Simulating Urban Dynamics Using a Combination of Cellular Automata and Activity-Based Models

Rachel Katoshevske\textsuperscript{1}, Theo Arentze\textsuperscript{2} and Harry Timmermans\textsuperscript{2}

\textsuperscript{1}Technion and \textsuperscript{2}Eindhoven University of Technology

Key words: Cellular automata, Activity-Based Modelling, Urban dynamics

Abstract: Cellular automata models of urban change have been criticised for their lack of behavioural theory and representation in simulating urban change. It has led to a plea for developing multi-agent models. As a first step into this direction, this paper discusses the formulation of a combination of a cellular automata representation for an initial configuration of land use. This is linked to an activity-based model of travel demand, which generates varying demand for facilities across space. Agents, representing facility providers respond to this demand in locating and resizing their facilities. The interplay between these components then results in urban dynamics. The model is illustrated using a hypothetical example of urban forms.

1. INTRODUCTION

Cellular automata models have been the dominant modelling approach in simulating urban dynamics. These models divide the area of interest into a number of grid cells, with each cell depicting a particular type of land use. In addition, a set of transition rules is specified, expressing the probability of transitions between alternative types of land use. The principle of cellular automate models then is to apply these rules to the configuration of land use, resulting in trajectories of urban change and evolution. Models
differ in terms of the specification of these rules, the classification of land use types, the spatial operation of the system, allowance for stochastic components, calibration of the model, use of constraints and other operational decisions.

A problem underlying cellular automata models is their behavioural weakness. The transition rules are supposed to capture the decision processes but often the rules are not founded in behavioural theory, but rather summarize dynamics. Moreover, grid cells are not the decision making unit.

As a step forward to developing a multi-agent system of land use change, we developed an approach that uses some characteristics of cellular automata and an activity-based model of travel demand. The cellular automata system is used to simulate the configuration of land use. The activity-based model replaces the transition rules, offering a behaviourally more valid component to the simulation system.

In this paper, we will summarize this simulation approach and discuss an illustration.

2. THE MODEL

We assume that urban dynamics are caused by decisions of at least three groups of actors. First the planning authority makes urban decisions about the allocation of land use across locations. The zoning plan defines which land uses are allowed where and based on that which facilities are to be developed where. Secondly, firms (supply agents) decide on the actual development and location of facilities. Finally, individuals and households travel in the area and use these facilities. Dynamics occur due to changes in the decisions of each of these groups of actors and due to the interactions between these three groups. The model offers a first step into modelling these decisions and the interaction of these groups of actors.

The model related to the planning agent needs as an input the definition of the plan area, that is, number of grid cells, a definition of the transport network and the total amount of land that needs to be reserved for each land-use assuming some classification of land uses. The suitability of a cell is defined for each land use, much in line with common practice of cellular automata models. The suitability of a cell for a specific land use is assumed to depend on: (1) Suitability of the land/ soil for the land use, (2) Accessibility to main roads, the city center and specific other land use categories, and (3) Adjacency (land use in neighborhood cells), referring to any direct negative or positive effect one land use may have on another,
adjacent land use (for example- noise, traffic, visibility etc.). The latter involves the four direct neighboring cells and four diagonal ones.

The suitability of a cell for a specific land use as described above is measured in terms of the following equations:

\[
\begin{align*}
    z_{lg} &= \sum_i w_i^g \sum_j x_{ij}^g \chi_{ij}^g (l) + \sum_{k \in G} z_{gh} \chi_h (l) \\
    \chi_{ij}^g (l) &= \begin{cases} 
        1, & \text{if } c_{i,j-1} \leq d_i (l) < c_{i,j} \\
        0, & \text{otherwise}
    \end{cases}
\end{align*}
\]

(1)

(2)

where:
- \(z_{lg}\) is the suitability of cell \(l\) for land use \(g\)
- \(w_i^g\) is the weight of the \(i\)-th distance variable for \(g\)
- \(x_{ij}^g\) is a suitability score assigned to the \(j\)-th interval of the \(i\)-th distance variable for \(g\)
- \(z_{gh}\) is the suitability of presence of land use \(h\) adjacent to \(g\)
- \(\chi_h (l)\) equals 1, if land use \(h\) is adjacent to \(l\) and 0 otherwise
- \(d_i (l)\) is the value of \(l\) on the \(i\)-th distance variable
- \(c_{i,j}^g\) is the \(j\)-th cut-off-point on the \(i\)-th distance variable defined for land-use \(g\) (\(c_{i,0} = 0, c_{i,6} = \infty\))

The accessibility variables include distances from the cell to any other land-use, to a main road and to the city centre. Note that accessibility is not measured as some continuous function of distance, but rather in terms of a set of distance bands (\(d_i (l)\)), allowing more flexibility in terms of defining the effect of accessibility on suitability. For each distance variable \(i\), parameters \(c_{i,j}^g\) divide the distance scale into 6 distance intervals. In addition, one also has to set parameters \(x_{ij}^g\) to define a score for each distance interval, expressing the effect of each distance category on the suitability of a particular cell, and parameters \(w_i^g\) to allow these effects to differ between different types of land uses.

The parameters \(z_{gh}\) represent adjacency, and allow one to define a bonus (\(z_{gh} > 0\)), representing a synergistic relationship, or penalty (\(z_{gh} < 0\)), indicating an antagonistic relationship, for the suitability of \(l\) for \(g\) if land-use \(h\) exists in one or more of the eight neighborhood cells of \(l\). A value of zero indicates that the adjacent land use has no effect at all (is neutral). Note that \(z_{gh}\) is not necessarily equal to \(z_{hg}\). For example, industry may have a negative impact on housing, whereas industry may not necessarily be
affected by housing. By setting high values to the diagonal $z_{gg}$ spatial clustering of a specific land-use is amplified.

In the present application, the suitability scores related to accessibility ($x_{ij}^f$) are specified as whole numbers between 0 and 5 to the $j$-th interval of the $i$-th distance variable. These accessibility scores are varied by land use type according to weight vector ($w_{ij}^f$). On the other hand, the adjacency parameters ($z_{gh}$) are represented as an integer matrix of bonus or penalty scores. The numbers used fall in the range of -10 to 10.

These principles can be used to generate a configuration of land use, consistent with this input using some land-use allocation heuristic\textsuperscript{1}. Next, regarding the individuals and households, we assume that these agents will organize their daily activities within this spatial context, generating demand for different facilities in time and space. This is simulated by formulating an activity-based model of travel demand. To that effect, first a synthetic population is created. It involves simulating the number of people living in each housing cell and allocating workers within cell populations to industrial and commercial cells. Households are described in terms of different attributes used in the activity-based model, such as: work status composition of individual member, marital status, age of oldest person in the household, socio-economic class, and others. The synthesis uses attribute data of an existing sample of the simulated population. For each household the system draws with replacement a person at random from the sample and adopts the attributes of that person. In addition to the personal attributes the system also assigns a random day of the week.

Both parents and children are generated. The presence of children is an attribute of an adult female person indicating whether there are children in the household and the age group of the youngest child (the age classification corresponds to the age group related to the different types of schools). The system uses this attribute to establish the family situation. The decisions are made by randomly drawing from an appropriate distribution: the number of children, the exact age of the youngest child and, for each next child in order of age the differences in age to the younger brother or sister. The distributions from which the numbers are drawn are based on national statistics (Israeli Central Bureau of Statistics). The school activities are allocated on an activity basis rather than person basis, that is, the school location is determined each time a school activity is generated.

As indicated the activity-based model generates demand for facilities varying across space. It creates the market demand for facilities. If demand is

\textsuperscript{1} The heuristic used searches for an optimal allocation through a procedure consisting of two phases: an initial allocation based on a greedy search and an optimization based on a swapping algorithm.
lower than a minimum threshold, the particular facility ceases to exist, triggering urban dynamics. The distribution of facilities – the behaviour of supplier agents – is simulated in two separate stages, using two sub-models. First, a facility-location model is used to simulate the behavior of suppliers of facilities in terms of their decisions to open outlets. In a later stage, using the locational configuration of facilities as input, the activity-based model (a modified version of Albatross, developed by Arentze and Timmermans, 2000) is used to simulate the implementation of activity travel-patterns of the synthetic population. Based on visits attracted to facilities, the agents then make adaptation decisions, if any, in terms of closing or resizing facilities.

In the facility-location model, each supplier agent is specialized in a certain facility type. All agents evaluate facility locations based on the number of visitors a (new) facility would attract in a given time period (e.g. a day), which they estimate based on a catchment area analysis. In each time step in the system, agents have the opportunity to submit development proposals. A proposal specifies the location, square meter floor space and the type of the facility. Agents base a proposal on an assessment of the market potential of each feasible location in the study area using a catchment area analysis. Application of the facility location model requires setting the following parameters for each facility type: the penetration rate in cells, the radius of the primary and secondary catchment area, the maximally allowed rate of cannibalism incurred by a facility, a road bonus/penalty, the center bonus/ penalty, and minimum size of floor space required for the facility to be viable. This market analysis is aimed to create a perceived value of a site in order to choose the preferred allocation place, this can be described as:

\[ Q_{lh} = B_{lh} - w_h B_{lh} \]  

where \( Q_{lh} \) is the perceived value of the site for supply \( h \), \( B_{lh} \) is the estimate of the size of demand attracted by the new facility at \( l \), \( B_{lh} \) is the total size of demand distracted from the existing chain and \( w_h \) is a penalty factor for cannibalism.

The estimated demand attracted to the new facility, \( B_{lh} \), is determined as:

\[ B_{lh} = B_{h}^{i}(l) + B_{h}^{l}(l) + B_{h}^{s}(l) + B_{h}^{c}(l) \]  

where the terms on the right hand side of the equation represent demands from different sources including the catchment area - \( B_{h}^{i}(l) \), interception of traffic - \( B_{h}^{l}(l) \), visitors of other existing facilities at the same location - \( B_{h}^{s}(l) \), and \( B_{h}^{c}(l) \) visitors to the city centre.
Having determined the performance of each feasible location, a proposal \( l \) is generated only if:

\[
Q_{lh} = \max_{l'} Q_{l'h} \geq Q_h^{\text{min}}
\]

where \( Q_h^{\text{min}} \) is a parameter set by the user which represents the size of demand that is minimally needed for exploiting the smallest unit of facility \( (h) \). The planned size for a new facility is found as:

\[
V_{lh} = \sigma_h Q_{lh}
\]

where \( \sigma_h \) is the size of floor space needed to accommodate a consummating visit per day.

Calculated in this way, the proposed (and implemented) size \( V_{lh} \) is the best estimation of the demand that a new facility will attract. However, this estimated demand is based on limited information about the behavior of the individuals, for example, the penetration rates, and action radius. Other parameters of the method are only proxies of actual behavior that determines the generation of activities and allocation of activities across locations. The actual behavior is governed in this system by a different set of rules which is the activity-based model, using the locational configuration of facilities as an input. The consequences are that only after some time of exploiting the facility, the actual size of demand attracted will be known. Periodically, the suppliers consider re-sizing facilities and possibly even closing facilities based on actual size of the attracted demand using the same equations as in the planning stage whereby the estimated \( Q_{lh} \) is replaced by its realized counterpart.

Activity patterns of the adult population in the system, determine, which activities are conducted where, for how long, when, and, if travel is involved, the transport mode used. Therefore, an activity episode, which is defined as an uninterrupted period of engaging in a certain activity at the same location, can be described as:

\[
i = (a, t', v, h, l, t', m)
\]

where \( a \in A \) is the activity type, \( t' \) is the start time, \( v \) is the duration, \( l \) is the location, \( h \in H \) is the facility type, \( t' \) is the travel time (the duration of the trips, but interpreted here as the time the individual is willing to travel), and \( m \) is the transport mode of episode \( i \).
The combination of origin location, maximal travel time and transport mode determine the locations that are within reach. Before a choice set can be delineated, the facility type is determined by drawing $h$ from the probability distribution $\alpha_a^h$ which defines the stochastic matching relations between activity types $a$ to facility types $h$. If the choice set is empty, then the nearest facility of $h$ is taken as the choice. Otherwise, the system considers two alternative heuristics as optional for determining a choice. The first heuristic simply describes a random choice from the choice set. The second heuristic selects the highest-order location and solves ties based on distance.

Based on visits attracted to facilities, the agents make adaptation decisions, if any, in terms of closing or re-sizing facilities.

3. ILLUSTRATION

As explained, the model requires as input the size of the planned area and cell size of the grid used. In addition, it needs a classification of land use categories and demand defined as the total number of cells for each land use. The following seven land use categories were distinguished: (i) Housing High density (to be denoted further as Housing $H$); (ii) Housing Low density (to be denoted further as Housing $L$); (iii) Industry High Tech (to be denoted further as industry $H$); (iv) Industry Low tech (to be denoted further as Industry $L$); (v) Commercial; (vi) Green Recreation, and (vii) Nature. The current plan area consists of an array of 2404 cells of 125 x 125 m divided as follows: 760 cells for Housing $H$, 400 cells for Housing $L$, 96 cells for Industry $H$, 96 cells for Industry $L$, 96 cells for Commercial land use, 80 cells for (Green) Recreation and 972 cells for Nature. These proportional land use requirements are derived from an anticipated population size of 150,000 people and planning standards. The size of a cell was determined such that it is small enough to accurately represent facilities, but not too small to avoid excessive computation times. The number of housing cells is derived from a predetermined density. The number of cells for industrial land use is also based on a density standard (number of workers per industry cell), while the total amount is based on the total population (the total number of workers in the residential population). The number of commercial and recreational cells was based on planning standards as a function of population size.

In the system the allocation of land use is done based on a set of suitability parameters. Because we have seven land uses, plus the main roads
and the city center, we defined \( c_{ij}^g \) as a 9 x 6 matrix of integers. The six cut-off points were defined between 100 and 12500 m, (where 12500 m is the maximal distance in the grid). The suitability score matrix, depicting the effect of accessibility \( x_{ij}^g \) is also a 9 x 6 matrix of integers for each land use \( g \), which assigns a score between 0 and 5 to the \( j \)-th interval of the \( i \)-th distance variable. These accessibility scores are varied by land use type according to weight vector \( w_i^g \). Weights were also varied between 0 and 5. Hence, these parameters setting allowed us to vary the impact of accessibility on the suitability of a cell across land use types between 0 and 25. The adjacency parameters \( z_{gh}^z \) are represented by a 7 x 7 integers matrix of bonus or penalty scores. The numbers used fall in the range of -10 to 10. These scores are based on the results of a conjoint analysis (Katoshevski and Timmermans, 2001), and expert knowledge.

Having generated a zoning plan, the next step concerns simulating the number of people living in each Ho using cell and allocating workers to Industrial and Commercial cells. In this illustration a 10\% fraction of the population is synthesized. Given the fraction used, the total number of households per cell equals 92 for high density Housing cells and 39 for low density Housing cells. These numbers follow from the assumptions that on average a house occupies 210 m\(^2\) and 500 m\(^2\) in respectively high density and low density cells, and households on average have 1.24 adult members. The number of workers (in fte) follows the ratio 2 (high tech industrial): 1 (low tech industrial): 3 (commercial).

Having synthesized the population, the next stage in the model deals with implementing facilities in the city, using the land use map as a basis for this stage. It is done in two stages which includes two different models: first, the facility location model (location decisions of supplier agents), and then the facility use model (activity-travel behaviour of individuals and households and responsive behaviour of supplier agents). The application of the facility location model requires setting the following parameters for each facility type:

1) **Penetration rate**: the percentage of the population present in a cell that will be attracted by the facility, determined for each facility type, differentiated by a primary and secondary catchment area. In the model, the parameters differ between the primary catchment area and the secondary one, where for the secondary level the rates are lower, reflecting typical distance decay effect.

2) **Radiuses of the catchment areas**: the area from which the facility would attract visitors. Those were set such that for the neighborhood facilities a
relatively small catchment area results, compared to higher-order, city-level facilities. This is meant to keep these facilities at short distance from Housing and to spread them across the city.

3) *Maximally allowed rate of cannibalism*: this aspect determines the extent to which overlap in the primary catchment areas between facilities of the same type is allowed. Also for this aspect values were determined for each facility type.

4) *Center bonus / penalty*: in addition to the catchment radius it is another feature that helps placing neighborhood facilities in the neighborhood area, on the one hand, and, on the other hand, setting city level facilities in the central area. Thus, in order to push neighborhood facilities to the neighborhood, negative scores were given to “push out” these facilities from the centre. In addition, facilities that are expected to be located in the central part of the city were given positive bonus values in order to “pull” these elements into the centre.

5) *Road bonus / penalty*: represent the extra demand attracted if a facility is located in the proximity of the main roads. These can help in “pulling” or “pushing” facilities close to or far from main roads. Values were specified for three distance bands: 125 m, 250 m, and 375 m.

6) *Space needed per 100 visitors (on a daily basis) and the minimum size of floor space required for the facility to be viable*: represent space needs and viability conditions for each facility type. Some facilities like a post office or a bank can operate in a small place, while others, such as a hospital, require large areas. In general, for facilities that have two or more levels (neighborhood and city level for example) a clear difference in minimum floor space was fixed. The floor space required per 100 daily visitors indicates the space needed to provide the service as a function of the
number of visitors. The range of this parameter across facility types is also large.

The model was applied to three scenarios, graphically displayed in Figure 1, and labeled Basic, Corridor and Connected form. Each scenario resulted in different urban dynamics and land use configurations. A set of performance indicators was calculated for each resulting configuration allowing a comparison of the impact of basic urban forms on the performance and dynamics of the urban system.

The land use pattern that emerges for the Basic City is portrayed in Figure 2. It shows the creation of a city center with primarily commercial land use and the development of housing along the arterial road, with high density development closer to the center. Industrial areas appear in the periphery of the area. The land use pattern that emerges for the Corridor City is displayed in Figure 2. It is characterized by a large central part, which is dense and includes housing, a shopping area, a park, and different facilities, and an outer part situated at some distance from the center where development is less dense. To some extent it develops similarly to the Basic City, in the sense that housing development takes place along the main roads. However, since the number of roads is doubled, the developed area concentrated in the center of the city is larger. Penetration of nature cells also takes place, but only on a limited scale, and the “fingers” of development into the nature areas are found only at some distance from the center. Compared with the Basic City, the industrial cells, which are developed in three different places – all in the outer part of the city – are
closer to the housing cells. However, Housing-L cells are never adjacent to industry cells and in most cases are adjacent to the green areas.

Figure 3 portrays the emerging land use pattern for the Connected City. This layout is based on circular as well as radial roads, creating different development in which the city evolves as a single dense area. Some Housing-L cells are developed along the roads, emphasizing the road structure. There is no penetration of nature areas into built-up inner areas, and only limited development has occurred in the outer areas. Industrial areas, which are situated in two separate locations, develop at some distance away from housing.

A comparison of these emerging patterns suggests some similarities and differences. Common to all urban forms is that commercial cells are clustered in the center of the city, creating a city business center. This is as expected because the accessibility functions and adjacency scores favored the city center for this type of land use. Also common is the emergence of clustered recreational cells in one area adjacent to the central commercial area, creating a city level park. In addition, industrial cells are clustered in a few industrial areas, at the edges of the city, and mostly away from housing and adjacent to nature areas. Across forms, nature cells are located mostly on the outskirts of the city, while low density housing is located mostly towards the edges of the city, although the latter finding is less prominent for the Basic City. Differences in the road network are mainly responsible for differences in the land use patterns. The pattern for the Basic City has a dense part in the center comprised of a commercial area in the center with an adjacent recreational area. The city is evolving outwards from the center, creating strips of development along the main roads, allowing penetration of nature cells into the city. The central area is surrounded by Housing-H cells.

However, there is also clear-cut development of Housing-L cells in the outlying parts of the city, where development is surrounded by nature cells, producing a less dense area. Industry is located in two areas, neither of which is adjacent to housing cells. Because of more roads in different directions, development along the radial roads is less articulated in the Corridor City scenario, creating a more extensive core area and this is even more observable for the Connected City. As a result, the other spatial distribution of the land uses also differs.

These different land use patterns provide the supply side that influences where suppliers will locate their facilities (based on catchment area analysis) and where individuals and households can conduct their activities. The implementation of activity-travel patterns changes the demand in each
cell and this may lead suppliers to close-down or resize facilities and thus change the location configuration of facilities.

In the Basic City facilities are spread all over the city. More specifically, neighborhood daily shopping units are dispersed throughout the city, whereas city level shops tend to gravitate towards the central district. The non-daily shopping facilities demonstrate a less efficient distribution, as most shops are developed in the inner part of the city, thereby limiting the number of shops in the outer fingers of development. Schools of all levels, sports facilities and parks are scattered throughout the city, implying convenient access to facilities. As expected, leisure and services facilities are developed in all city areas but are densest in the center.

In the Corridor City facilities also show a generally good dispersion. However, the shopping facilities and schools are not evenly distributed and tend to be concentrated in the dense part of the city, hampering their development in the outward-pointing “fingers”. Medical, leisure, sports and park facilities, on the other hand, are well dispersed across the city areas, suggesting a very efficient spatial distribution.
In the Connected City the spatial distribution of facilities is more spread out. The educational and medical facilities, leisure, services, sport facilities and parks are placed all over the city. Leisure and services facilities are denser in the center, as expected. The shopping facilities (daily and non-daily) also show a wide spreading, but however do not reach the edges of the city.

4. DISCUSSION AND CONCLUSIONS

The purpose of this paper was to suggest a combination of a cellular automata model and an activity-based model to simulate urban change and illustrate the model system in a hypothetical scenario study. The model can be seen as a first step towards a more comprehensive multi-agent simulation system. The study shows that principles of the suggested approach seems valuable and feasible.

It is also evident that the supply agents that were incorporated were relatively easy to model for a variety of reasons. First, it can be assumed that their location patterns are either strongly influenced by planners’ decisions or are strongly related to consumer demand. Secondly, the location theories
for these facilities have been well documented and are relatively well developed. Finally, it is relatively easy to collect data for these agents. All of this is may as well be true for agents representing land use such as farm land, offices, and industry, which would also need to be included in future versions of the system.

5. REFERENCES


