DOI: 10.1111/tgis.12756

RESEARCH ARTICLE

Transactions in GIS WILEY

Modeling urban regions: Comparing random forest and support vector machines for cellular automata

Andreas Rienow[1] | Ahmed Mustafa[2] | Leonie Krelaus[1] | Claudia Lindner[1]

[1]Institute of Geography, Ruhr-University Bochum, Bochum, Germany

[2]Urban Systems Lab, The New School, New York, NY, USA

Correspondence
Andreas Rienow, Institute of Geography, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany.
Email: andreas.rienow@rub.de

Funding information
US National Science Foundation, Grant/Award Number: 1934933

Abstract

Cellular automaton (CA) are important tools that provide insight into urbanization dynamics and possible future patterns. The calibration process is the core theme of these models. This study compares the performance of two common machine-learning classifiers, random forest (RF), and support vector machines (SVM), to calibrate CA. It focuses on the sensitivity analysis of the sample size and the number of input variables for each classifier. We applied the models to the Wallonia region (Belgium) as a case study to demonstrate the performance of each classifier. The results highlight that RF produces a land-use pattern that simulates the observed pattern more precisely than SVM especially with a low sample size, which is important for study areas with low levels of land-use change. Although zoning information notably enhances the accuracy of SVM-based probability maps, zoning marginally influences the RF-derived probability maps. In the case of the SVM, the CA model did not significantly improve due to the increased sample size. The performance of the 5,000 sample size was observed to be better than the 15,000 sample size. The RF-driven CA had the best performance with a high sample, while zoning information was excluded.

1 | INTRODUCTION

In his 1937 lecture on town planning, American landscape designer Earle Draper coined a term to describe the unaesthetic and uneconomic settlement structure of American cities. In reflecting on the obliteration of rural and urban spaces, this term became increasingly popular: "perhaps diffusion is too kind of word… in bursting its bounds, the city actually *sprawled* and made the countryside ugly, uneconomic [in terms] of services and doubtful social value" (Wassmer, 2002, p. 9). More than 80 years have passed since the term "sprawl" emerged in a geographical context, and the specific patterns and processes of growing cities have developed from an American to a European problem. Currently, approximately 75% of Europeans live in urban areas (European Environment Agency, 2006; IT.NRW, 2013; Lavalle, Demicheli, & Kasanko, 2002). Accordingly, land consumption, in which the conversion of open spaces is utilized for built-up land use, is a constantly evolving process spilling over the actual urban agglomeration borders and bequeathing a spatial footprint of sealed surfaces to rural regions. Urban sprawl is one of the most challenging land-use/cover changes, with implications for anthropogenic and geobiophysical spheres; it has become an inherent part of the discourse on international sustainability in the context of global change (Intergovernmental Panel on Climate Change, 2012; Millennium Ecosystem Assessment, 2005; Ramankutty et al., 2006; United Nations Environment Programme, 2012). The same holds true for geographic research that is characterized by a long history of diverse theories and models of land-use change in general and urban sprawl. Above all, these models attempt to simulate future patterns, predict growth rates, and understand the cause–effect relationships of the forces driving urban sprawl. One of the most important streams of urban models perceives urban areas as complex systems that are characterized by nonlinearity, irrationality, multilevel feedback loops, and emergence phenomena. These geosimulation techniques make use of artificial intelligence to model the micro-processes responsible for the macro-patterns of urban systems (Batty, 2005; Benenson & Torrens, 2004; Couclelis, 1989; Miller, Hunt, Abreha, & Salvini, 2004). In line with the well-known quote by Aristotle—"the whole is more than the sum of its parts"—urban sprawl models simulate urban systems using a bottom-up approach. Among other approaches, cellular automata (CA) stimulate the shift from aggregation to disaggregation, homogeneity to heterogeneity, and equilibrium to disequilibrium in urban modeling. The creation of CA has been credited to the mathematicians, von Neumann (1951) and Ulam (1952); "one can say that the 'cellular' comes from Ulam and the 'automata' comes from von Neumann" (Rucker, 1999, p. 69). The final breakthrough of CA came with John Conway's Game of Life in 1970 (Batty, Barros, & Júnior, 2006). Urban CA are often defined by five key traits: (a) a raster lattice representing the spatial context; (b) a set of states associated with a cell with a specific land-use type; (c) neighborhoods influencing the spatial configuration; and (d) transition rules regulating the conversion of a cell state with every (e) time-step (Batty, 2005; Benenson & Kharbash, 2005; Silva, 2004, 2011; Silva & Clarke, 2005; Wu & Silva, 2010). Interregional land-use and land-cover information is often derived from satellite-based classifications, that is, raster data (Verburg, Kok, Pontius, & Veldkamp, 2006). As such, the gridded two-dimensional (2D) character of a CA environment makes it conducive to the simulation of urban sprawl. Several studies have utilized the simulation power of CA to investigate future patterns of urban sprawl with sophisticated calibration methods (Abolhasani & Taleai, 2020; Alaei Moghadam, Karimi, & Habibi, 2018; Charif, Omrani, Abdallah, & Pijanowski, 2017; Kantakumar, Kumar, & Schneider, 2016; Lin & Li, 2016; Momeni & Antipova, 2020; Zhang, Wang, He, & Xia, 2020). The bottom-up approach of CA does not provide insight into how urban patterns were formed or which processes of the coupled human–environment system in metropolitan regions underpin such patterns (Batty, 2005; Benenson & Torrens, 2004; Wu & Silva, 2010).

This study aims to apply different machine-learning (ML) classifiers to analyze the relationship between urban sprawl and its driving forces. We test two popular classifiers, random forest (RF) and support vector machines (SVM), to generate urban change probability maps. These maps are compared in terms of sensitivity to input parameters, modeling accuracy, variable importance measurements, and CA implementation abilities. We selected Wallonia (Belgium) as a case study to evaluate both ML classifiers. In this case, Belgian cadastral data were used

for urban land use and a set of static variables relating to accessibility, geophysical features, policies, and socio-economic factors set the urbanization driving forces.

In recent years, an increasing number of studies have investigated how future urban sprawl patterns may be modeled with CA by gaining insights into their driving factors, focusing on time-consuming, albeit highly accurate, applications such as differential evolution and genetic algorithms (Feng & Tong, 2018), advancing classical methods such as geographically weighted logistic regression (Mirbagheri & Alimohammadi, 2017), or implementing behavior-oriented approaches such as subjective local weights (Karimi Firozjaei, Sedighi, & Jelokhani-Niaraki, 2020) and swarm intelligence (Cao, Huang, Xu, Lü, & Chen, 2019). In terms of ML, there are a large number of methods offering potentially effective and efficient classification. In their review, Maxwell, Warner, and Fang (2018) compared SVM, single decision trees (DTs), RF, boosted DTs, artificial neural networks, and *k*-nearest neighbors based on the choice of algorithm, training data requirements, user-defined parameter selection and optimization, feature space impacts and reduction, and computation costs. They concluded that SVM, RF, and boosted DTs demonstrated the best overall performance. They also found that boosted DTs were slow and thus less attractive than SVM and RF. Samardžić-Petrović, Dragićević, Kovačevic, and Bajat (2016) implemented an SVM-based approach to model urban land-use change without incorporating a CA component to simulate urban sprawl dynamically or including zoning information. Mustafa, Rienow, Saadi, Cools, and Teller (2018) compared a coupled CA and logistic regression (CA-logit), with a coupled CA and SVM (CA-SVM) model; the former is one of the most common CA-based models. They reported that the CA-SVM model significantly outperformed the CA-logit model. Yang, Li, and Shi (2008) and Okwuashi, McConchie, Nwilo, and Eyo (2009) combined SVM with CA-based approaches to model urban sprawl in a spatially explicit manner. Yang et al. (2008) applied their model to the city of Shenzhen, and Okwuashi et al. (2009) modeled the Lekki area of Lagos. Both of these studies derived nonlinear transition rules for CA simulation of urban land-use dynamics; however, they did not exploit the complementary advantages of SVM and CA, and in both cases urban sprawl was erratically directed. Rienow and Goetzke (2015) demonstrated how SVM can enhance the performance of urban CA and provide insights into the determinants underpinning urban sprawl. A sensitivity analysis of the integration of zoning information and sample sizes was not undertaken. Harris and Grunsky (2015) used RF for land-cover prediction. The research was lithologically oriented and an integration in a cellular automaton was not undertaken. Kamusoko and Gamba (2015) showed that the RF–CA combination outperformed the SVM–CA and logistic regression–CA combinations. However, they stated that their study was only suited to the study area, no zoning information was tested, and the models were only applied to one city (Harare, Zimbabwe). Additionally, some model uncertainties remain in the literature, and these will be clarified during this study by addressing the following research questions:

1. Is RF equally suitable for CA-based urban change modeling as SVM when modeling metropolitan regions?
2. What impact does zoning information have on RF- and SVM-based probability maps and does it automatically enhance CA simulation performance?
3. What impact does initial sample size have on ML-based probability maps and the CA outcome?
4. How does the calibrated cellular automaton simulate future urban expansion in the Wallonia region until 2030? To answer this question, we used the best-performing model to simulate future urban development in the study area.

Simulating possible future patterns is crucial to understand the spatial processes of urban development over time and serve as tools to evaluate the potential impacts of policies on urban spaces and the environment.

The remainder of this article is structured into five additional sections. Section 2 describes the study area, and Section 3 lists and discusses the various input data sets. This section also details the models applied and the methods chosen to assess accuracy. Section 4 presents the parametrization of RF and SVM, the resulting probability maps and their validation, and the variable importance and validation of CA. Section 5 provides a short discussion, offers a perspective for future research direction and concludes.

2 | STUDY AREA

The case study area is the Wallonia region in Belgium, in which we simulate urban expansion dynamics between 2000 and 2010. Wallonia (Figure 1) spans 16,844 km^2 with around 3.5 million inhabitants as of 2010; this amounts to one third of Belgium's population (Belgian Federal Government, 2013). The region is predominantly French-speaking. Its capital, Namur, and the most populous cities, Liège, Charleroi, and Mons, are located in the northern part of the region, while the remainder of the region is less urbanized. As such, the population of Wallonia is highly concentrated in the northern areas, following the nineteenth-century industrial axis, running from east (Liège) to west (Mons) (Thomas, Frankhauser, & Biernacki, 2008). North of this axis, urban development has been considerably influenced by the Brussels metropolitan area. Toward the far south of Wallonia, which is less densely populated, urban development is influenced by the city of Luxembourg (Thomas et al., 2008). In the early 1900s, rural and property interests were able to prevent a dispersed urban development through land-use plans in Belgium and plans specific to Wallonia, which had long-term consequences on urban sprawl. The development of land-use plans for the three regions in the 1970s and 1980s led to a slight reduction in the progression of urban sprawl (Mustafa & Teller, 2020).

3 | DATA AND METHODS

3.1 | Data

We use data that have been created and processed by Mustafa, Heppenstall, et al. (2018) and Mustafa, Rienow, et al. (2018) for the present study. The data set consists of built-up data and a set of auxiliary data. Belgian cadastral data (CAD) were used to generate two built-up maps for 2000 and 2010 using the construction date provided for each building. The vector CAD were rasterized with a cell dimension of 2 m × 2 m, and then aggregated up to

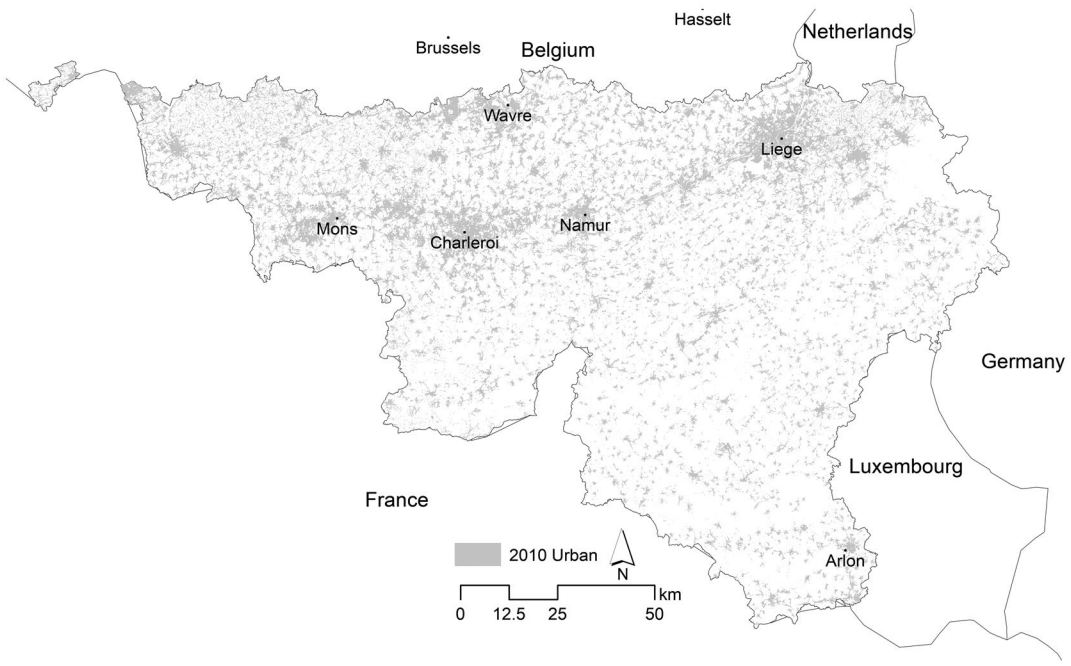

FIGURE 1 Study area: Wallonia region, Belgium

TABLE 1 List of urbanization driving forces

ID	Name	Description
1	DEM	Elevation data for the study area
2	Slope	Slope rise percentage
3	Dist large cities	Euclidean distance to major cities in Belgian (over 90,000 inhabitants)
4	Dist med cities	Euclidean distance to major cities in Belgian (20,000–90,000 inhabitants)
5	Dist RD1	Euclidean distance to highways
6	Dist RD2	Euclidean distance to main roads
7	Dist RD3	Euclidean distance to secondary roads
8	Dist RD4	Euclidean distance to local roads
9	Dist rail	Euclidean distance to railway stations
10	Zoning	Wallonia official zoning plan (code 1, urban development is legally permitted; code 0, otherwise)

100 m × 100 m raster maps following Mustafa, Heppenstall, et al. (2018). Each 100 m × 100 m cell shows the sum of 2 m × 2 m cells within its boundary. All 100 m × 100 m cells that contained 25 or more of the 2 m × 2 m cells (this corresponds to an average-sized building of 100 m^2) were classified as built-up cells and other cells as non-built-up cells. Table 1 presents a set of auxiliary data that was used as the urbanization driving forces. In simulating urban sprawl, the probability of land-use change is often determined by using factors that drive urban development, where proximity variables often represent factors of accessibility to markets (Wang et al., 2020). We normalized all driving forces between 0 and 1. The accuracy of many ML methods such as SVMs, can substantially deteriorate if the data are not normalized (Ben-Hur & Weston, 2010; Chang & Lin, 2001).

3.2 | Transition rules of the cellular automaton

The urban land-use change model used in this study is a CA raster-based model. The model inputs were two actual built-up maps for 2000 and 2010 (Section 3.1), and the urbanization driving forces (Table 1). The model used these built-up maps to generate a binary map showing changes between 2000 and 2010 (non-urban to urban transitions). The model simulates the 2000–2010 changes and compares it with the actual 2000–2010 period using the fuzzy similarity rate (Equation 3 below).

This article focuses on urban expansion as an example application. The transition rule contained two different components: a set of urbanization driving forces (Table 1); and the local neighborhood characteristics. The transition potential, *P*, for a cell, *ij*, changing its state was calculated as:

$$P_{ij} = \left(P_d\right)_{ij} \times \left(P_n\right)_{ij} \times con\left(.\right) \tag{1}$$

where $(P_d)_{ij}$ is the urban probability based on the driving forces; $\left(P_n\right)_{ij}$ is the neighborhood effect on the cell, *ij*; and *con*(.) is a restrictive condition for land-use change. In this case study, *con*(.) was 0 if the land-use class of cell, *ij*, was water, based on the official zoning plan, and 1 otherwise.

The value of $(P_n)_{ij}$ was calculated using the following equation (Mustafa, Rienow, et al., 2018):

$$\left(P_n\right)_{ij} = \frac{\sum \text{count}\left(s = \text{urban}\right)}{n\left(n-1\right)} \tag{2}$$

where count(*s*=urban) is the number of urban cells in the Moore 3 × 3 neighborhood.

The model changes a number of non-urban cells during each time-step, corresponding to one year. Cells are first ranked in descending order of transition potential (*P*) values. Then, non-urban to urban transitions are conducted beginning from the cell with the highest *P* and proceeding until the amount of new urban cells is satisfied. Our main goal is to evaluate the spatial allocation ability of the model calibrated using SVMs and RF. As such, the number of new urban cells was equal to the actual changes between 2000 and 2010; this value was then divided by the number of time-steps (10 years).

As the primary aim of this study is to compare the performance of SVMs with RF, the model estimated the urban probability based on the driving forces $(P_d)_{ij}$, in Equation (1), via two different approaches using SVMs (Section 3.3) and RF classifiers (Section 3.4). The dependent variable for both classifiers was defined as a binary map demonstrating the spatial pattern of actual urban expansion between 2000 and 2010. A value of 1 signified changes from non-urban to urban, while a value of 0 indicated that the non-urban cell did not change its land use. The independent variables were the urbanization driving forces listed in Table 1.

The evaluation process of the performance of our model consists of: (a) the evaluation of the probability maps computed by SVMs and RF, $(P_d)_{ij}$ in Equation (1), using the recording operator characteristics (ROC) procedure; and (b) the comparison of the simulated built-up maps of 2010 with that produced by the model with the real map of 2010 using the maximization of the fuzzy similarity rate (FSR) between simulated and observed maps. The comparison considers only changes in terms of new urban transitions. The fuzzy similarity rate was calculated as follows (Mustafa, Heppenstall, et al., 2018):

$$FSR_k = \frac{\sum_{ij_k \epsilon X_{k,\mathrm{sim}}} \left| I_{ij_k 0} \cdot (1/2)^{0/2}, I_{ij_k 1} \cdot (1/2)^{1/2}, \ldots\ldots, I_{ij_k d} \cdot (1/2)^{d/2} \right|_{\max}}{X_{k,\mathrm{actual}}} \times 100 \tag{3}$$

where FSR_k ($0 \le FSR_k \le 100$) is the fuzzy similarity rate for class *k*; $I_{ij_k d}$ is 1 if cell ij_k in the simulated map in zone *d* ($0 \le d \le 4$) has a similar land-use class to one cell at zone *d* in the observed map, and 0 otherwise; $X_{k,\mathrm{sim}}$ is the quantity of change for class *k* in the simulated map; and $X_{k,\mathrm{actual}}$ equals the quantity of change for class *k* in the observed map.

3.3 | Support vector machines

SVMs were first described in their contemporary form by Cortes and Vapnik (1995). Along with artificial neural networks and genetic programming, SVMs represent a new generation of ML algorithms. Fundamentally, SVM is a linear binary classifier that labels a sample of empirical data by constructing the optimal separating hyperplane. Traditional ML methods attempt to minimize the empirical training error, which prompts a tendency to overfit (Vapnik, 1998; Xie, 2006). As these methods are strongly tailored to the training data, it is difficult to extend them to additional data. According to the principles of structural risk minimization (Vapnik, 1995, 1998), the SVM function minimizes the upper bound of the expected generalization error by maximizing the margin between the separating hyperplane and data. The margin concept is a key element in the SVM approach as it indicates its generalizability (Burges, 1998; Huang, Xie, & Tay, 2010).

The principal advantage of SVM is the option to transform the model to solve a nonlinear classification problem without any a priori knowledge. Based on a scenario with a set of training vectors *T*, a hyperplane must be identified that separates the positive feature vectors from the negative. This *separating hyperplane H* may be parameterized linearly by *w* and *b*:

$$H: \langle w, x \rangle + b = 0 \tag{4}$$

where the *w* element of R^d is normal to *H*, and the *b* element of *R* is the bias.

The classification problem may be formalized as the decision function:

$$\operatorname{sgn}\left(f_{(x)}\right) = \operatorname{sgn}\left(\langle w, x\rangle + b\right) \tag{5}$$

For the linearly separable case, we defined two hyperplanes, H_+ and H_-, constructed by the closest positive and negative examples, termed the *support vectors*:

$$\begin{aligned} H_+ &: \langle w, x\rangle + b = 1 \\ H_- &: \langle w, x\rangle + b = -1 \end{aligned} \tag{6}$$

Through the perpendicular distance from the origin of H_+ and H_- the distance between the optimal separating hyperplane H_+ and H (or 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 where the margin between H_+ and H_- is the largest; as such, $||w||$ must be minimized.

The formulation of the constrained optimization problem is:

$$\begin{gathered} \min_{w,b} \frac{1}{2} w^2 + C \sum_{i=1}^{n} \xi_i \\ \text{subject to } y_i \left(\langle w, x_i\rangle + b\right) - 1 \geq 0 \text{ for } i = 1, \ldots, n \end{gathered} \tag{7}$$

where the constant, C, is termed the *penalty parameter*, and ξ_i is a slack variable representing error in the classification. The first part of the objective function attempts to maximize the margin between the two classes, while the second part minimizes the classification error. The optimization problem was solved through its formulation in a dual form derived by constructing a Lagrange function according to the Karush–Kuhn–Tucker optimality condition (Burges, 1998). The resulting classification rule was:

$$\operatorname{sgn}\left(f_{(x)}\right) = \left[\operatorname{sgn} \sum_{i=1}^{n} a_i y_i \left(x_i, x + b\right)\right] \tag{8}$$

where α_i is the corresponding Lagrange multipliers to x_i; b is the constant; and the support vectors are all x_i, where $\alpha_i \neq 0$.

If the classification problem was not linearly separable, the decision function could be solved by simply using this separation approach. In this instance, the data set was transferred or projected into a higher dimension, the *Hilbert space*. This process extends the methods of vector algebra from two- or three-dimensional spaces into spaces depicting any finite or infinite number of dimensions. By using the function f with $d_1 < d_2$ the number of possible linear separations has increased:

$$\mathbb{R}^{d1} \to \mathbb{R}^{d2}, x \to \Phi_{(x)} \tag{9}$$

SVM is conducive to this operation as the training data, x_i, only emerge in scalar products in terms of the optimization problem. The scalar product, $x_i\,(x)$ is calculated in the higher-dimensional space $f_{(xi)}\,(f_{(x)})$. The transfer was conducted with the use of a kernel function, k, according to Mercer's theorem (Burges, 1998):

$$k_{(x_i,x)} = \left\langle \Phi_{(x_i)}, \Phi_{(x)} \right\rangle \tag{10}$$

The Gaussian radial basis kernel function (RBF) was a reasonable first choice (Waske, van der Linden, Benediktsson, Rabe, & Hostert, 2010; Xie, 2006). RBF kernels are the most generalized form of kernelization

and are one of the most widely used kernels as they are similar to the Gaussian distribution (Waske et al., 2010; Xie, 2006). For two points, x and x_i, the RBF kernel function computes the similarity of the points. This kernel may be mathematically represented as

$$k_{(x_i,x)} = e^{-\gamma|x-x_i|^2} \tag{11}$$

The parameter γ defines the width of the Gaussian kernel function, where $|x - x_i|$ is the Euclidean (L_2-norm) distance between two points, x and x_i. After the kernel function, the hyperplane is set in a Hilbert space, and the decision function becomes.

$$\mathrm{sgn} f_{(x)} = \left[\mathrm{sgn}\sum_{i=1}^{n} a_i y_i k\left(x_i, x+b\right)\right] \tag{12}$$

The class probability was calculated (Equation 12), as opposed to directly predicting the class label; this is the basis for the probability maps of urban sprawl. Platt (1999) approximated the probabilities for a binary SVM using a sigmoid function,

$$P(y=1|x) = \frac{1}{1+e^{A+f_{(x)}B}}, \tag{13}$$

where A and B are parameters estimated by minimizing the negative log-likelihood function (Platt, 1999; Wu, Lin, & Weng, 2004). The n-fold cross-validation procedure is considered an effective method to balance the accuracy of "known" training data with "unknown" testing data (Hsu, Chang, & Lin, 2010). According to the "curse of dimensionality" and the Hughes phenomenon, which describes the degradation of classifier performance with an increasing number of features, selecting the optimal feature combination has been recommended (Hughes, 1968). This selection of relevant features may improve the prediction ability, generalization performance, and computational efficiency of an SVM model (Mewes, 2011; Nguyen & de la Torre, 2010). For this study, adept feature selection offers additional insights into the impacts of the various driving forces underpinning the local suitability of urban sprawl. A common element of SVM feature selection is forward feature selection (Hsu et al., 2010; Waske et al., 2010), which initially trains each feature of the input feature set. Here, the best-performing feature, the one with the lowest error rate, is selected (see Table 6), and remaining features are used for training in combination with the initially selected feature. This procedure is repeated until all features have been selected. The result is a functional ranking of different feature combinations, whereby features that weaken the SVM classifier may be eliminated.

3.4 | Random forest

In recent years, ensemble classifiers (e.g., bagging or boosting) have become increasingly attractive in remote sensing, as they are more robust and accurate compared to single classifiers (Gislason, Benediktsson, & Sveinsson, 2006). RF was developed by Breiman (2001) and is an ensemble classifier based on a set of classification and regression trees (CART). Bagging (or bootstrap aggregating) uses a bootstrapped data set considering all possible features for each node. In contrast, RF only uses a random subset of variables; as such, it may be described as advanced bagging (Bosch, Zisserman, & Munoz, 2007; Liaw & Wiener, 2002). Each DT independently assigns a class to a cell (i.e. each tree has one vote), and the majority of votes defines the final class (Breiman, 2001).

Compared to other ensemble classifiers (e.g. boosting), RF is computationally simple and rapid while achieving similar accuracy. It can handle large numbers of variables without expanding the computation time required

(Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012); the computation time required for an RF model is:

$$T\sqrt{MN\log(N)} \tag{14}$$

where T is the number of trees, M is the number of variables used for each split, and N is the number of training samples (Breiman, 2001). RF is also advantageous as it does not overfit data and is relatively robust to outliers and noise (Breiman, 2001; Rodriguez-Galiano et al., 2012). As there are only two selectable parameters (*ntree* and *mtry*), RF is very user-friendly (Liaw & Wiener, 2002) and can thus be used to estimate variable importance (Liaw & Wiener, 2002); however, it is still a black-box model (Gislason et al., 2006).

An RF contains *n*, individual binary trees (*ntree*). The default suggestions for *ntree* are most commonly 100 (e.g., Bosch et al., 2007; Eisavi, Homayouni, Yazdi, & Alimohammadi, 2015), or 500 (e.g., Elhadi, Mutanga, Odindi, & Abdel-Rahman, 2014; Gislason et al., 2006). An RF tree is constructed using a bootstrapped data set. It includes randomly selected cases (cells) from the training data set, allowing replacement and duplicates. Typically, approximately two thirds of data is used for building the trees (Liaw & Wiener, 2002). At each node, a random subset (*mtry*) out of all variables M (i.e., bands) is considered to split data, whereby all trees are fully grown without pruning (Breiman, 2001). This leads to uncorrelated trees with reduced computational capacity (Gislason et al., 2006). The default value for *mtry* is the square root of all variables, $\sqrt{P}$ (Breiman, 2001). RF uses the Gini index as a measurement of impurity to select the best splitting feature at each node. For a given training data set, T, this may be expressed as:

$$\sum \sum_{j \neq i} \left(f\left(C_i, T\right) / |T|\right)\left(f\left(C_j, T\right) / |T|\right) \tag{15}$$

with a probability $f\left(C_i, T\right) / |T|$ that the cell belongs to class C_i (Pal, 2005).

As only approximately two thirds of the training data builds up the bootstrapped data sets, the remaining one third (the "out-of-bag" (OOB) samples) may be used to validate the RF (Breiman, 2001). Running OOB samples down every tree, which they were not previously part of, reveals the OOB error (cross-validation), and the proportion of OOB samples incorrectly classified (Gislason et al., 2006). Therefore, it is possible to identify the best *ntree* and *mtry* combination with the lowest error rate (grid search approach; Adam et al. 2014).

RF is also able to estimate the importance of a variable, *m* (i.e., a spectral band). This is such that it can permute all values of the *m*th input variable and consider the change in class prediction. Breiman (2001) offered four different options to examine the amount of change, where the most commonly used option has been the increase in the OOB error (e.g., Gislason et al., 2006; Liaw & Wiener, 2002). Conducting this on every tree provides a reliable indication of the information content for the *m*th variable.

The end-product of an RF is a classification map that shows the predicted class for each cell based on the majority of votes. As a by-product, it also generates a probability map for each class, which indicates the likelihood that each cell belongs to a certain class. A high prediction rate results in a high class probability, and vice versa. Thus, information regarding the uncertainty of the classification is also provided (Harris & Grunsky, 2015).

3.5 | Parametrization of RF and SVM classifiers

Both classifiers were tested on the study area using either 5,000, 10,000, 15,000, or 20,000 sample cells; each sample cell consisted of 50% urban change and non-change cells. The driving forces (Table 1) defined the feature space in which the two classes, urban change and non-urban change, were discriminated. After reviewing

TABLE 2 RF parametrization

Features/bands	10 (see Table 1), 9 (without zoning)
Randomly selected features:	3
Training samples	5,000, 10,000, 15,000, or 20,000
Trees	100
Classes	2 (0, 1)
Impurity function	Gini coefficient

preliminary results, it became apparent that the zoning variable dominated the classifications in all combinations. As such, a set of classifications excluding this variable was also computed.

The settings for the RF and SVM classifiers were kept at the default setting, resulting in the parametrization detailed in Tables 2 and 3.

4 | RESULTS

4.1 | Machine-learning-based probability maps of urban sprawl

In addition to the classification, the output of the parametrization is a probability map depicting the likelihood of a cell transforming to urban; these are the main input factor for CA. Figure 2 shows the probability map based on SVM, RF, and the differences between SVM and RF for urbanized areas. This figure presents high SVM values with concurrent low RF (blue) values, or low SVM values concurrent with high RF values. The difference map reflects that many urbanized areas may be overlooked by SVM, while RF still shows possibilities of change from non-urban to urban.

Another key difference is the transition of areas with high probabilities to areas with low probabilities. Although the RF map shows smooth, gradual transitions covering extensive areas around agglomerations, the SVM map is characterized by strong, sharp edges. This first impression was confirmed by inspecting the histograms of both maps (Figure 3). The SVM model contained fewer probability values between the maximum and minimum characteristics, while the RF model computed more probabilities. The most interesting difference between both histograms was the steep frequencies around the minimum and maximum of the SVM model. Additionally, it was observed that the RF model rounds to two decimal places.

4.2 | Validation of probability maps

ROC is an approved index to assess the accuracy of binary categorical probability estimations (Pontius & Schneider, 2001). The ROC divides probability outcomes into percentile groups from high to low probability and compares individual probability groups with cumulative values. The ROC only considers positive values estimated by the model; in this study, all cells were labeled as changed to urban.

Table 4 presents the results for the SVM-based and RF-based probability maps. We tested four different samples with and without zoning information for both ML classifiers. It was clear that the RF model outperformed the SVM model for five different combinations, while only one RF model had underperformed five SVM models; however, this model did not exhibit the poorest performance. We observed that the higher the amount of training samples, the better the performance of the ML model in allocating cells that were changing to urban. The ROC results show that zoning provides valuable information for ML models in terms of forecasting urban sprawl. This

TABLE 3 SVM parametrization

Training data				
Number of samples	5,000, 10,000, 15,000, or 20,000			
Number of features	10 (see Table 1), 9 (without zoning)			
Number of classes	2 (0, 1)			
Parameter search settings (5,000)				
Gaussian RBF kernel parameters *g*	0.00–1000, multiplier 10			
Regularization parameters *C*	0.01–1000, multiplier 10			
Number of cross-validation folds	3			
Termination criterion	0.1			
Model parameters				
Gaussian RBF kernel parameters *g*:				
	Sample size			
	5,000	**10,000**	**15,000**	**20,000**
With zoning	1.0	0.01	0.1	1.0
Without zoning	0.1	1.0	1.0	1.0
Regularization parameters *C*:				
	Sample size			
	5,000	**10,000**	**15,000**	**20,000**
With zoning	1.0	1,000	10	1.0
Without zoning	10	1.0	1.0	1.0
Termination criterion: 0.001				
Number of support vectors:				
	Sample size			
	5,000	**10,000**	**15,000**	**20,000**
With zoning	3,949	4,752	7,395	12 850
Without zoning	3,530	7,892	11 317	14 699
Classwise support vectors (class 0, class 1):				
	Sample size			
	5,000	**10,000**	**15,000**	**20,000**
With zoning	2,075; 1,874	2,406; 2,346	3,748; 3,647	6,800; 6,150
Without zoning	1,788; 1,742	4,166; 3,726	5,983; 5,334	7,753; 6,946

holds true for all combinations except for SVM–10,000; this was the only model where the model fed with zoning information was outperformed by the model without this information.

4.3 | Variable importance and feature selection

The findings of the ROC assessment were confirmed by examining the importance of the variables that were used for model calibration. Tables 5 and 6 present the results for models with a training sample size of 20,000,

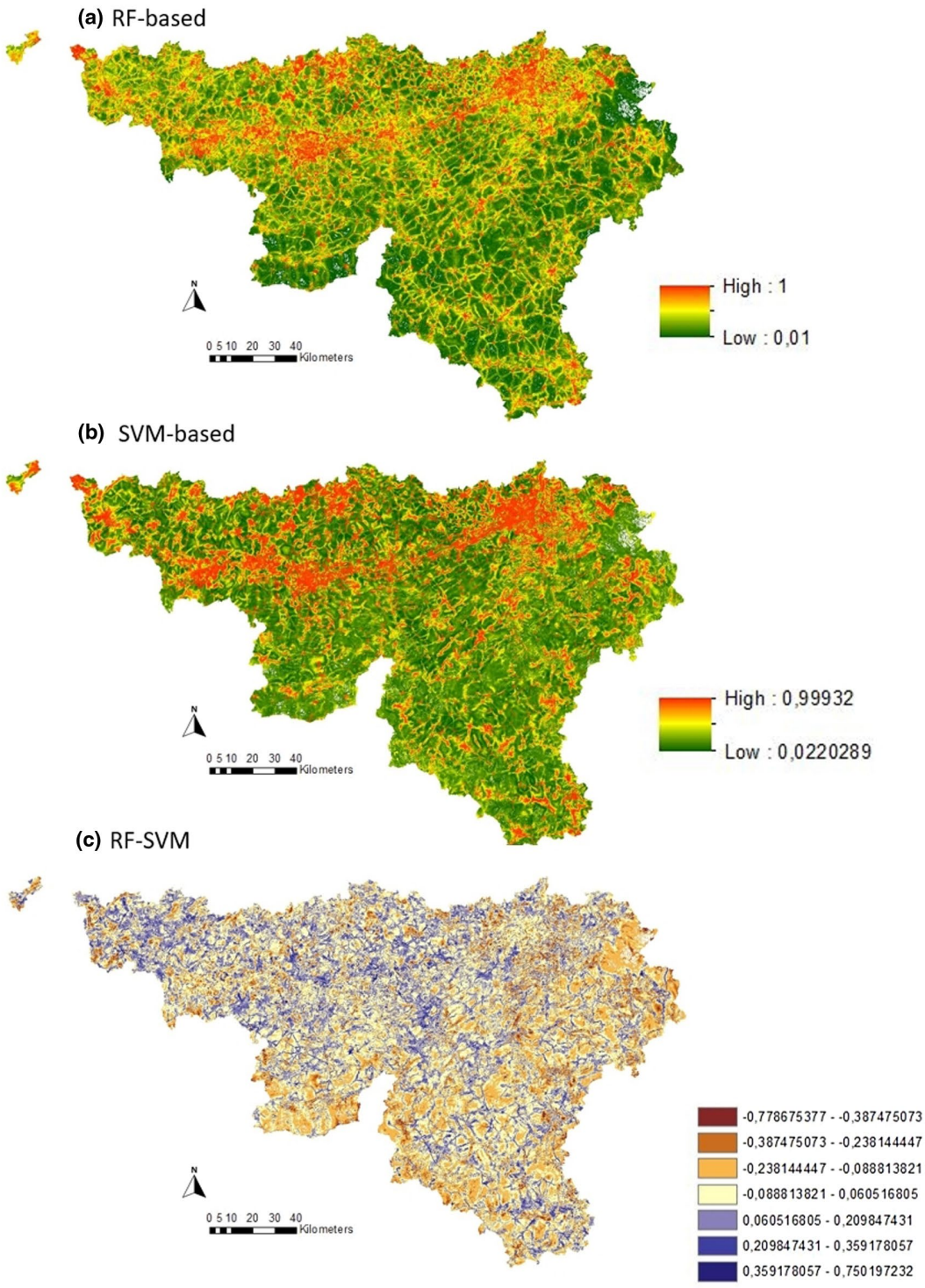

FIGURE 2 (a) RF-based and (b) SVM-based probability map of urban sprawl, and (c) their differences (RF–SVM, bottom)

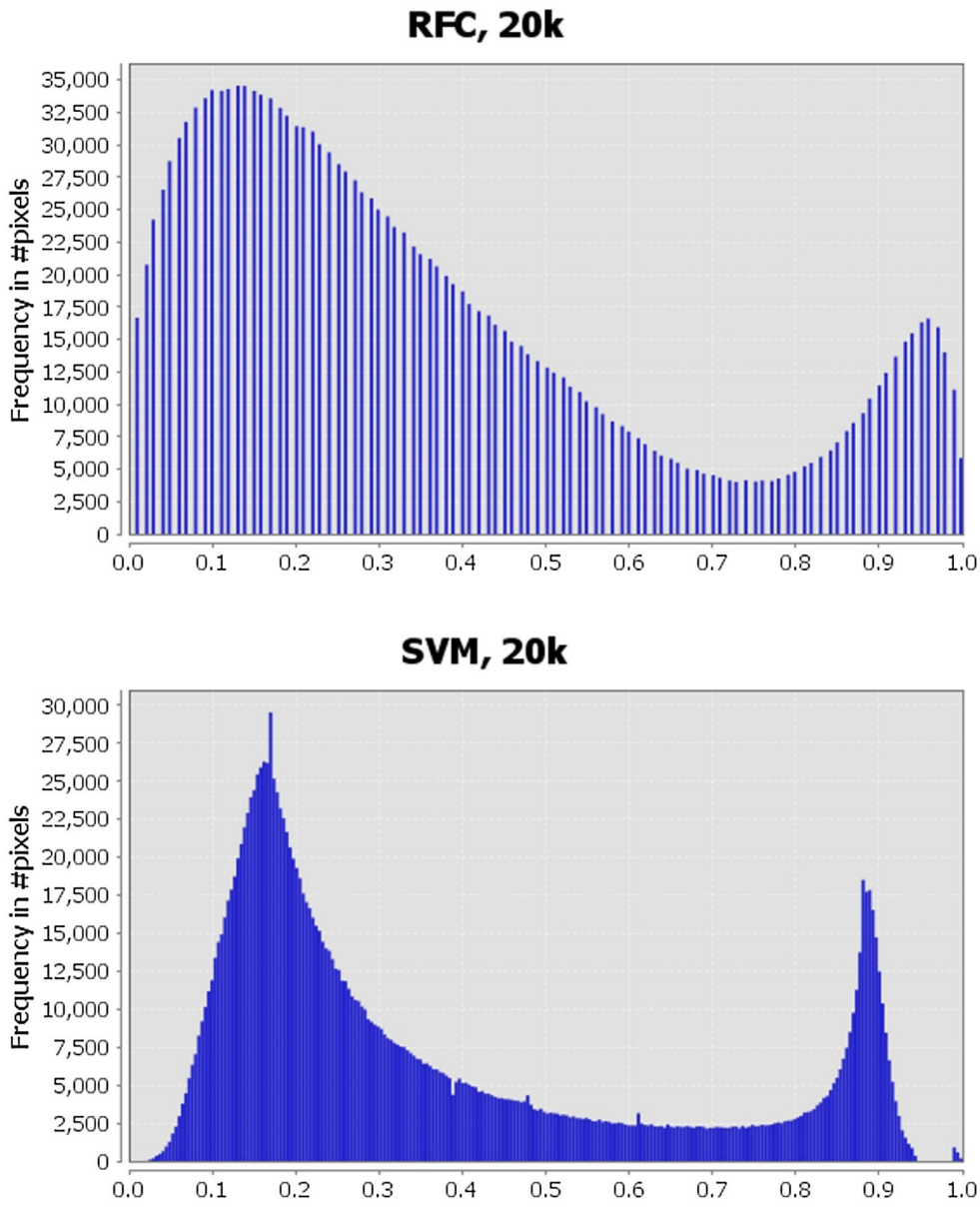

FIGURE 3 Histograms of the probability maps (derived with 20,000 training cells)

including models with zoning information. In both models, "zoning" was the most important variable; Figures 4 and 5 demonstrate the importance of this information, particularly for the RF model where the variable explains most of the allocation characteristics. This was followed by all distance information, where the distance to the next city was more important for the RF model than for its SVM counterpart. Here, the distance to different road types was more important, and both distance variables represent access to markets. Less important was the geophysical information on slope and elevation. Interestingly, the SVM model performance had improved when the road type reflecting highways was introduced, while the RF model prioritized secondary roads before highways, and local roads before main roads. This underlines the visual impression of the probability maps (Figure 2); the RF model-based probability map tends to a broader allocation area, while the SVM model-based probability map was more definitive in the selection of "urban change" or "non-urban change".

TABLE 4 ROC validation of probability maps

Rank	Probability map (input CA)	Sample size	Zoning	ROC
1	RF	20,000	Yes	0.972
2	RF	20,000	No	0.967
3	RF	15,000	Yes	0.948
4	RF	15,000	No	0.931
5	RF	10,000	Yes	0.920
6	SVM	20,000	Yes	0.898
7	RF	5,000	Yes	0.888
8	RF	10,000	No	0.885
9	SVM	15,000	Yes	0.861
10	SVM	20,000	No	0.847
11	SVM	5,000	Yes	0.838
12	SVM	15,000	No	0.826
13	RF	5,000	No	0.822
14	SVM	10,000	No	0.803
15	SVM	5,000	No	0.750
16	SVM	10,000	Yes	0.676

TABLE 5 Variable importance in RF model (sample size 20,000, with zoning information)

Variable	Raw	Normalized	*SD*	Variable/band name
1	14.55	20.93	0.70	Zoning
9	4.26	7.3	0.58	Dist large cities
8	2.95	4.66	0.63	Dist med cities
7	2.34	4.28	0.55	Dist rail
4	2.66	4.13	0.64	Dist RD3
6	2.80	3.99	0.70	Dist RD1
3	2.88	3.75	0.77	RD4
5	2.77	3.54	0.78	RD2
2	2.80	3.22	0.87	Slope
10	1.24	2.39	0.52	DEM

4.4 | Validation of cellular automata

The fuzzy similarity rate (FSR), Equation 3, was used to evaluate CA model performance. Table 7 shows that the FSR increased with the addition of the zoning variable for the low sample size (5,000), and decreased when the zoning variable was used for a larger sample size (15,000 and 20,000). For the SVM, the CA model with zoning was consistently better than the model without zoning; however, it does not significantly improve by increasing sample size, and note that the 5,000 sample outperformed the 15,000 sample. These results are interesting as they contradict the observations of the probability map assessment (see Section 4.1). The RF probability maps also showed performance improvements as soon as zoning information was included. The ML-based probability maps set a suitable pattern to allocate urban cells with the CA in a temporal dynamic manner; this was annual. The

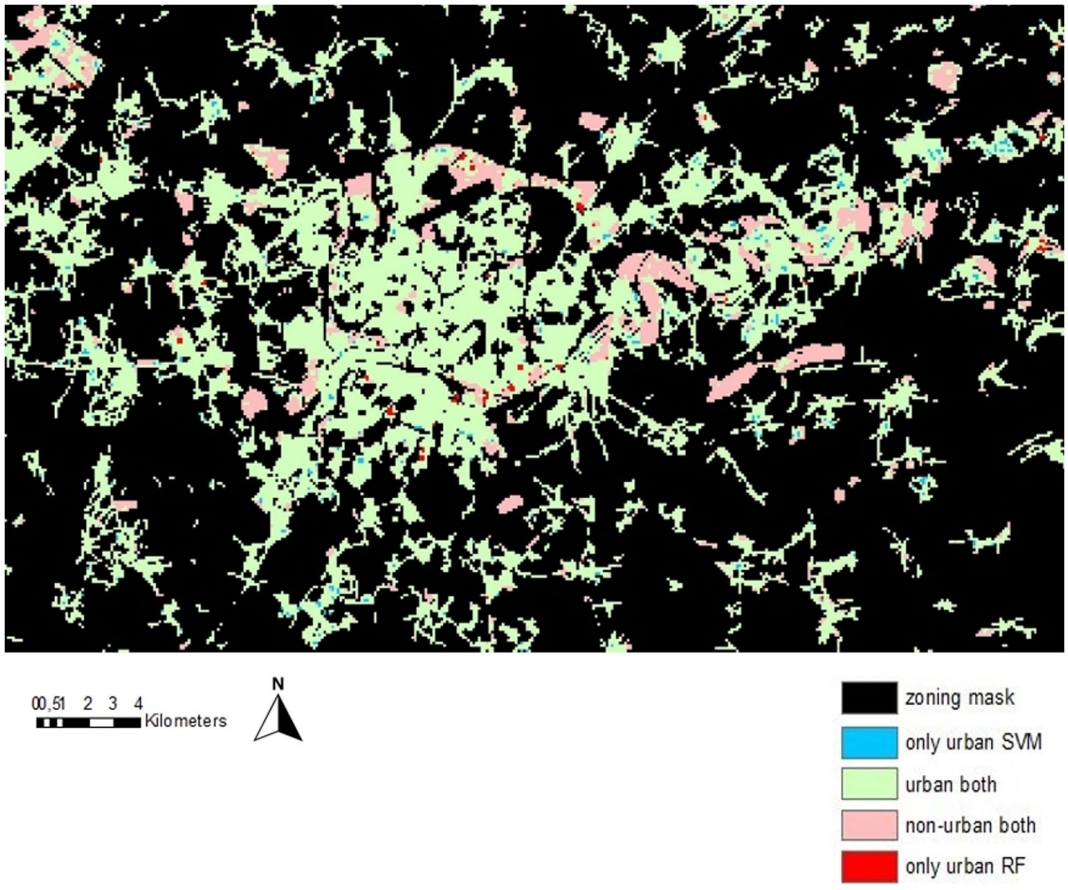

FIGURE 4 Comparison of CA results RF–20,000–no-zoning and SVM–20,000–zoning

CA determined which cells would become urbanized; as such, the performance of different training area numbers based on the inclusion or exclusion of zoning information in the probability maps is key if introduced to a cellular automaton. The inclusion of zoning information as an additional input for the CA itself was not tested, having already been examined in several studies (Kantakumar et al., 2016).

4.5 | Cellular automata outcome

Figure 4 compares the two CA results for models with the best performance for each ML algorithm, which were based on 20,000 training areas. While RF performed the best without zoning, the best SVM performance included zoning information. It was observed that most new urban areas were allocated in gaps close to existing urban patches, reflecting an edge or spread growth type. There were almost no urban areas identified in the rural regions.

We tested 16 models in total, which constitutes eight models per ML algorithm. The study reflects a high intensity of sensitivity analyses due to the variations in ML algorithms, sample sizes, and zoning information. The latter produced unexpected results in terms of the difference in probability map performance following implementation in the CA. However, this testing approach lacked differentiation of calibration and validation data. As we did not use independent validation data from another time slice, unbiased validation was not possible. In these

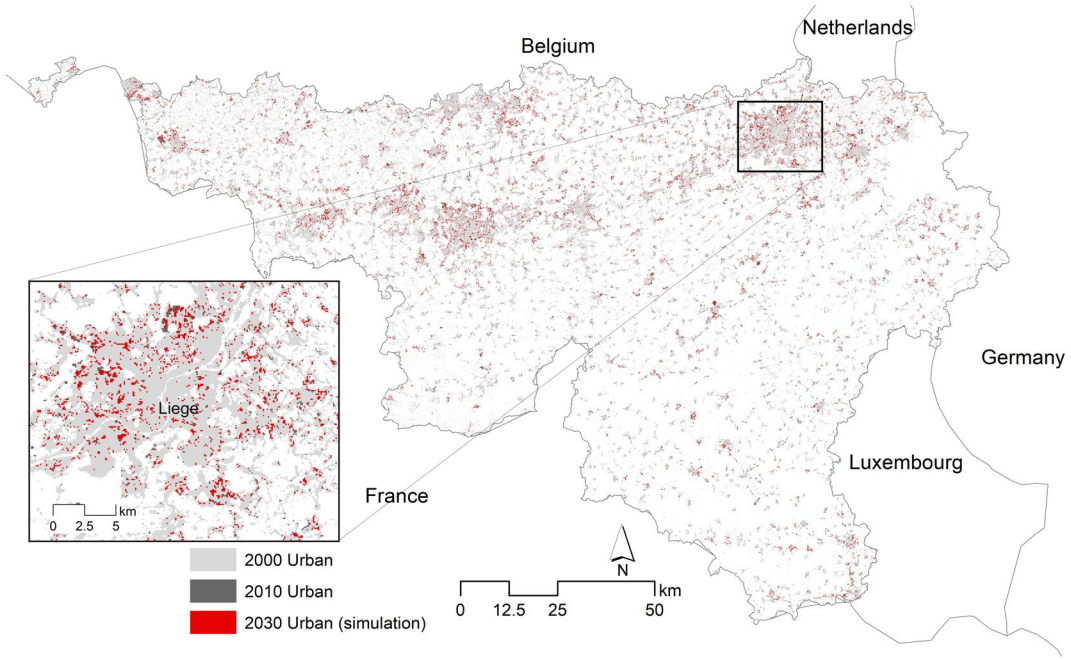

FIGURE 5 The 2030 simulation based on the RF probability map with a sample size of 20,000 and the exclusion of zoning information

TABLE 6 Forward feature selection in SVM model (20,000; with zoning information)

Variable/band name	Feature ranking	Performance
Zoning	1	75.85
Dist large cities	2	76.07
Dist RD1	3	76.78
Dist RD2	4	77
Dist med cities	5	77.27
Dist rail	6	77.81
Dist RD3	7	78.14
Dist RD4	8	78.26
DEM	9	77.88
Slope	10	78.34

terms, it may be postulated that ML approaches do not use the entire sample for training. On the contrary, they keep a small amount of the sample for testing, such that probability maps are generated without bias; this was the main purpose of this study.

The model can be used as a predictive tool to simulate future patterns based on past "calibrated"' trends. Figure 5 illustrates the projected urban expansion in 2030. This pattern corresponds to the business-as-usual scenario that extrapolates past trends in terms of quantity and location of changes observed between 2000 and 2010. The simulation shows that the study area will experience urban development within urban cores, and rural and remote locations, causing increased sprawl over time.

TABLE 7 FSR of the implemented CA

Probability map (input CA)	Sample	Zoning	FSR
RF	20,000	No	47.821
RF	20,000	Yes	45.974
RF	15,000	No	45.453
RF	15,000	Yes	44.531
RF	10,000	Yes	42.923
RF	5,000	Yes	41.809
RF	10,000	No	41.549
SVM	20,000	Yes	38.665
RF	5,000	No	38.603
SVM	5,000	Yes	38.217
SVM	15,000	Yes	37.493
SVM	20,000	No	36.502
SVM	15,000	No	35.483
SVM	10,000	No	34.884
SVM	10,000	Yes	34.433
SVM	5,000	No	32.681

5 | CONCLUSIONS AND OUTLOOK

This article contributes to the few studies that have calibrated the transition rules of CA models using ML algorithms in an effort to enhance their performance in simulating urban sprawl. We assessed and compared the performance of CA–SVMs and CA–RF models for modeling urban sprawl. Coupling CA models with ML algorithms enables the simultaneous, dynamic simulation of urban sprawl processes alongside the analyses of multiple controlling factors that determine change suitability. While the model developed was applied to Wallonia (Belgium) as a case study, it is generic and may be applied to other case studies. For application to other studies, an investigation into the transferability of model parameters is an interesting direction for future research.

We have examined two main aspects of the sensitivity of the ML-based probability maps of urban sprawl: the input sample size the algorithms were trained on; and the effect of including zoning information as an independent variable. The results show that RF-based probabilities outperformed those derived by SVMs in almost all sampling–zoning combinations. This is particularly important for case studies with a small number of land-use changes (sample size). Our analysis reveals that the RF-based CA model still yields good results with low sample size. Furthermore, the RF smooths the edges between low and high land-use change probability areas more efficiently than SVM. We rejected the hypothesis that probability maps are unlikely to differ and the sensitivity analysis of implemented CA models would not exhibit substantial disparities. However, both ML classifiers were suitable for improving the performance of a CA in metropolitan regions. The RF- and SVM-driven CA performed better in simulating urban sprawl with larger sample sizes and the inclusion of zoning information; this did not hold true after the introduction into the CA. The CA results demonstrated that RF probability maps that had high sample sizes and lacked zoning information produced more reliable CA results with higher allocation accuracy than RF probability maps including zoning information.

Although ML enriches the calibration methods of CA modeling urban sprawl, there are limitations to this method as the algorithms are relatively complex in theory and implementation. Moreover, due to the black-box nature of CA models, they do not enable an evaluation of the relative contribution of each independent variable

on urban sprawl after being introduced into the CA. As such, further research within the urban sprawl modeling domain is required to improve the current understanding of the requisite mathematical and computer knowledge for ML algorithms.

ACKNOWLEDGMENTS

This material was supported by US National Science Foundation "GCR: Social, Ecological, and Technological Infrastructure Systems for Urban Resilience" project (Award #1934933).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Due to the nature of this research, the participants in this study did not agree for their data to be shared publicly; as such, supporting data are not available.

ORCID

Andreas Rienow https://orcid.org/0000-0003-3893-3298
Ahmed Mustafa https://orcid.org/0000-0002-1592-6637
Claudia Lindner https://orcid.org/0000-0001-5318-7997

REFERENCES

Abolhasani, S., & Taleai, M. (2020). Assessing the effect of temporal dynamics on urban growth simulation: Towards an asynchronous cellular automata. *Transactions in GIS*, *24*, 332–354. https://doi.org/10.1111/tgis.12601

Alaei Moghadam, S., Karimi, M., & Habibi, K. (2018). Simulating urban growth in a megalopolitan area using a patch-based cellular automata. *Transactions in GIS*, *22*(1), 249–268. https://doi.org/10.1111/tgis.12309

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: MIT Press.

Batty, M., Barros, J., & Júnior, S. A. (2006). Continuity, transformation and emergence. In E. Garnsey & J. McGlade (Eds.), *Complexity and co-evolution* (pp. 61–76). Cheltenham, UK: Edward Elgar Publishing.

Belgian Federal Government. (2013). *Statistics Belgium*. Retrieved from http://statbel.fgov.be/fr/statistiques/chiffres/

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. In O. Carugo & F. Eisenhaber (Eds.), *Data mining techniques for the life sciences* (pp. 223–239). Totowa, NJ: Humana Press.

Benenson, I., & Kharbash, V. (2005). Geographic automata: From paradigm to software and back to paradigm. In *Proceedings of the Eighth International Conference on GeoComputation*, Ann Arbor, MI (pp. 1–12).

Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. Chichester, UK: John Wiley & Sons.

Bosch, A., Zisserman, A., & Munoz, X. (2007). Image classification using random forests and ferns. In *Proceedings of the 11th IEEE International Conference on Computer Vision*, Shenzhen, China (pp. 1–8). Piscataway, NJ: IEEE.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

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

Cao, M., Huang, M., Xu, R., Lü, G., & Chen, M. (2019). A grey wolf optimizer–cellular automata integrated model for urban growth simulation and optimization. *Transactions in GIS*, *23*(4), 672–687. https://doi.org/10.1111/tgis.12517

Chang, C.-C., & Lin, C.-J. (2001). *LIBSVM: A library for support vector machines*. Retrieved from https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf

Charif, O., Omrani, H., Abdallah, F., & Pijanowski, B. (2017). A multi-label cellular automata model for land change simulation. *Transactions in GIS*, *21*(6), 1298–1320. https://doi.org/10.1111/tgis.12279

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Couclelis, H. (1989). Macrostructure and microbehavior in a metropolitan area. *Environment and Planning B*, *16*(2), 141–154. https://doi.org/10.1068/b160141

Eisavi, V., Homayouni, S., Yazdi, A. M., & Alimohammadi, A. (2015). Land cover mapping based on random forest classification of multi-temporal spectral and thermal images. *Environmental Monitoring and Assessment*, *187*(5), 291. https://doi.org/10.1007/s10661-015-4489-3

Elhadi, A., Mutanga, O., Odindi, J., & Abdel-Rahman, E. M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, *35*(10), 3440–3458. https://doi.org/10.1080/01431161.2014.903435

European Environment Agency. (2006). *Urban sprawl in Europe: The ignored challenge* (EEA Report No. 10/2006). Luxembourg: Office for Official Publications of the European Communities.

Feng, Y., & Tong, X. (2018). Calibration of cellular automata models using differential evolution to simulate present and future land use. *Transactions in GIS*, *22*(2), 582–601. https://doi.org/10.1111/tgis.12331

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Harris, J. R., & Grunsky, E. C. (2015). Predictive lithological mapping of Canada's North using random forest classification applied to geophysical and geochemical data. *Computers & Geosciences*, *80*, 9–25. https://doi.org/10.1016/j.cageo.2015.03.013

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2010). *A practical guide to support vector classification*. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Huang, B., Xie, C., & Tay, R. (2010). Support vector machines for urban growth modeling. *GeoInformatica*, *14*(1), 83–99. https://doi.org/10.1007/s10707-009-0077-4

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, *14*(1), 55–63. https://doi.org/10.1109/TIT.1968.1054102

Intergovernmental Panel on Climate Change. (2012). *Renewable energy sources and climate change mitigation: Special report of the Intergovernmental Panel on Climate Change*. New York, NY: Cambridge University Press.

IT.NRW. (2013). *Landesdatenbank NRW*. Retrieved from https://www.it.nrw/

Kamusoko, C., & Gamba, J. (2015). Simulating urban growth using a random forest cellular automata (RF-CA) model. *ISPRS International Journal of Geo-Information*, *4*, 447–470. https://doi.org/10.3390/ijgi4020447

Kantakumar, L. N., Kumar, S., & Schneider, K. (2016). Spatiotemporal urban expansion in Pune metropolis, India using remote sensing. *Habitat International*, *51*, 11–22. https://doi.org/10.1016/j.habitatint.2015.10.007

Karimi Firozjaei, M., Sedighi, A., & Jelokhani-Niaraki, M. (2020). An urban growth simulation model based on integration of local weights and decision risk values. *Transactions in GIS*, *24*(6), 1695–1721. https://doi.org/10.1111/tgis.12668

Lavalle, C., Demicheli, L., & Kasanko, M. (2002). *Towards an urban atlas: Assessment of spatial data on 25 European cities and urban areas* (Environmental Issue Report Vol. 30). Luxembourg: Office for Official Publications of the European Communities.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Lin, J., & Li, X. (2016). Knowledge transfer for large-scale urban growth modeling based on formal concept analysis. *Transactions in GIS*, *20*(5), 684–700. https://doi.org/10.1111/tgis.12172

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Mewes, T. (2011). *The impact of the spectral dimension of hyperspectral datasets on plant disease detection*. Bonn, Germany: Universitäts- und Landesbibliothek Bonn.

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Washington, DC: Island Press.

Miller, E. J., Hunt, J. D., Abraham, J. E., & Salvini, P. A. (2004). Microsimulating urban systems. *Computers, Environment and Urban Systems*, *28*(1–2), 9–44. https://doi.org/10.1016/S0198-9715(02)00044-3

Mirbagheri, B., & Alimohammadi, A. (2017). Improving urban cellular automata performance by integrating global and geographically weighted logistic regression models. *Transactions in GIS*, *21*(6), 1280–1297. https://doi.org/10.1111/tgis.12278

Momeni, E., & Antipova, A. (2020). Pattern-based calibration of cellular automata by genetic algorithm and Shannon relative entropy. *Transactions in GIS*, *24*(6), 1447–1463. https://doi.org/10.1111/tgis.12646

Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., & Teller, J. (2018). Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computers, Environment and Urban Systems*, *67*, 147–156. https://doi.org/10.1016/j.compenvurbsys.2017.09.009

Mustafa, A., Rienow, A., Saadi, I., Cools, M., & Teller, J. (2018). Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. *European Journal of Remote Sensing*, *51*, 391–401. https://doi.org/10.1080/22797254.2018.1442179

Mustafa, A., & Teller, J. (2020). Self-reinforcing processes governing urban sprawl in Belgium: Evidence over six decades. *Sustainability*, *12*, 4097. https://doi.org/10.3390/su12104097

Nguyen, M. H., & de la Torre, F. (2010). Optimal feature selection for support vector machines. *Pattern Recognition*, *43*(3), 584–591. https://doi.org/10.1016/j.patcog.2009.09.003

Okwuashi, O., McConchie, J., Nwilo, P., & Eyo, E. (2009). Stochastic GIS cellular automata for land use change simulation: Application of a kernel based model. In *Proceedings of the 10th International Conference on GeoComputation*. Sydney, NSW, Australia (pp. 1–7).

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large-margin classifiers* (pp. 61–74). Cambridge, MA: MIT Press.

Pontius, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, *85*(1–3), 239–248. https://doi.org/10.1016/S0167-8809(01)00187-6

Ramankutty, N., Graumlich, L., Achard, F., Alves, D., Chhabra, A., DeFries, R. S., & Turner, B. L. (2006). Global land-cover change: Recent progress, remaining challenges. In E. F. Lambin & H. Geist (Eds.), *Land-use and land-cover change* (pp. 9–39). Berlin, Germany: Springer.

Rienow, A., & Goetzke, R. (2015). Supporting SLEUTH—Enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, *49*, 66–81. https://doi.org/10.1016/j.compenvurbsys.2014.05.001

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rucker, R. (1999). *Seek!: Selected nonfiction*. New York, NY: Four Walls Eight Windows.

Samardžić-Petrović, M., Dragićević, S., Kovačević, M., & Bajat, B. (2016). Modeling urban land use changes using support vector machines. *Transactions in GIS*, *20*(5), 718–734. https://doi.org/10.1111/tgis.12174

Silva, E. A. (2004). The DNA of our regions: Artificial intelligence in regional planning. *Futures*, *36*(10), 1077–1094. https://doi.org/10.1016/j.futures.2004.03.014

Silva, E. A. (2011). Cellular automata and agent base models for urban studies: From pixels to cells to hexa-dpi's. In X. Yang (Ed.), *Urban remote sensing: Monitoring, synthesis and modeling in the urban environment* (pp. 323–334). New York, NY: John Wiley & Sons.

Silva, E. A., & Clarke, K. C. (2005). Complexity, emergence and cellular urban models: Lessons learned from applying SLEUTH to two Portuguese metropolitan areas. *European Planning Studies*, *13*(1), 93–115. https://doi.org/10.1080/0965431042000312424

Thomas, I., Frankhauser, P., & Biernacki, C. (2008). The morphology of built-up landscapes in Wallonia (Belgium): A classification using fractal indices. *Landscape and Urban Planning*, *84*(2), 99–115. https://doi.org/10.1016/j.landurbplan.2007.07.002

Ulam, S. (1952). Random processes and transformations. In *Proceedings of the International Congress on Mathematics*, Cambridge, UK (pp. 264–275).

United Nations Environment Programme. (2012). *Global environment outlook: GEO4, Environment for development*. Nairobi, Kenya: UNEP.

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

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

Verburg, P. H., Kok, K., Pontius, R. G., & Veldkamp, A. (2006). Modeling land-use and land-cover change. In E. F. Lambin & H. Geist (Eds.), *Land-use and land-cover change* (pp. 117–135). Berlin, Germany: Springer.

von Neumann, J. (1951). The general and logical theory of automata. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon Symposium* (pp. 1–31). New York, NY: John Wiley & Sons.

Wang, R., Feng, Y., Wei, Y., Tong, X., Zhai, S., Zhou, Y., & Wu, P. (2020). A comparison of proximity and accessibility drivers in simulating dynamic urban growth. *Transactions in GIS*, *25*(2), 1–25. https://doi.org/10.1111/tgis.12707

Waske, B., van der Linden, S., Benediktsson, J. A., Rabe, A., & Hostert, P. (2010). Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(7), 2880–2889. https://doi.org/10.1109/TGRS.2010.2041784

Wassmer, R. W. (2002). *An economist's perspective on urban sprawl: Part 1—Defining excessive decentralization in California and other western states*. Sacramento, CA: California Senate Office of Research.

Wu, N., & Silva, E. A. (2010). Artificial intelligence solutions for urban land dynamics: A review. *Journal of Planning Literature*, *24*(3), 246–265. https://doi.org/10.1177/0885412210361571

Wu, T.-F., Lin, C.-J., & Weng, R. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, *5*, 975–1005. https://doi.org/10.5555/1005332.1016791

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

Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, *34*(6), 592–602. https://doi.org/10.1016/j.cageo.2007.08.003

Zhang, B., Wang, H., He, S., & Xia, C. (2020). Analyzing the effects of stochastic perturbation and fuzzy distance transformation on Wuhan urban growth simulation. *Transactions in GIS*, *24*(6), 1779–1798. https://doi.org/10.1111/tgis.12683

How to cite this article: Rienow A, Mustafa A, Krelaus L, Lindner C. Modeling urban regions: Comparing random forest and support vector machines for cellular automata. *Transactions in GIS*. 2021;00:1–21. https://doi.org/10.1111/tgis.12756