

**CENTRE FOR ADVANCED  
SPATIAL ANALYSIS  
Working Paper Series**



**Paper 68**

---

**INTEGRATED  
URBAN  
EVOLUTIONARY  
MODELING**

---

**Yichun Xie  
Michael Batty**



Centre for Advanced Spatial Analysis  
University College London  
1-19 Torrington Place  
London WC1E 7HB

[t] +44 (0) 20 7679 1782  
[f] +44 (0) 20 7813 2843  
[e] [casa@ucl.ac.uk](mailto:casa@ucl.ac.uk)  
[w] [www.casa.ucl.ac.uk](http://www.casa.ucl.ac.uk)

[http://www.casa.ucl.ac.uk/working\\_papers/paper68.pdf](http://www.casa.ucl.ac.uk/working_papers/paper68.pdf)

Date: October 2003

ISSN: 1467-1298

© Copyright CASA, UCL

## **INTEGRATED URBAN EVOLUTIONARY MODELING**

---

Yichun Xie<sup>†</sup> and Michael Batty<sup>‡</sup>

<sup>†</sup>Institute for Geospatial Research and Education, Eastern Michigan University  
125 King Hall, Ypsilanti, Michigan 48197, USA

<sup>‡</sup>Centre for Advanced Spatial Analysis, University College London,  
1-19 Torrington Place, London WC1E 6BT, UK

# INTEGRATED URBAN EVOLUTIONARY MODELING<sup>†</sup>

Yichun Xie<sup>1</sup> and Michael Batty<sup>2</sup>  
[yxie@emich.edu](mailto:yxie@emich.edu) [m.batty@ucl.ac.uk](mailto:m.batty@ucl.ac.uk)

<sup>1</sup>Institute for Geospatial Research and Education  
Eastern Michigan University  
125 King Hall  
Ypsilanti, Michigan 48197, USA  
<http://ceita.emich.edu/>

<sup>2</sup>Centre for Advanced Spatial Analysis  
University College London  
1-19 Torrington Place  
London WC1E 6BT, UK  
<http://www.casa.ucl.ac.uk/>

## ABSTRACT

Cellular automata models have proved rather popular as frameworks for simulating the physical growth of cities. Yet their brief history has been marked by a lack of application to real policy contexts, notwithstanding their obvious relevance to topical problems such as urban sprawl. Traditional urban models which emphasize transportation and demography continue to prevail despite their limitations in simulating realistic urban dynamics. To make progress, it is necessary to link CA models to these more traditional forms, focusing on the explicit simulation of the socio-economic attributes of land use activities as well as spatial interaction. There are several ways of tackling this but all are based on integration using various forms of strong and loose coupling which enable generically different models to be connected. Such integration covers many different features of urban simulation from data and software integration to internet operation, from interposing demand with the supply of urban land to enabling growth, location, and distributive mechanisms within such models to be reconciled. Here we will focus on developing better housing market and site subdivision processes within CA models, taking as our starting point the Dynamic Urban Evolutionary Model (*DUEM*) first proposed by Xie (1994) and operationalized through a graphical user interface by Batty, Xie and Sun (1999). We set our new model within a wider model-based infrastructure, devising a version which integrates the cellular approach to various residential models of traditional form. We call the resulting system *IDUEM*. This model system retains the cellular approach which is highly visual in terms of the way urban growth and change is conceived but uses this as the interface to different varieties of model, making the framework much more applicable to real policy problems.

---

<sup>†</sup> First presented at *Geocomputation 2003*, held at the University of Southampton, 8-10 September 2003; see <http://www.geog.soton.ac.uk/conferences/geocomp/>

# 1. Introduction

Urban areas have long been recognized as displaying nonlinear, dynamic properties with respect to their growth (Crosby, 1983). Capturing their dynamics, however, is one of the most delicate problems in urban modeling. Only very recently have the conceptual and mathematical foundations for substantive inquiry into urban dynamics been made possible due to our growing understanding of open systems and the way human decision processes feed back into one another to generate the kinds of nonlinearity that characterize urban growth and change. Applications have been made possible by fundamental advances in the theory of nonlinear systems, much of it inspired by theories of dissipative structures, synergetics, chaos and bifurcation in the physical sciences. In fact many of the originators of these new approaches have seen cities as being a natural and relevant focus for their work. Prigogine's work on dissipative structures, for example, has been applied to urban and regional systems by Allen (1997) while Haken's work on self-organization has been implemented for city systems by Portugali (2000) and Weidlich (2000). Many of these applications have built around traditional aggregate static approaches to urban modeling pioneered in the 1950s and 1960s, and were motivated as part of the effort to make these models temporally dynamic and consistent with new ideas in nonlinear dynamics (Wilson, 2000).

The development of complexity theory has proceeded in parallel where the concern has been less on spatial simulation *per se* but more on the way complex systems are composed of many individual and agents whose behavior drives change at the most local level. Ideas about how life can be created artificially have guided many of these developments and in this context, highly disaggregate dynamic models based on cellular change - cellular automata (CA) - have become popular as a metaphor for the complex system. CA models articulate a concern that systems are driven from the bottom up where local rules generate global pattern, and provide good icons for the ways systems develop in which there is no hidden hand in the form of top-down control. Again cities are excellent exemplars (Holland 1975). Despite the hype, CA has recently been proposed as a 'new science', articulated as the basis for taking a fresh look at a number of different fields of scientific inquiry (Wolfram, 2002).

In fact, the embedding of nonlinear dynamics into traditional cross-sectional static urban models has not led to a new generation of operational land use transportation models for policy analysis, despite the fact that this kind of dynamics is consistent with the way those models are formulated. What has happened is that cellular automata models have found much more favor but these models have all but abandoned the focus on socio-economic activity in favor of simulating physical change at the level of land use and development. Consequently most CA models do not treat the transportation sector in any detail whatsoever and thus their use in policy analysis is limited. This lack of an explicit transportation dimension is largely due to the way such CA models are structured. CA focuses on physical processes of urban systems and simulates land use changes through rules usually acting upon immediate neighboring cells or at best some wider set of cells which still restrict the neighborhood of spatial influence (Batty, 1998; Batty, Xie and Sun, 1999; Bell, Dean and Blake, 2000; Clarke and Gaydos, 1998; Li and Yeh, 2000; White and Engelen, 1993; Wu and Webster, 1998; Wu, 2002; Xie, 1996). Insofar as transportation enters these models, it is through notions of local diffusion which do not map well onto actual physical movements of short duration such as those characterizing trip-making in cities.

Though many innovative ideas such as genetic algorithms, neural network methods, and stochastic calibration for determining weights and parameters have been proposed and successfully developed, such CA models are still essentially heuristic and simplistic. The origins of CA modeling in urban systems also dictate some of their limitations. Raster-based digital data particularly from remote sensing and GIS software that readily works with such data, has given added weight to models that are composed of cells. The notion too that CA might be used to simulate the evolution of different physical landscapes has influenced their form and structure while the fact that many of the groups developing such models have been remote from policy, has not focused the effort on real planning applications.

Currently several profound challenges to CA-based urban simulation models exist. Firstly, both physical and socioeconomic processes interact with each other and their surroundings in complex, nonlinear and often surprising ways. These processes have subsystem elements that, in turn, may be complex and operate in different ways but in precisely the same geographical space. Different urban elements working in different ways contribute to the emergent properties of the entire

system (Wilson, 2000). Each component of a complex urban system may itself be complex. However, current CA models of cities are, to a large degree, limited to physical processes and land development. They ignore urban activities that comprise such spaces which are usually the focus of policy analysis. Secondly, cells defined as the basic unit of land development are often characterized by a binary state of developed or undeveloped land, or by a land use type within a cell which is usually restricted to only one such use per cell. The basic land unit does not usually carry attributions such as the number of people or households that reside on it, the behavior of its residents, the value of its property or rent, the amenity of the surrounding neighborhood, or the quality of its environment. The exclusion of such socioeconomic features is a serious limitation to realistic applications which adopt CA models for urban planning and related forms of decision-making.

Thirdly, it is difficult to establish compatibility between a cell and a real urban entity. Despite increasingly higher resolution with finer cell sizes adopted in CA models as increases in computer power has enabled larger and larger systems to be represented, pixel-based cellular dynamics seldom matches area-based socioeconomic phenomena. Scale and longitudinal change in the socio-economic geography of an area further complicate the calibration and validation of CA models. Fourthly, CA models are usually supply-driven with demand entirely function of supply, there being no feedback between demand and supply to reflect any market clearing. This is an important omission as it suggests that CA models do not react to the economic mechanisms that determine how land is actually developed, once again reflecting the disjunction between socio-economic models which form part of the urban economic, regional science, and transportation traditions, and this newer tradition of geographical modeling.

To put these criticisms in perspective, our starting point will be *DUEM*, the Dynamic Urban Evolutionary Model which we have developed in several places in South Eastern Michigan and which throws into stark relief the limitations of the cellular approach. This model like most CA models of urban development simulates the growth (and decline) of different land uses in cells representing the supply side which is determined by rules governing physical development. These models do not handle geo-demographics, site location, transportation and so on at all well. Having set the scene with a description of *DUEM*, we will then outline a new model framework *IDUEM* (Integrated Dynamic Urban Evolutionary Modeling) which begins

to resolve these problems. We will then examine its conceptual structure, improvements to the way demand and supply for land and housing are handled, links to forecasting models, and issues of detailed land parcel and site representation. Our paper is a first sketch of an ambitious framework for extending and making much more realistic and policy-relevant the CA approach. As such, this is a work in progress.

## 2. Starting Points: *DUEM* as a Cellular Model of Land Use Development

### 2.1 *An Outline of the Model*

*DUEM* is somewhat different from the standard cellular model of urban development in that it deals with a comprehensive series of land uses, one of which is the infrastructure associated with transportation. In fact like most other CA models, what goes on in each cell is physical development and there are no measurable attributes of cells such as population levels, rent, density, etc. *DUEM* is also unique in that it provides a strong life cycle focus to land use development reminiscent of Forrester's (1969) model of *Urban Dynamics* and this emphasis makes it highly suited for simulating long term evolution at the level of the city system. Finally the graphical user interface is well developed, putting it into a class in which the model is generic, hence applicable to a wide variety of situations by non-expert users. Nevertheless such generalization is difficult; most users find that standard software has to be refined for particular circumstances. Part of our current effort to develop more applicable model structures is to address the limitations of the generalized user interface that we are currently working with.

In essence, each land use is classified as belonging to one of three life phases - initiating, mature and declining - which reflect the life cycle of aging with the assumption that as a land use ages, it becomes increasingly less able to act as a generator of new land uses. In fact we assume that only initiating land uses spawn new uses while mature land uses simply exist *in situ* with declining uses moving to extinction where they disappear, the land they have previously occupied becoming vacant. In *DUEM*, there is an explicit life cycle which ages these land uses through different stages (but with the possibility that a mature or declining use can revert to an earlier category as indeed sometime occurs). Land uses can also make transitions in that a land use can change its type at any stage although for the most

part, this possibility is more likely the older the use, and thus once again relates to its life cycle.

Initiating land uses drive the growth of new land uses in the model. They spawn new land uses in their neighborhood which is a restricted field of cells usually symmetrically arrayed around the origin cell but sometimes with directional distortion. The probability of an initiating use spawning another use in this field is a function of the distance away from the central cell and what cells get developed will depend ultimately on how strong the competition is between different land uses being spawned by the particular land use in question. A land use has the potential to spawn any number of different uses but only one these will occur in each time period. The spawning process is subject to a series of constraints, some within a narrower traditional CA neighborhood around the cell in question which relates to density and type of uses in the neighboring cells, and also subject to regional constraints which limit what each cell might be used for.

CA models are difficult to present in a closed form that makes their operation transparent. This is because transitions from one state (land use) to another in any cell are determined by various rules which although uniformly applied across all cells, cannot usually be written in continuous algebraic form. Thresholding and counting for example are typical operations that make such rules work. Hence this makes the analysis of the dynamics of such models only possible through simulation. The mathematical structure of *DUEM* has been spelt out in some detail elsewhere (Batty, Xie and Sun, 1999) but we do need some formality in presentation if we are to make clear the limitations of this model and demonstrate how these might be resolved. We will define cells using subscripts  $i$  and  $j$ , land use states as  $k$  and  $l$ , life cycle aging as  $\tau$  where the range of  $\tau$  is subdivided into three classes - initiating, mature and declining, and time itself as  $t$  and  $t+1$ . A land use of type  $k$  in cell  $i$  with age  $\tau$  at time  $t$ ,  $S_i^k(\tau, t)$ , and this defines the transition as  $S_i^k(\tau, t) \rightarrow S_i^l(\tau+1, t+1)$  where aging and state change are clearly marked through the passage of time from  $t$  to  $t+1$ . However these transitions are not mainly defined by intrinsic changes within the cells but by changes that are taking places in the rest of the system, particularly in the local neighborhood around the cell in question but also in their wider region.



Change itself is in fact generated by two processes - initiating land use spawning new land uses usually in a different place and existing land uses mutating into others (which might be seen as new) which are usually in the same place. In fact land uses which are in their declining phase make the transition to vacant land *in situ* at some point in this cycle. The way changes take place for new land uses depends on three different sized regions. Most important is the field or district which is wider than the local neighborhood within which the spawning or initiating land uses sits. The distance from the spawning land use is a determinant of where new land use takes place but within the more local neighborhood around this land use, the composition of other lands uses is instrumental in determining any state change. At the level of the region which is at the system level, constraints on what are or are not allowed in terms of cells being developed or not, are imposed. We list the three typical land use transitions as follows:

- $S_i^k(\tau = \textit{initiating}, t) \rightarrow S_j^l(\tau = 0, \textit{initiating}, t + 1)$  where the new use is in its initiating phase,
- $S_i^k(\tau = \textit{initiating-mature}, t) \rightarrow S_i^l(\tau + 1 = \textit{initiating-mature}, t + 1)$  where the changed use can be at a later stage thus reflecting properties of the old use,
- $S_i^k(\tau = \textit{declining}, t) \rightarrow S_i^*(\tau = \textit{vacant}, t + 1)$  where the new use is vacant land ready to come back onto the market at a later time period and thus available for new land uses being initiated from existing ones.

The first set of transitions which determine the growth process are influenced by the region, the field and the neighborhood, the second simply by the field and the neighborhood, and the last simply by the cell itself.

The dynamics emerging from this process is complex in that it is impossible to predict other than through simulating the total land use which is generated from this process at any one time. Total land use activity for any type is given as  $\sum_{i\tau} S_i^k(\tau, t) = S^k(t)$  and for all types as  $\sum_k S^k(t) = S(t)$ . Not only are these totals controlled by the land development process which operates from the bottom up and whose total predictive capacity is unknown prior to each simulation but the relative proportion of different land uses  $S^k(t)/S^l(t)$  are not controlled in any way and can vary dramatically. In one sense, this is an extremely desirable property of CA models for it means they are in the business of predicting total growth or decline which is largely absent from land use transportation models where such

totals are predetermined. However this is still problematic because the mechanisms at the bottom level based on the land development process are not designed with such total predictions in mind. In short this is an ambitious goal but much too ambitious given what little we know about such relationships and the way such features are built into the current generation of models.

The problems with this CA model like many others are manifold but in particular there are three key issues. First there is no feedback between demand and supply. Supply is imposed from outside in that the rules that are used to determine land use transition, hence growth and decline determine what is supplied, and it is supply that preconditions demand. This leads to the second problem which we have already noted: there is no control over the demand that the model supplies. Total demand has to be scaled artificially if it is to meet certain external known limits and if this is required, the model does not determine how such totals are generated. As we have implied, we consider it almost impossible to devise models based on bottom-up relationships which would produce feasible and realistic totals, at least at this stage. Third, there is no explicit transportation in the model. In fact, in *DUEM* we do generate streets as a distinct land use; we usually define commercial, industrial, residential, vacant land and two kinds of streets - junctions and segments. The number and the location of streets determine how many other types of land use can be generated. In short, one needs streets as infrastructure to enable other uses to be put in place and vice versa but apart from the physical infrastructure, there are no explicit interaction models which assign traffic flows, for example, to such streets. Moreover, this is the only form of transportation in the model and thus other forms of movement - electronic, by air, by rail and so on, are excluded.

## ***2.2 Pedagogic and Real Applications to Small Towns and Metro Regions***

To give an idea of what this model can be used for, we present three brief examples. First we can use the model to generate hypothetical growth patterns and one of the most useful simulations is to show how capacitated growth occurs and how land uses cycle in time through the urban space. As the space fills up, then land uses age and eventually disappear opening up more space for development. In this way the capacitated system cycles up and down and this provides a very useful diagnostic to see how all the various rules for transition

between land uses are balanced in the wider simulation. We show such a simulation in Figure 1 where the graphs demonstrate how housing, industry and services oscillate through time and how particular land uses begin to get the upper hand as reflected in the bottom-up rules which are prespecified. This is an important way of figuring out their plausibility.

In fact, Figure 1 provides a clear demonstration of how we do not know in advanced how the model predicts the relative ratios of different land uses. As this example is capacitated, when all land is filled, what happens is that we see quite clearly how the ratios of total housing to industry to services change: in Figure 1 we see how industry is gradually increasing relative to housing whose proportion is falling with services more or less constant in total, that is  $S^{industry}(t)/S^{housing}(t) \rightarrow ++$ . However what this kind of demonstration does show is that there is no stability in the predictions. We have not run this example for a very large number of time periods but it is entirely possible that in the limit, one land use would dominate and occupy all the space. It might be said that having a simulation device to show this is extremely useful yet all this is actually showing is a limitation of the model which is undesirable.

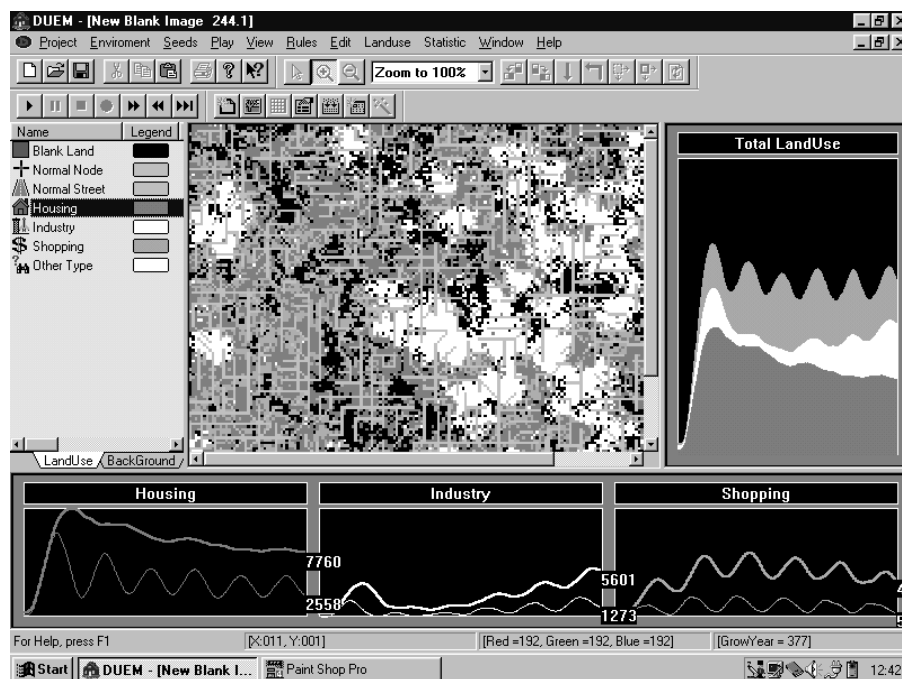


Figure 1: Forecasting Changes in Land Use Competition Using DUEM

We have also applied **DUEM** to the simulation of urban sprawl in Ann Arbor (Figure 2a) and to long term urban growth in the Detroit Metro Region (Figure 2b). In terms of the Ann Arbor application, when we run the model with the plausible default rules, we see immediately that housing growth is too focused along street and transport routes. Basically we find it hard to code into the model rules on clusters which must be in place if the sizes of housing development that characterize reality are to be simulated. This again is simply another way of showing how limited the model is. Our application to Detroit also points up the difficulty of this kind of model. In fact in the area shown in Figure 2b, the problem is that the simulations should show massive decline and abandonment in housing but again this is hard to simulate in the model. To develop such features, we need a much better supply side representation and we also need transportation and migration to explicitly represent socio-economic attributes and magnitudes associated with the relevant populations.

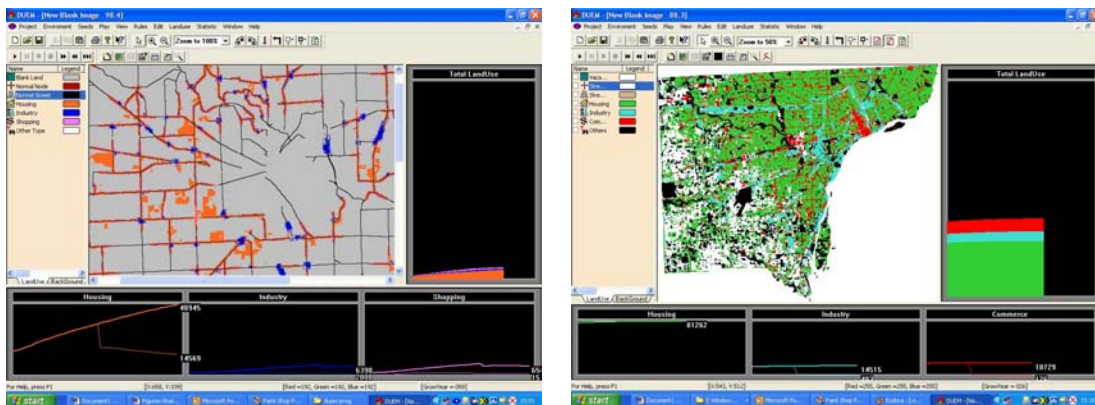


Figure 2: (a) left: *Simulating Sprawl in Ann Arbor MI:*

*Land parcels are hard to assemble into relevant clusters of real development and the simulation predicts growth along roads: the clusters are the existing changes between 1985 and 1990 which spawn the new growth along roads*

(b) right: *Simulating Long Term Housing Growth in Detroit:*

*Housing grows in this scenario but in fact Detroit is characterized by decline and abandonment, and it is simply an artifact of the closed space that growth takes place in this fashion*

A clear problem with **DUEM** and all cellular models involves the way cells are mapped onto real spatial units. Although the cellular grid is a fairly neutral means

of spatial representation in terms of CA models where the cells only contain one land use, this kind of representation is highly abstract. It might be possible to generalize cells into nodes which have differential sizes associated with them but this takes us into representing size in ways that CA models are unable to do. These problems are quite well-known and have been documented by O'Sullivan and Torrens (2000) but rather than dwell further on these limitations, we now need to sketch the way we are beginning to extend *DUEM* in integrating it with other models and new models.

### 3. The Design of *IDUEM*

We have presented *DUEM* in a number of previous papers (Xie, 1994; Xie, 1996; Xie and Batty, 1997; Batty, Xie and Sun, 1999). From our summary, it is clear that the model is highly physical in nature, built around a process of land supply driving urban development but without any of the detail of the economic decision processes which determine how land is supplied and then balanced with respect to consumer demand. In our extended model - *IDUEM* - we build a central core to the framework around the demand for and supply of residential use (housing and land) but feed the model with data and predictions from more well established disaggregate and aggregate models involving geo-demographic, geo-economic, transportation, migration and mover processes. The cellular representation is used as the visual interface to the simulation. In the model we are currently building which we sketch in this section of the paper, we will focus on five different issues

- the conceptual structure of *IDUEM* which will provide the reader with an immediate sense of what we intend.
- demographic and economic attributes of activities in cells which will show how the cell structure can be augmented in terms of these kinds of data and characteristics.
- tight and loose coupling to urban and regional planning models based on micro simulation, agent-based and integrated land use and transportation models.
- differentiated urban growth which is marked by the way space is filled and by new subdivision development which characterizes urban growth and sprawl, particularly in US cities.

- object-based simulation and programming which lies at the heart of how we are operationalizing and implementing the model as well as the construction of the graphical user interface.

### ***3.1 The Conceptual Structure of IDUEM***

There are two main directions in which we can extend CA models to make them more practically applicable in terms of different modeling styles. The first involves generating a much richer form of disaggregation to the level of the individual or agent and there is considerable momentum at the present time with this type of modeling (Parker, Manson, Janssen, Hoffman, and Deadman, 2003). There are several urban applications which show promise (Batty, 2003) but currently these are a long way from practical implementation and in some senses like any CA model, still tend to be pedagogic in nature rather than practical in the policy sense. The second and more conservative strategy is to link our CA models to traditional cross-sectional approaches based on land use and transportation models and more simplistic dynamics such as population and employment forecasting models. As we have indicated, these kinds of models are more pragmatic in structure but are operational and well-established. In fact what we will do is steer our developments to the latter while at the same time having regard to the former, particularly within the core of the model where we will build around mechanisms to make explicit demand and supply processes governing urban land at the individual household level.

This core is a model which might be regarded as agent-based but with links directly to more aggregate models. In essence we consider that there are three main types of model reflecting different sectors of the urban system that need to be formally represented. First, overall demand for urban activities, specifically employment and population, can be factored into different kinds of detailed activity such as services, entertainment, population types and so on, as well their related attributes such as incomes, rents etc. These are best simulated using conventional demographic and economic models such as those which are built around population cohort survival, spatial input-output/urban econometric models, and so on. These provide small area activity forecasts that dimension the more detailed demands which lead to land use change at the level of building blocks or the parcel. The second type of model reflects ways of simulating land supply and these are largely based on land suitability analysis. In a sense, the dictates of the market for land supply are not yet represented in our proposal for at this point we feel that adding issues involving

mortgage and capital markets that do clearly influence land supply, is beyond the capability of these models. Land suitability analysis extending to accessibility and related environmental issues is as far as we will go in the current proposal. However, the design philosophy of *IDUEM* (based on the common object model) will support the extensibility and interoperability needed for integration with types of model derived from mortgage and capital markets in future.

The singly biggest problem in existing CA models is a lack of a transportation sector but in our proposal because we will be modeling detailed movements within the housing stock, we will link these to accounting methods which will be made consistent with discrete choice. These will run within the background but will link to the geo-demographics and economic models and the land suitability analysis by dimensioning, keeping the quantities predicted within reasonable limits. Finally we see cellular representation as being much relaxed in the overall model in that we will abandon the strict neighborhood-field characterization in favor of tagging individual land parcels and groups of individuals. However the cellular approach is still useful in terms of visualization and to all intents and purposes, at least superficially, the model will continue to be a CA-like structure. In Figure 3, we show the general structure of the model as it is currently developing. We elaborate this in two ways below, in terms of its system architecture and in terms of its detailed submodels. Of course, there are many different ways of looking at such a rich structure but in this paper, we will not get down to specific ways of implementing the entire structure formally for this is very much a work in progress, whose theoretical structure and applications will be reported as they evolve.

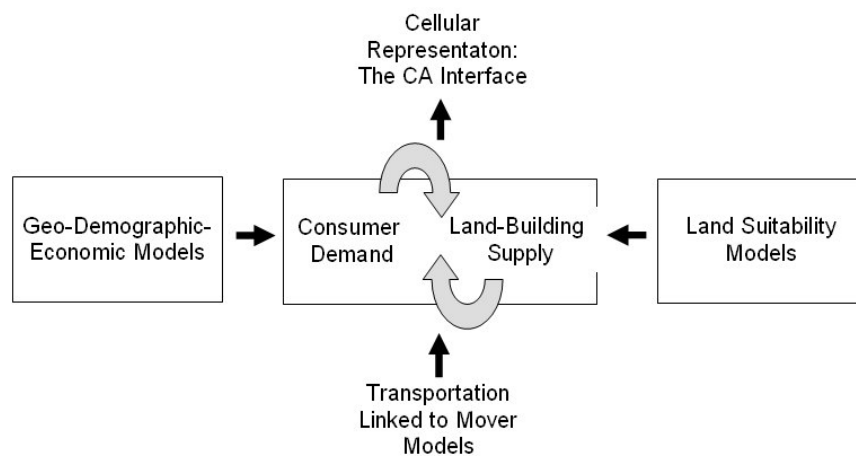


Figure 3: The Aggregate Structure of the *IDUEM*

### 3.2 Augmentation of Cell Attributes and Space

According to the classical definition of cellular automata, such a system consists of two basic elements: a cellular space (which is *structural*), and a set of transition rules (which are *functional*) defined over that space. The cellular space is an *infinite* *n*-dimensional regular Euclidean space on which a neighborhood relation is established. This neighborhood relation specifies a *finite* list of cells which are called neighbors. In applications, the CA space usually consists of a set of regular grids of the same shape and size, and the CA neighborhood is a subset embracing the same number of cells and displaying a similar structure. A neighborhood usually includes a very small number of cells for two reasons: first a large neighborhood can lead to tremendous difficulties when formulating CA transition rules; second, local rules where action-at-a-distance is minimal, with cells comprising only first nearest neighbors for example, give rise to global patterns which are unexpected and have macro structure. This is widely regarded as evidence that such structures are rather good at simulating emergence whose signature is fractal. For instance, von Neumann's construction of a neighborhood only considers four cells around a central fifth cell, while the most used neighborhood (after Moore) includes 8 cells surrounding a ninth.

When applying the CA paradigm to spatial systems with policy/planning applications, it is inevitable that the concept of the CA neighborhood be regarded as a crucial bond which connects cellular automata with geographical phenomena. In the context of urban growth, an ideal space unit is likely to be a property parcel as used to represent land in North American legal, real estate and urban planning applications. A parcel is the smallest cadastral mapping unit. It shows directly the property boundaries associated with land ownership (who owns what and where), property values, its development history, land use type, and the building type and structure which often occupies the land. It indirectly represents who occupies the plot as well as the occupants' demographic, social, economic, and personal characteristics and behaviors. Furthermore, a parcel is a dynamic commodity circulating on the market as a result of economic development or household change. We have thus decided that the parcel must be formally mapped onto the cell in building new simulation tools to take account of cadastral data. The integration of parcel space involves us in two development phases: first to "cellularize" attribute data based on parcel cadastral data (aggregating parcel demographic, social and economic data to cells), and then to replace cells with parcels in building a new generation



of CA with appropriate spatial neighborhoods for conducting dynamic simulation. The latter approach will be implemented and discussed at a future time, but to anticipate how action-at-distance is to be handled, the traditional notion of the restricted physical neighborhood will be much relaxed. It will still exist in *IDUEM* but it will only serve certain obvious functions related to site whereas situation will be a function of the general kinds of field theory that underlie spatial interaction modeling.

In conventional CA modeling, the state of a cell in the context of urban simulation, is often either binary (developed urban area or open space) or a type of land use as in *DUEM* (residential, industrial, commercial, agricultural, open, etc). CA in its classic form, simulates cell state changes based on existing cell states and the spatial configurations in their neighborhood (Batty, Xie and Sun, 1999; Li and Yeh, 2000). This is entirely different from traditional modeling practice where urban activities albeit associated with land use, are the objects of simulation. CA models thus miss demographic and socio-economic attributes and this makes them difficult to root in conventional urban theory. CA's traditional focus on cell states and restricted neighborhood configurations confines them to pedagogic uses, useful as metaphors for spatial exploration or sketch planning tools rather than tools for practical planning or prediction (Batty, 1997).

One of the most significant developments in *IDUEM* is to augment a cell's attribute from a single variable "state," to a comprehensive array of demographic, social and economic variables. A cell will thus take a new form as "an object". Physically the cell object will have a size dimension (100 meters or so), encompassing several parcels, containing several buildings (houses or factories), and household or employment types. The cell object thus represents several sets of attributes, such as household, building (housing), economic, land use and environmental data. These will in fact be managed by an external database and we envisage that a commercialized relational database management system (RDBMS) such as Microsoft Access will be chosen at this stage so that the *IDUEM* software package is easy to run and to maintain. Cell objects and other GIS datasets will be stored as feature data layers. External models have not been used very much so far in classic CA modeling. The viability of a cell (automaton) is usually determined by its spatial configuration, not its characteristics but our augmentation of cell attributes will transform this traditional notion of CA. CA attributes in *IDUEM* truly determine the

dynamics of cell objects. Moreover CA attributes are direct data inputs to urban and regional models that will be both tightly and loosely coupled in *IDUEM*.

One of technical breakthroughs in *IDUEM* is the seamless integration between the cellular space, which is the model infrastructure of CA or agent-based models, and geographical space of areas, on which aggregate socio-economic models are built. This technology makes possible the augmentation of cell attributes in the context of traditional modeling applications. As Figure 4 illustrates, the upper left panel of *IDUEM (DUEMPro)* Application (the current pilot model) is the content window of vector data layers, representing area-based socio-economic models. The lower left panel is the content window of raster data layers, typically called housing-building modeling (HBM) data layers in *IDUEM*. This kind of juxta-positioning has not been possible in traditional CA models which have tended to work quite literally at the cell/pixel level. *IDUEM* breaks with this tradition as much because data is hard to force into a cellular representation and the way the land market works requires cells to be configured as plots in that land assembly for development is very critical in enabling realistic allocations to be simulated in residential sector which is at the core of this framework.

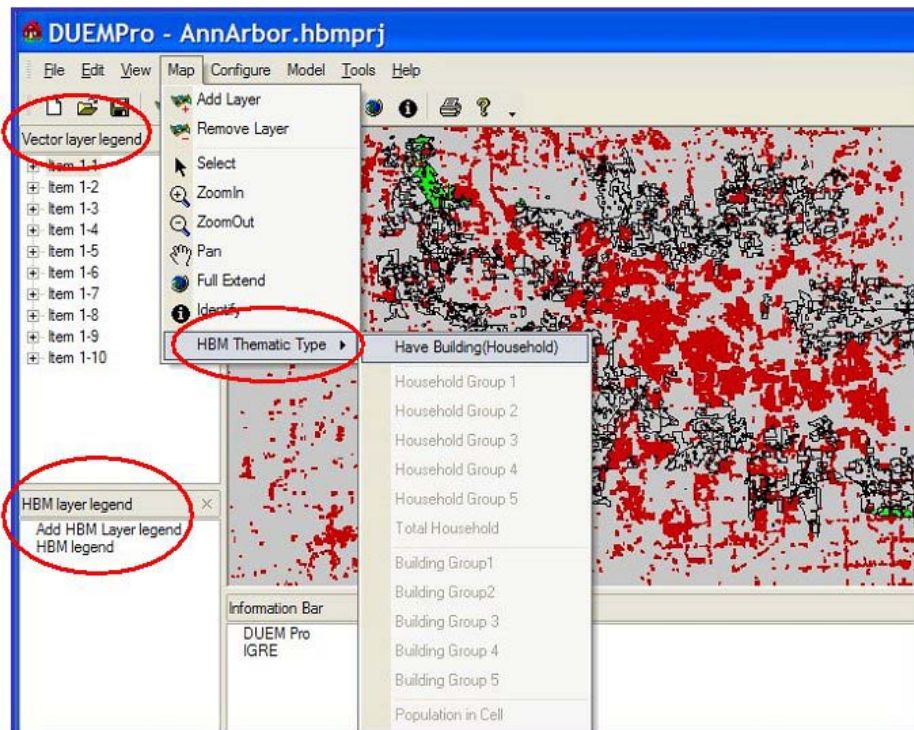


Figure 4: A Snapshot of the Vector and Raster/Cellular Layers Reconciled within the *IDUEM* Simulation for Ann Arbor, Michigan

### ***3.3 Tightness of Coupling with Urban and Regional Planning Models***

One of the major issues in modern software design involves the extent to which software developed for one purpose which might be entirely compatible with another can be linked to that other software in the most effective and seamless way. The loosest such coupling simply involves transferring files traditionally using manual means but more recently through automated desktop and network systems. However tighter coupling is usually more desirable especially if functions traditionally in one software package are to be split up between many. For example, in *IDUEM* the core demand-supply model leads to the movement of households, firms and shops etc and in turn sets up changes to the transportation flows that are involved, traditionally such flows being estimated using aggregate models that do not interface easily with migrations of any kind over however short a scale. In augmented and extended CA modeling of this kind, we thus need to consider such linkages and this suggests at least that in terms of the transportation models that we will build, these will be quite strongly coupled into the system. This is in contrast to more aggregate models such as population forecasting which can produce predictions at higher levels that can be easily factored into separate software packages as control totals.

External models are also used to produce simulation parameters or constraints on the control growth rate, location, or pattern of CA simulations, although there are many efforts reported in literature to integrate external models with CA (Xie, 1996; Batty, Xie, and Sun, 1999). *IDUEM* takes an integrated approach to rely on data and data-driven models to answer the question of why growth happens and what is the driving force of dynamic urban automata. In short *IDUEM* accepts the common notion in urban studies that growth and development is driven by economic development and associated demographic change.

There are three traditional types of cross-sectional static urban model that have been developed over the last 50 years. These are based on spatial-interaction, discrete choice, and spatial input-output analysis in regional econometric form. From an operational point of view, the most popularly referenced models in North America include generalized urban models of the Lowry vintage namely: the *DRAM/EMPAL* models developed by Putman (1983), and the spatial input-output

*TRANUS* and *MEPLAN* models, developed respectively by de la Barra (1989) and Echenique (1994); urban economic oriented structure such as the *CATLAS* (and later *METROSIM* and *NYMTC-LUM*) models developed by Anas (1982), and the *MUSSA* models developed by Martínez (1992); pragmatic land development models with substantially GIS-like functionality such as the California Urban Futures (*CUF*, *CUF-2*) Model (Landis, 1994); and the more comprehensive (and more recent) model structures incorporating discrete choice and disaggregate micro-simulation such as *UrbanSIM* (developed by Waddell, 2000). These models are discussed in detail in several recent reviews (Schock, 2000; Guhathakurta, 2003).

We are considering all these model structures and thinking about how we might be able to interface *IDUEM* with several of these. We think that a tight coupling with urban and regional models must take advantage of rich demographic, social and economic data that exists at a micro level for this is an important reflection of a cell object. This data decides the dynamic of each cell objects and the consequent simulation processes. Another important consideration is the “stimulus” that cell dynamics gives to growth and change. We want to model a cell’s viability for change or its “mobility” which is missing in current CA models. Therefore, *IDUEM* focuses very strongly on mobility type models which determine how urban growth is activated. We show this in more detail than in Figure 3 in Figure 5 below.

Population predictions are generated from macro regional socio-economic models and are reported over cities, or townships, or minor civil divisions. The technique of Iterative Proportional Fitting (IPF) is implemented in *IDUEM* to partition area-based predictions over the grid space and over time (Beckman *et al*, 1995). The Household Mobility Model (or the Employment Mobility Model) simulates households (or jobs) which involve decisions to move from current locations. Multinomial logit models are applied to historical data to determine movement probabilities as in *UrbanSIM* (Waddell, 2000). Once a household or a job decides to move, it no longer has a location in the study area, and is placed in a temporary allocation set or pool. The space it formerly occupied is made available for “CA space-filling” which we describe briefly below. This household (or job) in motion will be placed by CA space-filling into what we call ‘groundbreaking construction’ (of a new subdivision) simulation (discussed in next section). A similar approach is taken in implementing the Building Mobility Model.

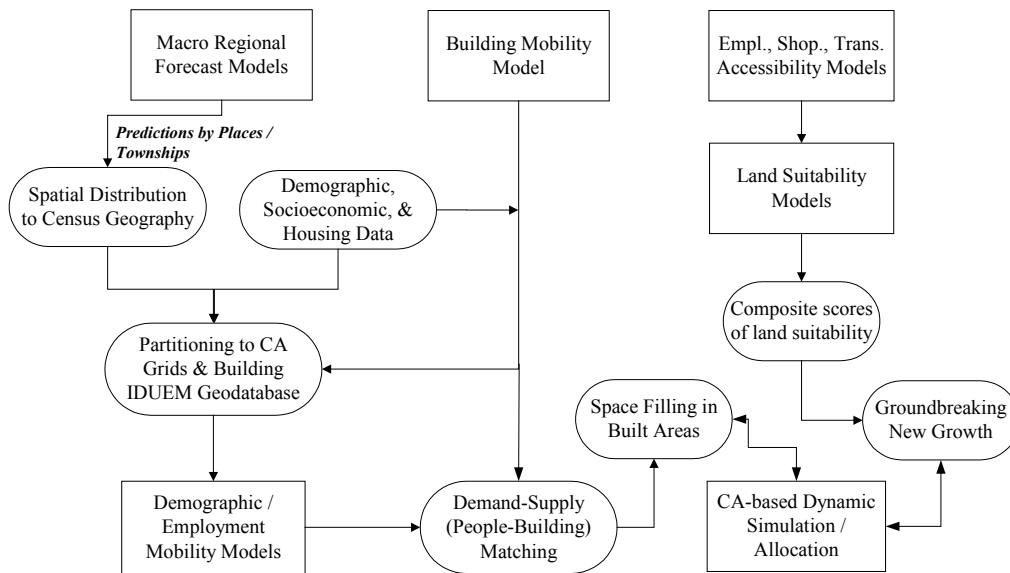


Figure 5: The Full Model Structure for IDUEM

### 3.4 CA Space Filling and New Subdivision Development

Urban change takes place in many forms across time and space. Inner city decline is often associated with suburban sprawl. Progressive change in the building stock through deterioration, renovation, demolition, and new construction in existing built areas is accompanied by groundbreaking construction of new subdivisions and new urban centers. *IDUEM* supports simulation of both progressive dynamics which is a more detailed staging of the life cycle effects of *DUEM* with groundbreaking growth which is largely determined on the supply side through GIS-based land suitability modeling (Figures 3 and 5). The regional macro demographic and economic models determine growth predictions for an area as a whole while the population and employment mobility models determine the number of people who are going to move.

The building mobility model provides the answers with respect to how existing buildings accommodate such moves. The CA simulation allocates people to existing buildings (through matching their socio-economic characteristics) or determine in new structures through ground breaking construction in which they are housed. Additional persons who are not balanced by the usual equilibrating movements of

demand and supply represent the sources for development of new subdivisions. GIS-based land suitability models are executed to rank sites with respect to their suitability. Such suitability models will include sets of tools for calculating accessibility scores for transportation convenience, employment opportunities, and shopping choice; assigning weights and ranks to relevant and available GIS data layers of interest; and composing final scores of suitability for sites available for new development. The suitability scores will determine the order of available sites in terms of the way the CA allocations take place.

The way in which demand and supply is reconciled within *IDUEM* is still under discussion. Because demand is conceived in terms of households and supply in terms of houses/dwelling (buildings) on plots, then the model operates at a much finer scale than *DUEM* where supply was predicted as  $S_i^k(\tau, t)$  and demand was simply assumed to be always equal to supply  $S_i^k(\tau, t) = D_i^k(\tau, t)$ . We can think of this supply and demand as being measured in terms of households and dwellings although in *IDUEM* one of the processes is a complex balancing which takes place within the temporal structure of the model with any imbalance - where the market has failed to clear - being left until the next time period. The alternative approach which is attractive is to reduce the length of the time interval to a sufficiently small unit to ensure that any imbalance need not be dealt with until the next time period, assuming that such imbalances were a realistic feature of the system (as they probably are). In this way housing market equilibrium would be an ever shifting target.

### ***3.5 An Object-Based Simulation And Programming Approach***

Our last foray into outlining this model structure will briefly note some programming details. As we have strongly emphasized, this paper is a progress report written to present the overall structure of *IDUEM* as well as to indicate how we are implementing this structure. *IDUEM* is a new generation of CA model coupling movement (demographic and employment) with building and land use development and as such, an object-oriented approach dominates the new design. First, cellular automata (cells now and parcels later) in *IDUEM* are ***objects representing basic urban units***. The urban objects are characterized by two groups of properties: first the physical properties which are characteristics of building stock, land/property values, land availability scores, environmental

amenities, accessibilities to development stimuli, and adjacencies to existing development; and second, the socioeconomic properties including the attributes of population, the number and size of households, age composition, economic situation, employment status, travel time to work, and recent changes in these demographic and socioeconomic characteristics.

The socioeconomic subsystem interacts with the physical subsystem following the classic equilibrium of demand and supply. The socioeconomic subsystem takes into consideration demographic and socioeconomic changes, predicts probable movements of people in terms of matching buildings with preferences, and determines the demand for housing and land development. The physical subsystem simulates the supply of buildings and probable locations of future development based on land suitability. *IDUEM* is thus the first object-oriented CA model to explore the interactions of two most important phenomena of urban growth, detailed migration patterns with respect to housing and subsequent land development. These are in fact the drivers of urban sprawl in particular and urban growth in general whose understanding and prediction remains the rationale for this kind of model, at least in the first instance.

Urban entities and their physical and socioeconomic properties are analyzed and processed in the style of objects. The match (equilibrium) between demand and supply (between household migration, building construction and land supply) is realized through multiple variance analysis of object properties. One simple illustration in the context of residential growth simulation is that households are classified according to the median household income, while housing is categorized by the type. The median household income is classified into five object types: very low, low, middle, upper middle, and high, according to the natural breaks within the data. Housing is divided into five types: apartment (AP), duplex (DUPLEX), low amenity single family (SF1), middle amenity single family (SF2), and high amenity single family (SF3). Probabilities of transition between types and between incomes are generated through intersections between the Block-Group Data Layer (containing the median household income data) with the Land Use Data Layer (including different types of housing information). The matching probability set is then used to determine where the household might move according to its household income. The CA simulation routine will finally place the household in a

location based on the CA rules and the physical and socioeconomic properties of a land cell object.

**IDUEM** is being coded as a suite of object oriented programs following the common object model (COM). The database management is being implemented as ActiveX database access objects. The user interface and graphic visualization are coded as VC++ multiple documents and VC++ DLLs. The models coupled with **IDUEM** are being slowly integrated through programming as either VC++ DLLs or Java Serverlets. In Figure 6, we show the typical user interface to generating an allocation of households to houses (as illustrated earlier in Figure 4) which illustrates the degree of control the user has over the simulation.

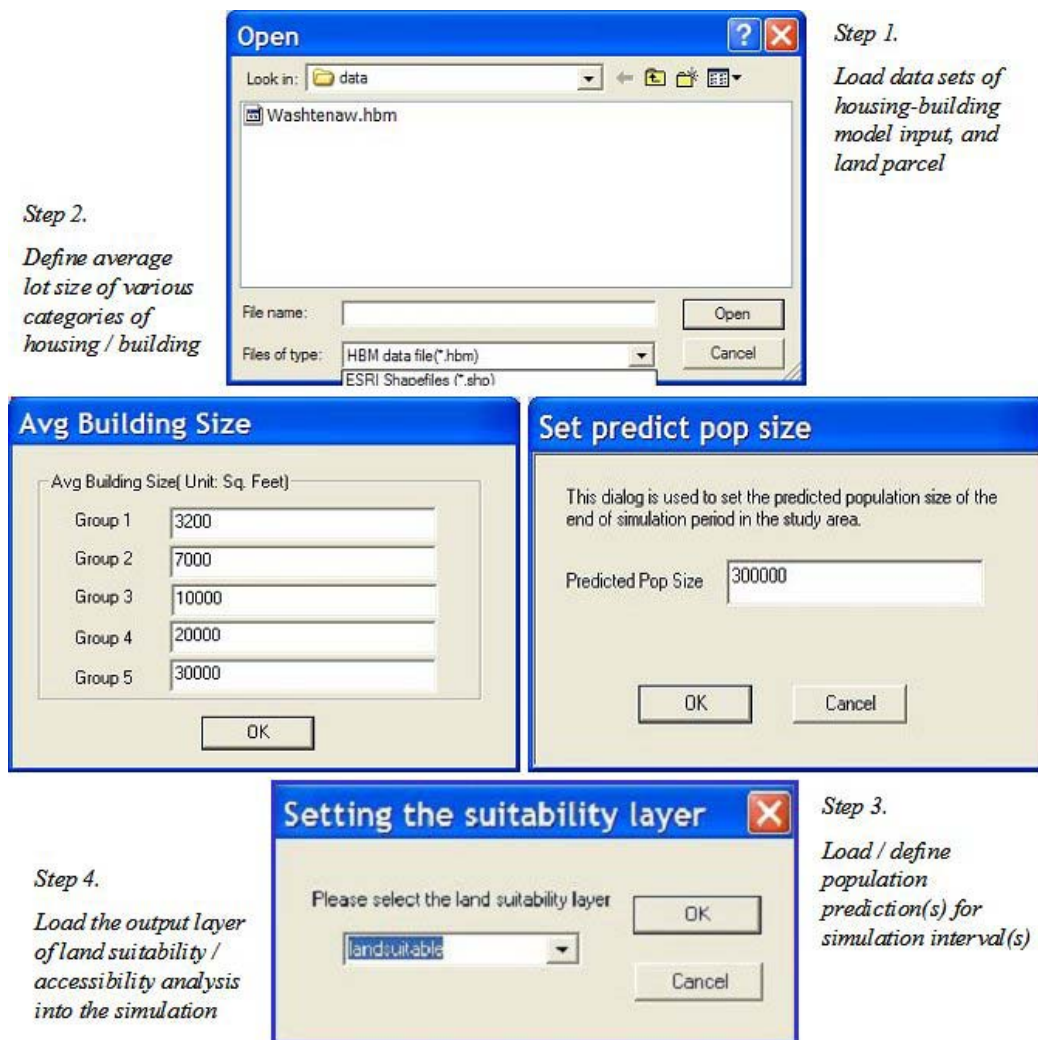


Figure 6: Screenshots of the **IDUEM** Simulation for the town of Ann Arbor, Michigan



At Step 1, we load both grid-based housing/building modeling (HBM) data, and area-based land suitability data. The HBM data layer is a composite data layer through several preprocesses. They include grouping and categorizing residence housing data, grouping and categorizing socio-economic data from Census Block Group Data, matching through multiple variance analysis. The land suitability data is the union outcome of land accessibility analyses and land suitability analysis. The tools of objectizing, categorizing and matching housing/building with household are not packaged in *IDUEM* at this moment, but will be in the future. The two pull-down menus (in non activated mode) in Figure 4 are prototype tools for viewing and re-configuring the matching table of housing/building with households.

Step 2 shows how to parameterize housing/building size constraints in the simulation. Step 3 illustrates an integration of population prediction in the simulation. Step 3 supports more complex situations, the predictions over multiple intervals and over subdivisions. A file browser will be opened for selecting a file that contains the prediction data in this complex simulation. After the specification of predicted population, the predicted value(s) will be partitioned over time and space by an IPF (iterative proportional fitting) routine and then a probability matrix of mobility will be calculated to guide the simulation. Step 4 allows users to confirm which layer contains the results of the land suitability analysis. Then *IDUEM* simulation will be launched. One simulation outcome was shown earlier in Figure 4 where a detailed analysis shows that already the model is successful in predicting sizeable developments which account for the physical properties of land parcels, households demand and the structure of the building process.

## 5. Next Steps: Current Development and Future Plans

The focus of our current work is on building the demand-supply core of the model, linking this to external geo-demographic and geo-economic models and to land supply suitability potentials. Currently we have not attempted to handle the transportation component in anything but a cursory and temporary way. All our applications are being tested using data from the urban areas of Ann Arbor and Ypsilanti in Michigan where our focus has been and continues to be on predicting the location of population growth and housing development between 1985 and

2020, as we illustrated in Figure 2(a). As Figure 4 reveals, we are also conscious that models such as these should be as tightly coupled as possible with contemporary and proprietary GIS software and data formats and to this end the user interface has many features which enable links to such external software.

What this paper has illustrated is that for large scale urban modeling projects, many different components need to be developed both sequentially and in parallel and at any point in the development of the wider framework, it is incomplete. This paper has simply sketched the first stages of this model development and offered some snapshots of progress. However we consider that for CA models to become applicable in the urban planning process specifically and more generally as part of land development, then it is necessary to move away from the literal cellular frame itself and begin to incorporate the detail of the geometry and geography of the real city as well as its linkages through transportation activity. To do this, many different modeling traditions need to merge and this suggests that integrating different models, establishing consistency between them, and making them work together should be at the forefront of this variety of urban simulation.

## 5. References

Allen, P. M. (1997) *Cities and Regions as Self-Organizing Systems: Models of Complexity*, Taylor and Francis, London.

Anas, A. (1982) *Residential Location Markets and Urban Transportation: Economic Theory, Econometrics, and Policy Analysis with Discrete Choice Models*, Academic Press, New York.

Batty, M. (1997) Cellular Automata and Urban Form: A Primer, *Journal of the American Planning Association*, **63**, 266-274.

Batty, M. (1998) Urban Evolution on The Desktop: Simulation using Extended Cellular Automata, *Environment and Planning A*, **30**, 1943-1967.

Batty, M., Xie, Y., and Sun, Z. (1999) Modeling Urban Dynamics through GIS-Based Cellular Automata, *Computers, Environments and Urban Systems*, **233**, 205-233.

Batty, M. (2003) Agents, Cells and Cities: New Representational Models for Simulating Multi-Scale Urban Dynamics, A paper presented to the Conference on *Framing Land Use Dynamics*, University of Utrecht, The Netherlands, April 16-18, 2003.

Beckman, R. J., et al. (1995) *Creating Synthetic Baseline Populations*, Transportation Research Board Annual Meeting, Washington, DC.

- Bell, M., C. Dean, and Blake, M. (2000) Forecasting the pattern of urban growth with PUP: A Web-based Model Interfaced with GIS and 3D animation, *Computers, Environment and Urban Systems*, **24**, 559-581.
- Clarke K, Gaydos L. (1998) Loose-Coupling a Cellular Automaton Model and GIS: Long-Term Urban Growth Prediction for San Francisco and Washington/Baltimore, *International Journal of Geographical Information Science*, **12**, 699 - 714.
- Crosby, R. W. (1983) Introduction and Asking Better Questions, in R.W. Crosby (Editor) *Cities and Regions as Nonlinear Decision Systems*, Westview Press, Boulder, CO, 1-28.
- de la Barra, T. (1989) *Integrated Land Use and Transport Modelling*, Cambridge University Press, Cambridge, UK.
- Echenique, M. H. (1994) Urban and Regional Models at the Martin Centre, *Environment and Planning B*, **21**, 517-534.
- Forrester, J. W. (1969) *Urban Dynamics*, MIT Press, Cambridge, MA.
- Guhathakurta, S. (Editor) (2003) *Integrated Land Use and Environmental Models*, Springer-Verlag, Berlin.
- Holland, J. H. (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.
- Landis, J. H. (1994) The California Urban Futures Model: A New generation of Metropolitan Urban Simulation Models, *Environment and Planning B*, **21**, 399-420.
- Li, X., and Yeh, A. G. O. (2000) Modelling Sustainable Urban Development by the Integration of Constrained Cellular Automata and GIS. *International Journal of Geographical Information Science*, **14**, 131-152.
- Martinez, F. (1992) The Bid-Choice Land Use Model: An Integrated Economic Framework, *Environment and Planning A*, **24**, 871-885.
- O'Sullivan, D., and Torrens, P. M. (2000) Cellular Models of Urban Systems, in S. Bandini and T. Worsch (Editors) *Theoretical and Practical Issues in Cellular Automata: ACRI'2000: Proceedings of the Fourth International Conference on Cellular Automata for Research and Industry*, Springer-Verlag, London, 108-117.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffman, M. J., and Deadman P. (2003) Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review, *Annals of the American Institute of Geographers*, **93**, 314-337.
- Portugali, J. (2000) *Self-Organization and the City*, Springer-Verlag, Berlin.
- Putman, S. (1983) *Integrated Urban Models: Policy Analysis of Transportation and Land Use*, Pion Press, London.
- Schock, S. (Editor) (2000) *Projecting Land Use Change*, EPA/600/R-00/98, National Exposure Research Laboratory, EPA, Washington DC.

Waddell, P. (2000) A Behavioral Simulation Model for Metropolitan Policy Analysis and Planning: Residential and Housing Market Components of UrbanSIM, *Environment and Planning B*, **27**, 242-263.

Weidlich, W. (2000) *Sociodynamics: A Systematic Approach to Mathematical Modelling in the Social Sciences*, Harwood Academic Publishers, Amsterdam, The Netherlands.

White, R., and 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**, 1175-1189.

Wilson, A. G. (2000) *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*, Prentice Hall, Harlow, Essex, UK.

Wolfram, S. (2002) *A New Kind of Science*, Wolfram Media, Inc., Urbana, IL.

Wu, F. (2002) Calibration of Stochastic Cellular Automata: The Application to Rural-Urban Land Conversions, *International Journal of Geographical Information Science*, **168**, 795-818.

Wu, F., and Webster, C. J. (1998) Simulation of Land Development through the Integration of Cellular Automata and Multi-Criteria Evaluation, *Environment and Planning B*, **25**, 103-126.

Xie, Y. (1994) *Analytical Models and Algorithms for Cellular Urban Dynamics*, unpublished PhD dissertation, State University of New York at Buffalo, Buffalo, NY.

Xie, Y. (1996) A Generalized model for Cellular Urban Dynamics *Geographical Analysis*, **284**, 350-373.

Xie, Y., and Batty, M. (1997) Automata-Based Exploration of Emergent Urban Form, *Geographical Systems*, **4**, 83-102.