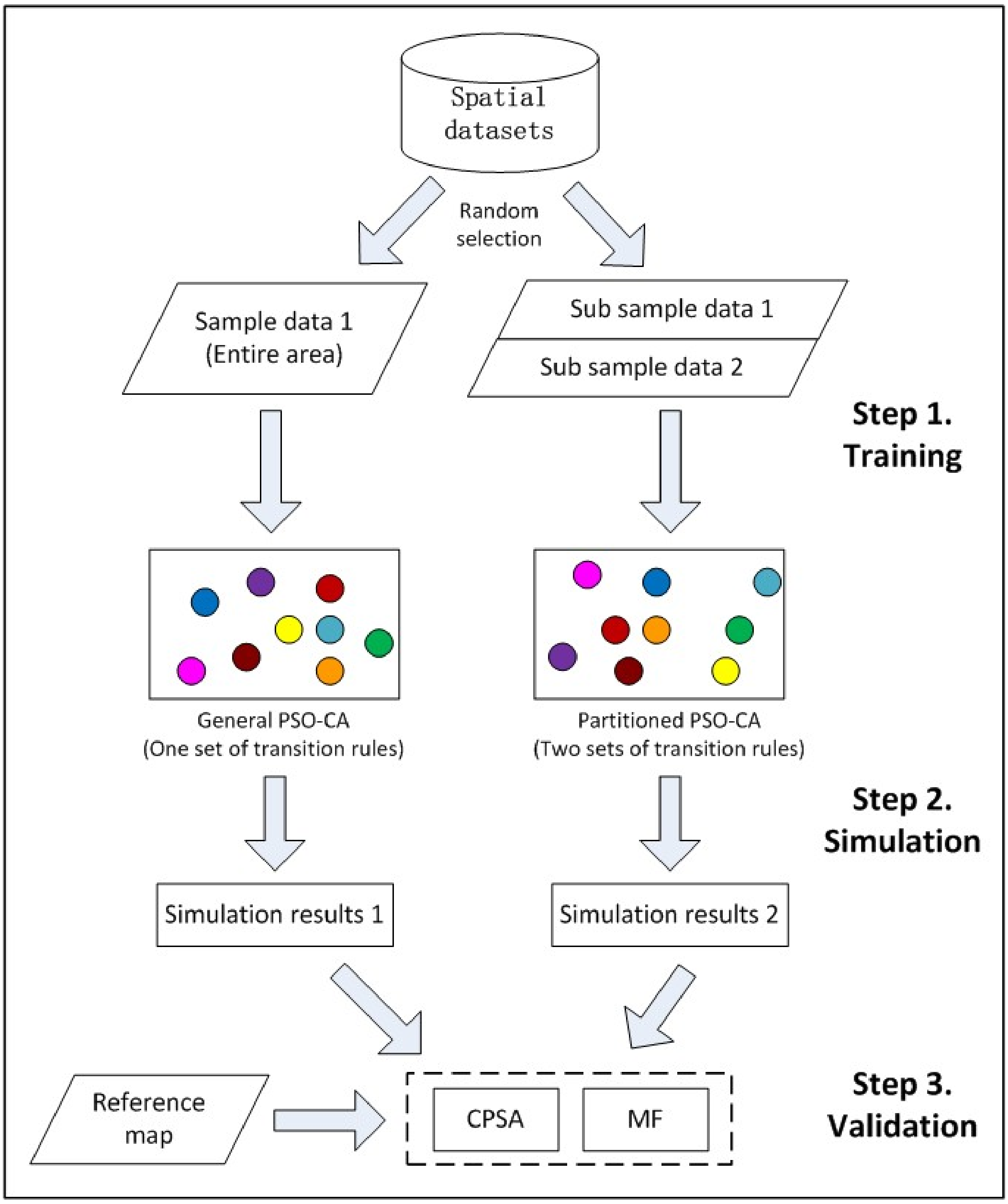

Figure 1: Model implementation procedures

3. Case study

3.1. Study area and data processing

Ipswich City is the second oldest local government area (LGA) of Brisbane-South East Queensland (SEQ) region. It is located approximately 35 km west of Brisbane, the capital city of Queensland, Australia. Two districts within Ipswich: Bellbird Park - Brookwater and Redbank Plains (Figure 2) are selected as the study area. These districts represent the typical trend of land use across Ipswich from 1999 to 2016. In 2016, the area of these two districts was 3,543.36 ha. The main land uses are "Intensive uses", "Conservation and natural environments" and "Production from natural environments", occupying 55.48%, 20.26% and 19.16% of the study area, respectively. The remainder

4.89% and 0.21% of the study area are classified as "Production from dryland agriculture and plantations" and "Water". Polygons of the study area were extracted from the LGA dataset of Queensland. This was obtained from QSpatial, a state-owned geospatial portal of Queensland (Queensland Government, 2016) which is also the source of the land use maps (1999 and 2016).

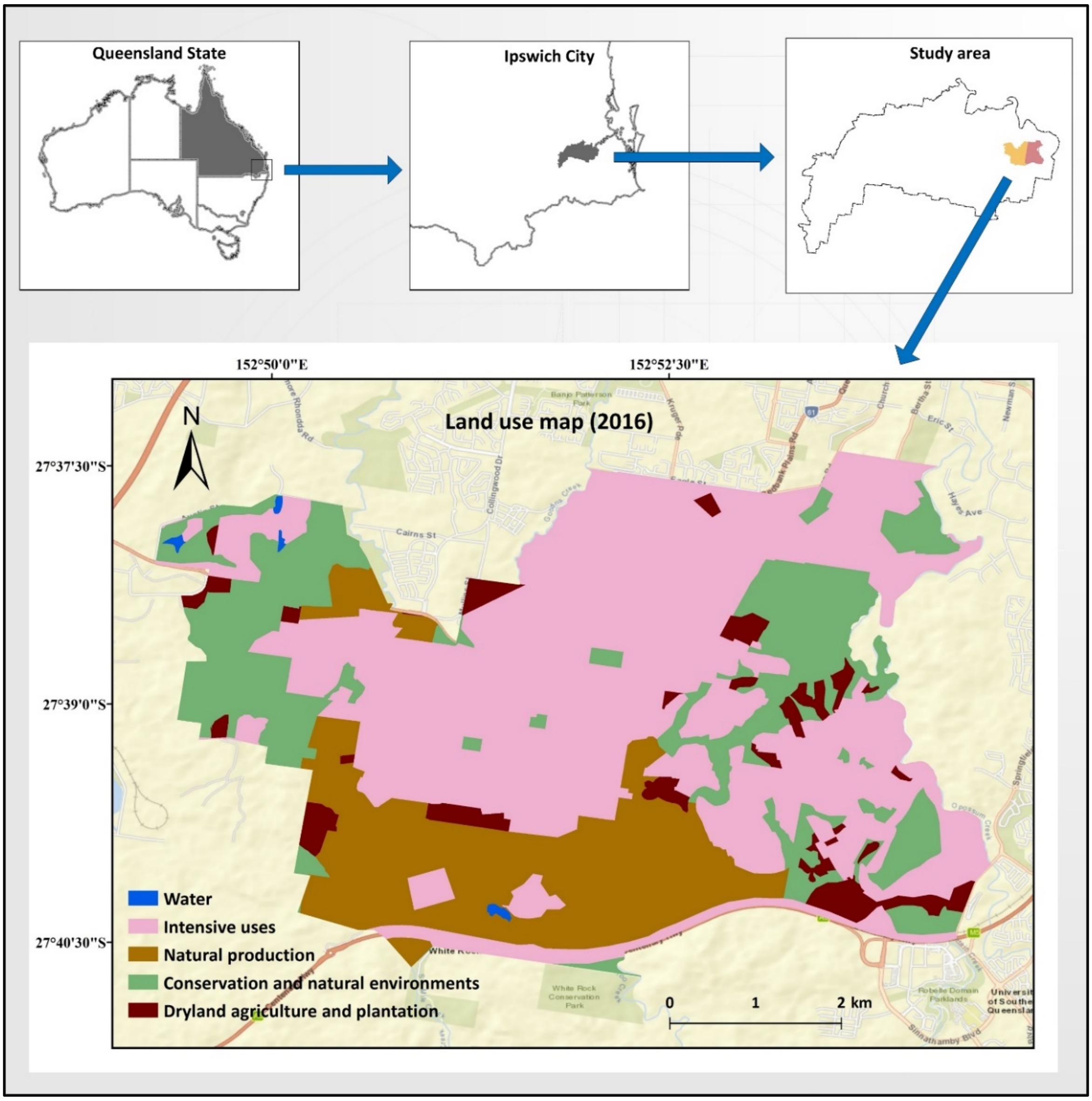

Figure 2: The study area is in Ipswich, Queensland, Australia

Of the land converted to residential by 2016, 92.57% were from secondary land use classes "Grazing native vegetation" and "Other minimal use", "Services (4.65%)" and "Land in transition (2.78%)" are the remaining sources. Specifically, the area of "Other minimal use" had been decreased by 450.73 ha between 1999 and 2016, representing 61.04% of all reduced land use. Besides, "Grazing native vegetation" suffers a 272.31 ha reduction, which is as much as 36.88% of the entire decreased category. Two land uses were excluded from the simulation: "Services" lands (mainly Schools and education institutions), are managed under local state regional planning (Queensland Government, 2017), while "Land in transition" refers to the unknown land use. For analysis, "Other minimal use" and "Grazing native vegetation" are classified as "non-residential", while "Intensive uses" is considered as "Residential". Here we focus on the transformation of parcels from non-residential to residential, as this is the most common land use change in such environments (Seto and Shepherd, 2009; Fragkias et al., 2013). The transformed (from non-residential to residential) and stable (remain non-residential during study period) cells are summarized in Table 1.

	Transformed		Stable	
	Count	**Area (ha)**	**Count**	**Area (ha)**
Bellbird Park - Brookwater	2,091	157.25	1,117	798.64
Redbank Plains	2,137	129.61	1,523	879.80
Entire study area	4,228	286.86	2,640	1,678.44

Table 1: The transformed and stable cells of study area during 1999 to 2016

3.2. Simulation process and result

Each simulation experiment was run for 100 iterations in order to reproduce the subtle patterns of land use change (Cao et al., 2015). In general, one iteration can be summarized as "calculate, select, and update". Calculate: at the beginning of an iteration, the transfer probabilities are calculated and assigned to non-residential layer. Select: newly transformed cells, where their attributes are converted to residential, are identified from the non-residential layer with relatively high transfer probabilities. Update: these selected cells are merged into the residential layer, and used to update the non-residential layer before next iteration. The main difference between general and partitioned CA is that the non-residential layer in partitioned CA will be updated by the residential layers in partitioned sub-regions in order to avoid edge effects on its simulation outputs.

Two model configurations were applied. In the first configuration, a single set of transition rules was calibrated and applied using the full study area. In the second, two sets of transition rules were calibrated, one for each sub-region (Bellbird Park - Brookwater, Redbank Plains). In every experiment, 20% of the sample cells (including both transformed and stable), which achieved a balance between information richness and over-training, were randomly selected to calibrate the transition rules. On the basis of initial experiments, along with existing studies on calibration of transition rule (Feng et al., 2011; Liao et al., 2014), the weight of velocity v, individual and global learning factors c1 and c2 (in Equation 3) were set as 1, 1.5 and 1.5, respectively. Similarly, two constants a_0 and b_0 were both set as 1. The initial position and velocity of each particle was generated with random values in the intervals [-5, 5] and [-2, 2], and the maximum particle velocity was restricted to 1 to reduce unrealistic results due to an extreme velocity. Following Harrison et al. (2019), 1000 iterations were used for PSO training to ensure all particles were well-trained.

Study area	**Dis_{com}**	**Dis_{cen}**	**Dis_{pub}**	**Slope**	**Dis_{road}**	**Dis_{sta}**	**Dens**	**$Area_{cell}$**
Bellbird Park - Brookwater	-78.61	-80.21	46.22	-14.32	66.48	-4.70	125.64	-36.06
Redbank Plains	-8.88	-36.64	30.00	7.58	-27.39	-7.93	89.87	-22.75
Entire study area	-72.23	-22.94	-2.25	-17.51	35.07	-23.25	165.06	-57.50

Table 2: Comparison of average weights

According to Table 2, the importance of factors is illustrated by the absolute values of their weights. Positive and negative values indicate whether a relatively larger or smaller value makes a greater contribution to the conversion from non-residential to residential. Specifically, the population density in 2016 (Dens) is the most important factor for the transformation from non-residential to residential in sub-regions and the entire study area. Nevertheless, the difference between general and partitioned CAs can be observed from the factor of 2nd largest contribution, which is identified as distance to district centre (Dis_{cen}) in two sub-regions, as well as distance to commercial (Dis_{com}) in entire study area by general transition rules. Similarly, Dis_{com} is also the factor with 3rd largest contribution in Bellbird Park – Brookwater. Considering this, it is confirmed by the difference of weights that heterogeneity existed between sub-regions and entire study area, which leads to a further diversity of simulation results in the following. After obtaining the weights of general and partitioned CA, all simulation experiments are implemented with 100 iterations. To ensure stability of the solutions, 30 separate model runs were applied.

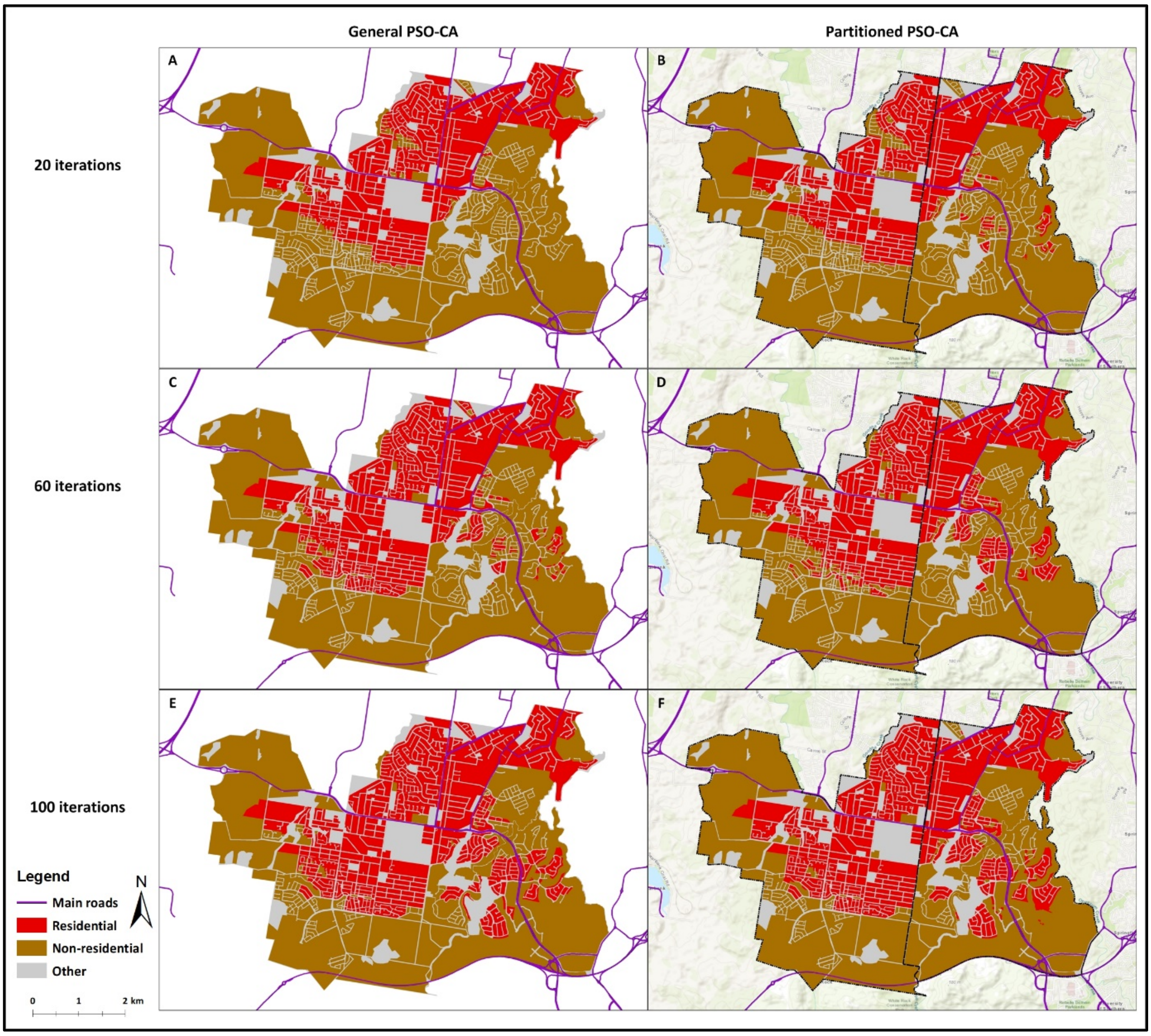

Figure 3: Simulation processes in entire/sub region(s)

Figure 3 illustrates the simulation processes of residential extension by general and partitioned PSO-CA. After the completion of 20 iterations, clustered new residential cells have been observed in southern part of previous residential area (Figure 3A) and along the Augusta Parkway of Bellbird Park - Brookwater (Figure 3B). As the simulation progresses, the new residential areas were further expanded from the previous time steps (Figures 3C and 3D). At the end of the simulation experiment, new residential cells are evident on both sides of State Road 61 and Augusta Parkway, the main roads that bisect the sub-regions. The remaining scattered non-residential cells were distributed in western and north-eastern parts of Redbank Plains (Figures 3E and 3F). Taking the real land use data as reference, the CPSA and MF of all simulation experiments are recorded in Table 3 and Figure 4.

		Experiment	Correct simulated area (ha)	PSA value (%)	CPSA value (%)	Standard deviation (%)
General PSO-CA	Bellbird Park – Brookwater	1	123.36	78.45	79.66	4.67
		2	131.32	83.51		
			...			
		29	110.91	70.53		
		30	122.31	77.78		
	Redbank Plains	1	119.46	92.17	90.45	1.41
		2	116.53	89.91		
			…			
		29	116.58	89.94		
		30	114.00	87.95		
Partitioned PSO-CA	Bellbird Park – Brookwater	1	134.25	85.37	82.99	1.97
		2	135.66	86.27		
			...			
		29	129.94	82.63		
		30	129.44	82.31		
	Redbank Plains	1	106.45	82.13	85.44	2.79
		2	107.84	83.20		
			...			
		29	111.24	85.53		
		30	113.90	87.88		

Table 3: The PSA and CPSA values of sub and entire study areas

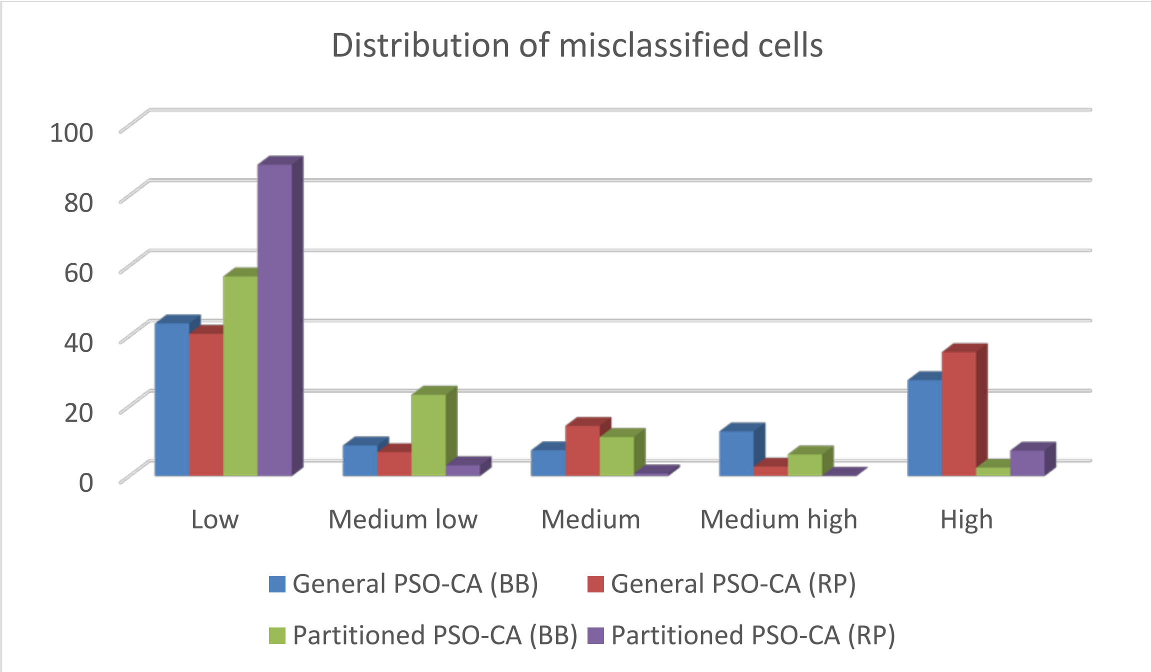

Figure 4: Ratios of misclassified cells with different frequencies (%)

* BB: Bellbird Park – Brookwater, RP Redbank Plains

4. Discussion and conclusion

In this paper, vector PSO-CA models with both general and partitioned transition rules have been proposed for the simulation of residential expansion in Ipswich City, Queensland, Australia during 1999 to 2016. According to the comparison between outputs and reference map, the values of PSA and CPSA of two sub-regions are 79.66% and 90.45% by general PSO-CA, as well as 82.99% and 85.44% in partitioned PSO-CA (Table 3). Specifically, while the difference between maximum and minimum PSA values is 25.03% for the general PSO-CA (92.48% and 67.45%), it is 13.07% (90.54% and 77.47%) for the partitioned PSO-CA. In addition, standard deviations of the simulation results are 4.67% (Bellbird Park – Brookwater, general PSO-CA), 1.41% (Redbank Plains, general PSO-CA), 1.97% (Bellbird Park – Brookwater, partitioned PSO-CA), and 2.79% (Redbank Plains, partitioned PSO-CA). Therefore, it can be concluded that general PSO-CA is more accurate in one sub-region while partitioned PSO-CA obtained a higher CPSA in another sub-region. Furthermore, the produced PSA values by partitioned PSO-CA are more stable.

Five levels of misclassification frequency have been selected, corresponding to ranges [1, 6] (Low), [7, 12] (Medium low), [13, 18] (Medium), [19, 24] (Medium high) and [25, 30] (High). As Figure 4 reveals, both general and partitioned PSO-CAs have demonstrated a “drop-rise” trend in misclassification frequency. For general PSO-CA, misclassified cells with low frequencies are the largest part, which occupied 43.61% and 40.64% in two sub-regions. Besides, 27.44% and 35.42% of the misclassified cells belong to high frequency, the second largest group in corresponding districts. The similar trend can also be detected in Redbank Plains for which partitioned PSO-CA produced 88.78% and 7.29% (Redbank Plains) misclassified cells with low and high frequencies. However, Bellbird Park – Brookwater is an exception where the misclassification frequency rates keep decreasing with the rise of frequency level. Overall, general transition rules make relatively spatially concentrated errors while the results produced by partitioned transition rules are more diverse, which corresponds to a broader range of potential solutions.

According to this case study, it is confirmed that different simulation results will be generated while vector PSO-CA models are integrated with partitioned transition rules. It is difficult to distinguish which one is better at this stage: partitioned rules improve the spatial accuracy in one sub-region (Bellbird Park – Brookwater) while general rules obtain a more accurate output in another sub-region (Redbank Plains). Future work will assess the effects across a broader study region, including the integration of both general and partitioned transition rules on a per-subregion basis.

5. Acknowledgements

Authors are thankful for valuable feedback from referees and the editor, as well as the data provided by the Australian Bureau of Statistics (ABS) and Queensland Spatial Catalogue (QSpatial). It is also appreciated for the financial support from China Scholarship Council (CSC).

6. References

Almeida, C., Gleriani, J., Castejon, E.F. and Soares ‐ Filho, B. 2008. Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science.* **22**(9), pp.943-963.

Cao, M., Tang, G.a., Shen, Q. and Wang, Y. 2015. A new discovery of transition rules for cellular automata by using cuckoo search algorithm. *International Journal of Geographical Information Science.* **29**(5), pp.806-824.

Feng, Y., Liu, Y., Tong, X., Liu, M. and Deng, S. 2011. Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landscape and Urban Planning.* **102**(3), pp.188-196.

Flache, A. and Hegselmann, R. 2001. Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. *Journal of Artificial Societies and Social Simulation.* **4**(4).

Fragkias, M., Güneralp, B., Seto, K.C. and Goodness, J. 2013. A synthesis of global urbanization projections. *Urbanization, biodiversity and ecosystem services: Challenges and opportunities.* Springer, Dordrecht, pp.409-435.

Harrison, K.R., Ombuki-Berman, B.M. and Engelbrecht, A.P. 2019. A Parameter-Free Particle Swarm Optimization Algorithm using Performance Classifiers. *Information Sciences.*

He, C., Okada, N., Zhang, Q., Shi, P. and Zhang, J. 2006. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Applied Geography.* **26**(3-4), pp.323-345.

Heppenstall, A.J., Crooks, A.T., See, L.M. and Batty, M. 2011. *Agent-based models of geographical systems.* Springer Science & Business Media.

Ke, X., Qi, L. and Zeng, C. 2016. A partitioned and asynchronous cellular automata model for urban growth simulation. *International Journal of Geographical Information Science.* **30**(4), pp.637-659.

Kocabas, V. and Dragicevic, S. 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems.* **30**(6), pp.921-953.

Li, X., Liu, X. and Yu, L. 2014. A systematic sensitivity analysis of constrained cellular automata model for urban growth simulation based on different transition rules. *International Journal of Geographical Information Science.* **28**(7), pp.1317-1335.
Li, X. and Yeh, A.G.-O. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science.* **16**(4), pp.323-343.
Liao, J., Tang, L., Shao, G., Qiu, Q., Wang, C., Zheng, S. and Su, X. 2014. A neighbor decay cellular automata approach for simulating urban expansion based on particle swarm intelligence. *International Journal of Geographical Information Science.* **28**(4), pp.720-738.
Liu, X., Li, X., Yeh, A.G.-O., He, J. and Tao, J. 2007. Discovery of transition rules for geographical cellular automata by using ant colony optimization. *Science in China Series D: Earth Sciences.* **50**(10), pp.1578-1588.
Moreno, N., Ménard, A. and Marceau, D.J. 2008. VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B: Planning and Design.* **35**(4), pp.647-665.
Moreno, N., Wang, F. and Marceau, D.J. 2009. Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers, Environment and Urban Systems.* **33**(1), pp.44-54.
O'Sullivan, D., Evans, T., Manson, S., Metcalf, S., Ligmann-Zielinska, A. and Bone, C. 2016. Strategic directions for agent-based modeling: avoiding the YAAWN syndrome. *Journal of land use science.* **11**(2), pp.177-187.
Pettit, C.J. 2005. Use of a collaborative GIS-based planning-support system to assist in formulating a sustainable-development scenario for Hervey Bay, Australia. *Environment and Planning B: planning and design.* **32**(4), pp.523-545.
Pettit, C.J., Klosterman, R.E., Delaney, P., Whitehead, A.L., Kujala, H., Bromage, A. and Nino-Ruiz, M. 2015. The online what if? Planning support system: A land suitability application in Western Australia. *Applied Spatial Analysis and Policy.* **8**(2), pp.93-112.
Pinto, N., Antunes, A.P. and Roca, J. 2017. Applicability and calibration of an irregular cellular automata model for land use change. *Computers, Environment and Urban Systems.* **65**, pp.93-102.
Queensland Government, A. 2017. *South East Queensland Regional Plan 2017.* [Online]. Available from: https://dsdmipprd.blob.core.windows.net/general/shapingseq.pdf
Şalap-Ayça, S., Jankowski, P., Clarke, K.C., Kyriakidis, P.C. and Nara, A. 2018. A meta-modeling approach for spatio-temporal uncertainty and sensitivity analysis: an application for a cellular automata-based Urban growth and land-use change model. *International Journal of Geographical Information Science.* **32**(4), pp.637-662.
Seto, K.C. and Shepherd, J.M. 2009. Global urban land-use trends and climate impacts. *Current Opinion in Environmental Sustainability.* **1**(1), pp.89-95.
Torrens, P.M. and O'Sullivan, D. 2001. *Cellular automata and urban simulation: where do we go from here?* : SAGE Publications Sage UK: London, England.
Verburg, P. and Overmars, K. 2007. Dynamic simulation of land-use change trajectories with the CLUE-s model. *Modelling land-use change.* Springer, pp.321-337.
Verburg, P.H., De Koning, G., Kok, K., Veldkamp, A. and Bouma, J. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological modelling.* **116**(1), pp.45-61.
White, R., Engelen, G. and Uljee, I. 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design.* **24**(3), pp.323-343.
Wu, F. and Webster, C.J. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design.* **25**(1), pp.103-126.

Xia, C., Zhang, A., Wang, H. and Zhang, B. 2019. Modeling urban growth in a metropolitan area based on bidirectional flows, an improved gravitational field model, and partitioned cellular automata. *International Journal of Geographical Information Science.* pp.1-23.
Xu, Z. and Coors, V. 2012. Combining system dynamics model, GIS and 3D visualization in sustainability assessment of urban residential development. *Building and Environment.* **47**, pp.272-287.
Yeh, A.G.-O. and Li, X. 2006. Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems.* **30**(1), pp.10-28.