Computers, Environment and Urban Systems 33 (2009) 35–43

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

# Modeling urban growth using a variable grid cellular automaton

Jasper van Vliet [a,*], Roger White [a,b], Suzana Dragicevic [c]

[a] Research Institute for Knowledge Systems, P.O. Box 463, 6200 AL Maastricht, The Netherlands
[b] Department of Geography, Memorial University of Newfoundland, St. John's, Canada A1B 3X9
[c] Spatial Analyses and Modeling Laboratory, Department of Geography, Simon Fraser University, Burnaby, Canada V5A 1S6

ARTICLE INFO

*Article history:*
Received 26 July 2007
Received in revised form 10 June 2008
Accepted 20 June 2008

*Keywords:*
Cellular automata
Variable grid
Urban growth models
Spatial dynamics

ABSTRACT

Constrained cellular automata (CA) are frequently used for modeling land use change and urban growth. In these models land use dynamics are generated by a set of cell state transition rules that incorporate a neighborhood effect. Generally, neighborhoods are relatively small and therefore only a limited amount of spatial information is included. In this study a variable grid CA is implemented to allow incorporation of more spatial information in a computationally efficient way. This approach aggregates land uses at greater distances, in accordance with a hierarchical concept of space. More remote areas are aggregated into consecutively larger areas. Therefore the variable grid CA is capable of simulating regional as well as local dynamics at the same time. The variable grid CA is used here to model urban growth in the Greater Vancouver Regional District (GVRD) between 1996 and 2001. Calibration results are tested for goodness of fit at the cellular level by means of the kappa statistic and for land use patterns by means of cluster size analysis and radial analysis. Kappa results show that the model performs considerably better than a neutral allocation model. Cluster and radial analysis indicate that the model is capable of producing realistic urban growth patterns.



## 1. Introduction

Tobler's first law of geography states that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). Translated to land use this implies that the surroundings of a location are related to the land use at that location, but close surroundings have a stronger influence than more remote surroundings. The notion that land uses are spatially related and that nearby land uses have a stronger relation than land use at a greater distance was confirmed by empirical analysis of neighborhood characteristics (Verburg, Nijs, Ritsema van Eck, Visser, & Jong, 2004a; Verburg, Ritsema van Eck, Nijs, Dijst, & Schot, 2004b). This influence of neighboring land uses is strongly embedded in cellular automata (CA) based land use models by their neighborhood effect.

CA models are used in several ways to model land use changes (Clarke, Hoppen, & Gaydos, 1997; White, Engelen, & Uljee, 1997; Wu, 1998), where they are found to be particularly applicable to simulate urban dynamics (Barredo, Demichelli, Lavalle, Kasanko, & McCormick, 2004; White & Engelen, 1993). The latter is predominantly so for the ability of CA to create complex patterns (Wolfram, 1984) that are not unlike urban patterns (Batty, 2005; Batty & Xie, 1994). More recently, CA land use models have been applied as tools to support land use planning and policy analysis (Geertman & Stillwell, 2004) as well as to explore scenarios for future development (Barredo et al., 2003; Engelen, White, & Nijs, 2003; Nijs, Niet, & Crommentuijn, 2004).

A CA essentially comprises the following elements: (1) a cell space or lattice, (2) a finite set of cell states, (3) a definition of a cell's neighborhood, (4) a set of transition rules to compute a cell's state change and (5) time steps in which all cell states are simultaneously updated (White & Engelen, 2000). To make CA applicable for geographical modeling, the strictly defined CA rules are frequently loosened. These models are therefore referred to as relaxed cellular automata models (Couclelis, 1997). In constrained CA models, the total amount of area per land use is not a function of the transition rules, but determined exogenously instead, while the allocation of these land uses is computed by the CA (White et al., 1997). For example in an urban growth model the total area for residential land use can be derived from historic data or extrapolations thereof. This area demand is then imposed on the CA model that allocates a corresponding number of cells on the map, based on the transition rules.

### 1.1. On a cell's neighborhood

A cell's neighborhood is the region that serves as an input to calculate the neighborhood effect in the transition rules. This effect is a

* Corresponding author. Tel.: +31 43 3883343; fax: +31 43 3253155.
*E-mail addresses:* Jvanvliet@riks.nl (J.van Vliet), Roger@mun.nl (R. White), Suzanad@sfu.ca (S. Dragicevic).

0198-9715/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2008.06.006

function of a cell's own state and the state of the cells within its neighborhood. In land use terms, this represents attraction or repulsion of neighboring land uses. Hence, the size of the neighborhood determines the amount of land use information that is considered in the neighborhood effect. Originally, in CA only directly and diagonally adjacent cells were included. In human induced land use change however information at greater distances also influences land use changes, although the effect typically decreases with increasing distance. Hence larger neighborhood configurations are used to model land use change and urban growth (White & Engelen, 1993). In current applications this size ranges up to an 8-cell radius, enclosing 196 cells (Barredo et al., 2004; Engelen et al., 2003). Since larger neighborhoods include more land use information, they allow for better models. The number of cells in a neighborhood is directly related to the radius of the neighborhood. Therefore, increasing this radius would include more land use information. However, the required computation time would increase dramatically, as the number of cell-to-cell relations grow with the square of the radius. At the same time, this approach would use spatial information at larger distances at a higher level of detail then required.

Still, intuitively, more distant areas also influence land use change (Andersson, Lindgren, Rasmussen, & White, 2002a). This notion that information can travel over greater distances and thus have influence further away than just adjacent areas is well established in Hägerstrand's innovation diffusion (1967). To incorporate effects operating over larger distances, it has been necessary to combine two or more models that operate on different spatial levels. In these integrated models, a gravity based regional model calculates regional demands for land uses and a constrained CA model then allocates these demands on the map (White & Engelen, 2000). To overcome this problem, a more complete hierarchical conceptualization of space was introduced in Andersson, Rasmussen, and White (2002b). The assumption is that humans intuitively use a similar indexation to interpret and divide space: A city has several parts, each part consists of several blocks and every block again has a number of houses. The closer a feature is, the more in detail we think of it. Close surroundings, like neighboring houses, are of prime importance in spatial decisions. The more remote environment is considered with respect to its place in a spatial hierarchy: the next block is less important then immediate adjacent houses, but more important than the other side of town (Andersson et al., 2002a). In analogy to this hierarchical notion of space, cells at a greater distance can be aggregated to larger areas, while detailed information is kept for areas close by. This aggregation to area averages of land uses considerably reduces the number of spatial relations and thus the required computation time (White, 2005). Consequently, spatial information over much larger distances can be incorporated in the neighborhood effect and interregional effects need no longer be calculated in a separate model.

The variable grid CA is an implementation of this concept in a CA environment that allows incorporating all available land use information when calculating an individual cell's propensity to change. This is done by enlarging the neighborhood to include cells at all distances by using a hierarchical representation of space in the neighborhood definition. Specifically, this method uses a variable grid to aggregate more remote areas to mean field approximations (White, 2005). More distant cells are aggregated into increasingly bigger fields. This limits the number of spatial relations to be computed while nevertheless incorporating the maximum amount of land use information. Thus the model incorporates long distant relations as well as local effects. In this study the variable grid CA is applied to simulate urban growth in the Metro Vancouver area (former Greater Vancouver Regional District – GVRD). Both its applicability to simulate actual urban growth and its ability to simulate regional dynamics were tested with this application.

Moreover, the variable grid as presented in White (2005) introduces levels of activity for land uses. In the present application these are not incorporated and therefore activities are not considered in this text.

## 2. The variable grid cellular automata model

For this study the variable grid neighborhood is implemented in a constrained CA model. Hence the demand per land use class is defined exogenously; for every year the demand for constrained land use classes is defined in terms of a number of cells for the whole area (White et al., 1997). The allocation of these cells is determined by the potential of each cell for all land use classes as computed by the CA transition rules and using the variable grid neighborhood configuration. Land uses are assigned to cells with the highest potential, until the demand for this land use is met. Potentials for each cell and for each constrained land use class are calculated as follows (White & Engelen, 2000):

$$P_{il} = v * A_{il} * S_{il} * Z_{il} * N_{il},$$

where $P_{il}$ is the potential for cell $i$ and land use type $l$; $v$ is a stochastic perturbation term equal to $1 + (-\log(random))^{\alpha}$, where $\alpha$ is a scaleable parameter and *random* is a randomly drawn number from a uniform distribution between 0 and 1; $A_{il}$ is the accessibility of cell $i$ for land use $l$ to transport networks; $S_{il}$ is the suitability of cell $i$ for land use $l$; $Z_{il}$ is the zoning status of cell $i$ for land use $l$; and $N_{il}$ is the neighborhood effect for cell $i$ for land use $l$ as computed with the variable grid method as explained below. Calculation of variables other than the neighborhood effect is discussed more fully in White and Engelen (2000).

The variable grid CA was implemented using the Geonamica spatial modeling framework. This modeling framework (without the variable grid) has been applied successfully in land use change models, for example the Environment Explorer (Engelen et al., 2003) and the MOLAND project (Barredo, Lavalle, Demichelli, Kasanko, & McCormick, 2003b), and in integrated spatial models, for example MedAction (van Delden, Luja, & Engelen, 2007).

### 2.1. Definition of the cell neighborhood effect

The basic lattice with the highest resolution is referred to as the level 0 grid. At this level, every cell has only one state that represents its actual land use, formalized as

$$C_k^0(x) =\in \{0, 1\},$$

where $C_k^0(x)$ is 1 if land use $k$ is present at location $x$ and 0 otherwise. Now each successive level ($L$) then contains $(3^2)^L$ level 0 cells. Thus level 1 cells are an aggregation of $3^2 = 9$ level 0 cells and a level 2 cell of $(3^2)^2 = 81$. As a result, higher level cells are represented with cell counts of level 0 land uses instead of having one single state, and $C_k^L(x)$ is the cell count of land use class $k$ in a square of $3^{2L}$ cells centered at $x$. Each level 0 cell has eight adjacent cells, 4 rook adjacent and 4 bishop adjacent. Around this level 0 neighborhood there are eight level 1 aggregated cells, which are again surrounded by eight level 2 cells, etc. More generally every level L contains four rook adjacent cells $D_i^{\text{rook}}(L) = \{(i, i+3^L), (i+3^L, i), (i, i-3^L), (i-3^L, i)\}$ and four bishop adjacent cells $D_i^{\text{bishop}}(L) = \{(i+3^L, i+3^L), (i+3^L, i-3^L), (i-3^L, i+3^L), (i-3^L, i-3^L)\}$. This neighborhood template, as shown in Fig. 1, is relative to each individual cell and therefore moves cell by cell over the entire grid. Each aggregated cell holds cell counts for all land uses $l, k \in \{1, 2, \ldots, \ldots, m\} = K$, where $K$ is the set of all possible land uses states.

Influence of land use is represented by a weight which represents the attraction or repulsion from one land use to another as a function of the distance. Since rook adjacent cells are closer than