International Congress on Environmental Modelling and Software

4th International Congress on Environmental Modelling and Software - Barcelona, Catalonia, Spain - July 2008

Jul 1st, 12:00 AM

An activity based cellular automaton model to simulate land use changes

Jasper Van Vliet

Hedwig Van Delden

iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

An activity based cellular automaton model to simulate land use changes

Jasper van Vliet[a] **and Hedwig van Delden**[a]

[a] *Research Institute for Knowledge Systems bv, Witmakersstraat 10, 6211 JB Maastricht, The Netherlands (jvanvliet@riks.nl, hvdelden@riks.nl)*

Abstract: This paper proposes a spatial explicit model for land use dynamics, based on activities, which can represent for example population or jobs. In land use models based on constrained cellular automata (CA), the total area per land use is defined exogenously, while the model computes the allocation based on transition rules. The activity based model is a CA model, but it is constrained by activities instead of areas. Each time step, activities are distributed over cells based on transition rules. The CA transition rules comprise the effects of the activity in the neighbourhood, the land use of a cell, externalities and a stochastic perturbation term. Land use then is computed based on the activity distribution. Hence land use and activities are mutually dependent and each cell has two values: a land use state and an amount of activity. The activity based model is applied to simulate population dynamics and land use changes in Spain. Simulation results show that the model can produce realistic land use dynamics. Moreover, we argue that the inclusion of activities closer resembles the process of real world land use dynamics and offers good opportunities for integrated modelling.

Keywords: Activity based modelling; cellular automata; land use change; population dynamics.

1 A WALK IN THE REAL WORLD

Imagine you are in the centre of a large urban area and you take a walk. Your starting point might be the central business district and you are surrounded by high rise office buildings, but the further you walk the more residential buildings you'll find. Eventually you will pass some suburbs to end in the agricultural area outside the town. Parts of the city you passed through each have some distinct features; some are densely populated while others are not, some are mainly for commercial use, while others are primarily residential but certainly not entirely. Although land uses can be classified, most areas actually have a mixed land use. When you try to find the boundary between commercial zones and residential areas, between the city and its surrounding, you will find that in reality these are not always clear. In fact geographical classifications as well as their spatial boundaries are often fuzzy (Fisher, 2000).

These fuzzy boundaries and mixed land uses are difficult to represent in computer models that simulate the real world. For example land use classes are usually discrete and dynamics are often Boolean processes. In the model proposed in this study, we aim to overcome some of these problems. It allows mixed land uses by allocating population and jobs separately from, and on top of, the predominant land use in a location. Moreover it simulates dynamics in population and jobs in an incremental way. Hence, land use changes are not sudden events, but the result of a more gradual process.

In the next part first a short overview is given of some methods that our model builds on. Section 2 describes in depth the computational scheme that is applied. Section 3 shows the case study for which the model is applied and discusses some preliminary results to end with conclusions and directions for further research in section 4.

1.1 On agents, cells and population density

Over the last decades several methods are proposed and applied to simulate spatial explicit land use change. Overviews of different methods are among others available in Veldkamp and Lambin (2001) and Parker et al. (2003). Instead of giving another overview we would like to place the proposed model in the spectrum of modelling methods. For this we briefly consider three methods specifically: multi agent systems (MAS), cellular automata (CA) models and we will shortly point at some characteristics of economic land use models..
Both MAS and CA are dynamic approaches that can simulate land use changes in a way that approaches human decision making (White and Engelen, 2000). Moreover, both allow for heterogeneity among agents or land uses. These aspects are important since they differ from several other models on these points. In this discussion, agents in MAS are actors that can act and move independently over space. Hence CA are not considered MAS since there agents are the cells themselves, and their location is fixed. The advantage of MAS is that they can represent the behaviour of agents in a very straightforward way, since agents can interact directly with each other and with the environment. It is precisely these local interactions between agents and differences among them that generate the patterns observed on a global scale. However, since the agents are the basic unit of computation, MAS are computationally demanding. This is illustrated by an overview presented in Parker et al. (2003) where several case study applications are summarized. The applications that use individuals or households are all on a relative small scale using a limited number of agents.
Cellular automata, although sometimes considered agent based as well, differ from MAS in that sense that the basic unit for computation is a cell, not an agent. They have in common that both methods simulate changes from the bottom up, since the eventual global land use pattern is a result from interactions at the local scale. Together cells make up the lattice on which the CA exists, which makes them inherently spatial and therefore very suitable for the simulation of land use dynamics. Since cell sizes can be adjusted according to the scope of the simulation, models can keep a computational efficiency. Hence, CA can be applied to simulate land use changes on larger scales, from urban systems to regions or countries. This advantage comes at the cost of detail. Individual actors are not considered. Instead cells have a state, which generally represents the predominant land use, but the number of possible land use classes is limited. Moreover, cells have only one land use where in reality mixed land use is the rule rather than the exemption.
From the side of regional economics, land use models are usually considered in a non spatial way. Starting from Alonso (1964) these models describe population, utilities, jobs, or prizes as a function of the distance to the city centre. In this they differ from models like MAS and CA as described above (Irwin and Geoghegan, 2001). A quick look at any city tells that this simplification hardly holds, since urban areas are usually not symmetric but show local differences instead. This local variation can continue to grow and eventually results in polycentric developments (Wu, 1998). Moreover, the eventual results of these economic models are equilibrium situations and the way towards this equilibrium is usually not considered. Hence this approach neglects the incremental steps that play a role in the evolution of a city and the fact that most urban systems are still evolving.

In this paper we propose an activity based CA model to simulate land use dynamics. For this we introduce the notion of “activities” which is used to indicate the general concept of the model, in the case study described below it will be given a specific meaning. Activity is used here as a general term to denote people or jobs or any other feature that can be quantified and located. However, activities are considered a cell property, not entities by themselves. Hence a cell has a land use state, and a numerical value for each activity. Therefore, the cell remains the basic unit of computation, and it is possible to build on the CA computational framework. On the other hand, we do not consider his model a MAS for the same reason; activities can represent individuals, but we do not simulate decisions from individual agents.

This activity based approach will generate more spatial information than original CA land use models while retaining their computational simplicity. Also it allows to explicitly model interactions among people and jobs, like externalities, without losing the explicit spatial extent as is the case in models from the economic background. This activity based approach overcomes some limitations of conventional CA models as mentioned above. First, activities can exist in locations that do not have the associated land use, like inhabitants living in agricultural areas. Therefore additional spatial information is available from model results and mixed and multifunctional land uses are possible. Second, the focus of applications needs not only to be on the predominant land use, but rather on the activities itself which allows for a higher level of detail. Moreover, it offers more direct coupling between the socio-economic and the bio-physical subsystems since impacts that origin from activities can be quantified more precisely now. The latter is an asset for integrated models that aim to simulate interactions between both sub-systems more elaborately.

2 THE ACTIVITY BASED CA MODEL

The activity based cellular automaton as proposed in this paper builds on the Metronamica constrained CA models as developed by White and Engelen (1993; 2000). Their model exists on a lattice of regular squares, which represents a land use map. Each cell can have one of a limited number of cell states, which represent the predominant land use. Each cell has a neighbourhood and there is a set of transition rules to compute state transitions based on this neighbourhood. All cell states are updated simultaneously within each time step. The "constrained" in constrained cellular automata refers to the notion that the demand per cell state is determined external to the CA (White, Engelen and Uljee, 1997). In land use terms, the total number of cells of, for example, residential land use in a certain time step is determined exogenously while the CA allocates these cells on the map. Hence an increase in population over time needs to be converted to a number of cells for the associated residential land use per time step.

To introduce activities to this model, we made a few additions to the constrained CA framework. By introducing activities, a cell has no longer one discrete cell state. Instead it has a land use state, as well as a quantity for each type of activity. Activities can represent for example people, jobs, or maybe more abstract terms like economic value. Moreover, more than one type of activity can be used at the same time. For example, when both jobs and population are modelled, a cell could have agricultural land use, contain 12 inhabitants (activity "population") and have 4 people working in it (activity "jobs"). Instead of a demand for a number of cells per land use, the model is constrained by an amount of activity that needs to be allocated.

Each activity has one associated land use. For example, population and jobs are directly associated with residential and commercial land use respectively. For computation of activity dynamics and land use changes we assume that activities and land use are mutually influential. This is implemented as a two step process and each step is executed for all cells simultaneously. First activities are redistributed over the map based on the land use and activity distribution from the previous time step. When more than one activity is considered, this is done for all activities separately. Then, land uses are assigned based on the new distribution of activities. Now all cells have an activity, but only those cells whose activity exceeds a threshold value will get the associated land use.

However, not all land uses can be associated with activities, at least not in the sense that this activity can represent the eventual demand in terms of cells. For example the land use agriculture can be expressed by a number of jobs in agriculture, but that cannot directly be translated to an amount of hectares that need to be allocated on the map. Consider for example the difference between intensive and extensive agriculture. Therefore we need area constrained land uses as well. In fact the model uses four types of land use classes: activity constrained land uses, area constrained land uses, vacant states and features.

Activity constrained land uses are land uses that are directly associated with an activity as described above. Area constrained land uses do not have activities. Instead they have an area defined per time step that needs to be allocated. Vacant states are land uses that can only change as a result of other changes in the model. A typical example of a vacant land use is natural vegetation. Features finally are those land uses that are supposed to remain constant over time, like water bodies.

2.1 Distribution of activities

To allocate activity, the potential for that activity for each cell is computed according to the equation:

$$TP_{x,i} = c_{l(i),x} \cdot Z_{x,i} \cdot S_{x,i} \cdot A_{x,i} \cdot f\left(N_{x,i} + E_{x,i} + \varepsilon\right)$$

Where:

$TP_{x,i}$	is the total potential for activity *x* in cell *i,*
$c_{l(i),x}$	is the compatibility coefficient for land use *l* in cell *i* and function *x,*
$Z_{x,i}$	is the zoning for activity *x* in cell *i,*
$S_{x,i}$	is the suitability for activity *x* in cell *i,*
$A_{x,i}$	is the accessibility for activity *x* in cell *i,*
$f(p)$	is a transformation function to avoid negative potentials: $f(p) = \log_2\left(1 + 2^p\right)$,
$N_{x,i}$	is the neighbourhood effect for activity *x* in cell *i,*
$E_{x,i}$	is the externalities effect for activity *x* in cell *i*,
ε	is a stochastic variable drawn from a Cauchy (0, *α*) distribution, where *α* is an adjustable scale parameter. This variable is drawn independently for each function and each cell.

The compatibility coefficient is an adjustable parameter, which results from calibration. Suitability is a characteristic that is defined per cell a priori, as is zoning, but the zoning status can change over time. Both result from an overlay of GIS base layers. Accessibility is also a cell characteristic, which is computed from the distance from a cell to the nearest element of the infrastructure network as well as the importance of this element for the function at stake (exits of highways are often important for industry, while office building have a tendency to be located near train stations). The neighbourhood effect is a function of the current vector of activities and current land use in each cell of the neighbourhood of a cell. It is computed as the sum of the effects of all cells *j* at each distance *d* in the neighbourhood. This includes the current activities and land use of cell *i* itself:

$$N_{x,i} = \sum_{d} \sum_{j \in J_{d,i}} \sum_{f} w_{f,x}(d) \cdot X_{f,j}$$

Where:

$N_{x,i}$	is the neighbourhood effect for activity *x* in cell *i,*
$J_{d,i}$	is the set of cells at distance d from cell *i*

$w_{f,x}(d)$ is the weight function representing the attraction or repulsion from activity or land use *f* on activity *x* at a distance *d,*

$X_{f,j}$ is the level of activity for function *f* in cell *j*. For those land use functions that do not have an activity associated with them, the value is 1 when that land use is present and 0 otherwise.

Externalities represent the, mostly negative, effects of agglomeration. It is computed as a function of the amount of activity already present in an area:

$$E_{x,i} = \gamma_1 \cdot \left(N_{x,i} - \gamma_2\right)^{\gamma_3}$$

Here, γ_1, γ_2 and γ_3 are parameters that need to be calibrated. Typically -1 << γ_1<0 and $\gamma 3$ >1 to make sure that externalities are small initially and grow more than proportional with the present activity.

Once the total potential for each activity in each cell is known, all activity is (re-) distributed proportionally. When there is more than one activity type, allocation is independent for each type of activity. Hence activities only influence each other through the neighbourhood effect, which is based on the distribution in the previous time step:

$$X_{x,i} = \frac{TP_{x,i}}{\sum_j TP_{x,j}}$$

where

$X_{x,i}$ is the level of activity for activity type *x* in cell *i,*

$TP_{x,i}$ is the total potential for activity *x* in cell *i,*

$\sum_j TP_{x,j}$ is the sum of the potentials for activity *x* in all cells *j*, including *i*.

We assume that the basic unit of activity is 1. Therefore activities in all cells are rounded to whole numbers. This concurs with the assumption that people have to make a choice where to live and where to work. If two locations are equally attractive, one has to be selected in the end and based on this decision future development will take place.

2.2 Allocation of land uses

Land use is now assigned to cells per type of land use class. First features are allocated, then activity constrained land uses, then area constrained land uses, and finally vacant land uses. Since features cannot change over time, they are by definition allocated on the location they already hold. For the other land uses the order of assignment corresponds with the influence associated with them. In reality agents that claim residential or commercial land uses usually have more power than those for agricultural or natural land use. Activity based land uses represent these more powerful types of land uses, while the agriculture is represented by an area constrained class. Vacant land uses are those land uses that hardly have any economic power, like abandoned land or semi natural vegetation.

The basis for the land use assignment is the distribution of activities as well as a reference value for each activity. This reference value is taken as a percentile of the activities of all cells that have the associated land use in the initial land use map. Each time step, the activity per cell is compared relative to this reference value. A cell gets the land use function for which it has the highest relative activity as long as this is greater than one. For function land uses that are not constrained by an amount of activity but by a number of cells instead, the total potential for each cell is computed similar to the total potential for activity. However land uses are not assigned proportional to this total potential, but instead to cells with the highest potential until the externally defined demands are met. Cells that already have a feature or activity constrained land use are excluded from this assignment. Finally, all cells that do not have a land use from either of the constrained land use types become vacant. Since the number of cells that is occupied by activity constrained land uses is not known a priori it is a computational necessity for the model to have at least one land use type that can occupy vacant cells, to be able to assign a land use to all cells.

3 A CASE STUDY ON SPAIN

The activity based CA model as described above was applied to simulate land use dynamics for the country of Spain. In this case study the model was defined on a regular grid of 912 by 1076 cells of 1 km2 each, and land use change was computed for yearly time steps starting from 2000. Population, with residential areas as the associated land use function was modelled as an activity constrained class. Forest, commercial and industrial, agriculture and recreational land uses are area constrained classes. Natural vegetation is a vacant land use and water, wetlands, airports and mining areas are feature land uses. Although some of the latter land uses can actually change over time, their dynamics are much less. Besides it is assumed that their behaviour depends on other influences than local interactions. Still they do influence other land use changes considerably.

A limitation to apply the activity based CA model on a real world case is its data requirements. For these applications we used land use data from the Corine Land Use database for the year 2000. Population data was not measured directly but derived from the Corine land use database (Gallego and Peedel, 2001). Moreover, since we could not find spatial explicit data on jobs, commercial land was not modelled using activities.

Eventually, the activity based land use change model will not be used as a stand alone model, but as a part of an integrated assessment model that is developed within the DeSurvey project. This integrated model aims to explore future land use scenarios in the context of land degradation and desertification. Since population is modelled as an activity, land abandonment can be simulated with this model. Moreover, information on the amount of population allows for more direct coupling between the socio-economic system and the physical effects of land degradation and it enables exploration of a wider range of scenarios.

3.1 Preliminary results and discussion

The aim of this case study is to assess whether the proposed model realistically simulates activity dynamics and land use change. Hence model results were assessed on realistic patterns rather than cell-to-cell accuracy of land use changes. A first way to assess model results is by visual interpretation. Although it is preferable to measure results objectively rather than interpret them subjectively, visual interpretation can still gives an important clue about the behaviour of the model. Visual interpretation of the population distribution indicates that the model retains local differences in density. This is an important aspect since it is exactly that what distinguishes a spatial explicit model from economic models. This is partly due to local characteristics that are incorporated in the model, and partly due to the internal feedback mechanisms in the model.

For clusters of residential land use, urban areas, scaling laws are found to apply (Roehner, 1995). To assess the outcomes of the model we choose one such measure to assess whether a realistic distribution for residential land use was obtained: the cluster size frequency distribution (White, 2006). Although specific urban areas can change their position on the ranking, the distribution over larger areas remains mostly constant over time. The cluster size frequency distribution confirms the visual interpretation that the model can produce realistic dynamics. Both large urban areas are growing and new clusters appear. Figure 1 shows the results of the cluster size frequency distribution, where cluster sizes are measured in cells. Only the largest clusters seem to grow disproportionally. This is due to the input data. Cells around the major urban areas that are not residential initially already have a high population. Hence, in the first time step these cells change their land use to residential.

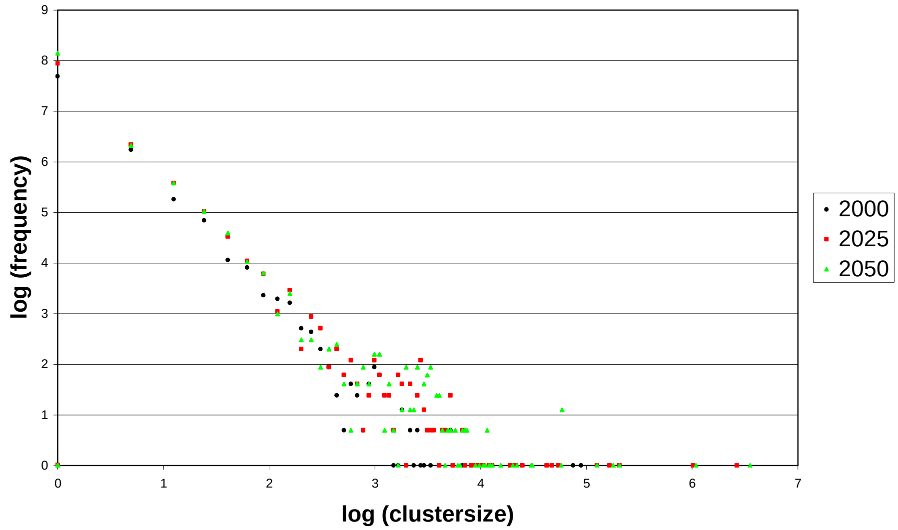

Figure 1: Cluster size frequency distribution for Spain. Blue points represent original data, while green and red are simulation results.

4 CONCLUSION AND DIRECTIONS FOR FURTHER RESEARCH

We believe that the important interactions among activities and from activities to land uses can be simulated using activities. In existing spatial explicit land use models the interface between a land use model and other models was mainly a land use map on the output side, and a multitude of information on the input side (see for example van Delden et al., 2007). However, many relations are not so much related or proportional to the predominant land use, but rather to the amount of activities on this location. The introduction of additional layers of information offers new opportunities for integrated modelling and more elaborate studies of land use dynamics. As an example Luck (2007) gives an overview of the relation between population density and its effect on biodiversity. He concludes that biodiversity and population density are significantly correlated and points at the need to focus on anthropogenic drivers of environmental changes. Still this case study only gives a first indication and the concept should be tested more extensively on other areas.

A major issue for the proposed model is data availability. Although synthetic applications can offer first insights in the mechanisms for land use change, real world data is required for meaningful studies. The model requires a land use map and a distribution for each activity. For proper calibration this data is even required for two points in time and independent validation would even require three such data sets. Due to developments in remote sensing, land use data is currently widely available. For population data this is certainly not the case, just as for jobs or other possible activities. Currently such data is at best available at more aggregate levels such as NUTS regions. The population data that was used in this research was derived from the land use map and is therefore biased. Alternatively, some remote sensing techniques are developed to estimate population from satellite imagery (Harvey, 2002), but this does not generate cellular values for population density yet. Finally there is the question how to assess the quality of the model results. Measures for land use maps as such are already topic of discussion (Hagen, 2003). Since the activity based model generates additional spatial results, more sophisticated measures are needed.

Acknowledgements

The activity based cellular automata model was developed within in the DeSurvey project, funded by the European Commission 6th Framework Program, Global Change & Ecosystems, Integrated Project Contract No. 003950.

References

Alonso, W. (1964) Location and land use. Harvard university press, Cambridge.

Anas, A., Arnott, R. & Small, K.A. (1998). *Urban spatial Structure.* Journal of Economic Literature 36. 1426 – 1464.

Delden, H. van, Luja, P., & Engelen, G. (2007) Integration of multi-scale dynamic spatial models of socio-economic and physical processes for river basin management. Environmental modeling & software 22, 223-238.

Fisher, P. (2000). *Sorites paradox and vague geographies.* Fuzzy sets and systems 113, 7 – 18.

Gallego J. & Peedell, S. (2001). *Using CORINE land cover to map population density.* In: European Environment Agency (2001). *Towards agri-environmental indicators; Integrating statistical and administrative data with land cover information.* EEA, Copenhagen, pp 92 - 104.

Hagen, A. (2003) *Fuzzy set approach to assessing similarity of categorical maps.* International journal of geographic information science 17 (3) 235 – 249.

Harvey, J.T. (2002) *Estimating census district populations from satellite imagery: some approaches and limitations.* International journal of remote sensing 23 (10), 2071 – 2095.

Irwin, E.G., & Geoghegan, J. (2001). *Theory, data, methods: developing spatial explicit economic models of land use change.* Agriculture Ecosystems and Environment 85, 7-23.

Luck, G.W. (2007). *A review of the relationship between human population density and biodiversity.* Biological review 82. 607 – 645.

Parker, D.C., Manson, S.M., Janssen, M.A., Hoffman, M.J. & Deadman, P. (2003) *Multi-agent systems for the simulation of land-use and land-cover change: A review.* Annals of the association of Americaan Geographers 83 (2). 314-337.

Roehner, B.M. (1995). *Evolution of urban systems in the Pareto plane.* Journal of regional science 35 (2), 277 – 300.

Veldkamp, A. & Lambin E.F. (2001) *Predicting land use change.* Agriculture, Ecosystems and Environment 85, 1 – 6.

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

White, R., Engelen, G. & Uljee, I. (1997). *The use of constrained cellular automata for high-resolution modeling of urban land use dynamics.* Environment and Planning B 24, 323 - 343.

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

White, R. (2006) *Pattern based map comparisons.* Journal of geographical systems 8, 145 – 164.

Wu, F. (1998). *An experiment on the generic polycentricity of urban growth in a cellular automatic city.* Environment and planning B 25, 731 – 752.