

# **On the consistency of urban cellular automata models based on hexagonal and square cells**

Aditya Tafta Nugraha, Ben J. Waterson, Simon P. Blainey, and Frederick J. Nash

## **Abstract**

Cellular Automata (CA) have found extensive applications in the modelling of urban systems. Calculations in CA models are based on cell centroids and therefore CA models are sensitive to the choices of cell size and shape. While the effect of cell size for urban simulation has been studied, discussions on the effect of cell shape on urban CA models have been limited. Applications in other fields suggest there are advantages of using hexagonal cells over square cells, yet most urban CA models use square cells. Using connectivity indices from Graph Theory, experiments in this study compared models based on hexagonal and square cells to examine the potential advantage of hexagonal cells in urban CA models. This paper finds that simulation results from the model with hexagonal cells are more consistent and concludes that hexagonal cells would increase the robustness of model simulation.

**Keywords:** Cellular Automata, urban simulation, hexagonal cells, Graph Theory.

## **Acknowledgement**

This work was supported by Indonesia's Ministry of Finance through The Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan – LPDP) Program.

## 1 Introduction

There has been an increasing use of Cellular Automata (CA) in modelling the urban form in cities. CA are made up of a lattice of cells in an  $n$ -dimensional plane with the state of each cell being determined by its neighbouring cells. Ever since their inception, CA have found applications in a wide range of fields, including the urban system, usually adapted as constrained CA (Couclelis, 1997; Torrens and O'Sullivan, 2001). The advantage of CA models lies in their ability to represent the emergence of global structure from local interactions (Batty and Xie, 1994), which is a prominent feature in urban evolution (Batty, 1997; Batty et al., 1997, Batty, 2007).

As calculations in constrained CA are based on distances between cell centroids, CA models are sensitive to the size and shape of cells. Ménard and Marceau (2005) argued that for every context and scale of CA modelling, there is a threshold value of cell size beyond which model performance would drop disproportionately to the increase of cell size. Samat (2006) found that for an urban system this threshold is around 270m by 270m for a square cell. In terms of cell shapes, two of the most commonly used cell shapes are squares and hexagons, with squares by far the most popular in urban CA models (Torrens and Benenson, 2005).

Studies in other fields of CA application, observed that hexagonal cells produce results that are more accurate than square cells. These applications include CA models to simulate forest fire (Iovine et al., 2005; Trunfio, 2004), debris flow (D'Ambrosio et al., 2003) and chemical reaction-diffusion system (Adamantzyk et al., 2006). However, apart from several early urban models (Bura et al., 1996; Sanders et al., 1997) the use of hexagonal cells in urban models has been limited, and no studies using urban CA models have compared the performance of square and hexagon cells. Indeed, a review of relatively recent urban CA model applications (Santé et al., 2010) had only mentioned a non-urban model, Iovine and colleagues' forest fire model (2005), as an example of CA model with hexagonal cells and had not discussed how hexagonal cells may affect urban CA model performance.

This paper, therefore, seeks to investigate whether there is any difference in performance between urban CA models based on square and hexagon cells. In the next section, the use of hexagonal cells in CA models and its theoretical advantages over square cells are discussed leading to the theory that urban CA models with hexagonal cells are more robust than those with square cells. The third section discusses the design of experiment to test this theory while the fourth section provides a mathematical representation of the urban CA model used in this paper. The penultimate section presents and discusses the results of the experiment and the final section provides the conclusion.

## 2 Potential advantage of hexagon cells in urban CA models

Neighbourhood definition is an important element in CA model. In a two dimensional CA with square cells, there are two ways to define the neighbours of a cell (Batty et al., 1997). The Von Neumann system regards four orthogonally adjacent cells to be the neighbours of a central cell, whereas the Moore Neighbourhood system also regards the four diagonally adjacent cells as its neighbours, therefore defining neighbours as all eight cells surrounding a central cell.

In the von Neumann system, the centroids of all neighbours are located at the same distance from the central cell, while in the Moore system the diagonally located neighbours are further from the central cell by a factor of  $\sqrt{2}$ . This difference raises the question of whether central cell should have the same effect on diagonally located neighbours as on the orthogonally located neighbours. Treating them differently would imply a relative ease of development-sprawling in the orthogonal directions compared to other directions. In contrast, by treating diagonal and orthogonal cells as the same, there is a danger that the effect of distant cells in the diagonal direction could be overestimated. In this regard, there is an advantage in using hexagonal cells, as all cells surrounding a central cell are located at the same distance from the centre (Figure 1).

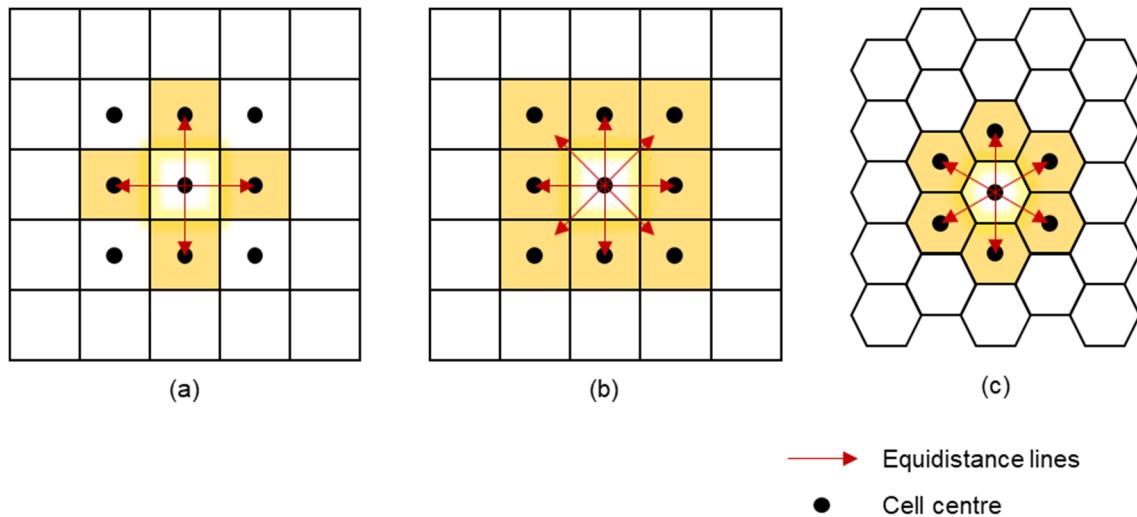


Figure 1 – Comparison of (a) von Neumann, (b) Moore, and (c) hexagonal CA configurations

Couclelis (1985) found that the conventions of original CA limit their ability to simulate complex geographical dynamics in a realistic manner. White and Engelen (1993) seconded this argument by stating that to simulate socio-economic systems such as urban and regional development, as opposed to wholly physical systems, the definition of neighbours in urban CA should go beyond immediately adjacent cells. Within this extended neighbourhood framework, the advantage of uniform distance to all adjacent neighbours in hexagonal cells becomes less clear-cut. For this reason, the use of hexagonal cells is more often found in models with more direct physical interpretations such as the forest fire model (Iovine et al., 2005) than in models with socio-economic interpretations such as urban systems.

Nevertheless, cell shapes may still have an important effect in urban CA models. This can be illustrated by taking as an example an initial urban location and two secondary points located at the same distance from the initial urban location but at a different direction in an isotropic homogenous surface. The first secondary cell is located parallel to one of the main axis of the cells while the second one is located at an angle (Figure 2 – top diagrams).

At a first time-step, the most likely cells to become urbanised are the ones located adjacent to the initial cell and thus urbanisation is likely to happen in waves from the initial location to the secondary cells through a series of adjacent cells. As these two secondary points are located at the same distance from the initial urban location, they should ideally become urbanised at the same time.

In the example given in Figure 2, the first secondary point would be likely to become urbanised after seven waves of urbanisation. With square cells, the second secondary point would likely be urbanised after either 5 or 10 waves depending on how diagonal neighbours are treated. If it is assumed that cells affect their orthogonally placed neighbours more than their diagonal neighbours (a), it will likely take 10 waves of urbanisation for the second point to be urbanised while if it were assumed that cells affect orthogonal and diagonal neighbours equally (b) it would likely only take 5 waves. With hexagonal cells, the second secondary cell would likely be urbanised at the 8<sup>th</sup> wave.

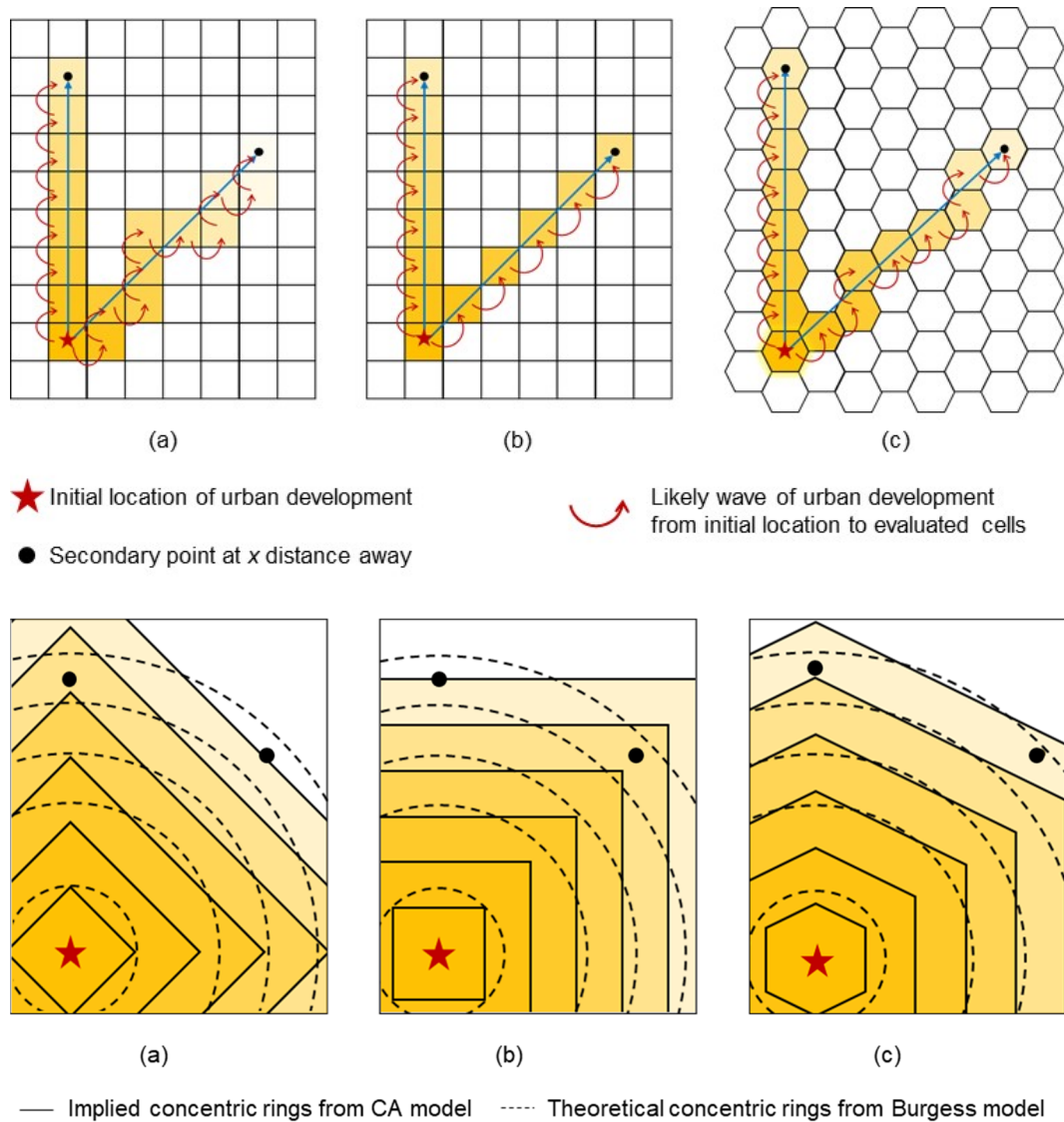


Figure 2 – Illustration on how cell shapes affect the urban concentric rings

These three examples can also be examined by comparing them with the Burgess urban land use model. In the first assumption of square-based CA model (a), the first secondary point is likely to be in a more central ring than the second one as the effect of urban centre will be more pronounced in the orthogonal than the diagonal directions. Meanwhile, a reverse condition is likely to be found in the second assumption (b). Hexagonal cells (c) are more likely to end up with the two secondary points being in the same ring as the implied concentric rings resemble more the theoretical rings from Burgess model (Figure 2 – bottom diagrams).

Note however, that the occurrence of these patterns is dependent on model formulation and cell resolution. Nevertheless, it is reasonable to expect the model with hexagonal cells to replicate the ideal situation more closely. The models with square cells, on the other hand, are more sensitive to rotation as two points located at the same distance but at a different angle are likely to be urbanised at different times. Therefore, this study proposes a theory that urban CA models with hexagonal cells are more robust than urban CA models with square cells. The next two sections will discuss the experimental design and the model used to test this theory.

### 3 Experimental design

The central objective of this experiment was to examine whether cell shapes affect the sensitivity of urban CA model to rotational placement of features. In order to do this, the model examined scenarios of urban evolution on two types of CA grids, one having hexagon cells and the other square cells. In each scenario, urban development initiated from two points separated by a distance of  $d$  on a flat featureless grid. The two initial seeds formed an axis located at angle of  $\theta$  to the north-south axis.

Calculations of CA model are based in cell centres. This meant that the model could not always obtain the exact distance of  $d$  and the exact angle of  $\theta$  for every scenario, but rather the closest approximate of them. This brought forward a question on whether one cell shape has higher ‘positioning accuracy’ than the other one. The term positioning accuracy here refers to the ability of CA model to accurately position features rather than its ability to predict the location of an activity in the future. The experiment in this study therefore had two main parts. The first part of the experiment compared the positioning accuracy and the second part examined the simulation consistency of the two cell shapes.

In this paper,  $d$  was set to 4,500m. In order to compare positioning accuracy, the study examined 91 scenarios based on the value of  $\theta$ , these were  $0^\circ$  to  $90^\circ$  inclusive with  $1^\circ$  increments. It is reasonable to assume that beyond  $90^\circ$  the simulation would repeat the same pattern for square-celled model while repetition would begin after  $60^\circ$  for hexagonal-celled model,  $90^\circ$  was chosen as it is the larger between the two numbers. For all scenarios, the model positioned the initial seeds for each scenario and examined their deviations from the target  $d$  and  $\theta$  values.

To examine simulation consistency, seven main scenarios out of 91 were chosen. These were  $0^\circ$ ,  $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ,  $75^\circ$ , and  $90^\circ$ . A hexagon and a square, respectively, would deviate furthest from  $0^\circ$  at rotational angles of  $30^\circ$  and  $45^\circ$ . After these angles, cells would begin to revert until their initial shapes reform in  $60^\circ$  (for hexagon) and  $90^\circ$  rotations (for square). An increment of  $15^\circ$  was chosen as it is the highest common factor between  $30^\circ$  and  $45^\circ$ . Seven test-points located equidistant from the initial seeds were used as references to gauge the consistency of the model (Figure 3).

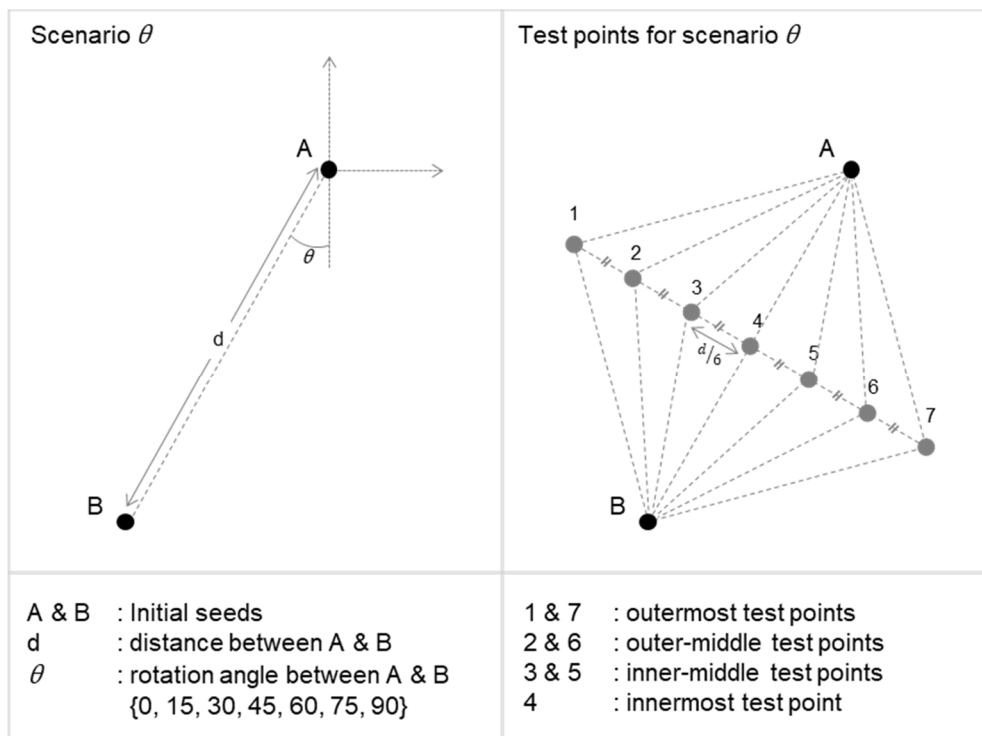


Figure 3 – Illustration of simulated scenarios and test-points

The distance between the first and last of these points was equal to  $d$  and the remaining points were distributed between these two points so that adjacent test-points were located  $d/6$  apart. The line formed of these points was deemed to be most sensitive to the effect of rotation in the given scenario as the equidistance to both initial seeds allows it to receive a balanced influence from both sides.

Robustness of the model was examined from the consistency of urban potentials in the test-points at the end of the simulation period across the seven simulated scenarios. Urban growth for these seven scenarios was simulated for 200 iterations using a potential-based urban CA model. The formulation and parameterisation of this model are discussed in the next section.

#### 4 Model formulation and parameterisation

The model used in this experiment is a CA model based on development potentials. The model calculates the potential of each cell to attract urban activities in one period and allocates new urban activities to cells based on these potentials. As in the urban CA model developed by Van Vliet and colleagues (2012), a given cell contains a set of non-negative integer values of  $Z_v$ , where  $Z_v$  represents the number of an urban activity type  $v$  populating the cell.

The model considers the effect of accessibility due to transport infrastructure, geographic features, and zone planning as well as neighbourhood interaction to calculate the potentials of a cell, in a similar way to the Metronamica model (a widely used potential-based urban CA model using square cells) (RIKS, 2010). Equation 1 generalises this by calculating total potentials of cell  $j$  for activity type  $v$  in the beginning of iteration  $t$  ( ${}^tP_{v,j}$ ) as a weighted multiplication of neighbourhood effect ( ${}^{t-1}N_{v,j}$ ), geographic suitability ( ${}^{t-1}G_{v,j}$ ), artificial (institutional) suitability ( ${}^{t-1}I_{v,j}$ ), and (transport) accessibility ( ${}^tT_{v,j}$ ) at the end of the previous iteration. Parameters  $w_1$  through  $w_4$  represent the weighting for each factors. The model calculates geographic and artificial suitability in the same way as the Metronamica model (RIKS, 2010), and uses the methods described by Aljoufie and colleagues (2013) to calculate accessibility.

$${}^tP_{v,j} = {}^{t-1}N_{v,j}^{w_1} \cdot {}^{t-1}G_{v,j}^{w_2} \cdot {}^{t-1}I_{v,j}^{w_3} \cdot {}^{t-1}T_{v,j}^{w_4} \quad (\text{Equation 1})$$

To isolate the impact of cell shapes, which would be most evident on the neighbourhood effect, the model runs on a flat featureless plane that is not under the jurisdiction of any planning authority. Therefore, the geographic and artificial suitability effects can be excluded from the model. The model assumes the same levels of transport infrastructure across all cells in the model and therefore the transport accessibility factor can also be excluded. The values for geographic suitability, artificial suitability, and accessibility weighting parameters are therefore set to zero, meaning that neighbourhood effect becomes the only determinant of cell potentials. Neighbourhood effect represents cell's attractiveness for a type of activity due to its proximity to activities in other cells. The total neighbourhood effect for cell  $j$  is the lower-bounded compound of attraction and repellent effects of activities on all cells  $i$  which are located within the neighbourhood of cell  $j$  ( $i \in J$ ) as in Equation 2.

$${}^tN_{v,j} = f \left( \sum_u \sum_{i \in J} \left( {}^tZ_{u,i \in J} \cdot A_{u,v,i \in J,j} - \lambda_1 ({}^tZ_{u,i \in J})^{\lambda_2} \cdot R_{u,v,i \in J,j} \right) \right) \quad (\text{Equation 2})$$

In this experiment, neighbourhood is defined as a circular area within a radius of 4,500m from a cell. All cells whose centre is contained within this radius are considered neighbours to the central cell. This distance was set to coincide with the distance between two initial seeds,  $d$ , so that the two initial seeds were at the neighbourhood border from each other. This was deemed the appropriate distance as it positioned the test-points in an area of medium attractiveness, which was suspected to be most sensitive to cell configuration. Cells too close to initial seeds are likely to be densely populated while cells that are too far are likely to be empty regardless of cell configuration.

The annotation ( ${}^tZ_{u,i\{i\in J\}}$ ) represents the amount of activity  $u$  in cell  $i$  at the end of iteration  $t$ . The annotations  $A_{u,v,i\{i\in J\},j}$  and  $R_{u,v,i\{i\in J\},j}$  respectively represent the attraction and repulsion effect of urban activity  $u$  in cell  $i\{i \in J\}$  for a development of urban activity  $v$  in cell  $j$ . Equation 2 implies a linear effect of the amount of activity on attraction and a non-linear one on repulsion. Van Vliet and colleagues (2012) suggest a value between zero and one for  $\lambda_1$  and a value higher than one for  $\lambda_2$ , as they would produce small initial negative externalities that grow exponentially with the amount of activity. In this experiment,  $\lambda_1$  and  $\lambda_2$  were equal to 0.3 and 1.3 respectively. The formula  $f$  transforms values lower than the lower bound to the lower bound value. As cells' potential reflect their relative probability to attract urban activity, a lower bound value of zero was used to ban negative probability values.

Every permutation pairing between activities  $u$  and  $v$ , including  $u = v$ , has one attraction and one repulsion function. These in general follow a logistic decay function with the effects being strong in short distances and weak in long distances. The attraction and repulsion effects of activity  $u$  on cell  $i(j)$  to development potentials of activity  $v$  on cell  $j$  are functions of the distance between  $i\{i \in J\}$  and  $j$ ,  $D_{i\{i\in J\},j}$ , and the compatibility between  $u$  and  $v$  which is defined by the defining parameters of the logistic decay function,  $\alpha$ ,  $\beta$ , and  $\gamma$ . These functions can be generalised as Equation 3, which is a modification of Wilson's (1967) "entropy-maximizing model" (cited by Horowitz, 2004).

$$\begin{aligned} A_{i\{i\in J\},j} &= (\alpha_A + \beta_A \cdot \text{Exp}(\gamma_A \cdot D_{i\{i\in J\},j}))^{-1} \\ R_{i\{i\in J\},j} &= (\alpha_R + \beta_R \cdot \text{Exp}(\gamma_R \cdot D_{i\{i\in J\},j}))^{-1} \end{aligned} \quad \text{(Equation 3)}$$

One general type of 'urban activity', which is best described as a combination of retail, offices, and housing, is used to isolate the model from the effect of compatibility between different land use types. The defining parameters for self-compatibility of urban activity are 0.73 and 1.87 for  $\alpha_A$  and  $\alpha_R$ , 0.30 and 2.70 for  $\beta_A$  and  $\beta_R$ , and 3.20 and 3.40 for  $\gamma_A$  and  $\gamma_R$ . These parameters, together with  $\lambda_1$  and  $\lambda_2$ , imply a land capacity of about 8,000 urban activity per cell, which corresponds to an allocation of 7.8m<sup>2</sup> land space per unit of urban activity. Beyond this level of activity, cell land potentials would be equal to zero and new activities would unlikely be assigned to that cell.

A pre-determined number of activities are added into a pool of unallocated activities at every iteration to represent urban growth. Along with these, the model takes out a pre-determined percentage of allocated activities from their current cells and put them into the pool to represent urban relocation. The model then allocates activities in the pool to cells based on cell potentials. At every iteration, the model converts potentials into relative probability of attracting a land use activity by assigning a range defined by a lower value and an upper value to each cell as defined in Equation 4.

$$\begin{aligned} {}^tLV_{v,j} &= 0 & \text{for } j = 1 \\ {}^tLV_{v,j} &= {}^tUV_{v,j-1} + 1 & \text{for } j \neq 1 \\ {}^tUV_{v,j} &= {}^tLV_{v,j} + [10^6 \cdot {}^tP_{v,j}] & \text{for all values of } j \end{aligned} \quad \text{(Equation 4)}$$

The annotations  ${}^tLV_{v,j}$  and  ${}^tUV_{v,j}$  represent the lower and upper values of cell  $j$  corresponding with land use  $v$ , at iteration  $t$ . The rounding and multiplication by 10<sup>6</sup> in the upper value calculation transform potential values from double into integer with six decimal points accuracy. The model then assigns each unallocated activity a random integer between zero and the upper value of the last cell. Activities are then allocated into a cell whose lower and upper values encompass their random number.

To compare square and hexagonal cell shapes, the CA model simulated urban growth of the seven scenarios in ten runs. In every run, the model simulated growth on two grids, one made up of regular hexagon cells and the other of square cells. Each side of a square cell is 250m long and thus a square cell has an area of 62,500m<sup>2</sup>. This cell size is within the maximum limit of 270m by 270m suggested by Samat (2005). Each hexagon cell has the same area as a square cell, in other words, each side of the hexagon is approximately 155.1m long.

## 5 Results and discussion

### 5.1 Positioning accuracy of hexagon and square cells

Across the 91 scenarios to evaluate positioning accuracy, the average absolute deviation of distances between the two initial seeds from the target value were 60m in the model with hexagon cells and 55m in the model with square cells. Meanwhile, the average absolute deviation of angle between the two seeds from the target angles were 0.75° and 0.74° respectively for hexagon and square cells. Statistical tests have been conducted to examine whether these deviations were significant from the target values.

The datasets violated normality assumption and therefore non-parametric tests, signed-rank tests, were used to evaluate distance and angle deviations. Table 1 presents the results of these tests and shows that, at 0.05 significance level, the distances and the angles between the initial seeds were not significantly different than the target values of  $d$  and  $\theta$  in both hexagon and square cells. Further, paired tests show that there is no significant difference between the absolute deviations obtained from hexagonal and square cells. It can be concluded that at the resolution used in this study, there is no difference in positioning accuracy between hexagon and square cells.

Table 1 – Positioning accuracy statistical test results

Null Hypotheses		p-value
<i>signed-rank tests</i>		
Hexagon	Median distance deviation = 0	0.211
	Median angle deviation = 0	0.672
Square	Median distance deviation = 0	0.170
	Median angle deviation = 0	0.555
Median absolute distance deviation for hexagon is equal to that of square		0.278
Median absolute angle deviation for hexagon is equal to that of square		0.723

The area of each cell used in this study is around 86% of the suggested maximum cell area size for urban CA model (Samat, 2006). Smaller cell sizes would result in higher positioning accuracy, but the results in Table 2 show that there are no positioning accuracy issues even at the level of resolution close to the maximum suggested cell size. At larger cell sizes, one of the cell shapes may come across positioning accuracy issues earlier than the other shape. However, Samat (2006) has suggested that for urban CA models, performances decrease significantly at slightly higher size. In other words, there is not much value in examining positional accuracy at larger or smaller cell size than the one used in this paper.

### 5.2 Simulation consistency

Model simulations were tested for consistency across the seven simulated scenarios. Robustness of the model was examined by a series of statistical tests to compare potentials on test-points across the seven simulated scenarios for each cell shape. As the model simulated urban growth on a flat

featureless plane with uniform transport infrastructure across all cells, given the same distance, the potentials of the test-points across all scenarios should not be affected by the angle between the initial seeds. A cell shape with less significantly different potentials on the test-points across the scenarios is less sensitive to rotation and will therefore produce a more robust model.

The average of potentials for cells containing the test-points at the final five iterations for every run populated the dataset for simulation consistency examination. By averaging, the experiment obtained a more stable measurement of potentials and thus became less likely to detect model inconsistency. Therefore, any inconsistency detected would be more likely to be caused by cell shapes rather than random factors.

Figure 4 presents the average potentials for the test-points over ten runs at every scenario. It appears that the potentials in the hexagon cells cluster more than they do in the square cells, supporting the theory that hexagon cells are more robust to rotations. Hexagon cells formed two main clusters with scenario  $30^\circ$  and  $90^\circ$  forming one cluster and the remaining scenarios forming the other. Meanwhile, square cells seemed to have at best three main clusters with scenarios  $0^\circ$  and  $90^\circ$  forming the first cluster, scenarios  $30^\circ$ ,  $60^\circ$ , and  $75^\circ$  forming the second, and scenario  $45^\circ$  being the sole member of the third cluster. Scenario  $15^\circ$  of square cells seemed to fluctuate between the first two clusters.

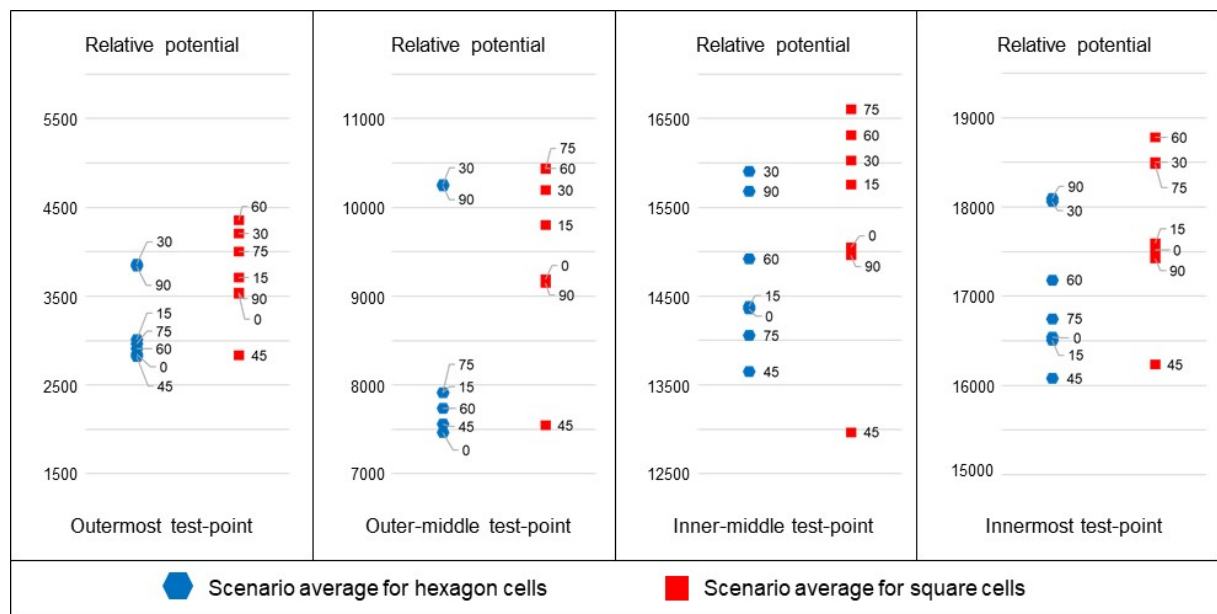


Figure 4 – Average of test-points' potentials

As the datasets were not normally distributed, a non-parametric method was employed to support this observation. A set of rank-sum tests compared the results of ten runs from pairs of scenarios to determine if there is significant difference of the test-points' potentials. For every test-point and every pair of scenarios, the null hypothesis was that there was no difference between the medians of test-point's potentials from the two scenarios, or in other words that the model predicted the potentials consistently across the two scenario. A significance level of 0.01 was used instead of 0.05 to reduce the probability of a type-I error. The resulting p-values of these tests were presented in Table 2.

Table 2 – Simulation consistency statistical test results (p-values)

Outermost		Hexagon Cells						Square Cells					
Rotation	15	30	45	60	75	90	15	30	45	60	75	90	
0	0.068	0.000 **	0.925	0.250	0.181	0.000 **	0.229	0.000 **	0.000 **	0.000 **	0.001 **	0.756	
15	-	0.000 **	0.041	0.379	0.797	0.000 **	-	0.003 *	0.000 **	0.000 **	0.034	0.120	
30	-	-	0.000 **	0.000 **	0.000 **	0.968	-	-	0.000 **	0.120	0.394	0.000 **	
45	-	-	-	0.208	0.050	0.000 **	-	-	-	0.000 **	0.000 **	0.000 **	
60	-	-	-	-	0.310	0.000 **	-	-	-	-	0.014	0.000 **	
75	-	-	-	-	-	0.000 **	-	-	-	-	-	0.002 *	

Outer-middle		Hexagon Cells						Square Cells					
Rotation	15	30	45	60	75	90	15	30	45	60	75	90	
0	0.022	0.000 **	0.561	0.050	0.017	0.000 **	0.002 *	0.000 **	0.000 **	0.000 **	0.000 **	0.903	
15	-	0.000 **	0.041	0.273	0.925	0.000 **	-	0.029	0.000 **	0.003 *	0.008 *	0.002 *	
30	-	-	0.000 **	0.000 **	0.000 **	0.756	-	-	0.000 **	0.239	0.310	0.000 **	
45	-	-	-	0.181	0.050	0.000 **	-	-	-	0.000 **	0.000 **	0.000 **	
60	-	-	-	-	0.351	0.000 **	-	-	-	-	0.925	0.000 **	
75	-	-	-	-	-	0.000 **	-	-	-	-	-	0.000 **	

Inner-middle		Hexagon Cells						Square Cells					
Rotation	15	30	45	60	75	90	15	30	45	60	75	90	
0	0.989	0.000 **	0.005 *	0.019	0.218	0.001 *	0.003 *	0.001 **	0.000 **	0.000 **	0.000 **	0.457	
15	-	0.000 **	0.001 *	0.017	0.096	0.001 *	-	0.181	0.000 **	0.022	0.002 *	0.002 *	
30	-	-	0.000 **	0.000 **	0.000 **	0.636	-	-	0.000 **	0.298	0.050	0.000 **	
45	-	-	-	0.000 **	0.053	0.000 **	-	-	-	0.000 **	0.000 **	0.000 **	
60	-	-	-	-	0.000 **	0.060	-	-	-	-	0.365	0.000 **	
75	-	-	-	-	-	0.000 **	-	-	-	-	-	0.000 **	

Innermost		Hexagon Cells						Square Cells					
Rotation	15	30	45	60	75	90	15	30	45	60	75	90	
0	0.970	0.001 *	0.212	0.140	0.571	0.011	0.678	0.031	0.009 *	0.005 *	0.021	0.678	
15	-	0.001 **	0.186	0.054	0.427	0.007 *	-	0.007 *	0.006 *	0.002 *	0.014	0.678	
30	-	-	0.000 **	0.017	0.001 *	1.000	-	-	0.001 **	0.427	0.850	0.031	
45	-	-	-	0.005 *	0.026	0.001 **	-	-	-	0.001 **	0.001 **	0.038	
60	-	-	-	-	0.162	0.140	-	-	-	-	0.473	0.006 *	
75	-	-	-	-	-	0.014	-	-	-	-	-	0.026	

(\*) significantly different at the 0.01 level

(\*\*) significantly different at the 0.001 level

The values in Table 2 support the theory that hexagon cells produce a more robust model as there are less significance differences in the potentials compared to the model based on square cells. Comparisons between 0° rotation and the other 6 scenarios confirm the initial expectation that the 30° and 90° rotations in the hexagonal-celled model produce the least similar results. However, the results from the other four scenarios in most cases were not significantly different from 0°.

Meanwhile, the potentials produced by square cells, were found to be different from the 0° scenario even before the rotation reached 45°. In the outer-middle and inner-middle test-points, they became significantly different even at 15° rotation and did not revert to similarity until the 90° rotation. Moreover, the potentials from 15°, 30°, 60°, and 75° rotations were in turn found to be significantly dissimilar in all test-points from 45° suggesting the existence of at least three clusters. These evidences support the previous argument that the hexagonal-celled model produced a lower number of clusters compared to square-celled model.

Further, graph theory was used to visualise and parameterise the results of statistical tests. For every test-point, the seven scenarios were represented as nodes and unweighted vertices connect scenarios which are not significantly different from each other in that test-point. Two graph connectivity parameters were used to parameterise the similarity of scenarios, diameter (D) and gamma-index ( $\Gamma$ ). Diameter (ranging from 1 to infinity) measures the shortest distance between the two most dissimilar scenarios in the graph. Meanwhile, gamma index (ranging from zero to one) compares the number of existing vertices to the number of potential vertices as a proportion and was used to measure how close the results were to perfect consistency. Lower diameter and higher gamma-index values indicate better performance. A model perfectly consistent across the seven scenarios would have a diameter and a gamma-index of 1 (Figure 5).

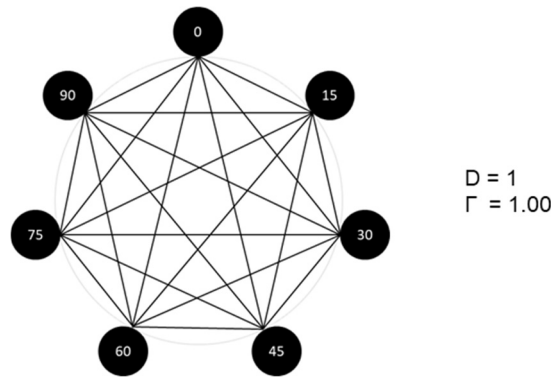


Figure 5 – Graph of a perfectly robust model across seven scenarios

Further, a weighted graph combines four unweighted graphs from different test-point locations for each model to provide combined measurements across the test-points. Weighted vertices connect scenarios not significantly different in at least one test-point. A vertex of weight 1 connects two scenarios similar across all four test-points. The weight increase by one for every test-point that the scenarios were found to be statistically different. Figure 6 provides graphical representations of the statistical tests. Thicker lines in the weighted graphs represent lower weight (higher connections) between scenarios.

From the graphs, it can be observed that diameters of model with hexagonal cells were never higher than model with square cells. Further, hexagonal cells performed consistently better in all four test-points for gamma index. In the outermost and outer-middle test-points, both cells failed to form a connected graph and therefore have infinite diameters. In the hexagon-based model, this was due to the dissimilarity of scenario  $30^\circ$  and  $90^\circ$  from any other scenarios while in the square-based model scenario  $45^\circ$  stood apart the remaining scenarios. Further, looking closely on the clusters formed, the hexagon-based model produced clusters which are perfectly connected within themselves, the subgraph formed by scenarios  $0^\circ$ ,  $15^\circ$ ,  $45^\circ$ ,  $60^\circ$  and  $75^\circ$  has a gamma index of 1.00 at the subgraph level. Meanwhile, the square cell model formed more and smaller clusters in the outer test-points ( $0^\circ$ ,  $15^\circ$  and  $90^\circ$  as one cluster and  $30^\circ$ ,  $60^\circ$ , and  $75^\circ$  as another) and lacked within-cluster consistency in the outer-middle test-point.

In the inner-middle test-point, the diameter of the square cell model remains infinite while the hexagon model's diameter improves to five. The gamma index of hexagonal-celled model was worst in the inner-middle but still outperformed the square-celled model. Both models performed best in the innermost test-point where for the first time the diameter of square-celled model dropped from infinity and its gamma index reached above 50%. In this test-point, the hexagonal-celled model matched the diameter at three and its gamma index was 67%.

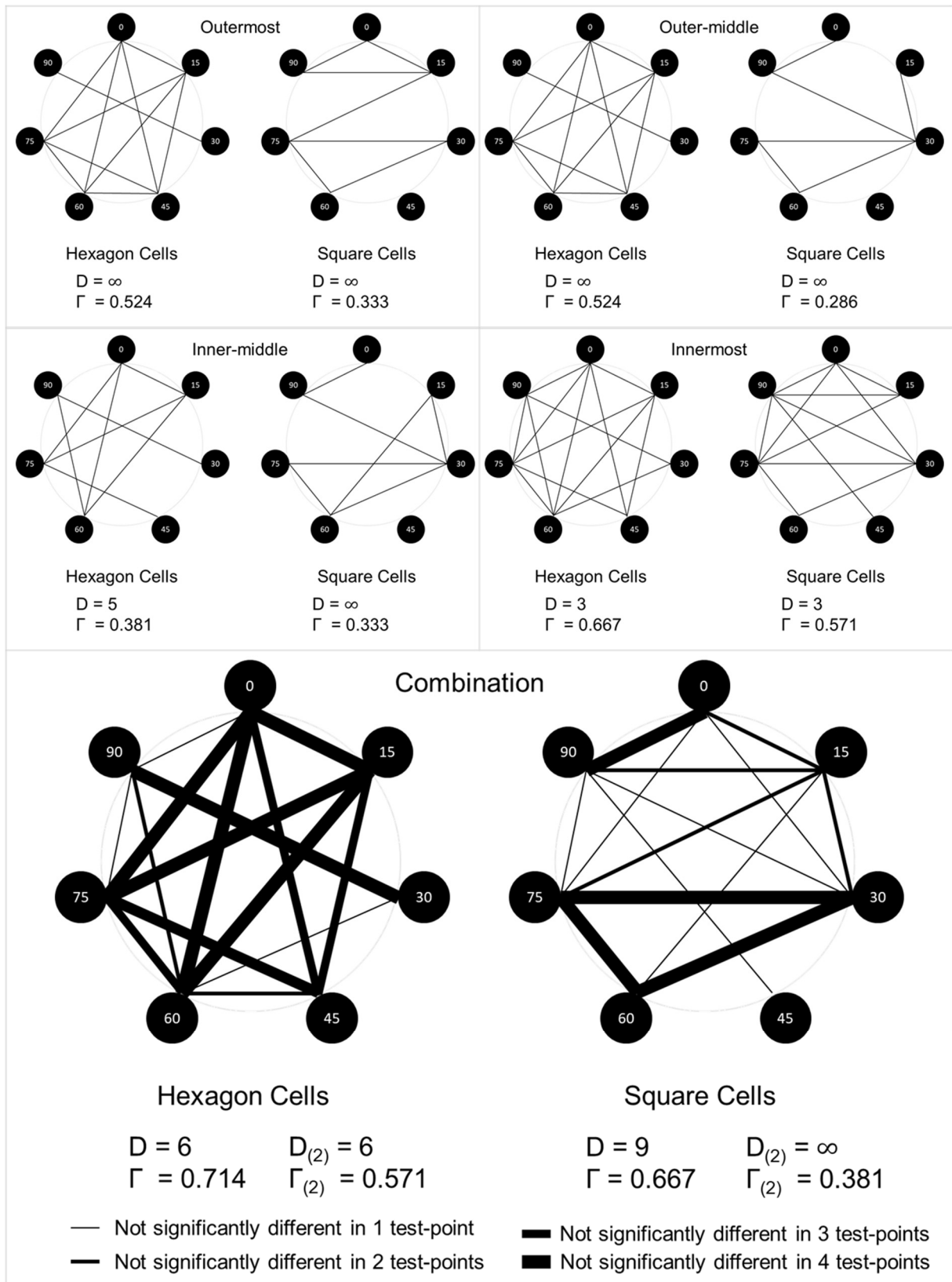


Figure 6 – Graph representations of model consistency

The gamma index of model with hexagonal cells increased to 71% in the combination graph, outperforming the square cell by 4%. The diameter of the combined graph for hexagonal-celled model is lower by 3 points compared to that of the square-celled model. Further, if the definition of 'connectedness' is made stricter so that two scenarios were declared similar if they were not significantly different in at least two test-points, the new diameter ( $D_{(2)}$ ) of the combined graph for square-celled model increased to infinity while the hexagonal-celled model retains the same diameter of 6. In this case, the new gamma indices ( $\Gamma_{(2)}$ ) for both models also suffer but more so for the square-celled model. These results suggests that hexagon cells produce more consistent simulation results than square cells.

Additionally, Figure 7 shows samples of resulting cell-populations from the simulation. A weakening of development along the axis between two urban centres can be examined in the rotated scenarios of both models but even more so in the model with square cells. This weakening is especially apparent in the 45° rotation of the square-celled model where cell population in the middle of the axis fell below 300. This was due to the cell-shape-effect pulling development orthogonally, thus weakening the development along the urban centres axis.

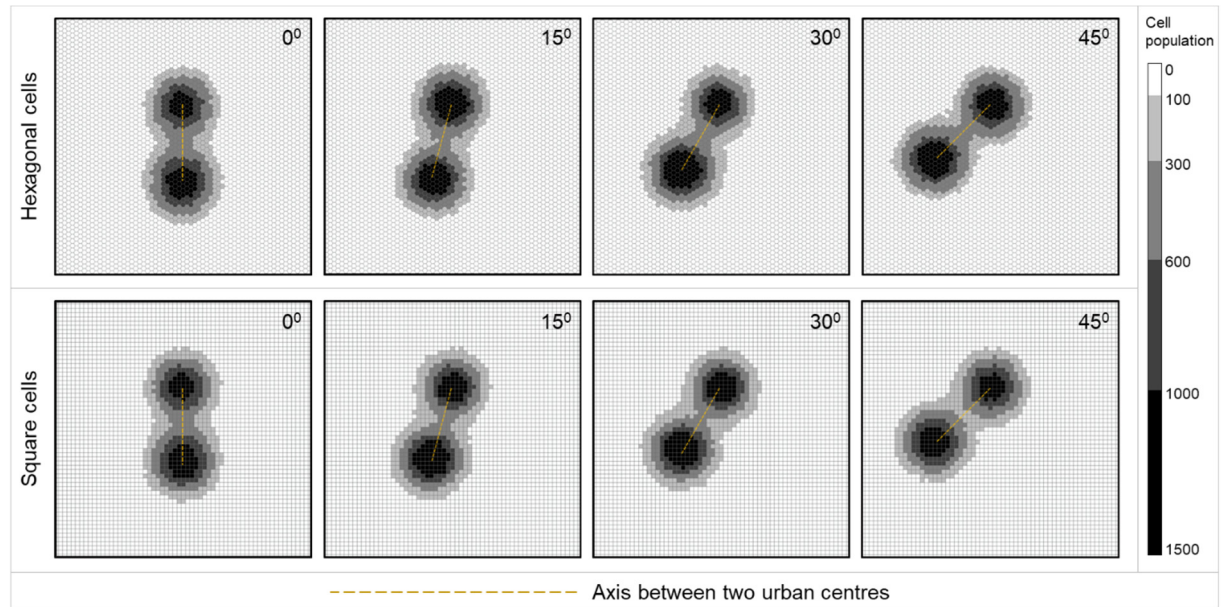


Figure 7 – Samples of simulation results

Some levels of weakening was also observed in the hexagon-celled model but development levels never fell as low even in the 30° rotation scenario which supposedly was the weakest point of the hexagon-celled model. Furthermore, development in the axis within the hexagon-celled model strengthened again in the 45° rotation. The six-fold rotational axis of hexagons, compared to a four-fold one of squares, allows a shorter weakening-strengthening cycle which in turn earns hexagon-celled model a higher consistency level.

## 6 Conclusions

This research represents one of the first studies to examine the use of both hexagonal and square cells in CA models with high socio-economic interpretation such as the urban system. The experiment used 0.01 significance level, which reduced the probability of declaring the results of two scenarios significantly different compared to the more widely used 0.05 significance level. Despite this, inconsistencies across scenarios were still found, suggesting that CA models are sensitive to rotations

of features. In this regard, more inconsistencies were found in the model made up of square cells. While the difference of test-points' potentials may be small across scenarios, as CA models simulate path-dependence of urban development (Brown et al., 2005; Batty, 2007) its effect could be magnified. If enough activities are simulated as occurring in particular areas due to cell-shape effect rather than actual urban-dynamics, those activities could attract further development in the next iterations and the emerging pattern could be misleading. While, within the simulation presented in this paper such major differences were not observed, in a model formulation with higher emphasis on distance-effects, more substantial differences are likely to occur.

This paper therefore support the theory that an urban CA model based on hexagonal cells is more robust than that based on square cells. In terms of simulation consistency, the hexagonal-celled model outperformed its square-celled counterpart across all test-points. There was no evidence to suggest that either model has higher positioning accuracy than the other does, therefore the higher consistency of hexagonal cells can be accounted to their homogenous neighbourhood layout.

The advantages of hexagonal cells are well known in highly physical CA models such as models to predict forest fire (Iovine et al., 2005; Trunfio, 2004), debris flow (D'ambrosio et al., 2003) and chemical reaction-diffusion system (Adamantzy et al., 2006). However, except for a few early geographic automata system models (for example Bura et al., 1996; Sanders et al., 1997), the use of hexagonal cells in urban CA model has always been limited. Since then, hexagonal cells seem to have further fallen out of use even though there have been no studies to justify this decline. Of course, there would be cases where square-cells would be preferred such as in the simulation of cities designed on a square grid, although this would place particular importance on ensuring cells' orientation matched the orientation of the grid. However, the use of square cells in urban CA literature is often the default and perhaps enforced by software capabilities rather than based on stronger justifications. This paper has presented evidence to suggest that the use of hexagonal cells should become the default choice for urban CA models instead, as this would provide a consistency advantage over the use of square cells in the majority of urban contexts. Therefore, urban researchers or practitioners employing the use of urban CA model need to seriously consider using hexagonal cells in their model, unless there is a clear indication that square cells are more appropriate for the urban area being modelled.

## References

- Adamantzy, A., Wuensche, A., De Lacy Costello, B. (2006). Glider-based computing in reaction-diffusion hexagonal cellular automata. *Chaos, Solitons & Fractals*, 27(2), pp.287-295.
- Aljoufie, M., Zuidgeest, M., Brussel, M., van Vliet, J., van Maarseveen, M. (2013). A cellular automata-based land use and transport interaction model applied to Jeddah, Saudi Arabia. *Landscape and Urban Planning*, 112, pp.89-99.
- Batty, M. (1997). Cellular automata and urban form: A primer. *Journal of the American Planning Association*, 63(2), pp.266-274.
- Batty, M. (2007). *Cities and complexity*. The MIT Press.
- Batty, M., Couclelis, H., and Eichen, M. (1997). Urban systems as cellular automata, *Environment and Planning B*, 24, pp.159-164.
- Batty, M., and Xie, Y. (1994). From cells to cities, *Environment and Planning B*, 21, pp.s31-s48.
- Brown, D.G., Page, S., Riolo, R., Zellner, M., and Rand, W. (2005). Path dependency and the validation of agent-based spatial models of land use. *International Journal of Geographic Information Science*, 19(2), pp.153-174.
- Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., and Sanders, L. (1996). Multiagent systems and the dynamics of a settlement system, *Geographical Analysis*, 28(2), pp.161-178.
- Couclelis, H., (1985). Cellular worlds: a framework for modeling micro-macro dynamics. *Environment*

- and *Planning A*, 17, pp.585–596.
- Couclelis, H., (1997). From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B*, 24, pp.165–174.
- D'Ambrosio, D., Di Gregorio, S., and Iovine, G. (2003). Simulating Debris Flows through a Hexagonal Cellular Automata Model: SCIDDICA S<sub>3-hex</sub>. *Natural Hazards and Earth System Science*, 3, pp.545-559.
- Horowitz, A.J. (2004). Lowry-Type Land Use Models. In Hensher, D.A., Button, K.J., Haynes, K.E. and Stopher, P.R. (Eds.) *Handbook of Transport Geography and Spatial Systems (Handbooks in Transport, Volume 5)*, pp.167-183.
- Iovine, G., D'Ambrosio, D., Di Gregorio, S. (2005). Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects. *Geomorphology*, 66, pp.287-303.
- Ménard, A., and Marceau, D.J. (2005). Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B*, 32, pp.693–714.
- RIKS (2010). *Metronamica Documentation*. RIKS BV: Maastricht.
- Samat, N. (2006). Characterizing the scale sensitivity of the cellular automata simulated urban growth: a case study of the Seberang Perai Region, Penang State, Malaysia. *Computer, Environment, and Urban Systems*, 30, pp.906-920.
- Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., and Bura, S. (1997). SIMPOP: a multiagent system for the study of urbanism, *Environment and Planning B*, 24, pp.287-305.
- Santé, I., García, A.M., Miranda, D., and Crecente, R. (2010). Cellular Automata Models for the Simulation of Real-World Urban Processes: A review and analysis. *Landscape and Urban Planning*, 96, pp.108-122.
- Torrens, P.M., and Benenson, I. (2005). Geographic automata systems, *International Journal of Geographical Information Science*, 19(4), pp.385-412.
- Torrens, P.M., and O'Sullivan, D. (2001). Cellular automaton and urban simulation: where do we go from here? *Environment and Planning B*, 28, pp.163-168.
- Trunfio, G.A. (2004). Predicting Wildfire Spreading Through a Hexagonal Cellular Automata Model. In: Sloat P.M.A., Chopard B., Hoekstra A.G. (Eds.) *Cellular Automata. ACRI 2004. Lecture Notes in Computer Science*, vol 3305. Springer: Berlin, Heidelberg.
- Van Vliet, J., Hurkens, J., White, R., and Van Delden, H. (2012). An activity based cellular automaton model to simulate land-use changes. *Environment and Planning B*, 39, pp.198-212.
- White, R., and Engelen, G., (1993). Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land use patterns. *Environment and Planning A*, 25, pp.1175–1199.