Paper accepted to the International Journal of Geographical Information Science (reference ID IJGIS-2016-0031.R2.) on 26/07/2016

Configuring the neighbourhood effect in irregular cellular automata based models

Barreira-González, Pablo (1) and Barros, Joana (2)

(1) Departmento de Geología, Geografía y Medio Ambiente, Universidad de Alcalá; Alcalá de Henares, Spain.

(2) Department of Geography, Environment and Development Studies, Birkbeck College, University of London, London, United Kingdom.

Abstract:

Cellular automata (CA) models have been widely employed to simulate urban growth and land use change. In order to represent urban space more realistically, new approaches to CA models have explored the use of vector data instead of traditional regular grids. However, the use of irregular CA-based models brings new challenges as well as opportunities. The most strongly affected factor when using an irregular space is neighbourhood. Although neighbourhood definition in an irregular environment has been reported in the literature, the question of how to model the neighbourhood effect remains largely unexplored.

In order to shed light on this question, this paper proposed the use of spatial metrics to characterise and measure the neighbourhood effect in irregular CA-based models. These metrics, originally developed for raster environments, namely the enrichment factor and the neighbourhood index, were adapted and applied in the irregular space employed by the model. Using the results of these metrics, distance-decay functions were calculated to reproduce the push-and-pull effect between the simulated land uses. The outcomes of a total of 55 simulations (five sets of different distance functions and eleven different neighbourhood definition distances) were compared with observed changes in the study area during the calibration period. Our results demonstrate that the proposed methodology improves the outcomes of the urban growth simulation model tested and could be applied to other irregular CA-based models.

Keywords: irregular space, neighbourhood, cellular automata, modelling, simulation

(1) Corresponding Author: Pablo Barreira-González, Facultad de Filosofía y Letras, c/Colegios n° 2, 28801 Alcalá de Henares, Spain. pablo.barreira@edu.uah.es

(2) Joana Barros, GEDS, Room 304, 32 Tavistock Square, WC1H 9EZ, London, j.barros@bbk.ac.uk

1. Introduction

Simulation models are a useful tool to study, understand and explore the behaviour of complex systems. Cellular automata (CA) have become one of the most widely used modelling frameworks in recent decades and have been employed to simulate phenomena such as land use change or urban growth (Barredo *et al.,* 2003; Batty, 2007; Santé *et al.,* 2010; Triantakonstantis and Mountrakis, 2012). Their success can be attributed to their capacity to reproduce the complex behaviour of dynamic systems such as cities, including aspects such as emergence, self-organisation, self-similarity and non-linear behaviour (Portugali 2000).

In CA models, space is usually divided into regular grids where each cell can have one of a set of possible states. The evolution of each cell over discrete time lapses is based on its state and the states of its neighbours, and is controlled by a set of transition rules (Wolfram, 1984). For instance, in urban growth models, states tend to represent urban land uses which affect each other, producing a new snapshot of the urban growth pattern after each iteration (Benenson and Torrens, 2004).

The rigid principles of CA models for urban systems have often been relaxed and their basic rules extended, generating more complex models (e.g. Stevens *et al.*, 2007; Petrov *et al.,* 2009) called CA-based models (Couclelis, 1997). A wide variety of relaxations have been applied to traditional CA models, including: different neighbourhood sizes and shapes (Liao *et al*., 2016), different scales (Farsaie and Hakimpour, 2014), addition of other relevant parameters such as land suitability or zoning status (White *et al*., 1997), implementation of stochastic disturbance to reproduce the uncertainty related to human activities (García *et al*., 2011; Mustafa *et al*., 2014) and alternative space representations that differ from the traditional regular grids (Stevens *et al*., 2007; Moreno *et al*., 2008).

The suitability of regular grids to represent geographic space in CA-based models is often a subject of debate within the scientific community. It has been suggested that regular grid structures are more appropriate for large scale rather than local scale models (Pinto and Antunes, 2010). The regular grid structure as an array of cells not only allows for easier calculations, but is also compatible with satellite imagery and other raster data sources, which is a very useful advantage. Nevertheless, the fact that cities consist of irregular blocks or features rather than regular cells is often highlighted in the literature (e.g. O'Sullivan, 2001b; Dahal and Chow, 2015) and there have been successful attempts to employ irregular space representation in CA-based models (e.g. Hu and Li, 2004; Stevens and Dragicevic, 2007; Moreno *et al.,* 2008; Dahal and Chow, 2014).

In addition, the basic unit for urban planning in many countries is the 'plot' or 'parcel', usually available in urban cadastres. As the aim of CA models is often to aid or improve urban planning, it would seem appropriate to adopt the same space representation as that used by urban planners. Early attempts to develop irregular CA-based models of land use change and urban growth have used cadastral parcels or similar types of space representation (Moreno *et al.*, 2008; Pinto and Antunes, 2010; Lugo and Valdivia, 2012; Dahal and Chow, 2014).

Nevertheless, models based on irregular structures can be very slow and inefficient; hence, incorporating this space effectively into CA models remains a two-fold challenge. The first of these is to deal efficiently with vector data, and the second is to define and calibrate parameters within this spatial structure. One solution to the efficiency problem, proposed by O'Sullivan (2001a) and detailed by Baetens and de Baets (2012), has been to use graph theory to abstract vector representation and reduce model processing time. This solution has also been tested and implemented by Barreira-González *et al.* (2015).

The second challenge concerns factor implementation and calibration. Neighbourhood, which is an intrinsic factor of CA models, is particularly affected by changes in spatial representation. According to O'Sullivan (2001a), neighbourhood is defined by relations of nearness between spatial elements (parcels in irregular space), where such relations depend on spatial location and the influential relationship (effects that one land use exerts on the others). Therefore, neighbourhood poses two problems in irregular CA-based models: (1) how to define neighbourhood in an irregular space and (2) how to model the neighbourhood effect between parcels with different land uses.

The first problem has recently been explored using several approaches (Stevens and Dragicevic, 2007; Moreno *et al.*, 2008; Baetens and de Baets, 2012; Dahal and Chow, 2015). In the case of CA-based models that employ regular grids, the definition of neighbourhood is very straightforward. The most usual types of neighbourhood definition are known as Von Neumann and Moore neighbourhoods, which consider some or all of the cells adjacent to a given cell. These neighbourhoods can also be extended using a radius around the central cell, as shown in Figure 1.

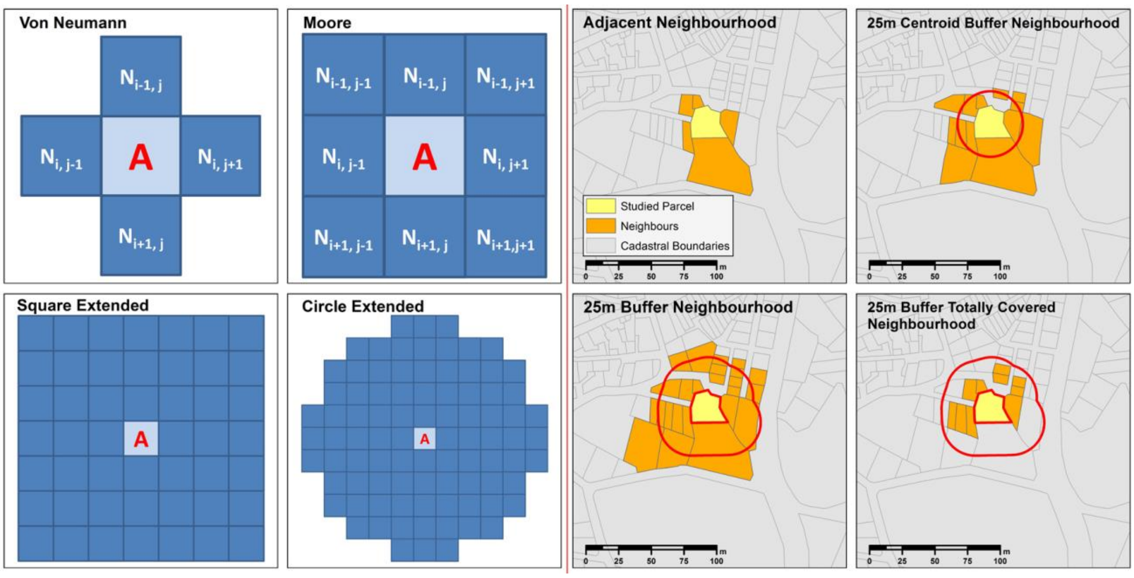


Figure 1: From regular to irregular neighbourhood definition. The neighbourhood defined for raster spaces is shown on the left, while some examples of neighbourhood in irregular spaces are shown on the right

Nevertheless, the problem becomes more complex when these same definitions of neighbourhood are implemented in irregular spaces. Stevens and Dragicevic (2007) have proposed three alternative solutions to this problem, defining the neighbours of each parcel as: a) adjacent parcels only, b) those parcels that are totally or partially covered by a distance buffer, or c) the area within a buffer. Dahal and Chow (2015) proposed another neighbourhood definition based on topological relations, proximity and intercepted buffers, and the extended neighbourhood, in which the entire study area is considered the neighbourhood of every single parcel. Figure 1 shows some examples of neighbourhood definition in irregular spaces: adjacent parcels, neighbourhood calculated by means of a distance buffer from the centroid of the studied parcel whereby parcels falling totally or partially inside the buffer are considered neighbours, neighbourhood calculated according to a buffer from the parcel boundary and another neighbourhood based solely on parcels totally covered by the buffer.

As shown in Figure 1, every cell or parcel in irregular CA-based models is potentially different from all the others in terms of size and shape, and consequently the definition of neighbourhood can also be different for each parcel: irregular parcels also present irregular neighbourhood definitions. Thus, for any given parcel, the nearest parcels would exert a greater effect than distant ones. Nevertheless, distance is not the only issue that influences neighbourhood. Each of the land uses present in a neighbourhood would exert different effects on a given cell. This combination of distance and the effect between land uses is known as the push-and-pull effect (or the neighbourhood effect). Assessment and calibration of the neighbourhood effect in irregular CA-based models remains largely unexplored.

The aim of the present study was to develop a method to measure the neighbourhood characteristics of a study area and, based on these characteristics, to generate functions that can reproduce the neighbourhood effect of a study area in an irregular CA-based model. The overall objective is to develop an approach able to model neighbourhood effect in irregular CA-based models. For this purpose, spatial metrics originally developed for raster environments, namely the enrichment factor (Verburg *et al.*, 2004a) and the neighbourhood index (Hansen, 2012), were adapted and applied to measure the neighbourhood effect in irregular space (Geertman *et al.*, 2007; Hagoort *et al.*, 2008) in a past time period. These metrics were calculated enabling the generation of functions that reproduce the neighbourhood effect that pushes parcels to change their land use. Obtained functions were implemented and tested in an irregular CA-based model.

The following section describes the methodology, providing an explanation of the irregular CA-based model employed. Section 3 presents the study area as well as a study of urban growth between 2000 and 2010 in the same region (3.1), and then gives the metrics results (3.2) and reports their implementation through functions in the model that reproduce the neighbourhood effect (3.3). These functions were tested in the model, simulating urban growth and comparing the results against observed growth (3.4).

Finally, section 4 discusses the results obtained and section 5 presents the conclusions of this study.

2. Methodology

The methodology proposed here explores the neighbourhood effect within an irregular CA-based model using vector spatial metrics. Configuring the neighbourhood effect accurately is part of the calibration process of every CA model. Thus, this study focuses on the period 2000–2010, which was also the model calibration period.

The first step consisted of analysing land use change in the study area in order to identify parcels in which land use had changed during the study period as well as those which had remained stable. Spatial metrics, namely the enrichment factor (Verburg *et al.*, 2004a) and the neighbourhood index (Hansen, 2012), were adapted from regular to irregular space and applied to measure neighbourhood characteristics. Averages were computed considering, a) all parcels and b) only parcels that had changed. Based on the results of the metrics, four sets of distance-decay functions were generated (one per metric and per options a and b) and implemented in the model for testing. Sensitivity tests were carried out by running the model for the period 2000–2010 with a combination of the set of functions and different distances as the neighbourhood definition. The results were compared with observed growth in the same period, and the set of functions and the distance which best fit the observed growth were then selected for use in the model.

2.1. The irregular graph CA model

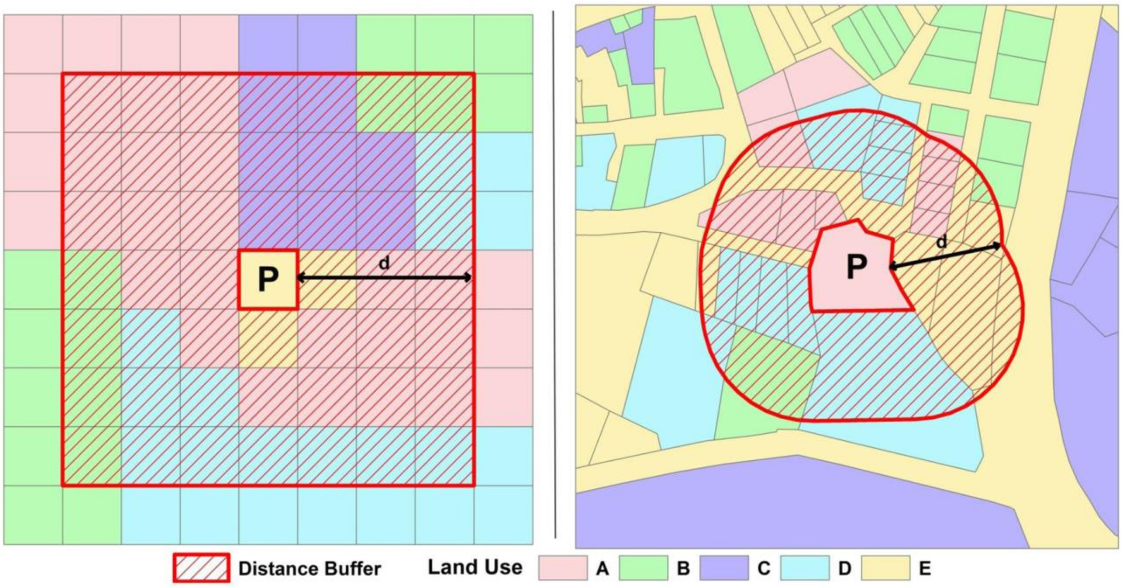
The irregular CA-based model employed here (Barreira-González *et al.*, 2015) was developed to simulate land use change in urban areas in Spain. The model simulates land use change for two urban land uses: residential and productive (commercial and industry). Rather than using a regular grid, it employs the irregular structure of cadastral parcels, where each individual plot or parcel is represented by a polygon. Although the version of the model used here is a prototype which has not yet been fully validated, it provides a suitable platform to conduct an in-depth study of the neighbourhood effect, which was the aim of the present study. The model uses four parameters (neighbourhood, accessibility, suitability and zoning status) to model urban growth, based on the NASZ modelling schema (White *et al.*, 1997).

In the model, *neighbourhood* effect reproduces the push-and-pull effect exerted by a land use in a specific location on other locations. It has two components: definition and effect. The model uses buffers around each parcel to define neighbourhood, which is computed from the intersection of the buffer with the vector dataset of parcels. Parcels that fall partially or completely within the buffer of a given parcel are considered neighbours. Distance-decay functions are used to reproduce the neighbourhood effect. *Accessibility* measures the ease with which someone could access the road network from a specific location. For a given parcel, the model calculates the Euclidean distance to the closest road network measured from the edge of the parcel.

Suitability represents the intrinsic capacity of a given location to develop a specific land use due to its characteristics. In the model, a suitability map that includes factors such as slope, height, hydrography or current land uses yields a suitability value for each parcel. Finally, the *zoning status* specifies where the law allows land uses to be developed or not. This information is usually obtained from urban plans. For each parcel, the model assigns its real zoning status according to the current urban plan available for the municipality.

The use of irregular space requires significantly more computing power and memory than grid-based models in order to store all relationships between parcels (neighbourhood definition) in the vector dataset. For this reason, the model proposed by Barreira-González *et al.* (2015), following O'Sullivan's approach (2001a), uses graph theory to reduce computational time as well as to extract neighbourhood definition in combination with vector analysis tools, which are standard in most GIS packages.

A graph is a combination of elements named nodes (usually denoted as V) and their relationships, called edges (E). Edges connect two nodes that present a relationship condition (Felzenszwalb and Huttenlocher, 2004). In the case of the model reported here, a graph was built from the cadastral structure in which each parcel is represented as a node and the edges in the graph represent the neighbourhood relationship between parcels. If two parcels are linked by a common edge, they are considered neighbours (Figure 2). As shown in Figure 2, nodes are located in the centroid of each parcel solely to provide a spatial representation.



	RASTER (cells)				VECTOR(m^2)			
LAND USE	BUFFER	STUDY AREA	$F_{P,d}$	$NI_{P,d}$	BUFFER	STUDY AREA	$vF_{P,d}$	$vNI_{P,d}$
A	21	31	1.14	0.44	849	1,841	1.82	0.15
B	6	16	0.63	0.13	476	2,891	0.65	0.08
C	8	10	1.35	0.17	0	4,495	0.00	0.00
D	11	21	0.88	0.23	2,279	4,202	2.14	0.41
E	2	3	1.13	0.04	2,010	8,754	0.91	0.36
TOTAL	48	81			5,614	22,183		

Figure 2: The figure on the left represents the neighbourhood in the irregular CA-based model. The figure on the right represents abstraction of the neighbourhood and parcels to a graph.

The first step in the model setup is to calculate suitability, accessibility and zoning status (which will remain stable throughout model iterations) for each parcel and update this information in the attribute table. A graph is then generated, creating one node per each parcel and saving the parcel's attributes in the node. Next, edges are created to generate the neighbourhood definition. Based on a distance entered by the user, the model defines the neighbourhood for each parcel as those parcels that fall completely or partly within the buffer computed around the parcel. This neighbourhood definition is based on the one proposed by Stevens and Dragicevic (2007). The model works iteratively: each iteration corresponds to a calendar year and in every iteration, a fixed amount of land (demand) per land use will be developed. There is a specific demand for residential land use and another different for productive land use.

The first iteration begins with the calculation of the neighbourhood effect from distance-decay functions. For a given parcel, each of its neighbours will exert different effects depending on their land use and relative distance. Subsequently, the model identifies parcels in the graph which are candidates for development into an urban use, and a potential development value is calculated for each of these candidates by combining the four parameters implemented (one potential value for residential land use and another for productive use). The model ranks potentials from the highest to the lowest and then decides which parcels will be converted to urban use until the annual demand is met. A parcel may be developed for either residential or productive land use depending on which potential is higher. Parcel development is independent of its size (area). This version of the model does not implement a parcel subdivision algorithm. As the cadastre is a land property database (each parcel belongs to an individual or entity), the model assumes that each parcel has a single use, and this land use is assumed to be homogeneously distributed throughout the parcel. Parcels that are developed in each iteration are updated in the database, serving as input for following iterations. A more detailed account of model implementation is given in Barreira-González *et al.* (2015).

2.2. Characterising the neighbourhood effect using spatial metrics

Spatial metrics can be used to analyse land use patterns, spatial distribution and, in this case, the neighbourhood effect (Geertman *et al.*, 2007). Verburg *et al.* (2004a) proposed the enrichment factor as a generic metric to measure, quantify and understand neighbourhood characteristics in regular grid spaces (raster). This metric, denoted as $F_{i,k,d}$ in Equation (1), is defined as the occurrence of a land use type k in the neighbourhood defined by distance d of a location i relative to the occurrence of the same land use in the study area. In other words, it compares the proportion of a land use in the local neighbourhood with the proportion of the same land use in the entire study area:

$$F_{i,k,d} = \frac{n_{i,k,d}/n_{i,d}}{N_k/N} \tag{1}$$

where $n_{i,k,d}$ represents the number of cells with land use k in a neighbourhood defined with radius d, $n_{i,d}$ represents the number of cells in the neighbourhood of radius d, N_k represents the number of cells with land use k in the entire study area and N is the number of cells in the study area. This metric is often used as an indicator of neighbourhood characteristics (see Verburg *et al.*, 2004a; Verburg *et al.*, 2004b; Geertman *et al.*, 2007; Hansen, 2012; Pan et al., 2010; van Vliet *et al.*, 2013) and to define the push-and-pull effect functions as part of the neighbourhood rules (Hagoort *et al.*, 2008).

Hansen (2012) proposed an alternative metric called the neighbourhood index (*NI*), as shown in Equation (2). This provides the proportion (values ranging from 0 to 1) of a specific land use within a given neighbourhood definition. The metric is basically calculated by dividing the number of cells in the neighbourhood with a specific land use ($n_{i,k,d}$) by the total number of cells in that neighbourhood ($n_{i,d}$), as shown below:

$$NI_{i,k,d} = \frac{n_{i,k,d}}{n_{i,d}} \tag{2}$$

Although *F* and *NI* have been successfully employed for raster, they have not yet been adapted to suit irregular structures. In a regular space, calculation of these metrics is merely a question of counting cells, but in irregular spaces the number of cells can be replaced by the amount of area measured with vector analysis tools: a distance buffer (reproducing the extended neighbourhood definition) can be generated from the edge of the parcel in question and the area that each land use occupies within the buffer can be quantified through the intersection of the buffer and the parcels shapefiles. Equations (3) and (4) show how the enrichment factor and neighbourhood index have been adapted to irregular space and renamed the vector enrichment factor (*vF*) and vector neighbourhood index (*vNI*).

$$vF_{i,k,d} = \frac{a_{i,k,d}/a_{i,d}}{A_k/A} \tag{3}$$

$$vNI_{i,k,d} = \frac{a_{i,k,d}}{a_{i,d}} \tag{4}$$

In Equations (3) and (4), for a given parcel i, a buffer is generated with radius d measured from the parcel boundary (Ballestores and Qiu, 2012). The area a specific land use k occupies in the buffer is calculated ($a_{i,k,d}$), where $a_{i,d}$ is the total area covered by the buffer. For the vector enrichment factor, A_k represents the total area occupied by land use k in the study area, and A is the total area of the study area. Figure 3 shows an example of how the original metrics would be calculated (on the left hand side) and how they would be adapted to vector spaces.

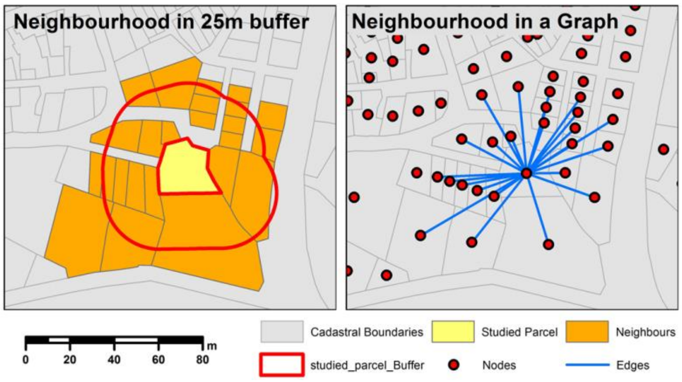


Figure 3: Spatial metrics calculation in regular and irregular environments.

It is important to note that the use of *F* to measure the neighbourhood characteristics of a cell can generate biased results. The values obtained for this metric can be read as over- or under-representing a land use in a specific location in relation to the proportion of the same land use in the overall study area. Thus, when the study area presents a dominant land use, the values derived from *F* for land uses with a lower presence can be extremely high. Therefore, although the use of two metrics might initially seem redundant, their results are complementary. In other words, the combined results of these two metrics facilitate accurate measurement and understanding of neighbourhood characteristics from both global (*vF*) and local (*vNI*) perspectives.

2.3. Converting spatial metrics results into neighbourhood effect functions

The neighbourhood characteristics and metrics serve as basis for developing neighbourhood rules in the irregular CA-based model, following the methodology proposed by Hagoort *et al.* (2008). The model includes the neighbourhood effect implemented through distance-decay functions which attempt to reproduce the effect that land uses exert on neighbouring locations. These functions can be obtained from the spatial metrics results, where a polynomial function is adjusted to the distribution of the values of a spatial metric over different distances. The function is different for every land use simulated (e.g. one function can represent the effect that residential use has on other parcels as regards development for residential use).

Distance-decay functions are easily entered into the model code. To assess the model's capacity to accurately reproduce urban growth between 2000 and 2010, four sets of distance-decay functions were entered: A) functions derived from *vNI* values for parcels that had changed during the calibration period per land use; B) functions derived from *vNI* values obtained for all parcels per land use; C) functions derived from log(*vF*) values for parcels that had changed during the calibration period per land use, and finally; D) functions derived from log(*vF*) values obtained for all parcels per land use.

2.4. Implementing neighbourhood effect functions in the simulation model

The four sets of functions (A, B, C and D) were tested in the model in combination with different neighbourhood definitions (by using different buffer distances), which can highlight the sensitivity of the model to changes in this parameter (Al-Ahmadi *et al.*, 2009). This can also be seen as the calibration process for the neighbourhood effect. Calibration can be understood as adjusting the parameters to improve the model's goodness of fit (Rykiel Jr., 1996; Petrov *et al.*, 2009; van Vliet *et al.*, 2011). In the case of the model employed here, the aim of calibration was to obtain the most suitable values for the transition rule parameters in order to reproduce the land use change processes that had occurred in the past (Santé *et al.*, 2010).

In this case, the most suitable values depended on the distance selected for the neighbourhood definition and on the set of functions employed to reproduce the neighbourhood effect. A total of 11 different distances combined with the 4 sets of functions (A, B, C and D) yielded 44 simulations for 2010 that were then compared against observed growth for the period 2000–2010. In order to demonstrate model performance using neighbourhood metrics as the calibration method, results from a previous version of the model, the Business as Usual simulation (BAU), were also compared against observed growth. For the BAU, the same definition of neighbourhood was employed. The neighbourhood effect functions were derived from neighbourhood effect masks implemented in a previous version of this CA-based model, which employed a regular grid to represent space and was used to analyse the same study area (see Barreira-González *et al.*, 2015).

A number of methods can be used to compare the model's resulting maps against reality, including visual comparison, which can be used to conduct an exploratory analysis (Pinto and Antunes, 2007). For the present analysis, a quantitative method was deemed necessary to evaluate model performance appropriately. Thus, the percentage of agreement (PA) shown in Equation (5), similar to producer's and user's accuracies (Congalton and Green, 2008), was adopted to quantify agreement between simulated and observed growth.

$$PA_k = \frac{\sum_{i=o}^{i=m} Area_{i,k,sim}}{\sum_{j=o}^{j=n} Area_{j,k,obs}} \quad ; \quad \forall Parcel\ where\ k, sim = k, obs \tag{5}$$

In Equation (5), PA for a land use k is calculated as the sum of the areas of those parcels in which observed and simulated growth for 2010 of land use k coincides, divided by the total area of observed growth for the same land use for 2010. PA was selected for this study, rather than producer's and user's accuracies, because the results obtained by the latter would have the same value for both indices in each simulation (see Barreira-González *et al*., 2012), whereas PA provides a single value.

3. Results

3.1.Land use change

Los Santos de la Humosa is a municipality located in a region of Spain which has experienced intense urban sprawl in recent decades (Díaz-Pacheco and García-Palomares, 2014), rendering it particularly suitable for the study of urban growth. Its size allows for a detailed study and measurement of neighbourhood characteristics. This municipality, shown in the map in Figure 4, has a total area of 34.9 km^2 and contains approximately 4,000 parcels.

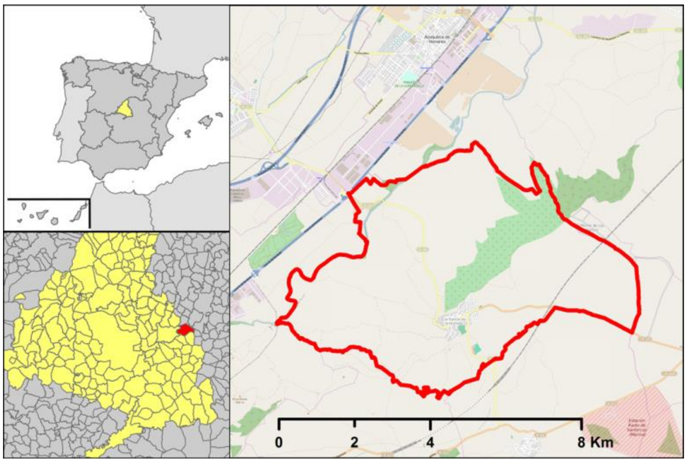


Figure 4: Study area. Source: compiled from Open Street Maps 2015.

The data source employed is the cadastral database in shapefile format provided by General Directorate for Cadastre (2013), which corresponds with land registry, each parcel represents an administrative unit that belongs to an owner, who pays taxes that vary according to its size and land use. The information contained includes area, cadastral ID and the year of development of each parcel, enabling identification of land use change in each parcel over the period 2000–2010. Historical satellite imagery (Nomecalles, 2015) was employed to determine the land use present in each parcel in the years 2000 and 2010. Zoning status areas were obtained from the Urban

Development Plan (Gómez-Vilarino and Gómez-Orea, 2013), commonly named as PGOU in Spain, and then assigned to each parcel in the municipality.

During this period, cumulative urban growth for the study area was over 38%, of which 88% was residential growth and 12% was productive growth (industrial and commercial areas) (see Figure 5). Most of the new residential parcels are located in the urban centre, infilling previously existing gaps, but there are also new residential areas in the surroundings of the urban centre, where vacant land has been converted into urban land. The growth of industry was not very significant between 2000 and 2010, and the few new industrial parcels are all located outside the urban centre.

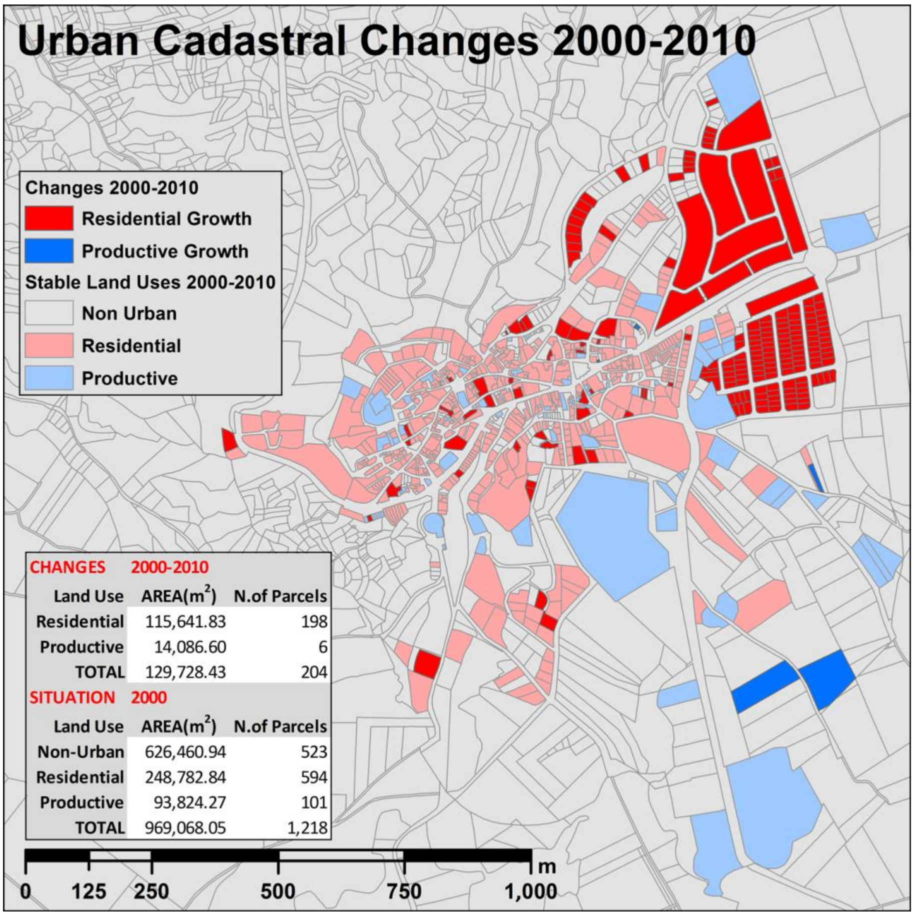


Figure 5: Urban land use changes over the period 2000–2010. The amount of area that has changed is shown in the table, where "No. of Parcels" refers to the number of cadastral parcels. Source: General Directorate for Cadastre (2013).

3.2.Spatial metrics results

Both spatial metrics (*vF* and *vNI*) were calculated for each urban cadastral parcel in the study area. Different distances ranging from 25 m to 500 m were selected for buffers in order to gain a better understanding of how distance affects the tendency towards land use change. Values of *vF* are presented in logarithmic scale for ease of understanding: values over 0 indicate that a land use is over-represented in a specific neighbourhood and values below 0 indicate that it is under-represented in the same neighbourhood. Figure 6 shows the global tendency of these spatial metrics for the urban parcels in the study area.

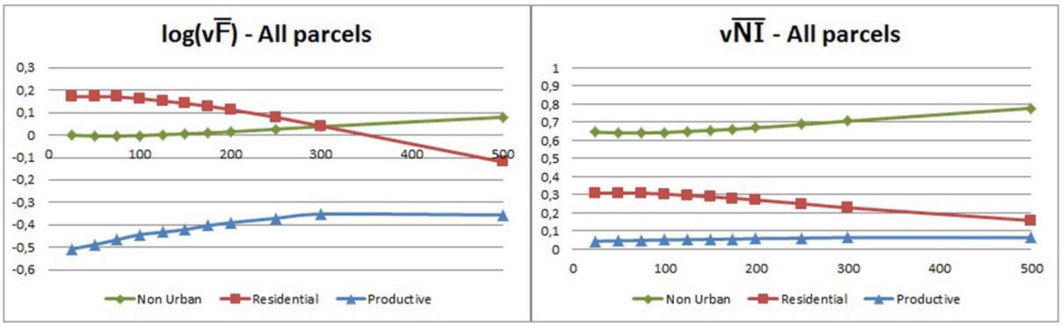


Figure 6: Average of vF and vNI for urban parcels. Each line represents trends in each land use in the neighbourhood of parcels from distance 25 m to 500 m.

In this study, residential land use was over-represented in the neighbourhood of parcels up to a distance of 300 m (values of log(*vF*) over 0). Values for non-urban land use were close to 0 until reaching 300 m, where the line for non-urban land use intersected with that for residential land use. This could be related to the size and shape of the urban centre. Values of log(*vF*) for productive use showed that this land use was under-represented throughout the entire range of distances as a consequence of its low presence in the study area. The proportion of these three land uses over the distances 25 m to 500 m is represented by the *vNI* metric: productive land use appeared stable throughout the entire range of distances, residential use varied from 32% to 17% and non-urban use varied from 65% to 78%.

Figure 7 shows the metrics which were calculated differentiating parcels by their land use. Parcels with a non-urban land use presented a high proportion of the same land use at any distance, as shown in the *vNI* metric as an almost horizontal line (80% of the land in the neighbourhood was non-urban). Within their neighbourhood definition, they also presented the same proportion at any distance of residential (18%) and productive (2%) land uses. Parcels with a residential land use showed a constant decrease in terms of *vNI* and *vF* throughout the selected range of distances. This indicates that residential parcels tend to be located close together, and these values fall when the urban centre boundaries are examined. The *vF* values obtained indicate that residential land use was over-represented until reaching a distance of 400 m, which would be the urban boundary. They also revealed that productive land use was considerably under-represented, suggesting that productive parcels tend to be located outside the urban centre. Finally, parcels with a productive land use were also under-represented in their neighbourhood (*vF* values below 0), which suggests they are dispersed, but that they also have some residential areas nearby (*vF* value of 0.2 from 25 m to 200 m).

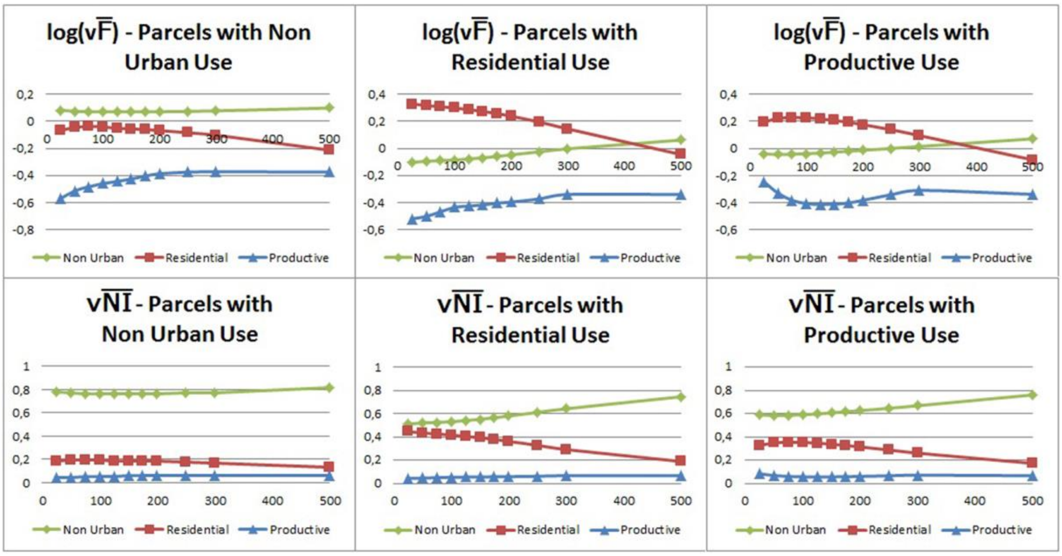


Figure 7: Average of vF and vNI for urban parcels divided into the land use that they present.

The metrics employed seem to provide a good basis for calibrating the push-and-pull effect in the model. Figure 8 shows the spatial metrics *vF* and *vNI* for parcels that were developed for urban land uses over the period 2000–2010.

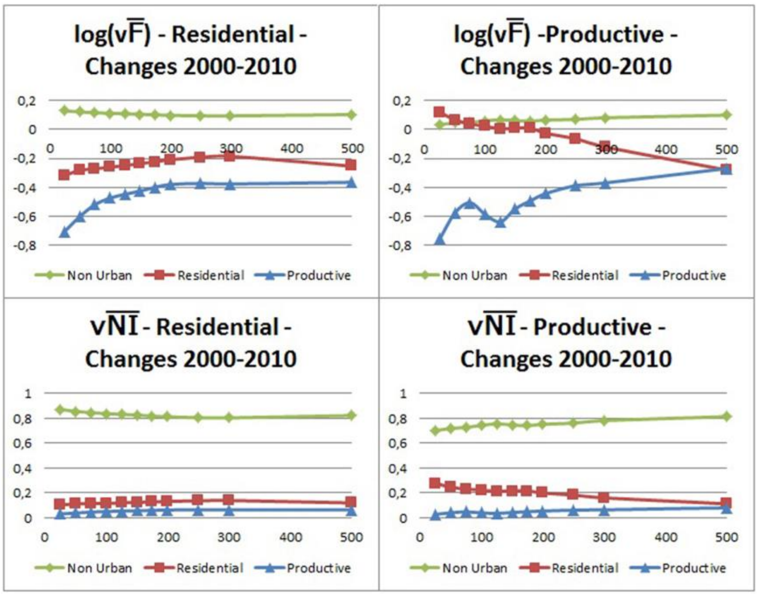


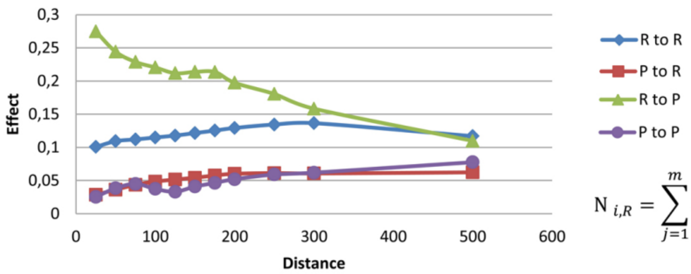
Figure 8: Average of vF and vNI for parcels that were developed for an urban use between 2000 and 2010.

Parcels that were converted into residential land use showed neighbourhoods in which the most representative use was non-urban (see non-urban values of *vNI* over 0.80 throughout the range of distances). This could be interpreted as indicating dispersed growth: new residential or productive parcels are not located adjacent to pre-existing urban uses, but seem instead to be more isolated. Nevertheless, the high proportion of non-urban land use in relation to the other land uses in the study area explains these high values of non-urban land use for *vNI*.

Similarly, parcels which were converted into productive land use also presented *vNI* values over 0.75. However, residential land use was slightly over-represented in their neighbourhood (*vF* value of 0.15 for the shortest distances), which suggests that productive growth is located close to the urban centre. In this case, *vNI* values were also higher for residential land use.

3.3. *Generating neighbourhood effect functions*

Four sets of functions were tested in the model. In order to explain how they were obtained, the set of functions A will be used as example. Note that the same methodology was employed to obtain the other sets of functions (B, C and D). The set of functions A consists of 4 functions (1, 2, 3 and 4). The neighbourhood effect measured by *vNI* in parcels developed for a residential land use (Figure 8, bottom left) provided two functions (1, 2): the residential line (in red) was used to obtain the function that reproduces the effect exerted by a residential use on a given location in favour of developing a residential land use, while the productive line (in blue) was used to obtain the function that reproduces the effect exerted by parcels with a productive use on a given location in favour of developing a residential land use. The neighbourhood effect measured by *vNI* in parcels developed for a productive land use (Figure 8, bottom right) provided two more functions (3,4): the function which reproduces the effect exerted by residential use on a given location in favour of developing a productive land use was obtained from the residential line, while the function that reproduces the effect exerted by parcels with a productive use on a given location in favour of developing a productive land use was obtained from the productive line (Figure 9).



$$E_{R\to R} = -4 \cdot 10^{-7} \cdot d^2 + 0.0003 \cdot d + 0.0944$$
$$E_{P\to R} = -3 \cdot 10^{-7} \cdot d^2 + 0.0002 \cdot d + 0.0271$$
$$E_{R\to P} = 2 \cdot 10^{-7} \cdot d^2 - 0.0004 \cdot d + 0.2694$$
$$E_{P\to P} = -5 \cdot 10^{-8} \cdot d^2 + 0.0001 \cdot d + 0.0268$$

$$\mathrm{N}_{i,R} = \sum_{j=1}^{m} E_{R\to R} + \sum_{k=1}^{n} E_{P\to R} \quad \mathrm{N}_{i,P} = \sum_{j=1}^{m} E_{R\to P} + \sum_{k=1}^{n} E_{P\to P}$$

Figure 9: Set of functions A, derived from values of vNI for parcels developed for an urban use between 2000 and 2010. Each equation expresses the effect that a location with a land use (residential or productive) at a distance ‘d’ exerts on a given location ‘i’

These four functions constitute the set of functions A. The model was run using one set of functions at a time (A, B, C or D). Initially, the model identifies the neighbours of a given parcel, and calculates each neighbour’s effect on the parcel in favour of developing a residential land use ($E_{i,R}$), and a productive use ($E_{i,P}$), using the corresponding function according to its land use. Each neighbour is located at a different distance and has a different land use, so each one will exert a different effect on the given parcel ‘i’. Effects exerted by non-urban land use remain constant in this version of the model. As there is more than one value of $E_{i,R}$ and $E_{i,P}$ per parcel, these must be summed to obtain a global value that expresses the neighbourhood effect that contributes to developing a residential use ($N_{i,R}$, Figure 9). The same calculation would be performed for productive land use ($N_{i,P}$).

3.4. Testing the neighbourhood effect functions and simulation results

Table 1 shows the percentage of agreement between each simulation for 2010 and observed urban use in the same year. The best PA values for residential use were obtained at medium distances. The values for productive use were distorted by two large parcels that represented 95.90% of productive growth, which is why PA values were only 0%, 52.98% and 95.90%. Overall, the best agreement for both land uses was given by the set of functions B at a distance of 200 m, in which the PA between simulated land uses and real growth was 64% and over 90% for residential and productive uses, respectively. The model reproduced almost 68% of the real growth. When stable residential and productive land uses were included in the comparison, PA values increased up to 88% and 99%, respectively. A comparison of the results obtained for functions A, B, C and D, with those obtained using the BAU model clearly shows that the methodology improved overall model performance.

Table 1: PA between simulated and real urban growth for 2010. Four different sets of functions were tested (A, B, C and D). A previous version of the model (BAU) was also compared against real growth, using different neighbourhood definitions from 25 m

Distance(m) / Functions Set	25	50	75	100	125	150	175	200	250	300	500
A	13.31	19.64	26.96	53.87	46.75	61.19	63.20	52.40	53.31	45.88	35.22
	0	0	0	0	0	0	52.98	0	52.98	52.98	52.98
B	15.39	31.06	44.03	55.78	55.46	60.92	52.06	64.12	55.90	49.50	37.08
	0	0	0	52.98	95.90	95.90	95.90	95.90	52.98	52.98	52.98
C	45.54	42.99	42.47	42.59	42.59	42.80	42.18	42.16	41.99	43.93	42.14
	95.90	95.90	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98
D	40.51	40.02	46.01	35.93	36.51	56.83	48.55	41.25	42.56	39.21	41.80
	52.98	95.90	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98
BAU	32.12	32.13	32.19	31.59	31.59	31.59	31.59	31.59	31.59	32.13	32.13
	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98	52.98

Figure 10 shows the spatial location of the best PA values (200 m and set of functions B), combining both simulated land uses into what we have called urban. In terms of location, agreement was mostly located on the eastern side of the urban centre, where new concentrated residential areas were built. However, the model failed to simulate growth filling gaps in the urban centre.

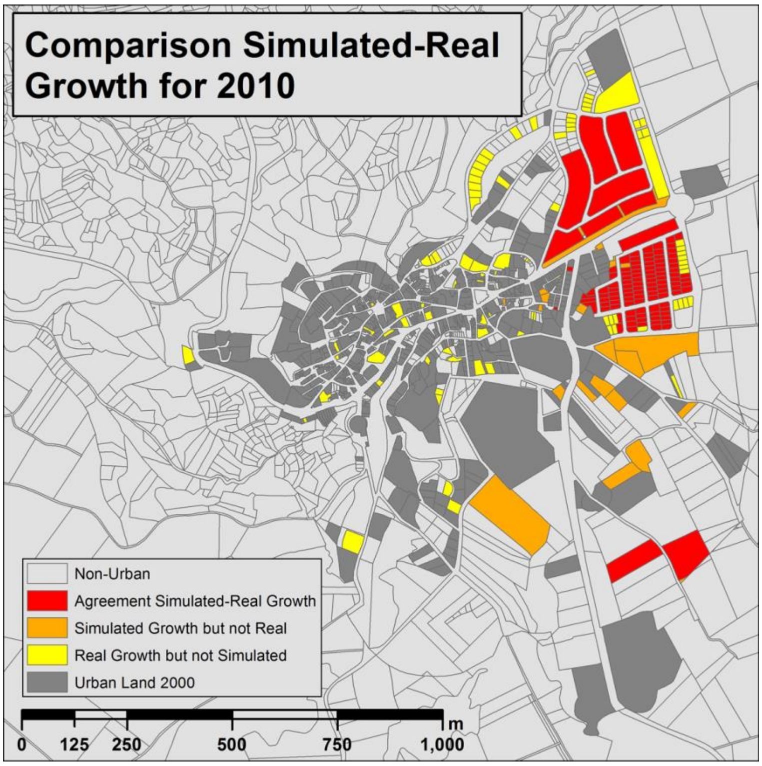


Figure 10: Comparison of the agreement between real and simulated growth in 2010 using the 200 m neighbourhood and the set of functions B.

4. Discussion

This study has explored the ability of vector spatial metrics to capture neighbourhood characteristics in an irregular urban environment and how their results can be translated into neighbourhood effect functions with CA modelling and simulation purposes. Using the proposed methodology, the average neighbourhood conditions for residential or productive parcels were quantified. The results obtained for non-urban use remained almost stable regardless of distance variations. Such results, however, were obtained using a small municipality as case study area. In order to obtain more robust results, in the next stage of this research the study area will be extended.

Another possible improvement concerns the neighbourhood effect functions, which were obtained from the metrics results taking into account: (1) the distance between the studied parcel and each of its neighbours, and (2) the land use of the neighbour. An alternative method would be to calculate the neighbourhood effect using a weighted system which considers the area of each parcel covered by the neighbourhood buffer. The inclusion of barriers such as roads, as proposed by Dahal and Chow (2015), could also be useful to simulate urban change dynamics more realistically. Different neighbourhood definitions, such as using people's perception of neighbourhood (defining parcels along the same road as neighbours rather than ones backing onto a parcel), could also be explored.

The neighbourhood effect functions were tested by running the model under four different sets of functions as well as running a BAU simulation. The results were compared against observed urban growth using PA. Although PA has been useful in demonstrating the degree of agreement with observed reality, other metrics and indices could have been used. Landscape metrics (generally known as spatial metrics) such as number of patches, shape indices or relative distance indices, could also be applied to compare results by measuring fragmentation, dispersion or relative distance between simulated parcels (see McGarigal *et al.*, 2002; Geertman *et al.,* 2007). Landscape metrics can be particularly useful when comparing different simulations with similar PA values. For example, when the amount of area simulated that coincides with observed growth is the same, thus resulting in similar PA values, change can be located in different places. Landscape metrics can be useful in those cases and, thus, should be included on the validation procedures for such models. Other metrics such as *Khisto* or *Kloc* (Hagen 2002; van Vliet *et al.*, 2009; van Vliet *et al.,* 2011) could be also adapted to an irregular environment. They would also contribute to detect differences in terms of location when simulations present differences in the amount of area developed.

Regarding the spatial structure, irregular CA-based models allow for a more realistic representation of the urban environment, where the spatial unit corresponds to the unit used in urban planning: a parcel or plot. Land use data at cadastral level allow for individual units to be represented, avoiding the use of aggregate data which is typical of socio-economic vector representations. Irregular approaches have been successfully adopted for other kind of land uses, such as agricultural, based on the understanding that 'decisions underlying these changes are typically not made for pixels, but for all parcels managed by a single actor or institution' (Zelaya et al., 2016, pages 95-96). However, models using irregular structures entail some technical and conceptual limitations.

Technically, very large study areas might be computationally demanding and time intensive, even when using graph representation. Vector representation at this level of detail also presents a considerable challenge when modelling urban expansion, since the difficulties entailed in simulating urban morphology realistically are avoided when using raster representation. Larger parcels might be subdivided into smaller ones when developing a new urban use or even when remaining stable. This issue combined with changes in roads and plot layout, are very difficult to simulate realistically. A possible avenue for future research is to explore the parcel subdivision algorithm.

One of the benefits of working at such detailed scale is that by avoiding the use of aggregate data, the modifiable areal unit problem (MAUP) is also avoided. It is important to note that the methodology proposed here to study the neighbourhood effect on land use change could also be used with aggregate data, in which case the MAUP effects would have to be considered in the results.

Conceptually, it is important to reflect on the benefits that working at this level of detail entail. The parcel level makes it possible to study urban growth and change making use of the planning smallest unit as well as analysing the neighbourhood effect more realistically than using regular cells. The use of the parcel could help to achieve better communication between modellers and planners, and consequently more effective use of models as tools to support decision-making (Pinto and Antunes, 2010).

5. Conclusions

This paper proposes a methodology to implement the neighbourhood effect in an irregular CA-based model based on the neighbourhood characteristics of parcels. Neighbourhood effect functions can be obtained from the results of spatial metrics that characterise the local neighbourhood of each parcel. The functions were introduced in a model in order to test its ability to reproduce past urban growth. A comparison of the spatial outcomes of the model with and without the implementation of these functions provides evidence that using the proposed methodology the model results improve. Nevertheless, it is necessary to expand the study area in order to obtain more robust results that support the applicability of the methodology presented here.

The results obtained suggest that this methodology could be used as part of the calibration process in other models employing irregular spatial representation, such as CA- and agent-based models. Furthermore, the methodology could also serve as a standalone method for land use change studies using land use data at cadastral (plot) level. Quantitative studies of neighbourhood dynamics are often limited by the availability of data at parcel level, but could nevertheless provide a very useful contribution to our understanding of local urban dynamics.

Acknowledgements

This research was performed within the context of SIMURBAN2 project (Geosimulation and environmental planning on metropolitan spatial decision making. Implementation to intermediate scales) (CSO2012-38158-C02-01), funded by the Spanish Ministry of Economy and Competitiveness. The Universidad de Alcalá supports the first author within the “Ayudas para la Formación de Personal Investigador (FPI) 2012” framework and the “Ayudas de Movilidad para el personal Investigador en Formación 2014” to perform the research at Birkbeck College, University of London. The authors would also like to thank the “Servicio de Traducción de la Universidad de Alcalá” which provided language help and to the anonymous reviewers for their very helpful comments.

References

Al-Ahmadi, K. et al., 2009. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecological Complexity*, 6(2), pp.80–101.

Baetens, J. M., and De Baets, B., 2012. Cellular automata on irregular tessellations. *Dynamical Systems*, *27*(4), pp. 411-430.

Ballestores, F. and Qiu, Z., 2012. An integrated parcel-based land use change model using cellular automata and decision tree. *Proceedings of the International Academy of Ecology and Environmental Sciences*, 2 (2), pp. 53-69.

Barredo, J.I. et al., 2003. Modelling dynamic spatial processes: simulation or urban future scenarios through cellular automata. *Landscape and Urban Planning*, 64(3), pp .145–160.

Barreira-González, P., González Cascón, V. and Bosque Sendra, J., 2012. Detección de errores temáticos en el CORINE Land Cover a través del estudio de cambios: Comunidad de Madrid (2000-2006). *Estudios Geográficos*, 73(272), pp.7–34.

Barreira-González, P., Gómez-Delgado, M. and Aguilera-Benavente, F., 2015. From raster to vector cellular automata models: A new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems*, 54, pp.119–131.

Batty, M. 2007. Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. The MIT press.

Benenson, I. and Torrens, P.M., 2004. Geosimulation : automata-based modelling of urban phenomena, Hoboken, NJ: John Wiley & Sons.

Congalton, R. G., & Green, K., 2008. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.

Couclelis, H., 1997. From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B: Planning and Design*, 24, pp.165–174.

Dahal, K.R. and Chow, T.E., 2014. An agent-integrated irregular automata model of urban land-use dynamics. *International Journal of Geographical Information Science,* 28(11), pp. 2281–2303.

Dahal, K.R. and Chow, T.E., 2015. Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. *International Journal of Geographical Information Science*, 29(3), pp.475-497.

Díaz-Pacheco, J. and García-Palomares, J., 2014. Urban Sprawl in the Mediterranean Urban Regions in Europe and the Crisis Effect on the Urban Land Development: Madrid as Study Case. *Urban Studies Research, 2014*.

Farsaie, M., and Hakimpour, F., 2014. Evaluation of Scale Change Effect on Simulating Urban Expansion Using Continuous Cellular Automata. *Journal of Geomatics Science and Technology*, *4*(1), pp. 67-78.

Felzenszwalb, P.F. and Huttenlocher, D.P., 2004. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2), pp.167–181.

García, A.M. et al., 2011. An analysis of the effect of the stochastic component of urban cellular automata models. *Computers Environment and Urban Systems*, 35(4), pp.289–296.

Geertman, S., Hagoort, M. and Ottens, H., 2007. Spatial-temporal specific neighbourhood rules for cellular automata land-use modelling. *International Journal of Geographical Information Science*, 21(5), pp.547–568.

General Directorate for Cadastre, Ministry of Finance and Public Administration, Government of Spain. Available from: http://www.catastro.meh.es/ [Accessed October 2013].

Gómez-Vilarino, A. and Gómez-Orea, D., 2013. *Ordenación Territorial.* Mundi-Prensa Libros.

Hagen, A., 2002. Multi-method assessment of map similarity. In *Proceedings of the 5th AGILE Conference on Geographic Information Science*. Mallorca (SPAIN): Universitat de les Illes Balears Palma, pp. 171–182.

Hagoort, M., Geertman, S. and Ottens, H., 2008. Spatial externalities, neighbourhood rules and CA land-use modelling. *The Annals of Regional Science*, 42(1), pp.39–56.

Hansen, H.S., 2012. Empirically derived neighbourhood rules for urban land-use modelling. *Environment and Planning B: Planning and Design*, 39(2), p.213-228.

Hu, S. and Li, D., 2004. Vector cellular automata based geographical entity. In *12th Int. Conf on Geoinformatics - Geospatial Information Research: Brindging the Pacific and Atlantic*. University of Gävle, Sweden, pp. 249–256.

Liao, J. et al., 2016. Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. *Environmental Modelling & Software*, 75, pp. 163-175.

Lugo, I. and Valdivia, M., 2012. Geospatial cellular automata programmed in python for social sciences. In N. Norte Pinto, J. Dourado, and A. Natálio, eds. *International Symposium on Cellular Automata Modeling for Urban and Spatial Systems* (CAMUSS). Oporto, Portugal: Department of Civil Engineering of the University of Coimbra, pp. 39–52.

McGarigal, K. et al., 2002. FRAGSTATS: spatial pattern analysis program for categorical maps.

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.

Mustafa, A. et al., 2014. Measuring the effect of stochastic perturbation component in cellular automata urban growth model. *Procedia Environmental Sciences*, 22, pp. 156-168.

Nomecalles , Nomenclátor Oficial y Callejero, 2015. Instituto de Estadística, Comunidad de Madrid. Available from: http://www.madrid.org/nomecalles/Inicio.icm [Accessed June 2015]

O'Sullivan, D., 2001a. Graph-cellular automata: a generalised discrete urban and regional model. *Environment and Planning B: Planning and Design*, 28(5), pp.687–706.

O'Sullivan, D., 2001b. Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis*, 33(1), pp.1–18.

Pan, Y. et al., 2010. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34(5), pp.400–408.

Petrov, L.O., Lavalle, C. and Kasanko, M., 2009. Urban land use scenarios for a tourist region in Europe: Applying the MOLAND model to Algarve, Portugal. *Landscape and Urban Planning*, 92(1), pp. 10-23.

Pinto, N.N. and Antunes, A.P., 2007. Cellular automata and urban studies: A literature survey. *ACE: Arquitectura, Ciudad y Entorno*, 1(3), pp. 367-398.

Pinto, N.N. and Antunes, A.P., 2010. A cellular automata model based on irregular cells: application to small urban areas. *Environment and Planning B:Planning and Design*, 37(6), pp.1095–1114.

Portugali, J., 2000. *Self-organization and the city*. Springer.

Rykiel Jr, E.J., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling*, 90(3), pp.229–244.

Santé, I. et al., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96(2), pp.108–122.

Stevens, D. and Dragicevic, S., 2007. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B:Planning and Design*, 34(4), pp.708–724.

Stevens, D., Dragicevic, S. and Rothley, K., 2007. iCity: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling and Software*, 22, pp. 761-773.

Triantakonstantis, D. and Mountrakis, G., 2012. Urban Growth Prediction: A Review of Computational Models and Human Perceptions*. Journal of Geographic Information System*, 4(6), pp.555–587.

Verburg, P.H.et al., 2004a. A method to analyse neighbourhood characteristics of land use patterns*. Computers, Environment and Urban Systems*, 28(6), pp.667–690.

Verburg, P.H., et al., 2004b. Land use change modelling: current practice and research priorities. *GeoJournal*, 61(4), pp.309–324.

van Vliet, J. et al., 2013. Measuring the neighbourhood effect to calibrate land use models. Computers, *Environment and Urban Systems*, 41, pp.55–64.

van Vliet, J., Bregt, A.K. and Hagen-Zanker, A., 2011. Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modelling*, 222(8), pp.1367–1375.

van Vliet, J., White, R. and Dragicevic, S., 2009. Modeling urban growth using a variable grid cellular automaton. *Computers Environment and Urban Systems*, 33(1), pp.35–43.

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.

Wolfram, S., 1984. Cellular Automata as Models of Complexity. *Nature*, 311(5985), pp.419–424.

Zelaya, K., van Vliet, J. and Verburg, P.H. 2016. Characterization and analysis of farm system changes in the Mar Chiquita basin, Argentina. *Applied Geography*, 68, pp. 95–103.