

# **An UrbanSim Model of Brussels within a Short Timeline**

**Zachary Patterson**

**Michel Bierlaire**

**Transport and Mobility Laboratory, EPFL**

**Conference paper STRC 2007**

**STRC**

**7<sup>th</sup> Swiss Transport Research Conference**

Monte Verità / Ascona, September 12. – 14. 2007

## An UrbanSim Model of Brussels within a Short Timeline

Zachary Patterson  
Transport and Mobility Laboratory, EPFL  
Lausanne

+41 (0)21 693 93 29  
+41 (0)21 693 80 60  
zachary.patterson@epfl.ch  
September 2007

Michel Bierlaire  
Transport and Mobility Laboratory, EPFL  
Lausanne

+41 (0)21 693 25 37  
+41 (0)21 693 80 60  
michel.bierlaire@epfl.ch

### Abstract

UrbanSim is an increasingly popular integrated transportation land-use model. A particularly attractive feature of UrbanSim is its disaggregate frame of analysis. This characteristic means it requires much land-use and socio-economic data at a fine scale. This research was conceived as a feasibility study. In particular, it asked 'Is it possible to mount an UrbanSim model with relatively aggregate data (from a TRANUS application) and limited human resources in three to four months?' The conclusions are as follows. It was possible to mount an UrbanSim model with the limited resources and data available. The UrbanSim sub-models estimated were not very robust, compromising simulation results. Weaknesses of the sub-models were not due to the use of disaggregate data. The weaknesses were due to insufficient exploitation of existing data and the lack of some critical data. It was not possible to mount a travel model AND an UrbanSim model given the time and resource constraints.

### Keywords

UrbanSim, integrated transportation and land-use modeling

# 1 Introduction

The interrelationship between urban form, transportation infrastructure and transportation demand has long been recognized. The development of Integrated Transportation and Land-use Models (ITLUMs or integrated models) to better understand and quantify these relationships has received the greatest attention since the 1980s. As a result, several integrated models are available for application.

UrbanSim is an increasingly popular alternative for integrated land-use modeling. Some of the advantages of UrbanSim as an integrated model are: it is open source and thus freely available; its disequilibrium, dynamic modeling framework; its extremely disaggregate approach. This last characteristic allows for a rich analysis. It also presents significant challenges since data preparation can take up to two years. Recently, an application of the TRANUS integrated model was undertaken for the city of Brussels. As a result, a fair bit of land-use (albeit much more aggregate than required by UrbanSim) and transportation data already exist for this region.

Given the demanding data requirements of UrbanSim, the purpose of this research was to test the feasibility of developing and applying UrbanSim using TRANUS data in the context of a short timeline with limited resources. In particular, the aim was to see whether a master's student in 3-4 months could: run an application of UrbanSim using data from TRANUS; and if UrbanSim could perform reasonably well in terms of its land-use predictions in this context. The project was undertaken in collaboration with Stratec in Brussels.

The paper begins with some background on integrated land-use modeling and continues with a description of UrbanSim. This is followed by a description of data preparation and the development of the UrbanSim model of Brussels. The relevant sub-models are presented as well as simulation results based on these sub-models. An evaluation of the results are presented and next steps provided. The paper finishes with some concluding remarks.

## 2 Traditional Transportation Demand Modeling

Travel demand modeling has been, and continues to be, dominated by the so-called 'Four-step Model' (FSM). The literature on the FSM is vast. For good descriptions, see for example Ortuzar and Willumsen (2001) or McNally (2000b). The FSM methodology was initially developed in the 1950s for the now famous *Chicago Area Transportation Study*. Federal US legislation requiring 'continuous, comprehensive, and cooperative' urban transportation planning made it a *de facto* standard in the 1960s, a position it has maintained since (McNally 2000). The FSM responded to the demands of rapidly growing cities with increasing rates of automobile ownership. It represented an intuitive, tractable and effective method of predicting future traffic conditions based on demand projections and infrastructure investments. The persistence of the FSM is likely due to its logical appeal, institutional inertia and the ready availability of FSM software (Bates, 2000).

The four 'steps' of the method are: trip generation, trip distribution, mode split and traffic assignment. In the first step, the aim is to predict the number of trips entering and leaving each traffic analysis zone (TAZ). This is done in any number of ways: regression analysis; cross-classification; or through trip-rate models. Initial data on which this analysis is performed is vehicle count and/or survey data, as well as socio-economic and land-use characteristics of the

the TAZs.

The objective of trip distribution is to estimate the most likely origin-destination or flow matrix of trips across a system of zones. The first two steps can be obviated if data from a comprehensive origin-destination survey exist. When trip generation data are used (e.g. total productions or attractions by zone) there is a need to estimate the number of trips between each of the zone pairs. The most common model for doing this is the gravity model. Gravity models use an entropy maximizing approach to distribute trips. Trips are distributed as a function of: origins in the origin zone, destinations in the destination zone, and travel impedance (i.e. time or generalized costs) between the TAZs.

The third step is mode split estimation. This is normally done using discrete choice models such as the multinomial logit or multinomial probit. These models estimate the probability that a given mode will be chosen for a particular trip, given characteristics of travelers, the modes and the trip.

The final stage of the analysis is trip assignment. This is the stage where trips between the different zones are loaded onto the transportation network. Trips are most commonly loaded onto the network under the assumption of user equilibrium (i.e. no user of the system can unilaterally change routes and improve his/her travel time), so that all paths used for a given O-D pair have equal impedances.

Another stream to develop in transportation demand modeling is that known as Activity-based Modeling (see McNally, 2000a). Instead of focusing on trips as the unit of analysis, the activity-based approach focuses on the behavioural underpinnings affecting demand for trips. This approach can overcome a number of limitations of the FSM. Among other things it overcomes the FSM limitation that assumes overall travel demand to be fixed and independent of the transportation system. It is also much more intellectually appealing since it looks at the drivers of transportation demand as opposed to just the results (i.e. trips).

While different in approach, both take as a starting point the spatial distribution of population and employment (land-use). That is, land-use is considered an important determinant of transportation demand and transportation system performance. However, both ignore the fact that land-use patterns are partially determined by the performance of the transportation system. In other words, they ignore the interaction between land-use and the transportation system.

It is this weakness that integrated models try to address. They seek to do this by explicitly modeling not only how land-use affects the transportation system, but also processes of urban development and how these are affected by the transportation system.

This weakness of the traditional approach has been recognized for almost as long as the FSM has been used. In response, the first integrated models were built in the 1960s. Due to computational, methodological and data constraints the first generation models were not particularly successful. In the following decades, enabled by the computer revolution, GIS, new modeling methods and richer data, development of more sophisticated models advanced (Miller, 2003). Currently, there are many integrated models that are in use or under development around the world.

A detailed description of the different integrated models is outside of the scope of this paper, although there are a number of good overviews (see for example Hunt et al., 2005, Kanaroglou and Scott, 2002 and Southworth, 1995). The focus of the present paper is the integrated model UrbanSim.

### 3 UrbanSim and Research Motivation

UrbanSim has been under development since the late 1990s by Paul Waddell at the University of Washington. More formal documentation of the model can be found in Waddell et al. (2007), Waddell (2002) and Waddell et al. (2003). Since UrbanSim is constantly evolving, most of the description here comes from the reference manual for *The UrbanSim Project* (2006). Some of the description also comes from experience and close monitoring of the UrbanSim mailing list. Ample information and documentation on UrbanSim can also be found at [www.urbansim.org](http://www.urbansim.org).

Since its introduction, UrbanSim has attracted a fair bit of attention and is being applied in more and more locations in the US, Europe and the Middle East. Three features of UrbanSim in particular set it apart from other integrated models. It is these features that elicited our interest. First, UrbanSim is Open Source under the GNU General Public License. This means that anyone can freely access the code, modify and redistribute it. This is appealing since some land use models can prove extremely expensive. Being able to access the code directly can also help in understanding how the model functions. It also facilitates tailoring the software to specific applications.

Second, UrbanSim uses a dynamic disequilibrium framework. Most other models take a system equilibrium approach. That is, they assume that individual markets within the urban system have reached equilibrium (e.g. the real estate market). While this is a useful assumption for mathematical tractability, it is a strong assumption that is quite likely wrong - it is doubtful that an urban system can ever really be considered to be in equilibrium. UrbanSim does not assume equilibrium. It can be considered as 'equilibrium chasing' in the sense that the system *tends* towards equilibrium from one period to another, but is never assumed to actually reach it. Another aspect of 'dynamic disequilibrium' is the time step that is used in future predictions. Whereas many models consider equilibrium states separated by several years, UrbanSim conducts its simulations on an annual basis.

The third distinguishing feature of UrbanSim is its disaggregate approach. There are two elements to this. First, much of the model implementation works at the level of individual households and jobs. Household and employment behaviour is determined by the application of discrete choice models. Second, UrbanSim conducts its analysis at an extremely fine level of geographical detail. Traditionally this has been at the 'gridcell' level of 150 by 150 metres. This means that the model generally works on thousands, or hundreds of thousands of geographical units as opposed to more aggregated models using hundreds of zones. Of course, the fact that UrbanSim operates at such a fine level of detail also means that it requires a great deal of data. While allowing for a rich analysis this can present significant challenges to model implementation, with data collection and preparation generally taking two years.

This characteristic relates to the motivation for this work, as well as to the selection of the case study region (Brussels, Belgium). The broader motivation of the research was to evaluate the difficulty of mounting an UrbanSim model. In order to gauge the difficulty, limited resources (one supervised master's student for 4 months) were available for model implementation. The research team also had data on Brussels that was used for another integrated model, TRANUS. TRANUS works with much more aggregated zones than UrbanSim (152 zone for the Brussels region). As such, a secondary motivation was to see whether it would be possible to use the transportation, land-use, real estate, economic and demographic data from TRANUS in Urban-

Sim and to obtain reasonable results.

## 4 Functional Description of UrbanSim

UrbanSim is best thought of as a land-use model that is ‘coupled’ to a transportation model. That is, the transportation model is not ‘part’ of UrbanSim. Land-use predictions from UrbanSim are input into external travel models and travel conditions from the transportation model are input for use in UrbanSim for subsequent simulations. Any traditional transport model can be used.

The primary (baseyear) data required for UrbanSim relates to land-use, demography and transportation. In UrbanSim a city is divided into ‘gridcells’. Each gridcell is associated with different characteristics including location, current development, surface area and real-estate value by development type, zoning, environmental characteristics (slope, location within aquatic buffers, etc.) and its proximity to roads. Transportation performance information comes from an overlay with the traffic analysis zones (TAZs) of the external transport model. Transportation performance information includes distance or travel time and composite utilities (generally a logsum) between zones.

Households and jobs are associated with particular gridcells as well as particular buildings. Households are described according to socio-demographic characteristics. Jobs are distinguished by their economic sector and the type of building in which they are found. For simulations of future years, the model needs the following exogenous data - control totals for population and employment, and transportation performance data from the transportation model.

In reality, UrbanSim is not really ‘a’ model, but a modeling system. The modeling system is composed of a series of submodels. The UrbanSim sub-models are the: economic, demographic, and development project transition models; household and employment mobility models; accessibility model; household and employment choice models; real estate development model; and land-price model.

### 4.1 Economic Transition Model

Jobs are classified into industrial sectors by the user. Data on aggregate employment by sector are provided exogenously. The economic transition model then determines whether there has been employment growth or decline in the given sectors. In sectors where there has been employment growth, jobs for that sector are placed in a queue to be assigned locations. In sectors where there is decline, jobs are randomly removed and the space that they occupied is added to the pool of vacant space to which jobs can be added by the employment location choice model.

### 4.2 Demographic Transition Model

Households are classified by type. The demographic transition model works in a similar fashion to the Economic Transition Model. Control totals of the population and households by type (if available) are provided to the model. By comparing the control totals to the current population and number of households, the model determines whether there have been increases or

decreases in the number of households of each type. New households are added to a list of households to be located by the Household Location Choice Model. A decline in the number of households results in households being removed and the dwelling they occupied becomes available to other households to be placed by the household location choice model.

### 4.3 Development Project Transition Model

This model creates development projects in order to match the desired (user-defined) vacancy rate. If actual rates are lower than target rates, development projects are created. The characteristics of the new projects are based on historical development events. Once created, the development projects are put in a queue to be placed.

### 4.4 Employment and Household Relocation Models

The Employment Relocation Model predicts the probability that jobs from each sector will move from their current location or stay during the following year. This is intended to reflect the fact that a certain number of jobs will change location from year to year due to different factors such as employee turnover, layoffs, business relocations, etc. The probability of a job moving is a function of user defined rates of job relocation and is proportional to the spatial distribution of jobs in the sector. Once a job is selected for moving, it is removed from its current location and added to the same queue as new jobs to be placed by the Employment Location Choice Model.

Analogously, the Household Relocation Model predicts the probability that households of each type will move from their current residence to another. The probability of moving is user defined and allows for different mobility rates for different types of households. Once households are selected for moving, they are placed in the same queue as new households from the Demographic Transition Model. These households are then placed by the Household Location Choice Model.

### 4.5 The Accessibility Model

The accessibility model calculates the distribution of opportunities weighted by the travel impedance or utility of travel. The utility of travel is measured as the composite utility across all modes of travel between each origin destination pair (the logsum from the transport model). If composite utility is used, the access measure for each location can be written as:

$$A_i = \sum_{j=1}^J D_j e^{L_{ij}} \quad (1)$$

Where  $D_j$  is the quantity of activity in location  $j$  and  $L_{ij}$  is the logsum for vehicle ownership level  $\alpha$  households, from location  $i$  to  $j$ . Many accessibility measures can be calculated based on this formulation and they are used as explanatory variables in several of the UrbanSim submodels, including the: location choice models, developer model and land price model. An

example of such an accessibility measure is the accessibility to jobs of a particular gridcell. It would measure the distribution of jobs weighted by the utility of travel accessible to the gridcell.

#### **4.6 Employment and Household Location Choice Models**

Once the list of new jobs that need to be located is determined by the Economic Transition Model and the Employment Relocation Model, UrbanSim needs to place the jobs. This is done by the Employment Location Choice Model. This model predicts the probability that a new job will be located at a particular location. The gridcell is used as the unit of analysis. The number of locations available for a job depends mainly on total non-residential surface area in the cell and the spatial requirements of the jobs (square feet per employee).

The model processes each job in the queue individually. It queries the gridcells for alternative locations to consider. Alternatives are sampled in proportion to the capacity of the built space in the cell for accommodating jobs. A multinomial logit model is used to estimate the probability that the current job would move to each of the alternative job spaces under consideration. Monte Carlo simulation is used to generate a decision about where the job will be placed among the alternatives. Once this decision is made, the job is assigned to the cell and the quantities of vacant land and used space are updated.

The logit model used is calibrated using the location of jobs in the base year dataset. A sample of jobs in each sector is used to estimate the logit model. UrbanSim allows for the use of many different explanatory variables in model estimation. The explanatory variables available for use are inspired from the Urban Economics literature on employment choice. These include real estate characteristics, various measures of accessibility to population and activities as well as measures of accessibility to the transportation network.

The Household Location Choice Model is analogous to the Employment Location Choice Model. It predicts the probability that a household (from the list created by the Demographic Transition and Household Mobility models) will choose to move to a particular gridcell. As before, a multinomial logit model is used to allocate households to locations in a random sampling of alternatives from existing vacant housing.

The model used is calibrated using base year data of household location. UrbanSim allows for the use of many different explanatory variables in the model estimation. The explanatory variables available for use are inspired from the Urban Economics literature. These include housing characteristics, various measures of accessibility and neighbourhood characteristics.

#### **4.7 Real Estate Development Model**

The real estate development model simulates a process where development projects of a specific type choose locations to be built. This model is similar to the employment and household location choice models. The real estate development model depends on a list of development projects that comes from the Development Project Transition Model. Development projects are located with a multinomial logit. It is calibrated using historical development data. The variables used are similar to those of the other models. I.e. real estate characteristics, accessibility characteristics, etc. Different development types (residential, commercial, etc.) are treated separately.

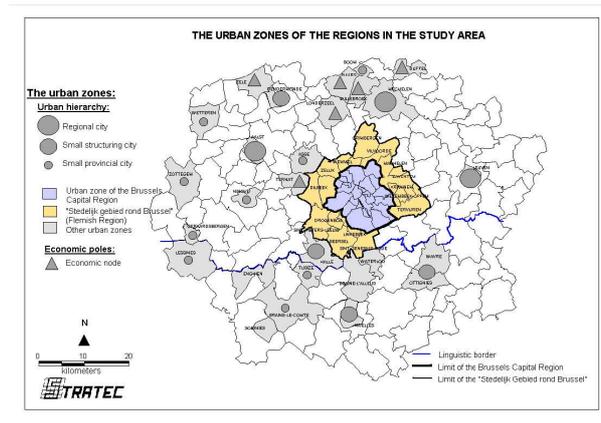


Figure 1: The Greater Brussels Region

## 4.8 Land-price Model

This is the last model to be executed. It estimates land prices after all jobs, households and developments have been placed. These end-of-year prices are then used as the values of reference for each of the sub-models in the subsequent year. The model is based on a hedonic regression. Most of the explanatory variables available for the hedonic land price regressions are similar to those used in the other models. I.e. site characteristics (current land-use) and regional accessibility. One exception is vacancy rate. In theory, lower vacancy rates should result in higher land prices.

## 4.9 Overall Functioning of UrbanSim

Intuitively, UrbanSim can be seen to function in the following way over the course of a simulation year. Exogenous household and employment data are used as an input for the demographic and economic transition models. These models either remove jobs and households or create new jobs and households to be located later by the location models. Based on land-use, accessibility data, households and jobs are assigned locations. Based on the vacancy rate, the Development Project Transition Model creates a list of development projects to be placed. The location for these developments is then chosen by the Real Estate Development Model. Finally, the Land Price Model is executed to estimate updated land-values that will be used in subsequent simulation years.

This concludes the description of UrbanSim and its functioning. The following section goes into greater detail about the case study region of Brussels in Belgium.

# 5 The UrbanSim Model of the Greater Brussels Region

## 5.1 Data for the Greater Brussels Region

As mentioned above, the research team had data previously used for an application of the integrated model TRANUS for the Greater Brussels Region. Brussels is the capital of Belgium and

home to many international organizations. It is perhaps best known for its importance relative to the European Union. Among other EU institutions, the European Commission and the Council of the European Union are located in Brussels.

The TRANUS data covered an area of roughly 4,300 km<sup>2</sup> centred around the city of Brussels (refer to Fig. 1). The study region included 139 townships in parts of Wallonia (French-speaking area to the south of the region) as well as the Flemish Region (Flemish-speaking area to the north). The region also incorporated a number of other important cities with Mechelen, Aalst and Leuven being the largest. As such, the study region represented roughly 15% of the entire country of Belgium. Figures 2 and 3 show the distribution of population and jobs across the region.

The highest concentrations of population (represented as households) and jobs are found in and around the Brussels Capital Region. There is also significant job and household density near and around the larger population centres of the Greater Brussels region. Circles, generally with higher densities, are found inside the larger townships. These circles represent the central area of these townships. With the inclusion of these township centres there were 152 zones all together in the original TRANUS dataset.

Household and jobs location data were available for the years 1991 and 2015. Land value (price per m<sup>2</sup>) and transportation data were also available for the analysis. Transportation data included impedance matrices and other transport network data. With respect to impedance matrices, both logsum origin-destination matrices and travel-time matrices were available. The team also had GIS layers that included the location main arterials and highways.

## 5.2 The Research Project

As discussed above, the purpose of the project was to gauge two things. First was the difficulty of mounting an UrbanSim model. Second was to see whether relatively aggregate data from TRANUS could be used to produce sensible results from UrbanSim.

In order to evaluate these two different elements, the project was given to a master's student. The student's task was to, with significant guidance, use the data available to calibrate and

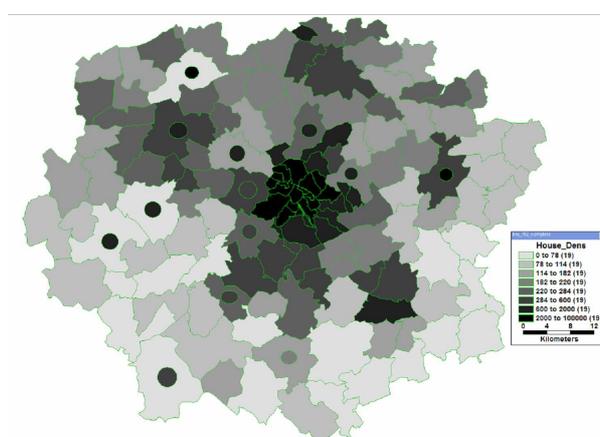


Figure 2: Distribution of Households in the Greater Brussels Region in 2001

mount an UrbanSim model of the Brussels region. The student had four months in which to complete the work. A sample UrbanSim database for the city of Eugene in Oregon is available for download. This database was used as a reference for the construction of the database developed for Brussels. UrbanSim version 4.0 of 21 November 2006 was used in the analysis.

## 6 UrbanSim Data Preparation

Since UrbanSim works primarily at the gridcell level, the first step in the process was to create a layer of gridcells. A 'standard' grid of 150m by 150m was applied to the study region. This resulted in just under 200,000 gridcells covering the study region.

As mentioned in Section 5.1, land-use, demographic and transportation data were available for the 152 TRANUS zones. Since UrbanSim requires land-use and demographic data at a much finer level of detail, it was necessary to disaggregate this data. Moreover, the land-use data available was not sufficiently rich for use with UrbanSim (e.g. it did not contain information on buildings). As a result, other adjustments to the data were required. A description of land-use and demographic data preparation follows. With respect to transport data, disaggregation was not required, but the transport data presented its own challenges described in the following section.

### 6.1 Transportation Data

As mentioned in Section 4, UrbanSim is normally 'coupled' to a transportation model. That is, land-use data from UrbanSim are input into an external travel model and travel conditions from the transportation model are input for use in UrbanSim for subsequent simulations. Any traditional transport model can be used. EMME/2 is the most commonly used transportation model and currently the best supported as well.

Various possibilities for the transport model were considered early on in the project. In TRANUS, the transport model is so closely coupled to the land-use side that integration with UrbanSim was beyond the scope of the project. Since no stand-alone model existed, the alternative was to

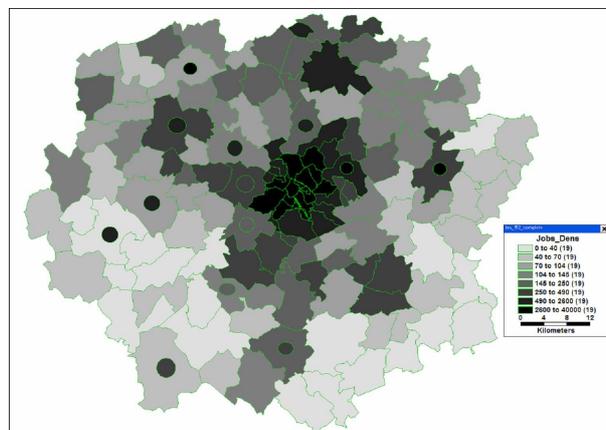


Figure 3: The Distribution of Jobs in the Greater Brussels Region in 2001

mount an easily supported transport model based on the TRANUS data. This was considered the most logical approach. It was clear early in the project that it was outside of what could be accomplished given resource constraints. As a result, intrazonal travel impedance data (travel time and composite utility matrices) from the TRANUS baseyear (2001) was used for all simulation years. This amounts to assuming that travel conditions did not change over the course of the simulations. As such, land-use evolution could be affected by transportation network performance, but not the inverse.

Apart from zonal-level impedance data, UrbanSim requires gridcell-level information on proximity to transport infrastructure - highways and main arterials. Using GIS layers for highways and main arterials, this information was easily obtained for each gridcell.

## **6.2 Household, Employment and Land-use Data**

### **6.2.1 Gridcell Development Type**

At the time of the analysis, no information was available relative to zoning or geographical constraints to development (e.g. slope, water cover, etc.). It appears that in the version of UrbanSim used, it was necessary to have at least seven development types (residential, commercial, mixed, industrial, governmental, undevelopable and vacant). As a result, the gridcells of the zones were randomly assigned to one of the seven development types in equal numbers.

### **6.2.2 Creation of Buildings**

UrbanSim requires the existence of buildings that contain households and jobs. Each of these buildings needs to be characterized by several variables (value, number of units, etc.). At the same time, no information on buildings was available. As such, buildings were created in order to accommodate the households and jobs in each zone. For simplification, and based on the Eugene example, one building was assigned to each gridcell, apart from those gridcells identified as 'vacant' or 'undevelopable.'

General hypotheses about: spatial requirements for households and jobs, building age, value and vacancy rates were formulated with Stratec. Vacancy rates were particularly important for the construction of buildings. For non-residential buildings, a uniform 10% was adopted for the region as a whole. For residential buildings, vacancy rates grew linearly towards the city centre. That residential vacancy rates are higher in the centre than the periphery is a peculiarity of Brussels. Residential vacancy rates were assumed to vary between 2% (extreme periphery) and 10% (city centre). As such, the capacity of buildings was set uniformly across each zone based on the vacancy rate and the number of households and jobs per zone.

### **6.2.3 Creation of Households**

In the TRANUS data, households were classified into seven broad types. There were three 'inactive' groups: inactive, students and retired. There were also four 'active' groups distinguished by income (high or low) and number of persons (single and two or more). UrbanSim is designed for more precise data including income, number of children, etc. As a result, families needed

to be constructed. Based on demographic characteristics of the Brussels region as well as general hypotheses about the other characteristics (e.g. income), synthetic households were created according to what group they belonged to. Each of the characteristics was drawn from a distribution that was different for each of the household types. For example for income, students had incomes of 500-1000 Euros a month. High-revenue families had incomes of 4000-6000 Euros per month.

#### **6.2.4 Location of Households and Jobs**

Households (by type) and jobs (by sector) needed to be assigned to buildings able to house them. Having no more precise information, households and jobs were assigned randomly to appropriate buildings in the zone to which they were identified. That is, households were assigned to 'Residential' buildings, public sector jobs to 'Government' buildings and so on. In addition, households were assigned characteristics based on a random selection from the synthetic families generated before.

### **6.3 Real Estate and Macroeconomic Data**

#### **6.3.1 Land Value**

Land value (in Belgian Francs) was available per m<sup>2</sup> per zone for three different categories of land-use: high-density residential, low-density residential and mixed. The mixed values were used for non-residential gridcells and the two residential land-use categories were used for residential gridcells. Land-value per m<sup>2</sup> for the zone was applied to all of the relevant gridcells in the zone.

#### **6.3.2 Historical Land Development Data**

In order to calibrate the Real Estate Development Model, UrbanSim requires information on historical development events. Although no data on developments existed, there was data on household and employment change from 1991-2001. This data was used to estimate where construction had taken place. This was done by first establishing in which zones there had been a population increase. In those zones, buildings were randomly selected as having been built. The number of buildings equaled the number of buildings required to house the new population to those zones. Non-residential construction was estimated in an analogous way.

#### **6.3.3 Household and Job Relocation Rates**

UrbanSim requires data on the relocation rates of households and jobs. Based on discussions with Stratec, a household relocation rate of 2% was used for all household types. This implies that 2% of households in Brussels can be expected to move every year. A relocation rate of 10% for jobs was used.

### 6.3.4 Macroeconomic Data

Stratec provided population and employment data for the years 2001 and 2015. Population data was by household type and employment data by sector. Constant growth rates were used to estimate population and employment figures for the other simulation years.

## 7 Sub-model Estimation

The following section describes the various models that were estimated and used in the UrbanSim simulations. These include the location choice models, real-estate development models and the land-price model. Standard multinomial logit formulations are used for the location choice and real-estate development models. That is, the utility of a given alternative (indexed  $j$ ) is expressed as  $V_j$ .

$$V_j = \beta' x_j + \epsilon_j \quad \forall_j \quad (2)$$

It is a linear combination of the different characteristics ( $x$ s) of the alternative. In the case of a household location choice, the  $x$ s would be, for example, distance from the CBD, access to employment, etc.  $V_j$  also includes an 'error' term denoted as  $\epsilon_j$ . In the case of the MNL,  $\epsilon_j$  is assumed to be independently and identically extreme value distributed (iid). Due to this assumption, the probability that a particular alternative is chosen is the well-known logit formulation.

$$P_i = \frac{e^{V_i}}{\sum_{j=1}^J e^{V_j}} \quad (3)$$

The land price model is a linear regression estimated using ordinary least squares (OLS). It can be formulated as:

$$Price_j = \beta' x_j + \epsilon_j \quad \forall_j \quad (4)$$

In the case of the land price model, the  $x$ s are similar to those for the location choice models. In the case of OLS,  $\epsilon_j$  is distributed normally with mean zero and variance  $\sigma$  estimated in the regression.

These models are estimated within UrbanSim which includes many different pre-defined explanatory variables. It is also relatively easy to program and use user-defined variables.

The models presented are compared with two different sets of models. The first set comes from the models included with Eugene sample dataset. The second comes from a recent first-in-class UrbanSim application to Salt Lake City, Utah reported in (Waddell et al., 2007 and Waddell, Franklin and Britting, 2003). Naturally, there are tremendous differences between Brussels and these two other cities. As such, the comparison is intended to look at the overall performance of the models and not at their specifics.

Before continuing, it is worth mentioning some important features of the data available for these other applications. The Utah example included fine data on: gridcell development information (22 development types); households (by income (9 types), by size (5 sizes), by age

(4 categories), etc.); and geographical features. The Eugene dataset also had finer information on gridcell development, households and geographic features than Brussels. This is important because the primary difference between the Brussels models and those from Eugene and Salt Lake City has to do with the number of statistically significant variables used in the models. Since the data were much richer, it was possible to include more variables for Eugene and Salt Lake City than for Brussels.

## 7.1 Location Choice Models

For the household and employment location choice models, it is possible to estimate different model formulations for different industry sectors or household types. In order to keep the project within the timeline, the number of models was kept to a minimum. As a result, one Household Location Choice model was estimated for all of the household types. This followed Eugene example. With respect to employment location choice, two models were estimated. One for commercial and the other for industrial jobs. No further model segmentation was undertaken, unlike the Eugene example which included models for subgroups of the commercial and industrial sectors.

The location choice models perform reasonably well. Although some of the coefficient signs seem unintuitive, they compare well with models from the Eugene example. Take for example the Household Location Choice Model in Table 1. All of the variables are highly significant and for the most part meet *a priori* expectations. The odds of choosing a particular location decreases with cost. The odds increase with the proportion of similar household-income types. There is one problematic variable - travel time to CBD. Theoretically, households should prefer being closer to the CBD than farther away because of the many work, shopping and other opportunities found there. This implies a negative coefficient for travel time to CBD. The resulting coefficient, however, is positive.

Brussels is an exception among European cities in that the downtown is perceived as less desirable than surrounding areas. It is characterized by higher vacancy rates and lower incomes. As such, it is perhaps not surprising that this variable could come out positive. At the same time, it does suggest omitted variable bias.

The model compares favourably with the Eugene example. The Eugene example contains eight coefficients and like the Brussels model these include variables on cost and characteristics of the surrounding neighbourhood. Unlike the Brussels model it does not contain any accessibility variables. The Brussels model does not compare very well with the Utah model. It has close to twenty statistically significant variables. The majority of the extra variables in the Utah model concern variables not fully exploited (household characteristics) or unusable (development types) in the Brussels data.

It is possible that the Brussels model could be improved through further segmentation of the data along household-type lines. Inclusion of more accurate development type data could help to improve it as well. That some important variables come out as statistically significant implies that not all important information is lost through disaggregation.

The other location choice models are about as successful as the Household Location Choice Model. Each contains several significant variables. The variables concern primarily accessibility to employment, population and the CBD, with land value also being important. Some of the variables have the wrong sign. For example, land value is positive in both the models. This is

Table 1: Household Location Choice Model

Variable	Coefficient	Std. Error	t-value
Cost to Income Ratio	-0.064	0.025	-2.599
% high income if high income	0.032	0.001	32.800
% low income if low income	0.059	0.001	60.167
% mid income if mid income	0.027	0.001	26.517
Travel Time to CBD	4.106E-04	0.000	2.796
Initial Log-likelihood:	-4.397E+05		
Log-likelihood	-4.371E+05		
Log-likelihood ratio index is :	0.006		
Likelihood Ratio Test:	0.000		
Number of observations :	129269		
Convergence statistic is :	0.000		

likely due to omitted variable bias. It is interesting to note that land value is also positive in the Eugene models. As with the household location choice models, the models could likely be improved by considering different industry sectors separately and better development type data. As with household location, not all is lost using disaggregated data.

## 7.2 Real-estate Development Models

Based on the Eugene example, three different real-estate development models were used. The use of disaggregate data is less critical than the availability of data in the calibration of these models. With around 1,300 observations, the Residential Real-estate Development model performed reasonably well. It had six significant variables whose signs met *a priori* expectations. Table 2 presents this model.

Table 2: Residential Real-estate Model

Variable	Coefficient	Std. Error	T-value
Log commercial sqft w.w.d.	-0.313	0.048	-6.569
Log Access to population	0.247	0.142	1.745
Log industrial sqft w.w.d.	-0.118	0.019	-6.193
Log total population w.w.d.	0.447	0.054	2.765
Travel Time to CBD	0.006	0.002	3.111
Average Income w.w.d.	-0.217	0.045	-4.837
Initial Log-likelihood :	-4530		
Final Log-likelihood :	-4424		
Log-likelihood ratio index is :	0.024		
Likelihood Ratio Test :	0.000		
Number of observations :	1332		
Convergence statistic is :	0.000		

Essentially it shows that residential development is more likely in areas with high access to population with little commercial and industrial development that is away from the CBD. One

odd-seeming result is that development appears to be inversely correlated with average income. While the positive coefficient on travel time to the CBD seems odd, a similar result is used in the Eugene example. Another troublesome result is that there is no measure of land value in this model, something that surely is an important factor in development choice and that has an important effect on model simulations.

This model compares relatively well with the Eugene example that has eight statistically significant variables, but less well with the Utah example. It has around twenty variables for its development models. Again, the extra variables come primarily from data unavailable in the Brussels model.

Calibration of the commercial and industrial development models was less successful. Based on the data from Stratec, there were very few development events that could be used as observations to estimate models. There were only 81 observations of commercial developments and 26 observations of industrial developments. The result was that the two models had only two, and one variables respectively.

The fact that the residential development model was reasonably successful suggests that even disaggregate data can be used successfully in the real estate development models. One caveat is that there needs to be enough data.

### 7.3 Land-price Model

The Land-price Model performs relatively well. It has a strong fit with an Adjusted  $R^2$  of 47% - quite high for a cross-sectional dataset. Eight variables come out statistically significant and they all have the 'right' sign. Table 3 shows the land-price model. Land value increases with accessibility to employment and population and with the proportion of high-income households. It decreases with the presence of basic sector employment and with travel time to the CBD.

Table 3: Land-price Model

Variable	Coefficient	Std. Error	T-value
Log total empl. basic sect. w.w.d.	-0.239	0.002	-110.714
Log Access to empl.	0.770	0.001	660.032
Log residential units	0.116	0.001	113.787
Log total empl. w.w.d.	0.461	0.003	170.515
Log Access to population	0.014	0.002	6.678
% High income w.w.d.	0.001	0.000	5.168
Travel Time to CBD	-0.001	0.000	-11.291
Number of observations :	165780		
R-Squared :	0.473		
Adjusted R-Squared :	0.473		

This model compares relatively well with the Eugene model that contained ten statistically significant variables. It performs much worse than the Utah model. It had more than forty variables and an Adjusted  $R^2$  of 75%. Again, the majority of the extra variables were not available for the Brussels model.

The results of the Land-price Model suggest that using the disaggregate data, some reasonable

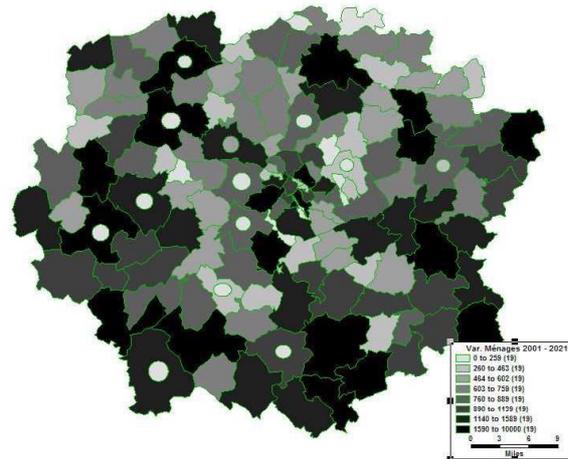


Figure 4: Household Variation by Zone - Population Growth

results can be obtained. Naturally, the Brussels model could be improved by incorporating data on geographical characteristics, development and zoning types.

## 8 Simulation Results

Since UrbanSim was not coupled with a transport model, these simulations assume that performance of the transportation network was fixed over the entire period. The first conclusion to reach is that it was possible within the short timeline and given the data constraints to run UrbanSim simulations. Many different simulations were compared and only a few of them will be shown here. Two figures are used to describe the primary conclusions about the simulations as a whole.

Figure 4 shows the simulation results for household variation by zone from the period 2001-2021. This simulation uses the TRANUS population growth forecasts. It shows the highest changes in number of households is found in outlying areas. If we consider vacancy rates for the year 2021 (see Figure 5) it is obvious that outlying zones have almost no vacancy.

This implies that households are choosing to live outside of the central area of Brussels. It turns out that the Residential Development Model has built the largest number of residences in these outlying areas. Households prefer to live outside of the city and are choosing to live in the new houses that are available to them there, despite very low vacancy rates. That this is the case is shown when comparing these simulations with a no-population-growth simulation. In the no-growth simulation, households choose to live where there are more residences available, towards the centre. I.e. household location choice is driven by residential development location. Similar results are found for commercial and industrial jobs i.e. there are very low vacancy rates everywhere except the city centre.

In reality, it seems unlikely that there would be such low vacancy rates over such large areas. Very low vacancy rates should increase land values and thus help to equilibrate the housing market. However, vacancy rates do not enter the Land-price Model. That this equilibration does not take place has two possible explanations.

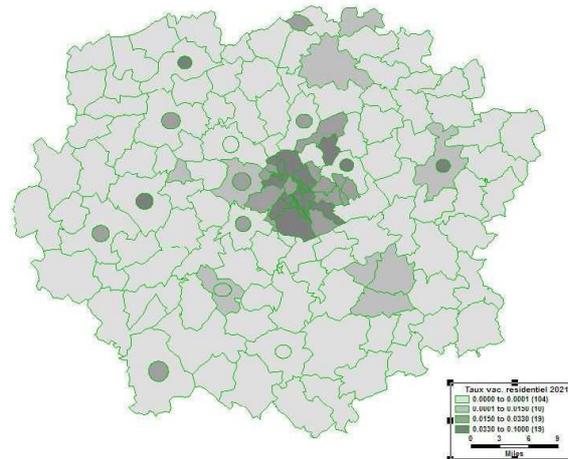


Figure 5: Residential Vacancy Rates in 2021 - Population Growth

Perhaps vacancy rates do not affect land-prices. The land-price model reported in Waddell et al. 2007 does not contain a vacancy rate variable. Neither does the Eugene model. If this is the case, however, then the model seems to produce rather extreme results. The second possibility is that the land-price model should include vacancy rates, but that the vacancy rate data is not sufficiently detailed to come out significant in the model.

Another possibility for which households settle almost exclusively outside the centre is the weakness of the development models. The likelihood of building a residence by the Residential Development model increases with access to population, the residential nature of surroundings, distance from the CBD and does not include land price. Thus, development takes place in desirable locations regardless of price. In reality, we would expect land price to be an important factor. Moreover, prices should reduce the probability of developing in desirable locations in favour of cheaper locations. As such, if the vast majority of residences are built in outlying areas of a residential nature with high accessibility to population, it is not surprising that households choose to live there.

## 9 Discussion

The prime motivation of this research was as a feasibility study. I.e. was it possible to mount an UrbanSim model within a short timeline, with limited resources using aggregate data from a TRANUS model? As such, the discussion is divided into three subsections: integration with a transport model; understanding and mounting UrbanSim; and the sufficiency of data from TRANUS.

### 9.1 Integration with a Transport Model

A complete UrbanSim model requires integration with a transport model. EMME/2 is the transport model which is the most commonly used with UrbanSim. It is also the best supported. It might have been possible to integrate UrbanSim with an EMME/2 model were one available.

None was available so it was not possible given the project resources to integrate UrbanSim with a transport model.

## 9.2 Understanding UrbanSim and Mounting a Model

UrbanSim is not 'plug-and-play' software. Even with the documentation that exists and the active listserve, a fair bit of learning how to use UrbanSim requires 'going in' to the Python code. Python is a relatively accessible programming language making this task easier, although not automatic. Implementing an UrbanSim model also requires familiarity with MySQL and GIS software. That said, it is possible for an UrbanSim novice with programming experience to understand and apply UrbanSim in a relatively short timeline - in this case a few months. Critical to being able to understand UrbanSim is using the Eugene database.

## 9.3 Sufficiency of TRANUS Data

TRANUS data is more aggregate than is required for operation with UrbanSim. It is more aggregate, not only geographically, but also in almost all dimensions. Demographic information is limited and land-use and zoning information is inexistant. In this project, TRANUS data was successfully disaggregated. The use of this data in UrbanSim was, however, only a partial success. There are three reasons for this.

First, given the limited time available for the project, only a hasty analysis of the available data was possible. As a result there was information in the existing data that was not exploited. The household location models could have been improved by exploiting data available about the different household types. The employment location models could have been improved by allowing different coefficient estimates for the different industries. The ability to better use this data can only be determined once a more thorough analysis has been conducted.

The disaggregation of the TRANUS data appears to be less of a challenge than the lack of some important data. The most crucial is data used to calibrate the development models. The small number of development events (particularly for commercial and industrial events) made it difficult to calibrate robust development models. The development of stronger models would need additional data.

Other data that could help to improve the models would include zoning and other geographical data (e.g. development types). The research team now has some GIS data on land development types that could be incorporated into the data set. Also potentially useful would be more accurate vacancy rate data that might help to improve the land-price model and overall UrbanSim performance.

## 9.4 Next Steps

The next steps of this project involve two directions: continued evaluation of using TRANUS data for UrbanSim; and mounting an adequate UrbanSim model of Brussels. Continued evaluation of the use of TRANUS data will require further development of the location choice, land-price and real-estate development models. The purpose is to see whether the information in the data can be better used to develop more robust models.

Successfully mounting an UrbanSim model of Brussels will require a number of things. First, it is necessary to develop a transport model and to integrate it with UrbanSim. The transportation data from TRANUS will shortly be received by the research team and this step will be undertaken.

Second, it would be ideal to have more data. Three types of data are most urgent. More data on historical real-estate development is most important - particularly for commercial and industrial development. This would allow for the estimation of more robust development models. Second, zoning data could help to improve the quality of many models by providing better geographical information for the location of housing and employment. GIS layers of zoning types are now available and will be used in further model development. Finally, improved data on vacancy rates could prove invaluable in improving the UrbanSim model. For the time being, it is uncertain whether improved vacancy rate data will be available.

## 10 Conclusion

UrbanSim is an integrated transportation and land-use model that is increasing in popularity. Its disaggregate nature provides the potential for very rich analysis on the interaction between urban development and the transport system. It is its disaggregate nature that provides the biggest challenges to its use.

The purpose of this research was to test whether more aggregate data from another integrated model (TRANUS) could be used to successfully mount an UrbanSim model with limited resources and a short timeline. As such, a master's student was given three months to use the available data to mount an UrbanSim model.

The main conclusions are as follows. First, it was not possible to integrate UrbanSim with a transport model and the analysis focused on the land-use side. If a transport model in EMME/2 had been available this might have been possible. Second, it was possible to understand UrbanSim and to get it to run simulations with the data available. Third, while possible to run UrbanSim, the underlying models developed were lacking and resulted in simulations that did not inspire enormous confidence. Fourth, the weaknesses of the submodels did not stem from the use of disaggregated data. Rather, it stemmed from the insufficient exploitation of existing data and the lack of particularly critical data. The lacking critical data included historical development information as well as zoning, development type and vacancy rate data.

## References

- Bates, J. (2000). *History of Demand Modeling*, number 2 in *Handbooks in Transport*, Pergamon, New York, pp. 11–33.
- Hunt, J., Kriger, D. and Miller, E. (2005). Current operational urban land-use transport modelling frameworks: A review, *Transport Reviews* **25**(3): 329–376.
- Kanaroglou, P. and Scott, D. (2002). *Integrated Urban Transportation and Land-use Models for Policy Analysis*, Ashgate Publishing Limited, Aldershot, Hampshire, UK, chapter 3, pp. 42–72.
- McNally, G. (2000a). *The Activity-based Approach*, Handbook of Transport Modelling, Pergamon, chapter 4, pp. 53–69.
- McNally, G. (2000b). *The Four-step Model*, Handbook of Transport Modelling, Pergamon, chapter 3, pp. 35–52.
- Miller, E. (2003). *Land Use: Transportation Modeling*, CRC Press, Boca Raton, FLA, chapter 5, pp. 5–1 – 5–24.
- Ortuzar, J. and Willumsen, L. (2001). *Modelling Transport*, third edn, John Wiley, Chichester.
- Southworth, F. (1995). A technical review of urban land use - transportation models as tools for evaluating vehicle travel reduction strategies, *Technical Report ORNL-6881*, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- The UrbanSim Project (2006). Opus: The open platform for urban simulation and urbansim version 4 - reference manual and users guide, *Technical report*, Center for Urban Simulation and Policy Analysis, University of Washington, Seattle, Washington.
- Waddell, P. (2002). Urbansim: Modeling urban development for land use, transportation, and environmental planning, *Journal of the American Planning Association* **68**(3): 297–314.
- Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M. and Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of urbansim, *Networks and Spatial Economics* **3**(1): 43–67.
- Waddell, P., Franklin, J. and Britting, J. (2003). Urbansim: Development, application and integration with the wasatch front regional travel model, *Technical Report CUSPA-03-01*, Center for Urban Simulation and Policy Analysis, University of Washington, Seattle.
- Waddell, P., Ulfarsson, G. F., Franklin, J. P. and Lobb, J. (2007). Incorporating land use in metropolitan transportation planning, *Transportation Research Part A* (41): 382–410.