Vector-based Cellular Automata: exploring new methods of urban growth simulation with cadastral parcels and graph theory.

Pablo Barreira González, Montserrat Gómez-Delgado and Francisco-Aguilera Benavente

Abstract

Raster space representation in Cellular Automata (CA) based models that simulate land use change or urban growth have been widely employed. However, when simulating at local scales, it offers an oversimplification of the reality being modelled. The use of the same space representation employed by urban plans would put these models closer to them. In this respect, a new prototype of a vector-based CA model to simulate urban growth is presented in this work using cadastral parcels as the cells of the model. Graph theory is employed in order to reduce computational time costs. Neighbourhood factor and its effect can be defined more flexibly. The rest of the typical parameters in CA-modelling of urban growth can be easily calculated. The prototype is tested in a municipality located within the Region of Madrid, one of the most dynamic spaces, in terms of urban growth, in Europe over the last decades.

Pablo Barreira González
pablo.barreira@edu.uah.es

Montserrat Gómez-Delgado
montserrat.gomez@uah.es

Francisco-Aguilera Benavente
f.aguilera@uah.es
Departamento de Geología, Geografía y Medio Ambiente, Universidad de Alcalá, c/Colegios nº2, 28801 – Alcalá de Henares, España

1. Introduction.

Land use changes, especially those related to urban growth, are one of the most important factors of environmental change, since one of their main effects is loss of natural land cover [1]. The study of these phenomena and their related dynamics may help in mitigating these effects. In this context, land and urban planning play a key role as regards proposing alternatives to promote sustainable development.

In this respect, there has been a notable increase over the last two decades in the use of simulation models to study urban processes and explore future scenarios (e.g.[2,3]). There are many approaches to modelling land use change and urban systems, such as Cellular Automata (CA), Multicriteria Evaluation, Agent-based models and Neural Networks, among others. However, the use of CA-based models in particular has proliferated widely since the 1970s, and they have been used to study urban growth processes at different scales and to propose alternatives to current urban planning approaches [4].

This success stems from their simplicity, intuitiveness and capacity to model systems such as cities with specific characteristics: emergence, self-similarity, self-organisation or non-linear behaviour [5,6]. Formal or strict CA models can be defined in several different ways [7,8], but there are many similarities between these definitions: the space being modelled is represented as a grid of regular cells, each cell has a state (land use) from among a set of possible states and their evolution over time (using discrete representation of time) depends on a set of transition rules based in large part on neighbourhood configuration and the state of neighbouring cells.

As Couclelis [9,10] has indicated, the application of strict CA models entails certain limitations when simulating complex systems such as cities, suggesting that there is a need for greater flexibility in the formal CA structure in order to adapt these models and render them more realistic. This approach has been incorporated into many models to explore different alternatives, obtaining important results demonstrating that outputs derived from CA-based models are sensitive to cell size [11] and to different neighbourhood configurations [12], among other factors. These two problems are strongly linked to space representation in CA.

The question thus arises of how space should be represented in a CA-based model of urban growth. Traditionally, CA models have represented urban space as a regular grid [13]. This is not only an optimum solution in terms of computational time but it also provides an overview of how urban systems behave at global scales. However, when focusing on a local scale for the purposes of urban planning, we may question the use of this grid representation since it does not represent urban morphology properly; consequently, there is a need to adapt urban model representations to the existing representation used in urban planning for each territory.

In this study, we developed a new prototype of a CA-based model for simulating urban growth and tested it in order to determine whether it would be possible to use cadastral parcels as the cellular space. In Spain, these parcels are used as the basic unit in local urban planning development; therefore this representation would be more realistic in terms of representing urban morphology. The model was tested on a municipality in the Region of Madrid, which has been one of the most dynamic urban spaces in Europe over the last decade.

The prototype follows the NASZ model scheme framework proposed by White and Engelen [7], where the parameters employed are neighbourhood, accessibility, suitability and zoning status. In order to reduce computational time and based on the O'Sullivan's [15] approach where graphs are used to represent urban neighbourhood gentrification, the prototype abstracts the representation through a graph to compute the transition rules. The neighbourhood is redefined as a non-static parameter of the CA, where each parcel can have different neighbours in terms of number and distance, based on land use within the parcel. The prototype has been developed in its entirety in Python, combining commercial and open source libraries.

This paper is organised as follows: Section 2 discusses the representation of space in CA modelling and the use of graphs to reduce computational time. Section 3 reports implementation of the prototype, describing the parameters employed and the sequence of operations. Section 4 presents a study to test the prototype and finally, Sections 5 and 6 present the discussion and conclusions.

2. Cellular space representation

2.1. From Cells to Parcels

The raster format or regular grids employed in CA modelling have been extremely useful since they can be treated as an array (computationally speaking), where each cell in the array has the same size and topological characteristics. The facility of computers to perform calculations on this structure is totally assumed. They are also a good method of representation when studying complex systems with the research goal of understanding their global behaviour.

Nevertheless, as the scale becomes more detailed in urban planning, ever finer resolution is required. One might think that smaller cells would provide more detail; however, the problem here is that at that scale, it would be unrealistic to represent the land or the space as a regular set of abstract cells rather than using objects, irregular elements or any other kind of geographical representation that better fits reality. Thus, the question arises of how best to represent the space in relation to the purpose of the model.

In the case of CA-based models that simulate urban growth, a raster grid can be a good choice when simulating at larger scales. However, a regular grid is less attractive when more detailed scales are employed, especially when simulating at the same scale considered in urban planning. Thus, several kinds of vector approaches have been employed to develop new and more flexible CA-based models in order to address this question.

The first approaches employed Voronoi polygons for several purposes: to represent spatial interactions among objects [15], urban clusters [16] or other geographical objects [17]. Related to this form of representation, Delaunay triangulation has also been employed, derived from the centroids of land plots in the urban gentrification model proposed by O'Sullivan [14,18]. However, neither Voronoi polygons nor Delaunay triangulations solve the problem presented here since they use a real object to represent space in an abstract (non-real) way.

Other representations have been developed in order to overcome this problem. For instance, it has been suggested that census data might be a much more realistic way to represent urban space rather than triangles or random polygons [19]. Alternatively, another set of geographical objects has been employed as the cellular space [20,21], the improvement here being that it is possible to implement change in their structure. Lastly, land use parcels have also been used [22], as well as cadastral plots [23].

If we now consider how urban plans are developed, most of them use spatial representations of areas that could be developed as urban land, have the potential for construction or should be protected due to their own natural characteristics or other human interests, and in many cases, cadastral parcels are employed to draw up local urban planning schemes. It seems evident that if models are developed to support local urban planning, they should represent space in the same way as urban plans do, and therefore we adopted the use of cadastral parcels as cellular space. In the country where the prototype was tested (Spain), land is structured into cadastral parcels and these are always used in urban plan development.

2.2. From Parcel to Graph

The main problem with the use of vector data is the computational time required to read each layer and perform a spatial analysis of them. Raster datasets entail lower computational costs than other kind of formats, which is why most CA-based models have used this format. When implementing a vector-based CA model, the need to reduce computational time emerges and some approaches have achieved this.

One of the most classical approaches, the use of graph theory, resolves this problem, yielding more efficient models. A graph G can be defined as a set of entities (nodes) V whose inter-relationships are established through links or connections (edges) E (Eq. 2.2.1).

$$G = (V, E);\ \text{where } E \subseteq [V^2] \tag{2.2.1}$$

A graph basically contains a set of elements and their inter-relationships, as shown in Fig.1. This structure has been used to solve several problems and questions such as shortest path calculation, optimal paths and road network management for transport, among others. Its use has also been explored in the context of urban systems, as in the case of representing neighbourhood between buildings to model urban gentrification [14].

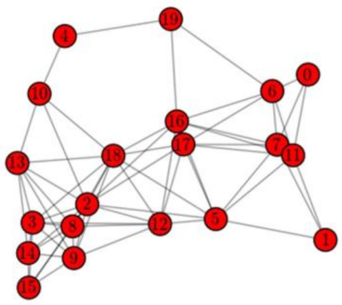

Fig. 1. Usual representation of a graph. Numbers represent the identifier of nodes and lines are the edges that connect nodes..

In the context of a vector-based CA model, a graph may be used to abstract the representation of the entities employed as cells in the model. In our case, each node in the graph represents one cadastral parcel with its complete data and characteristics (Fig. 2). However, it is not only the nodes in the graph which are stored, but also the spatial relationships between parcels (which would be the neighbourhood), using edges. An edge connects two parcels that are considered as neighbours; thus, the effect of neighbourhood can be calculated using the graph, taking all the parcels considered as neighbours into account for each parcel. The rest of the parameters implemented in the CA can be computed using this structure, and the transition rules can also be applied. Outputs derived from running the model can easily be transferred to the cadastral parcel structure in order to obtain a spatial result.

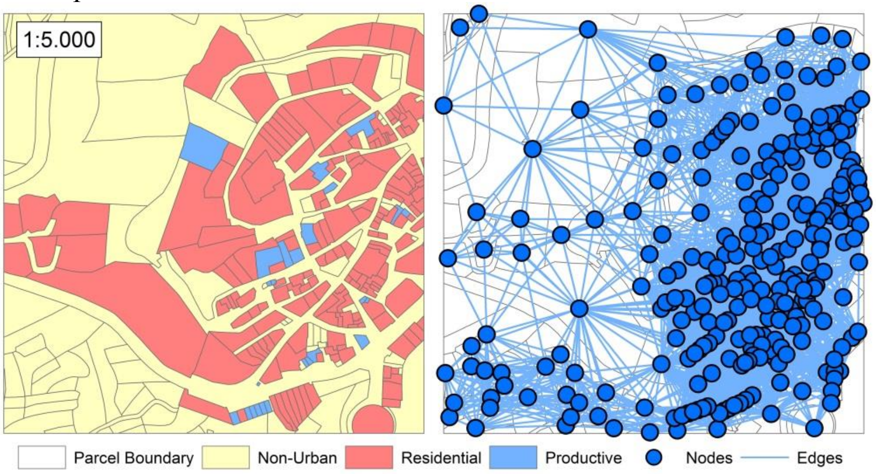

Fig. 2. Example of graph abstraction from cadastral parcels.

3. Prototype Implementation

3.1. The Parameters

The conceptual basis of the prototype presented here is the NASZ model schema [24], which uses four CA-based model parameters of urban growth: neighbourhood, accessibility, suitability and zoning status. This set of parameters has been widely applied in urban modelling, and a stochastic parameter has also been added to reproduce the uncertainty related to urban and social processes. However, this latter factor was not included in the present study since the aim was simply to test the viability of the model structure.

Neighbourhood is the most critical parameter in a vector-based CA model since every cell (parcel) is irregular and differs from the others; consequently, neighbourhood is a non-static parameter, as it used to be in raster models. It defines the influence that land use in the rest of the parcels exerts on the parcel under study, and its effect depends on land use and distances between parcels. Thus, there are two objectives that must be addressed here: identification of neighbours and computation of their effect. The effect of each parcel is calculated using distance decay functions.
Accessibility measures the degree to which a parcel has better or worse access to a road network. An accessibility value is calculated per parcel using a layer that combines national, urban and car-passable roads.

To measure the intrinsic capacity of a parcel to harbour an urban land use, a *suitability* factor is required. As with many other studies, a suitability map should include variables that are important for or driving forces of urban land development, such as terrain height, slope, type of soil, etc. Here, we employ an urban suitability map that includes a value per parcel. High values indicate that a parcel is more suitable for urban land use development, while parcels with low values are less likely to be developed.

Finally, *zoning status* relates to the legal status of parcels in urban planning terms. In the context of planning, parcels can be urban land, building land or non-building land because they are protected or because of some other kind of consideration. In the context of a CA-based model, this parameter acts as a mask, impeding new development in protected areas and allowing it in those that can be developed.

These four parameters are combined to yield a potential transition or development value per parcel as shown in Eq. 3.1.1, where a potential transition value P to develop a land use k in parcel i is calculated by combining the values at parcel level for neighbourhood N, accessibility A, suitability S and zoning status Z. The summation of N values means that the effect of every neighbouring parcel is summed.

$$P_{i,k} = \left(\sum_{j=1}^{n} N_{i,j} \right) A_i S_i Z_i \tag{3.1.1}$$

3.2. Basic Prototype Operation

The prototype was developed in Python, combining commercial libraries such as ArcPy from ArcGIS [25], and open source libraries such as Networkx for graph creation and management [26] and OGR from GDAL to deal with vector entities [27], among others.

The purpose of the prototype is to simulate urban growth divided into two land uses: residential and productive (industrial and commercial areas), and to obtain a map of future land uses at parcel level, depending on the desired number of iterations (1 iteration is equivalent to a natural year). The demand for new residential and productive land is introduced as a threshold in each iteration and is based on trends over the previous 10 years.

The prototype starts from the situation of the cadastral parcels in 2010. Several items of information are stored in the attribute table for each parcel: ID, area, land use, year of development, etc. Information about suitability, zoning status and accessibility is added in order to use these values in the graph. Neighbours of each parcel are calculated with a 100m distance buffer for each parcel, that is to say, every parcel that is partially or totally covered by the buffer is considered as a neighbour and stored in the attribute table of the vector file.

Once all the required information has been added to the cadastral parcels, the prototype abstracts the vector representation to a graph (as shown in Fig. 3), with one node per parcel. Edges are created between those parcels considered as neighbours. The effect of neighbourhood is calculated as the sum of all the effects that every neighbouring parcel exerts on the parcel under study. The effect is computed using distance decay functions from and to each land use. Once the graph is complete, the prototype is ready to start running: land with 'vacant' zoning status is obtained, transition rule thresholds are applied and developed parcels are obtained. This would be the first iteration.

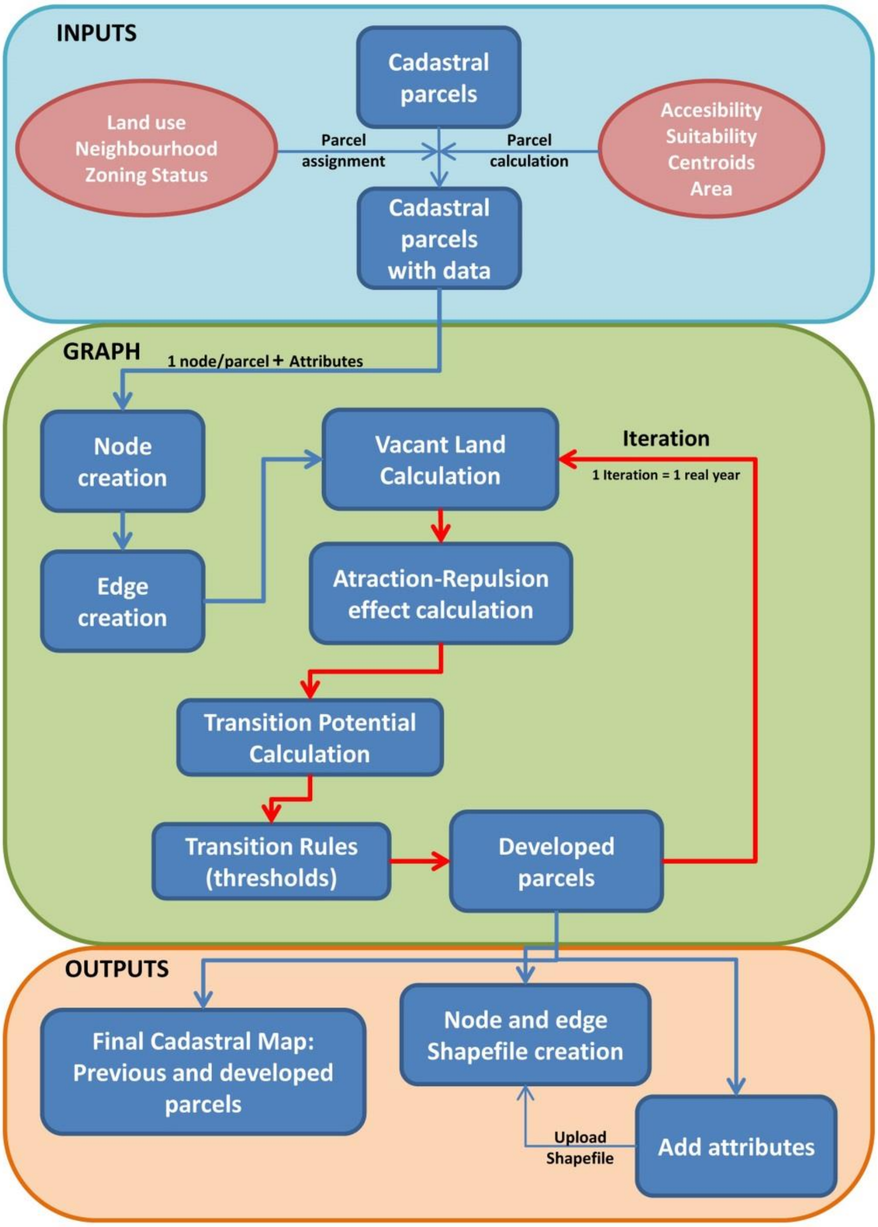

Fig. 3. Flowchart of the prototype.

For subsequent iterations, a loop is implemented so that developed parcels derived from the previous iteration affect ensuing iterations. Vacant land is recalculated according to the parcel's new status. The effect of neighbourhood is recalculated and transition rules are applied. Once the last iteration has finished, the results are transferred to the cadastral map and a node and edge map is created containing all the information obtained from the prototype.

4. A Case Study

The prototype was tested on one of the most urbanised areas in Europe, the Region of Madrid. With more than 6 million inhabitants, this region has experienced substantial urban growth over the last two decades, primarily due to the housing bubble that burst in 2008 [28]. One of the areas of most interest within the Region of Madrid is the Corredor del Henares, where many industrial areas have been developed along the national A-2 road connecting Madrid with Barcelona via Guadalajara. Cities located nearby have also contributed to this massive urban growth. Consequently, we selected Los Santos de la Humosa (Fig. 4), a small municipality within the Corridor which has not been developed to the same extent as the rest of the municipalities, and there is still a chance of preventing unsuitable development and proposing a more sustainable form of urban planning.

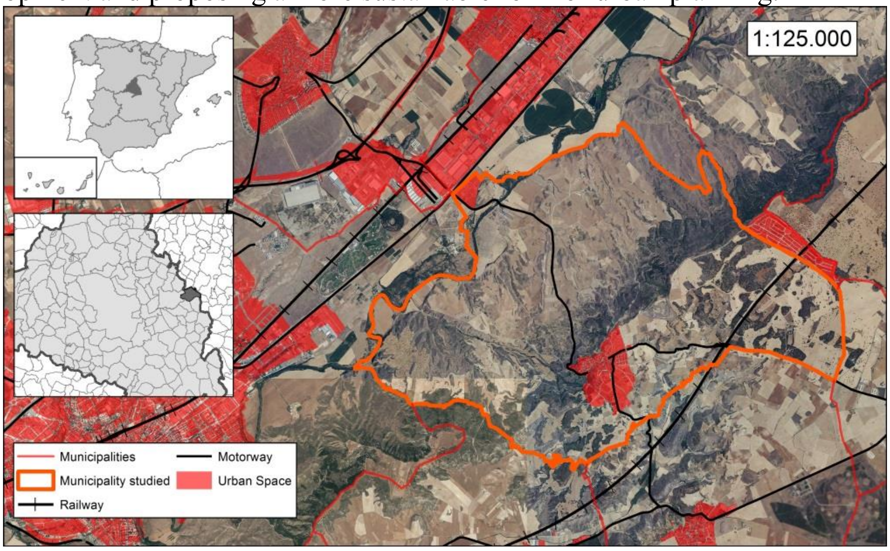

Fig. 4. Location of the municipality of Los Santos de la Humosa.

Starting from the status of the cadastral parcels in 2010, the prototype was run from that year to 2020 (10 iterations) in order to test the viability of the prototype, introducing a demand based on the trend observed between 2000 and 2010. The highest computational cost occurred when the edges were generated, since in this case, there were more than 3,000 parcels and over 200,000 edges. Once the edges had been created, the transition rules, neighbourhood effect and the rest of the operations were computed rapidly. The results derived from the graph (prototype outputs) are shown in Fig. 5.

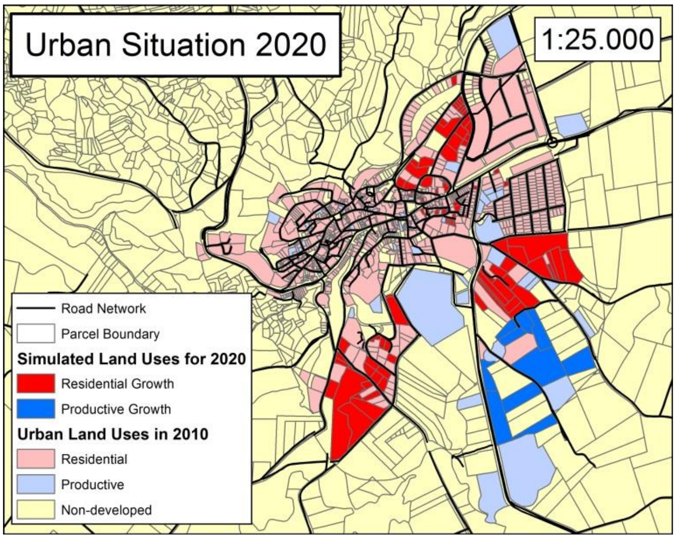

Fig. 5. Prototype's simulation from 2010 to 2020 (10 iterations).

Most of the new developed parcels were located around the urban areas already in existence in 2010. Due to the neighbourhood effect, the pre-existing productive land generated new productive land parcels close by in the south of the studied area. Zoning status prevented the location of new urban parcels in rural areas, preserving their natural characteristics.

5. Discussion and Limitations

The success of the approach presented here resides in the fact that relaxation of space representation within the formal structure of classical CA-based models has been computationally solved. In this case, a vector representation of space is proposed that uses cadastral parcels. Graphs are a good way to represent neighbourhood and to abstract vector representation to a more operative one, and thus represent an easy way to reduce computational time. The graph structure facilitates faster computation by the prototype and provides a stable topology. Graph status can be explored at any time during the prototype run, and all the attributes (from nodes and edges) can be modified if necessary.

Neighbourhood is defined within the prototype as a function of distance from each parcel. Accessibility is now computed as the shortest distance to a road network, but not every road is the same as the others. More information about the type of road, the traffic or commuting could be added in order to obtain a more realistic parameter.

In the current version, application of the transition rules is based on thresholds. In each iteration, vacant parcels store a value for transition potential to residential land use and another value for transition potential to productive land use. In this respect, if two parcels present the same potential value to develop both land uses, there should be a mechanism that decides. Parcels are developed completely after each iteration, and thus it would be interesting to establish a percentage of development of the parcels with the highest potential for development in order to prevent the massive growth of the largest parcel in the cadastre. It seems obvious that calibration of parameters and transition rules is required in order to obtain a robust model where trends can be reproduced and simulation under several scenarios could subsequently be performed.

Lastly, randomness has not yet been introduced as a prototype parameter, although it has been used in other urban CA-based models do [29]. However, this parameter could be included in the current prototype in several ways. For instance, randomness could be introduced when calculating the transition potential, modifying the value calculated for the rest of the factors. It could also be introduced when applying zoning status, since as has happened in Spain, illegal buildings and proscribed land uses have been developed in some parcels even though they were not programmed in the local urban plan.

The approach presented here has an evident limitation in that no subdivision is introduced in the cadastral structure; thus it is not yet possible to subdivide larger parcels. It would be of great interest to find ways to subdivide parcels in relation to the amount of land demanded.

6. Conclusion

The purpose of the present study was to test the viability of using vector representation as the cellular space of a CA-based model to simulate urban growth. In this case, cadastral parcels were employed as the smallest unit of space since they are also employed in urban planning; thus the model could be run at local planning scales. A graph was introduced to abstract the vector space and to compute the transition rules and all the required data. The prototype runs within a reasonable time and the outputs are logical, so we can affirm that the goal has been reached. More research is required to calibrate and validate the prototype in order to obtain a complete vector-based CA model. It would also be useful to compute a sensitivity analysis to determine how current implementation of the parameters entered in the prototype influences the outputs. Main and total effects of the parameters could easily be computed to learn more about the individual and combined contribution of the factors.

Acknowledgements

This research was conducted within the context of the SIMURBAN2 project (Geosimulation and environmental planning in metropolitan spatial decision making. Implementation at 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.

Bibliography

1. Lauf, S., Haase, D., Hostert, P., Lakes, T., Kleinschmit, B., 2012. Uncovering land-use dynamics driven by human decision-making–A combined model approach using cellular automata and system dynamics. *Environmental Modelling & Software* 27 71-82.

2. Sante, I., Garcia, A.M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning* 96(2) 108-122.

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

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

5. Barredo, J.I., Kasanko, M., McCormick, N., Lavalle, C., 2003. Modelling dynamic spatial processes: simulation or urban future scenarios through cellular automata. *Landscape and Urban Planning* 64(3) 15 (145-160).

6. Frankhauser, P., 1998. Fractal geometry of urban patterns and their morphogenesis. *Discrete dynamics in Nature and Society* 2(2) 127-145.

7. White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems* 24(5) 383-400.

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

9. Couclelis, H., 1985. Cellular worlds: a framework for modeling micro-macro dynamics. *Environment and Planning A* 17(5) 585-596.

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

11. Ménard, A., Marceau, D.J., 2005. Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B*: 32 21.

12. Kocabas, V., Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems* 30 32.

13. White, R., Engelen, G., 1993. Cellular-Automata and Fractal Urban Form - a Cellular Modeling Approach to the Evolution of Urban Land-Use Patterns. *Environment and Planning A* 25(8) 1175-1199.

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

15. Shi, W.Z., Pang, M.Y.C., 2000. Development of Voronoi-based cellular automata - an integrated dynamic model for Geographical Information Systems. *International Journal of Geographical Information Science* 14(5) 455-474.

16. Semboloni, F., 2000. The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment and Planning B* 27(4) 549-564.

17. Shiyuan, H., Deren, L., 2004. Vector cellular automata based geographical entity, *12th Int. Conf on Geoinformatics - Geospatial Information Research: Brindging the Pacific and Atlantic*: University of Gävle, Sweden, pp. 249-256.

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

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

20. Moreno, N., Ménard, A., Marceau, D.J., 2008. VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B* 35(4) 647-665.

21. Moreno, N., Wang, F., Marceau, D.J., 2009. Implementation of a dynamic neighborhood in a land-use vector-based cellular automata. *Computers Environment and Urban Systems* 33(1) 44-54.

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

23. Stevens, D., Dragicevic, S., Rothley, K., 2007. iCity: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling & Software* 22 12.

24. White, R., Engelen, G., Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B* (3) 323-343.

25. ArcGIS, E., 2012. 10.1. Redlands, California: ESRI.

26. Hagberg, A., Swart, P., S Chult, D., 2008. Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Laboratory (LANL).

27. Fundation, O.S.G., 2008. GDAL-OGR: Geospatial Data Abstraction Library/Simple Features Library Software.

28. Plata-Rocha, W., Gomez-Delgado, M., Bosque-Sendra, J., 2011. Simulating urban growth scenarios using GIS and multicriteria analysis techniques: a case study of the Madrid region, Spain. *Environment and Planning B* 38(6) 1012-1031.

29. Garcia, A.M., Sante, I., Crecente, R., Miranda, D., 2011. An analysis of the effect of the stochastic component of urban cellular automata models. *Computers Environment and Urban Systems* 35(4) 289-296.