Comparing support vector machines with logistic regression for calibrating cellular automata land use change models

Ahmed Mustafa, Andreas Rienow, Ismaïl Saadi, Mario Cools & Jacques Teller

Published online: 06 Mar 2018.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS Check for updates

Comparing support vector machines with logistic regression for calibrating cellular automata land use change models

Ahmed Mustafa [a,b], Andreas Rienow [b], Ismaïl Saadi [a], Mario Cools [a] and Jacques Teller [a]

aLEMA, Urban and Environmental Engineering Dept., Liège University, Liège, Belgium; bGeomatics Research Group, Ruhr-University Bochum, Bochum, Germany

ABSTRACT

Land use change models enable the exploration of the drivers and consequences of land use dynamics. A broad array of modeling approaches are available and each type has certain advantages and disadvantages depending on the objective of the research. This paper presents an approach combining cellular automata (CA) model and supported vector machines (SVMs) for modeling urban land use change in Wallonia (Belgium) between 2000 and 2010. The main objective of this study is to compare the accuracy of allocating new land use transitions based on CA-SVMs approach with conventional coupled logistic regression method (logit) and CA (CA-logit). Both approaches are used to calibrate the CA transition rules. Various geophysical and proximity factors are considered as urban expansion driving forces. Relative operating characteristic and a fuzzy map comparison are employed to evaluate the performance of the model. The evaluation processes highlight that the allocation ability of CA-SVMs slightly outperforms CA-logit approach. The paper also reveals that the major urban expansion determinant is urban road infrastructure.

Introduction

Several land use change models are developed to explore the drivers of land use/land cover change and to simulate future land use patterns (e.g. Hallowell & Baran, 2013; Kryvobokov, Mercier, Bonnafous, & Bouf, 2015; Puertas, Henríquez, & Meza, 2014; Wang & Maduako, 2018). The existing modeling approaches generally adopt cellular automata (CA), Agent-based (AB), urban-economic discrete-choice and/or statistical models. CA modeling framework (e.g. Batty, Xie, & Sun, 1999; Troisi, 2015) is particularly useful in encompassing spatial autocorrelation effects by considering local neighborhood dynamics. AB models (e.g. Mustafa et al., 2017; Zhang, Zeng, Bian, & Yu, 2010) examine agents as goal-oriented entities capable of responding to their environment and taking independent actions, where these agents may represent individuals, institutions etc. In AB models, solutions have been designed to explore the emergent properties of systems with relatively simple behavioral rules representing individual agents. The urban-economic discrete choice models emerged from an integration of urban economic analysis with agents choices in the urban environment. UrbanSim is an example application of this approach (e.g. Kryvobokov et al., 2015; Waddell, 2002). This application works with agents and integrates works with agents and integrates discrete choice approach

and statistical methods to estimate model parameters (Ševčíková, Raftery, & Waddell, 2007). Another approach relies on statistical methods (e. g. Mustafa et al., 2018b; Hu & Lo, 2007; Vermeiren, Van Rompaey, Loopmans, Serwajja, & Mukwaya, 2012) that help identify drivers behind land use change dynamics.

Among the abovementioned approaches, CA has received considerable attention due to its simplicity, transparency and its ability to represent the evolution of land use, particularly urban expansion (Clarke & Gaydos, 1998; Troisi, 2015). Aburas et al. (2016) and Santé, García, Miranda, and Crecente (2010) have reviewed CA models and have concluded that CA approach is one of the most appropriate techniques for simulating land use change. CA models focus on the simulation of spatial patterns by explicitly considering the immediate neighbors of each landscape unit, that is, e.g. cell, rather than on the interpretation of driving factors of the land use change. Due to this limitation of CA models, huge research effort has been made in order to improve CA modeling structure by incorporating a variety of driving forces into the model (e.g. Jokar Arsanjani, Helbich, Kainz, & Darvishi Boloorani, 2013; Munshi, Zuidgeest, Brussel, & Van Maarseveen, 2014). The key challenge in such approach is the calibration of the transition rules. Recently, logistic regression method (logit) has become one of the

---

CONTACT Ahmed MUSTAFA ✉ a.mustafa@uliege.be 🖃 LEMA, Urban and Environmental Engineering Dept., Liège University, Belgium; Geomatics Research Group, Ruhr-University Bochum, Bochum, Germany

most popular techniques for calibrating CA models (e.g. Chen, Li, Liu, & Ai, 2014; Munshi et al., 2014; Poelmans & Van Rompaey, 2010; Wu, 2002). Logit requires less demand for computational resources and can include several driving forces. In addition, it measures the relative contribution of each driving forces which is of great value for policymakers. Despite these strengths, logit assumes that the occurrence probability is linearly and additively related to the independent variables on a logistic scale (Cheng & Masser, 2003). If this assumption cannot be satisfied, the performance of the model may decline.

Proposed by Vapnik et al. in the 1990s (Boser, Guyon, & Vapnik, 1992; Schölkopf, Burges, & Vapnik, 1996), the support vector machines (SVMs) is a supervised algorithm that can model nonlinearity relationships (Martens, Baesens, Van, & Vanthienen, 2007). A number of researchers argue that SVMs are an effective method for defining transition rules for CA models, owing to their ability to model nonlinear relationships with good generalization performance (Rienow & Goetzke, 2015; Yang, Li, & Shi, 2008). The basic idea of SVMs algorithm is quite different from that of logit method, while logit employs a maximum likelihood algorithm, SVMs, in contrast, tries to project input vectors on a binary (i.e. two classes) hyperplane that is linearly separable. If the linear separation is not possible, SVMs algorithm is still able to find a separation boundary for classification by a curved (nonlinear) separation. In the SVMs, nonlinear solutions can be found by increasing the dimensionality of the input variable space (Verplancke et al., 2008). Being able to recognize patterns reliably, the SVM algorithms are applied for regression challenges like the prediction of hospital mortality (e.g. Verplancke et al., 2008) or financial time series (e.g. Van Gestel et al., 2001). These techniques are also heavily used to solve classification problems, for example, in the context of satellite imagery (Raczko & Zagajewski, 2017; Vogel, 2013; Waske, Linden, Benediktsson, Rabe, & Hostert, 2010).

There are limited research efforts reported on performance differences between SVMs and logit within land use change domain. Huang, Xie, and Tay (2010) compared the performance of SVMs to logit without integration with CA. Rienow and Goetzke (2015) and Yang et al. (2008) compared CA-SVMs with CA-logit. However, both studies exhibit a stochastic disturbance term. Since the stochastic term is integrated into the model, the results may not demonstrate a fair comparison of the performance of both approaches. However, these studies concluded that SVMs outperformed logit. This paper contributes to the research efforts that examine the performance of CA-SVMs model and compare it with CA-logit model. In comparison with the previous work, a major differentiation of our work is comparing the performance of CA-logit model with CA-SVMs with and without introducing a stochastic term to get a more reliable comparison. This study separately introduces SVMs and logit as methods for defining the transition rules for CA model. Both approaches are developed and tested for Wallonia, southern Belgium as a case study. We simulate the spatio-temporal process of urban expansion from 2000 to 2010, using time steps of 1 year. Our model is a predictive model, which simulates future land use change based on the calibration results. Explorations of future land use change are important to define potential change areas. However, that is outside the scope of the present paper. The urban class in our model configuration consists of land that is covered by buildings and does not consider all other artificial uses such as transport infrastructure. The simulation outcomes are evaluated with the relative operating characteristic (ROC) (Aldrich & Nelson, 1984) and the fuzziness comparison index.

Materials and methods

Study area

Wallonia, Figure 1, encompasses the southern part of Belgium with a total area of 16,844 km^2. It comprises five provinces: Hainaut, Liège, Luxembourg, Namur and Walloon Brabant. The main urban areas are Charleroi, Liège, Mons and Namur. These urban areas are all characterized by a historical city-center, around which the urban development expanded (Mustafa, Saadi, Cools, & Teller, 2018c). The total population of Wallonia in 2010 was 3,498,384 inhabitants, corresponding to one-third of the Belgium population (Belgian Federal Government, 2013). Urban development in the Northern part of Wallonia is strongly influenced by the presence of Brussels especially in the province of Walloon Brabant. In the southernmost part of Wallonia, the presence of the city of Luxembourg affects urban development (Thomas, Frankhauser, & Biernacki, 2008).

Wallonia typifies a growing debate regarding the trade-offs between socioeconomic development and their impacts on the landscape. It is characterized by a strong urban sprawl and resulting landscape fragmentation (EEA, 2011). This, in turn, increases environmental impacts. In order to tackle those impacts, the authorities in Wallonia set a planning policy to reduce the conversion rate of non-urban to urban lands from 20 km^2/year to 12 km^2/year by 2020 and to 9 km^2/year by 2040 (SPW, 2013). Such policies require a holistic vision of the urban development process.
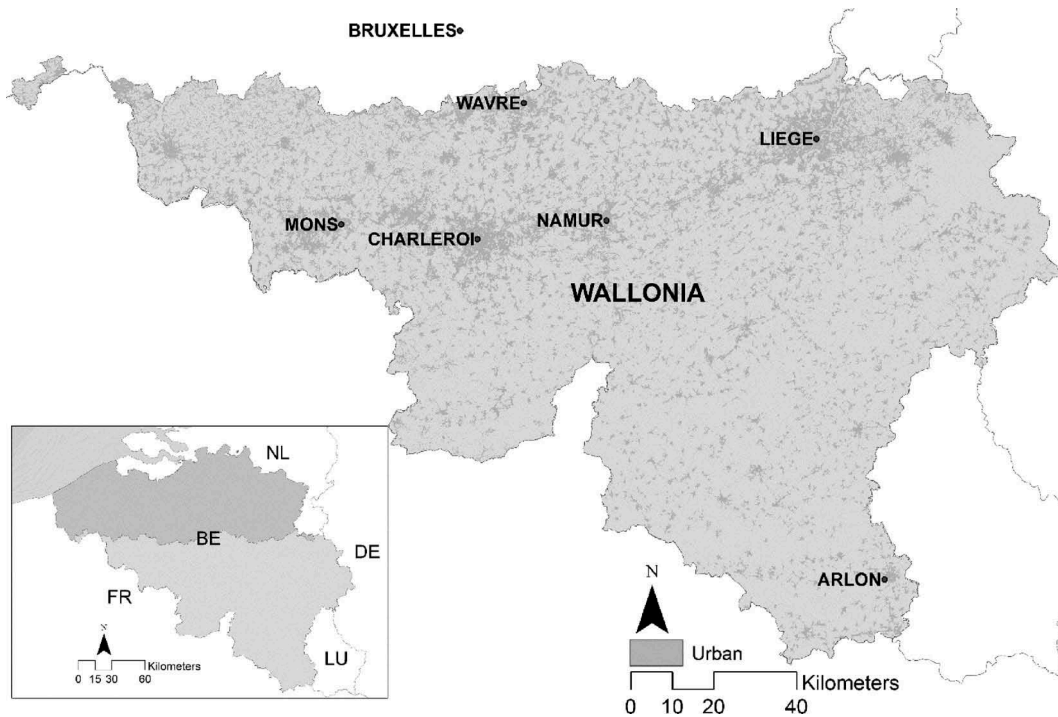
Figure 1. Study area.

Data

Belgian cadastral data (CAD) are used to prepare land use maps. CAD, made available by the Land Registry Administration of Belgium, is a vector data representing buildings as polygons. Each building comes with different attributes including its construction date. Using the construction date, two urban raster-grids were generated for 2000 and 2010. The vector data is rasterized at a fine cell dimension of 2 m × 2 m. The rasterized cells were then aggregated to obtain a 100 m × 100 m raster-grid. The aggregated data consider cells as urban, as soon as one 2 m × 2 m cell is built-up within its boundary. As a result, the amount of urban area might be overestimated. To overcome this problem, all aggregated cells with a density less than 25 of 2 m × 2 m cells were considered as non-urban cells. The threshold of 25 (representing a building of 100 m^2) corresponds to an average-sized residential building in Belgium (Tannier & Thomas, 2013). Aggregated urban lands are assigned a value of 1, while other land uses are assigned a value of 0.

Existing literature introduced a wide range of urbanization driving forces, including geophysical, proximity, policies and socioeconomic factors. However, the geophysical and proximity factors are included in most studies (Berberoğlu, Akın, & Clarke, 2016; Chen et al., 2014; Mustafa, Cools, Saadi, & Teller, 2017; Mustafa et al., 2018a). Based on the best available data, we select six factors related to proximities and geophysical aspects. Elevation and slope are introduced as geophysical drivers.

Proximity to highways, main roads, secondary roads and local roads are introduced as accessibility indicators. They act as proxies for socio-economic driving forces like market access in a "von Thünen" model (Verburg, Van Eck, De Nijs, Dijst, & Schot, 2004). The Navteq streets of 2002 are used to calculate Euclidean distances to the four road classes. Digital Elevation Model provided by the Belgian National Geographic Institute is used to calculate slope in percentage rise for each cell. All maps are created as raster grids with a resolution of 100 m × 100 m. The variance inflation factors (VIF) test has been performed to ensure that there is no multicollinearity between the selected driving forces. The driving forces show VIF values between 1.06 and 1.33, which means that there is no potential multicollinearity (Montgomery & Runger, 2003).

The model structure

This paper presents a CA land use change model with a focus on urban expansion process. Among other CA models, the model we propose has some overlaps with a previous scheme proposed in Iannone, Troisi, Guarnaccia, D'agostino, and Quartieri (2011), Iannone and Troisi (2013), and Troisi (2015) where a holistic urban potential-based approach has been introduced.

The model consists of two principal modules with distinct functions, namely a non-spatial demand module and a spatially explicit allocation module. The non-spatial module calculates the demand of

new urban cells at each time step at the aggregate level. Within the second module, these demands are translated into changing the state of a specific number of non-urban cells into urban ones at different locations over the study area. In order to draw attention to the allocation ability of the model, the demand module assumes that the amount of new urban cells is equal to that of the actual urban development occurring during the simulated period divided evenly by 10 (the number of years).

The allocation module is the key part of the model. Figure 2 highlights the module workflow. This module starts generating two urbanization probability maps based on logit and SVMs. This is done by associating 2000–2010 urban changes with the driving forces. The module then measures the potential for urban expansion on a yearly basis by considering the effects of the neighboring land uses using CA model. Finally, CA is coupled with logit and SVM approaches in which the potential for urban expansion was defined as follows (Feng et al., 2011; Wu, 2002):

$$Purbn_{ij}^t = PD_{ij}^t \times N_{ij}^t \times con(.) \qquad (1)$$

where $Purbn_{ij}^t$ is the urbanization potential of a cell ij at time t, PD_{ij}^t is the transition probability for a cell ij based on the driving forces, N_{ij}^t is the neighborhood potential for a cell ij based on the immediate

neighborhood interactions and $con(.)$ is the restrictive cases for urban development. In our case study, $con(.)$ is 0 if cell ij is occupied by water, defined by the official zoning plan, or 1 otherwise. The model then changes the cells with the highest $Purbn_{ij}^t$ scores to urban cells until meeting the required new urban cells, i.e. expansion demand. The PD_{ij}^t is calculated based on two different ways using logit (PD_{logit}) and SVMs (PD_{svm}).

The dependent variable for logit and SVMs is a binary map showing the spatial pattern of observed urban expansion between 2000 and 2010. A value of 1 in the map indicates that the non-urban cell has changed its land use to urban where a value of 0 means that the cell did not change its use. The independent variables are the selected urbanization driving forces. As the independent variables are measured in different units, we normalized them between 0 and 1. This is especially important for SVMs as the accuracy can severely deteriorate if the data are not normalized (Ben-Hur & Weston, 2010; Chang & Lin, 2001).

In order to minimize the potential effects of spatial autocorrelation on the logit results, both models were calibrated using a random sample (S) of 4000 cells with a minimum distance of 500 m between each cell within the sample, Figure 3. The same sample set is used on SVMs and logit. All existing urban cells in 2000 are excluded from the samples.

Definition of cell neighborhood

The value of the neighborhood potential, N_{ij}^t, is calculated as follows (Feng et al., 2011; Wu, 2002):

$$N_{ij}^t = \frac{\sum U}{n \times n - 1} \qquad (2)$$

where U is the number of urban cells among the Moore $n \times n$ neighborhood. The proper size of neighborhood is selected based on a sensitivity analysis of the model performance with different neighboring sizes ranging from 3×3 to 9×9.

Logistic regression

Logistic regression (logit) is an empirical modeling technique in which the selection of the independent variables is data-driven rather than knowledge-driven. Logit can readily identify the impact of independent variables and provides a degree of confidence regarding their contributions (Hu & Lo, 2007). This type of regression analysis is usually employed in estimating a model that defines the relationship between one or more independent variable(s) to a binary dependent variable. It considers the urbanization driving forces to be independent variables. Dependent
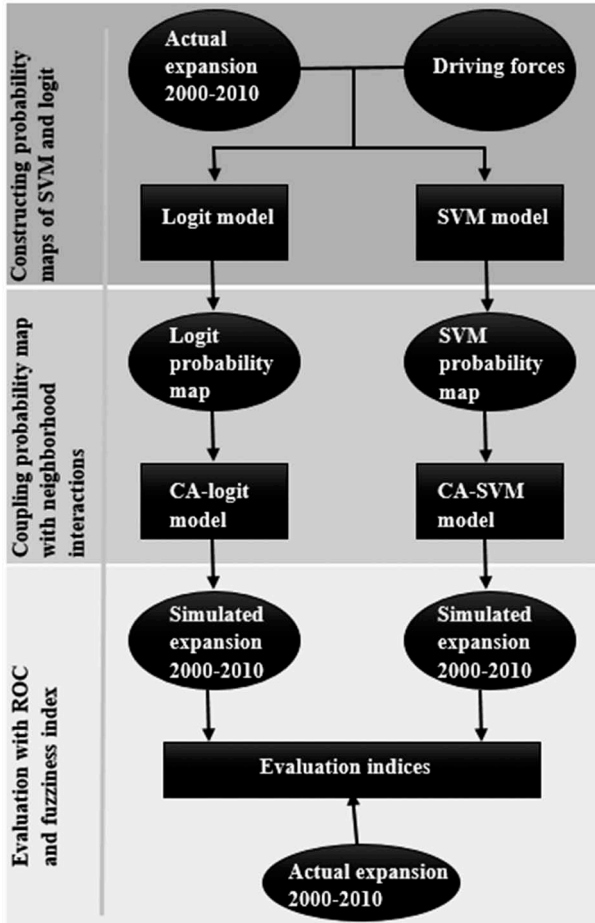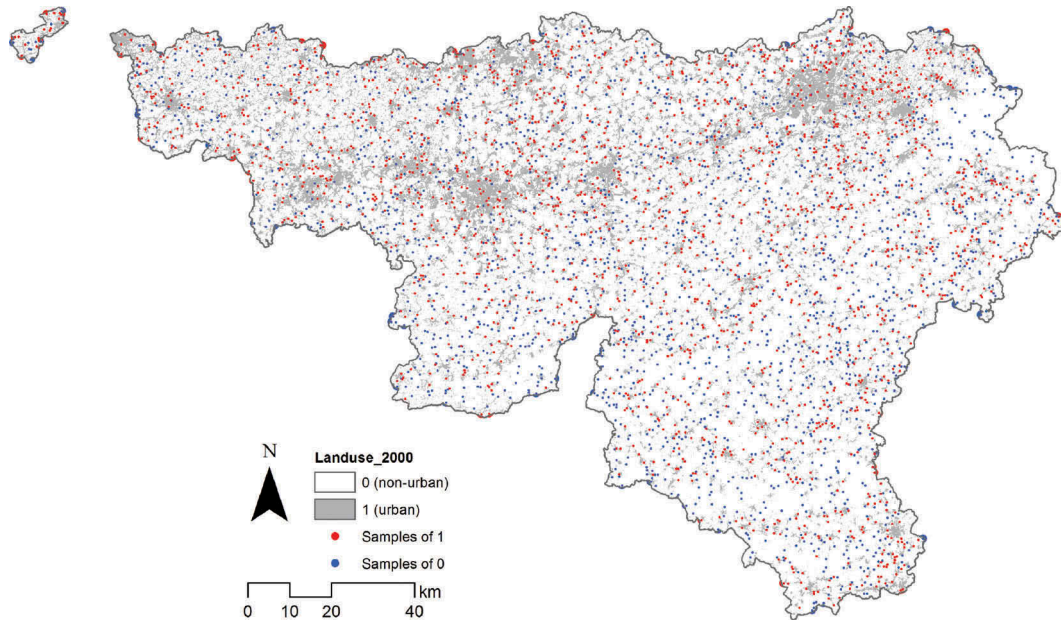
Figure 2. The model's flowchart.

Figure 3. The selected samples (2000 cells of 0 and 2000 cells of 1).

variable takes the values of 1 (positive response) and 0 (negative response) following the logistic curve. The logistic function can be estimated by means of the following equation:

$$PD_{logit} = P(Y = 1 | \chi_1, \chi_2, \ldots, \chi_n)$$
$$= \frac{\exp(\alpha + \beta_1\chi_1 + \beta_2\chi_2 + \ldots + \beta_n\chi_n)}{1 + \exp(\alpha + \beta_1\chi_1 + \beta_2\chi_2 + \ldots + \beta_n\chi_n)} \quad (3)$$

where PD_{logit} is the probability of a non-urban cell being urban, $P(Y = 1 | x_1, x_2, \ldots, x_n)$ the probability of the dependent variable Y being 1 given independent variables ($x_1, x_2, \ldots, x_n$), which can be either categorical or continuous, α is the intercept representing the value of Y when the values of the independent variables are zero and ($\beta_1, \beta_2, \ldots, \beta_n$) are the regression coefficients. The logit employs the procedure of the maximum likelihood (Pace & LeSage, 2002) to encounter the α and β.

Support vector machines

Along with artificial neural networks and genetic programming, SVM algorithms represent a new generation of machine learning algorithms. To put it simply, SVMs are a linear binary classifier that labels a sample of empirical data by constructing the optimal separating hyperplane. Traditional machine learning methods try to minimize the empirical training error so that they tend to overfit (Vapnik & Vapnik, 1998; Xie, 2006). They are strongly tailored to the training data, so transferring them to further data turns out to be difficult. Considering the principles of structural risk minimization (Vapnik, 1995; Vapnik & Vapnik, 1998), SVMs aim at minimizing the upper bound of the expected generalization error through maximizing the margin between the separating hyperplane and the data (Figure 4, left). The concept of margin plays a key role in SVM algorithm as it indicates the generalization capability of SVMs (Burges, 1998; Huang et al., 2010). The main
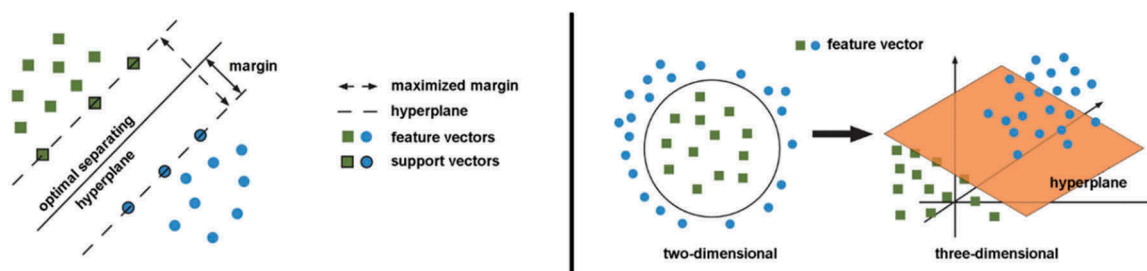

Figure 4. An optimal hyperplane constructed by separating the training data (left). Having a nonlinear classification problem, the input data is projected onto a higher-dimensional Hilbert space (right) (Vogel, 2013).

advantage of SVMs is the ability to transform the model to solve a nonlinear classification problem without any prior knowledge. The input vectors are re-projected to a higher-dimensional space in which they can be classified linearly using the so-called kernel trick (Eq.8–9) (Figure 4, right).

We need to find a hyperplane which separates the positive from the negative feature vectors. The separating hyperplane H can be parameterized linearly by w and b:

$$H : \langle w, x \rangle + b = 0 \qquad (4)$$

where w, element of R^d, is a normal to H, and b, element of R, the bias. In case of the linearly separable, SVMs can define two hyperplanes H_+ and H_- constructed by the closest positive and negative examples – the so-called support vectors:

$$\begin{aligned} H_+ &: \langle w, x \rangle + b = 1 \\ H_- &: \langle w, x \rangle + b = -1 \end{aligned} \qquad (5)$$

As H_+ and H have the same normal and no training points fall between them, they are parallel. The distance between the optimal separating hyperplane H_+ and H, resp. $H-$ and H, is $1/||W||$' where $||W||$ is the Euclidean norm of w. Thus, the margin between H_+ and $H-$ is $2/||W||$. The optimal separating hyperplane is found where the margin between H_+ and $H-$ is the largest and therefore $||W||$ has to be minimized. The outline of the constrained optimization problem is

$$\begin{aligned} &\min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \xi_i \text{ subject to } y_i(w, x_i + b) \\ &- 1 \\ &\geq 0 \text{ for } i = 1, \ldots, n \end{aligned} \qquad (6)$$

The constant C is called penalty parameter and ξ_i is a slack variable representing the error in the classification. The first part of Eq. 6 maximizes the margin between the two classes whereas the second part minimizes the classification error. The optimization problem is solved by formulating it in a dual form derived by constructing a Lagrange function according to the Karush–Kuhn–Tucker optimality condition (Burges, 1998). If the classification problem is not separable linearly, the data set has to be transferred or projected respectively into a higher dimension: the Hilbert space. It extends the methods of vector algebra from two-/three-dimensional spaces to spaces depicting any finite or infinite number of dimensions. By using the function ϕ with $d_1 < d_2$ the amount of possible linear separations is increased as follows:

$$\mathbb{R}^{d1} \rightarrow \mathbb{R}^{d2}, x \rightarrow \phi_{(x)} \qquad (7)$$

SVMs are appropriate for the nonlinearity problems since the training data x_i emerge only in scalar products. The scalar product x_i, x is calculated in the higher dimensional space $\phi_{(xi)}, \phi_{(x)}$. This transfer is performed with the use of a kernel function k according to Mercer's theorem (Burges, 1998):

$$k_{(x_i, x)} = \phi_{(x_i)}, \phi_{(x)} \qquad (8)$$

The Gaussian radial basis function kernel is used in this study (Waske et al., 2010; Xie, 2006):

$$k_{(x_i, x)} = e^{-\gamma |x - x_i|^2} \qquad (9)$$

where γ defines the width of the Gaussian kernel function. Instead of predicting the label directly, the class probability is calculated (Eq. 8) delivering the basis for the probability maps of urban expansion. Platt (1999) approximates the probabilities for binary SVMs using a sigmoid function as follows:

$$P(y = 1|x) = \frac{1}{1 + e^{A + f_{(x)}B}} \qquad (10)$$

where A and B are parameters estimated by minimizing the negative log-likelihood function (Platt, 1999).

The SVMs is implemented using the software tool imageSVM® in the EnMAP Toolbox® developed at Humboldt University of Berlin. Initially, imageSVM tool has been developed to solve classification problems in the context of multi- and hyperspectral satellite imagery (Waske et al., 2010). The output of SVMs classification with imageSVM is not only a classified binary image but also a probability image based on the principles of Eq. 10.

It is important to determine the best parameter values for constructing a probability map based on SVMs algorithm, including appropriate values for the penalty parameter C (Eq. 6) and the kernel parameter γ defining the width of the RBF kernel (Eq. 9). We use the n-fold cross validation procedure (Hsu, Chang, & Lin, 2010) as it is an effective method for balancing the accuracy results of known training data with unknown testing data. According to the curse of dimensionality and the Hughes phenomenon, which describes the degradation of the classifier performance when increasing the number of features, it is additionally advisable to select the optimal feature combination (Hughes, 1968). This selection of relevant features can improve prediction ability, generalization performance, and computational efficiency of SVMs (Nguyen & De La Torre, 2010). We employ feature selection which provides additional insights into the impacts of the various driving forces. A common method of SVMs feature selection is a forward feature selection (FFS) (Hsu et al., 2010; Waske et al., 2010), which initially trains each feature of the input feature set. The best performing feature is selected and the remaining features are used for training in combination with the initially selected one. The procedure is repeated until all features have been selected. The result is a functional ranking of the different feature combinations and those features which weaken SVMs classifier can be eliminated.

Model evaluation

Various methods of map comparison have been proposed to evaluate the outcomes of land use change models. Fuzzy map comparison (Bandemer & Gottwald, 1995) is one of these methods which offers potential for avoiding the problems of traditional cross-tabulate method and spatial metrics (Bandemer & Gottwald, 1995; Power, Simms, & White, 2001). A key factor in the fuzzy map comparison is that it considers the neighborhood of a cell to measure similarity of that cell in a value between 0 and 100 (fully similar). A number of studies have evaluated model performance based on the ROC (e.g. Achmad et al., 2015; Vermeiren et al., 2012) and spatial metrics summarizing the whole landscape (García, Santé, Crecente, & Miranda, 2011; Liu, Li, Shi, Wu, & Liu, 2008).

In this study, the process of evaluation is based on the following criteria: (i) the ROC statistic which is used to evaluate the obtained probability maps of logit and SVMs and (ii) the fuzzy map comparison which is employed to evaluate the allocation performance of CA-logit versus CA-SVMs model. First, ROC method is used to compare the probability maps of logit and SVMs with the observed 2010 map. 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 varies between 0.5 (random fit) and 1 (perfect fit).

Second, the 2010 simulations (CA-logit and CA-SVMs) are evaluated against observed 2010 map using fuzzy map comparison. The average fuzzy map index is an exponential decay with a halving distance of two cells and a neighborhood with a four-cell neighbor extent as in Hagen (2003) and Mustafa et al. (2018a) and calculated as follows:

$$A_k = \frac{\sum\limits_{x_k \in X_{k,sim}} \left| I_{x_k 0} \cdot (1/2)^{0/2}, I_{x_k 1} \cdot (1/2)^{1/2}, \ldots, I_{x_k d} \cdot (1/2)^{d/2} \right|_{max}}{X_{k,actul}} \times 100$$

(11)

where A_k $(0 \le A \ge 100)$ is the average fuzzy map index for class k, I_{xkd} is 1 if cell x_k in the simulated map at zone d $(0 \le d \ge 4)$ is identical to one cell at zone d in the observed map otherwise is 0, $X_{k,sim}$ is the total number of changed cells of class k in the simulated map and $X_{k,actul}$ is the total number of changed cell of class k in the observed map.

Results and discussion

The proportion of urban land use increased from 15.9 to 16.5 percent, an area increase of 112 km^2 between 2000 and 2010. Table 1 shows the calibrated coefficients of logit model.

Table 1. Logit model coefficients and FFS rank.

Driver	Coefficient	Odds ratio	FFS rank
Intercept	1.60	-	-
Slope	−1.48	0.23	5
Elevation	0.29	1.34	6
Distance to highways	−1.56	0.21	4
Distance to main roads	−1.81	0.16	3
Distance to secondary roads	−2.44	0.09	2
Distance to local roads	−7.05	0.00	1

The goodness-of-fit of the logit model is evaluated using McFadden pseudo R-square, and its value is 0.227. Clark and Hosking (1986) reported that a McFadden pseudo R-square value greater than 0.2 indicates a good model fitness.

The relative contribution of each driver to urbanization is measured with the Odds Ratio (OR), that equals $exp(\beta)$. An OR greater than 1 indicates a positive effect, whereas a value of less than 1 indicates a negative effect. Logit model assesses an overall model performance and the significance of individual explanatory variables. All selected driving forces are statistically significant at $p\text{-value} \le 0.05$ except for elevation, which has $p\text{-value}$ of 0.139.

The rank according to SVMs FFS is given in Table 1. The results show that the FFS rank follows the magnitude of the logit coefficients. According to the results, the major driving forces of the urban expansion process are related to the road network especially local roads. According to the OR values, distance to roads show a negative effect on the urban process so that the non-urban to urban transitions generally occur close to roads as in Rienow and Goetzke (2015). Slope also shows a negative effect on urban expansion.

In order to exclusively evaluate different models performance, all persistent urban areas in 2000 were excluded. When using the ROC statistic to compare logit and SVMs approaches the curve of the SVMs model gives the best result. It clearly reaches a stable level much earlier than logit curve, Figure 5.

The ROC value of logit and SVMs are 0.689 and 0.723, respectively. Qualitative analysis of the probability maps can provide some explanation for the varying performances of the two approaches. Figure 6 presents the probability maps based on SVMs and logit. The major difference between the two maps is the transition areas between high and low probability. Logit map renders these areas as gradual transitions whereas SVMs map renders these areas as sharp edges.

We investigate the performance of both models in the dynamic environment of some random noises by incorporating a stochastic perturbation (SP) term in Equation 1 as follows:

$$Purbn_{ij}^t = PD_{ij}^t \times N_{ij}^t \times con(.) \times SP \qquad (12)$$

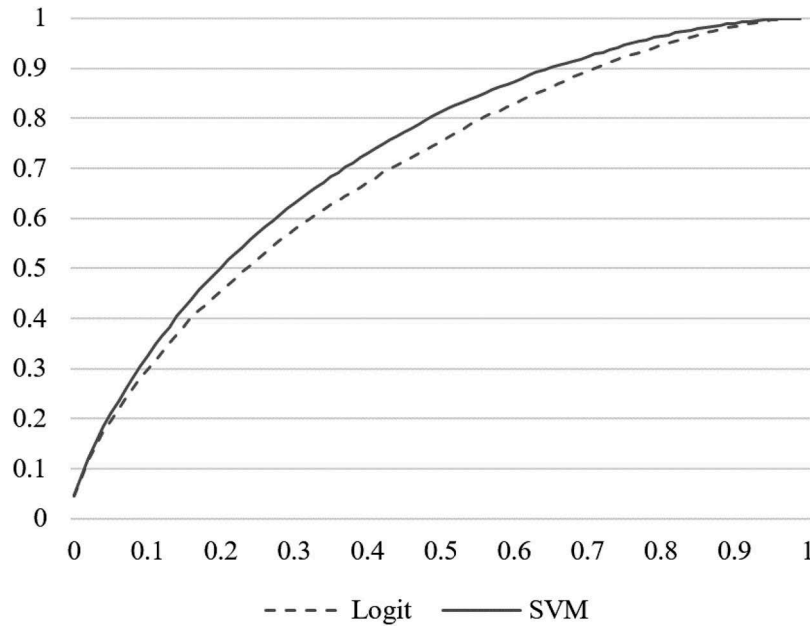The SP term is calculated as follows (White & Engelen, 1993):
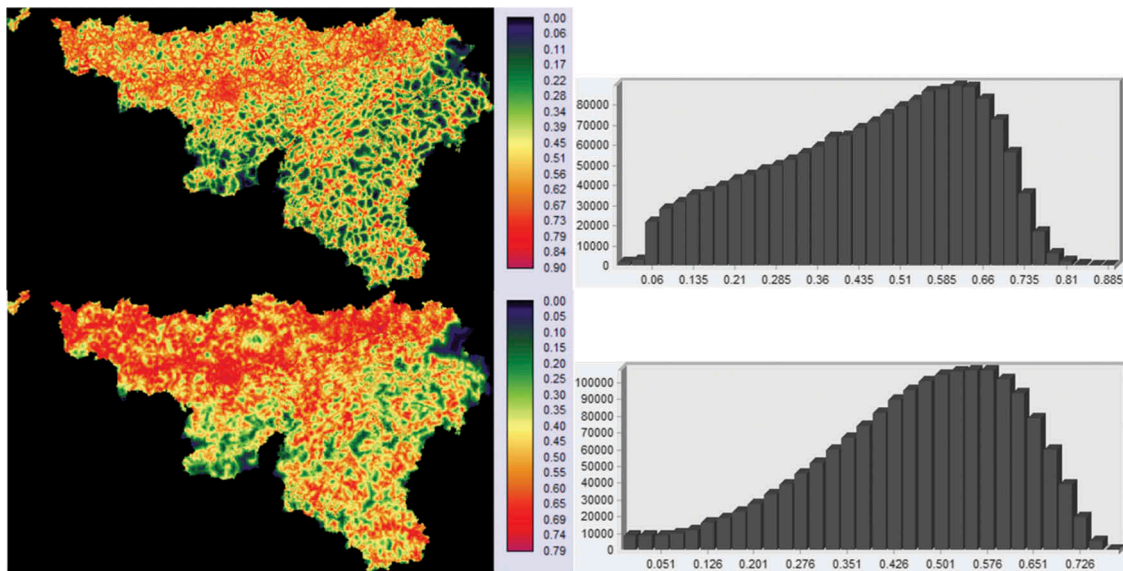
Figure 5. ROC curves of logit and SVMs.

Figure 6. Probability maps and histograms of SVMs (top) and logit (bottom).

$$SP = 1 + (-\ln \rho)^{\alpha} \qquad (13)$$

where ρ is a uniform random number between 0 and 1, and α is a parameter that allows to control the degree of the SP. We set α at 0.05 as recommended by Mustafa et al. (2014). Table 2 lists the maximum, the average and the minimum fuzzy accuracy rates for 200 runs (100 each approach). The results reveal that

Table 2. Average fuzzy accuracy rates of CA-SVMs and CA-logit for 100 runs per approach considering stochastic perturbation term.

	CA-SVMs	CA-logit
Original model	31.46	29.86
Maximum	31.65	29.84
Average	31.50	29.74
Minimum	31.38	29.68

the performance of CA-SVMs is slightly improved by introducing SP term in contrast to CA-logit. One explanation is that CA-SVMs differentiates between cells with higher probabilities and cells with lower probabilities in a better way than CA-logit as shown in Figure 6. Still the observed improvement is not spectacular, especially when one considers the uncertainty related to such models and the indirect cost of the CA-SVMs approach. One of these costs is related to the fact that the relative weight of the different explanatory variables in the result is no longer made explicit, as opposed with the logit approach.

Table 3 shows the average fuzzy accuracy rates between the simulated urban map in 2010 predicted by CA-SVMs runs with different neighboring sizes and the observed urban pattern in 2010. The results

Table 3. Average fuzzy accuracy rates of CA-SVM considering different neighborhood window sizes.

Neighbor size	Average fuzzy accuracy (%)
11 × 11	28.49
9 × 9	29.03
7 × 7	29.95
5 × 5	30.52
3 × 3	31.46

reveal that model run with the window size of 3 × 3 produces the highest accuracy rate. This is line with Chen et al. (2014) and Poelmans and Van Rompaey (2009) who analyzed several neighbors window sizes and concluded that the model run with the 3 × 3 neighborhood window produces a land use pattern that most fits the actual pattern.

The average fuzzy accuracy of simulated urban expansion by CA-SVMs and CA-logit based on 3 × 3 neighborhood window size in comparison to the observed urban expansion is 31.46% and 29.86%, respectively. One of the main reasons for the moderate accuracy rate is that we selected a set of urbanization driving forces without any insights into the urbanization process in Wallonia, as the main focus of the present study is evaluating the performance of CA-SVMs vs CA-logit. Another possible source of this moderate accuracy rate is related to uncertainties in the decision of urban developers. However, it is routine for CA urban expansion models, to show low accuracy rates due to the complexity of urban environment (Jantz, Goetz, & Shelley, 2003; Mustafa et al., 2017; Wang et al., 2013).

Conclusions

This paper has been contributed to the few number of studies that calibrated transition rules of CA models using SVMs. We also have assessed the performance of CA-SVMs in comparison with CA-logit model. Coupling CA models with SVMs or logit enables the simultaneous dynamic simulation of land use change process along with the analyses of a number of controlling factors that determine change suitability. Our model has been applied to Wallonia (Belgium) as a case study, but the model is generic and can be applied to other case studies nonetheless. In such a case, an investigation of the transferability of the model parameters is an interesting direction for future research.

We have examined two main aspects of the accuracy of the model: (i) the goodness-of-fit of probability maps and (ii) fuzziness similarity of CA-logit and CA-SVMs models. The results show that SVM-based probabilities exhibit a better performance compared to those derived by logit. Furthermore, the SVMs render the edges between low and high land use change probability areas in a more efficient way than logit.

Although SVMs enriches the calibration methods of CA models, limitations of this method exist because SVMs are relatively complex in its theory and implementation. Moreover, due to their black-box nature, they do not allow to ponder relative contribution of each explanatory variable, which is a key element for policymaking. Therefore, more studies within land use change modeling domain are needed to improve our understanding of the requisite mathematical and computer knowledge of SVMs.

ORCID

Ahmed Mustafa http://orcid.org/0000-0002-1592-6637
Andreas Rienow http://orcid.org/0000-0003-3893-3298
Ismaïl Saadi http://orcid.org/0000-0002-3569-1003
Mario Cools http://orcid.org/0000-0003-3098-2693
Jacques Teller http://orcid.org/0000-0003-2498-1838

References

Aburas, M.M., Ho, Y.M., Ramli, M.F., & Ash'aari, Z.H. (2016). The simulation and prediction of spatio-temporal urban growth trends using cellular automata models: A review. *International Journal Applications Earth Obs Geoinformation*, 52, 380–389.

Achmad, A., Hasyim, S., Dahlan, B., & Aulia, D.N. (2015). Modeling of urban growth in tsunami-prone city using logistic regression: Analysis of Banda Aceh, Indonesia. *Applications Geographic*, 62, 237–246.

Aldrich, J.H., & Nelson, F.D. (1984). *Quantitative Applications in the Social Sciences: Linear probability, logit, and probit models*. Thousand Oaks, CA: SAGE Publications Ltd doi: 10.4135/9781412984744.

Bandemer, H., & Gottwald, S. (1995). *Fuzzy sets, fuzzy logic, fuzzy methods*. Chichester: Wiley.

Batty, M., Xie, Y., & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Computation Environment Urban Systems*, 23, 205–233.

Belgian Federal Government. 2013. Statistics Belgium [WWW Document]. *Stat. Belg. URL*. Retrieved from http://statbel.fgov.be/fr/statistiques/chiffres/(Accessed April 29.14).

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines, in: Data mining techniques for the life sciences, methods in molecular biology. *Humana Pressure*, 223–239. doi:10.1007/978-1-60327-241-4_13

Berberoğlu, S., Akın, A., & Clarke, K.C. (2016). Cellular automata modeling approaches to forecast urban growth for adana, Turkey: A comparative approach. *Landscape Urban Planning*, 153, 11–27.

Boser, B.E., Guyon, I.M., & Vapnik, V.N., 1992. *A training algorithm for optimal margin classifiers*. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92, ACM, New York, NY, USA, pp. 144–152. doi:10.1145/130385.130401

Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 121–167.

Chang, -C.-C., & Lin, C.-J. (2001). *LIBSVM: A library for support vector machines*. National Taiwan University, Taipei, Taiwan.

Chen, Y., Li, X., Liu, X., & Ai, B. (2014). Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy. *International Journal Geographic Information Sciences*, 28, 234–255.

Cheng, J., & Masser, I. (2003). Urban growth pattern modeling: A case study of Wuhan city, PR China. *Landscape Urban Planning*, 62, 199–217.

Clark, W.A.V., & Hosking, P.L. (1986). *Statistical methods for geographers* (1 ed.). New York: Wiley.

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 Geographic Information Sciences*, 12, 699–714.

EEA. (2011). *Landscape fragmentation in Europe (Publication)*. European Environment Agency, Copenhagen.

Feng, Y., Liu, Y., Tong, X., Liu, M., & Deng, S. (2011). Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landscape Urban Planning*, 102, 188–196.

García, A.M., Santé, I., Crecente, R., & Miranda, D. (2011). An analysis of the effect of the stochastic component of urban cellular automata models. *Computation Environment Urban Systems*, 35, 289–296.

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal Geographic Information Sciences*, 17, 235–249.

Hallowell, G.D., & Baran, P.K. (2013). Suburban change: A time series approach to measuring form and spatial configuration. *Journal Space Syntax*, 4, 74–91.

Hsu, C., Chang, C., & Lin, C. (2010). *A practical guide to support vector classification*. National Taiwan University, Taipei, Taiwan

Hu, Z., & Lo, C.P. (2007). Modeling urban growth in Atlanta using logistic regression. *Computation Environment Urban Systems*, 31, 667–688.

Huang, B., Xie, C., & Tay, R. (2010). Support vector machines for urban growth modeling. *GeoInformatica*, 14, 83–99.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions Information Theory*, 14, 55–63.

Iannone, G., & Troisi, A. (2013). Ca-pri, a cellular automata phenomenological research investigation: Simulation results. *International Journal Modern Physical C*, 24, 1350027.

Iannone, G., Troisi, A., Guarnaccia, C., D'agostino, P.P., & Quartieri, J. (2011). An urban growth model based on a cellular automata phenomenological framework. *International Journal Modern Physical C*, 22, 543–561.

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. *Environment Plan B Plan Design*, 31, 251–271.

Jokar Arsanjani, J., Helbich, M., Kainz, W., & Darvishi Boloorani, A. (2013). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal Applications Earth Obs Geoinformation*, 21, 265–275.

Kryvobokov, M., Mercier, A., Bonnafous, A., & Bouf, D. (2015). Urban simulation with alternative road pricing scenarios. *Case Studies Transp Policy*, 3, 196–205.

Liu, X., Li, X., Shi, X., Wu, S., & Liu, T. (2008). Simulating complex urban development using kernel-based non-linear cellular automata. *Ecology Model*, 211, 169–181.

Martens, D., Baesens, B., Van, G., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal Operational Researcher*, 183, 1466–1476.

Montgomery, D.C., & Runger, G.C. (2003). *Applied statistics and probability for engineers* (4th ed.). New York: John Wiley & Sons.

Munshi, T., Zuidgeest, M., Brussel, M., & van Maarseveen, M. (2014). Logistic regression and cellular automata-based modelling of retail, commercial and residential development in the city of Ahmedabad, India. *Cities*, 39, 68–86.

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*, 69C, 529–540. https://doi.org/10.1016/j.landusepol.2017.10.009

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.

Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., & Teller, J. (2018a). Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computation Environment Urban Systems*, 67, 147–156.

Mustafa, A., Rompaey, A.V., Cools, M., Saadi, I., & Teller, J. (2018b). Addressing the determinants of built-up expansion and densification processes at the regional scale. *Urban Studies (Edinburgh, Scotland)*, 1–20. doi:10.1177/0042098017749176

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.

Mustafa, A., Saadi, I., Cools, M., & Teller, J. (2018c). Understanding urban development types and drivers in Wallonia. A multi-density approach. *International Journal of Business Intelligence and Data Mining*, 13, 309–330.

Nguyen, M.H., & De La Torre, F. (2010). Optimal feature selection for support vector machines. *Pattern Recognition*, 43, 584–591.

Pace, R.K., & LeSage, J.P. (2002). Semiparametric maximum likelihood estimates of spatial dependence. *Geographic Analysis*, 34, 76–90.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advancement Large Margin Classifier*, 10, 61–74.

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. *Landscape Urban Planning*, 93, 10–19.

Poelmans, L., & Van Rompaey, A. (2010). Complexity and performance of urban expansion models. *Computation Environment Urban Systems*, 34, 17–27.

Power, C., Simms, A., & White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal Geographic Information Sciences*, 15, 77–100.

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.

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal Remote Sensing*, 50, 144–154.

Rienow, A., & Goetzke, R. (2015). Supporting SLEUTH – Enhancing a cellular automaton with support vector machines for urban growth modeling. *Computation Environment Urban Systems*, 49, 66–81.

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

Schölkopf, B., Burges, C., & Vapnik, V., 1996. *Incorporating invariances in support vector learning machines, in: Artificial Neural Networks — ICANN 96*. Presented at the International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg, pp. 47–52. doi:10.1007/3-540-61510-5_12

Ševčíková, H., Raftery, A.E., & Waddell, P.A. (2007). Assessing uncertainty in urban simulations using Bayesian melding. *Transp Researcher Particle B Methodol*, 41, 652–669.

SPW. (2013). *Schéma de Développement de l'Espace Régional-Une vision pour le territoire wallon (Regional space development plan – A vision for Wallonia)*.Service Public de Wallonie. Namur, Belgium.

Tannier, C., & Thomas, I. (2013). Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian cities. *Computation Environment Urban Systems*, 41, 234–248.

Thomas, I., Frankhauser, P., & Biernacki, C. (2008). The morphology of built-up landscapes in Wallonia (Belgium): A classification using fractal indices. *Landscape Urban Planning*, 84, 99–115.

Troisi, A. (2015). Can CA describe collective effects of polluting agents? *International Journal Modern Physical C*, 26, 1550114.

Van Gestel, T., Suykens, J.A., Baestaens, D.-E., Lambrechts, A., Lanckriet, G., Vandaele, B., … Vandewalle, J. (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions Neural Network*, 12, 809–821.

Vapnik, V.N. (1995). *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc..

Vapnik, V.N., & Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.

Verburg, P.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. *Environment Plan B Plan Design*, 31, 125–150.

Vermeiren, K., Van Rompaey, A., Loopmans, M., Serwajja, E., & Mukwaya, P. (2012). Urban growth of Kampala, Uganda: Pattern analysis and scenario development. *Landscape Urban Planning*, 106, 199–206.

Verplancke, T., Van Looy, S., Benoit, D., Vansteelandt, S., Depuydt, P., De Turck, F., & Decruyenaere, J. (2008). Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC Medica Informatics Decisions Mak*, 8, 56.

Vogel, R. 2013. *Entwicklung eines automatisierten Wolkendetektions- und Wolkenklassifizierungsverfahrens mit Hilfe von Support Vector Machines angewendet auf METEOSAT-SEVIRI-Daten für den Raum Deutschland* (Text.PhDThesis).

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal American Planning Association*, 68, 297–314.

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. *Landscape Urban Planning*, 110, 99–112.

Wang, J., & Maduako, I.N. (2018). Spatio-temporal urban growth dynamics of Lagos metropolitan region of Nigeria based on hybrid methods for LULC modeling and prediction. *European Journal Remote Sensing*, 51, 251–265.

Waske, B., Linden, S.V.D., Benediktsson, J.A., Rabe, A., & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions Geoscience Remote Sensing*, 48, 2880–2889.

White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land-use patterns. *Environment Plan A*, 25, 1175–1199.

Wu, F. (2002). Calibration of stochastic cellular automata: The application to rural-urban land conversions. *International Journal Geographic Information Sciences*, 16, 795–818.

Xie, C. (2006). *Support vector machines for land use change modeling*. University of Calgary, Calgary, Canada.

Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computational Geoscience*, 34, 592–602.

Zhang, H., Zeng, Y., Bian, L., & Yu, X. (2010). Modelling urban expansion using a multi agent-based model in the city of Changsha. *Journal Geographic Sciences*, 20, 540–556.