Computers, Environment and Urban Systems 34 (2010) 17–27

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

# Complexity and performance of urban expansion models

Lien Poelmans *, Anton Van Rompaey

*Physical and Regional Geography Research Group, K.U. Leuven, Celestijnenlaan 200E, 3001 Heverlee, Belgium*

A R T I C L E   I N F O

*Article history:*
Received 28 November 2008
Received in revised form 10 June 2009
Accepted 12 June 2009

*Keywords:*
Urban expansion
Belgium
Logistic regression
Cellular automata
Validation

A B S T R A C T

Urban expansion and spatial patterns of urban land have a large effect on many socioeconomic and environmental processes. A wide variety of modelling approaches has been introduced to predict and simulate future urban development. These models are often based on the interpretation of various determining factors that are used to create a probability map. The main objective of this paper is to evaluate the performance of different modelling approaches for simulating spatial patterns of urban expansion in Flanders and Brussels in the period 1988–2000. Hereto, a set of urban expansion models with increasing complexity was developed based on: (i) logistic regression equations taking various numbers of determining variables into account, (ii) CA transition rules and (iii) hybrid procedures, combining both approaches. The outcome of each model was validated in order to assess the predictive value of the three modelling approaches and of the different determining variables that were used in the logistic regression models. The results show that a hybrid model structure, integrating (static) determining factors (distance to the main roads, distance to the largest cities, employment potential, slope and zoning status of the land) and (dynamic) neighbourhood interactions produces the most accurate probability map. The study, however, points out that it is not useful to make a statement on the validity of a model based on only one goodness-of-fit measure. When the model results are validated at multiple resolutions, the logistic regression model, which incorporates only two explanatory variables, outperforms both the CA-based model and the hybrid model.

## 1. Introduction

Urban land is roughly estimated to occupy only 2% or 3% of the Earth's land surface and has therefore often been ignored in traditional land change studies. A large part of the earliest studies on land use and land cover change patterns have focussed on deforestation (Kaimowitz & Angelsen, 1998; Mertens & Lambin, 1997; Schneider & Pontius, 2001) and agricultural intensification (Lambin, Rounsevell, & Geist, 2000; Van Dessel, Van Rompaey, Poelmans, & Szilassi, 2008; Veldkamp & Fresco, 1996). However, during the last decades it has been increasingly recognised that the loss of land to urban expansion has resulted in important environmental and socioeconomic impacts (Lambin et al., 2001). As a consequence, a series of empirical and theoretical modelling techniques has been introduced for the simulation and prediction of urban expansion patterns. The modelling procedures used in urban and rural environments were however distinctly different (Verburg & Overmars, 2007): urban models have dominantly been based on neighbourhood functions such as cellular automata (CA) and on infrastructure access (Clarke, Hoppen, & Gaydos, 1997; Ward, Murray, & Phinn, 2000; White, 1998; Wu, 2002), while rural land use models primarily used intrinsic suitability assessments as the basis for simulation (Pontius, Cornell, & Hall, 2001; Serneels & Lambin, 2001; Veldkamp & Fresco, 1996). Recently, integrated land use models, such as the CLUE-s model (Verburg et al., 2002) and the Environment Explorer (White & Engelen, 2000) have been developed, which combine different modelling approaches and focus on the interplay and competition between different land use activities (Schaldach & Priess, 2008). Since these 'hybrid' models combine different modelling approaches, such as empirical analysis, cellular automata and decision rules based on theory and expert knowledge, it is not possible to classify them into a single model category as proposed by Lambin et al. (2000) or Agarwal, Green, Grove, Evans, and Schweik (2002). Table 1 shows the different modelling approaches that were used in a small selection of recent urban model applications.

In spite of the variety of modelling approaches that have been developed and used in the last couple of years, it is possible to identify a common structure valid for a large number of spatially explicit land use models. In most model structures the prediction of the quantity or rate of land use change is explicitly separated from the prediction of the location or the spatial pattern of change (Verburg, Kok, Pontius, & Veldkamp, 2006). The quantity of change can either be estimated by using simple approaches such as linear

* Corresponding author. Tel.: +32 16 326406; fax: +32 16 326400.
*E-mail addresses:* lien.poelmans@ees.kuleuven.be (L. Poelmans), anton.vanrompaey@ees.kuleuven.be (A. Van Rompaey).

0198-9715/$ - see front matter 
doi:10.1016/j.compenvurbsys.2009.06.001

**Table 1**
Overview of the explanatory variables and modelling approaches used in seven recent urban model applications.

| Reviewed model applications | Explanatory variables | | | | | Modelling approach |
|---|---|---|---|---|---|---|
| | Biophysical factors | Social factors | Economic factors | Neighbourhood interactions | Spatial policies | |
| Clarke et al. (1997) | x | | x | x | x | CA-rules |
| White and Engelen (2000) | (x) | (x) | x | x | x | CA-rules<br>Expert knowledge |
| Wu (2002) | x | | x | x | | Logistic regression<br>CA-rules |
| Cheng and Masser (2003) | | | x | x | x | Logistic regression |
| Verburg, Ritsema van Eck et al. (2004) | x | | x | x | x | Logistic regression |
| Hu and Lo (2007) | x | x | x | x | x | Logistic regression |
| Dendoncker et al. (2007) | x | | x | x | | Logistic regression |

(x) Optional.

extrapolation of the observed change in the past (Poelmans & Van Rompaey, 2009; Pontius, Huffaker, & Denman, 2004) or by integrating models that predict economic and/or demographic growth (White & Engelen, 2000). The prediction of the location of change, on the other hand, is often based on the interpretation of several independent determining factors that are used to create a 'suitability' or 'probability' map. Such a probability map indicates the suitability of a location for a certain land use relative to the suitability of other locations.

Verburg, Ritsema van Eck, de Nijs, Dijst, and Schot (2004) have identified five types of determinants that can be used to explain the spatial patterns of land use change: (i) biophysical factors, (ii) social factors, (iii) economic factors, (iv) spatial policies and (v) spatial interactions and neighbourhood characteristics. While selection of the factors that are included in the analysis often relies on different disciplinary theories and prior understanding of the underlying processes of land use change, the quantification of the relations between land use change and the different driving factors is usually based on empirical methods and expert knowledge

2004). Different approaches have already been used to take these spatial autocorrelation effects into account. Poelmans and Van Rompaey (2009) have used a neighbourhood constraint that restricts new urban land to locations near to existing urban areas. Dendoncker, Rounsevell, and Bogaert (2007) and Verburg, Ritsema van Eck et al. (2004) have modelled urban expansion by including simple neighbourhood variables such as the proportion of urban land in the neighbourhood of a cell, as an explanatory factor in logistic regression models. The most popular method to take neighbourhood interactions directly into account is by using cellular automata (CA) techniques. In CA-based models neighbourhood interactions are translated into a set of transition rules, which represent push and pull effects between different land use types. The future use of a land unit is not only dependent on exogenous predicting variables, but on the previous state of the cell and on the states of the neighbouring cells. One of the most widespread CA-based models is the SLEUTH model, designed by Clarke et al. (1997) to predict urban expansion in North-American cities. White and Engelen (2000) have developed a CA-based model application

(Verburg, Schot, Dijst, & Veldkamp, 2004). Biophysical factors (i) can have an impact on the suitability of a location for a certain land use type and are often taken into account in land use change models that predict changes in rural areas (Serneels & Lambin, 2001; Van Rompaey, Govers, & Puttemans, 2002; Verburg, Schulp, Witte, & Veldkamp, 2006). In urban modelling, biophysical variables are sometimes incorporated because they may be correlated with the suitability and costs of a parcel for urban construction. Clarke et al. (1997), for instance, included a slope layer in an urban growth model that was applied to simulate urban expansion in a hilly region in North-America. Social factors (ii) can be taken into account in land use models by means of simple indicators such as population density, racial composition or average income (Hu & Lo, 2007). Economic (iii) factors are often implemented in land use models by means of simple accessibility indicators, such as distance to urban centres, distance to the road network and distance to open water (Irwin & Geoghegan, 2001). These variables can be considered as a proxy variable for market access and trade. Spatial policies (iv) on national or regional level largely control urban development (Dieleman & Wegener, 2004). In particular, policies that outline conservational or protected areas or designated areas have a large influence on urban patterns and are therefore often included as an explanatory variable in urban models. Finally, interactions between neighbouring land use types (v) are often implemented as a determining factor in land use models because land use patterns nearly always show spatial autocorrelation caused by a number of centripetal and centrifugal forces (Overmars, de Koning, & Veldkamp, 2003). In urban environments it is especially important to incorporate neighbourhood interactions because of the fact that urban development can be regarded as a self-organising system (Verburg, de Nijs, Ritsema van Eck, Visser, & de Jong,

that is widely used to develop scenarios of urban development. In their model, they adopted the concept of transition probability and defined it as a function of the intrinsic suitability for a certain land use, zoning status, local accessibility and the neighbourhood effect.

Table 1 shows the explanatory variables that have been taken into account in a set of recent model applications to simulate spatial patterns of urban expansion. All of the reviewed model applications take neighbourhood interactions into account; most models consider economic and policy-related variables, while social factors are not often included.

In many cases it is, however, not known to what extent relatively complex model structures, that incorporate several determining factors and combine different modelling approaches, result in more accurate predictions. Recent validation studies have indicated that most spatial models still contain a high level of uncertainty or have not been validated at all. Pontius et al. (2008) have applied a specific set of validation techniques in order to evaluate the performance of nine widely used land use models. They found that in six out of 13 investigated applications the model prediction was less accurate than the prediction of a null model, which predicts no change. This uncertainty of the model results can be caused by error in the input data and by model error, due to an incomplete representation of reality in the model structure (Van Rompaey & Govers, 2002). Moreover, both sorts of error will propagate through the simulation process of different types of models. Van Dessel, Van Rompaey, Szilassi, Jordan, and Csillag (submitted for publication) have quantified the impact of input data quality on error propagation using logistic regression techniques. Yeh and Li (2006) have found that similar issues of data error, model uncertainty and error propagation can be described in urban CA simulations, but they demonstrated that CA models show some unique characteristics as well. Model developers should therefore be aware that the optimal model complexity is dependent on the specific application and on the quality of the available input data (Van Rompaey & Govers, 2002).

The main objective of this paper is to evaluate the added value of using more complex models in terms of model performance. Hereto a series of modelling approaches with an increasing complexity is implemented to simulate the urban expansion patterns in the Flanders–Brussels area in the period 1988–2000. A set of urban expansion models is developed based on: (i) logistic regression equations taking various numbers of predicting variables into account, (ii) CA transition rules and (iii) hybrid procedures, combining both approaches. The model complexity of these different models varies in terms of the number of determining factors that is taken into account and the modelling approach that is used. The model results are validated for the entire study area and for two smaller regions within the study area with distinct urbanisation patterns in order to assess the uncertainty and predictive value of the different models. Sensitivity analysis is performed in order to assess the contribution that the different determining factors make to the accuracy of the predictions.

## 2. Material and methods

### 2.1. Study area

The Flanders–Brussels region (northern part of Belgium) is situated in the heart of Europe and is characterised by a heterogeneous landscape with a high urbanisation rate (Fig. 1). The region (13,700 km$^2$) has a population of approximately 7.1 million people, making it one of the most densely populated regions in the world. Since the 1960s, the main population growth has shifted gradually from the historical city centres into the surrounding countryside and has created a so-called 'rurban' landscape, characterised by a highly fragmented complex mosaic of different land uses (Antrop, 2004). This specific form of urban expansion, sometimes referred to as urban sprawl, has become a common phenomenon throughout Europe. Within the European context, the situation of Flanders and Brussels is however quite remarkable because of the high proportion of built-up land and the strong interwovenness of urban land and open space (Poelmans & Van Rompaey, 2009).

The high road density in Flanders and Brussels has been the most important trigger of the diffusion process of the urban land. The availability of good accessible, cheap open space has attracted large-scale retail activities, which at their part attracted residential areas leading to a widespread ribbon development (Antrop, 2000, 2004). Moreover, the rather permissive spatial policy during most of the 20th century even enhanced this landscape fragmentation and ribbon development. In 1995, Flemish policy makers have implemented the Flemish Structure Plan (RSV), which is a conceptual plan aiming at a 'deconcentrated clustering' of residential, commercial and industrial areas in the existing urban centres, while preserving the remaining open space in the countryside (Albrechts, 1998, 1999; Faludi, 2005; Ministry of the Flemish Community, 2004). In the Brussels–Capital Region, a similar Regional Development Plan (GewOP) has been implemented in

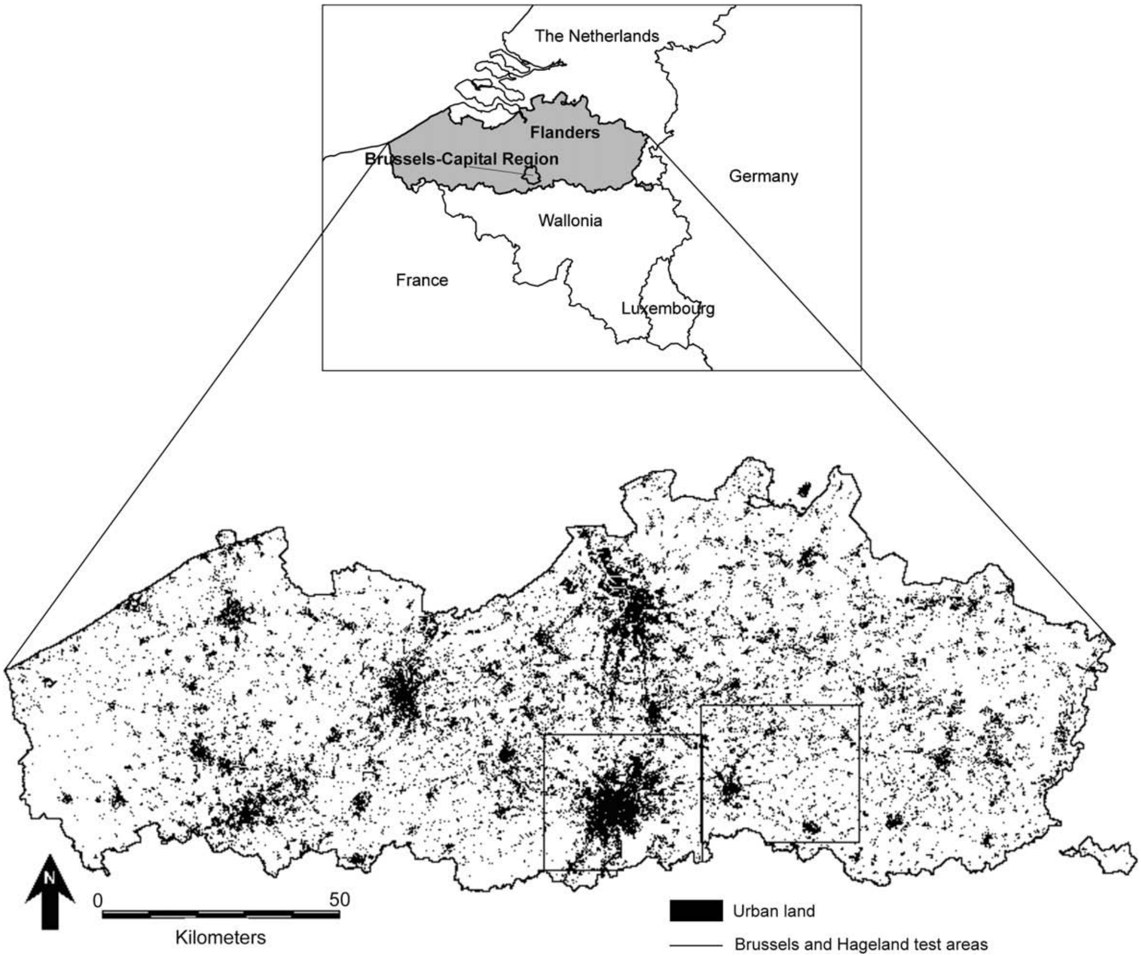


**Fig. 1.** Situation and overview of the Flanders–Brussels study area.

2001 (Albrechts, 1998). One of the shortcomings of both plans, however, is the lack of realistic scenarios of future urban development. A better understanding of the mechanisms of urban expansion in the Flanders–Brussels region is necessary in order to develop a more feasible conceptual planning.

In order to evaluate the model performance in diverse landscapes with different urbanisation patterns, two small test areas within the Flanders–Brussels region were selected: the city of Brussels and its surroundings and the Hageland region, east of Brussels (Fig. 1). Although the two test areas are situated adjacent to each other, they show a quite contrasting urbanisation and urban expansion pattern. The Brussels area is highly urbanised with very few open spaces left and shows a rather compact urban growth. The Hageland area, on the other hand, is less urbanised, but shows a highly fragmented and complicated landscape that consists of a number of urban centres, connected by elongated 'urban corridors' (Poelmans & Van Rompaey, 2009).

### 2.2. Land use data

Landsat satellite images of the study area were acquired for the years 1976 (Landsat-MSS), 1988 (Landsat-TM) and 2000 (Landsat-ETM+). The original spatial resolution of the 1976 image was $60 \times 60$ m, while the 1988 and 2000 images had a $30 \times 30$ m resolution. Land use maps for all selected years were compiled by means of a supervised maximum likelihood classification. The classified land use classes are built-up (urban) land, arable land, grassland, forest and water surfaces. The built-up category contains all paved areas, including residential and commercial zones, industrial

### 2.3. Modelling urban expansion

In this study, two different modelling approaches were implemented and combined with each other in order to create a number of different urban expansion probability maps. Firstly, logistic regression equations were used to associate urban expansion patterns with a number of explanatory variables. The dependent variable was a binary map showing the spatial pattern of urban expansion between 1976 and 1988. A value of 1 in the map indicates that the non-urban cell has changed its land use to urban between 1976 and 1988, a value of 0, on the other hand, means that the cell was already urban in 1976 or that it did not change its use from non-urban to urban between 1976 and 1988. The logistic function calculated the probability of urban expansion for each unit of observation (cell) by means of the following equations:

$$P_{LR} = P(Y = 1|x_1, x_2, \ldots, x_k) = \exp\left(\alpha + \sum \beta_j x_j\right) \Big/ \left(1 + \exp\left(\alpha + \sum \beta_j x_j\right)\right) \quad (1)$$

and

$$\text{logit}P(Y = 1|x_1, x_2, \ldots, x_k) = \ln(P/(1 - P)) = \alpha + \sum \beta_j x_j \quad (2)$$

where $P_{LR}$ is the probability of a cell being built up calculated through the logistic regression procedure, $P(Y = 1\ |x_1, x_2, \ldots, x_k)$ the probability of the dependent variable $Y$ being 1 given $(x_1, x_2, \ldots, x_k)$, i.e. the probability of a cell changing its use to urban land, $x_j$ an independent variable representing a driving force of urbanisation, which can be either categorical or continuous and $\alpha$ and $\beta_j$ are the model parameters that need to be estimated.

areas, roads and parking spaces. The overall accuracy of the three classified land use maps was verified through a comparison of the maps with aerial photographs (at a scale of 1:20,000). The percentage of correctly classified pixels is 77.6% for 1976 and 82.8% for 1988 and 82.3% for 2000. The user's and producer's accuracy of the built-up category is more than 85% for all of the selected years (Poelmans & Van Rompaey, 2009).

Next, the original land use maps were resampled to a resolution of $100 \times 100$ m using the nearest neighbourhood approach. This resampling procedure did not result in a significant alteration of the landscape composition for the selected years: the change in area of the different land use classes was less than 0.2%. The $100 \times 100$ m resolution land use maps were used for all further analyses.

Fig. 2 shows the compiled land use maps at a $100 \times 100$ m resolution.

Logistic regression is an empirical modelling technique in which the selection of the explanatory variables is data-driven rather than knowledge-driven. Nevertheless, an informed selection of independent explanatory variables was made. The independent variables that were used in this study are: (i) distance to cities with more than 50,000 inhabitants in Flanders and the Brussels–Capital region, (ii) distance to the main roads, (iii) slope, (iv) employment potential (Poelmans & Van Rompaey, 2009) and (v) a zoning map, discerning zones where urban development is only partially permitted such as natural reserves, agricultural areas and forested areas (code 0) and zones that are designated for residential, industrial or commercial development (code 1). The five independent variables are illustrated in Fig. 3.

All possible combinations of the five independent variables were analysed, resulting in 1 full model, including all five independent variables, and 30 simplified models, taking only one, two, three or

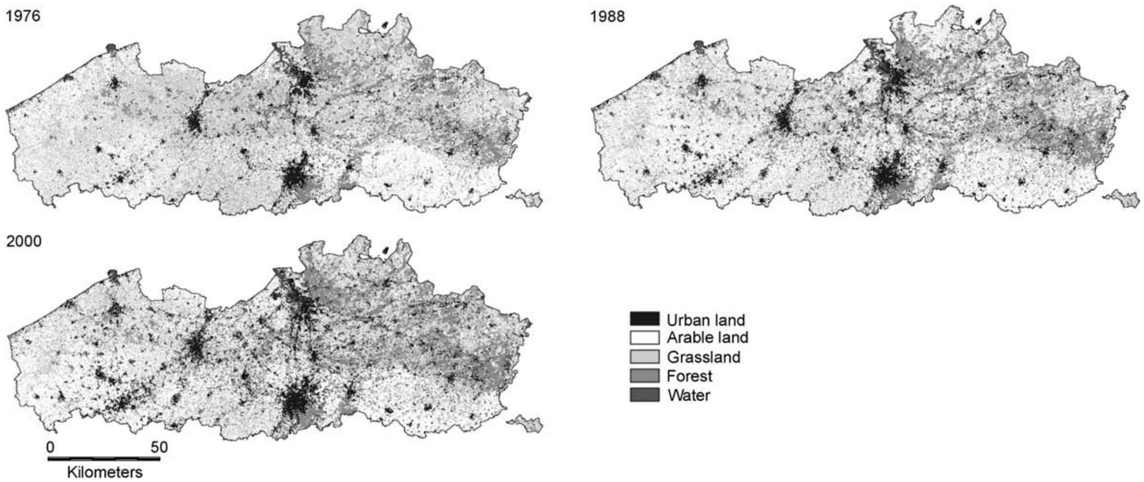


**Fig. 2.** Land use in the Flanders–Brussels study area in 1976, 1988 and 2000.

four of the independent variables into account at the same time. Hence, 31 different probability maps were created based on the logistic regression procedure. This was done in order to evaluate the contribution of each independent variable to the explanation of the spatial pattern of urban expansion. Fig. 5 shows the probability map based on the full complexity logistic regression model for two small test areas within the Flanders–Brussels study area.

In order to minimise the potential effects of spatial autocorrelation on the regression results, all logistic regression models were calibrated using a (stratified) random sample of 30,000 observations (2.5% of the study area) with an equal number of 0 and 1 observations of the dependent variable. Unequal sampling rates do not affect the estimation of the model coefficients $\beta_j$, but only affect the intercept $\alpha$ (Allison, 1999). When using the model results to compute probability maps, the intercept was corrected by distracting $\log(p_1/p_0)$, where $p_1$ (22%) and $p_0$ (1.2%) are the proportions of observations chosen from the two groups for which the dependent variable is 1 and 0, respectively (Allison, 1999). The model parameters $\alpha$ and $\beta_j$ of Eqs. (1) and (2) were estimated by using a maximum likelihood procedure, implemented in SAS® (SAS Institute, 2003).

Secondly, a dynamic CA-based model was developed in order to calculate the probability for urban expansion at a yearly basis. At each iteration (1 year) the CA transition rules calculate the probability of a cell to convert to urban land based on the effects of the neighbouring land uses and then change the cells with the highest probability for urban land. The quantity of conversion was constrained to the observed quantity of new urban land in the studied time period divided evenly over the number of years in the studied time period. The transition probability for each time step was calculated according to a procedure proposed by White and Engelen (2000):

$$P_{CA} = N_{bu} = \sum_{x} \sum_{d \in D} w_{kxd} \cdot I_{xd} \tag{3}$$

where $P_{CA}$ is the probability of a certain cell being built up, based on the CA transition rules, $N_{bu}$ the neighbourhood effect on this cell, $w_{kxd}$ the weighting parameter applied to a cell with land use $k$ (urban land, arable land, grassland or forest) at position $x$ in distance zone $d$ of the neighbourhood $D$ and $I_{xd}$ is 1 if neighbouring cell $x$ in distance zone $d$ is occupied by land use $k$ or 0 otherwise.

The weighting parameters that define the neighbourhood's attraction (positive weights) or repulsion (negative weights) for urban land needed to be calibrated. The neighbourhood space $D$ was defined as a square region around the central cell and contains 168 cells that are arranged in six square distance zones $d$. In the calibration phase, the 'best set' of weighting parameters for the period 1976–1988 was assessed. Hereto, the model was run for a range of $w_{kxd}$ values. For each model run, the goodness-of-fit between the obtained probability map for 1988 and the actual change map (between 1976 and 1988) was calculated by means of the ROC procedure as proposed by Pontius and Schneider (2001) for the validation of land use change models. The set of $w_{kxd}$ coefficients that produced the highest agreement was kept in order to produce a probability map of urban expansion.

Theoretically, it is possible to include neighbourhood interactions directly in an empirical approach such as logistic regression as an explanatory variable (Verburg, de Nijs et al., 2004; Verburg, Ritsema van Eck et al., 2004). However, because the logistic regression method is essentially static, it is not able to reveal the path-dependent and self-organising development that is typical for urban expansion (Wu, 2002).

Finally, the two modelling approaches were combined to create a so-called 'hybrid model', which takes both the independent explanatory variables and the dynamic neighbourhood interactions into account. The hybrid model is a dynamic modelling approach, similar to the CA-based model, since it updates the land use maps at each time step of the simulation. In the case of the hybrid model, the probability for urban expansion was defined as follows:

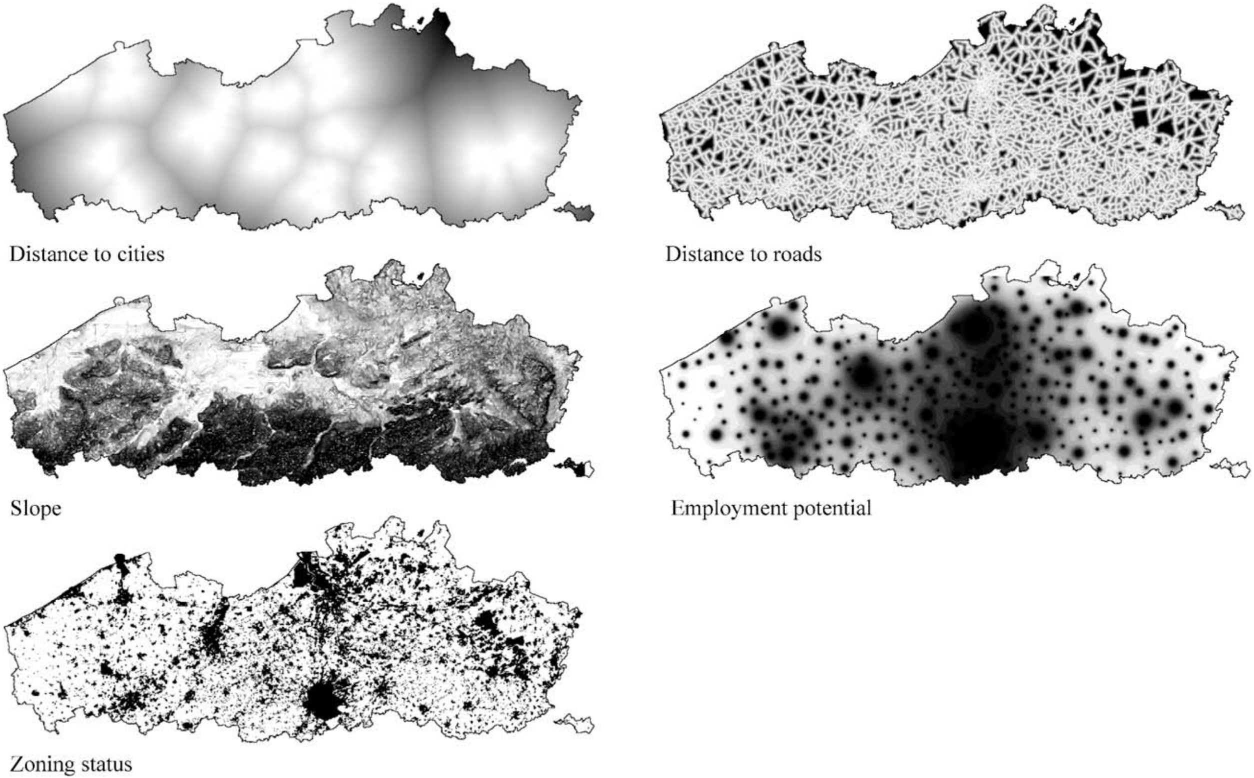


**Fig. 3.** Independent variables applied in the logistic regression procedure (darker colours correspond with higher values).

$$P_{hyb} = P_{LR} \cdot N_{bu}^{\sigma} \quad (4)$$

where $P_{hyb}$ is the transition probability for a certain cell based on both (exogenous) explanatory variables and (endogenous) neighbourhood interactions, $P_{LR}$ the probability map produced by the full complexity logistic regression model (Eq. (1)), $N_{bu}$ the neighbourhood effect based on the calibrated weights (Eq. (3)) and $\sigma$ is a parameter which expresses the importance of the neighbourhood effect $N_{bu}$. In this study, the value of parameter $\sigma$ was set to 1 because this produced the highest agreement according to the ROC procedure. The probability map for the hybrid model at the last iteration is shown in Fig. 5. Similar approaches to create hybrid model structures were applied by White and Engelen (2000), Barredo, Kasanko, McCormick, and Lavalle (2003) and Wu (2002).

### 2.4. Model validation and comparison

In order to assess the predictive value of an urban model, it is important to make a clear distinction between the calibration phase of model and the validation of the model results (Pontius et al., 2004). Calibration refers to the process of creating a model such that it is consistent with the data used to create the model (Verburg, Schulp et al., 2006). In this paper, calibration of the model parameters for the different modelling approaches was based on the land use maps from 1976 and 1988. The calibrated parameters $\alpha$ and $\beta_j$ of the logistic regression model (Eq. (1)) and the calibrated weights $w_{kxd}$ of the CA-based model (Eq. (3)) were subsequently used to simulate urban expansion patterns for the 1988–2000 period. Model results for the 1988–2000 period were compared to the observed urban pattern in 2000 and validated by using two complementary goodness-of-fit measures: (i) the ROC statistic (Pontius & Schneider, 2001), which was used to evaluate the obtained probability maps and (ii) the null resolution (Pontius et al., 2004), which was used to examine the spatial patters of urban expansion of model performance than a goodness-of-fit measure such as the percentage of correctly assigned pixels (Fawcett, 2006).

By itself, the ROC statistic can only provide a relative indication of how well a certain model performs, but it remains a useful tool when comparing the performances of several models in order to obtain an optimal probability map.

Secondly, for each simulation the percentage of pixels that are classified correctly as urban land was calculated. Hereto, non-urban cells with the highest probability for urban land were converted to urban land until the required area of urban land in the simulated land use map, based on the actual land use map of 2000, was met. Because for each simulation the exact number of new urban cells was allocated, no quantification error occurred and all observed error on the predicted map could be attributed to location error, related to the probability maps and the allocation procedure (Pontius, 2000). The percentage correct is the most commonly used goodness-of-fit measure because it is simple to compute and relatively easy to interpret. However, pixel-to-pixel analysis can fail to detect spatial patterns because it ignores the land use in the neighbourhood. Therefore, it is advisable to use multiple resolution techniques when comparing two land use maps (Pontius, 2002; Pontius et al., 2004). This multiple resolution comparison was carried out by averaging neighbouring pixels at the original resolution of $100 \times 100$ m into coarser pixels both on the simulated and on the reference land use map and by calculating the percentage correct on the newly developed maps.

Finally, Pontius et al. (2004) have stated that the agreement between a simulated land use map and a reference land use map should be better than the agreement between the land use map resulting from a null model and the reference map. A null model is a naïve model that predicts 'no change' or pure persistence of the existing land use. The null resolution of a certain model run is defined as the spatial resolution at which the predictive model is as accurate as the corresponding null model.

by comparing them to a ‘no change’ scenario.

Firstly, the probability maps were evaluated by using the Relative Operating Characteristic-procedure (ROC) as it was proposed by Pontius and Schneider (2001). The ROC compares the predicted probability maps to a map with the observed expansion of urban land between 1988 and 2000 in order to validate the model’s ability to specify the location of change, independently from the simulated amount of change. The procedure first calculates the proportion true-positives and false-positives for a range of thresholds and relates them to each other in a graph. The ROC statistic measures the area under the curve and varies between 0.5 (complete random assignment of the probabilities) and 1 (perfect probability assignment). Since the ROC procedure results in a figure that integrates several contingency tables, it is able to provide a richer measure

## 3. Results and discussion

### *3.1. Model interpretation*

The calibrated model coefficients of the full logistic regression model and of the five logistic regression models that only take one of the explanatory variables into account are shown in Table 2. The model coefficients of the other simplified models, that each include two, three or four of the explanatory variables are not shown. In the full model all explanatory variables, except the employment potential, are statistical significant at the 0.01 level. Based on the Wald statistics calculated for each parameter in the full model and on the likelihood ratios of the simplified models,

**Table 2**
Logistic regression results of six models.

| | | Intercept | Distance to cities | Slope | Employment potential | Distance to roads | Zoning status |
|---|---|---|---|---|---|---|---|
| Full model | Coefficient (odds ratio) | −1.864 | −0.00002 (1.000) | −0.0822 (0.921) | 0.000012 (1.000) | −0.00191 (0.998) | 2.5672 (13.030) |
| | Wald | 227.8 | 126.7 | 66.6 | 1.4 | 1293.0 | 6816.8 |
| Simplified model 1 | Coefficient (odds ratio) | −1.062 | −0.00004 (1.000) | n.i. | n.i. | n.i. | n.i. |
| | Wald | 251.1 | 882.2 | – | – | – | – |
| Simplified model 2 | Coefficient (odds ratio) | −1.378 | n.i. | −0.0776 (0.925) | n.i. | n.i. | n.i. |
| | Wald | 6.3 | – | 100.2 | – | – | – |
| Simplified model 3 | Coefficient (odds ratio) | −1.487 | n.i. | n.i. | 0.0003 (1.000) | n.i. | n.i. |
| | Wald | 128.6 | – | – | 39.9 | – | – |
| Simplified model 4 | Coefficient (odds ratio) | −0.638 | n.i. | n.i. | n.i. | −0.00258 (0.997) | n.i. |
| | Wald | 1416.4 | – | – | – | 2780.0 | – |
| Simplified model 5 | Coefficient (odds ratio) | −1.5228 | n.i. | n.i. | n.i. | n.i. | 2.7955 (16.371) |
| | Wald | 4940.4 | – | – | – | – | 8711.8 |

n.i. Not included.

the major determinant of the spatial pattern of urban expansion seems to be the zoning status. The model coefficients ($\beta_j$) of this variable range from 2.5 to 2.8 for the different simplified models and the full model, resulting in odds ratios ranging from 12 to 16.5 (odds ratio = $e^{\beta}$). This means that it is around 12–16 times more likely to find new urban land in areas designated to urban land than in areas designated to other land uses. The model results also demonstrate that, as in Cheng and Masser (2003) and Hu and Lo (2007), the distance to the main roads plays an important role in explaining urban development. For each kilometre away from a road, it is 13 times less likely to find new urban development (odds ratio = $e^{1000^{*}-0.00258} \approx 1/13$). In the full model, the influence of the distance to the roads decreases somewhat (odds ratio = $e^{1000^{*}-0.00191} \approx 1/6.5$). The road pattern contributes to the spatial pattern of ribbon development, typical for Flanders and Brussels and many other regions in the world. Furthermore, new urban development tends to occur near to the largest cities and on relatively flat terrains. However, the contribution of these variables to the explanation of the urban expansion patterns is quite small as is demonstrated by their lower Wald statistics. Finally, the employment potential, which shows the lowest Wald statistic of the included explanatory variables, has a minor influence on the patterns of urban expansion. Both the estimated model coefficients of the employment potential and the distance to the cities are close to zero, which proves that urban expansion is not restricted to the city centres or the periphery, but is practically affecting the whole countryside. The obtained probability map of the full logistic regression model is shown in Fig. 5 for the Hageland and Brussels test areas.

According to the CA transition rules, urban land is attracted by other urban land in the neighbourhood. The calibrated weights that define the transition rules of the CA-based model are illustrated in Fig. 4. The graph shows the distance functions that from the existing land use. The weighting parameters show that the conversion from forested land to urban land is quite uncommon, while arable land and grassland will urbanise more easily. The pull effect of urban land for new urban development is strongest in the immediate neighbourhood of the cell, decreases exponentially and becomes neutral at a distance of around 600 m. This distance-decay effect is in line with the results reported in other studies. Hagoort, Geertman, and Ottens (2008) used survey-based neighbourhood rules, based on expert knowledge and calibration-based rules, based on spatial metrics to simulate land use change in four urban regions in the Netherlands. They reported that existing residential land was the most attractive neighbour for new residential areas up till a distance of 800 m. Barredo et al. (2003) applied an urban CA model in the city of Dublin. The calibration of the parameter weights was carried out by means of: (i) a visual comparison of the simulation map with the actual land use map, (ii) a comparison between the fractal dimensions of the maps and (iii) a quantitative evaluation using comparison matrices. They found that residential land was attracted to itself up till a distance of 600 m.

Fig. 5 shows the transition probability map $P_{CA}$ of the year 1988 (i.e. the last iteration of the calibration period) for the Hageland and Brussels test areas. The probability is higher in the immediate neighbourhood of urban land, gradually decreases in the open space and shows the lowest values in large forested areas.

Finally, Fig. 5 shows the transition probability map of 1988 based on the hybrid modelling approach with equal importance for the neighbourhood effects and for the logistic regression approach ($\sigma = 1$).

### 3.2. Model validation and comparison

Model performance was assessed: (i) by evaluating the probaexpress the interaction between four land use categories (urban land, arable land, grassland and forest) and urban land. The interaction effect between the land use categories and urban land is limited to relatively short distances. Arable land, grassland and forest only play a substantial role at the zero-distance, which can be interpreted as the ease with which urban land will take over

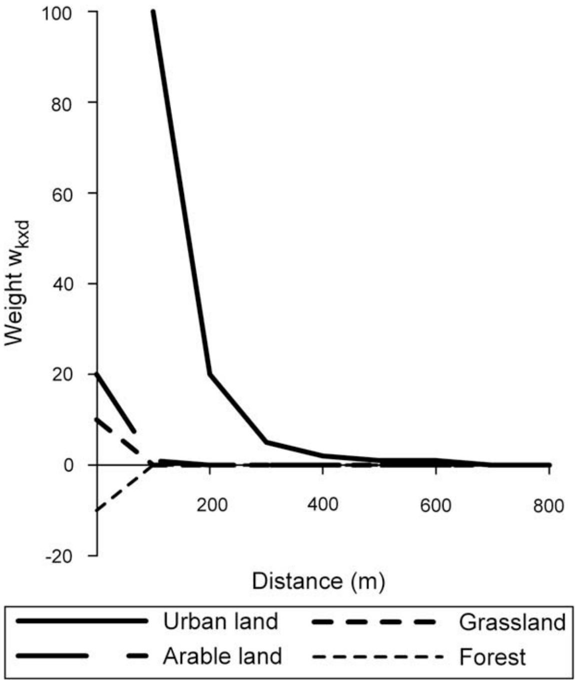


**Fig. 4.** Calibrated CA transition rules.

bility maps computed by the different modelling approaches using the ROC and (ii) by comparing the simulated land cover maps of 2000 with the reference map of 2000 at multiple resolutions. The ROC statistic was calculated for the Flanders and Brussels area and for two smaller test areas within Flanders and Brussels (Hageland and Brussels region).

When using the ROC statistic to compare the different modelling approaches, the hybrid model gives the best results (Table 3; Fig. 6, left). This is in agreement with Dendoncker et al. (2007), Verburg, Ritsema van Eck et al. (2004) and Wu (2002) who all reported that adding neighbourhood characteristics, such as the interactions between neighbouring land use classes, to logistic regression models improved the performance of these models. While Dendoncker et al. (2007) and Verburg, Ritsema van Eck et al. (2004) used static neighbourhood relations (e.g. proportion of urban land in the neighbourhood at a certain point in time), Wu (2002) used a dynamic approach, similar to the one applied in this study. The performance of the hybrid model is only slightly better than the performance of the CA-based model and the most accurate logistic regression models, which show comparable results. This is against the expectations because neighbourhood relations are only able to capture part of the processes that govern urban expansion, while the logistic regression approach, based on several driving factors, is in principle much better to capture the full complexity of the involved processes (Verburg & Overmars, 2007). Despite of the fact that driving factors are not taken directly into account they are not completely ignored in the CA-based approach. This is due to the fact that the simulations are based on the existing land use pattern of 1988, which indirectly reflects the influence of the determining factors.

When comparing the 31 logistic regression models based on the ROC statistic, it can be seen that there is a lot of variability between the performances of the different models in the Flanders–Brussels

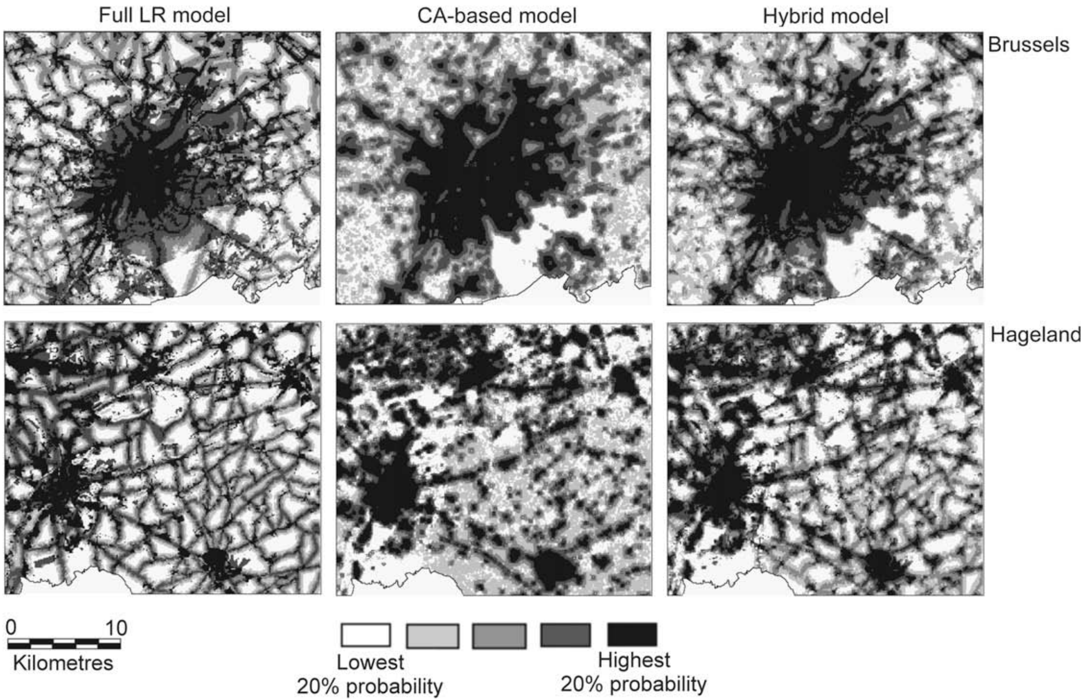


**Fig. 5.** Probability maps produced by the different urban expansion models (illustrated for two test areas within the study area).

region (Table 3; Fig. 6, left). Some of the models that only take two, three or four of the explanatory variables into account, perform as well as the full complexity logistic regression model, while other simplified models show significantly lower ROC values. The model performance thus is highly sensitive to the explanatory variables that are taken into account. Table 3 shows that the model shows greatest sensitivity to the exclusion of the zoning map: the ROC values of the models excluding the zoning status are all smaller

**Table 3**
ROC values (%) and null resolutions of the 31 logistic regression models, the CA-based model and the hybrid model.

| Distance to cities | Slope | Employment potential | Distance to roads | Zoning status | CA-rules | ROC (%) | Null resolution (m) |
|---|---|---|---|---|---|---|---|
| x | x | x | x | x | | 82.6 | 622 |
| | x | x | x | x | | 82.6 | 602 |
| x | | x | x | x | | 82.6 | 636 |
| x | x | | x | x | | 82.5 | 670 |
| x | x | x | | x | | 79.9 | 10,338 |
| x | x | x | x | | | 71.4 | 3874 |
| | | x | x | x | | 82.5 | 624 |
| | x | | x | x | | 82.3 | 655 |
| | x | x | | x | | 79.8 | 17,252 |
| | x | x | x | | | 68.9 | 1204 |
| x | | | x | x | | 82.5 | 662 |
| x | | x | | x | | 79.8 | 13,519 |
| x | | x | x | | | 71.0 | 2886 |
| x | x | | | x | | 79.6 | 10,667 |
| x | x | | x | | | 70.4 | 2125 |
| x | x | x | | | | 63.2 | 20,975 |
| | | | x | x | | 82.2 | 674 |
| | | x | | x | | 80.0 | 21,760 |
| | | x | x | | | 70.5 | 4389 |
| | x | | | x | | 77.6 | 4667 |
| | x | | x | | | 68.6 | 1239 |
| | x | x | | | | 62.3 | 43,886 |
| x | | | | x | | 79.5 | 11,247 |
| x | | | x | | | 70.3 | 2092 |
| x | | x | | | | 63.0 | 22,497 |
| x | x | | | | | 61.5 | 18,598 |
| x | | | | | | 61.8 | 16,661 |
| | x | | | | | 51.8 | 11,055 |
| | | x | | | | 62.2 | 60,543 |
| | | | x | | | 68.5 | 2245 |
| | | | | x | | 77.5 | 10,861 |
| | | | | | x | 82.5 | 2557 |
| x | x | x | x | x | x | 85.5 | 1400 |

than 71.5%. In contrast, exclusion of the variables slope, employment potential and distance to the cities has less effect on the accuracy of the probability map.

The ROC statistics were also calculated for two contrasting test areas within Flanders and Brussels. Fig. 6 (right) shows that the ROC values for all three modelling approaches are somewhat higher in the Brussels area than in the Hageland area. In the Hageland area, the most accurate logistic regression model produces a probability map with a ROC of 81.3%. This is slightly better than the performance of the CA-based model, which results in a ROC of 79.4%. In the Brussels test area, in contrast, the CA-based model and the best performing logistic regression models show comparable ROC statistics around 84%. In both test areas, distance to the roads and zoning status are the most important variables to include in the logistic regression, just like in the larger Flanders–Brussels study area.

When looking at the null resolutions calculated for the different models, the conclusions are somewhat different (Table 3; Fig. 7). The logistic regression models which incorporate at least the variables distance to the roads and the zoning status show the lowest null resolution: at a resolution of approximately 600–700 m the null model is outperformed by the model predictions. The CA-based model results in a null resolution of 2.5 km, while the null resolution of the hybrid model is situated somewhere in between, around 1.4 km. This means that when the variables distance to the roads and zoning status are incorporated in the model the location errors on the simulated land cover maps occur over relatively small distances: they resolve when the resolution becomes only slightly coarser.

Based on the estimated null resolutions one could say that the performance of all tested models is relatively poor: not one of the model simulations is able to perform better than the null model at high resolutions. Pontius et al. (2008), however, have indicated that it is common for land cover models to be less accurate than their null models at fine resolutions. In their study, they reported this finding for six out of 13 investigated model applications. The seven model applications that performed better than their null model at all resolutions either had a very coarse spatial resolution (>15,000 m) or a relatively high amount of observed change in the reference maps. Furthermore, the Flanders–Brussels region is a quite unique setting with very high landscape fragmentation and a scattered urban development. This fragmented appearance of the landscape makes the region a rather challenging study area for doing model simulations (Poelmans & Van Rompaey, 2009).

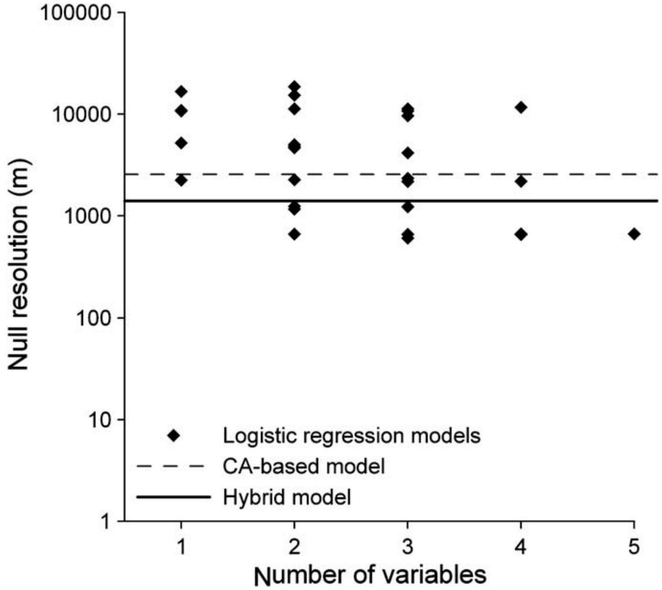


**Fig. 7.** Null resolution of the different urban expansion models calculated for the Flanders–Brussels study area.

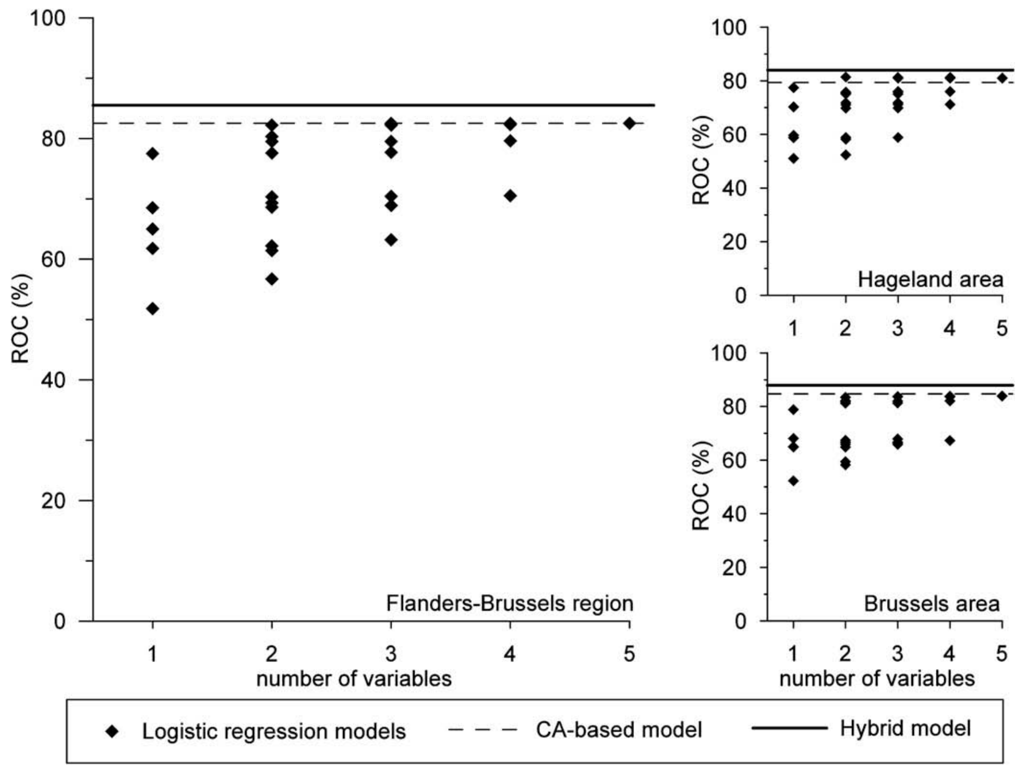


**Fig. 6.** ROC values (%) of the different urban expansion models calculated for the Flanders–Brussels study area (left) and for the Hageland and Brussels test areas (right).

## 4. Conclusions

In this paper, three different frequently used modelling approaches were applied in order to predict spatial patterns of urban expansion in the highly fragmented Flanders–Brussels region in Belgium. All three applications were calibrated based on the observed urban expansion in the 1976–1988 period and used to simulate urban growth in the period 1988–2000. Firstly, five exogenous explanatory variables were used and included in logistic regression equations in order to create 31 urban probability maps. The full complexity model, including all five predicting variables, distance to the main roads, distance to the largest cities, slope, employment potential and zoning status resulted in the best possible probability map. Sensitivity analysis revealed that the zoning status was the most important determinant of the spatial pattern of urban expansion in Flanders and Brussels. Excluding the zoning map from the model resulted in probability maps with significantly smaller ROC statistics. In many urban studies however, spatial policies, such as the zoning status or legally protected areas, are not included because the necessary data is not available in spatially explicit digital form. The distance to the road network seemed to be the only variable that was able to capture the highly fragmented appearance of the urban landscape. Exclusion of this variable from the logistic regression model decreased the ROC statistic of the probability map to a lesser extent than the exclusion of the zoning status did, but raised the null resolution of the simulated land cover map dramatically. This finding supports the theory of urban sprawl, which is defined as a low density outward expansion of urban areas, mainly related to the road network (Ewing, Pendall, & Chen, 2002). The distance to the largest cities and the employment potential, on the other hand, only showed little additional predictive value. While in the 1950s and 1960s, most of the urban growth took place in the city centres and the inner urban fringe, often near to the centres of employment, urban development is now practically affecting the whole countryside (Antrop,

to be the best way to deal with these different forms of urban expansion since it combines the benefits of the composing modelling approaches: empirical analysis of the determining factors of urban growth, incorporation of neighbourhood interactions and a dynamic simulation of the spatial patterns of urban sprawl. However, when validating the three different modelling approaches for two smaller test areas with distinct spatial patterns of urbanisation (compact vs. fragmented), similar conclusions were drawn. The hybrid model was the most accurate model in both the Hageland as in the Brussels region.

Finally, this study pointed out that it is not useful to make a statement about the validity of a model based on only one goodness-of-fit measure. Moreover, it is important to recognise that no single modelling approach can meet with all possible validation criteria, in all different sorts of environments and at all spatial scales. In this study, the hybrid model seemed to be the most accurate model based on the ROC statistic, but showed a significantly higher null resolution than a simple logistic regression model based on only two different explanatory variables. Land use modellers should select the modelling approach and validation criteria that best fit the research question and characteristics of the study area.

## References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). *A review and assessment of land-use change models: Dynamics of space, time and human choice (General technical report NE-297 ed.)*. Newton Square, PA: USDA Forest Service.

Albrechts, L. (1998). The Flemish diamond: Precious gem and virgin area. *European Planning Studies, 6*(4), 411–424.

Albrechts, L. (1999). Planners as catalysts and initiators of change. The new structure plan for flanders. *European Planning Studies, 7*(5), 587–603.

Allison, P. D. (1999). *Logistic regression using SAS®: Theory and application*. Cary, NC: SAS Institute.

Antrop, M. (2000). Changing patterns in the urbanized countryside of Western Europe. *Landscape Ecology, 15*, 257–270.

ment is now practically affecting the whole countryside (Antrop, 2004).

The ROC statistic of the CA-based model, which was only based on neighbourhood interactions, was similar to the numbers reported for the best performing logistic regression models. This is due to the fact that CA-based models are driven by the actual spatial patterns of the land use, in which the underlying determining factors of urban expansion are explicitly embedded. The major advantage of using a CA-based approach, like the one applied in this study, is that the neighbourhood interactions, which drive the model, can be derived using a single land use dataset for the whole study area, without having to collect additional data layers. Moreover, as opposed to the static regression models, dynamic CA-based models are able to simulate path dependence and self-organisation, which are important features of the urban environment. However, like logistic regression models CA models are very sensitive to error propagation. Yeh and Li (2006) have demonstrated that potential input errors can be reduced during a CA simulation because of the averaging effect of the neighbourhood functions implemented in CA models. Nevertheless, CA models contain a lot of inherent model uncertainties that are related to a number of elements such as the neighbourhood, the transition rules, the cell size and the computation time, which define the CA model. Both the input errors and the model related uncertainties can propagate through the simulation process.

Combining both approaches in a so-called hybrid model resulted in the most accurate probability map for the Flanders–Brussels region. The study area consists of several urban centres, a large semi-urbanised and highly fragmented area and some remote parts of open land in the countryside, which are all characterised by different forms of urban expansion. The hybrid model seemed


Antrop, M. (2004). Landscape change and the urbanization process in Europe. *Landscape and Urban Planning, 67*(1–4), 9–26.

Barredo, J. I., Kasanko, M., McCormick, N., & Lavalle, C. (2003). Modelling dynamic spatial processes: Simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning, 64*(3), 145–160.

Cheng, J., & Masser, I. (2003). Urban growth pattern modeling: A case study of Wuhan city, PR China. *Landscape and Urban Planning, 62*(4), 199–217.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design, 24*, 247–261.

Dendoncker, N., Rounsevell, M., & Bogaert, P. (2007). Spatial analysis and modelling of land use distributions in Belgium. *Computers, Environment and Urban Systems, 31*(2), 188–205.

Dieleman, F., & Wegener, M. (2004). Compact city and urban sprawl. *Built Environment, 30*(4), 308–323.

Ewing, R., Pendall, R., & Chen, D. (2002). *Measuring sprawl and its impact*. Washington, DC: Smart Growth America.

Faludi, A. (2005). The Netherlands: A culture with a soft spot for planning. In B. Sanyal (Ed.), *Comparative planning studies* (pp. 285–307). New York/London: Routledge.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Hagoort, M., Geertman, S., & Ottens, H. (2008). Spatial externalities, neighbourhood rules and CA land-use modelling. *The Annals of Regional Science, 42*, 39–56.

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

Irwin, E. G., & Geoghegan, J. (2001). Theory, data, methods: Developing spatially explicit economic models of land use change. *Agriculture, Ecosystems and Environment, 85*(1–3), 7–24.

Kaimowitz, D., & Angelsen, A. (1998). *Economic models of tropical deforestation: A review*. Bogor, Indonesia: Centre for International Forestry Research.

Lambin, E. F., Rounsevell, M. D. A., & Geist, H. J. (2000). Are agricultural land-use models able to predict changes in land-use intensity? *Agriculture, Ecosystems and Environment, 82*, 321–331.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., et al. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change, 11*(4), 261–269.

Mertens, B., & Lambin, E. F. (1997). Spatial modelling of deforestation in southern Cameroon: Spatial disaggregation of diverse deforestation processes. *Applied Geography, 17*(2), 143–162.

Ministry of the Flemish Community (2004). *Spatial structure plan for Flanders*. Brussels: Ministry of the Flemish Community (in Dutch).

Overmars, K. P., de Koning, G. H. J., & Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling, 164*, 257–270.

Poelmans, L., Van Rompaey, A. (2009). Detecting and modelling spatial patterns of urban sprawl in the Flanders–Brussels region (Belgium). *Landscape and urban planning*, in press.

Pontius, R. G. Jr., (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing, 66*(8), 1011–1016.

Pontius, R. G. Jr., (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing, 68*(10), 1041–1049.

Pontius, R. G., Jr., Boersma, W., Castella, J., Clarke, K., de Nijs, T. C. M., Dietzel, C., et al. (2008). Comparing the input, output and validation maps for several models of land change. *Annals of Regional Science, 42*, 11–37.

Pontius, R. G. J., Cornell, J. D., & Hall, C. A. S. (2001). Modeling the spatial pattern of land-use change with GEOMOD2: Application and validation for Costa Rica. *Agriculture, Ecosystems and Environment, 85*, 191–203.

Pontius, R. G., Jr., Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling, 179*(4), 445–461.

Pontius, R. G., Jr., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment, 85*(1–3), 239–248.

SAS Institute (2003). *SAS® 9.1.3. Help & Documentation*. Cary, NC: SAS Institute Inc..

Schaldach, R., & Priess, J. A. (2008). Integrated models of the land system: A review of modelling approaches on the regional to global scale. *Living Reviews in Landscape Research, 2*, 1–34.

Schneider, L. C., & Pontius, R. G. Jr., (2001). Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment, 85*(1–3), 83–94.

Serneels, S., & Lambin, E. F. (2001). Proximate causes of land-use change in Narok District, Kenya: A spatial statistical model. *Agriculture, Ecosystems and Environment, 85*(1–3), 65–81.

Van Dessel, W., Van Rompaey, A., Poelmans, L., & Szilassi, P. (2008). Predicting land cover changes and their impact on the sediment flux in the Lake Balaton catchment. *Landscape Ecology, 23*, 645–656.

Van Dessel, W., Van Rompaey, A., Szilassi, P., Jordan, G., Csillag, G. (submitted for publication). Sensitivity analysis of logistic regression parameterization for land use and land cover probability prediction. *International Journal of GIS*.

Van Rompaey, A. J. J., & Govers, G. (2002). Data quality and model complexity for regional scale soil erosion prediction. *International Journal of Geographical Information Science, 16*(7), 663–680.

Van Rompaey, A. J. J., Govers, G., & Puttemans, C. (2002). Modelling land use changes and their impact on soil erosion and sediment supply to rivers. *Earth Surface Processes and Landforms, 27*, 481–494.

Veldkamp, A., & Fresco, L. O. (1996). CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological Modelling, 91*(1–3), 231–248.

Verburg, P. H., de Nijs, T. C. M., Ritsema van Eck, J., Visser, H., & de Jong, K. (2004). A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems, 28*, 667–690.

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

Verburg, P. H., & Overmars, K. P. (2007). Dynamic simulation of land-use change trajectories with the CLUE-s model. In E. Koomen, J. Stillwell, A. Bakema, & H. J. Scholten (Eds.), *Modelling land-use change* (pp. 321–335). Dordrecht, The Netherlands: Springer.

Verburg, P. H., Ritsema van Eck, J. R., de Nijs, T. C. M., Dijst, M. J., & Schot, P. (2004). Determinants of land-use change patterns in the Netherlands. *Environment and Planning B: Planning and Design, 31*, 125–150.

Verburg, P. H., Schot, P., Dijst, M., & Veldkamp, A. (2004). Land use change modelling: Current practice and research priorities. *GeoJournal, 61*(4), 309–324.

Verburg, P. H., Schulp, C. J. E., Witte, N., & Veldkamp, A. (2006). Downscaling of land use change scenarios to assess the dynamics of European landscapes. *Agriculture, Ecosystems and Environment, 114*(1), 39–56.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. A. (2002). Modeling the spatial dynamics of regional land use: The CLUE-s model. *Environmental Management, 30*(3), 391–406.

Ward, D. P., Murray, A. T., & Phinn, S. R. (2000). A stochastically constrained cellular model of urban growth. *Computers, Environment and Urban Systems, 24*, 539–558.

White, R. (1998). Cities and cellular automata. *Discrete Dynamics in Nature and Society, 2*, 111–125.

White, R., & Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems, 24*(5), 383–400.

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

Yeh, A. G. O., & Li, X. (2006). Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems, 30*, 10–28.