International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Development of an Activity-based Cellular Automata Land-use Model: the case of Flanders, Belgium

Tomas Crols[a,b], Roger White[c], Inge Uljee[b], Guy Engelen[b], Frank Canters[a], Lien Poelmans[b]
[a]VUB – Vrije Universiteit Brussel, Cartography and GIS Research Group, Department of Geography, Pleinlaan 2, 1050 Brussel – Belgium
[b]VITO NV – Flemish Institute for Technological Research, Unit Spatial Environmental Modelling, Boeretang 200, 2400 Mol – Belgium
[c]MUN – Memorial University of Newfoundland, Department of Geography, St. John's, NL A1B 3X9 – Canada
[a]tcrols@vub.ac.be

Abstract: Cellular automata (CA) models are increasingly applied for simulating land-use change in urban areas. However, in areas with strongly mixed land uses, like Flanders, Belgium, different types and intensities of human activity occur within a single dominant land use. This is in conflict with the discrete and dominant land-use states applied in CA. The direct modelling of the intensity of activities (population density and employment in different sectors) within a CA grid environment is an interesting alternative to model mixed and multifunctional land use. In this research, an activity-based cellular automata (ACA) model, developed by White et al. [2012] will be further enhanced, applied and calibrated for Flanders. Its resolution will be increased to 1 ha to effectively address environmental, socio-economic and spatial planning problems. It should be able to cope with the complex multi-nodal structure and messy urbanised morphology of Flanders, typified as it is by multifunctional land use and diffuse, fragmented urban development strung out along roads. This paper shows the results of an initial application of the model of White et al. to a sub-region of Flanders and discusses how the model should be further developed in the future. The effect of diseconomies of agglomeration, accounting for high costs and congestion in dense urban areas, was investigated, as well as the capacity of the ACA to allocate both activity-based land uses and non-activity based land uses (e.g. protected nature).

***Keywords**:* Land-use change; CA model; Activity-based model; Variable grid; Network distances.

1 INTRODUCTION

Land-use change models are increasingly used to simulate the future spatial development of regions. Cellular automata (CA) models perform well because of their dynamic and realistic character [Poelmans and Van Rompaey 2010]. A big disadvantage of all classical CA models (and most other land-use change models) is that they define land-use categories in a strictly discrete manner – one cell is one state – which is not optimal for regions exhibiting mixed land uses like Flanders, Belgium.

Additionally, many environmental or economical applications are better served by the detailed spatial location and level of distinct activities (e.g. population and employment levels) rather than the location of fairly abstract and aggregated land-use types. Four examples are (1) the location of new power plants in regions exhibiting population and/or economic growth, (2) the generation of waste products by individuals, (3) the exposure to health risks for an increasing number of individuals living in highly polluted areas or (4) predicting population living close to major roads, a category of people that would increase by 21% in Flanders in 2030 according to the Flemish Environment Outlook 2030 [Van Steertegem et al. 2009]. Multi agent systems (MAS) directly simulate the interaction between actors but are computationally slow and as such not useful for application to large regions [Parker et al. 2003]. A possible solution are activity-based cellular automata (ACA) [van Vliet et al. 2011; White et al. 2012] that model directly interactions between multiple types of activity and between activities and land uses. This results in a prediction of population and employment levels as well as the associated land uses, at a high spatial resolution.

Another problem of classical CA models is that they don't account for long distance spatial interactions. Traditionally this is solved by linking the CA model to a spatial-interaction based model representing dynamics among larger administrative regions [e.g. White and Engelen 2000; Engelen et al. 2007]. However, this is suboptimal as many parameters are needed to specify and couple both types of models. Moreover, the regions are typically compacted into centroids which poorly represent the full regions. Also, intermediate spatial scales are not represented [White et al. 2012].

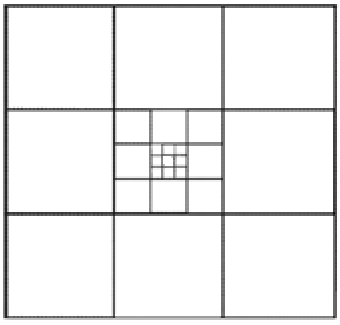

Figure 1. Neighbouring cells of level 0 through level 2 around a unit-cell in the variable grid approach.

This can be solved by using a variable grid approach where the entire modelling area serves as the neighbourhood around a unit-cell [White 2005; van Vliet et al. 2009]. For computational reasons, rings of super-cells are defined which become larger and larger further away from the unit-cell (central level 0 cell). Each super-cell of level *L* consists of 9 (super-)cells of level $L - 1$, and of 3^{2L} unit-cells (cf. figure 1).

In our research, the model developed by White et al. [2012] was applied to Flanders. A full calibration exercise and a major enhancement of the model are scheduled for the future. Research goals include a realistic computation of distances needed to calculate the variable grid neighbourhood effect, a methodology to disaggregate population statistics from a government level to the model resolution, improved historical calibration techniques, and the development of new policy-oriented spatially-dynamic indicators based on activities rather than on land uses.

The current model will be briefly explained in the next section. The practical application to Flanders and some problems encountered in the current early research stage will be addressed in the following section. Next, model results will be presented for a sub-region centred on Antwerp. Here we will focus on the impact of congestion in densely settled areas.

2. METHODOLOGY

2.1 A multiple activity-based cellular automaton

In the model of White et al. [2012], some land uses can be easily associated with an activity: e.g. residential land use is associated with population. Each land use can host several activities, but the associated activity is considered the primary activity and other activities present on the cell are secondary. Land uses not easily associated with an activity are considered to be an activity themselves (e.g. the activity of a forest cell is forest) and the value of this land-use "activity" is always binary (present or not). However these land uses will generally also host some secondary activities (a forest cell may also house people).

For each activity K, the proportion q_K to be located as primary activity (i.e. on the associated land use), is calibrated. Currently this parameter is calculated from the initial state, though changes over time are conceptually possible in the model. The rest of the activity proportion ($1 - q_K$) will be located on all other land uses as secondary activity. The proportion of secondary activity to be located on a specific land-use is known from calibrated compatibility factors between activities and land uses. Next, to calculate transition potentials for land uses and activities, the neighbourhood effect of all activities and land uses on all activities as well as several other factors are needed: the zoning status and suitability for all activities, an accessibility measure, a random factor and a diseconomies of agglomeration factor. This last component accounts for high land costs and congestion in densely settled areas. It reduces the transition potential if activity in a cell is higher than a critical value, composed of a calibrated parameter and average population.

The potential for activities can be calculated as:

$$V_{Ki} = r\, Z_{Ki}\, R_{Ki}\, S_{Ki}\, N_{Ki} \tag{1}$$

with V_{Ki} the activity potential for activity K on cell i, r the random perturbation, Z_{Ki} the zoning status for activity K on cell i, R_{Ki} the accessibility measure for activity K on cell i, S_{Ki} the suitability of cell i for activity K, and finally N_{Ki} the neighbourhood effect. Land-use transition potentials can then be calculated as:

$$VT_{Ki} = D_{Ki}\, (V_{Ki})^{mK} + I_{Ki} \tag{2}$$

with VT_{Ki} the land-use transition potential for activity K on cell i, D_{Ki} the diseconomies factor, m_K a parameter to be calibrated and I_{Ki} the inertia value for activity K on cell i.

Briefly, the transition process is as follows. For each cell, transition potentials are ranked, and then all cells are ranked on the basis of the highest value for each cell. Starting with the highest ranked cell, each cell is assigned the land use for which it has the highest transition potential, except that once the demand for a particular land use is satisfied, no further cells are assigned that use. Next, for each cell with a land use corresponding to an activity, primary activity is allocated in proportion to the cell's activity potential. Finally, secondary activities are allocated proportionally to activity potentials, as adjusted in accordance with the compatibility factors.

2.2 Application of the ACA model to a test area in Flanders

The first stage of this research consists of applying the current model of White et al. [2012] to Flanders at an improved resolution of 1 ha. We consider two activity-driven land uses (the residential urban fabric associated with population, and industrial and commercial areas, associated with employment in all economic sectors), two area-driven land uses (protected nature and agriculture), two passive land uses (unprotected nature and other, a rest category), and three static land uses (recreational areas, infrastructure and water). Forecasted population and employment trends for the Belgian arrondissements (NUTS3-level), obtained from the Federal Planning Agency, were used as future total activities. The proportion of primary activities and the compatibilities for secondary activities were calculated from an overlay between the initial activity map and the initial land-use map.

Several problems still need to be solved in future stages of this research: (1) there exists only one high resolution land-use map for the initial state (2010), (2) applying an ACA model at a 1 ha resolution for an area as large as Flanders (1,350,000 ha) is computationally heavy given the number of runs required, and (3) activity data are only available at a coarse spatial resolution.

Firstly, calibrating a land-use model requires high quality and high resolution land-use data for several discrete time steps [Straatman et al. 2004]. Unfortunately, there is only one land-use map for Flanders at the required resolution of 100 m (generalised from a 10 m map) available, for the year 2010. A similar map exists for the year 2006 at a resolution of 150 m (generalised from a 15 m map) featuring a mixed residential-commercial land-use class. In the future we intend to derive a map for 2006 at a 100 m resolution with separate residential and commercial classes. Currently, in the absence of a second land-use map, the model was not historically calibrated on the past, rather run forward from 2010 until 2040. For this paper, the general behaviour of the model was only visually examined.

Secondly, applying the model to the whole of Flanders is time consuming. Therefore we developed and tested a preliminary model on a region centred on Antwerp, in the central northern part of Flanders, consisting of the arrondissements Antwerp, Mechelen, Turnhout, Sint-Niklaas and Dendermonde (cf. figure 2). This test area was chosen because it is representative for Flanders and because extra data sets, including remote sensing products, are available for this region. The extra data sets could be useful for solving a third problem, namely the lack of activity data at the detailed spatial resolution of the model.

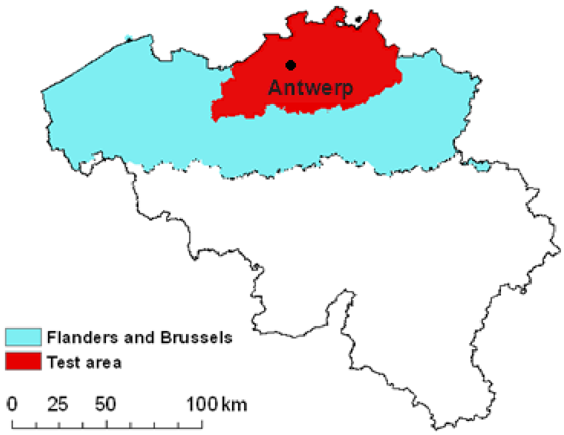

Figure 2. Location of the test area around Antwerp within the targeted modelling area (Flanders and Brussels) and Belgium.

Good input data of the activities themselves are indispensable for the appropriate functioning of the ACA model. Here the calibration problem is in essence not a time resolution problem – activity data are available for many time steps – rather a spatial resolution problem. Population and employment numbers for different economic sectors are needed for every cell, yet, these statistics can only be retrieved from the official sources on an aggregated scale (municipalities or statistical sectors). Dasymetrical mapping, a multiple regression-based spatial-allocation technique, can be applied on the aggregated data with a view to obtain the required cellular representation. The individual cell (activity) values are the dependent variable in such a model, while other, known, cell-based geographical data (e.g. the land use, distance to city centres) are used as independent variables [Langford et al. 1991; Wu et al. 2005].

White et al. [2012] propose a methodology that uses all components of the activity transition potential in the start year to estimate the location of the population. So far, in this research, we used a simple dasymetrical mapping model with only the input land-use map of the model as the independent variable. In the future, we want to raise its level of sophistication by using additional geographical variables, including remote sensing data.

3 FIRST RESULTS AND DISCUSSION

The initial land use and population density in 2010 are shown in Figures 3 and 4. For all parameters, several values were tried out with a view to verify and assess the behaviour of the model in the Flemish context. In the remainder of this paper, we will focus on the effects of the diseconomies of agglomeration factor and the impact of accessibility on the results.

A low diseconomies factor should logically result in more people living nearby the current largest concentrations of population and employment, while a high factor should push people away to remote areas because of an excessive congestion in the urban areas. In the results, this is more or less the case in suburban and remote areas, however in the city centre of Antwerp, population density in most cells is higher for high diseconomies (cf. figure 5).

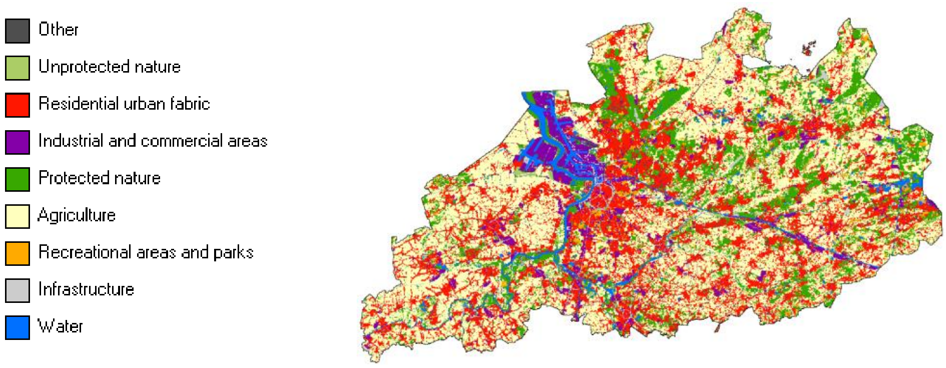

Figure 3. Actual land use around Antwerp in 2010.

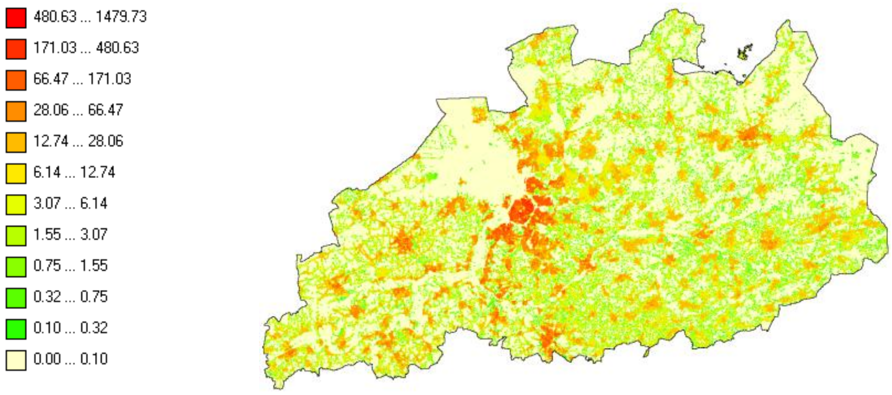

Figure 4. Actual population density (people per cell) around Antwerp in 2010.

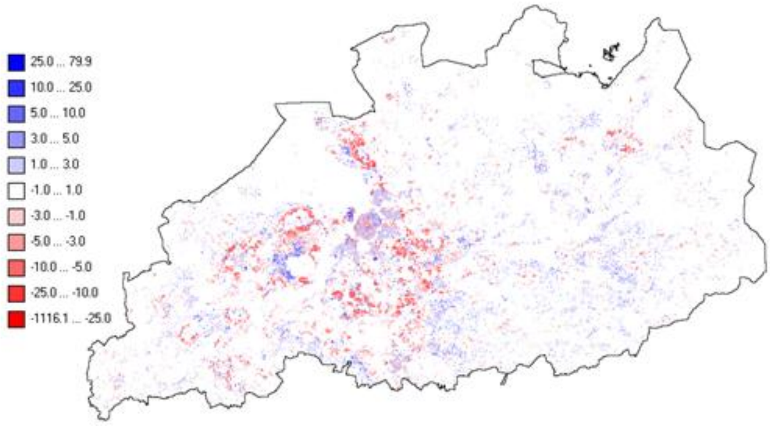

Figure 5. Simulated population density difference around Antwerp in 2040 between a high and low diseconomies of agglomeration scenario. Blue cells refer to a higher density for a high diseconomies scenario, red cells refer to a higher density for a low diseconomies scenario.

The effects are more noticeable in the resulting land-use maps for 2040. A possible explanation is the fact that in the current model of White et al. the diseconomies factor directly influences the land-use potential but not the activity potential. Next, the model also seems to find appropriate locations for the growth of protected nature, an area-driven land-use category which grows between 2010 and 2040 (cf. figure 6).

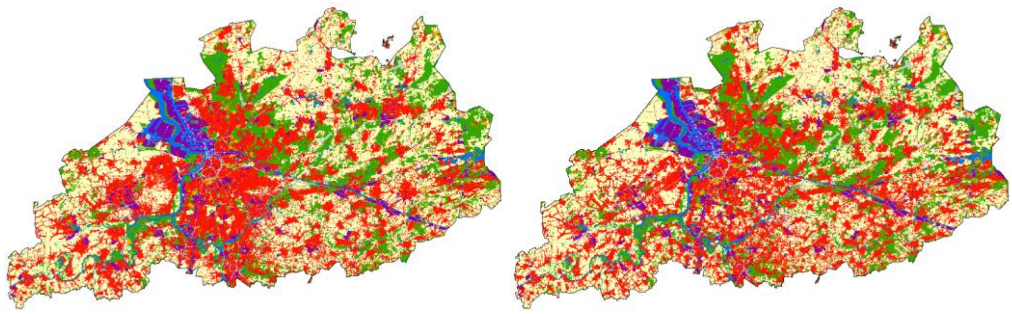

Figure 6. Simulated land use around Antwerp for 2040, with a low (left) or high (right) diseconomies of agglomeration effect. Legend as in figure 3.

It would be logical that clusters of new built-up area can grow further away from the centre of large cities in zones close to motorways or primary roads. This is not always clear from the results. Remote areas without major roads seem to grow too, especially in the high diseconomies scenario (cf. figure 7). The diseconomies factor seems to be too strong, and the classical accessibility factor too weak to have a significant impact. Measuring distances between variable grid super-cells using simple Euclidian distances, thus ignoring the transport network, might be another cause.

In the classic CA model, interregional distances and accessibility are dealt with at the regional level by means of a simple land use-transportation model. In the variable grid approach the interactions among the super-cells in the neighbourhood effect are to handle this. Therefore this research also entails a search for novel distance (or average travel time) calculation models to solve this problem.

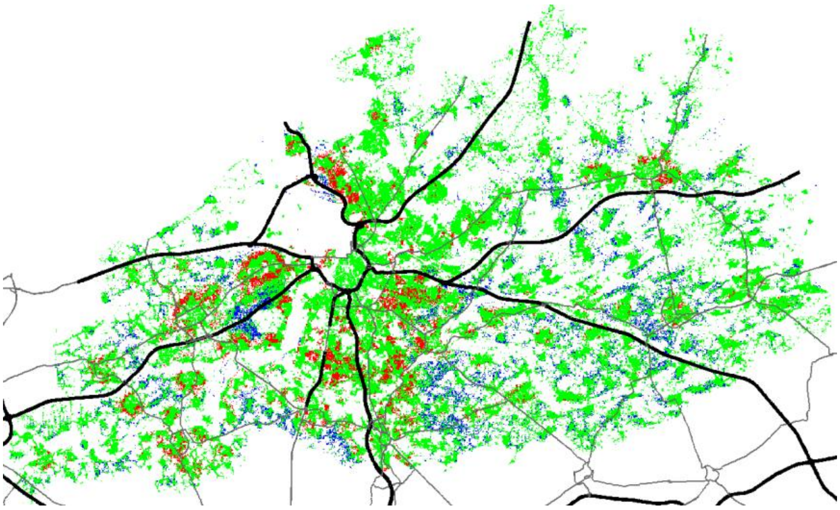

Figure 7. Simulated residential urban fabric around Antwerp in 2040 for low (in red) and high (in blue) diseconomies of agglomeration (in green for both). Overlay with motorways (thick black lines) and major roads (thin black lines).

Theoretically these distances should be measured over a transport network to simulate the behaviour of real world actors. It may even be better to make use of the relative time needed for a displacement between the cells. However, simple Euclidian distances are generally used in CA-based models. A classical CA neighbourhood only consists of a limited number of immediately neighbouring cells, hence, distances (or times) between the central cell and its neighbours are so small that the error made by using Euclidian distances becomes negligible.

The distance (or time) calculation problem gets more complicated in the variable grid environment [White 2005]. Distances are now needed between the central cell and the centre of each super-cell in its neighbourhood, consisting of the entire modelling area. Thus, representing large displacements by means of Euclidian distances may introduce substantial errors. On the other hand, the calculation of network distances between all points of the modelled area is computationally very intensive, and/or requires the storage of extremely large distance matrices.

The chosen method should enable the prediction of anisotropic urban growth patterns depending on accessibility between pairs of cells, instead of only the direct access of each cell to the road system, as captured by the accessibility factor in the model.

4 CONCLUSIONS

In this paper we discussed some critical issues related to the adaptation of the activity-based cellular automata (ACA) model proposed by White et al. [2012] to Flanders. Several socio-economic and ecological applications could benefit from a direct modelling of population and employment levels at a high spatial resolution (1 ha) for the Flanders region, yet several problems remain to be solved: more land-use data are needed for the calibration of the model; also a better method is required to disaggregate activity data to the model resolution.

First results show that the model can generate realistic behaviour and spatial patterns of both activity-based and non-activity based land-use categories, and that it can deal with the diseconomies of agglomeration effect. This is especially clear in land-use maps, less in activity maps. We therefore will have to examine whether the diseconomies computation should be included in the activity potential. Additionally, real network distances should be used in the calculation of the variable grid neighbourhood effect instead of Euclidian distances to capture the effects of accessibility and network quality in a more realistic way.

ACKNOWLEDGEMENTS

This work is supported by a PhD-scholarship financed by the Flemish Institute for Technological Research.

REFERENCES

Engelen, G., C. Lavalle, J. Barredo, M. van der Meulen, and R. White, The MOLAND modelling framework for urban and regional land use dynamics, in Koomen, E., J. Stillwell, A. Bakema, and H. Scholten (eds.), *Modelling Land-Use Change: Progress and Applications*, Springer, 297–320, 2007.

Langford, M., D.J. Maguire, and D.J. Unwin, The areal interpolation problem: estimating population using remote sensing in a GIS framework, in Masser, I., and M. Blakemore (eds.), *Handling Geographical Information: Methodology and Potential Applications*, Wiley, New York, 55–77, 1991.

Parker, D.C., S.M. Manson, M.A. Janssen, M.J. Hoffman, and P. Deadman, Multi-Agent Systems for the simulation of land-use and land-cover change: A review, *Annals of the Association of American Geographers*, *93*, 314–337, 2003.

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

Straatman, B., G. Engelen, and R. White, Towards an automatic calibration procedure for constrained cellular automata, *Computers, Environment and Urban Systems*, *28*, 149–170, 2004.

Van Steertegem, M., M. Bossuyt, J. Brouwers, C. De Geest, S. Maene, F. Maes, S. Opdebeeck, S. Overloop, B. Peeters, H. Van Hooste, L. Vancraeynest, and E. Vander Putten (eds.), *Milieuverkenning 2030, Milieurapport Vlaanderen*, *MIRA 2009,* Vlaamse Milieumaatschappij, Erembodem, 2009.

van Vliet, J., J. Hurkens, R. White, and H. van Delden, An activity based cellular automaton model to simulate land use dynamics, *Environment and Planning B., 29*, 431–450, 2011.

van Vliet, J., R. White, and S. Dragicevic, Modeling urban growth using a variable grid cellular automaton, *Computers, Environment and Urban Systems*, *33*, 35–43, 2009.

White, R, Modelling multi-scale processes in a cellular automata framework, in Portugali, J. (ed.), *Complex Artificial Environments*, Springer, 165–178, 2005.

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

White, R., I. Uljee, and G. Engelen, Integrated Modelling of Population, Employment, and Land Use Change with a Multiple Activity Based Variable Grid Cellular Automaton, *International Journal of Geographical Information Science*, 2012, DOI:10.1080/13658816.2011.635146.

Wu, S., X. Qiu, and L. Wang, Population estimation methods in GIS and remote sensing: a review, *GIScience and Remote Sensing, 42*, 80–96, 2005.