A microscale simulation of land prices, household and employment growth using UrbanSim in Yongsan-Gu, Seoul

Mohammad Mehedy Hassan & Chulmin Jun


To link to this article: http://dx.doi.org/10.1080/12265934.2011.626175

Published online: 15 Dec 2011.

Submit your article to this journal

Article views: 257

View related articles
A microscale simulation of land prices, household and employment growth using UrbanSim in Yongsan-Gu, Seoul

Mohammad Mehedy Hassan and Chulmin Jun*

Department of Geoinformatics, The University of Seoul, Korea

(Received 11 July 2011; revised version received 29 August 2011; final version accepted 7 September 2011)

Since cities continue to suffer from traffic congestion, high growth population, increased pressure of upgrading infrastructure and deterioration in environmental quality; many metropolitan organizations have switched their focus on growth modeling at micro level in order to estimate patterns, evolution and behavior of major elements involved in urban processes. The aim of this study is to employ the UrbanSim microsimulation land use model for Seoul city. This study attempts to simulate a set of UrbanSim land use models based on developed a dataset from Yongsan-Gu. The employed method includes microscale simulation of land prices, household and employment growth in Yongsan-Gu. In addition, by using the statistical model, various factors regarding location, neighborhood, built environment, economics, and demographics are analyzed in order to assess spatial variation of residential and non-residential location patterns. The input data preparation, model estimation, and strategy that were employed within the microsimulation model on different land use characteristics are described in this working paper.

Keywords: UrbanSim; microsimulation; land use model

Introduction

In recent years, dynamic microsimulation land use and transport models have been increasingly applied to examine and mitigate the accelerated growth of urban complexity, policy analysis and many other decisions regarding urban infrastructure and transportation investment (Brown & Harding, 2002). In current practices one can find a relatively common land use model such as Putman’s (1983) gravity-based ITLUP, input–output models like de la Barra’s (1989) TRANUS, Hunt and Abraham’s PECAS (2003), and microsimulation-based land use models such as Landis and Zhang’s (1998) CUF and ILUTE, and Waddel’s (2002) UrbanSim. Other models include Simmond’s (1999) DELTA, Martinez’s (1996) MUSSA, and UPlan (Johnston & Garry 2003).

In fact, all earlier land use models were spatially aggregated (Moeckel, Spiekermann, Schurmann, & Wegener, 2003), static equilibrium (Wegener, 2004) and used very coarse geographic units of analysis, typically by aggregating in zone or larger administrative districts (Wagner & Wegener, 2007).

*Corresponding author. Email: cmjun@uos.ac.kr
Many of these models forecasted poorly and are not very responsive to policy analysis (Moeckel et al., 2003). None of them essentially represent land markets which lacked in spatial details (Icono, Levinson, & El-Geleidy, 2008). In current practice, the most common geographic unit of analysis for microscopic land use models is grid cell, using resolutions ranging from 30 meters to 150 meters. Figure 1 shows the typical geographical units used for urban land use and transport modeling.

The ultimate goal of this study is to apply UrbanSim land use model to the Seoul region. UrbanSim (http://www.urbansim.org/Download/WebHome) is an open source microsimulation model for urban land use in larger metropolitan regions that predicts the growth patterns of household, employment, land price, and development events for 20 years or more (Waddell, 2002; Waddell & Ulfarsson, 2004). The model consists of six sub-models: accessibility, household location choice and relocation model, employment location choice and relocation model, real estate development model and land price model. Economic and demographic transition models are external models as well. Compared to other land use models, In the UrbanSim framework, the modelers can develop their database using any of the convenient spatial raster units described in Figure 1. In addition, model behavior and specific cases can be analyzed at the level of individual households and buildings.

Part of this research project includes a list of models calibrated on developed data sets from Yongsan-Gu, Seoul. While this study is the first attempt in using a microsimulation land use model in Korea, model estimation and result analysis are restricted, as a result of insufficient disaggregate data on households and business space. However, results from model calibration and validation from observed data provide evidence that, despite a number of overestimations and underestimations of population and employment growth, the employed modeling framework is able to capture the overall land use patterns at both aggregate and disaggregate levels. This study provides an outlook on what microsimulation models look like and what is required for calibration on different land use characteristics. Finally, some directions for future research are indicated, which focus on required resources for microsimulation models on different land use patterns.

Materials and methods
Like any microsimulation land use model, UrbanSim requires huge, substantial amounts of disaggregate input data. Much of this study revolved around developing the required data inputs for UrbanSim. Spatial data processing and analysis were performed using ArcGIS 9.3, while tabular data was processed using a combination

Figure 1. A typical geographical unit use for land use and transportation modelling.
of Microsoft Excel, Access and the MySQL database program. The custom software tool, Navicat, was used to facilitate MySQL data to improve the workflow and ensure consistency in data handling.

The study area Yongsan-Gu was divided into 150 meter × 150 meter (see Figure 2). Each gridcell corresponds with various land use attributes, these include building information (commercial, industrial, governmental, residential, and year built), land information (residential, non-residential land value, zoning types, and development types), environmental features (wetland, floodplain, etc.) and location information (distance to highway, arterials, urban growth boundary, X–Y coordinate, and other accessibility). These land use characteristics were then classified and stored into each corresponding gridcell.

The residential units in the gridcell were obtained from household travel surveys (2002; 2006). Building locations were traced out by using a daum map (http://local.daum.net/map) and GIS point data from the Building Inventory 2005. However, no data exist as to the characteristics of individual households at specific locations. To resolve this problem, an effort was made to make synthetic household data by using demographic and economic survey data from 2005. The synthesized household data include: age of head of household; household income; size; number of cars; and the number of workers. Although all supplied data were indeed in zone aggregation, GIS and other statistical techniques were used to disaggregate zone data to a 150 square meter gridcell. In this process, actual population size and the total number of households of the study area were maintained. The employment tables for the base year simulation were developed from the non-residential square divided by average square feet per job. Then these tables were classified into eight sectors following Koran Standard Industrial Classification 2000 (http://kostat.go.kr/kssc) and further aggregated into one of three sector types: (1) Commercial, (2) Industrial, and (3) Governmental. The non-residential square feet were obtained from the Building Register (2002, 2003, and 2005). This building inventory contained information on business location and a minimal amount of square feet but lacked the information required for an input database. Nonetheless, non-residential square feet were proportionally developed in relation to total employment in the study area. The construction year of each building was significantly important since location, choice of household, jobs, and model estimation relatively varied depending on the building construction year. This study attempts to develop year-built data by taking samples from 100 buildings from the study area, and then year-built data were generated by using data interpolation methods. This was essentially the strategy pursued, although clearly it would not be as accurate as it would have been if the assessor

Figure 2. Spatial and non-spatial data construction process.
files had maintained these data. This study intends, however, to create as clean as possible base year data, in order to maximize the utility of simulation processes for learning about the model system performance.

A set of distance measures were computed at the gridcell scale, since it is believed that individuals will choose to live in closer proximity to positive amenities and services, while maintaining distance from nuisance and undesirable land uses. In ArcGIS9.3, proximity measures from each grid centroid to the nearest road, arterials, bus stops, subway, and Han River were carried out. These distance metrics were then written into gridcell table and used for model estimation.

Model estimation and simulation process

UrbanSim sub-models rely on coefficients that are estimated through liner regression, logistic regression, or probabilistic functions. Model estimation is an iterative process with each step yielding a unique set of parameters, coefficient values, and metrics for evaluating model fit. UrbanSim provides model estimation tools within its environment, and this study made use of these tools. The set of estimated equations were stored in a database for further evaluation. The following describes UrbanSim sub-models which were employed in Yongsan-Gu.

Land price model

Empirical studies suggest that land prices in urban areas are quantified based on various factors. These are location utilities (Dubin & Sung, 1990; Isakson, 1997; Jordaan, Drost, & Makgata, 2004), neighborhood characteristics (Wingo, 1963; Lin & Jhen, 2009; McCain, Jensen, & Meyer, 2003) and environmental qualities (Kim, Park, & Kweon, 2007). Hedonic regression models (Rosen, 1974) are commonly applied in real estate and urban economics (Bible & Hsieh, 1999; Liu, Zheng, Turkstra & Huang, 2010; Vural & Fidan, 2009; Cervero & Kang, 2010). It is also the common method used to analyze land markets, property prices, and land use and transport (Waddell & Ulfarsson, 2004). The price of land in UrbanSim was estimated using a hedonic regression analysis. The simple form of the hedonic model is as follows:

\[
P_{ilt} = \alpha + \delta \left( \frac{V^c_t}{V^s_i} \right) + \beta X_{ilt}
\]

Where, \(P_{ilt}\) is the price of land per gridcell of development type \(i\) at location \(l\) at time \(t\), \(V^c_t\) is the current vacancy rate at time \(t\), weighting local and regional vacancy, \(V^s_i\) is the long-term structural vacancy rate, \(X_{ilt}\) is a vector of locational and site attributes, and \(\alpha\), \(\delta\), and \(\beta\) are estimated parameters. Apendix 1 shows list of explanatory variables used to estimate the land price model.

In this study the dependent variable (land price) was estimated by a set of proxy measures that characterize gridcells and its surrounding neighborhood. The explanatory variables include location effects (distance to nearest highway, distance to arterials, and distance to Han River), neighborhood characteristics (number of residential units, improvement value), and cell characteristics. Data are summarized at the gridcell scale and the value of each gridcell is regressed against the subset of the chosen variables.
Figure 3. Simulation (a) population density (b) household location pattern (c) employment distribution and (d) land price.
As expected, and the simulation result indicates, an increase in commercial
square feet, population density, access to highway, and high-income households is
associated with higher land prices. The model outputs in Figure 3(d) show other
factors of a site and its surrounding neighborhood influenced by non-residential land
value. As many urban economic studies suggest, non-residential land values tended
to decrease with the distance to main transportation networks and urban centers, the
nearest subway stations, arterial roads, and bus stops due to the erosion of accessible
transportation. The effect of distance to the subway station and bus stops on the land
value was higher than that of other variables. In addition, poor road access lowered
non-residential land value and higher employment agglomeration tended to reduce
residential land value.

Increased distance to arterial roads had a similar influence on residential
properties. Further, distance to the Han River had a higher impact on residential
land values. The improved access and synergetic amenities of the urban stream
combined with the Han River helped to boost residential land value around that
region.

Conversely, land values are lower when the gridcell is located within a block
of conservation land, industrial space is present, and travel time to the central
business district is greater. This makes sense because land in a conservation area
is likely to have more restrictions on the type and amount of development that is
allowed.

Household location model
In urban space, changes to existing household location were influenced by various
factors including economic, demographic, and life-cycle changes within a household
and desired quality of life (Clark, Deurloo, & Dieleman, 2006; Kim, Park, & Kewon,
2007; Krizek & Waddell, 2003; McCarthy, 1976). Decision about whether a
household chooses a location as a result of its utility function and the rational
decision of individual households depends on the household’s characteristics its
location attributes (Alonso, 1964; McFadden, 1978). Household rational decisions
often come from their basic needs and restrictions in terms of housing (Clark et al.,
2006) which are mostly related with budget issues (McCarthy, 1976). In classical
urban economic theory, the location preference of the household is determined
mainly on variables such as accessibility, neighborhood quality, housing quality,
land use attributes, and environmental amenities (Fujita, 1989). Accessibility
accounts for transportation facilities including the traveling cost (Anas, 1982),
mode of transport (Simma & Axhausen, 2001; Boehm, Herzog, & Schlottmann,
1991), and other advantages associated with a household’s daily activities, which
generally take place at different locations in the city (Bhat & Guo, 2007).
Neighborhood quality includes a range of variables such as the presence of crime
(Cullen & Levitt, 1999), household income level (Clark, Deurloo & Dieleman, 2006),
quality of housing, quality of life, and race or ethnic composition (Gabriel &
refers to the age of the building, size of the residential units (Rossi, 1955; Clark,
Deurloo & Dieleman, 1997), and prices of individual units. Environmental amenities
include pollutions, landscape design (Caruso, Rounsevell, & Cojocaru, 2005), and
other environmental amenities such as sunshine exposure, views, lakes, river fronts,
and so on (Loechl & Axusen, 2009).
The household location model in this study was estimated by using a set of explanatory variables; these include household income level, distance factors from commercial land and road networks, parks, open space, and urban stream – the Han River. Appendix 1 shows the estimated variables of the household model. In addition, Figure 2(b) provides a graphical presentation of simulated household location patterns in Yongsan-Gu.

In terms of location and transportation variables, in all household location categories, the observed case was far from bus stops and subway stations. Close proximity to open space and the Han River present a relatively favorable location choice for higher income households. Specifically, relative to other areas, the surrounding areas featured higher residential land value, greater population than employment density, and were comparatively small in household size.

In Yongsan-Gu, household simulation results indicate that households prefer to live near households with similar incomes. Higher income households prefer locations that are not close to low income households but show some affinity to being near to mid-income households.

In addition, Ichon dong, Wonhyoro (2) dong, Yongmun dong and Bogwang-dong, locations close to the Han River are the most preferred locations for higher income households. Conversely, cells located close to commercial and industrial space are regarded as less attractive for household location preference. The model outputs demonstrate, Ichon-dong (1) accounts for having the highest population and living households, followed by Bogwang-dong and Huam-dong. While the lowest population and households reside in Hanganro (1), (2), and (3), after Wonhyoro-dong (1).

Moreover, the population and demographic data suggest more than a quarter (29%) of households are single family. The most common household types are three person households, accounting for 46% of total households in year 2005.

Employment location model

The purpose and function of the employment or business location choice model is the same as for the household location choice model. Empirical studies suggest that a large number of factors influence the location preferences of business and employment. In general, the main factors affecting business location preference will depend upon the characteristics of business, physical infrastructure, and also agglomeration of economies (Krugman, 1998). Like household location choice, accessibility continues at the forefront of discussion in the employment location model. The only difference between household and employment location choice models is that the latter allocates jobs using sector-specific preference functions, while households are assigned using variables such as race, age, income, and so on.

In this study, employment location preferences were estimated using a set of explanatory factors including transportation factors, households, and population attributes as well as land use characteristics. The estimated results are tabulated in Appendix 1.

The model estimation and simulation result (see Figure 3(c), Figure 4 and 6) reveals that Hanganro (3) accounts for the highest employment agglomeration site. Hanganro (1), (2), and (3) hold nearly 35% of total employment of the Yongsan-Gu. While nearly one-third of the jobs are in retail, making up 40% of the total. The presence of industrial jobs in those regions is also significant. Among them, 19% of
the total employees were found to work in industrial jobs. In terms of location and transportation variables, in most cases commercial spaces were in close proximity to main roads, subway stations, and bus stops (see Figure 5).

Specifically, relative to other areas, the surrounding areas featured higher land value, greater employment than population density and greater presence of retail business.

The overall job scenarios demonstrate that nearly one-third of the jobs are in retail business, accounting for 28% of the total, followed by official jobs at 20%. Presence of industrial jobs is also remarkable in Yongsan-Gu, as nearly 15% of the

Figure 4. Three-dimensional representation of employment growth in Yongsan-Gu.

Figure 5. Greater number of business location within 200 meters from the main road.
Result validation and discussion

Based on the model calibration introduced in the previous section, evidence suggests that UrbanSim is able to capture dynamic land use patterns at both aggregated and disaggregate levels. All households and employments of the study area were simulated and stored in a database, which provided detailed information for further spatial analysis. In order to test the performance of UrbanSim, the result of simulation should be validated against census data. The model results from the base year were validated against the observed data from 2005.

This study attempts to validate the model outputs at zone level. The outputs were aggregated in each zone and validation was performed against demographic census data. Table 1 shows regression analysis between simulation and census data.

The regression analysis between simulation results and population census data reveals that, apart from a number of over-estimations and underestimations, population growth simulation closely represents the actual census data throughout the region.

Conclusions

The purposes of this study is to apply the microsimulation land use model to Korean cities, and understand how the micorsimulation model works and what is required for model calibration on different land use patterns as well as model performance evaluation. The model’s outputs can be summarized as follows: gridcells within 150 meters of the main road are more likely to have commercial use. Retail business is common, but the presence of industrial jobs is also noticeable in the study area. The
household model demonstrates high-rise residential apartments are found along the bank of the Han River and an increasing distance from commercial space. Higher income households tend to have more cars, associated with apartment buildings, and maintain relatively small-size households. Lower and mid income households are found around the gridcells which are classified as mixed use and not close to high-rise residential apartments, although affluent households were found close to middle income neighbors. Non-residential land values tended to decrease with the distance to main transportation networks and business centers, nearest subway stations, arterial roads and bus stops. Further, distance to the Han River had a higher impact on the residential land values.

Since, the study area was small in size, most attention was given on the application side, with limited work on data analysis and validation. Moreover, real estate development models and transitional models remain inactive. In addition, effects from regional accessibility were in exclusion of result analysis. This study is a first attempt to employ a micosimulation urban land use model in Korea. Clearly much more work has to be carried out in order to see full application. Despite these concerns, this experience has illustrated that our employed methodology is able to simulate land use elements at a micro-level, and further outcomes of this study will likely be used in some capacity by planners in the region as a tool for planning. In coming months, the study will be carried out through the whole Seoul region and explore new directions, including development of new environmental modules, integration of activity based traffic microsimulation models and expansion to various alternatives scenarios addressing contemporary planning issues. To assess the consistency of predicted model outputs to actual conditions, a robust validation process will be carried out in the future.

Acknowledgement

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF- 2009-413-D00001).

References


Appendix 1. Descriptive statistics for household, employment and land price model.

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Household location model</th>
<th>Employment location model</th>
<th>Land price model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>t-stats</td>
<td>Coef.</td>
</tr>
<tr>
<td>Constant</td>
<td>-13.827799</td>
<td>-3.19836</td>
<td>94.60523704</td>
</tr>
<tr>
<td>Distance to open space</td>
<td>0.007250338</td>
<td>0.39698</td>
<td>-0.071686913</td>
</tr>
<tr>
<td>Distance to subway</td>
<td>0.071686913</td>
<td>2.50743</td>
<td>5.778</td>
</tr>
<tr>
<td>Distance to Han River</td>
<td>7.772</td>
<td>7.395</td>
<td></td>
</tr>
<tr>
<td>Distance to main road</td>
<td>7.772</td>
<td>7.395</td>
<td></td>
</tr>
<tr>
<td>Log of the number of residential units in the gridcell</td>
<td>18.79135465</td>
<td>5.124461</td>
<td>0.159671</td>
</tr>
<tr>
<td>Log of the number of residential units within walking distance</td>
<td>-0.204343565</td>
<td>-4.35892</td>
<td>0.00250169</td>
</tr>
<tr>
<td>% high income households within walking distance</td>
<td>0.00250169</td>
<td>5.39457</td>
<td></td>
</tr>
<tr>
<td>Log of total improvement value</td>
<td>0.0990791</td>
<td>53.1253</td>
<td></td>
</tr>
<tr>
<td>% developed within walking distance</td>
<td>0.00951135</td>
<td>19.5447</td>
<td></td>
</tr>
<tr>
<td>% mid income household</td>
<td>0.393438224</td>
<td>6.82792</td>
<td></td>
</tr>
<tr>
<td>Distance to nearest road</td>
<td>0.010273922</td>
<td>1.63376</td>
<td>-0.091702829</td>
</tr>
<tr>
<td>Distance to nearest bus stop</td>
<td>0.000613238</td>
<td>0.07724</td>
<td>0.00281282</td>
</tr>
<tr>
<td>Total car per gridcell</td>
<td>0.790805051</td>
<td>102.283</td>
<td></td>
</tr>
<tr>
<td>Log of total non residential SQFT</td>
<td>1.265353008</td>
<td>3.33173</td>
<td></td>
</tr>
<tr>
<td>% low income household</td>
<td>2.653327493</td>
<td>5.06107</td>
<td></td>
</tr>
<tr>
<td>Population density</td>
<td>-0.397538925</td>
<td>-2.85376</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.9340</td>
<td>0.52485</td>
<td>0.6558</td>
</tr>
</tbody>
</table>