NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ

Bulletin of Geography. Socio-economic Series, No. 55 (2022): 107-121
http://doi.org/10.12775/bgss-2022-0008

BULLETIN OF GEOGRAPHY. SOCIO–ECONOMIC SERIES

journal homepages:
https://apcz.umk.pl/BGSS/index
https://www.bulletinofgeography.umk.pl/

ISSN 1732–4254 quarterly

Urban growth models and calibration methods: a case study of Athens, Greece

Tsagkis Pavlos

National Technical University of Athens (NTUA), Faculty of Rural and Surveying Engineering, 9 Iroon Polytexneiou street, 15780 Athens, Greece; phone: +30 210 112 606, e-mail: p.tsagkis@gmail.com

How to cite:
Pavlos, T. (2022). Urban growth models and calibration methods: a case study of Athens, Greece. *Bulletin of Geography. Socio-economic Series*, 55(55): 107-121. DOI: http://doi.org/10.12775/bgss-2022-0008

Abstract. A number of urban growth models have been developed to simulate and predict urban expansion. Most of these models have common objectives; however, they differ in terms of calibration and execution methodologies. GIS spatial computations and data processing capabilities have given us the ability to draw more effective simulation results for increasingly complex scenarios. In this paper, we apply and evaluate a methodology to create a hybrid cellular-automaton-(CA) and agent-based model (ABM) using raster and vector data from the Urban Atlas project as well as other open data sources. We also present and evaluate three different methods to calibrate and evaluate the model. The model has been applied and evaluated by a case study on the city of Athens, Greece. However, it has been designed and developed with the aim of being applicable to any city available in the Urban Atlas project.

Article details:
Received: 12 June 2021
Revised: 29 September 2021
Accepted: 3 February 2022

Key words:
Urban growth,
agent-based model,
cellular automata,
land-use simulation

Contents:

1. Introduction

Rapid urban growth leads to the transition of green and open land to urban land. As a result, the microclimate of the area changes, which in turn affects food security, public health and water quality and finally leads to global environmental change (Musa et al., 2017). Consequently, the environmental impact of urban growth must be studied by urban scientists and town planners because it degrades the ecosystem and affects human lives.

Over the last two decades, the scientific community has paid great attention to this phenomenon. In turn, various models and techniques have been developed to cope with urban growth or urban sprawl and predict its behaviour over time and space. These models can be defined as spatio-temporal models.

Moreover, various application platforms have been developed to cope with such spatio-temporal models. Some focus on urban growth, providing a clear methodology for the dynamics of the model, but lack the ability to adjust the model dynamics or to enrich the model with customised agents or local factors. Others provide a generic interface to implement such models but lack the urban growth orientation and adjustment, making them difficult to use for users without prior programming experience.

In this article, we use GAMA Platform in conjunction with data from the Urban Atlas project, as well as other open-source data, in order to simulate urban growth for the greater Metropolitan Area of Athens, Greece. We test and calibrate a simple model, taking into account global and local aspects but excluding extensive local knowledge, so that it can be adjustable to any city available in the Urban Atlas project. In addition, three different calibration methods were combined to calibrate our model, and finally, to perform a ten-cycle simulation up to 2072.

The article is organised as follows. The literature concepts are discussed in Section 2, while study area and related data are presented in Section 3. Model description, dynamics and calibration process are discussed in Section 4, whereas results are presented in Section 5. Finally, in Section 6 and 7 we discuss the limitations and strengths of this study along with some thoughts about the results.

2. Urban growth models

Urban growth models are spatial computer simulations that use economics, sociology, geography and statistics in order to predict land-use changes and thus urban spatial evolution. They combine theories, data and mathematical algorithms to predict spatial changes within the urban environment (Triantakonstantis & Mountrakis, 2012) (Sante et al., 2010).

Over the last two decades, several different models have been developed based on theoretical assumptions scientists have made upon studying the urban growth literature. However, most of these models lack the ability to adapt to new theories or to be enriched with different aspects of related theory. As a consequence, several different techniques have been developed and applied to address this disadvantage. According to (Li & Gong, 2016), modern geographic urban growth models may be classified into two main categories:

2.1. Cellular Automata (CA)

A cellular automaton is a computational model of systems with emergent complexity. Cellular automata are studied in computational theory, physics and theoretical biology, as well as in many other scientific fields. They were devised in the 1940s by mathematician John von Neumann in order to describe the functions of the biological cell (von Neumann & Burks W, 1966). However, they became better known during the 1970s with Conway's Game of Life (Conway, 1970), a two-dimensional cellular automaton that captured the interest of the academic community. Cellular automata gained much popularity in the 1980s thanks to US computer scientist Christopher Langton (Langton, 1997), a founder of the cognitive field of artificial life.

Cellular automaton models focus on the dynamics of the core elements of the urban system, as well as the interactions among them. A society of elements moves from cell to cell in predetermined

neighbourhoods while neighbours defined by transition rules determine the distance traction or repulsion these dynamics may develop (Liu, Zheng, & Wang, 2020). In other words, each cell represents a land-use value; its future land-use value depends on the neighbouring cells' values and the transition rule values related to the cell.

2.2. Agent-Based Models (ABM)

Agent-based modelling is a computational method that gives researchers the ability to create, analyse and experiment with models composed of agents that interact within space. ABM is the newest method of model development and has gained great popularity during the last decade (Gilbert, 2008) (Matthews et al., 2007). ABMs have several implementations related to social geography, ecology, biology and urban studies. They are also known as individual-based models (IBMs) or multi-agent systems (MAS), and individuals within IBMs may be simpler than fully autonomous agents within ABMs.

Agents are autonomous decision-making entities that can make decisions and execute various behaviours within the environment in which they live and interact. For that reason, ABMs are an excellent fit for the urban growth phenomenon because their spatial distribution is highly dependent on the decisions of humans living within the study area (Martinez & Viegas, 2017), (Bonabeau, 2002). Furthermore, ABMs can represent multiple spatial relationships among their agents, which in turn make them an appropriate fit for the complex city environment.

However, ABMs have certain drawbacks as well. In most cases, behaviour rules are subjective and are based on researchers' assumptions, and calibration appears to be a very complicated procedure (Baptista, et al., 2016), requiring a large amount of computing resources.

In this paper we present a hybrid cellular-automata-and-agent-based model. We also provide a methodology to calibrate the model from historical data for the study area, using Gama Platform as the model development toolkit.

3. Study area: data description and manipulation

3.1. Athens Metropolitan Area

The proposed model is applied, tested and calibrated in the Greater Athens Metropolitan Area. Athens has been an urban centre from its birth in antiquity until the present, occupying a space of approximately 300 km^2 and accommodating more than three million residents. It is located in the centre of Greece, in the Mediterranean region and has been a cultural crossroads of Asia, Europe and Africa for more than 2000 years.

During the last three decades, the city has expanded considerably, creating new peri-urban areas and sub-centres. However, the expansion that took place during the period 1950 to 1970 due to extensive rural immigration was unplanned and unregulated with no accompanying infrastructure projects, resulting in traffic congestion and environmental pollution (Beriatos, 2006) as well as high concentrations of industrial and production-related activities (Grekousis et al., 2013). Furthermore, the metropolitan area of Athens contains three mountains and several hills, which present a wide range of slopes and altitudes.

In terms of economic activity, Athens experienced dramatic growth between 2000 and 2004, the year Athens hosted the Olympic Games; however, the city has suffered from an economic crisis from 2010 until the present. As a result of the economic crisis, construction activity dramatically decreased, resulting in a low urban growth rate during the study period (2006–2012). For all these reasons, Athens' urbanisation doesn't follow a generic pattern, making the city very complex in terms of urban growth prediction and also a challenging case study.

3.2. Land-use map: Urban Atlas dataset

This study utilises land cover maps for the years 2006 and 2012 acquired from the Urban Atlas dataset. The Urban Atlas project is a pan-European dataset produced by analysing thousands of different satellite images. It provides land-use and land-cover

Table 1. Land-use data

Class value	Land use	Colour
1	Urban	Grey
2	Not urban / Free to urbanise	Yellow
3	Forbidden to urbanise	Red
4	Forest areas	Green

Source: own elaboration

data for large urban cities with more than 100,000 inhabitants as defined by the Urban Audit (Buttner et al., 2014). These are the 697 most populated cities in the EU. The GIS data can be downloaded together with a map for each urban area covered along with a report containing the metadata (Montero et al., 2014).

Urban Atlas data are provided free of charge by the European Environment Agency in a shapefile format. The projection is ETRS 1989 Lambert Azimuthal Equal Area L52 M10 and the scale is 1:10,000; the minimum mapping unit is 0.25 ha for the artificial surfaces and 1 ha for any other surface (Herold & Gamba, 2009). The project consists of two time-series datasets – one for 2006 and one for 2012. However, a new dataset for 2018 is expected, but at the time of writing (2021) this was not yet available.

To adjust the Urban Atlas dataset to our model, we have applied a reclassification in order to produce four different classes (Table1) out of the 27 existing classes: To support our model, two raster files have been created for the referenced years of 2006 and 2012, respectively. Vector polygons have been converted to raster grids of 1000 m × 1000 m resolution. Four different possible values have been defined on each cell. The produced raster files will be used to study the urban growth evolution of the area and thus calibrate our model to fit the past evolution. The raster file for 2012 will also be the starting point of our model run.

3.3. Other data used for the model

Several different factors may influence an urban growth model; thus, several different datasets must be collected in order to express those factors. We have chosen those that can be spatially expressed and that are easily found and freely downloadable online. These include:

- Shapefile of roads in 2016 (vector lines), divided into five different classes (motorways, primary, secondary, tertiary and residential roads). Data downloaded from the OpenStreetMap (OSM) project (OpenStreetMap, 2019). Each road class will have a different weight during the implementation of our model.
- Shapefile of Greater Athens census data collected from Hellenic Statistical Authority, expressing population changes between 2001 and 2011. For each municipality, a new field is added and populated with integer values representing the "population-change class" as described in Table 2:
- Raster file of 1000 m × 1000 m cell projects a slope map for the study area. The file was produced from a digital elevation model (DEM file). Slope classes have been defined as described in Table 3.
- Shapefile of Athens city centre (vector point). A single point represents the city centre of Athens.
- Shapefile of Greater Athens sub-centres (vector points). Multiple points represent the sub-centres of the Greater Athens Metropolitan Area. Data downloaded from http://geodata.gov.gr/ (community, n.d.)

4. Model

4.1. Model description

Our model was inspired by the models proposed by (Taillandier, et al., 2016) (Chaudhuri & Clarke, 2013) (Raimbault et al., 2016). The model proposed by (Taillandier, et al., 2016) is also included within the default demo models of Gama Platform, labelled

Table 2. Population-change classes

Class value	Description (pop. Changes in 2001–11)
1	Population percentage change<-20%
2	-20%<Population percentage change<-10%
3	-10%<Population percentage change<0%
4	0%<Population percentage change<10%
5	10%<Population percentage change<20 %
6	Population percentage change>20%

Source: own elaboration

Table 3. Slope classification

Slope class	Description: slope as percentage (%)
1	0–10%
2	10–20%
3	20–30%
4	30–40%
5	40–50%
6	50–60%
7	60–70%
8	70–80%
9	80–90%
10	90–100%
11	>100%

Source: own elaboration

"raster model". This model proposes three dynamics: urban density, road network and travel distance to city centre. We have also added the distance to sub-centres (as the study area is a large area including various sub-centres), the earth slope and the population-change trends as extra agents.

Additional agents influencing urban growth models, such as unemployment, economic growth trends and housing rent, which have previously been used for such models (Hu & Lo, 2007), could not be used in our case study. Finding such data was a challenge in our case and would have led us outside of our main scope, which was to keep the model easily adaptable to any city available on the Urban Atlas project. On completion of the calibration process, the model will run for ten cycles, resulting in the urban growth prediction for the year 2072.

For the implied model, the main agents are cells that represent a homogeneous area of space. A cell agent has two attributes:

- Building options: These may have five different values: 1 (built), 2 (not built – free for urbanisation), 3 (not built and forbidden to be built, 4 (currently forbidden to be built, but with a few changes can be built [e.g. forest]), 5 (newly built [acquired during model stimulation]).
- Constructability: The level of constructability of the cell is a float value between 0 and 1. Zero means that it is not at all probable to become urban and 1 that it is very probable to become urban. The constructability of each cell is estimated during every cycle of our model and its calculation is explained in detail in the following section.

Over the last five decades, the study area has suffered severe deforestation, and affected areas have since been transformed into urban areas. For this reason, we exclude forest areas from the non-urbanised areas and include them in our scenario as being available for urbanisation. However, the rate at which forest land is turned into urban land will be determined during the simulation process of our model.

4.2. Model dynamics

The general process of the model is described in the steps below:

A. Each cell computes its level of constructability using the flow below:

Values calculated only once, on start-up, while defining our model

- For each pixel, we compute the distance to the closest road.
- For each pixel, we compute the travel distance to the city centre.
- For each pixel, we compute the travel distance to the closest sub-centre, if this is not the main city centre.
- For each pixel, we assign a value for slope.
- For each pixel, we assign a value for population-change class.

Values calculated at every step of our model as it changes over time and after each cycle

- For each pixel, we compute the urban density. The density is determined by the number of neighbouring "urban pixels". It should be a number between 0 and 8.

B. The n number of cells with the highest level of constructability becomes urbanised at each cycle. The n number of cells that should be built is estimated using the comparison between the two images, making a simple assumption that the same percentage of land that has become urban during the past period will become urban during future periods.

TR1. Neighbourhood urban density (the higher, the better).

$$TR1 = \frac{number\ of\ built\ cells\ surrounding\ study\ cell}{total\ cells\ surrounding\ study\ cell}$$

TR2. Euclidean distance to the closest road (the closer, the better). We also multiply the distance using the class of road in order to give fewer possibilities for lower-class roads and more possibilities for higher-class roads such as highways and motorways:

$$TR2 = 1 - \left(\frac{(distance\ to\ closest\ road) \times (road\ class\ (number\ between\ 1\ and\ 5))}{max.dist.to\ closest\ road\ (among\ all\ cells) \times 5}\right)$$

TR3. Travel distance to city centre using road network (the closer, the better):

$$TR3 = 1 - \left(\frac{distance\ to\ city\ centre}{max.dist.to\ city\ centre\ (among\ all\ cells)}\right)$$

TR4. Travel distance to closest sub-centre (the closer, the better):

$$TR4 = 1 - \left(\frac{distance\ to\ subcentre\ (metres)}{max.dist.to\ subcentre\ (among\ all\ cells)}\right)$$

TR5. Earth-slope class the cell belongs to (the lower, the better):

$$TR5 = 1 - \left(\frac{slope\ class\ value\ (integer\ between\ 1\ and\ 11)}{11\ (max.slope\ class)}\right)$$

If slope class is more than 60%, meaning a value of >7, we consider the cell as not available for urbanisation, but this value is adjustable during model simulation.

TR6. Population-change class this cell belongs to (the higher, the better):

$$TR6 = \left(\frac{population\ change\ class\ \ (integer\ between\ 1\ and\ 6)}{6\ (max.pop.change\ class)}\right)$$

The level of constructability is the weighted average of these six indices. To this end, for every index, we must assign a weight expressing the importance of the index to the final model (Tsagkis & Photis, 2018). As a final step, and in order to take into consideration cells classified as forest land, we use a seventh weight: the forest weight. If we want to exclude forest land from urbanising, we set it to 0, whereas if we want forest cells to behave like any other cell available for urbanisation, we set it to 1.

The seven weights (w1, w2, w3, w4, w5, w6 and w7) are parameters of the model, while the

$$Constructability = w7 \times \frac{(w1 \times TR1) + (w2 \times TR2) + (w3 \times TR3) + (w4 \times TR4) + (w5 \times TR5) + (w6 \times TR6)}{w1 + w2 + w3 + w4 + w5 + w6}$$

constructability value is the weighted average of the six or seven criteria mentioned depending on the type of pixel calculating. The constructability value is expressed using the equation below, turning out a float number between 0 and 1:

C. Model simulation step is of six years' because data from Urban Atlas dataset are published every six years. Thus, each step in our simulation represents a six-year cycle of real-world time.

4.3. Model calibration

One of the most important challenges when developing simulation models is to clarify the validity of the outputs, commonly referred to as "model calibration" (Pijanowski, et al., 2014; Olmedo et al., 2015). The general idea is to study the past in order to predict the future (Clarke et al., 1996) (Hagen-Zanker & Martens, 2008).

Comparing the "predicted" present to the "known" present, using a logistic regression method should give the best match and thus the best weight combination to use for the agents influencing the model. To achieve that, we have produced several different images expressing the "predicted" present. To produce those images, we have used the batch mode of Gama Platform, assigning a range of values for every weight and running the model from the starting year of 2006 for one cycle until the year 2012. This results in a large number of images that will eventually be compared, one by one, with the real-life image of 2012 and, according to the similarities or differences, we will determine the optimal combination of weight values. The seventh weight of forest land probability is outside the calibration scope, so modellers may use a float value between 0 and 1. For the model proposed, we have used the following range of values, resulting 87,846 different prediction images:

w1: [0.01, 0.02, 0.03, 0.04, 0.05, 0.06];
w2: min: 0.0 max: 1.0 step: 0.1
w3: min: 0.0 max: 1.0 step: 0.1
w4: min: 0.0 max: 1.0 step: 0.1
w5: min: 0.0 max: 1.0 step: 0.1
w6: min: 0.0 max: 1.0 step: 0.1
w7: 0.8

The image map comparison is using the grid-based approach, comparing the real image with the predicted image, pixel by pixel. Various methods and algorithms have been developed to quantify the agreement between two raster maps. These include kappa statistic, fuzzy kappa (Kuhnert et al., 2005) (Pontius Jr & Cheuk, 2006) (Pontius Jr R.G., 2002) (Visser & De Nijs, 2006) (van Vliet et al., 2011), Relative Operating Characteristics (ROC), and Total Operating Characteristics (TOC) (Chakraborti et al., 2018) (Shafizadeh-Moghadam, 2019). One of the most well-known approaches for raster map comparison is Cohen's kappa index (Cohen, 1960) (Foody, 2007), which expresses the agreement between two categorical datasets corrected for the expected agreement (Fleiss et al., 1969).

In general, every pixel of the reality map is compared to the pixel in the same position on the prediction map. If their values match, we add 1; if not, we add 0. The sum is divided by the total number of pixels that are expected to get changed, which in turn yields a float number between 0 and 1:

$$RealAgreement\ (ReA) = \frac{number\ of\ matched\ pixels}{total\ number\ of\ changing\ pixels}$$

Using the same technique, we use a random raster map to compare against the reality map:

$$RandomAgreement\ (RaA) = \frac{number\ of\ matched\ pixels}{total\ number\ of\ changing\ pixels}$$

Using the two values obtained, we can finally calculate the kappa statistic using the equation:

$$Kappa = \frac{ReA - RaA}{1 - RaA}$$

Kappa index is a float number ranging from -1 to 1. If the value of kappa is close to 1, then the agreement between the maps is very high; if it is close to -1, the maps are totally different (Eugenio & Glass, 2004).

A similar approach that has gained popularity among the scientific community is the fuzzy kappa (Hagen-Zanker, 2009) (van Vliet et al., 2013). Predicting an urban pixel at the exact location is not always possible, due to the complex nature of urban growth processes. The fuzzy kappa approach goes beyond the cell-by-cell comparison and takes into account the neighbouring cells, giving credits to cells that are located nearby but not in the exact

location of the cell observed. In our case, if an urbanised pixel is not found in the exact location then we search within the surrounding eight pixels, and we use the sum of those urbanised divided by the length*2 of urbanised pixels:

$$Fuzzy\ Pixel\ Credit = \frac{\mathbf{SUP\ (number\ of\ Surrounding\ Urbanised\ Pixels)}}{\mathbf{SUP} * \mathbf{2}}$$

Consequently, we use this value to compute real and random agreement and finally draw the kappa fuzzy index (Mustafa et al., 2018).

Real map agreement:

$$Fuzzy\ RealAgreement\ (FReA) = \frac{number\ of\ matched\ pixels + surrounding\ matched}{total\ number\ of\ changing\ pixels}$$

Random map agreement:

$$Fuzzy\ Random\ Agreement\ (FRaA) = \frac{number\ of\ matched\ pixels + surrounding\ matched}{total\ number\ of\ changing\ pixels}$$

Fuzzy kappa statistic:

$$KappaF = \frac{FReA - FRaA}{1 - RaA}$$

One more method that is easy to apply is the distance index approach proposed by (Kuhnert, Voinov, & Seppelt, 2005). The distance index tells us how far apart in space the disagreeing cells are, using the mean Euclidean distance among all the compared cells. Thus, for each raster comparison we have a mean distance for all the changing cells, and the raster images with the lowest mean distance are considered to draw the best simulation results (Meratnia & de By, 2002). To normalise these values and get a more representative value, we use the equation proposed by (Kuhnert, Voinov, & Seppelt, 2005): One more method that is easy to apply is the distance index approach proposed by (Kuhnert et al., 2005). The distance index tells us how far apart in space the disagreeing cells are, using the mean Euclidean distance among all the compared cells. Thus, for each raster comparison we have a mean distance for all the changing cells, and the raster images with the lowest mean distance are considered to draw the best simulation results (Meratnia & de By, 2002). To normalise these values and get a more representative value, we use the equation proposed by (Kuhnert et al., 2005):

$$F = \frac{Nc\ (number\ of\ comparing\ cells\ )}{Nc + \sum_{Nc} Ed\ (Euclidean\ distance)}$$

The resulting value is a float number between 0 and 1. The highest value is considered to express the simulated image that best fit the image observed in reality.

Fuzzy kappa appears to be the most sophisticated and accurate method for map comparison, as it uses all the power of the kappa statistic but also takes into account small spatial errors. However, within the scientific community there seems to be no universally accepted method to use as a standard, and thus our previous assumption is based on the authors' experience and research study.

In our case study we have implemented all three methods mentioned and explained in this paper (kappa, fuzzy kappa and distance index), resulting in three indices, as well as a final index using the mean value of the three outputted indices. To achieve that, we compared 87,846 prediction images with the real-changes image. For each single image, four different index float numbers were created to express the level of agreement among the changing cells. The aim of the comparison was to identify the best weight for each transition rule.

5. Results

To have a fair result for our model evaluation, we use only the changing pixels between the years of 2006 and 2012 and compute all indices using only these pixels (Mustafa et al., 2017). According to the results, kappa statistic, fuzzy kappa and overall summary selected the same weights for the best fitted image. However, fuzzy kappa picked a single image for the highest score, whereas kappa picked 12 different images having equally high scores. One of those 12 highest kappa index images had the same weights as the one picked by the fuzzy kappa metric. Furthermore, the distance index showed some differences among weights but followed almost the same pattern. Therefore, we use this image as the best weights combination because three out of four metrics picked the same weights.

Within Table 4: Best weights matrix, we summarize the output results of the comparison for the best index values obtained compared to the

Table 4. Best weights matrix

Index name	Fuzzy kappa	Kappa	Distance index	Overall
Index value (highest score)	0.24	0.18	0.17	0.17
Urban density weight (w1)	0.06	0.06	0.04	0.06
Road distance weight (w2)	0.8	0.8	0.7	0.8
Main city centre distance weight (w3)	0.0	0.0	0.0	0.0
Sub-centres distance weight (w4)	0.5	0.5	0.6	0.5
Slope weight (w5)	0.7	0.7	0.3	0.7
Population trends weight (w6)	0.0	0.0	0.0	0.0

Source: own elaboration

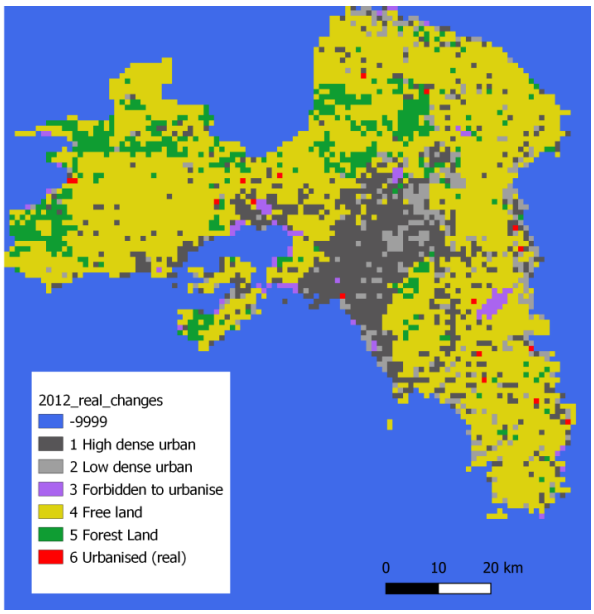

Fig. 1. Real image changes (2006–2012)
Source: own elaboration

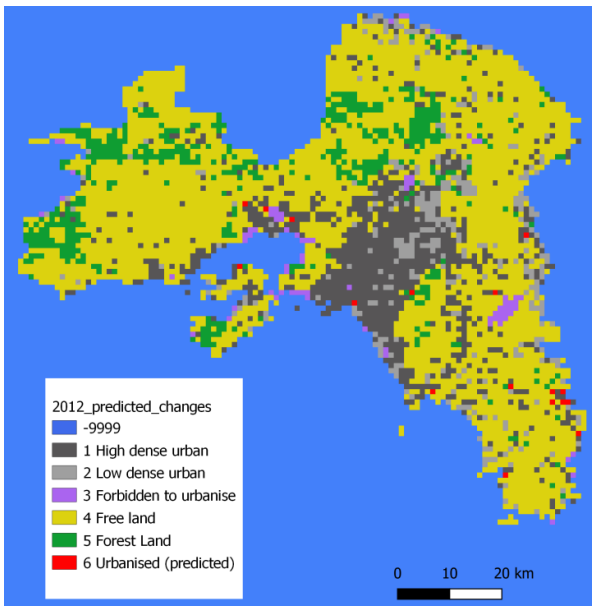

Fig. 2. Best score predicted image changes (2006–2012)
Source: own elaboration

weights chosen for the predicted image and among the newly created urban cells. Road network (w2) and earth slope (w5), having weights of 0.8 and 0.7 respectively, had the greatest influence on the model, while urban density (w1), with a value of 0.06, had a small but real influence on the model. On the other hand, main city centre (w3) and population growth (w6) both had no impact on the model (Table 4).

In the case of the main city centre, this seems reasonable because most of the surrounding main city area is already urbanised; thus, there is no remaining space available to urbanise. Furthermore, the study area is a large area consisting of various local urban centres, which are used as a different transition rule in our model. This is the sub-centres distance weight (w4) which in turn, with a value of 0.5, had a representative influence on the model. The low value of the population trend weight may be explained by the inconsistency between census data and the rest of the datasets. Census data are classified using the spatial entity of municipality (vector polygons), which is a large area that can hardly be projected in a cell of 1000 m × 1000 m; unfortunately, this leads to generalisation of data. In some cases, a single municipality polygon may contain more than 200 different pixels. Shown in red in Figure 1 and Figure 2 are all the pixels that changed to urban pixels in reality from 2006 to 2012, compared to those predicted for the same period.

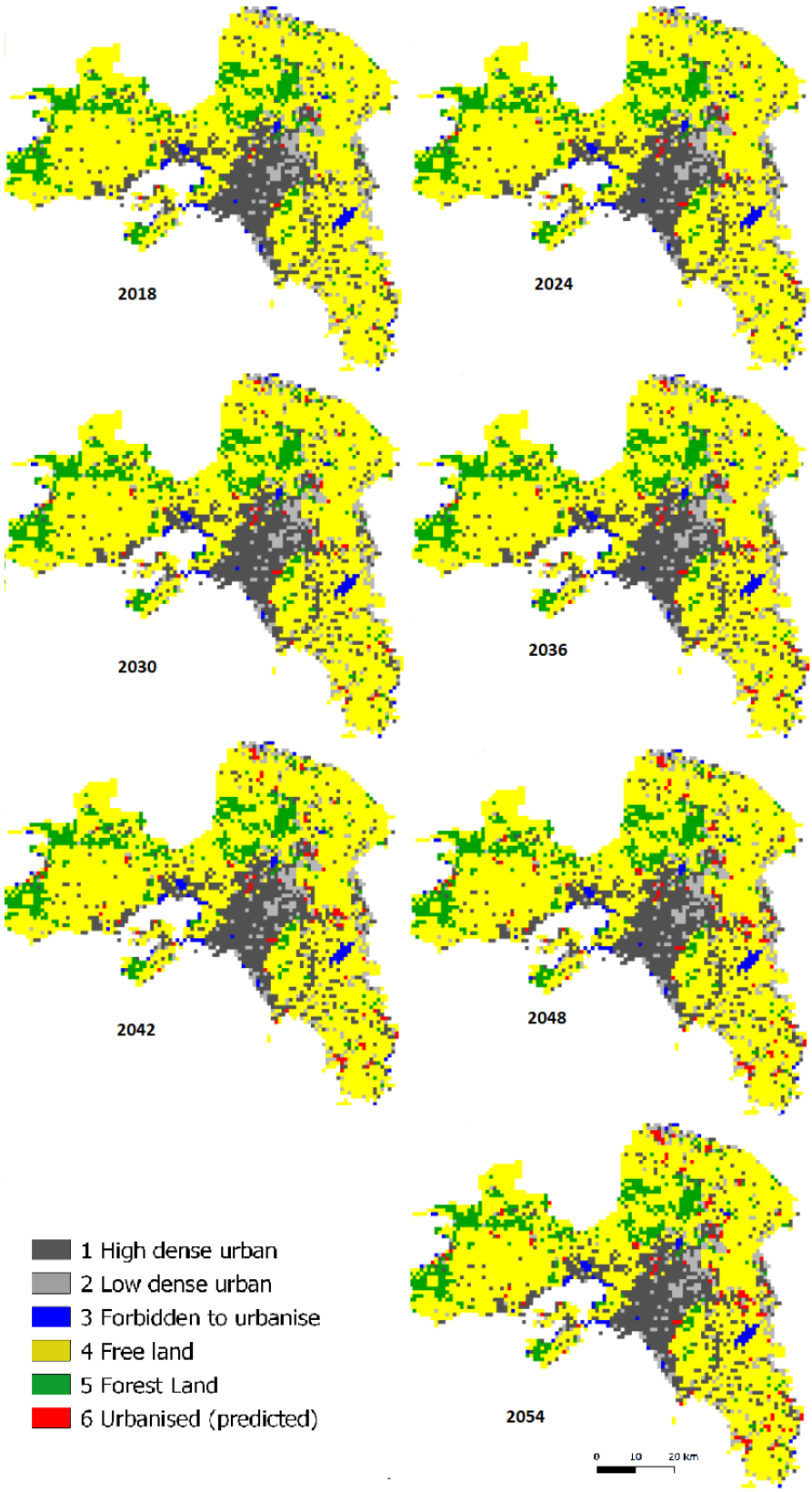

Fig. 3. Simulation 2012–2054
Source: own elaboration

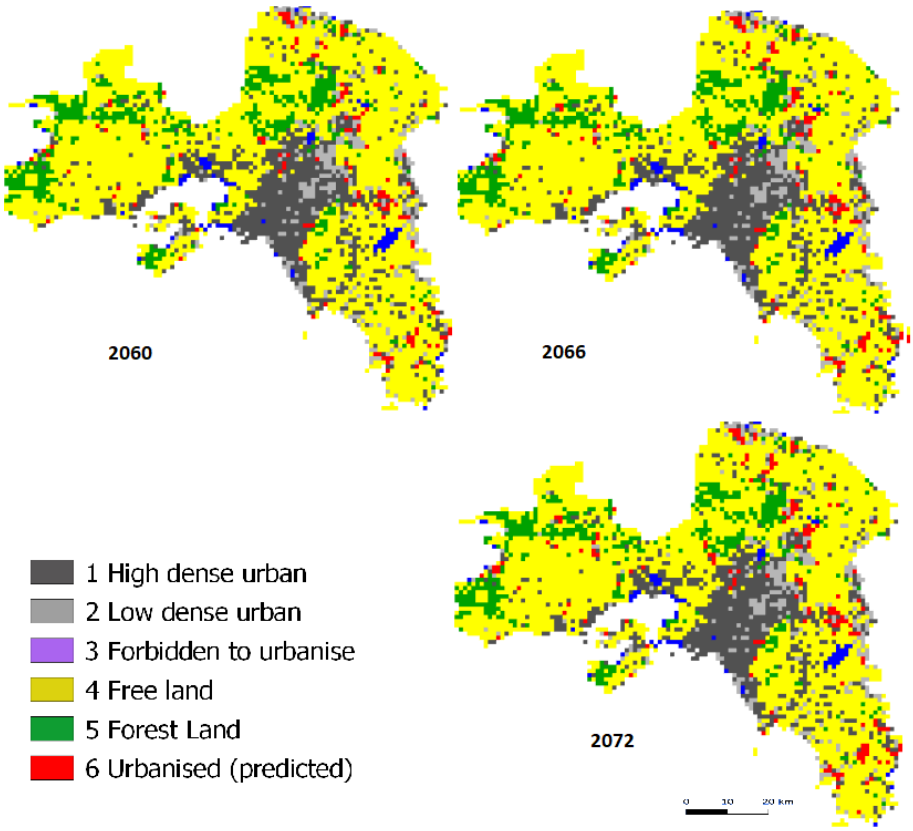

Fig. 4. Simulation 2066–2072
Source: own elaboration

Model simulation to the year 2072

Our simulation will run for ten cycles of six years each, resulting in ten land-use map snapshots from 2012 to 2072. At each step, a grid map will be generated with the following classes: 1) High-density urban land, 2) Low-density urban land, 3) Forbidden to urbanise (lakes, sea, etc.), 4) Not urban / Free to urbanise, 5) Forest areas, 6) New urban areas (urbanised).

Classes 1, 2 and 6 will be considered as urban areas during the model run time.

We are running our model using the weights generated by the calibration process:

- w1 (weight_density). Urban density weight = 0.06
- w2 (weight_road_dist). Road distance weight = 0.8
- w3 (weight_cc_dist). Main city centre distance weight = 0.0
- w4 (weight_sc_dist). Sub-centres distance weight = 0.5
- w5 (weight_slope). Earth slope weight = 0.7
- w6 (weight_pop). Population-change weight = 0.0

The urbanised cells created at each cycle are used, along with those already urbanised, to compute the urban density. Some emerging urban areas appear far from existing urban centres, while most of the newly urbanised land is close to roads and sub-centres (Fig. 3–4). Moreover, our model seemed to have a preference for areas of low earth slope and neighbouring areas with existing urban cells. The ten-step time-series future urban distribution map can be seen in Figure 3 and Figure 4. Red pixels represent newly urbanised pixels, and light and dark grey represent pixels urbanised before 2012, which is our starting point.

6. Limitations and strengths

There are three major limitations in this study that may be addressed in future research. The first is the generality of population data used for the study area. This weakness was observed during the calibration process, as the population weight had a value of 0, meaning it had zero influence on the model. The problem lies in that census data are classified using the spatial entity of municipality, which is a large area compared to the cell of 1000 m × 1000 m used for the rest of our data. To address this problem, a more accurate census dataset must be used.

The second limitation concerns the absence of the 2018 Urban Atlas dataset, which has not yet been published but would enrich our model with an additional sample with which to train and calibrate it. Finally, the third limitation concerns the economic crisis the study area suffered during the study time (2006–2012); this is expressed as a low urban growth rate (1.6%) during the study period. However, some might characterise this limitation as a challenge.

On the other hand, several strengths are worth addressing. Most importantly, all three calibration and evaluation methods used drew approximately the same result. As a consequence, all three methods have selected the same weights to use for the future simulation. Moreover, to apply these methods and estimate kappa, fuzzy kappa and distance index, we used only the pixels that changed between the years 2006 and 2012 and not those that remained unchanged during that time span. This yields a fairer result in terms of calibration method index value.

7. Conclusions

In this paper we implemented a hybrid CA and ABM model using Urban Atlas data for the Athens Metropolitan Area in Greece. A literature review was presented at the beginning of the paper, and described the main model types that modellers can use to carry out urban growth prediction simulations. Moreover, a brief description of the study area was provided, along with the data and sources related to it. Finally, we presented and analysed all the transition rules applied to our model that were used for the calibration and simulation processes.

Kappa statistic, fuzzy kappa and distance index methods were used and combined for the calibration process and thus identify the best combination of weights. To achieve that, we produced 87,846 different prediction images using weight ranges for each transition rule of the model. Furthermore, we compared every produced image against the real change image to calculate kappa, fuzzy kappa and distance index metrics for each image. Kappa fuzzy and overall metric selected the same single image with the highest score. However, kappa selected twelve images holding the highest score; one of them was equal to the one picked by the overall and kappa fuzzy metrics. We used Gama Platform as the toolkit to carry out the whole procedure of model calibration and implementation and, finally, to draw the future prediction images.

According to the calibration results, street network and earth slope were the most important agents of our model, whereas travel distance to the city centre of Athens and population trends had the lowest impact on the model. On the other hand, distance to sub-centres had a significant impact on the urban evolution, while urban density had a low but representative impact on the final shape of urban transformation.

Furthermore, 0.0506% of the free-to-urbanise cells became urban, which translated to a 1.6% real urban growth rate. None of the forest pixels turned into urban, but some of them turned to free-to-urbanise; thus, forest pixels were used in our model implementation as forbidden-to-urbanise. Finally, we ran our model for ten steps of six years at each step and presented the results as a time-series image map set.

In short, this study presents an urban growth model accompanied by its methodology, data structure and adaptation, calibration methods, model application and simulation results. The simulation results can be used to evaluate the impact of the transition rules to the urban evolution as well as to predict future urban changes – not only in Athens, but also in other cities – which in turn may provide valuable references to policy makers.

Acknowledgements

My dear supervisor Yorgos N. Photis, professor at NTUA (National Technical University of Athens), who unfortunately passed away recently, contributed to the conceptualisation and the methodology of this paper and should be cited among the authors. However, editorial rules do not allow this because obtaining his consent for publication was no longer possible when the paper was completed.

References

Baptista, R., Farmer, J. D., Hinterschweiger, M., Low, K., Tang, D. & Uluc, A. (2016). Macroprudential policy in an agent-based model of the UK housing market. Bank of England Staff Working Paper No 216.

Beriatos, E.A. (2006). Glocalising'urban landscapes: Athens and the 2004 Olympics. In: Dialogues in Urban and Regional Planning, 83-116, Routledge.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99: 7280-7287.

Buttner, G., Soukup, T. & Kosztra, B. (2014). CLC2012 addendum to CLC2006 technical guidelines. Copenhagen (EEA): European Environmental Agency.

Chakraborti, S., Nath Das, D., Mondal, B., Shafizadeh-Moghadam, H. & Feng, Y. (2018). A neural network and landscape metrics to propose a flexible urban growth boundary: A case study. *Ecological indicators*, 93: 952-965.

Chaudhuri, G. & Clarke, K. (2013). The SLEUTH land use change model: A review. *Environmental Resources Research*, 1(1): 88-105.

Clarke, K.C. & Gaydos, L.J. (1998). Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International journal of geographical information science*, 12(7): 699-714.

Clarke, K.C., Hoppen, S. & Gaydos, L. (1996). Methods and techniques for rigorous calibration of a cellular automaton model of urban growth. V Third International Conference/Workshop on Integrating GIS and Environmental Modeling, 21-25, Citeseer.

Cohen, B. (2004). Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. *World development*, 32(1): 23-51.

Cohen, J. (1960). Kappa: Coefficient of concordance. *Educ Psych Measurement*, 20: 37-46.

community, O.S. (brez datuma). GEODATA.gov.gr. (geodata.gov.gr) Pridobljeno iz http://geodata.gov.gr/

Conway, J. (1970). The game of life. *Scientific American*, 223(4): 4.

Dongya, L., Xinqi, Z. & Wang, H. (2020). Land-use Simulation and Decision-Support system (LandSDS): Seamlessly integrating system dynamics, agent-based model, and cellular automata. *Ecological Modelling*, 417(1): 108924.

Drogoul, A., Taillandier, P., Gaudou, B., Grignard, A., Nghi, H., Minh, T.T., Marilleau, N., Caillou, P., Vo, D-A. & Xuan, V.T. (2019). GAMA Platform. Available at: https://gama-platform.github.io/ (Access 1 July 2019).

Eugenio, B.D. & Glass, M. (2004). The kappa statistic: A second look. *Computational linguistics,* 30(1): 95-101.

Fleiss, J.L., Cohen, J. & Everitt, B.S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological bulletin*, 72(5): 323.

Foody, G.M. (2007). Map comparison in GIS. Progress in *Physical Geography*, 31(4): 439-445.

Gilbert, N. (2008). *Agent-based models*. Sage.

Goldstein, N.C. (2004). Brains versus brawn-comparative strategies for the calibration of a cellular automata-based urban growth model. *GeoDynamics*, 249-272.

Grekousis, G., Manetos, P. & Photis, Y.N. (2013). Modeling urban evolution using neural networks, fuzzy logic and GIS: The case of the Athens metropolitan area. *Cities*, 30: 193-203. DOI: https://doi.org/10.1016/j.cities.2012.03.006.

Hagen-Zanker, A. (2009). An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. *International Journal of Geographical Information Science*, 23(1): 61-73.

Hagen-Zanker, A. & Martens, P. (2008). Map comparison methods for comprehensive assessment of geosimulation models. *International Conference on Computational Science and Its Applications*, 194-209.

Herold, M. a. (2009). Global Mapping of Human Settlement: Experiences, Datasets, and Prospects.

Herold, M., Hemphill, J., Dietzel, C. & Clarke, K. (2005). Remote sensing derived mapping to support urban growth theory. URS.

Hu, Z. & Lo, C. (2007). Modeling urban growth in Atlanta using logistic regression. *Computers, Environment and Urban Systems,* 31(6): 667-688.

Iacono, M., Levinson, D. & El-Geneidy, A. (2008). Models of transportation and land use change: A guide to the territory. *Journal of Planning Literature*, 22(4): 323-340.

Kazemzadeh, A., Zanganeh, S., Salvati, L. & Neysani Samani, N. (2016). A spatial zoning approach to calibrate

and validate urban growth models. *International Journal of Geographical Information Science*, 31(4): 1-20.

Kuhnert, M., Voinov, A. & Seppelt, R. (2005). Comparing raster map comparison algorithms for spatial modeling and analysis. *Photogrammetric Engineering & Remote Sensing*, 71(8): 975-984.

Langton, C.G. (1997). *Artificial life: An overview*. Mit press.

Li, X. & Gong, P. (2016). Urban growth models: progress and perspective. *Science Bulletin*, 61(21): 1637-1650. DOI: https://doi.org/10.1007/s11434-016-1111-1.

Liu, D., Zheng, X. & Wang, H. (2020). Land-use Simulation and Decision-Support system (LandSDS): Seamlessly integrating system dynamics, agent-based model, and cellular automata. *Ecological Modelling*, 417: 108924.

Martinez, L.M. & Viegas, J.M. (2017). Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal. *International Journal of Transportation Science and Technology*, 1: 13-27. DOI: https://doi.org/10.1016/j.ijtst.2017.05.005.

Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G. & Gotts, N.M. (2007). Agent-based land-use models: a review of applications. *Landscape Ecology*, 22(10): 1447-1459.

Meratnia, N. & de By, R.A. (2002). Aggregation and comparison of trajectories. *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, 49-54.

Moeckel, R., Llorca Garcia, C., Moreno Chou, A.T. & Okrah, M.B. (2018). Trends in integrated land use/transport modeling: An evaluation of the state of the art. *Journal of Transport and Land Use,* 11(1): 463-476.

Montero, E., Van Wolvelaer, J. & Garzon, A. (2014). *The European urban atlas. V Land use and land cover mapping in Europe*, 115-124, Springer.

Musa, S.I., Hashim, M. & Reba, M.N. (2017). A review of geospatial-based urban growth models and modelling initiatives. *Geocarto International*, 32(8): 813-833. DOI: https://doi.org/10.1080/10106049.2016.1213891.

Mustafa, A., Rienow, A., Saadi, I., Cools, M. & Teller, J. (2018). Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. *European Journal of Remote Sensing*, 51(1): 391-401.

Mustafa, A., Cools, M., Saadi, I. & Teller, J. (2017). Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM). *Land Use Policy*, 69: 529-540. DOI: https://doi.org/10.1016/j.landusepol.2017.10.009.

Olmedo, M.T., Pontius Jr, R.G., Paegelow, M. & Mas, J.-F. (2015). Comparison of simulation models in terms of quantity and allocation of land change. *Environmental Modelling & Software*, 69: 214-221.

OpenStreetMap. (2019). Open Street Map. (OSM) Available at: https://www.openstreetmap.org.

Pijanowski, B.C., Tayyebi, A., Doucette, J., Pekin, B.K., Braun, D. & Plourde, J. (2014). A big data urban growth simulation at a national scale: configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. *Environmental Modelling & Software*, 51: 250-268.

Pontius Jr, R.G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, 68(10): 1041-1050.

Pontius Jr, R.G. & Cheuk, M.L. (2006). A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. *International Journal of Geographical Information Science*, 20(1): 1-30.

Raimbault, J., Banos, A. & Doursat, R. (2016). *A Hybrid Network/Grid Model of Urban Morphogenesis and Optimization*. CoRR.

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.

Shafizadeh-Moghadam, H. (2019). Improving spatial accuracy of urban growth simulation models using ensemble forecasting approaches. *Computers, Environment and Urban Systems*, 76: 91-100. DOI: https://doi.org/10.1016/j.compenvurbsys.2019.04.005.

Silva, E.A. & Clarke, K.C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, environment and urban systems,* 26(6): 525-552.

Stathakis, D. & Triantakonstantis, D. (2015). Urban Growth Prediction Using Artificial Neural Networks in Athens, Greece. *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering,* 9(3): 234-238.

Taillandier, P., Banos, A., Drogoul, A., Gaudou, B., Marilleau, N. & Truong, Q.C. (2016). Simulating Urban Growth with Raster and Vector Models: A Case Study for the City of Can Tho, Vietnam. In: International conference Autonomous Agents and Multiagent Systems (AAMAS 2016), 9 May 2016 - 10 May 2016 (Singapore, Singapore). DOI: https://doi.org/10.1007/978-3-319-46840-2_10.

Taillandier, P., Gaudou, B., Grignard, A., Huynh, Q-N., Marilleau, N., Caillou, P., Philippon, D. & Drogoul, A. (2018). Building, composing and experimenting

complex spatial models with the GAMA platform. *GeoInformatica*, 23(2): 299-322.

Torrens, P.M. (2000). How land-use-transportation models wor. Centre for Advanced Spatial Analysis, London.

Triantakonstantis, D. & Mountrakis, G. (2012). Urban growth prediction: a review of computational models and human perceptions. *Journal of Geographic Information System*, 4(6): 555.

Tsagkis, P. & Photis, Y. (2018). Using Gama platform and Urban Atlas Data to predict urban growth. *The case of Athens. 11th International Conference of the Hellenic Geographical Society (ICHGS - 2018), Athens*.

Tsoularis, A. & Wallace, J. (2002). Analysis of Logistic Growth Models. Mathematical biosciences, 179: 21-55.

Un.Nations. (2016). The World's Cities in 2016 – Data Booklet. Department of Economics & Social Affairs.

van Vliet, J., Bregt, A. K. & Hagen-Zanker, A. (2011). Revisiting Kappa to account for change in the accuracy assessment of land-use change models. E*cological modelling*, 222(8): 1367-1375.

van Vliet, J., Hagen-Zanker, A., Hurkens, J. & van Delden, H. (2013). A fuzzy set approach to assess the predictive accuracy of land use simulations. *Ecological Modelling,* 261-262: 32-42.

Visser, H. & De Nijs, T. (2006). The map comparison kit. *Environmental Modelling & Software*, 21(3): 346-358.

Wegener, M. (2004). Overview of land-use transport models. *Handbook of transport geography and spatial systems*, 5: 127-146.

Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science*, 16(8): 795-818.