



UNIVERSITY OF LEEDS

This is a repository copy of *Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/122357/>

Version: Accepted Version

Article:

Mustafa, A, Heppenstall, A orcid.org/0000-0002-0663-3437, Omrani, H et al. (3 more authors) (2018) Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computers, Environment and Urban Systems*, 67. pp. 147-156. ISSN 0198-9715

<https://doi.org/10.1016/j.compenvurbsys.2017.09.009>

© 2017 Elsevier Ltd. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm

Ahmed MUSTAFA (Corresponding Author)

LEMA, Urban and Environmental Engineering department, Liège University, Belgium.
Department of Computer Science, Purdue University, USA.

Permanent address:

Liège University, Allée de la Découverte 9, Quartier Polytech 1, 4000 Liège, Belgium
Tel.: +32 43 66 9394

Email: a.mustafa@ulg.ac.be

Present address:

Purdue University, Department of Computer Science, 305 N University St, West Lafayette,
47907 IN, USA.

Tel.: +1 765 250 1765.

Email: a-mustafa@purdue.edu

Alison HEPPENSTALL

School of Geography, University of Leeds, UK

E-mail: a.j.heppenstall@leeds.ac.uk

Hichem OMRANI

Urban Development and Mobility Department, LISER, Luxembourg

E-mail: hichem.omrani@liser.lu

Ismail SAADI

LEMA, Urban and Environmental Engineering department, Liège University, Belgium.

E-mail: ismail.saadi@ulg.ac.be

Mario COOLS

LEMA, Urban and Environmental Engineering department, Liège University, Belgium.

Email: mario.cools@ulg.ac.be

Jacques TELLER

LEMA, Urban and Environmental Engineering department, Liège University, Belgium.

Email: jacques.teller@ulg.ac.be

Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm

Abstract: This paper presents a model to simulate built-up expansion and densification based on a combination of a non-ordered multinomial logistic regression (MLR) and cellular automata (CA). The probability for built-up development is assessed based on (i) a set of built-up development causative factors and (ii) the land-use of neighboring cells. The model considers four built-up classes: non built-up, low-density, medium-density and high-density built-up. Unlike the most commonly used built-up/urban models which simulate built-up expansion, our approach considers expansion and the potential for densification within already built-up areas when their present density allows it. The model is built, calibrated, and validated for Wallonia region (Belgium) using cadastral data. Three 100×100m raster-based built-up maps for 1990, 2000, and 2010 are developed to define one calibration interval (1990-2000) and one validation interval (2000-2010). The causative factors are calibrated using MLR whereas the CA neighboring effects are calibrated based on a multi-objective genetic algorithm. The calibrated model is applied to simulate the built-up pattern in 2010. The simulated map in 2010 is used to evaluate the model's performance against the actual 2010 map by means of fuzzy set theory. According to the findings, land-use policy, slope, and distance to roads are the most important determinants of the expansion process. The densification process is mainly driven by zoning, slope, distance to different roads and richness index. The results also show that the densification generally occurs where there are dense neighbors whereas areas with lower densities retain their densities over time.

Keywords: Built-up density; cellular automata; multinomial logistic regression; multi-objective genetic algorithm

1. Introduction

Built-up development is the most typical form of land-use change. Without policy interventions, built-up developments may cause destructive impacts on the environment, on natural resources and on human health (Zhang et al., 2011). Consequently, modelling built-up development is attracting attention of scientists, urban planners and politicians alike. Most built-up/urban models (e.g. Han and Jia, 2016; Liao et al., 2014; Liu et al., 2014; Puertas et al., 2014; Vermeiren et al., 2012) are raster-based with a coarse cell space ranging from 30×30m to 300×300m. Whilst many authors advocate a larger grid cell for land-use modelling, for example 100×100m (e.g. Jiang et al., 2007; Munshi et al., 2014; Poelmans and Van Rompaey, 2010), land-use cells with these dimensions usually comprise a mix of different land-uses (Omrani et al., 2015). For example, a cell classified as built-up land may be occupied by 80% built-up surface and 20% arable surface. With increases in the spatial resolution of data, researchers have begun to use grid cells as small as 10×10m, such as Berberoğlu et al. (2016) model for Adana city (Turkey). However, the drawback to using such a fine resolution is that it requires intensive computational resources to model larger study areas such as regions where 100×100m cell dimensions are commonly used (e.g. Omrani et al., 2015; Poelmans and Van Rompaey, 2010). One solution to address the trade-off between coarse regular cell spaces and heterogeneity is examining several built-up densities instead of a binary classification (i.e. non built-up/built-up).

Although built-up densification processes, transitions from low-density to high-density, is critically important for policy makers who are concerned with restricting sprawl (Nabielek, 2012; Tachieva, 2010), the literature on urban/built-up expansion models highlights that many of the models focus only on expansion process (e.g. Poelmans and Van Rompaey, 2009; Wang et

23 al., 2013). However, there are a limited number of studies that consider the expansion of several
24 urban densities and/or densification in a variety of ways. Mustafa et al. (2015), Robinson et al.
25 (2012), Sunde et al. (2014), Xian and Crane (2005), Yang (2010) and Zhang et al. (2011) model
26 the expansion of different urban/built-up densities. Crols et al. (2015), Loibl and Toetzer (2003)
27 and White et al. (2015, 2012) model the processes of urban expansion as well as of densification.
28 They define densification as an increase in population and/or several economic sectors density.

29 One of the most popular techniques of existing urban/built-up expansion models which are
30 employed to analyze and/or predict the built-up pattern is cellular automata (CA) (e.g.
31 Berberoğlu et al., 2016; Feng et al., 2011; Han et al., 2009; Tian et al., 2016; Wang et al., 2013).
32 CA is a dynamic discrete space and time bottom-up modelling approach. CA is widely used in
33 urbanization modeling due to its simplicity, transparency and powerful capacities for dynamic
34 spatial simulation (Clarke and Gaydos, 1998). Aburas et al. (2016) and Santé et al. (2010)
35 reviewed CA urbanization models concluding that the CA modelling approach is one of the most
36 appropriate techniques for simulating urban/built-up patterns. However, key challenges in CA
37 are calibrating the transition rules of built-up development probability as a function of (i) a series
38 of causative factors (driving forces) and (ii) spatial (neighborhood) characteristics. Early
39 methods for CA calibration are based on trial and error (e.g. White and Engelen, 1997) and/or a
40 visual test, to determine the model's parameters (e.g. Clarke et al., 1997; Ward et al., 2000).
41 Recently, a variety of automated methods based on statistics (e.g. García et al., 2013), machine
42 learning (e.g. Rienow and Goetzke, 2015), artificial neural networks (e.g. Berberoğlu et al.,
43 2016) and search algorithms for optimization such as genetic algorithms (e.g. Al-Ahmadi et al.,
44 2009) and particle swarm optimization (e.g. Feng et al., 2011) have begun to be widely
45 employed.

46 Validation of CA models is another challenge. A common validation method is based on
47 pixel-by-pixel location agreement (Poelmans and Van Rompaey, 2009). This approach cannot
48 discriminate between “near-miss” and “far-miss” errors which limits its ability to detect spatial
49 patterns (Mustafa et al., 2014). Another approach is based on spatial metrics (Roy Chowdhury
50 and Maithani, 2014). Spatial metrics can be potentially misleading, for example, two areas with
51 distinctly different infrastructures may show the same spatial index (White and Engelen, 2000).
52 A third method is based on a fuzzy set theory. Fuzzy map comparison provides a method of
53 dealing and comparing maps containing a complex mixture of spatial information (Ahmed et al.,
54 2013). It takes into account local variations meaning that matches found at shorter distances are
55 given a higher agreement. It measures the similarity of a cell in a value between 0 (fully-distinct)
56 and 1 (fully-identical). Thus, it can easily distinguish areas of minor errors from areas of major
57 errors. Van Vliet et al., 2016 present a comprehensive survey of calibration and validation
58 practices in land use change modeling.

59 This study contributes to research efforts that model built-up expansion and densification
60 processes. We model the built-up expansion (non built-up to one of built-up density classes) and
61 densification (lower built-up densities to higher ones). The model is based on a hybrid approach
62 which integrates logistic regression and CA modelling approaches. The model is applied to
63 Wallonia (Belgium). Belgian cadastral data (CAD) are used to generate three built-up maps for
64 the years 1990, 2000 and 2010. These maps represent four built-up classes: non built-up (class-
65 0), low-density (class-1), medium-density (class-2) and high-density (class-3). Three maps can
66 define one calibration interval (1990-2000) and one validation interval (2000-2010). The model
67 considers a set of static causative factors related to accessibility, geo-physical features, policies
68 and socio-economic factors. Another important factor is neighborhood interactions because of

69 the fact that urbanization can be regarded as a self-organizing system (Poelmans and Van
70 Rompaey, 2010).

71 The model's parameters are calibrated based on a logistic regression model and genetic
72 algorithm. The logistic regression is employed to set the parameter of 12 built-up development
73 causative factors: elevation, slope, zoning status, employment rate, richness index and Euclidian
74 distances to highways, main roads, secondary roads, local roads, railway stations, large-sized and
75 medium-sized Belgian cities. The richness index is calculated as the average income per capita
76 for each municipality divided by the average income per capita in Belgium. The built-up
77 causative factors are selected according to a literature survey of common factors involved in
78 urban/built-up expansion models (e.g. Achmad et al., 2015; Cammerer et al., 2013; Dubovyk et
79 al., 2011; Li et al., 2013; Poelmans and Van Rompaey, 2010; Verburg et al., 2004) as well as the
80 finding of previous studies conducted for Wallonia (Beckers et al., 2013; Mustafa et al., 2015).
81 The dependent variable for the logistic regression model represents the changes from class-0 to
82 class-1, class-2 or class-3, the changes from class-1 to class-2 and the changes from class-2 to
83 class-3.

84 As the dependent variable is a multi-level, i.e. with more than two possible outcomes, we
85 should consider a non-binary logistic regression. The most common logistic regression types that
86 handle multiple levels of an outcome are ordered logistic regression and multinomial logistic
87 regression. Ordered logistic regression assumes that the levels of dependent status have a natural
88 ordering (i.e. low to high). This is known as the proportional odds model or parallel regression
89 assumption (Kim, 2003). To evaluate this assumption, the test of the proportional odds
90 assumption is performed. The null hypothesis of the test is that the relationship, i.e.
91 coefficients, between each pair of dependent levels is the same. The significance of Chi-Square

92 statistic of the proportional odds test is < 0.001 . Given the assumption of having a natural
93 ordering in the dependent variable is violated, thus a non-ordered multinomial logistic regression
94 model (MLR) is adopted for this study.

95 A multi-objective genetic algorithm (MGA) is employed to calibrate the neighborhood
96 interactions on a dynamic basis. García et al., (2013) reported that the GA is one of the most
97 robust heuristic automated methods to solve optimization problems. A number of studies have
98 used GA to calibrate CA models (e.g. Al-Ahmadi et al., 2009; García et al., 2013; Shan et al.,
99 2008). The MGA objective function is the maximization of allocation accuracy rates for all built-
100 up classes. The accuracy rate function is defined as a fuzzy membership function of exponential
101 decay with a halving distance of two cells and a neighborhood window of four cells. The
102 accuracy rate function is also employed to validate the model.

103 **2. Materials**

104 **2.1 Study area**

105 The model is applied to Wallonia region, the southern part of Belgium. Wallonia occupies an
106 area of 16,844 km² and administratively consists of five provinces: Hainaut, Liège, Luxembourg,
107 Namur, and Walloon Brabant. The total population in 2010 was 3,498,384 inhabitants,
108 corresponding to one third of the Belgium population (Belgian Federal Government, 2013). The
109 population is mainly concentrated on the northern areas, following the 19th century industrial
110 axis, running from east (Liège) to west (Mons) (Thomas et al., 2008). The rest of the territory is
111 less densely inhabited. Consequently, several densities can be easily detected in the region and
112 thus we can examine the transitions between different densities. The built-up development is

113 mainly characterized by low, slow rates, which makes the calibration of the model more difficult
 114 because there is less information on the built-up process (García et al., 2012). The expansion
 115 rates were 1.18% and 0.79% from 1990 to 2000 and from 2000 to 2010, whereas the
 116 densification rates were 12.18% and 9% respectively.

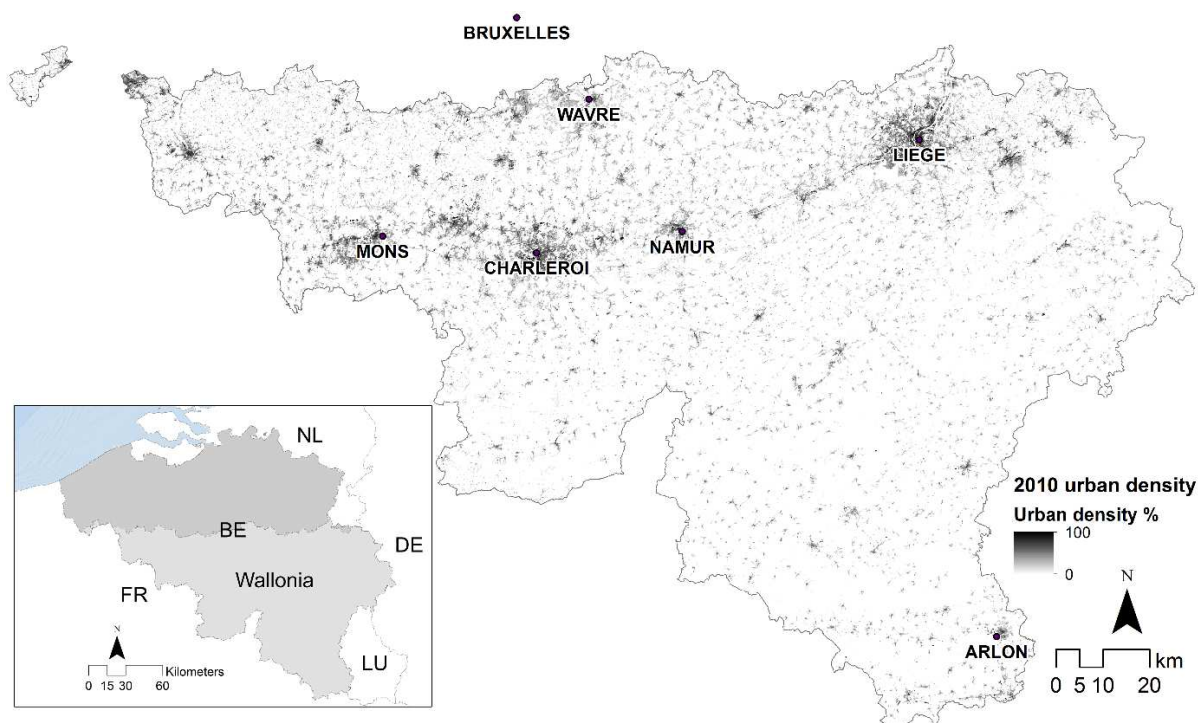


Figure 1: Study area.

117 Table 1 gives the actual built-up transitions over the modeled period for four density classes
 118 (Table 2). As in Xian and Crane (2005), the table suggests that the predominant built-up
 119 processes have been the development of low-density and medium-density areas. The majority of
 120 the new developments have a form of built-up sprawl. This development process had resulted in
 121 a highly fragmented built-up pattern. Table 1 indicates that the transitions from class-1 to class-3
 122 over the study period are marginal. Thus, the densification is considered as the transitions from

123 class-1 to class-2 and from class-2 to class-3, whereas the expansion are the transitions from
 124 class-0 to classes 1, 2 and 3.

Table 1. Class (column) to class (row) changes (% of the reference class).

| 1990-2000 | Class-0 | Class-1 | Class-2 | Class-3 |
|-----------|------------------|----------------|-----------------|---------|
| Class-0 | 1422166 (98.82%) | 0 | 0 | 0 |
| Class-1 | 10841 (0.75%) | 85142 (89.25%) | 0 | 0 |
| Class-2 | 5153 (0.36%) | 10102 (10.59%) | 128929 (98.57%) | 0 |
| Class-3 | 1016 (0.07%) | 151 (0.16%) | 1872 (1.43%) | 25284 |
| 2000-2010 | Class-0 | Class-1 | Class-2 | Class-3 |
| Class-0 | 1410959 (99.21%) | 0 | 0 | 0 |
| Class-1 | 7120 (0.50%) | 88341 (92.04%) | 0 | 0 |
| Class-2 | 3450 (0.24%) | 7535 (7.85%) | 142687 (98.96%) | 0 |
| Class-3 | 637 (0.04%) | 107 (0.11%) | 1497 (1.04%) | 28323 |

125 2.2. Datasets

126 The built-up maps for 1990, 2000 and 2010 are generated based on the Belgian cadastral
 127 database (CAD) in a shapefile format. CAD is provided by the Land Registry Administration of
 128 Belgium. The information contained includes the construction date for each building. CAD
 129 vector data were rasterized at a cell size of 2×2 m. The rasterized cells were then aggregated to a
 130 100×100 m raster-grid. The density values were calculated for the aggregated cells (100×100 m)
 131 by counting the smallest cells (2×2 m). All aggregated cells with a density values less than 25
 132 were considered as non built-up cells. The threshold of 25 (representing a building of 100m^2)
 133 corresponds to an average-sized residential building in Belgium (Tannier and Thomas, 2013).

134 All 100×100 m cells have a density index ranging between 0 and 2500. The density index is
 135 then used to set four classes: non-built-up (class-0), low-density (class-1), medium-density
 136 (class-2) and high-density (class-3). A geometrical interval classification method is used to set
 137 the density ranges that define the different classes. This classification method works very well on

138 continuous data (Arlinghaus and Kerski, 2013). The resulting density ranges are listed in Table
 139 2.

Table 2. Built-up density classes range in number of 2×2 cells (% of 100x100 cell area).

| Class | Minimum | Maximum |
|--------------------------|---------|------------|
| Class-0 (non built-up) | 0 | 24 (1) |
| Class-1 (low-density) | 25 | 102 (4.1) |
| Class-2 (medium-density) | 103 | 499 (20) |
| Class-3 (high-density) | 500 | 2500 (100) |

140 The built-up development causative factors were operationalized to be included in the MLR.
 141 Table 3 gives the selected factors for this study. The socio-economic data (employment rate and
 142 richness index) come from the Belgian statistics, published by The Walloon Institute for
 143 Evaluation, Prospective and Statistics. The elevation data are derived from the Belgian National
 144 Geographic Institute. The distance to the different road categories are derived from a vector
 145 dataset made available by Navteq Company. The Navteq dataset identifies the following
 146 categories of roads: Road1 (highways), Road2 (main roads), Road3 (secondary roads), Road4
 147 (local roads). The location of railway stations are provided by Walphot SA Company. This study
 148 considers distance to large-sized cities (population greater than 90,000) and medium-sized
 149 Belgian cities (population between 20,000 and 90,000). The distance-based factors are based on
 150 the Euclidean distance to selected features. Euclidean distance is widely used in land-use change
 151 models (Poelmans and Van Rompaey, 2009; Roy Chowdhury and Maithani, 2014). Zoning areas
 152 were obtained from the regional zoning plan, commonly named as PDS (plan de secteur) in
 153 Wallonia. A zoning map was developed by discriminating between the zones where built-up
 154 development is legally permitted and those where it is not.

Table 3. List of selected built-up causative factors.

| Factor | Name | Type | Unit |
|----------|-----------------------------|-------------|-------------------------------------|
| X_1 | Elevation (DEM) | Continuous | Meter |
| X_2 | Slope | Continuous | Percent rise |
| X_3 | Dist. to Road1 | Continuous | Meter |
| X_4 | Dist. to Road2 | Continuous | Meter |
| X_5 | Dist. to Road3 | Continuous | Meter |
| X_6 | Dist. to Road4 | Continuous | Meter |
| X_7 | Dist. to railway stations | Continuous | Meter |
| X_8 | Dist. to large-sized cities | Continuous | Meter |
| X_9 | Dist. to med-sized cities | Continuous | Meter |
| X_{10} | Employment rate | Continuous | Percent |
| X_{11} | Richness index | Continuous | Percent |
| X_{12} | Zoning | Categorical | Binary (0 non built-up, 1 built-up) |

156 3. Methodology

157 In this study, an integrated MLR and CA model is developed. The model considers two built-
 158 up processes: (1) built-up expansion (transitions from non-built-up to built-up) and (2) built-up
 159 densification (transitions from lower built-up densities to higher ones). This section discusses the
 160 main characteristics of the model. The quantity of change during calibration (1990-2000) and
 161 validation (2000-2010) phases was constrained to the actual quantity of new built-up lands, table
 162 1, divided evenly by 10 (the number of years).

163 3.1 The transition rules

164 The quantity of change is spatially allocated based on a transition rule which has two
 165 components. The first component concerned the main built-up causative factors as determined
 166 using MLR (section 3.1.1). The second component dealt with the neighborhood characteristics
 167 (section 3.1.2). The transition potentials P for a cell ij changing its state from non-built-up to one

168 of the built-up densities or low density built-up to a higher one at specific time-step is calculated
 169 as follows:

$$P_{ij} = \sqrt{(P_c)_{ij} \times (P_n)_{ij}^\sigma} \quad (1)$$

170 where $(P_c)_{ij}$ is the built-up probability based on built-up causative factors, $(P_n)_{ij}^\sigma$ is the
 171 neighborhood effect on the cell ij and σ expresses the relative importance of the neighborhood
 172 effect. Figure 2 demonstrates an example of how the final transition potential P matrix is
 173 calculated.

| | | | | | | | | | | |
|------|------|------|---|-------------------|------------------|------------------|---|------|-----|-----|
| 0.31 | 0.97 | 0.88 | ⊙ | 14.1 ² | 2 ² | 6.2 ² | → | 7.9 | 2 | 5.8 |
| 0.42 | 0.92 | 0.76 | | 16.2 ² | 2.8 ² | 7.1 ² | | 10.5 | 2.7 | 6.2 |
| 0.02 | 0.94 | 0.52 | | 7.8 ² | 4.2 ² | 8.6 ² | | 1.1 | 4.1 | 6.2 |

Figure 2: An example of built-up transition potentials matrix (right) which equals the square root of pairwise multiplication of P_c (left) and P_n (middle) matrices. The relative importance (σ) of P_n is assumed to be 2.

174
 175 The model selects the top-scoring cells from the built-up transition potentials matrix for each
 176 density class and changes their state to the appropriate class until meeting the required quantity.
 177 The transition potential matrices are calibrated for 1990-2000. The calibration results are then
 178 used to simulate 2000-2010 built-up pattern. The simulated map of 2010 is compared against the
 179 actual 2010 map to validate the model allocation ability (section 3.2).

180 3.1.1. Built-up development causative factors calibration

181 The $(P_c)_{ij}$ can be determined through a set of factors described in Table 3 using the MLR. The
 182 MLR is a model to discover the empirical relationships between a multi categories dependent
 183 variable and several independent variables (built-up development causative factors). The model
 184 performed for class-0 (dependent variable represents non-changes/changes from class-0 to class-

185 1 or class-2 or class-3), for class-1 (dependent variable represents non-changes/changes from
 186 class-1 to class-2) and for class-2 (dependent variable represents non-changes/changes from
 187 class-2 to class-3).

188 The general form of the MLR can be represented as:

$$\begin{aligned} \log(k_1) &= \alpha_{k_1} + \beta_{k_1 1} X_1 + \beta_{k_1 2} X_2 + \dots + \beta_{k_1 v} X_v \\ &\dots \\ \log(k_n) &= \alpha_{k_n} + \beta_{k_n 1} X_1 + \beta_{k_n 2} X_2 + \dots + \beta_{k_n v} X_v \end{aligned} \quad (2)$$

189 where $\log(k_n)$ is the natural logarithm of class k_n versus the reference class k_0 , X is a set of
 190 explanatory variables (X_1, X_2, \dots, X_v), α_{k_n} is the intercept term for class k_n versus the reference
 191 class and β is the slopes for the classes (the coefficient vector). Thus, the probabilities of each
 192 class can be obtained using the following formula:

$$\begin{aligned} ((P_c)_{ij}, Y = k_0) &= \frac{1}{1 + \exp(\log(k_1)) + \exp(\log(k_2)) + \dots + \exp(\log(k_n))} \\ ((P_c)_{ij}, Y = k_1) &= \frac{\exp(\log(k_1))}{1 + \exp(\log(k_1)) + \exp(\log(k_2)) + \dots + \exp(\log(k_n))} \\ &\dots \\ ((P_c)_{ij}, Y = k_n) &= \frac{\exp(\log(k_n))}{1 + \exp(\log(k_1)) + \exp(\log(k_2)) + \dots + \exp(\log(k_n))} \end{aligned} \quad (3)$$

193 where $((P_c)_{ij}, Y = k_n)$ is the probability of change from the reference class to class k_n occurring in
 194 cell ij . The MLR employs the maximum likelihood estimation method to achieve the best fit sets
 195 of coefficients for each X .

196 The MLR outcomes are a set of coefficients that define the relative contribution of each factor
 197 to the built-up process, as well as a set of maps of probability of built-up for each class that are
 198 generated by inserting the coefficients of the MLR model into Equation (3).

199 The goodness-of-fit of the MLR is evaluated using the relative operating characteristic (ROC)
200 method. The ROC is an excellent method to estimate the quality of a model that predicts the
201 occurrence of an event by comparing a probability map of that event occurring and a binary map
202 presenting the actual changes (Hu and Lo, 2007). A ROC value of 0.5 means a completely
203 random discrimination and 1 means a perfect one.

204 All the data layers were resampled to the same cell resolution of 100×100m. The *X-variables*
205 are measured in different units and therefore we standardized all continuous *X-variables*. If some
206 of *X-variables* relatively measure the same phenomena, then strong collinearities will cause the
207 erroneous estimation of the parameters. A multicollinearity test was examined in the initial stage
208 using variance inflation factors (VIF) to ensure that there are not two or more causative factors
209 measuring the same phenomena. (Montgomery and Runger, 2003) recommended that the VIF
210 values should not exceed 4.

211 The dependent variables may show spatial autocorrelation, which biases the results of the
212 regression analysis (Overmars et al., 2003). This issue can be addressed through a data sampling
213 approach (Cammerer et al., 2013; Poelmans and Van Rompaey, 2010; Rienow and Goetzke,
214 2015). A sample of 29300 cells was randomly selected. For each reference class, other existing
215 classes in 1990 are excluded from the sampling, e.g. expansion (class-0) sampling procedure
216 considers new transitions from class-0 to class-0, class-1, class-2 and class-3. The selection of
217 samples is based on 100 runs of the MLR with different random samples. The best sample set,
218 evaluated by ROC, is then selected.

219 3.1.2. Cell neighborhood calibration

220 Neighborhood interactions can also be calibrated in MLR model by including them as part of
 221 the explanatory variables (Hu and Lo, 2007; Verburg et al., 2004). However, because MLR
 222 models are not temporally explicit, they cannot reveal the path-dependent and self-organizing
 223 development which is typical for urban expansion (Poelmans and Van Rompaey, 2010; Wu,
 224 2002). The most common approach to explicitly calibrate the neighborhood interactions on a
 225 dynamic basis is by using a cellular automata (CA) modelling approach.

226 In some studies (e.g. Chen et al., 2014; Poelmans and Van Rompaey, 2009; Wu, 2002) the
 227 neighborhood is defined as a square region, the Moore neighborhood, around the central cell
 228 with many square sizes from 3×3 to 11×11 . Chen et al. (2014) and Poelmans and Van Rompaey
 229 (2009) analyzed several square sizes and concluded that the model run with the 3×3
 230 neighborhood window produces a land-use pattern that most fits the actual pattern. These studies
 231 use a coarse cell resolutions. However, it might be different for finer cell resolutions. In this
 232 study, a 3×3 neighborhood window is used to consider neighborhood interactions. The $(P_n)_{ij}$ is
 233 calculated according to the method proposed by White and Engelen (2000):

$$(P_n)_{ij} = \sum_k \sum_x \sum_d w_{kxd} \cdot I_{kxd} \quad (4)$$

234 where w_{kxd} is the weighting parameter assigned to a cell with class k , which represents one of the
 235 built-up classes listed in table 2, at position x at distance zone d and I_{kxd} is 1 if a cell in distance d
 236 is occupied by class k or 0 otherwise.

237 Our objective is to define the CA parameters that achieve the best allocation accuracy rate for
 238 the expansion process (transitions from class-0 to class-1, class-2 and class-3 simultaneously)

239 and for the densification process (transitions from class-1 to class-2 and transitions from class-2
240 to class-3). In order to automatically calibrate the neighborhood weighting parameters, a multi-
241 objective genetic algorithm (MGA) is used for the expansion and a genetic algorithm (GA) is
242 used for the densification process. The genetic algorithm is a highly effective algorithm for
243 solving both constrained and unconstrained optimization problems that has been inspired by the
244 mechanisms of evolution and genetics (Al-Ahmadi et al., 2009; Holland, 1975). MGA attempts
245 to portray a trade-off among multiple, possibly conflicting objectives at once. In this paper,
246 MGA is a variant of a non-dominated sorting genetic algorithm II (NSGA-II) proposed by (Deb,
247 2001). NSGA-II favors individuals with an elitist strategy and individuals that can help increase
248 the diversity of the population (Yijie and Gongzhang, 2008). The output of the MGA is a set of
249 solutions that is also known as Pareto front optimized solutions, among which we can select the
250 most preferable solution. Pareto front is a set of feasible solutions that are non-dominated to each
251 other but are significantly better than the rest of solutions.

252 The MGA/GA initializes a random initial population in which many solutions participate in an
253 iteration (generation). It then uses stochastic operators to generate new generations and direct a
254 searching process based on a fitness function. Each individual in the population corresponds to a
255 chromosome made up of a set of genes, where each gene represents one parameter that requires
256 calibration. In each generation, every individual in the population is evaluated through a fitness
257 function. Once the initial population is generated and evaluated, the parents for the next
258 generation are selected by using a tournament procedure based on a relative fitness score. In this
259 paper, the tournament randomly selects two individuals, and the individual with the highest
260 fitness value becomes a parent. Each two parents are combined based on a crossover operator.
261 We proposed that the crossover operator generates two children that lie on the line representing

262 both parents and inherit at least 70% genes from the parent with the better fitness value. Once the
263 new generation is obtained, each child is then perturbed in its vicinity by a mutation operator that
264 adds a small random number to each gene.

265 This study tries to achieve a proper balance between exploration and exploitation ability of the
266 MGA/GA. Exploration enables the MGA/GA to explore a broader search space, while
267 exploitation enables MGA/GA to focus on one direction which is an optimal solution or close to
268 it (Hansheng and Lishan, 1999). The mutation operator is used to provide exploration ability
269 whereas the crossover operator is used to lead the population to the global optimal solution so
270 far. In our case, the mutation operator selects a random number from a Gaussian distribution
271 with a center of zero and a standard deviation of 2 at the first generation. This standard deviation
272 is shrunk to 0 linearly as the last generation is reached. Consequently, the MGA/GA explores
273 much more search space at the beginning of the optimization process and ensures the
274 convergence of the population towards the global optimal solution by the end of the process.

275 MGA/GA is initialized with a random population. Stochastic operators are applied to this
276 population and a large number of generations evolved to obtain a favorable solution. Each
277 individual solution takes about 19 seconds in case of MGA and 8 seconds in case of GA to be
278 evaluated using a good PC (Intel Core i7-4700 CPU @ 2.4GHz) implying that large population
279 and generation numbers require considerable time to be processed. To minimize the computing
280 time, we implement a two phase MGA/GA. First, the MGA/GA starts with a low number of
281 population and generations. Second, the outcome of the first run is used to set the initial
282 population, initial range and number of generations. In addition, the first run is used to determine
283 values for the crossover and mutation operators. Based on this, a set of 500 generations (300 for

284 expansion, 100 for densification of class-1 and 100 for densification of class-2) with 500
285 individuals for each generation are used for the final MGA/GA.

286 The objective function for the genetic algorithms for the calibration is based on a fuzzy
287 membership function, as discussed further below. The parameter values that maximize the
288 objective function will be selected as the best calibration outcome.

289 **3.2. Validation**

290 The ability of the model to locate transitions from non-built-up to one of built-up densities and
291 lower densities to higher densities is validated by comparing the simulated map of 2010 with the
292 actual map of 2010. The comparison considers only new built-up transitions between 2000 and
293 2010. The fuzziness index of a cell location depends on the cell itself and the cells in its
294 neighborhood. There is no universally agreed extent to which the neighboring cells influence the
295 fuzzy representation and a type of decay function among land-use modelers. Although it may be
296 advantageous to experiment with different neighboring sizes and decay functions to define the
297 best alternative, this experiment is beyond the scope of this paper as it would require too much
298 space to adequately discuss such analyses. However, a number of authors proposed an
299 exponential decay function with a halving distance of two cells and a neighborhood with a four-
300 cell radius to evaluate (Ahmed et al., 2013; Hagen, 2003; Loibl et al., 2007). Likewise, the
301 average fuzziness index used in this paper is an exponential decay with a halving distance of two
302 cells and a neighborhood with a four-cell neighbor extent and calculated as follows:

$$A_k = \frac{\sum_{x_k \in X_{k,sim}} \left| I_{x_k 0} \cdot (1/2)^{0/2}, I_{x_k 1} \cdot (1/2)^{1/2}, \dots, I_{x_k d} \cdot (1/2)^{d/2} \right|_{\max}}{X_{k,actul}} \quad (5)$$

303 where A_k ($0 \leq A \leq 1$) is the average fuzziness index for class k , $I_{x_k d}$ is 1 if cell x_k in the simulated
 304 map in a neighborhood at zone d ($0 \leq d \leq 4$) is identical to one cell in neighborhood at zone d in
 305 the actual map otherwise is 0, $X_{k,sim}$ is the total number of changed cells of class k in the
 306 simulated map and $X_{k,actul}$ is the total number of changed cell of class k in the actual map. The
 307 fuzziness index is also employed as the objective function for MGA/GA.

308 **4. Results and discussion**

309 In this section, the built-up pattern resulted from classification of CAD data, the calibration
 310 results and the validation of the model are discussed. In general, the built-up pattern visible in
 311 Wallonia resembles the classical built-up pattern from across a wide range of regions worldwide
 312 (Kumar et al., 2012). A high level of built-up density was found in the major built-up cores
 313 surrounded by medium-density built-up areas. A large majority of low-density lands are likely to
 314 be found in scattered rural areas and remote locations. Figure 3 illustrates different densities for
 315 Charleroi and Namur metropolitan areas as an example.

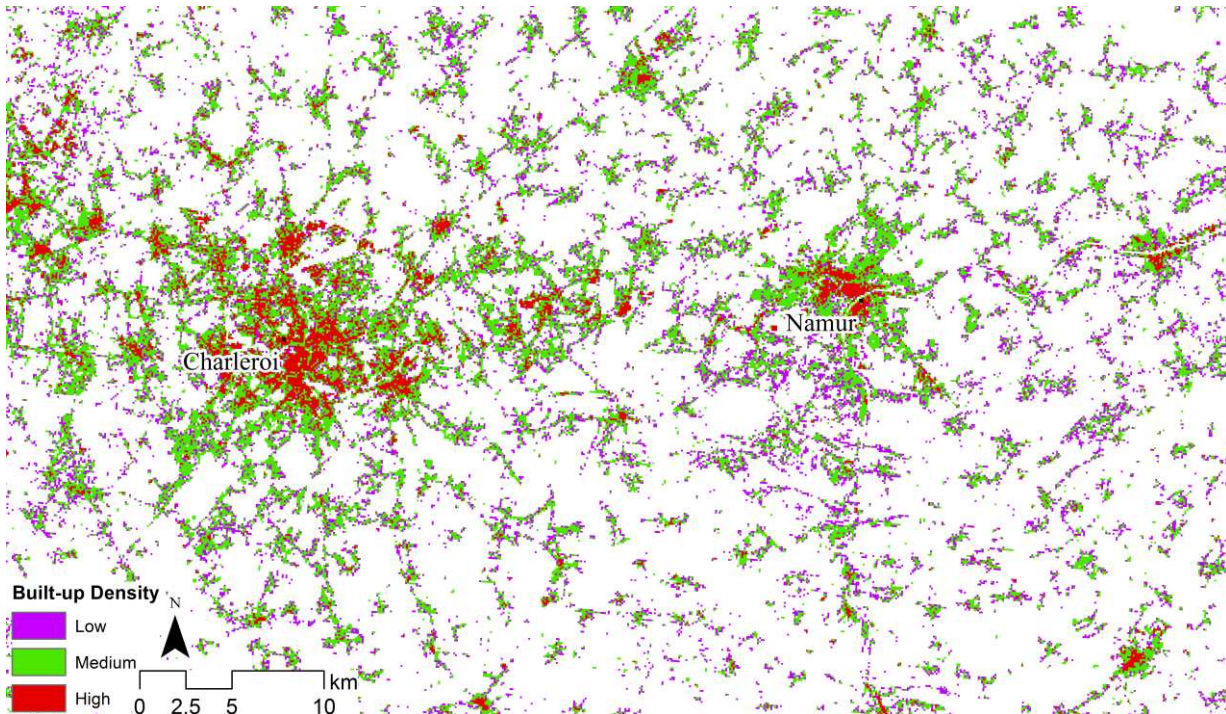


Figure 3: Built-up classes of 2010 for Charleroi and Namur metropolitan areas.

316 Variance inflation factors test, with values of less than 1.33, shows no problems with
 317 multicollinearity suggesting that all causative factors can be incorporated in the MLR model. The
 318 MLR parameter sets calibrated in the 1990–2000 are shown in Figure 4. According to the results,
 319 the major causative factor of the expansion process is the zoning status and that is in-line with
 320 Poelmans and Van Rompaey (2010). Zoning impact shows a steady upward trend along with
 321 density. High-density developments are located in areas where the legally-binding plan allows
 322 such developments, to avoid any possible administrative and financial risks. On the other hand,
 323 built-up developments in areas adjacent to urban cores (class-2) like suburbs do not strictly
 324 follow policies. The impact of policy on low-density developments is low compared to other
 325 classes. This class can be considered as remote built-up areas, consisting in scattered buildings,
 326 which can sometimes deviate from zoning plans. The magnitude of the zoning status influence

327 on the densification process is remarkably low compared to the expansion process. The fact that
328 the densification process is done within existing built-up areas, merely means that the zoning
329 plan does not have a strong effect on the densification processes. As in Poelmans and Van
330 Rompaey (2010) slope shows a negative effect on the development of built-up areas. Distance to
331 roads shows a negative effect on the built-up developments so that built-up transitions generally
332 occur close to roads as reported in Cammerer et al., (2013) and Poelmans and Van Rompaey
333 (2010). Distance to railway stations is statistically significant for the expansion of high-density
334 built-up suggesting that parcels nearby train stations are attractive for new dense developments.
335 Although the richness index is insignificant, as in Hu and Lo (2007), except for all medium-
336 density transitions, it implies that the medium-density developments are linked closely to the
337 income distribution. Medium-density can generally represent urban sprawl and suburbanization
338 which replace non-built-up lands with single-family houses on large lots. The richness index has
339 a positive impact on transition from non-built-up and low-density to medium-density implying
340 that affluent and middle-class people settle in medium-density built-up areas. In contrast, the
341 richness index has a strong negative impact on the transition from medium-density to high-
342 density so that most such transitions can be found in somewhat poor neighborhoods. Distance to
343 cities especially the medium-sized cities indicates a moderate negative impact on built-up
344 expansion processes and densification of low-density areas. That is in-line with Poelmans and
345 Van Rompaey (2010) who reported that urban development tends to occur near to the cities. As
346 in Hu and Lo (2007) and Poelmans and Van Rompaey (2010), employment rate has insignificant
347 impact on the expansion of most built-up densities and the densification process.

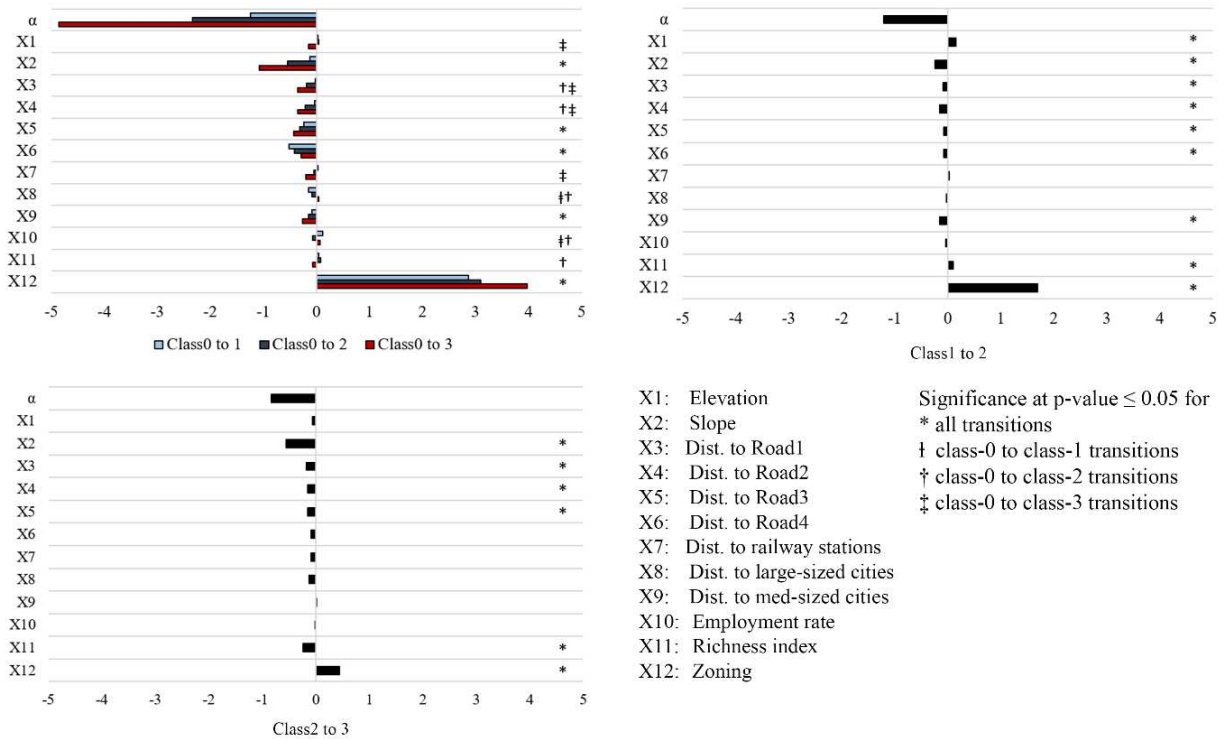


Figure 4: The MLR parameters coefficients for 1990-2000.

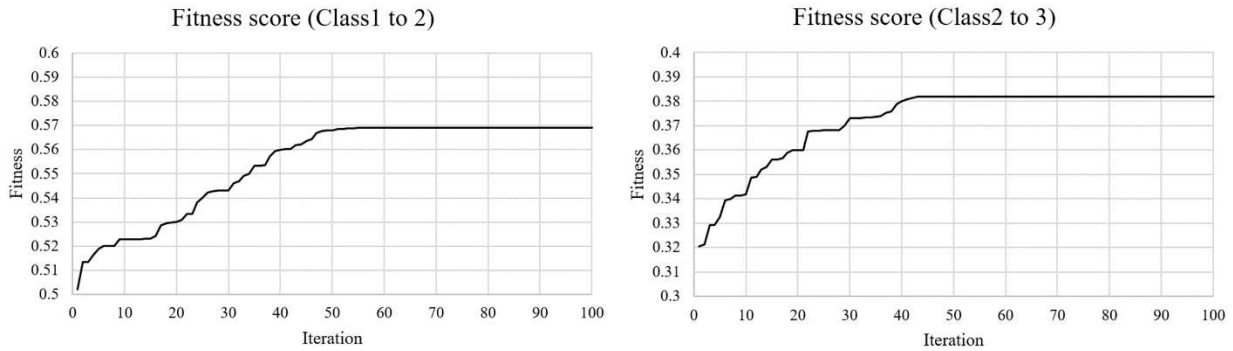


Figure 5: The convergence of the fitness score during the GA optimization.

348 The GA optimization module for the densification of class-1 and class-2 began to converge
 349 when reaching iteration 56 and 50 respectively (Figure 5). After 228 iterations, average
 350 change in the spread of Pareto solutions for MGA was less than 0.00001. The MGA/GA
 351 optimal weighting values that define neighborhood interactions are given in Figure 6 (a, b and c).

352 The calibration shows that the likelihood of low-density expansion is highly increased by
353 increasing the number of existing low-density and medium-density lands and decreasing the
354 number of high-density lands in the immediate neighborhood of the cell. The probability of
355 medium-density expansion is increased with increasing number of all land-uses, especially
356 medium-density cells. This study finds a positive relationship between expansion of high-density
357 and the number of existing high-density cells in the neighborhood of the cell. In contrast, the
358 expansion of high-density lands is negatively impacted by increasing the number of non built-up,
359 low and medium-density lands. The probability of low to medium-density built-up transitions is
360 positively linked with the existing non built-up, low and medium-density built-up neighbors and
361 negatively linked with high-density neighbors, whereas the densification of medium-density
362 areas is negatively related to the increasing number of non built-up and low-density cells and
363 positively related to the increasing the number of high-density cells in the neighborhood of the
364 cell. Together, these findings suggest that existing residents of low and medium-density areas
365 tend to protest dense developments near their home, whereas most new densified areas are
366 located within or close to already high-density neighbors. This causes a highly fragmented and
367 low-density built-up landscape. One of the main factors leading to this situation is the spatial
368 planning policy (Dieleman and Wegener, 2004; Poelmans and Van Rompaey, 2009).

369 The ROC values of the MLR outcomes are 0.81, 0.85, 0.94, 0.73 and 0.72 for class-0 to class-
370 1, class-0 to class-2, class-0 to class-3, class-1 to class-2 and class-2 to class-3 respectively. ROC
371 values higher than 0.70 are considered as a reasonable fit and the estimates can be used in further
372 analyses (Cammerer et al., 2013; Jr and Lemeshow, 2004).

373 The calibration and validation of allocation accuracy rates are given in figure 6 (d). The
374 relative importance of the neighborhood effect (σ) parameter is calibrated using MGA. The

375 MGA of σ converges when reaching iteration 35 for expansion process, 27 and 24 respectively
 376 for densification of class-1 and class-2. The value of parameter σ shows neutral effect, i.e. equals
 377 1, on the expansion of class-2, class-3 and the densification of class-2. For the expansion and
 378 densification of low-density class the values of σ are 1.97 and 0.53 respectively.

379 The calibration accuracy rates are larger than the validation rate. The possible source of this
 380 variation is potentially due to the uncertainty associated with the future values of modeling
 381 parameters. Most CA models (e.g. Al-Ahmadi et al., 2009; García et al., 2013) introduced a
 382 stochastic disturbance term to represent unknown errors and uncertainty. The extension of this
 383 study necessitates a more comprehensive framework that explicitly quantifies and models
 384 uncertainty related to future values of the model's parameters.

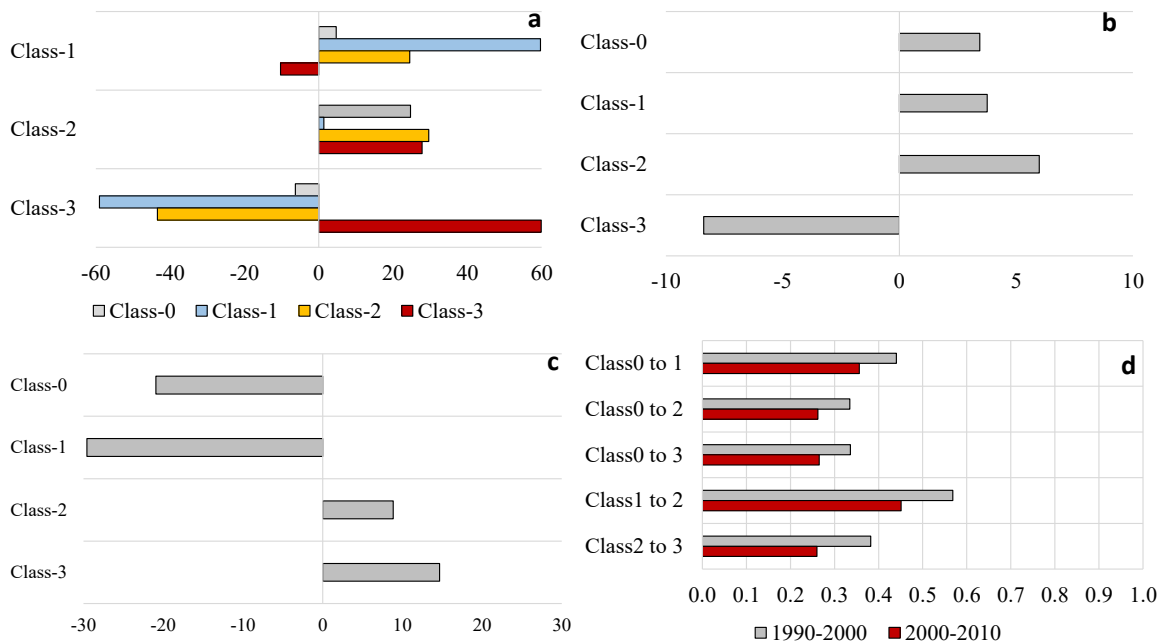


Figure 6: Weighting values that define neighborhood parameters values for (a) transitions from class-0 to class-1, class-2 and class-3, (b) transitions from class-1 to class-2 and (c) transitions from class-2 to class-3. (d) The average fuzzy similarity rates for calibration and validation.

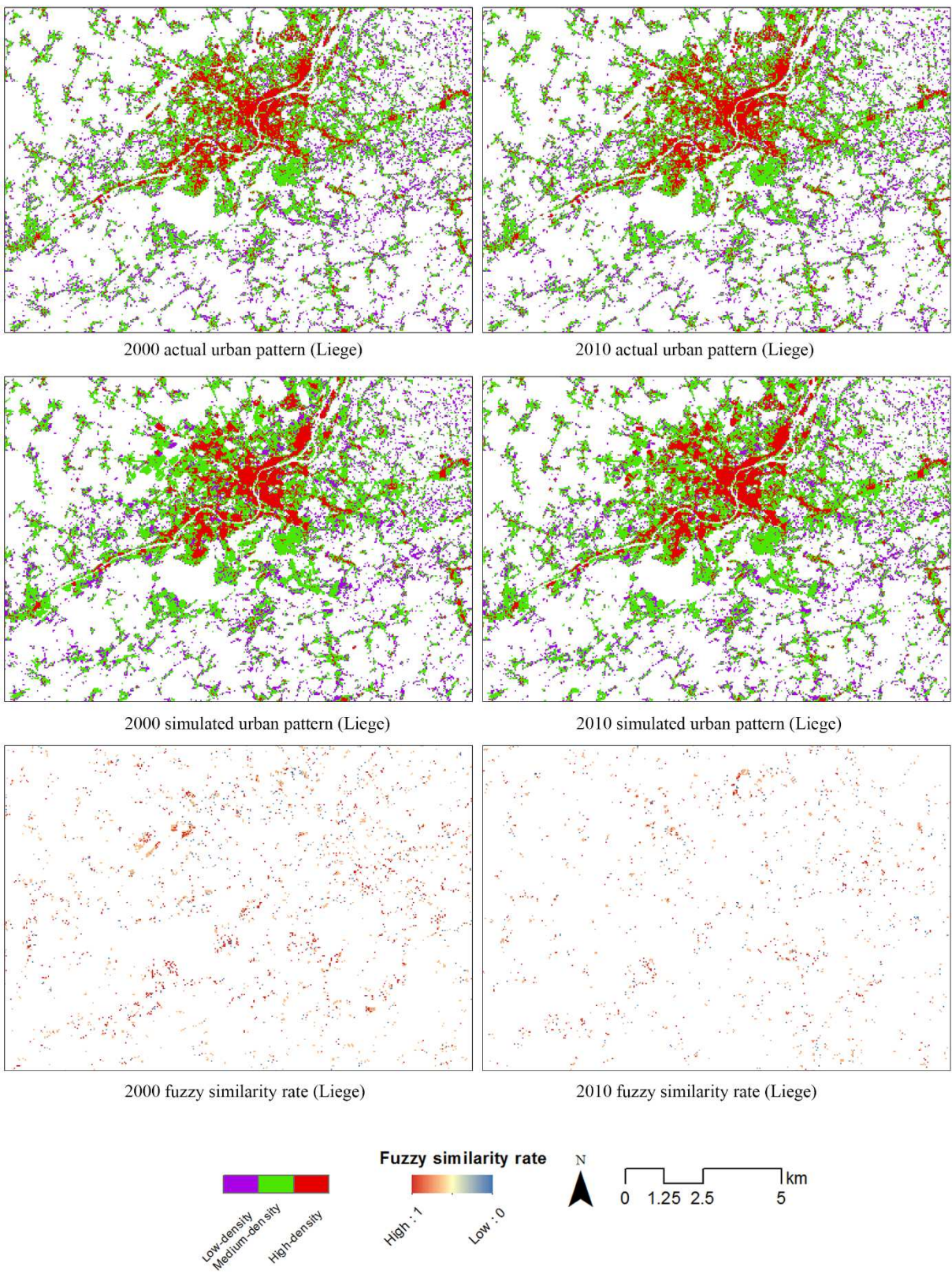


Figure 7: The actual and simulated 2000 and 2010 built-up patterns for Liege metropolitan.

387 The simulation of the 2000 and 2010 built-up patterns in the major metropolitan area (Liege),
388 as an example, are shown in Figure 7.

389 **5. Conclusion**

390 One of the central limitations of most existing built-up/urban expansion models is that
391 urbanization is considered as a binary process (built-up/non built-up). This research has
392 demonstrated that the built-up development process is heterogeneous, with links between density
393 and the impact of different built-up development causative factors. We propose an integrated
394 multinomial logistic regression (MLR) and cellular automata (CA) model to examine the built-up
395 development trends in Wallonia (Belgium). The built-up development considers both expansion
396 and densification. Considering the densification is an essential component of sprawl-fighting
397 land-use policies. In this study, built-up densities (non built-up, low-density, medium-density
398 and high-density) for 1990, 2000 and 2010 and geophysical and socioeconomic data that are
399 referred to as causative factors were gathered and processed.

400 The MLR allows to automate the calibration of the causative factors whereas the CA model is
401 used to simulate the neighborhood interactions. A multi-objective/genetic algorithm is employed
402 to calibrate neighborhood interactions parameters. The calibration is done for built-up transitions
403 between 1990 and 2000. The calibration results are then used to validate the model by simulating
404 the 2010 built-up pattern and compare it with the actual 2010 built-up. The model evaluates the
405 MLR outcomes using relative operating characteristic and validates the simulated built-up
406 patterns by means of fuzzy set theory. The model reveals a good overall accuracy. However,
407 calibration and validation processes provide information on the uncertainties in the model
408 outcomes over time. In later work we intend to pursue the analysis further by quantifying and

409 modelling uncertainty in the future built-up simulations. Therefore, our model can effectively
410 develop future built-up scenarios considering the uncertainty.

411 The findings drawn from this study suggest that all selected factors have impacts on the
412 expansion in Wallonia, but their relative importance varied with density. However, zoning status,
413 slope, and distance to local roads are the most important determinants of the expansion process.
414 In regards to the densification process, it is mainly driven by zoning, slope, distance to different
415 roads and richness index. The magnitude of the effect of land-use policies (zoning) decline along
416 with the densification process. The neighborhood effect weights imply that the densification
417 occurs in already dense areas whereas low-density and medium-density areas tend to retain their
418 densities over time. Public authorities clearly should play a role in the development of a more
419 balanced densification policy, considering the densification of very accessible (transport,
420 services, etc) low/medium density nodes besides a further densification of already dense areas.
421 This is not contradictory with a concentration spatial policy provided that low/medium density
422 nodes where densification occurs are well connected to city centers (as for instance promoted
423 through transit-oriented development).

424 This study identifies the most notable built-up development factors at different densities. Our
425 analysis does not consider building use or height. There are several missing of buildings uses and
426 heights within the cadastral data. Consequently, population and employment density indices
427 cannot be considered here. However, this study prompts a series of further research questions
428 regarding the relation between built-up density and land-use policy, spatial, geophysical and
429 socioeconomic factors. Hopefully this study should provide a useful context for policy makers
430 and the ongoing research.

431

432 **Acknowledgments:** The research was funded through the ARC grant for Concerted Research
 433 Actions, financed by the Wallonia-Brussels Federation.

References

- 434 Aburas, M.M., Ho, Y.M., Ramli, M.F., Ash'aari, Z.H., 2016. The simulation and prediction of spatio-temporal urban growth trends
 435 using cellular automata models: A review. *Int. J. Appl. Earth Obs. Geoinformation* 52, 380–389.
 436 doi:10.1016/j.jag.2016.07.007
- 437 Achmad, A., Hasyim, S., Dahlan, B., Aulia, D.N., 2015. Modeling of urban growth in tsunami-prone city using logistic regression:
 438 Analysis of Banda Aceh, Indonesia. *Appl. Geogr.* 62, 237–246. doi:10.1016/j.apgeog.2015.05.001
- 439 Ahmed, B., Ahmed, R., Zhu, X., 2013. Evaluation of Model Validation Techniques in Land Cover Dynamics. *ISPRS Int. J. Geo-*
 440 *Inf.* 2, 577–597. doi:10.3390/ijgi2030577
- 441 Al-Ahmadi, K., See, L., Heppenstall, A., Hogg, J., 2009. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi
 442 Arabia. *Ecol. Complex.* 6, 80–101. doi:10.1016/j.ecocom.2008.09.004
- 443 Arlinghaus, S.L., Kerski, J.J., 2013. *Spatial Mathematics: Theory and Practice through Mapping*. CRC Press.
- 444 Beckers, A., Dewals, B., Ercicum, S., Dujardin, S., Detrembleur, S., Teller, J., Piroton, M., Archambeau, P., 2013. Contribution
 445 of land use changes to future flood damage along the river Meuse in the Walloon region. *Nat Hazards Earth Syst Sci* 13, 2301–
 446 2318. doi:10.5194/nhess-13-2301-2013
- 447 Belgian Federal Government, 2013. *Statistics Belgium [WWW Document]*. Stat. Belg. URL <http://statbel.fgov.be/fr/statistiques/chiffres/> (accessed 4.29.14).
- 448 Berberoğlu, S., Akın, A., Clarke, K.C., 2016. Cellular automata modeling approaches to forecast urban growth for adana, Turkey:
 449 A comparative approach. *Landsc. Urban Plan.* 153, 11–27. doi:10.1016/j.landurbplan.2016.04.017
- 451 Cammerer, H., Thieken, A.H., Verburg, P.H., 2013. Spatio-temporal dynamics in the flood exposure due to land use changes in the
 452 Alpine Lech Valley in Tyrol (Austria). *Nat. Hazards* 68, 1243–1270. doi:10.1007/s11069-012-0280-8
- 453 Chen, Y., Li, X., Liu, X., Ai, B., 2014. Modeling urban land-use dynamics in a fast developing city using the modified logistic
 454 cellular automaton with a patch-based simulation strategy. *Int. J. Geogr. Inf. Sci.* 28, 234–255.
 455 doi:10.1080/13658816.2013.831868
- 456 Clarke, K.C., Gaydos, L.J., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San
 457 Francisco and Washington/Baltimore. *Int. J. Geogr. Inf. Sci.* 12, 699–714. doi:10.1080/136588198241617
- 458 Clarke, K.C., Hoppen, S., Gaydos, L., 1997. A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San
 459 Francisco Bay Area. *Environ. Plan. B Plan. Des.* 24, 247–261. doi:10.1068/b240247
- 460 Crols, T., White, R., Uljee, I., Engelen, G., Poelmans, L., Canters, F., 2015. A travel time-based variable grid approach for an
 461 activity-based cellular automata model. *Int. J. Geogr. Inf. Sci.* 29, 1757–1781. doi:10.1080/13658816.2015.1047838
- 462 Deb, K., 2001. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons.
- 463 Dieleman, F., Wegener, M., 2004. Compact City and Urban Sprawl. *Built Environ.* 1978- 30, 308–323.
- 464 Dubovyk, O., Sliuzas, R., Flacke, J., 2011. Spatio-temporal modelling of informal settlement development in Sancaktepe district,
 465 Istanbul, Turkey. *ISPRS J. Photogramm. Remote Sens., Quality, Scale and Analysis Aspects of Urban City Models* 66, 235–
 466 246. doi:10.1016/j.isprsjprs.2010.10.002
- 467 Feng, Y., Liu, Y., Tong, X., Liu, M., Deng, S., 2011. Modeling dynamic urban growth using cellular automata and particle swarm
 468 optimization rules. *Landsc. Urban Plan.* 102, 188–196. doi:10.1016/j.landurbplan.2011.04.004
- 469 García, A.M., Santé, I., Boullón, M., Crecente, R., 2013. Calibration of an urban cellular automaton model by using statistical
 470 techniques and a genetic algorithm. Application to a small urban settlement of NW Spain. *Int. J. Geogr. Inf. Sci.* 27, 1593–
 471 1611. doi:10.1080/13658816.2012.762454
- 472 García, A.M., Santé, I., Boullón, M., Crecente, R., 2012. A comparative analysis of cellular automata models for simulation of
 473 small urban areas in Galicia, NW Spain. *Comput. Environ. Urban Syst.* 36, 291–301. doi:10.1016/j.compenvurb-
 474 sys.2012.01.001
- 475 Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. *Int. J. Geogr. Inf. Sci.* 17, 235–249.
 476 doi:10.1080/13658810210157822
- 477 Han, J., Hayashi, Y., Cao, X., Imura, H., 2009. Application of an integrated system dynamics and cellular automata model for
 478 urban growth assessment: A case study of Shanghai, China. *Landsc. Urban Plan.* 91, 133–141. doi:10.1016/j.landur-
 479 bplan.2008.12.002
- 480 Han, Y., Jia, H., 2016. Simulating the spatial dynamics of urban growth with an integrated modeling approach: A case study of
 481 Foshan, China. *Ecol. Model.* doi:10.1016/j.ecolmodel.2016.04.005

- 482 Hansheng, L., Lishan, K., 1999. Balance between exploration and exploitation in genetic search. *Wuhan Univ. J. Nat. Sci.* 4, 28–
483 32. doi:10.1007/BF02827615
- 484 Holland, J.H., 1975. *Adaptation in natural and artificial systems*. U Michigan Press, Oxford, England.
- 485 Hu, Z., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. *Comput. Environ. Urban Syst.* 31, 667–688.
486 doi:10.1016/j.compenvurbsys.2006.11.001
- 487 Jiang, F., Liu, S., Yuan, H., Zhang, Q., 2007. Measuring urban sprawl in Beijing with geo-spatial indices. *J. Geogr. Sci.* 17, 469–
488 478. doi:10.1007/s11442-007-0469-z
- 489 Jr, D.W.H., Lemeshow, S., 2004. *Applied Logistic Regression*. John Wiley & Sons.
- 490 Kim, J.-H., 2003. Assessing practical significance of the proportional odds assumption. *Stat. Probab. Lett.* 65, 233–239.
491 doi:10.1016/j.spl.2003.07.017
- 492 Kumar, A., Pandey, A.C., Jeyaseelan, A.T., 2012. Built-up and vegetation extraction and density mapping using WorldView-II.
493 *Geocarto Int.* 27, 557–568. doi:10.1080/10106049.2012.657695
- 494 Li, X., Zhou, W., Ouyang, Z., 2013. Forty years of urban expansion in Beijing: What is the relative importance of physical, socio-
495 economic, and neighborhood factors? *Appl. Geogr.* 38, 1–10. doi:10.1016/j.apgeog.2012.11.004
- 496 Liao, J., Tang, L., Shao, G., Qiu, Q., Wang, C., Zheng, S., Su, X., 2014. A neighbor decay cellular automata approach for simulating
497 urban expansion based on particle swarm intelligence. *Int. J. Geogr. Inf. Sci.* 28, 720–738.
498 doi:10.1080/13658816.2013.869820
- 499 Liu, X., Ma, L., Li, X., Ai, B., Li, S., He, Z., 2014. Simulating urban growth by integrating landscape expansion index (LEI) and
500 cellular automata. *Int. J. Geogr. Inf. Sci.* 28, 148–163. doi:10.1080/13658816.2013.831097
- 501 Loibl, W., Toetzer, T., 2003. Modeling growth and densification processes in suburban regions—simulation of landscape transition
502 with spatial agents. *Environ. Model. Softw.*, *Applying Computer Research to Environmental Problems* 18, 553–563.
503 doi:10.1016/S1364-8152(03)00030-6
- 504 Loibl, W., Tötzer, T., Köstl, M., Steinnocher, K., 2007. Simulation of Polycentric Urban Growth Dynamics Through Agents, in:
505 Koomen, E., Stillwell, J., Bakema, A., Scholten, H.J. (Eds.), *Modelling Land-Use Change*, The GeoJournal Library. Springer
506 Netherlands, pp. 219–236. doi:10.1007/978-1-4020-5648-2_13
- 507 Montgomery, D.C., Runger, G.C., 2003. *Applied Statistics and Probability for Engineers*, Fourth. ed. John Wiley & Sons, New
508 York.
- 509 Munshi, T., Zuidgeest, M., Brussel, M., van Maarseveen, M., 2014. Logistic regression and cellular automata-based modelling of
510 retail, commercial and residential development in the city of Ahmedabad, India. *Cities* 39, 68–86. doi:10.1016/j.cit-
511 ies.2014.02.007
- 512 Mustafa, A., Cools, M., Saadi, I., Teller, J., 2015. Urban Development as a Continuum: A Multinomial Logistic Regression Ap-
513 proach, in: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O.
514 (Eds.), *Computational Science and Its Applications -- ICCSA 2015, Lecture Notes in Computer Science*. Springer Interna-
515 tional Publishing, pp. 729–744.
- 516 Mustafa, A., Saadi, I., Cools, M., Teller, J., 2014. Measuring the Effect of Stochastic Perturbation Component in Cellular Automata
517 Urban Growth Model. *Procedia Environ. Sci.*, 12th International Conference on Design and Decision Support Systems in
518 Architecture and Urban Planning, DDSS 2014 22, 156–168. doi:10.1016/j.proenv.2014.11.016
- 519 Nabielek, K., 2012. The Compact City: Planning strategies, recent developments and future prospects in the Netherlands - PBL
520 Netherlands Environmental Assessment Agency, in: *Proceedings of the AESOP 26th Annual Congress*. Presented at the
521 AESOP 26th Annual Congress, Ankara.
- 522 Omrani, H., Abdallah, F., Charif, O., Longford, N.T., 2015. Multi-label class assignment in land-use modelling. *Int. J. Geogr. Inf.*
523 *Sci.* 29, 1023–1041. doi:10.1080/13658816.2015.1008004
- 524 Overmars, K.P., de Koning, G.H.J., Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. *Ecol. Model.* 164,
525 257–270. doi:10.1016/S0304-3800(03)00070-X
- 526 Poelmans, L., Van Rompaey, A., 2010. Complexity and performance of urban expansion models. *Comput. Environ. Urban Syst.*
527 34, 17–27. doi:10.1016/j.compenvurbsys.2009.06.001
- 528 Poelmans, L., Van Rompaey, A., 2009. Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A
529 case study in the Flanders–Brussels region. *Landsc. Urban Plan.* 93, 10–19. doi:10.1016/j.landurbplan.2009.05.018
- 530 Puertas, O.L., Henríquez, C., Meza, F.J., 2014. Assessing spatial dynamics of urban growth using an integrated land use model.
531 Application in Santiago Metropolitan Area, 2010–2045. *Land Use Policy* 38, 415–425. doi:10.1016/j.landusepol.2013.11.024
- 532 Rienow, A., Goetzke, R., 2015. Supporting SLEUTH – Enhancing a cellular automaton with support vector machines for urban
533 growth modeling. *Comput. Environ. Urban Syst.* 49, 66–81. doi:10.1016/j.compenvurbsys.2014.05.001
- 534 Robinson, D.T., Murray-Rust, D., Rieser, V., Milicic, V., Rounsevell, M., 2012. Modelling the impacts of land system dynamics
535 on human well-being: Using an agent-based approach to cope with data limitations in Koper, Slovenia. *Comput. Environ.*
536 *Urban Syst.*, Special Issue: Geoinformatics 2010 36, 164–176. doi:10.1016/j.compenvurbsys.2011.10.002
- 537 Roy Chowdhury, P.K., Maithani, S., 2014. Modelling urban growth in the Indo-Gangetic plain using nighttime OLS data and
538 cellular automata. *Int. J. Appl. Earth Obs. Geoinformation* 33, 155–165. doi:10.1016/j.jag.2014.04.009
- 539 Santé, I., García, A.M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes:
540 A review and analysis. *Landsc. Urban Plan.* 96, 108–122. doi:10.1016/j.landurbplan.2010.03.001

- 541 Shan, J., Alkheder, S., Wang, J., 2008. Genetic Algorithms for the Calibration of Cellular Automata Urban Growth Modeling.
542 Photogramm. Eng. Remote Sens. 74, 1267–1277. doi:10.14358/PERS.74.10.1267
- 543 Sunde, M.G., He, H.S., Zhou, B., Hubbart, J.A., Spicci, A., 2014. Imperviousness Change Analysis Tool (I-CAT) for simulating
544 pixel-level urban growth. *Landsc. Urban Plan.* 124, 104–108. doi:10.1016/j.landurbplan.2014.01.007
- 545 Tachieva, G., 2010. *Sprawl Repair Manual*, 2 edition. ed. Island Press, Washington.
- 546 Tannier, C., Thomas, I., 2013. Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian
547 cities. *Comput. Environ. Urban Syst.* 41, 234–248. doi:10.1016/j.compenvurbsys.2013.07.003
- 548 Thomas, I., Frankhauser, P., Biernacki, C., 2008. The morphology of built-up landscapes in Wallonia (Belgium): A classification
549 using fractal indices. *Landsc. Urban Plan.* 84, 99–115. doi:10.1016/j.landurbplan.2007.07.002
- 550 Tian, G., Ma, B., Xu, X., Liu, Xiaoping, Xu, L., Liu, Xiaojuan, Xiao, L., Kong, L., 2016. Simulation of urban expansion and
551 encroachment using cellular automata and multi-agent system model—A case study of Tianjin metropolitan region, China.
552 *Ecol. Indic., Navigating Urban Complexity: Advancing Understanding of Urban Social – Ecological Systems for Transfor-*
553 *mation and Resilience* 70, 439–450. doi:10.1016/j.ecolind.2016.06.021
- 554 van Vliet, J., Bregt, A.K., Brown, D.G., van Delden, H., Heckbert, S., Verburg, P.H., 2016. A review of current calibration and
555 validation practices in land-change modeling. *Environ. Model. Softw.* 82, 174–182. doi:10.1016/j.envsoft.2016.04.017
- 556 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
557 Netherlands. *Environ. Plan. B Plan. Des.* 31, 125–150. doi:10.1068/b307
- 558 Vermeiren, K., Van Rompaey, A., Loopmans, M., Serwajja, E., Mukwaya, P., 2012. Urban growth of Kampala, Uganda: Pattern
559 analysis and scenario development. *Landsc. Urban Plan.* 106, 199–206. doi:10.1016/j.landurbplan.2012.03.006
- 560 Wang, H., He, S., Liu, X., Dai, L., Pan, P., Hong, S., Zhang, W., 2013. Simulating urban expansion using a cloud-based cellular
561 automata model: A case study of Jiangxia, Wuhan, China. *Landsc. Urban Plan.* 110, 99–112. doi:10.1016/j.landurb-
562 *plan.2012.10.016*
- 563 Ward, D.P., Murray, A.T., Phinn, S.R., 2000. A stochastically constrained cellular model of urban growth. *Comput. Environ. Urban*
564 *Syst.* 24, 539–558. doi:10.1016/S0198-9715(00)00008-9
- 565 White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Comput.*
566 *Environ. Urban Syst.* 24, 383–400. doi:10.1016/S0198-9715(00)00012-0
- 567 White, R., Engelen, G., 1997. Cellular Automata as the Basis of Integrated Dynamic Regional Modelling. *Environ. Plan. B Plan.*
568 *Des.* 24, 235–246. doi:10.1068/b240235
- 569 White, R., Engelen, G., Uljee, I., 2015. The Cellular Automaton Eats the Regions: Unified Modeling of Activities and Land Use
570 in a Variable Grid Cellular Automaton, in: *Modeling Cities and Regions As Complex Systems: From Theory to Planning*
571 *Applications*. The MIT Press, Cambridge, Massachusetts.
- 572 White, R., Uljee, I., Engelen, G., 2012. Integrated modelling of population, employment and land-use change with a multiple
573 activity-based variable grid cellular automaton. *Int. J. Geogr. Inf. Sci.* 26, 1251–1280. doi:10.1080/13658816.2011.635146
- 574 Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. *Int. J. Geogr. Inf. Sci.*
575 16, 795–818. doi:10.1080/13658810210157769
- 576 Xian, G., Crane, M., 2005. Assessments of urban growth in the Tampa Bay watershed using remote sensing data. *Remote Sens.*
577 *Environ.* 97, 203–215. doi:10.1016/j.rse.2005.04.017
- 578 Yang, X., 2010. Integration of Remote Sensing with GIS for Urban Growth Characterization, in: Jiang, B., Yao, X. (Eds.), *Geo-*
579 *spatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library. Springer Netherlands, pp. 223–250.
580 doi:10.1007/978-90-481-8572-6_12
- 581 Yijie, S., Gongzhang, S., 2008. Improved NSGA-II Multi-objective Genetic Algorithm Based on Hybridization-encouraged Mech-
582 *anism*. *Chin. J. Aeronaut.* 21, 540–549. doi:10.1016/S1000-9361(08)60172-7
- 583 Zhang, Q., Ban, Y., Liu, J., Hu, Y., 2011. Simulation and analysis of urban growth scenarios for the Greater Shanghai Area, China.
584 *Comput. Environ. Urban Syst., Geospatial Analysis and Modeling* 35, 126–139. doi:10.1016/j.compenvurbsys.2010.12.002