Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM)

Ahmed Mustafa*, Mario Cools, Ismaïl Saadi, Jacques Teller
Local Environment Management & Analysis (LEMA), UEE Dept., University of Liège
* a.mustafa@ulg.ac.be

Abstract

Several methods for modeling urban expansion are available. Most of them are based on a statistical, a cellular automaton (CA) and/or an agent-based (AB) approach. Statistical and CA approaches are based on the implicit assumption that people's behavior is not likely to change over the considered time horizon. Such assumption limits the ability to simulate long-term predictions as people's behavior changes over time. An approach to consider people's behavior is the use of an AB system, in which the decision-making process of agents needs to be parameterized. Most existing studies, which make use of empirical data to define the agents' decision-making criteria, rely on intensive data collection efforts. The considerable data requirements limit the AB-system's ability to model a large study area, as the number of agents for which data on decision-making criteria is required, increases with the size of the study area. This paper presents a hybrid urban expansion model (HUEM) that integrates logistic regression (Logit), CA and AB approaches to simulate future urban development. A key feature of HUEM lies in its ability to address various people behaviors that are variable over time through AB relying on a sample approach that combines Logit and CA. Three agent sets are defined; developer agents, farmer agents and planning permission authority agent. The agents' decision-making process is parameterized using CA and Logit models. The interactions of the agents are simulated through a series of rules. To assess HUEM performance, it is calibrated for Wallonia (Belgium) to simulate urban expansion between 1990 and 2000. Calibration results are then assessed by comparing the 2000 simulated map and the actual 2000 land-use map. Furthermore, the performance of HUEM is compared to a number of typical spatial urban expansion models, i.e. Logit model, CA model and CA-Logit to assess the added-value of HUEM. The comparison shows the performance of HUEM is better than other models in terms of allocation ability.

Keywords: logistic regression; cellular automata; agent-based; genetic algorithm; Wallonia

1. Introduction

The urban environment is a complex system, which includes a large number of inconstant parameters and several actors (e.g. households, developers, government, etc.). The complexity of such a system is well explored in (Batty, 2007, 2008). Urban expansion models are a tool to gain insight into the mechanisms of the urban environment. These models can project the expected future demands of urban lands and/or a geographical distribution of these demands. Urban expansion models have wide range applications, which expands from global warming (e.g. Haggert, 1995) to response to flood risks (e.g. Beckers et al., 2013; Mustafa et al., 2016; Poelmans et al., 2010).

Several statistical and geospatial approaches have been proposed and developed to model urban expansion, including logistic regression models (Logit) (e.g. Hu and Lo, 2007; Vermeiren et al., 2012), cellular automata (CA) (e.g. Al-Ahmadi et al., 2009; Mitsova et al., 2011; Mustafa et al., 2014) and agent-based models (AB) (e.g. Hosseinali et al., 2013; Zhang et al., 2010).

Often, the urbanization likelihood of a non-urban land is determined by static drivers related to accessibility, geophysical features, policies and socio-economic factors. Another important driver is neighborhood interactions because of the fact that urbanization can be regarded as a self-organizing system (Poelmans and Van Rompaey, 2010). The relative importance of different drivers as determinants of the urbanization likelihood can be based on different methods such as Logit and CA. In this study, we refer to the static drivers as global factors and to the neighborhood interactions as local factors.

Logit models are a common approach to model urban expansion. They predict the outcome of a categorical variables using a set of quantitative and/or qualitative predictors. Logit can include geophysical as well as socio-economic factors. The model's ability to include as many factors as necessary allows us to better understand the main drivers behind urbanization processes. Neighborhood interactions can also be captured in Logit models by including them as part of the explanatory variables as in Hu and Lo (2007) and Verburg et al. (2004). However, because Logit models are not temporally explicit, they cannot reveal the path-dependent and self-organizing development which is typical for urban expansion (Poelmans and Van Rompaey, 2010; Wu, 2002). The most well-known approach to calculating the neighborhood interactions on a dynamic basis is cellular automata (CA) based model, in which the neighborhood state is updated during each simulation step. Cellular models are simple and widely available (Clarke and Gaydos, 1998). However, pure CA models focus on the calculation of urbanization transitions by explicitly consider the immediate neighbors of each landscape unit, i.e. cell, rather than on the interpretation of urbanization drivers. Several studies try to overcome this limitation of CA models by integrating CA with other modeling methods to consider several urbanization drivers. In this context, Logit and CA are commonly combined to create a so-called 'CA-Logit model', which considers both the urbanization static drivers and the dynamic neighborhood interactions (e.g. Poelmans and Van Rompaey, 2010).

One of the clear drawbacks of Logit, CA and CA-Logit approaches is related to the lack of the theoretical link between the spatial rules and agents' decisions within the urban environment. Agent-Based (AB) models, which are less frequently used in the context of urban expansion modeling, forecast agents as goal-oriented entities capable of responding to their environment and interacting with each other. Agents in the model can play a role of individuals or groups of people, institutions, etc. They can exhibit different characteristics: they can be heterogeneous (e.g. economic state, age, family structure), autonomous (they take their own decisions based on analytical functions) and dynamic (they can learn and adapt to different conditions) (Valbuena et al., 2008). The agents are commonly grouped into homogeneous sets of individuals with comparable characteristics and behaviors. Generally, the decision-making criteria of agents require a large amount of data stemming from surveys that depict people's choices and utilize experts' knowledge. In a large study area, such an intensive data gathering is limited by a large number of agents (Valbuena et al., 2008).

This paper introduces an urban expansion model, namely a hybrid urban expansion model (HUEM), combining the simulation capabilities of Logit, CA and AB approaches. HUEM is a predictive model, which simulates future urban expansion. Agents' decisions are governed by a series of different possible behaviors, which are themselves variable over time. The non-urban to urban conversions of Wallonia (south Belgium) between 1990 and 2000 is used as a case study to demonstrate the applicability of HUEM to urban expansion modeling. Engelen et al. (2016) developed a spatial land-use change model for Flanders (north Belgium), called RuimteModel (Poelmans et al., 2013; White et al., 2015). It is a CA-based model that simulates annual changes of several land-use classes, with a resolution of 1 hectare. When compared to this approach, our model couples AG and CA, which allows us to compare the performances of different modeling approaches.

The behavior of urban agents is established based on AB model. In AB model, urban developers (DevA) seek to develop non-urban cells with the highest urbanization probability and they do so if the urbanization probability exceeds a farmer (FarmA) satisfaction threshold and is approved by the planning permission authority (PPA) who tries to ensure that future urban expansions are in accordance with the official zoning plan. CA and Logit are embedded into the AB for calculating the urbanization probability and farmers satisfaction instead of gathering data from surveys. In addition, HUEM facilitates the incorporation of different ancillary data (e.g. how strictly should urban regulations enforce urban expansion).

The main contribution of this paper is the added-value of combining agents' behavior and decisions into a typical CA-Logit model, in which spatial entities are the basic units of simulation. To this end, HUEM and CA-Logit models are compared. In addition, both models are compared with Logit and CA models to decide whether the added complexity of the combination is worth reduction in degrees of freedom.

The following sections describe model specifications, case study, results, and then give conclusions as well as suggestions for future study.

2. The Hybrid Urban Expansion Model (HUEM)

In this section, we describe the main components of HUEM model. The overall workflow is shown in Fig. 1.

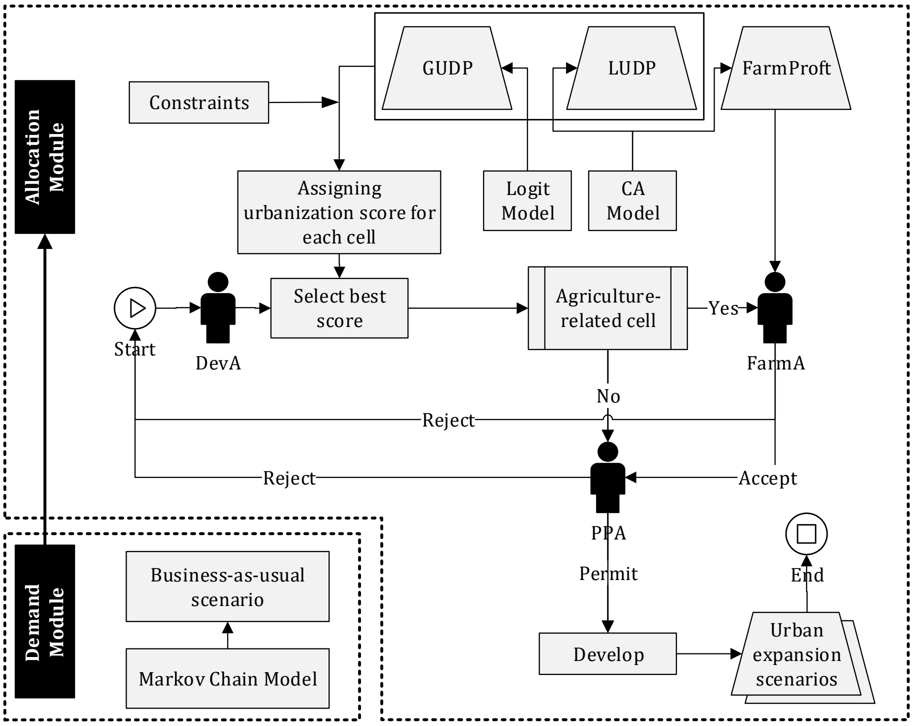


Fig. 1. Methodological flowchart of the Hybrid Urban Expansion Model (HUEM).

The model's space consists of a 2D array of cells of the same dimensions. Three groups of agents are included in the model: developer (DevA), which represent firms and households; farmer (FarmA) and the planning permission authority (PPA). Actually one of the predominant forms of land-use change is the transformation of agriculture-related lands to built-up lands (e.g. Poelmans and Van Rompaey, 2009; Sang et al., 2011). Consequently, some FarmAs may decide to stop their activity as farmers and switch their type to DevAs.

In HUEM model, the development of non-urban cells is realized by DevA and controlled by PPA. FarmAs, owning agriculture-related cells, will decide to keep or to sell their own cells.

A non-urban cell can be developed when three conditions are simultaneously satisfied: (i) the profitability of urban development is high, (ii) PPA allows construction in this cell and (iii) there are no constraints. The constraints are restrictive cases for urban development. Such constraints could include but are not limited to, flood-prone zones. Besides, it is defined that if a cell state is urban in a specific time-step, it automatically remains the same in the next time-steps.

HUEM is first calibrated and assessed with real land-use data of at least two-time frames and is then used to project possible future urban expansion scenarios at a specific time horizon. Generally, the key features for understanding the design concept of HUEM are observation, tuning, and uncertainty.

- *Observation*. We consider that the evolution of urban development is based on several socioeconomic, geographic and even political aspects that are referred to as urbanization driving factors. HUEM calculates the probability of urban development combining three layers that define cell probability for urban development: the local urban development probability (LUDP), the global urban development probability (GUDP) and the farming profitability (FarmProft). The LUDP and the FarmProft layers are developed using CA model. The GUDP layer is developed using Logit model. These layers are based on the exploring of past land-uses.

- *Automatic calibration*. The model does not require any prior knowledge about a specific study area. It employs Logit model and genetic algorithm (GA) to automatically calibrate all model's parameters.
- *Uncertainty*. Urban expansion models have inherent uncertainties related to the future values of model parameters. HUEM considers uncertainties through a set of various possible agents' behaviors.

HUEM consists of two modules: (i) a demand module and (ii) an allocation module. The demand module calculates the quantity of new urban cells at each time-step, whereas the allocation module spatially distributes this quantity over space. Generally, the quantity of new urban cells can be computed by several means including the Markov chain model (MC) (e.g. Sang et al., 2011; Yang et al., 2014), linear extrapolation (e.g. Mustafa et al., 2014; Poelmans and Van Rompaey, 2009) and/or based on socioeconomic factors (e.g. White and Engelen, 2000). It is hard to estimate a highly accurate projection of urban land demand in the future because of the complexity of the urban system and its related socioeconomic dynamics (He et al., 2008).

HUEM can either be fed with the expected quantity of new urban cells or computes the quantity based on past trend using the MC model to develop a so-called business-as-usual scenario. The MC model is explored in a number of studies such as (Guan et al., 2011; Puertas et al., 2014; Sang et al., 2011; Shafizadeh Moghadam and Helbich, 2013; Yang et al., 2014).

In the model, each time-step corresponds to one year which would be adequate in a model of land-use change (White and Engelen, 2000). At the initialization of the model, the actual land-use maps of at least two time-steps are uploaded into the model and the agents are created. FarmA controls all agricultural-related cells. PPA controls other land-uses except for urban cells and sets zoning constraints for the entire study area based on three categories of urban development; (1) permitted, (2) severely restricted and (3) forbidden.

2.1. Allocation module

The allocation module is the key part of the model representing the decision-making criteria of the agents to address the location of the estimated quantity of new urban cells between different points in time. Once the estimated quantity is reached, the module stops the allocation process. This module is typically calibrated using training data (i.e., past land-use maps).

2.1.1. Agents' decisions and interactions

The first step of the allocation module is the determination of the ideal non-urban cells to be developed in the next time-step to meet the required demand. To this end, the agents have to interact and decide which cells to develop. DevAs visit all non-urban cells and calculate the probability score of urban development for each cell. DevAs record the positions, the states and the probability scores of the visited cells and learn FarmAs and PPAs.

Commonly, authors consider various parameters representing decision-making criteria of agents to select cells for urban development based on qualitative and/or quantitative approaches (e.g. Matthews et al., 2007; Parker and Meretsky, 2004; Ralha et al., 2013). A quantitative approach is used in HUEM to parametrize the decision-making criteria. When DevAs have the opportunity to make a decision about urban development, they first form an urban development probability score. The probability score is calculated as follows:

$$score_{c_{i,j}}^{t} = LUDP_{c_{i,j}}^{t} \times GUDP_{c_{i,j}}^{t} \qquad (1)$$

where $score_{c_{i,j}}^{t}$ is the probability score of the urban development assigned to cell $c_{i,j}$ at time t, $LUDP_{c_{i,j}}^{t}$ is the local urban development probability according to the neighborhood effects on the cell and $GUDP_{c_{i,j}}^{t}$ is the global urban development probability according to the geo-physical and socio-economic factors.

In this stage, DevAs know the expected demands of urban area based on demand module and understand the profit-maximizing of global and local factors. When DevA selects an agricultural-related cell to develop, FarmA will make a decision on selling or preserving her/his cell. In principle, FarmA aims to maintain or increase her/his profitability, and to keep or even expand her/his cropped area. FarmAs imitate the land-use of their neighbors and therefore they are highly affected by urban neighbors. Increasing urban neighbors of a farmland may result in FarmAs dissatisfaction as operating small farmlands are economically infeasible (Bert et al., 2011). We assumed that the FarmAs' cells are negatively or positively affected in terms of agriculture profits by spatial externalities generated their neighbors. These externalities result in a loss or gain in FarmA's profitability ω. If FarmA's profitability drops below the probability score of urban development, s/he must exit farming at the current time-step as follows:

$$FarmADec_{c_{i,j}}^{t} = \begin{cases} accept, & \omega_{c_{i,j}}^{t} < score_{c_{i,j}}^{t} \\ reject, & \omega_{c_{i,j}}^{t} \geq score_{c_{i,j}}^{t} \end{cases} \qquad (2)$$

where $FarmADec_{c_{i,j}}^{t}$ is FarmA decision on selling or keeping her/his cell. When DevAs determined which cells to develop, they have to ask for a development permission from PPA. PPA realizes that policies are not always strictly enforced. If a cell is in a permitted or in a forbidden zone, PPA will instantaneously grant or reject the permission respectively. Otherwise, if the cell is in a severely restricted zone, a sort of competition will be carried out to find the development decision. The model defines the winner of the competition based on the number of times that PPA has lost cells in the previous competitions. In other words, PPA will give permissions for a specific percentage of the amount of required new urban cells (allowed rate) to be developed within the severely restricted zones as the follows:

$$GovADec_{c_{i,j}}^{t} = \begin{cases} accept, & LR_t < AR_t \\ reject, & LR_t \geq AR_t \end{cases} \qquad (3)$$

where $GovADec_{c_{i,j}}^{t}$ is the decision within the severely restricted zones, LR_t is the loss rate and AR_t is the allowed rate.

2.1.2. LUDP and FarmProft layers

In human-based systems, the idea of locality is hard to realize clearly, since agents are aware of their surroundings in a wide space. Thus, it is desirable to set a neighborhood large enough to capture the operational range of the local processes being modeled (White and Engelen, 2000). In some land-use change models (e.g. Poelmans and Van Rompaey, 2009; White and Engelen, 2000; Wu, 2002) the neighborhood is defined using all surrounding cells within a radius between one to eight cells.

A CA model is applied to define the LUDP for each cell at the next time-step according to the procedure proposed by White and Engelen (2000):

$$LUDP_{c_{i,j}}^{t} = \sum_{d} \sum_{x} uw_{kxd} \qquad (4)$$

where uw_{kxd} is the weighting parameter applied to land-use k at position x in distance zone d to represent the interaction with urban cell. The CA model is also applied to set the FarmProft according to the following formula:

$$\omega_{c_{i,j}}^{t} = \sum_{d} \sum_{x} aw_{kxd} \qquad (5)$$

where aw_{kxd} is the weighting parameter applied to land-use k at position x in distance zone d to represent the interaction with agricultural-related cell.

The weighting values that define the neighborhood's attraction or repulsion for urban and agriculture land-uses are calibrated based on GA.

2.1.3. GUDP layer

We consider that the GUDP are driven by several social, economic, geographic and politic factors. Many of these factors are difficult to be modeled and predicted. Notwithstanding certain factors, referred to as urbanization driving factors, can be taken into account to predict future urban expansion. (Bičík et al., 2001; Bürgi et al., 2005; Li et al., 2013; Mustafa et al., 2015; Verburg et al., 2004), among others, reviewed such factors.

Logit model is used to capture the relative contribution of each factor, focusing on the changes from non-urban to urban land-use. The input dependent variable (Y) is a is a binary map showing the observed changes from non-urban to urban cells (coded as 1) and cells whose status remains non-urban (coded as 0). The independent variables (X_n) are a set of urban development driving factors. Logit analysis yields coefficients for each X_n, which can be interpreted as weights in a formula that generates a GUDP map depicting the probability of each cell to be developed into urban as:

$$GUDP_{c_{i,j}}^{t} = \frac{exp(\alpha + \sum_n \beta_n X_n)}{1 + exp(\alpha + \sum_n \beta_n X_n)} \qquad (6)$$

where α is the intercept and β_n are the regression coefficients. HUEM evaluates the goodness-of-fit using the relative operating characteristic (ROC) procedure.

Prior to estimating Logit model parameters, it is important to check for three aspects that may exist in X_n: disparity in units, autocorrelation, and multicollinearity (Mustafa et al., 2015). It is quite common to have a disparity in units and even scale of X_n, for instance, some X_n may be measured in meter (such as distances to roads) and others in percentage (such as slope). As a result, all continuous X_n will be standardized before performing Logit model.

Spatial autocorrelation in one or more X_n will bias the results of the regression analysis. Autocorrelation is the propensity of a cell value to be nearly similar to other nearby cells. Normally, almost all X_n can show a strong degree of spatial autocorrelation (Cammerer et al., 2013; Crk et al., 2009; Li et al., 2013). To overcome this problem, a number of authors suggested to selecting a structured or random sample from the study area (Cammerer et al., 2013; Li et al., 2013). HUEM selects a random sample of the study area with an equal number of 0 (no change) and 1 (change) observations of the dependent variable. Unequal sampling rates do not affect the estimation of β_n, but only affect the intercept (Allison, 1999).

Multicollinearity shows a high degree of dependency among a number of X_n because some of X_n may measure the same phenomena (Mustafa et al., 2015). Strong degree of multicollinearity causes the erroneous estimation of parameters (Lin et al., 2014). HUEM uses variance inflation factors (VIF) to detect multicollinearity. Montgomery and Runger (2003) recommended the VIFs should not exceed 4. HUEM suppresses all X_n with VIF of 4 or larger.

After performing Logit model, the GUDP layer is computed based on the β_n of the X_n that represent agents' responses in terms of global urban development attractiveness. In order to capture the extensive range of agents' responses, we pick 1000 random samples of cells and estimate the β_n for each set. By using a range of possible values of β_n, we can capture a more realistic picture of agents' responses. Selecting a value from the 1000 different sets of coefficients to compute the GUDP layer can be done by using a measure of central tendency, e.g. the mean or median value, or by selecting a value from the samples randomly.

2.2. Calibration of model parameters

The purpose of the calibration process is to set the optimal values of parameters combination that can achieve the highest accuracy rate. The accuracy rate is measured in this phase using the cell-to-cell location agreement (CTC).

Calibration of model parameters includes the allowed rate of urban development within the restricted zones (Eq. 3), neighborhood weights of the LUDP and the FarmProft layers (Eq. 4, 5) and the GUDP parameters (Eq. 6). HUEM considers the influence of uncertainty about future behaviors through a combination of various possible agents' behaviors. The possible agents' behaviors can be captured through ranges of the model's parameters. To set the best ranges, a comprehensive uncertainty sensitivity analysis should be done which is outside of the scope of this paper. However, for our case study, we select the optimal values of parameters in order to develop HUEM, CA-Logit, CA and Logit simulations of 2000. GUDP parameters α and β_n are calibrated using Logit based on a maximum likelihood estimation procedure. Other parameters are automatically calibrated using the genetic algorithm (GA).

Recently, GAs are employed to calibrate urban expansion models (e.g. Al-Ahmadi et al., 2009; García et al., 2013; Shan, Alkheder, & Wang, 2008). García et al. (2013) claimed that the GA is one of the most robust heuristics automated methods to calibrate urban expansion models. GA is an evolutionary algorithm and is inspired by natural selection and adaptation (Holland, 1975). It seeks to find the global, or near global, optimal solution without ever requiring knowledge of search space being optimized. GA begins with a random initial population in which many solutions participate in an iteration (generation). It then employs a set of operators to reveal interesting regions of the search space using fitness function of the solutions at hand to produce a new generation. These operators are the selection of parents for the next generation, crossover, and mutation.

GA selects the best individuals in the current generation for mating so as to produce superior solutions by combining parts of parent solutions. Tournament selection is a robust selection method commonly used by GAs (García et al., 2013; Miller et al., 1995). Tournament method selects a number of individuals from the population at random and selects the best out of these to become a parent. Each two parents are combined based on a crossover operator and generate two children. Each child is then perturbed in its vicinity by a mutation operator that adds a small random number to each gene.

There is no general guide available to set the GA parameters. One approach for parameter settings is by undertaking empirical experiments on different values of the parameters using a small number of generations and population and choosing the best ones (Al-Ahmadi et al., 2009). Based on these empirical experiments, we set GA parameters for the final run. In the final GA run, the population size is set at 100 per generation while the algorithm terminates the run if the weighted average change in the best fitness value for 10 consecutive generations is less than 0.0001. The tournament selection is set at 4 individuals. The crossover operator generates two children that lie on the line representing both parents and inherit at least 65% genes from the parent with the better fitness value. In order to fulfill a good balance between the exploration of the entire search space and the convergence of the population towards the globally optimal solution, the mutation operator selects a random number from a Gaussian distribution with a center of zero and a standard deviation of 1.2 at the first generation. This standard deviation is shrunk to 0 linearly as generation 100 is reached. Consequently, the GA explores much more search space at the beginning of the optimization process and ensures the convergence of the population towards the global optimal solution by the end of the process.

The objective function for the GA is based on CTC. The parameters' values that lead to maximizing the objective function will be selected as the best calibration outcome.

2.3. Model assessment

The assessment of the model is the process of measuring the model predictive performance. The assessment procedure consists of (i) the evaluation of the GUDP layer computed by Logit model using the ROC procedure and (ii) the comparison of the simulated urban maps of 2000 with the real map of 2000.

First, the ROC is used to compare the outcomes of Eq. 6 to a map with the real changes of urban cells from 1990 to 2000. ROC calculates the proportion true-positives and false-positives for a number of thresholds and relates them to each other in a graph. It then measures the area under the curve which should vary between 0.5 (random fit) and 1 (perfect fit).

Second, to evaluate the simulated urban map, we applied two statistical techniques of map comparison: (i) CTC and (ii) evaluation of the structure of new urban pattern in terms of landscape compactness and complexity. CTC is one of the most explicit ways to evaluate the outcome of urban expansion models. It produces a stringent test of simulation as it measures on a cell basis (Wu, 2002). Consequently, it cannot evaluate the morphology of the urban spatial structures. To address the landscape morphology of our model outcomes, we evaluated how a model simulates spatial properties. Two matrices measuring fragmentation (number of patches and mean patch area), one matrix measuring the complexity (area-weighted mean shape index) and one matrix measuring dispersion (patch cohesion index) are selected to evaluate the model's outcome landscape pattern. Small differences on these metrics show a good correspondence between the simulated and real patterns in terms of landscape structure.

3. Case study: Wallonia, Belgium

3.1. Study area

To demonstrate the feasibility of HUEM model, Wallonia, Belgium is taken as an example application. Wallonia is situated in the southern part of Belgium at 49°28' to 50°49' N latitudes and 2°50' to 6°28' E longitudes, Fig. 2. Wallonia is the predominantly French-speaking region of Belgium. It accounts for 55% of the territory of Belgium with a total area of 16,844 km^2. The population in 2010 was 3,498,384 inhabitants that makes up a third of Belgium population (Belgian Federal Government, 2015). Administratively, it comprises five provinces: Hainaut, Liège, Luxembourg, Namur, and Walloon Brabant. With its 866 km roads, 1,605 km of railway lines, 453 km waterway network and two regional airports, Wallonia is so very accessible. Wallonia has a pronounced undulating topography. The topography goes from flat to hilly with altitude ranges from 0 to 693 m above sea-level. This means cycling is almost non-existent in Wallonia (Dujardin et al., 2012). Major cities in Wallonia are characterized by a strong center–periphery structure with well-off households located in the peripheries (Verhetsel et al., 2010). The main urban areas are Charleroi, Liège, Mons and Namur. They are all characterized by a historical city-center, around which the urban development was spread. Urban sprawl has affected Wallonia for decades leading to fragmented and isolated landscapes that were developed in space and time (Antrop, 2004).

During the 1970ies and early 1980ies, Belgium adopted a zoning plan covering the entire territory of the country (plan de secteur). This plan regulates the types of activities that can be accommodated on a specific zone. This plan is routinely updated. One of the deficiencies of this plan is the lack of realistic scenarios of future urban expansion. Consequently, there are serious failures to comply with the plan. This situation necessitates a better understanding of the mechanisms of urban expansion in Belgium to develop a more feasible zoning plan. Table 1 summarizes zoning information and urban expansion in Wallonia.

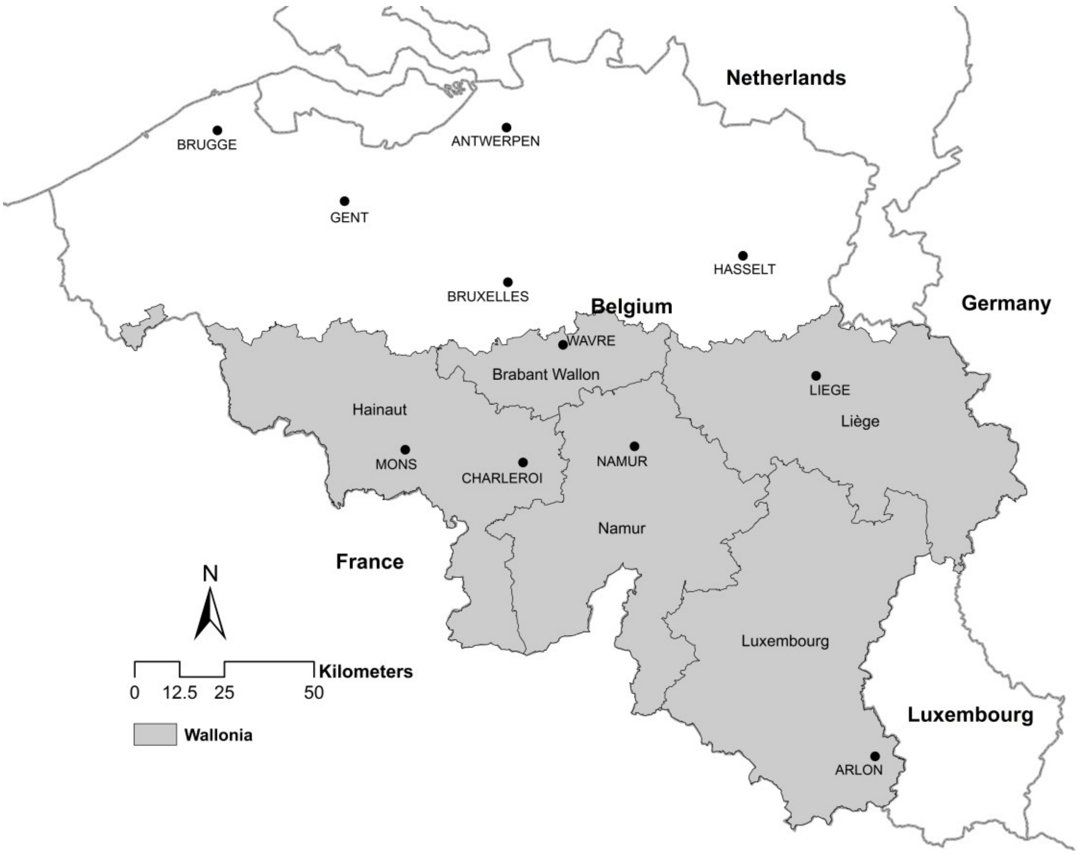


Fig. 2. Study area

Table 1. Zoning and urban expansion between 1990 and 2000.

	1990 (cells in thousands)	2000 (cells in thousands)	Expansion rate (in percentage)
Total	1689.69	1689.69	
urban	241.08	281.23	-
permitted	132.10	108.09	59.79
severely restricted	1313.10	1297.20	39.91
forbidden	3.42	3.18	0.30

3.2. Data

The CORINE Land-Cover (CLC) datasets give a detailed inventory of the biophysical land cover in Europe using 44 classes. It is made available by the European Environment Agency (EEA) (http://www.eea.europa.eu/data-and-maps) at resolutions of 100×100m and 250×250m grid cells. In this case study, the original 44 land-use classes are reclassified into seven aggregate land-use classes: 1.Urban lands, 2.Arable lands, 3.Grasslands, 4.Forests, 5.Wetlands, 6.Water bodies and 7.Others. The Navteq streets of 2002 dataset are used to calculated Euclidean distances to four functional road classes in meters: 1.high speed roads, 2.quick travel between and through cities, 3.moderate speed travel within cities and 4.moderate speed travel between neighborhoods.

Euclidean distances to cities are calculated for the major 11 Belgian cities, including major cities in Brussels and Flanders regions (Fig. 2). The border effect of Brussels and Flanders regions is implicitly considered in our case study through this variable. Access to jobs is measured as the number of jobs available within 20km for each municipality.

Digital Elevation Model (DEM) provided by the Belgian National Geographic Institute is used to calculate slope in percentage for each cell.

According to the most recent zoning plan of Wallonia, urban development is only allowed in those zones that are designated for residential, economic or leisure development. In other zones, such as agricultural and forest areas, urban development is not permitted unless specific conditions. The zoning map is developed by discerning zones where urban development is not permitted (code 0) and zones that are designated for urban development (code 1). All maps are created as raster grids with a resolution of 100×100m.

3.3. Results and discussion

The CLC of years 1990 and 2000 are used in this paper to calibrate and assess the model framework. The urban class in our model configuration consists of land that is covered by buildings and other man-made elements such as residential areas and related functions services, industries, firms, and transport infrastructure. Among the 1,448,553 cells that can be converted into urban land-use between 1990 and 2000, 40,151 cells were converted into urban lands over those ten years. The PPA sets three zones categories: (1) permitted (urban zones), (2) severely restricted (arable lands, grasslands, forests, wetlands and other classes) and (3) forbidden (water bodies).

The result of a calibration shows that the optimal value of the allowed rate of development within the restricted zones (Eq. 3) is 0.16%. With regard to the LUDP and the FarmProft layers, the neighborhood space is set as a square region around the cell under evaluation and contains nine cells, including the central cell, that are arranged in one square distance zones *d*. The best weighting values that define neighborhood interactions for the LUDP and the FarmProft are shown in table 2.

Table 2. Calibrated weighting values of the neighborhood.

	LUDP		FarmProft	
Distance (cells)	0	1	0	1
Urban land	-	12.91	-	-6.17
Arable land	10.31	-6.77	1.60	3.83
Grassland	1.81	-6.21	1.28	1.84
Forest	-3.27	-6.62	-	10.78
Wetlands	0.09	-4.59	-	1.91
Water bodies	-	1.54	-	0.43
Others	-0.22	-1.52	-	-1.52

The LUDP weighting values that represent the interaction between different land-uses and urban cell imply that arable lands play an obvious role at the zero-distance. The urban development of arable lands is quite common. Grasslands are also easy to be developed into urban land but less common than arable lands. On the contrary, the conversion from forestland to urban land is rare.

This calibration is somewhat in line with the actual number of changed cells from each land-use. The original land-uses in 1990 of the new urban cells in 2000 were 62%, 22%, 12% and 4% arable lands, grasslands, forests and others respectively. The calibration also shows that the impact of existing urban land on new urban development is extremely significant in the immediate neighborhood of the cell.

Concerning the FarmProft, that defines externalities effects on FarmA's profitability, arable and grasslands show a positive effect at distance zero. Urban land in the immediate neighborhood has a strong negative effect on FarmA profitability, while grassland, arable land, and forest have a positive effect.

The urban development driving factors (X_n) considered to develop the GUDP layer are distance to four road classes, distance to major cities, slope, access to jobs and zoning. All X_n are standardized and shown a very low degree of multicollinearity (variance inflation factors ranging from 1.01 to 2.76). Consequently, all selected X_n are used in Logit model. Logit is calibrated using a random sample of 50,000 cells with an equal number of 0 (non-urban cells in 1990 and 2000) and 1 (non-urban cells in 1990 and urban cells in 2000) observations of the dependent variable (Y) to minimize spatial autocorrelation, after standardization of X_n. The model selects the median value of each coefficient set. Table 3 gives the mean values, standard deviation, mean P-values and mean standard errors of coefficient sets.

Table 3. Coefficient values of the driving factors.

Driving factor	Mean coefficient	StDev*	Mean P-value	Mean S.E.**
Intercept	-0.4887	0.0070	-	0.0135
Slope	0.0002	0.0005	0.5330	0.0005
Dist to cities	-0.1979	0.0114	0.0000	0.0143
Dist to road 1	-0.1965	0.0125	0.0000	0.0152
Dist to road 2	-0.2292	0.0132	0.0000	0.0155
Dist to road 3	-0.3187	0.0123	0.0000	0.0141
Dist to road 4	-0.5678	0.0155	0.0000	0.0178
Access to jobs	0.0005	0.0026	0.5103	0.0009
Zoning	1.0379	0.0113	0.0000	0.0122

*StDev: standard deviation

**Mean S.E.: mean standard error

The small standard deviations in table 3 indicate that the mean tendency of the coefficient sets is very stable. Thus, the impact of the sampling procedure is negligible.

These coefficients reveal that the location of a new urban development is strongly correlated with the zoning status. Distances to different road classes and cities also play an important role in explaining urban development at a specific location. Furthermore, urban expansion tends to occur close to job locations and on relatively hilly terrains. However, the contribution of the variables slope and access to jobs to the urban development is small.

The ROC value of the GUDP layer is 0.78. The modest ROC value is understandable, as there will be other factors that can influence the location decision of urban development. However, the GUDP layer can still be used for reliable predictions of the future urban development in the Wallonia. ROC values higher than 0.70 are considered as a reasonable fit and can be introduced in further analyses (Cammerer et al., 2013; Jr and Lemeshow, 2004).

To evaluate the added-value of HUEM model for simulating urban expansion, a number of urban expansion simulations are tested based on (1) HUEM model, (2) CA-Logit (3) CA, and (4) Logit. Logit model is based on the GUDP layer, CA model is based on the LUDP layer and CA-Logit is based on the probability map produced by Eq.1. CA-Logit can be viewed as a matter of complexity. The AB rules which are applied by DevA, FarmA and PPA agents can be viewed as a methodology rather than just a combination of different methods. The initial state and the number of changed cells are kept constant in all simulations.

In the four models, cells with the best urban probability scores are selected at each time-step. In order to set the change rate per time step, a number of studies define the change rate by considering the total quantity of new urban cells divided evenly over the number of time steps (e.g. Mustafa et al., 2014; Poelmans and Van Rompaey, 2010). However, one could ask why the number of changed cells should be the same each year instead of, for example, being higher during earlier stages of development so that the more attractive development sites get developed earlier. We examine seven cases in which the number of changed cells at the first time-step are +/- 25%, +/- 10%, +/- 5% and +/- 0% of the number of changed cells in case of equal change rate per time-step. For instance, in case of +25%, the model converts 5018 non-urban cells in the year 1991 (time-step 1), comparing to 4015 cells in case of equal change rate per time-step, and decreases this number linearly till the year 2000. The results reveal that in all modeling approaches, different change rates produce almost the same results as an equal change quantity per time step. Figure 3 illustrates the CLC accuracy rates for different change rates per time step for HUEM as an example.

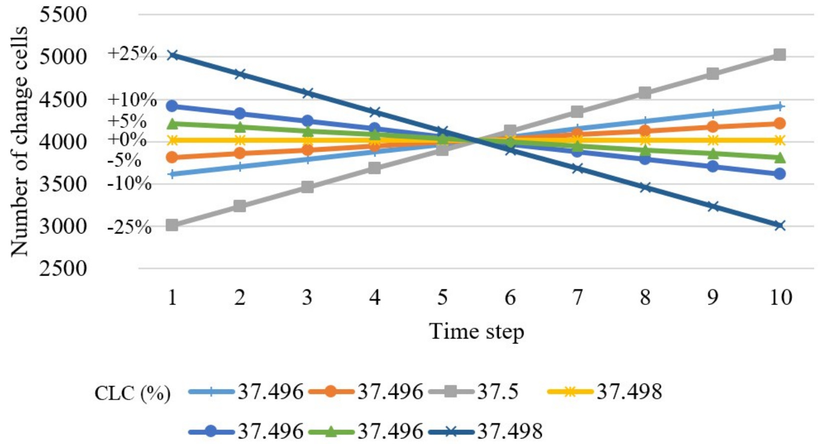


Fig. 3. Different change rates per time step (from +25% to -25% of the equal change rate per time-step)

Table 4. CTC agreement (%)between 1990 and 2000 (simulation vs. actual); and for a number of previous studies.

	Overall (all urban cells)	New urban cells
HUEM	91.08	37.50
CA-Logit	90.64	34.44
CA	90.18	31.25
Logit	89.01	22.97
Wang et al. (2013)	-	16.40
Poelmans and Van Rompaey (2009)	93.99	-
Liu et al. (2008)	78.30	-
Yang et al. (2008)	71.09	-
Jantz et al. (2003)	93.1	19
Wu (2002)	76.6	-

The outcome of each model is assessed under the same conditions in order to assess the performance of each model. Table 4 gives the CTC agreements for all simulations. The CTC of HUEM of the entire urban cells is 91%. This high CTC agreement is a result of the persistence of cells, which were already urban in 1990. To have a more fair comparison of the real performance, we focus only on the newly developed urban cells between 1990 and 2000. The CTC agreements reveal that Logit model showed the lowest agreement rate. The case study described in this paper is based on a set of predicting variables without any insights into the urbanization driving factors in the Wallonia. That could result in underestimating the process of urban development in the Wallonia through Logit analysis. The performance of CA model is better than Logit model, which is against expectations. Generally, CA models are only able to capture the part of the processes that govern urban expansion, while Logit models are better able to capture the full complexity of the urban expansion processes (Verburg and Overmars, 2007). A possible explanation for the results in our paper is the fact that CA calibration is based on the real land-uses of 1990 and 2000, which implicitly considers the urbanization driving factors.

The CTC results, for new urban cells, of all simulations are somewhat poor. It is common for urban expansion models, to have a low accuracy rate due to the complexity of the urban environment. Table 4 presents results of a number of developed urban expansion models. Surely, the results of other studies listed in table 4 are not conducted for our study area and cannot be directly compared because they are dependent on the purpose of the model, the model context, and the performance criteria. However, table 4 could roughly indicate the common accuracy rates in the urban expansion modeling domain.

Many urban expansion models have employed spatial metrics to analyze their results (e.g. García et al., 2011; Liu et al., 2008; Mustafa et al., 2014). We analyze the spatial pattern of different simulations focusing on landscape compactness and complexity. Fig. 4 indicates that HUEM performs well in terms of landscape structural conformity. The fragmentation rate (the number of patches (NP) and the mean patch area (MPA)) in HUEM simulation is close to the reality. CA-Logit and CA show moderate and high fragmentation rate comparing to the reality respectively (higher NP and lower MPA). Contrary, Logit model shows a very low rate of fragmentation (lower NP and higher MPA).

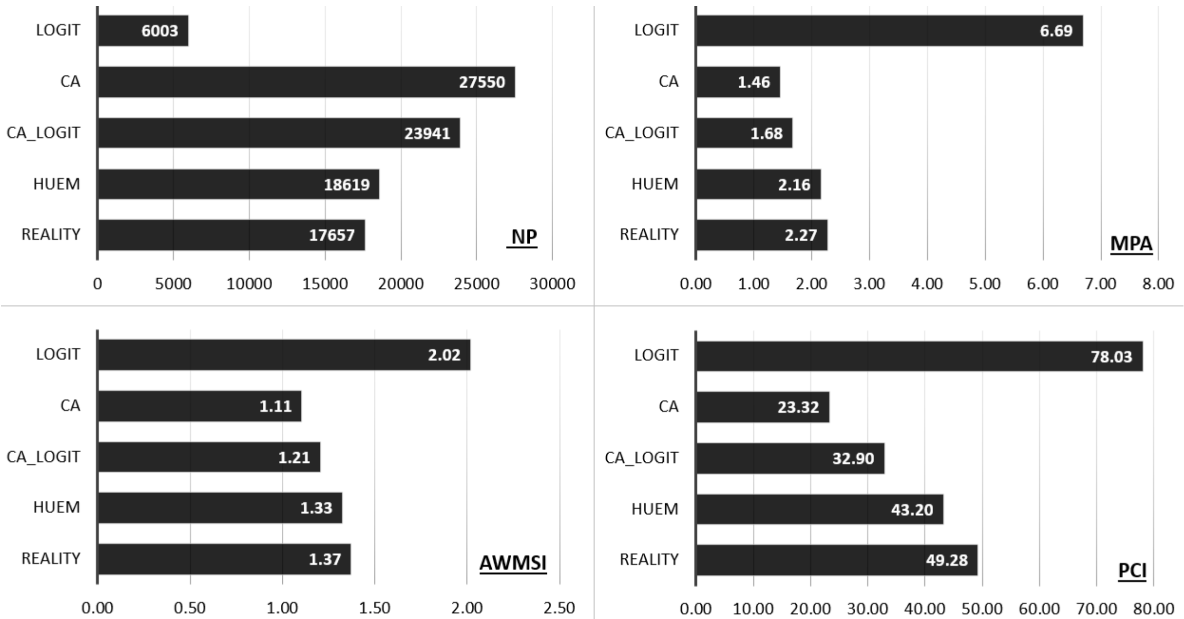


Fig. 4. Spatial matrices outcomes in Wallonia (simulation VS actual changes).

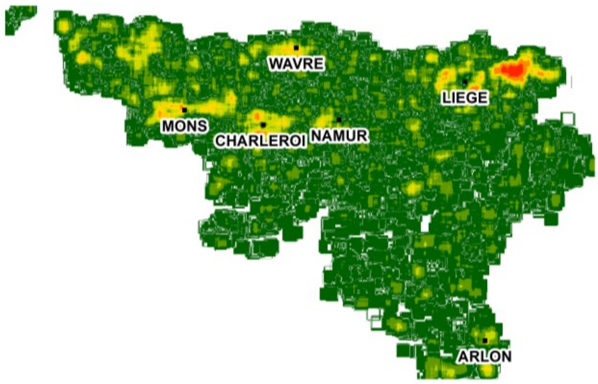
An area-weighted mean shape index (AWMSI) value of 1 represents a perfect regular shape (i.e. rectangle). HUEM generates urban patches close to the actual urban patches between 1990 and 2000 in terms of complexity. Both CA-Logit and CA have a rate of complexity smaller than the reality. Logit model presents a too high level of complexity.

In Logit model, according to the patch cohesion index (PCI), about 78% of the simulated cells are confined in patches, which results in a highly cohesive urban pattern. HUEM simulation presents a cohesion rate close to the reality. On the other hand, CA-Logit and CA generate low cohesive urban patterns in comparison to HUEM.

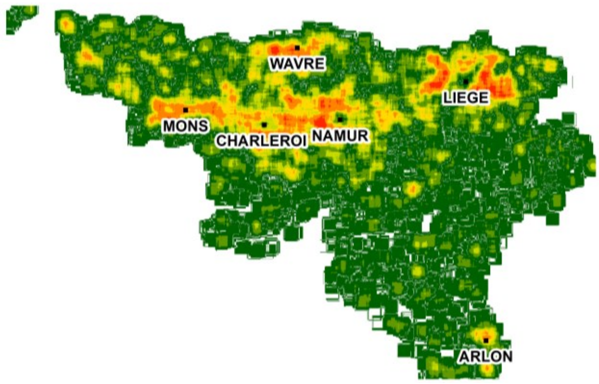
The analysis of landscape structural indices highlights the role of zoning status in Logit model. According to section 3.3, the location of new urban cells is strongly influenced by the zoning status. That means the new urban cells tend to be allocated within the permitted urban zones. Consequently, Logit model simulation pattern tends to be less fragmented, highly complex and very cohesive.

In order to examine spatial variability of the differences between actual urban change pattern and simulated patterns, a series of moving windows each sized 50x50 cells are used to calculate the landscape indices along with the abstract indices presented in Fig. 4. The results show that the absolute errors between actual change pattern and simulated patterns vary over space as Fig. 5. HUEM also produces areas with zero errors larger than other models. Fig. 5 demonstrates the absolute differences between simulated change patterns and the real one.

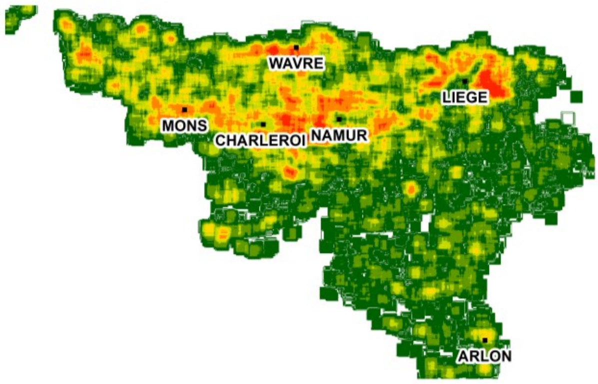
Fig. 6 shows allocation misclassification of the new urban cells between 1990 and 2000 in Namur metropolitan area, as an example.



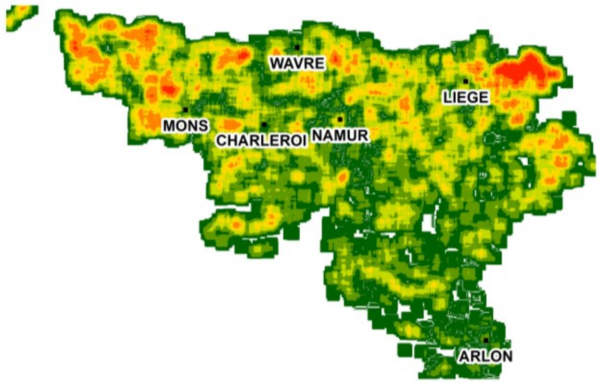
NP (HUEM)



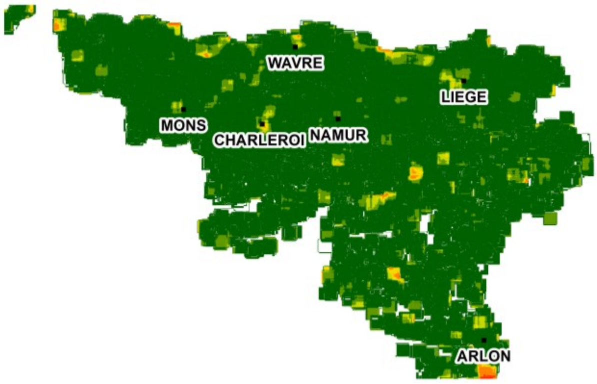
NP (CA_Logit)



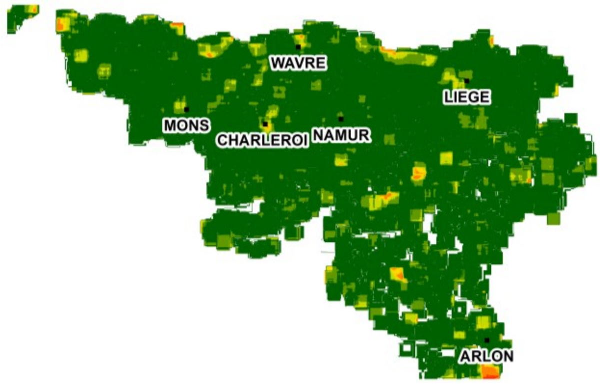
NP (CA)



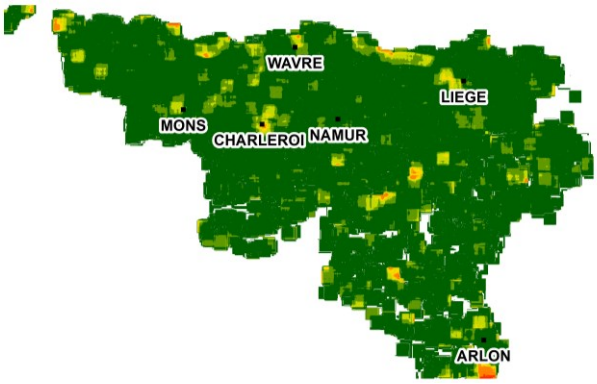
NP (Logit)



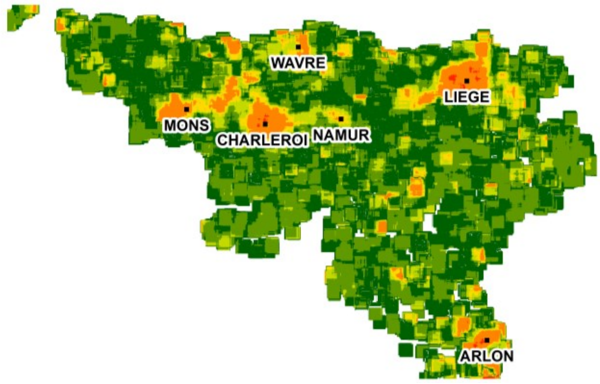
MPA (HUEM)



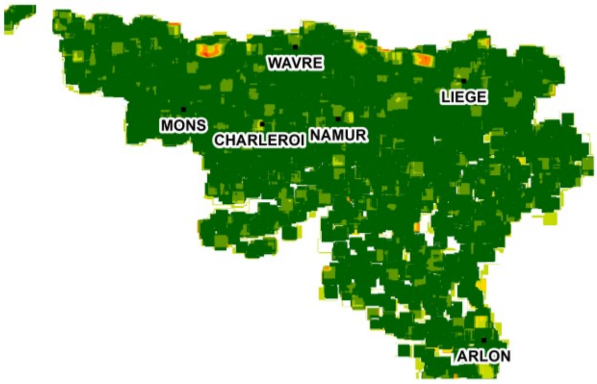
MPA (CA_Logit)



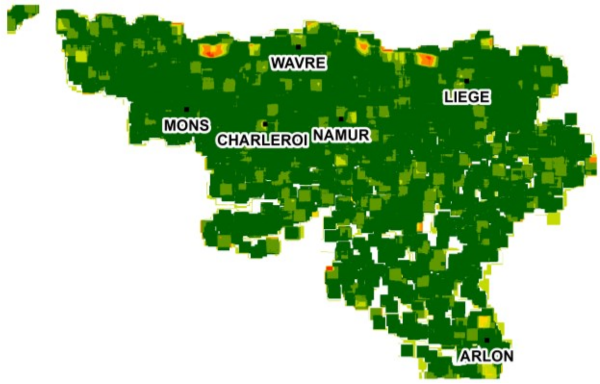
MPA (CA)



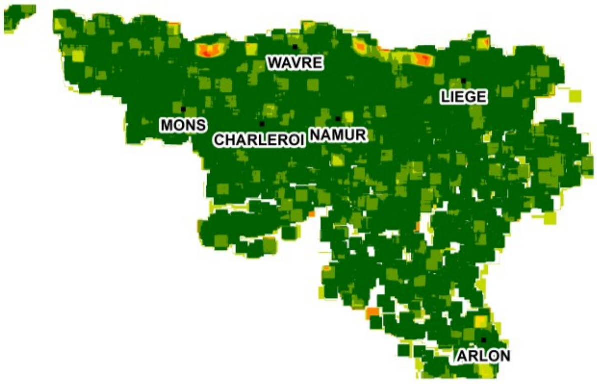
MPA (Logit)



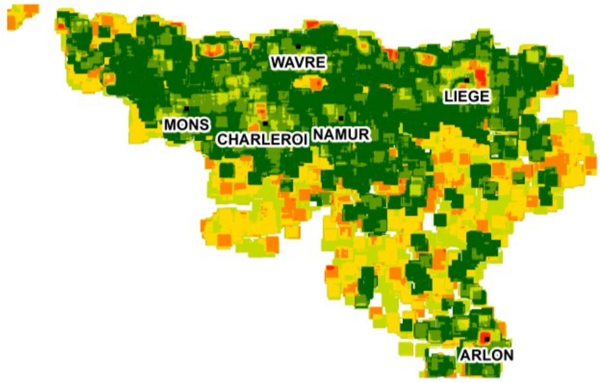
AWMSI (HUEM)



AWMSI (CA_Logit)



AWMSI (CA)



AWMSI (Logit)

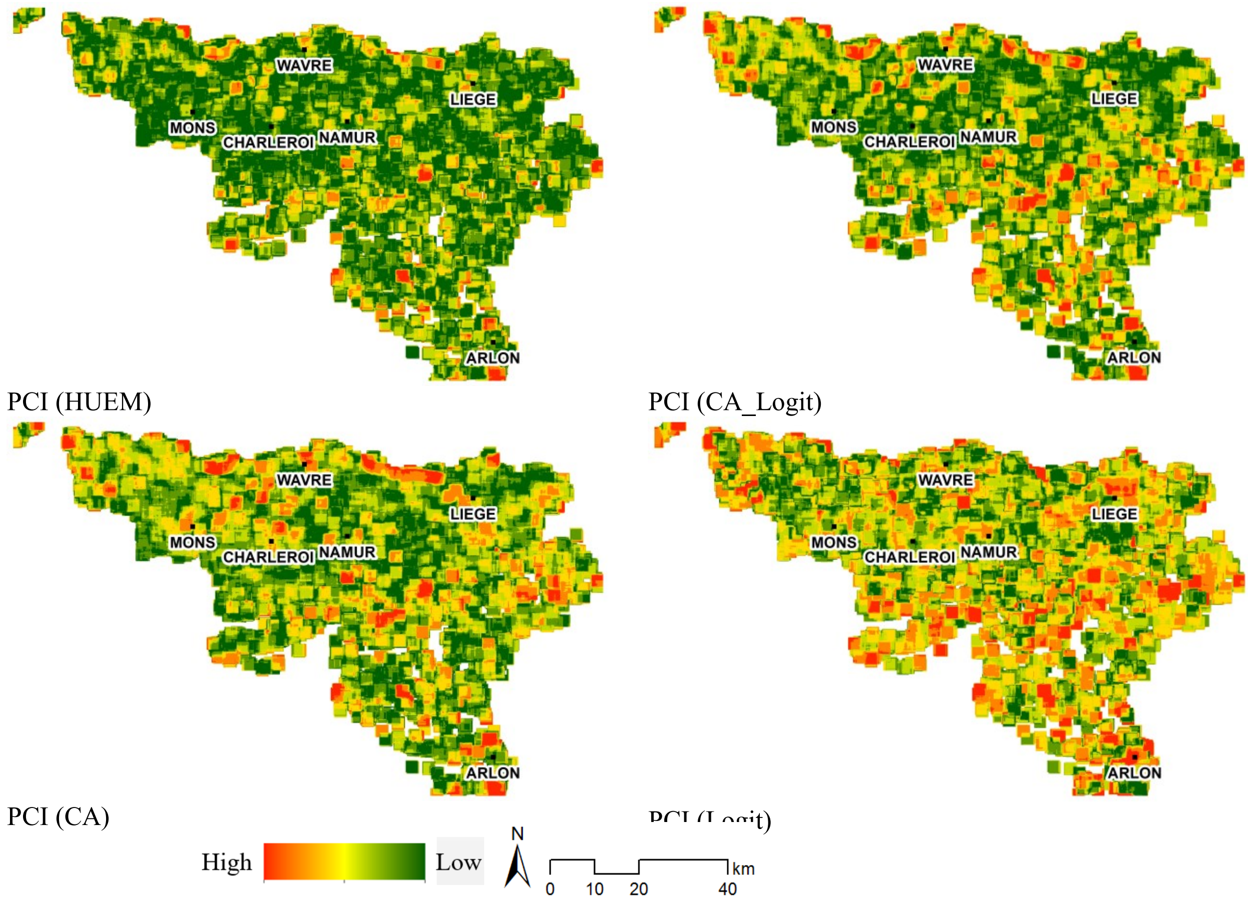


Fig. 5. The absolute differences between simulated change patterns and the real one based on a series of moving windows each sized 50x50 cells.

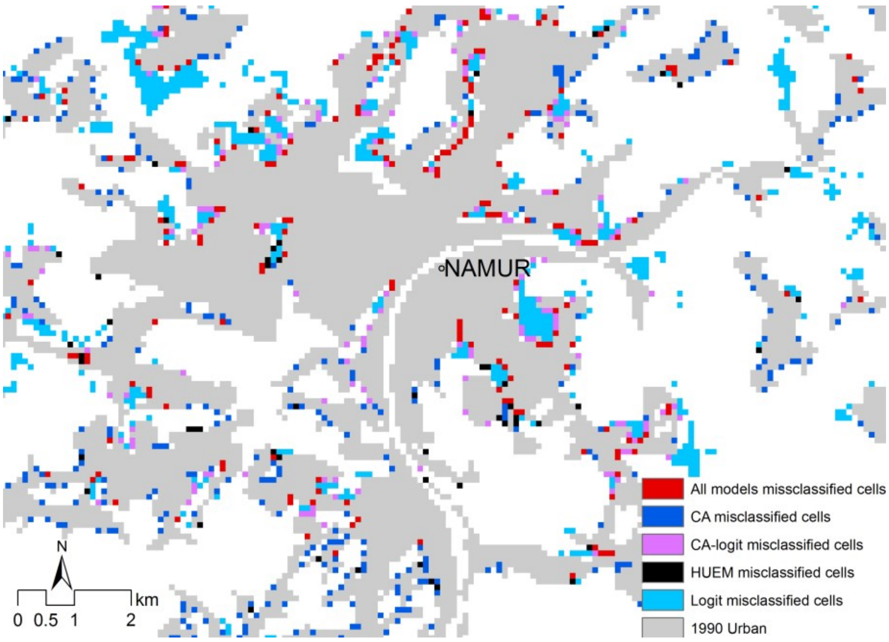


Fig. 6. Allocation misclassification of the new urban cells between 1990 and 2000 in Namur metropolitan area.

4. Conclusions

The expansion of urban areas is a global issue affecting water quality degradation, air pollution, socio-economic disparities, etc. Thus, there is a need to monitor urban expansions to support efficient planning visions for judicious use of natural resources and environment protection. The present study attempts to advance the applications of urban expansion modeling.

This paper presents a new model, named HUEM, to simulate future urban expansion. The model is based on an integrated approach that combines logistic regression, cellular automata, and agent-based approaches. The model has advantages in dealing with complex relationships among many of spatial variables, as well as stakeholders in the urban environment, which can capture the complexity of urban system better than traditional urban expansion models. Nonetheless, HUEM cannot capture all urbanization drivers and dynamics.

HUEM is successfully applied to Wallonia, Belgium to simulate the known urban expansion from 1990 to 2000. It is assessed using the cell-to-cell location agreement and landscape indices. In addition, HUEM is compared with typical spatial urban expansion models including Logit, CA and CA-Logit. Logit models are useful in explaining and determining the most important urbanization drivers because of the ability to consider several geophysical, socioeconomic and policy factors. However, Logit models are static and therefore they are not able to simulate self-organization of urban system over time (Poelmans and Van Rompaey, 2010). By contrast, CA models are dynamic and able to simulate self-organizing urban system by considering local neighboring interactions at each time step of the simulation. This study shows that CA model produces a better result than Logit model. Combining CA and Logit improves the results. This is in line with (Poelmans and Van Rompaey, 2010; Wu, 2002) who claimed that combining CA and Logit model produces a better result. HUEM integrates human behavior into a spatial CA-Logit model by considering interactions of various stakeholders who have contradictory values and priorities. The findings of this study confirm that these interactions, which are addressed by agents, can provide a better understanding, analysis, and forecasts of the future urban expansion.

The calibration of HUEM model is an automatic process based on Logit and genetic algorithm which makes the model generic and can be applied to other case studies. In this case, an explicit investigation of the transferability of the model parameters is an interesting direction for further research. Logit considers 1000 different sets of random samples. Each set represents different agents' responses. In addition, GA is a population-based algorithm implying that it has a multiple start research points. This nature of GA allows the optimization process maintain a population of possible solutions, which resulted in obtaining a number of best solutions (Pérez et al., 2003). Considering a series of possible agents' behavior is important in handling modifications of those behaviors over time. This is an essential feature for developing a methodology that will address the influence of uncertainty about future behaviors in our model.

Currently, the number of agents included in HUEM has been limited to three categories: urban developers, farmers, and planning permission authority. It might be interesting to include more agents, for instance, urban developers can be re-categorized into two different types of agents, namely households, and developers. This would require a better understanding of the settlement preferences of each of these agents in the specific case study, which is outside of the scope of this paper.

Acknowledgments: The research was funded through the ARC grant for Concerted Research Actions and through the Special Fund for Research, both financed by the Wallonia-Brussels Federation.

References

Al-Ahmadi, K., See, L., Heppenstall, A., Hogg, J., 2009. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. Ecol. Complex. 6, 80–101. doi:10.1016/j.ecocom.2008.09.004

Allison, P.D., 1999. Logistic Regression Using the SAS System: Theory and Application. WA, Cary, N.C.

Antrop, M., 2004. Landscape change and the urbanization process in Europe. Landsc. Urban Plan., Development of European Landscapes 67, 9–26. doi:10.1016/S0169-2046(03)00026-4

Batty, M., 2008. The Size, Scale, and Shape of Cities. Science 319, 769–771. doi:10.1126/science.1151419

Batty, M., 2007. Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals. MIT Press.

Beckers, A., Dewals, B., Erpicum, S., Dujardin, S., Detrembleur, S., Teller, J., Pirotton, M., Archambeau, P., 2013. Contribution of land use changes to future flood damage along the river Meuse in the Walloon region. Nat Hazards Earth Syst Sci 13, 2301–2318. doi:10.5194/nhess-13-2301-2013

Belgian Federal Government, 2015. Population [WWW Document]. Stat. Belg. URL http://statbel.fgov.be/fr/modules/publications/statistiques/population/population_-_chiffres_population_1990-2010.jsp (accessed 4.19.15).

Bert, F.E., Podestá, G.P., Rovere, S.L., Menéndez, Á.N., North, M., Tatara, E., Laciana, C.E., Weber, E., Toranzo, F.R., 2011. An agent based model to simulate structural and land use changes in agricultural systems of the argentine pampas. Ecol. Model. 222, 3486–3499. doi:10.1016/j.ecolmodel.2011.08.007

Bičík, I., Jeleček, L., Štěpánek, V., 2001. Land-use changes and their social driving forces in Czechia in the 19th and 20th centuries. Land Use Policy 18, 65–73. doi:10.1016/S0264-8377(00)00047-8

Bürgi, M., Hersperger, A.M., Schneeberger, N., 2005. Driving forces of landscape change - current and new directions. Landsc. Ecol. 19, 857–868. doi:10.1007/s10980-005-0245-3

Cammerer, H., Thieken, A.H., Verburg, P.H., 2013. Spatio-temporal dynamics in the flood exposure due to land use changes in the Alpine Lech Valley in Tyrol (Austria). Nat. Hazards 68, 1243–1270. doi:10.1007/s11069-012-0280-8

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. Int. J. Geogr. Inf. Sci. 12, 699–714. doi:10.1080/136588198241617

Crk, T., Uriarte, M., Corsi, F., Flynn, D., 2009. Forest recovery in a tropical landscape: what is the relative importance of biophysical, socioeconomic, and landscape variables? Landsc. Ecol. 24, 629–642. doi:10.1007/s10980-009-9338-8

Dujardin, S., Boussauw, K., Brévers, F., Lambotte, J.-M., Teller, J., Witlox, F., 2012. Sustainability and change in the institutionalized commute in Belgium: Exploring regional differences. Appl. Geogr. 35, 95–103. doi:10.1016/j.apgeog.2012.05.006

Engelen, G., Poelmans, L., Uljee, I., Van der Meulen, M., 2016. A cellular automata based spatial optimisation model to delineate new locations for economic activity in Limburg Province, Belgium, in: Proceedings of CAMUSS 2016. Presented at the Second International Symposium On Cellular Automata Modeling For Urban And Spatial Systems, Québec, Canada.

García, A.M., Santé, I., Boullón, M., Crecente, R., 2013. Calibration of an urban cellular automaton model by using statistical techniques and a genetic algorithm. Application to a small urban settlement of NW Spain. Int. J. Geogr. Inf. Sci. 27, 1593–1611. doi:10.1080/13658816.2012.762454

García, A.M., Santé, I., Crecente, R., Miranda, D., 2011. An analysis of the effect of the stochastic component of urban cellular automata models. Comput. Environ. Urban Syst. 35, 289–296. doi:10.1016/j.compenvurbsys.2010.11.001

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T., Hokao, K., 2011. Modeling urban land use change by the integration of cellular automaton and Markov model. Ecol. Model. 222, 3761–3772. doi:10.1016/j.ecolmodel.2011.09.009

Haggert, B.A., 1995. Review of Land Use and the Causes of Global Warming. Trans. Inst. Br. Geogr. 20, 518–520. doi:10.2307/622983

He, C., Okada, N., Zhang, Q., Shi, P., Li, J., 2008. Modelling dynamic urban expansion processes incorporating a potential model with cellular automata. Landsc. Urban Plan. 86, 79–91. doi:10.1016/j.landurbplan.2007.12.010

Holland, J.H., 1975. Adaptation in natural and artificial systems. U Michigan Press, Oxford, England.

Hosseinali, F., Alesheikh, A.A., Nourian, F., 2013. Agent-based modeling of urban land-use development, case study: Simulating future scenarios of Qazvin city. Cities 31, 105–113. doi:10.1016/j.cities.2012.09.002

Hu, Z., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. Comput. Environ. Urban Syst. 31, 667–688. doi:10.1016/j.compenvurbsys.2006.11.001

Jantz, C.A., Goetz, S.J., Shelley, M.K., 2003. Using the Sleuth Urban Growth Model to Simulate the Impacts of Future Policy Scenarios on Urban Land Use in the Baltimore-Washington Metropolitan Area. Environ. Plan. B Plan. Des. 31, 251–271. doi:10.1068/b2983

Jr, D.W.H., Lemeshow, S., 2004. Applied Logistic Regression. John Wiley & Sons.

Li, X., Zhou, W., Ouyang, Z., 2013. Forty years of urban expansion in Beijing: What is the relative importance of physical, socioeconomic, and neighborhood factors? Appl. Geogr. 38, 1–10. doi:10.1016/j.apgeog.2012.11.004

Lin, Y., Deng, X., Li, X., Ma, E., 2014. Comparison of multinomial logistic regression and logistic regression: which is more efficient in allocating land use? Front. Earth Sci. 1–12. doi:10.1007/s11707-014-0426-y

Liu, X., Li, X., Shi, X., Wu, S., Liu, T., 2008. Simulating complex urban development using kernel-based non-linear cellular automata. Ecol. Model. 211, 169–181. doi:10.1016/j.ecolmodel.2007.08.024

Matthews, R.B., Gilbert, N.G., Roach, A., Polhill, J.G., Gotts, N.M., 2007. Agent-based land-use models: a review of applications. Landsc. Ecol. 22, 1447–1459.

Miller, B.L., Miller, B.L., Goldberg, D.E., Goldberg, D.E., 1995. Genetic Algorithms, Tournament Selection, and the Effects of Noise. Complex Syst. 9, 193–212.

Mitsova, D., Shuster, W., Wang, X., 2011. A cellular automata model of land cover change to integrate urban growth with open space conservation. Landsc. Urban Plan. 99, 141–153. doi:10.1016/j.landurbplan.2010.10.001

Montgomery, D.C., Runger, G.C., 2003. Applied Statistics and Probability for Engineers, Fourth. ed. John Wiley & Sons, New York.

Mustafa, A., Bruwier, M., Teller, J., Archambeau, P., Erpicum, S., Pirotton, M., Dewals, B., 2016. Impacts of urban expansion on future flood damage: A case study in the River Meuse basin, Belgium, in: Sustainable Hydraulics in the Era of Global Change. Taylor & Francis Group.

Mustafa, A., Cools, M., Saadi, I., Teller, J., 2015. Urban Development as a Continuum: A Multinomial Logistic Regression Approach, in: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (Eds.), Computational Science and Its Applications -- ICCSA 2015, Lecture Notes in Computer Science. Springer International Publishing, pp. 729–744.

Mustafa, A., Saadi, I., Cools, M., Teller, J., 2014. Measuring the Effect of Stochastic Perturbation Component in Cellular Automata Urban Growth Model. Procedia Environ. Sci., 12th International Conference on Design and Decision Support Systems in Architecture and Urban Planning, DDSS 2014 22, 156–168. doi:10.1016/j.proenv.2014.11.016

Parker, D.C., Meretsky, V., 2004. Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. Agric. Ecosyst. Environ. 101, 233–250. doi:10.1016/j.agee.2003.09.007

Pérez, E., Herrera, F., Hernández, C., 2003. Finding multiple solutions in job shop scheduling by niching genetic algorithms. J. Intell. Manuf. 14, 323–339. doi:10.1023/A:1024649709582

Poelmans, L., Engelen, G., Uljee, I., van der Meulen, M., 2013. RuBeLim – Ruimte voor bedrijvigheid in Limburg (No. 2013/RMA/R/0255).

Poelmans, L., Van Rompaey, A., 2010. Complexity and performance of urban expansion models. Comput. Environ. Urban Syst. 34, 17–27. doi:10.1016/j.compenvurbsys.2009.06.001

Poelmans, L., Van Rompaey, A., 2009. Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A case study in the Flanders–Brussels region. Landsc. Urban Plan. 93, 10–19. doi:10.1016/j.landurbplan.2009.05.018

Poelmans, L., Van Rompaey, A., Batelaan, O., 2010. Coupling urban expansion models and hydrological models: How important are spatial patterns? Land Use Policy 27, 965–975. doi:10.1016/j.landusepol.2009.12.010

Puertas, O.L., Henríquez, C., Meza, F.J., 2014. Assessing spatial dynamics of urban growth using an integrated land use model. Application in Santiago Metropolitan Area, 2010–2045. Land Use Policy 38, 415–425. doi:10.1016/j.landusepol.2013.11.024

Ralha, C.G., Abreu, C.G., Coelho, C.G.C., Zaghetto, A., Macchiavello, B., Machado, R.B., 2013. A multi-agent model system for land-use change simulation. Environ. Model. Softw. 42, 30–46. doi:10.1016/j.envsoft.2012.12.003

Sang, L., Zhang, C., Yang, J., Zhu, D., Yun, W., 2011. Simulation of land use spatial pattern of towns and villages based on CA–Markov model. Math. Comput. Model., Mathematical and Computer Modeling in agriculture (CCTA 2010) 54, 938–943. doi:10.1016/j.mcm.2010.11.019

Shafizadeh Moghadam, H., Helbich, M., 2013. Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model. Appl. Geogr. 40, 140–149. doi:10.1016/j.apgeog.2013.01.009

Shan, J., Alkheder, S., Wang, J., 2008. Genetic Algorithms for the Calibration of Cellular Automata Urban Growth Modeling. Photogramm. Eng. Remote Sens. 74, 1267–1277. doi:10.14358/PERS.74.10.1267

Valbuena, D., Verburg, P.H., Bregt, A.K., 2008. A method to define a typology for agent-based analysis in regional land-use research. Agric. Ecosyst. Environ. 128, 27–36. doi:10.1016/j.agee.2008.04.015

Verburg, P.H., Overmars, K.P., 2007. Dynamic Simulation of Land-Use Change Trajectories with the Clue-S Model, in: Koomen, E., Stillwell, J., Bakema, A., Scholten, H.J. (Eds.), Modelling Land-Use Change, The GeoJournal Library. Springer Netherlands, pp. 321–337. doi:10.1007/978-1-4020-5648-2_18

Verburg, Peter H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004. Land use change modelling: current practice and research priorities. GeoJournal 61, 309–324. doi:10.1007/s10708-004-4946-y

Verburg, Peter H, van Eck, J.R.R., de Nijs, T.C.M., Dijst, M.J., Schot, P., 2004. Determinants of Land-Use Change Patterns in the Netherlands. Environ. Plan. B Plan. Des. 31, 125–150. doi:10.1068/b307

Verhetsel, A., Thomas, I., Beelen, M., 2010. Commuting in Belgian metropolitan areas: The power of the Alonso-Muth model. J. Transp. Land Use 2. doi:10.5198/jtlu.v2i3.19

Vermeiren, K., Van Rompaey, A., Loopmans, M., Serwajja, E., Mukwaya, P., 2012. Urban growth of Kampala, Uganda: Pattern analysis and scenario development. Landsc. Urban Plan. 106, 199–206. doi:10.1016/j.landurbplan.2012.03.006

Wang, H., He, S., Liu, X., Dai, L., Pan, P., Hong, S., Zhang, W., 2013. Simulating urban expansion using a cloud-based cellular automata model: A case study of Jiangxia, Wuhan, China. Landsc. Urban Plan. 110, 99–112. doi:10.1016/j.landurbplan.2012.10.016

White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Comput. Environ. Urban Syst. 24, 383–400. doi:10.1016/S0198-9715(00)00012-0

White, R., Engelen, G., Uljee, I., 2015. The Cellular Automaton Eats the Regions: Unified Modeling of Activities and Land Use in a Variable Grid Cellular Automaton, in: Modeling Cities and Regions As Complex Systems: From Theory to Planning Applications. The MIT Press, Cambridge , Massachusetts.

Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. Int. J. Geogr. Inf. Sci. 16, 795–818. doi:10.1080/13658810210157769

Yang, Q., Li, X., Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. Comput. Geosci. 34, 592–602. doi:10.1016/j.cageo.2007.08.003

Yang, X., Zheng, X.-Q., Chen, R., 2014. A land use change model: Integrating landscape pattern indexes and Markov-CA. Ecol. Model. 283, 1–7. doi:10.1016/j.ecolmodel.2014.03.011

Zhang, H., Zeng, Y., Bian, L., Yu, X., 2010. Modelling urban expansion using a multi agent-based model in the city of Changsha. J. Geogr. Sci. 20, 540–556. doi:10.1007/s11442-010-0540-z