A Prototype Integrated Transportation Land-use Model for the Lausanne Region

Zachary Patterson * Ricardo Hurtubia *

29 August 2008

Report TRANSP-OR 080829-2
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*École Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory, Station 18, CH-1015 Lausanne, Switzerland. E-mail: {zachary.patterson, ricardo.hurtubia}@epfl.ch

Contents

1 Introduction **3**

2 Transportation and Land-Use Modeling **4**

3 How UrbanSim Works **4**
3.1 Economic Transition Model 6
3.2 Demographic Transition Model 6
3.3 Development Project Transition Model 6
3.4 Employment and Household Relocation Models 6
3.5 The Accessibility Model 7
3.6 Employment and Household Location Choice Models 7
3.7 Real Estate Development Model 8
3.8 Land-price Model 8
3.9 Overall Functioning of UrbanSim 8

4 Required and available data **9**
4.1 UrbanSim data requirements 9
4.2 Data available for Lausanne 10

5 Data Preparation **11**
5.1 Constructed and used data 11
5.1.1 Gridcells Data 11
5.1.2 Households Data 12
5.1.3 Jobs Data 12
5.1.4 Buildings Data 12
5.1.5 Development Event History Data 13
5.1.6 Development Constraints 14
5.1.7 Other Data 14
5.2 The Lausanne-Morges Region in 2000 14

6 UrbanSim Submodels Estimated for the Lausanne-Morges Region **17**
6.1 Household Location Choice Model 18
6.2 Employment Location Choice Model 19
6.3 Real Estate Development Models 23
6.4 Land Price Model 26

7 Simulation Results **26**
7.1 Base Case Scenario (model validation) 30
7.2 Transport Scenario 32
7.2.1 Land-Use Scenario 34

8 Evaluating the Lausanne UrbanSim Prototype 36
8.1 Feasibility 37
8.2 Evaluation of Performance 38
8.3 Further Work Required for an Operational Model 39

9 Conclusion 41

List of Tables

1 Household Location Choice Model 19
2 Industrial Employment Location Choice Model 20
3 Commercial Location Choice Model (Retail) 21
4 Commercial Location Choice Model (Hotels) 22
5 Commercial Location Choice Model (Financial Services) 22
6 Commercial Location Choice Model (Real Estate) 23
7 Commercial Location Choice Model (Public Services) 24
8 Commercial Development Location Choice Model 25
9 Industrial Development Location Choice Model 25
10 Residential Development Location Choice Model 26
11 Land Price Model 27
12 Simulation scenarios 27
13 Developable and undevelopable plan types for the two zoning variants 29

List of Figures

1 The “Lausanne-Morges” Region 15
2 Location of population by hectare in the Lausanne-Morges region (2000) 16
3 Location of jobs by hectare in the Lausanne-Morges region (2000) 16
4 Accessibility in the Lausanne-Morges region (2000) 17
5 UrbanSim integration with the EMME travel model for Lausanne 28
6 Difference in household growth rate by commune (2000 -2007) . . 30
7 Household location increase between 2000 and 2020 31
8 Job location increase between 2000 and 2020 32
9 Access to population difference between base and transport scenario 33
10 Difference in household location (base scenario vs transport scenario) 34
11 Difference in job location (base scenario vs transport scenario) . 35
12 Difference in high income household location (base scenario vs transport scenario) 36
13 Location of households in agricultural and intermediary zones . . 37
14 Location of jobs in agricultural and intermediary zones 38

1 Introduction

Acknowledgment of the relation between transport and land-use is fundamental to fully understanding the complex dynamics of cities. Land use determines travel demand while, at the same time, characteristics of the transportation system affect the location choice of households and firms; the traditional approach to transport modeling does not consider these relationships. This has motivated the development of several Integrated Transportation and Land-use Models (integrated models), with the aim of better understanding and quantifying the effects of land use on transportation and vice-versa.

UrbanSim is an increasingly popular alternative for integrated land-use modeling. The primary advantages of UrbanSim as an integrated model are that it is open source and thus freely available and its disaggregate approach. This last characteristic allows for a rich analysis, but also presents significant challenges since data preparation can take up to two years.

This technical report serves as a description of the prototype UrbanSim model developed for the region of Lausanne-Morges. The project was undertaken with funding from the Matching Funds initiative of the Institute of Territorial Development (INTER) of the EPFL. The purpose of the project was fourfold. First was to test the feasibility of developing a prototype UrbanSim model for the region, and to identify the necessary effort (both in time and resources) to achieve it, given the use of available data and a pre-existing transport model for Lausanne (EMME). This work builds on previous research (Patterson and Bierlaire, 2007 and 2008) that developed a prototype UrbanSim model for the region of Brussels in Belgium with pre-existing data used previously for an aggregate integrated model (TRANUS). The second purpose was to identify model weaknesses and to evaluate the additional data and effort that would be required to develop a fully operational UrbanSim model for the region. Third was to forge and foster interdisciplinary and interlaboratory links at the EPFL by using the expertise and experience of Transport and Mobility Laboratory (TRANSP-OR), CHOROS and the Laboratoire de systèmes d'information géographique (LaSIG). Finally, an important purpose of the research was to demonstrate to public authorities how such a model could be used in Lausanne and to encourage them to consider the possibility of adopting such a model for planning purposes in the Canton of Vaud.

The report begins with a short introduction to transportation and land-use modeling and continues with a detailed description of the operation of UrbanSim. A third section provides a more detailed explanation on the data requirements of UrbanSim models and data sources available for the present application. Section Four describes how the UrbanSim model was implemented. This includes information on how available data were used to produce the UrbanSim baseyear database. Section Five and Six describe the main features of the baseyear data and the location choice and land-price models that were estimated with it. Section Seven reports on various UrbanSim simulations until the year 2020. The purpose of this section is show the results that were obtained, but also to demonstrate how UrbanSim can be used to model and predict future urban development and transportation system performance. The final section provides conclusions about the experience with the Lausanne prototype and describes what would need to be done to develop a fully operational model that could be used for actual planning purposes.

2 Transportation and Land-Use Modeling

The traditional approach to transport system modeling separates the problem in four stages: trip generation, trip distribution, mode split and traffic assignment. This is known as the 'Four-step Model' (Ortuzar and Willumsen, 2001).

The first step focuses on predicting aggregated travel demand, such as the number of trips generated and attracted in every traffic analysis zone. The second step distributes those trips to destinations (for generated trips) and origins (for attracted trips), this produces an origin-destination matrix. The third step estimates the probability that a given mode will be chosen for a particular trip, given the characteristics of the travelers and the trip. Finally the fourth step consists of loading the trips onto the transportation network.

Usually the first step takes into account land-use characteristics at an observed period of time and predicts the number of trips using regression analysis, cross-classification or trip-rate models. This approach ignores the fact that land-use patterns are affected by the performance of the transportation system. For example: if a metro line is built in a city, one should expect changes in travel patterns (mostly on mode split), but also (in a mid or long-term analysis) the existence of the metro line should affect its surroundings, making some areas more or less attractive to households and firms. This has an effect on the number and distribution of located agents and, in consequence, affects travel demand (specifically trip generation). This complex, recursive, interaction between land-use and the transport system is what integrated models try to address.

Most Integrated Transport and Land-use Models can be understood as adding a fifth step to the classic Four-step Model which estimates the location of agents (households and firms), given exogenous demographic and economic data, and travel conditions determined in the other four steps. These location patterns are then used to determine travel demand and travel conditions, which are used as input for land-use estimation, and so on. The specific methodology used to model the interaction between the land-use and transportation system depends on the model being used. For a good review of different integrated models see Hunt et al., 2005.

3 How UrbanSim Works

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). Ample information and documentation on UrbanSim can also be found at `www.urbansim.org`.

Since its introduction, UrbanSim has attracted a fair bit of attention and is now being applied in several locations in the US, Europe and the Middle East. Three features of UrbanSim in particular set it apart from other integrated models. 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 very 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. This means that a required assumption is 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* toward 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 behavior is determined by the application of discrete choice models. Second, UrbanSim conducts its analysis at a fine level of geographical detail. Traditionally this has been at the "gridcell" level of 150 by 150 meters. This means that the model works on thousands, or hundreds of thousands of geographical units as opposed to more aggregate 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, especially for data collection and preparation.

UrbanSim is not exactly an integrated model but a land-use model that works together with a transportation model. This means that 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 transportation model can be used for this.

UrbanSim divides the city into "gridcells". Each gridcell is associated with different attributes such as location, current development (buildings), land value, physical characteristics (surface area, slope, proximity to roads) and transportation performance (obtained from the associated transportation model). Households and jobs are associated with particular gridcells and buildings. Households are described by socio-demographic characteristics (income, size, car ownership). Jobs are categorized by economic sector.

To model agent decisions and development in the land-use system, Urban-

Sim uses a series of sub-models, which are described next. Newer versions of UrbanSim attempt to include firms as agents in their modeling frame while, at the same time, they use the much more disaggregated spatial division of parcels instead of gridcells. However, all models described here correspond to the "traditional" version of UrbanSim which only considers households and jobs as "location choosers" and uses the gridcell as spatial division.

3.1 Economic Transition Model

Jobs are classified into industrial sectors. 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.

3.2 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.

3.3 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.

3.4 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.

3.5 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:

$$A_i = \sum_{j=1}^{J} D_j e^{L_{\alpha ij}} \tag{1}$$

Where D_j is the quantity of activity in location j and L_{ij} is the logsum for vehicle ownership level α households, from location i to j. Many accessibility measures can be calculated based on this formulation and they are available as explanatory variables in several of the UrbanSim sub-models.

3.6 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 models used are 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 to be used are inspired from the Urban Economics literature. These include housing characteristics, various measures of accessibility and neighborhood characteristics.

3.7 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. Each of the development projects (buildings) generated by the Development Project Transition Model "choose" their location, based on characteristics and attributes of gridcells. This is done with a multinomial logit probability, which 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.

3.8 Land-price Model

This model 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.

3.9 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.

4 Required and available data

Given the level of detail at which most models work it is not surprising that a great deal of data is required. Since this work is focused on the development of a prototype UrbanSim model, and since obtaining all of the data normally required by UrbanSim would have been out of the scope of this project, not all of the standard data were used in this prototype. Emphasis was placed on using readily available data. The following section compares specific data requirements for UrbanSim with data available for the Lausanne model.

4.1 UrbanSim data requirements

Required data for UrbanSim can be separated in two groups: base data and primary data. Base data include overall model parameters (e.g. gridcell dimensions, measurement units, etc.), user defined parameters (target vacancy rate, yearly rate of relocating jobs and households, etc.) and control totals (population and employment projections).

Primary data consist of a detailed characterization of the households, jobs, buildings and gridcells contained in the city. This can be better understood if described as the series of six tables that need to be built in order to calibrate and run UrbanSim. These tables are the gridcells, households, jobs, buildings, development event history and development constraints tables.

The gridcells table is the central table that links all the other tables. It identifies and characterizes each gridcell in the urban system. The characteristics of a gridcell include:

- Location relative to other gridcells (spatial coordinates)
- Political characteristics (zoning, county, city, etc.)
- Traffic analysis zone correspondence (from the transport model)
- geographical characteristics (distance to transportation infrastructure, gridcell slope, location relative to stream buffers, etc.)
- characterization of land use (e.g. number of residential units, surface area of non-residential activities, etc.)

Each observation of the households table represents one household. Every household is characterized by socio-economic attributes such as number of persons, number of vehicles, number of children, age of household's head and household income. The households table also includes data on location for each household (identifying the gridcell in which they are found).

Each observation of the jobs table represents one job. Jobs are characterized by industrial sector, building type and location.

Each observation of the buildings table identifies the building location and their characteristics such as, construction year, building type, number of residential units in the building, non residential surface area, etc.

The development event history table contains information on historical real-estate developments and its characteristics. This information is used by the Development Project Transition Model, which samples from this table to create the queue of buildings to be placed in each modeling period.

The last primary table is the development constraints table. It identifies what constraints are placed on different types of gridcells. These can be zoning constraints, physical constraints (e.g. no building in stream buffers) or idiosyncratic individual gridcell constraints. The development project location choice model uses this to table to identify gridcells to which new developments can be placed.

4.2 Data available for Lausanne

The sources of data for the Lausanne-Morges model were the following:

- The Swiss Federal Census for the year 2000 (RFP for *Recensement Fédéral de la Population*) [1] contains information on households identifying their location and characteristics. Data on residential buildings is also available from the RFP.

- The Swiss Federal Census for Enterprises for the year 2001 (RFE Recensement Fédéral des Enterprises) [2]: reports information on firms including their location, number of employees and industrial sector.

- The Transport Model for Lausanne (EMME) includes information on the transport network and transport system performance measures.

- GIS map layers containing geographic information, zoning (plan types from the *Plan d'Affectation Cantonal* [3]) and transport network information (localization and characteristics for transport infrastructure).

Due to resource constraints, little additional data were were collected. As such, the data used lacked some information that is normally required for a fully-operational model. In particular, there were some land-use and household

[1]See: Office fédéral de la statistique Suisse, 2008a

[2]See: Office fédéral de la statistique Suisse, 2008b

[3]See: Administration Cantonale Vaudoise, 2008

information that was not used. On the land-use side, missing data included: residential and non-residential land value, improvement values (residential and non-residential) and building information (e.g. surface area, location) for non-residential buildings. With respect to household information, there was no information available on household income or household car ownership. Due to the fact that these data were not readily-available, and that it was beyond the scope of this project to find or collect them, it was necessary to construct part of the data required by UrbanSim. The following section describes how these data were constructed. It should be noted that although some data were not used in the prototype model, it was not because they would not be possible to get, only that they were not quickly available within the time constraints of the project. Section 8.3 discusses the sources for the missing data that would need to be collected for an operational model.

This issue is addressed in the next section. Resources permitting effort in future work should be placed on collecting this data from existing sources or implementing adequate models to estimate it, however this was out of scope for this project.

5 Data Preparation

In order to develop an UrbanSim Model for the city of Lausanne it was first necessary to build the primary tables (described in Section 4.1). The data construction processes are described in the following sections, together with an analysis of the Lausanne-Morges area for the calibration year (2000).

5.1 Constructed and used data

This section describes the processes used to construct missing data and build the primary tables that UrbanSim needs in order to calibrate models and run simulations.

5.1.1 Gridcells Data

Gridcells were defined to fit the hectare system used by the Swiss Federal Government and at which most of the data were available. The number of residential units is directly calculated as the sum of dwellings for each gridcell from the buildings table (explained ahead). The total surface by non-residential sector was also calculated from the buildings table, as the sum of surfaces of non-residential buildings.

Geographical information for each gridcell was obtained from different GIS layers (see Bettex, 2008), including percentage of water cover, planning type (zoning), commune ID and distance to highways and arterials. GIS layers were also used to associate each gridcell with with its EMME traffic analysis zone.

The available sources did not provide information on land value. It was assumed that land value is proportional to density. Residential land value was estimated as population density per square kilometer. Non-residential land value was estimated as job density per square kilometer. Future work should consider gathering information on these variables. There are estimations on land value for the Lausanne area done by private consulting firms (e.g. Wuest & Partner) that could be used for this purpose.

5.1.2 Households Data

Household data came mostly the Swiss Federal Census, but some variables needed to be constructed. The census lacks information on income for each household, but reports the job sector or occupation for the household head. Reasonable estimates of average income by job sector were associated to each household. Further work should consider a better way to assign income to households. One option would be to use estimates of income by occupation that could be obtained from public sources (e.g. *Départment fédéral du finance*), or even private sources such as Kelly Employment Services. The rest of the variables (number of persons, number of children, location) were directly extracted from the census.

Since there was no information available on car ownership this variable was included with the value of 0 for every household. For the moment this does not represents a problem since models can be calibrated without considering household's car ownership. Also the available transportation model does not disaggregate households by car ownership. Future work should include gathering information or constructing sound estimates for this variable. Automobile ownership data is normally collected in the census, but was unavailable for the data at our disposition. This data could be obtained from the *Office fédéral des statistiques* to complement the data on households currently available at the EPFL. It would also be possible to estimte automobile ownership through the use of count data models. A possible source for data for such a model is the Swiss *Microrecensement sur le comportement de la population en matière de transports* Office fédéral de la statistique Suisse OFS, 2008.

5.1.3 Jobs Data

Job data is obtained from the Federal Census for Enterprises (RFE); each job from each enterprise is listed as an observation. Jobs are classified by the industrial sector of the enterprise which contains them. Building type categories (commercial, industrial, governmental) are constructed according to the sector of the jobs they contain using the NOGA 2002 Industrial Classification.

5.1.4 Buildings Data

Data for residential buildings from the RFP was quite comprehensive. Residential buildings in each gridcell were obtained directly from from the RFP, which included the geocoded location of each residential building. Data on the number of residential units per building was also obtained from the RFP.

There was no data available for non-residential buildings. Instead, synthetic buildings were "created." Non-residential buildings were created so as to be large enough to house the number of jobs by hectare. These synthetic buildings data were sufficient to allow the development of the prototype model. Three different types of non-residential buildings were considered: commercial, industrial and governmental. Jobs were classified by the type of building that they occupied (e.g. commercial jobs were considered to be housed in commercial buildings). Buildings were then created to house the jobs present. This is certainly an aspect to improve in future work since this assumption forces jobs to be only in their specific kind of building (commercial, industrial or governmental). Building size (surface area) was determined as a function of the number of jobs and an assumed vacancy rate. As such, each gridcell contained at most one building for each of the non-residential building types.

No information on "improvement value" was available for residential or non-residential buildings. As a result, improvement value for residential buildings was estimated as a direct function of the number of residences and non-residential improvement value a function of building surface area.

A possible source of more detailed information on building improvement value are private companies such as Wuest & Partners who collect data on unit surface area values. A possible source for non-residential buildings (location, size, etc.) is the *Registre cantonal de bâtiments* of the Canton of Vaud. This is maintained by the *Office d'information sur le territoire* of the Canton.

5.1.5 Development Event History Data

Information on residential development event history was obtained from the Swiss Federal Census. Each residential building built between 1990 and 2000 was considered a development event and the number of residential units and construction year are attributes for the building. There was no information available on non-residential building developments so data needed to be constructed. Information on the number of jobs by industrial sector in each gridcell was obtained from the Federal Census for Enterprises for years 1995 and 2005[4]. If the increase in the number of jobs in a particular industrial sector inside one gridcell was greater than 5 then it was assumed that a development event had occurred and a new building (of the building type associated to the job sector) was included in the development event history. Surface area of non-residential buildings was estimated as the number of new jobs multiplied by the assumed average surface usage per worker including hypothesized vacancy rate (see 5.1.4). The construction year for every non-residential building was set to 1998.

[4]The year 2005 was used instead of 2001 in order to increase the number of development events. Experience from the Brussels model identified small numbers of development events as a major hindrance. The use of this data does not affect any other aspects of the model

5.1.6 Development Constraints

UrbanSim needs to know how new real-estate development is constrained for each gridcell. This was determined by the physical or geographical characteristic of the gridcell. The *Plan d'affectation du sol* for the Canton of Vaud was used to attribute a plan type to deach gridcell. There was no readily-available information on building or density restrictions by plan type so it was assumed that the observed situation (for year 2000) defined upper and lower bounds for each kind of development (number of residential units, non-residential surface for each type of building). The lower bound was assumed to be 0 for every plan type. For the upper bound (maximum number of units or maximum surface by type) it was first evaluated to use the observed maximums in each type plan. After initial analysis maximum residential units and surface areas were defined as the average plus two standard deviations for each plan type, by commune. For undevelopable plan types, such as agricultural or green protected areas, the maximum was set to 0 for every development type.

5.1.7 Other Data

In addition to the six primary tables, other data is required by UrbanSim. Demographic and economic estimates (number of households and jobs) for future years are based on Cantonal Office of Mobility projections used in the EMME model. The average of actual population growth rates for the 45 communes in the model was used for the years 2001-2007.

Transportation system performance data was obtained from the EMME Lausanne-Morges transportation model. Travel impedance data, such as interzonal travel times or generalized cost, was obtained for each traffic analysis zone. Accessibility measures were calculated according to Equation 1. The EMME model was calibrated for 2005, so Furness OD-matrix updating was used to estimate a year 2000 origin-destination matrix, the base year for the Lausanne UrbanSim model. There were some non-negligible inconsistencies between the TAZ population figures used to calibrate the EMME model and those from the UrbanSim baseyear. Given the time available, it was not possible to resolve all of these inconsistencies. It would, however, be ideal to resolve these inconsistencies for a fully operational model for Lausanne.

5.2 The Lausanne-Morges Region in 2000

UrbanSim works at the level of gridcells located in the *internal* zones of the travel model with which it is coupled. Travel models generally contain two types of zones within the TAZ system. Travel models and travel model analysis concentrate on those areas that are found at the center of the region interest. These central areas are divided into relatively small TAZs (e.g. the size of a neighborhood). In order to control for traffic moving into and out of the central region, more aggregate zones are used towards the extremities of the region. These are known as the *external* zones of the model. Such zones might make

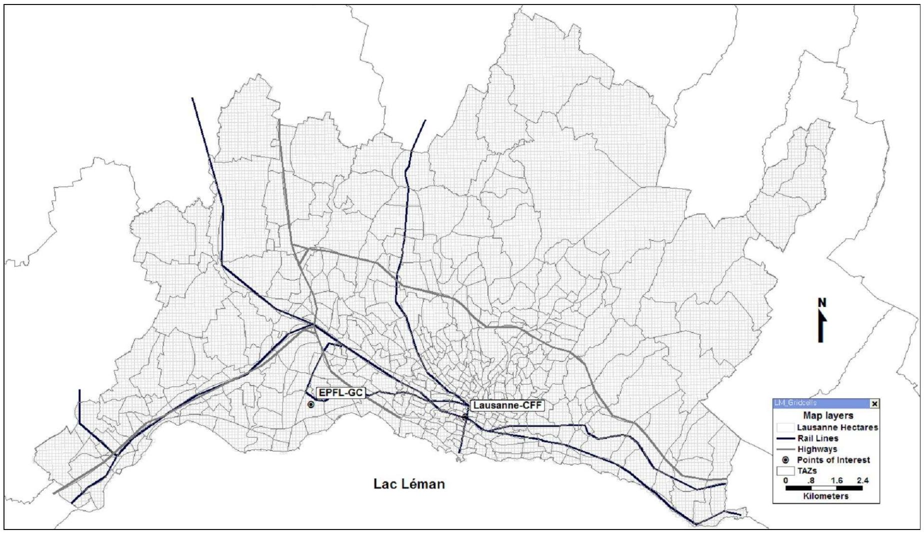

Figure 1: The "Lausanne-Morges" Region

up entire districts or even cantons. The Lausanne EMME model is composed of 497 zones.

There are 485 internal zones that include 45 communes that make up a triangle between Lully in the West, Bottens to the North and Cully to the East. Figure 1 shows a map of the study region. This map shows the delimitations of the TAZs. The hectare gridcells are also present and overlap with the internal TAZs. A few points of interest are included to orient the reader. The main highways and rail (metro and tram included) show the main transportation infrastructure for the region. The study region covers $\approx 200\ km^2$.

In 2000, the study region counted a population of roughly 277,000 people. Most of this population was found towards the center of the region surrounding the city of Lausanne (see Figure 2). For the most part, population collects along the shores of the lake and near to main highways and rail infrastructure.

The location of enterprises and jobs shows a similar pattern to that for population (see Figure 3), i.e. centered around Lausanne and following the shores of the lake and transportation infrastructure. One difference is that jobs tend to be more compactly located than households and therefore population.

Given the interest in the interaction between transportation and land use Figure 4 shows a measure of accessibility (see Section 3.5) for the hectares of the region. The units of the accessibility measure are not terribly intuitive, except that the higher the accessibility, the more "accessible" the location is. In general, hectares located closest to jobs and households and near to transportation infrastructure have the highest accessibility.

The characteristics (e.g. accessibility) of the locations of households and jobs are the primary factors included in the estimation of the location choice

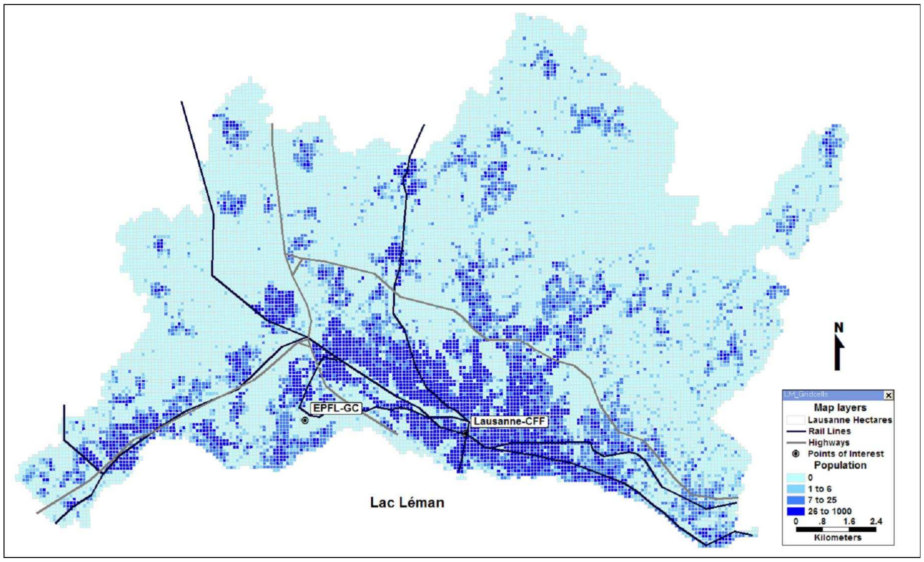

Figure 2: Location of population by hectare in the Lausanne-Morges region (2000)

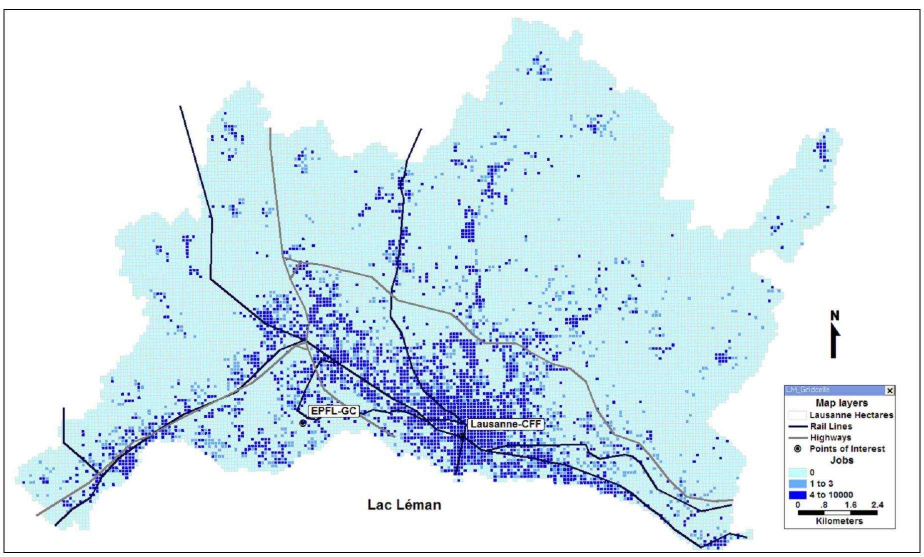

Figure 3: Location of jobs by hectare in the Lausanne-Morges region (2000)

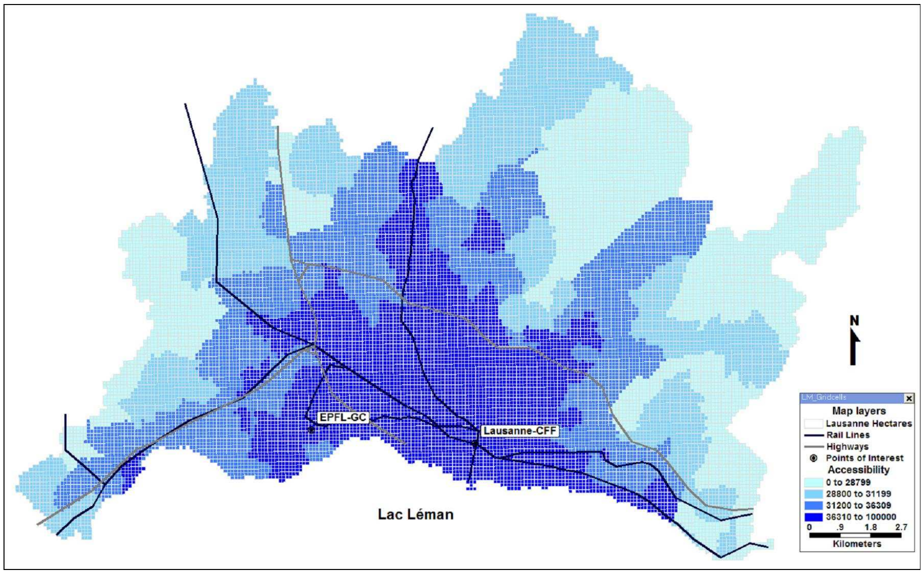

Figure 4: Accessibility in the Lausanne-Morges region (2000)

and land price models of the Lausanne UrbanSim prototype.

6 UrbanSim Submodels Estimated for the Lausanne-Morges Region

The following section describes the various submodels 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.

$$V_j = \beta' x_j + \epsilon_j \ \forall_j \tag{2}$$

It is a linear combination of the different characteristics (xs) of the alternative. In the case of a household location choice, the xs would be, for example, distance from the CBD, access to employment, etc. $V_j$also includes an 'error' term denoted as ϵ_j. In the case of the MNL, ϵ_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.

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

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

$$Price_j = \beta' x_j + \epsilon_j \; \forall_j \tag{4}$$

In the case of the land price model, the xs are similar to those for the location choice models. In the case of OLS, ϵ_j is distributed normally with mean zero and variance σ estimated in the regression.

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

Results of the models estimated for Lausanne are compared with a previous application of UrbanSim for the Greater Wasatch Front area of Utah. The Utah Model and its estimation results are described in Waddell et al., 2007, Waddell and Nourzad, 2002, Waddell and Ulfarsson, 2003a and Waddell and Ulfarsson, 2003b. The Lausanne models generally have fewer variables than those in the Utah example. The primary reason for this is that fewer variables were available for use in the models. This stems from the lack of data which would have rendered the use of many variables meaningless. Consider the variable for non-residential surface area - a variable used in many of the Utah location choice models. As explained above, in Lausanne there was no data available on non-residential surface area so this data was estimated a function of number of jobs. Use of non-residential surface area as a variable would have been synonymous with including the number of jobs in the model. Only variables that had reliable data were tested in the models.

6.1 Household Location Choice Model

Only one Household Location Choice Model was estimated for every household. Model specification was based on the classical urban economics model that assumes a trade-off between transportation, land cost and location attributes (Alonso, 1964; Muth, 1969). The model included 10 independent variables, which seems to be enough given the amount and quality of available data; but is considerably less than the number of variables included in the Utah model (24 variables). Estimation results are shown in Table 1.

All of the variables are highly significant and meet a priori expectations (most of them also have the same sign that the coefficients for the Utah model).

The probability of choosing a particular location decreases with cost, which is a very reasonable behavior. The probability also decreases with the number of residential units in the same gridcell if the household has children; this is also expected since families tend to prefer low density neighborhoods. The choice probability increases with the number of households in the same income level (for mid and high income households) and with retail jobs in the surroundings (which most likely means easy access to shopping facilities). The probability also increases if the household is young and the location is in a mixed use gridcell and has high density.

Variable	Coefficient	Std Error	T-value
Cost to income ratio	-4.278	0.661	-6.472
% high income if high income w.w.d.	0.030	0.001	48.543
Log access to population	0.432	0.046	9.416
% low income if low income w.w.d.	0.023	0.001	20.472
Log retail employment w.w.d.	0.036	0.003	11.078
Residential units if hh has children	-0.005	0.000	-44.207
Travel time to CBD	-0.021	0.003	-7.782
Travel time to train station	0.032	0.002	14.997
Young hh in high density residential	0.438	0.018	24.661
Young hh in mixed use	0.462	0.022	21.315
Initial Log-Likelihood	-4.44E+05		
Log-Likelihood	-4.41E+05		
Likelihood ratio test	6.93E+03		
Log-Likelihood ratio index	0.008		
Number of observations	130655		
Convergence statistic	0.001		

Table 1: Household Location Choice Model

Regarding the transportation variables, households prefer locations with high access to population. This particular result differs from the Utah model, which indicated that access to population was a negative attribute for household location. While this could be explained by omitted variable bias, its also reasonable to assume that this is product of differences between the two modeled areas (mostly on density and dwelling agglomeration).

Locations near the Central Business District are attractive while, at the same time, households tend to prefer locations away from the train station. This is very reasonable since the CBD concentrates commerce and service facilities which are attractive to households in general; usually train stations also concentrates this kind of activity but they also generate negative externalities (noise, visual intrusion) and tend to have more deteriorated surroundings, due to intense people-traffic, which makes it a less attractive location for households.

6.2 Employment Location Choice Model

Several employment location choice models were calibrated, one for all industrial jobs and one for each commercial sub-sector. Estimation results are shown in Tables 2 to 7.

The Industrial Employment Location Model presented good results. Ideally, different models would have been estimated for each of the industrial sectors. This was not done because the large number of industrial sectors resulted in some sectors having very few observations for estimation. As a result, one model was estimated for all industrial sectors. A better approach would be to estimate models for subgroups of industrial sectors, but this is not automatic in

Variable	Coefficient	Std Error	T-value
Is near highway	-0.039	0.014	-2.830
Log residential units w.w.d.	-0.024	0.005	-4.731
% low income w.w.d.	0.005	0.001	3.926
Gridcell is plan type Indutrial 4	0.059	0.014	4.095
Gridcell is plan type Indutrial 1	0.111	0.023	4.927
Gridcell is plan type Indutrial 2	0.097	0.022	4.356
Comercial jobs	-1.019E-04	0.000	-2.302
Industrial jobs w.w.d.	6.884E-05	0.000	9.868
Log total land value w.w.d.	-5.129E-08	0.000	-3.651
Initial Log-Likelihood	-1.399E+05		
Log-Likelihood	-1.397E+05		
Likelihood ratio test	4.07E+02		
Log-Likelihood ratio index	0.001		
Number of observations	41140		
Convergence statistic	0.000		

Table 2: Industrial Employment Location Choice Model

UrbanSim and time constraints meant the only feasible approach was to estimate one model. It is difficult to compare results of this model with those of the Utah model since it did include models for each one of the different job sectors inside the industrial category. At the same time, this model seems acceptable with all of the variables being significant and coefficient signs are consistent with *a priori* assumptions.

The probability of an industrial job choosing a particular location decreases with the number of residential units and commercial jobs in the surroundings, which is very expectable since industry tends to locate in low population-density areas. The probability also decreases with land value, while it increases with the number of low income households; this is also very likely to happen since industrial-use land is unattractive to high income households but may be suitable for low income households (due to the lower land price). The probability increases with the number of industrial jobs nearby and when the gridcell has an industrial development plan, which is also very reasonable. The only unintuitive result is the “Is near highway” coefficient, which is negative. One should expect industry to be attracted to locations near transport facilities, such as highways; however this is likely due to omitted variable bias or possibly to the use of population and job density instead of actual land value.

The Commercial Employment Location Model also presented good results. For this economic sector several sub-models were calibrated, each for a different sub-sector (retail, hotels, financial services, real estate and public services). In this case is also difficult to compare results with the Utah model since the job sub-sector disaggregation differs from one model to the other. However, every sub-model performed well and each one of them is revised separately next.

The retail sub-model’s coefficients indicate that the probability of a retail

Variable	**Coefficient**	**Std Error**	**T-value**
Log distance to highway	-0.034	0.004	-9.500
Log retail employment w.w.d.	0.358	0.010	35.716
Log service employment w.w.d.	-0.085	0.006	-14.145
Log total land value	-0.103	0.007	-14.640
Log work access to employment	-6.482	0.377	-17.192
Log work access to population	6.965	0.394	17.660
High income households w.w.d.	-0.001	0.000	-24.777
Low income households w.w.d.	0.000	0.000	18.964
Travel time to CBD	0.018	0.003	5.822
Initial Log-Likelihood	-9.60E+04		
Log-Likelihood	-9.42E+04		
Likelihood ratio test	3.58E+03		
Log-Likelihood ratio index	0.019		
Number of observations	28236		
Convergence statistic	0.000		

Table 3: Commercial Location Choice Model (Retail)

job choosing a specific location increases with the existence of other retail employments in the surroundings. This is very reasonable due to agglomeration economies and the same behavior was observed in the Utah model. The number of low income households also has a positive effect on the location probability. This is reasonable since mixed-use gridcells, that might hold retail jobs, are more likely to be used by lower income households than rich households; this also explains the negative coefficient for the land value variable. The probability also increases when the gridcell has high access to population (customers) but decreases with access to employment, this is explained by the fact that, in general, employments tend to be concentrated on specific areas of the city while retail tends to spread through the city in order to be accessible to population, so, this coefficients are reasonable (this also explains the positive coefficient for "travel time to CBD"). It is worth mentioning that the Utah model presented complete opposite results for the access variables. Again, this can be explained by omited variable bias or by structural differences between the cities.

The hotel sector sub-model shows that the probability of a hotel related job choosing a location increases with the number of service jobs and with access to employment. The probability also increases when the gridcell has a commercial plan type. On the other hand, land value, retail employment and access to population are attributes that makes the probability decrease. Regarding travel times, the model indicates that hotels like to be near to the train station (which makes sense) but they also prefer to be away from the CBD. This last result seems counterintuitive and can only be explained by a compensating effect between the "travel time to CBD" and "travel time to train station" variables, given that in this particular case the CBD and train station are very close to each other.

Variable	**Coefficient**	**Std Error**	**T-value**
Log retail employment w.w.d.	-0.104	0.016	-6.446
Log service employment w.w.d.	0.153	0.011	14.413
Log total land value	-0.271	0.014	-19.063
Log work access to employment	4.033	0.610	6.615
Log work access to population	-3.486	0.634	-5.497
Gridcell is plan type Commercial 2	2.616	0.181	14.461
Gridcell is plan type Commercial 1	1.572	0.347	4.525
Travel time to CBD	0.118	0.007	16.794
Travel time to train station	-0.102	0.007	-15.480
Initial Log-Likelihood	-2.630E+04		
Log-Likelihood	-2.565E+04		
Likelihood ratio test	1.31E+03		
Log-Likelihood ratio index	0.025		
Number of observations	7733		
Convergence statistic	0.000		

Table 4: Commercial Location Choice Model (Hotels)

Variable	**Coefficient**	**Std Error**	**T-value**
Log distance to highway	-0.038	0.010	-3.756
Log service employment w.w.d.	-0.074	0.017	-4.429
Log total land value	-0.073	0.024	-3.108
Log work access to employment	8.650	0.575	15.039
Log work access to population	-9.867	0.542	-18.194
Log total population w.w.d.	0.867	0.044	19.649
High income households w.w.d.	-0.001	0.000	-14.383
number of jobs same sector	0.004	0.000	70.917
Travel time to train station	-0.037	0.006	-5.728
Initial Log-Likelihood	-3.804E+04		
Log-Likelihood	-2.959E+04		
Likelihood ratio test	1.69E+04		
Log-Likelihood ratio index	0.222		
Number of observations	11184		
Convergence statistic	0.001		

Table 5: Commercial Location Choice Model (Financial Services)

Variable	Coefficient	Std Error	T-value
Is near highway	-0.122	0.022	-5.461
Log total land value	-0.167	0.011	-15.789
Log work access to employment	-7.000	0.334	-20.945
Log work access to population	6.131	0.358	17.134
Log total population w.w.d.	0.166	0.016	10.575
High income households w.w.d.	0.000	0.000	9.249
number of jobs same sector	0.003	0.000	70.191
Commercial jobs w.w.d.	0.000	0.000	-10.841
Industrial jobs w.w.d.	0.000	0.000	11.214
Travel time to CBD	-0.026	0.004	-6.675
Initial Log-Likelihood	-7.901E+04		
Log-Likelihood	-7.429E+04		
Likelihood ratio test	9.43E+03		
Log-Likelihood ratio index	0.060		
Number of observations	23229		
Convergence statistic	0.000		

Table 6: Commercial Location Choice Model (Real Estate)

The financial services sub-model indicates that jobs in this sector are more likely to locate in zones with high access to employment, population and jobs in the same sector this is very reasonable due to scale and scoop economies that can be found in this economic activity. Financial jobs also prefer locations near the train station. On the other hand, the location probability decreases with distance to highway, service employments, land value, high income households and access to population. All these results are consistent with the expected behavior of this kind of economic activity.

For real estate jobs, the probability of choosing a particular location increases with population, access to population, number of high income households, industrial jobs and jobs in the same sector. The odds decrease with land value, number of commercial jobs, access to employment and highway nearness. This means that real estate jobs prefers to be accessible to households, particularly rich ones, than to employments in general, while at the same time they prefer to be near the CBD.

For public services jobs, the probability of choosing certain location increases with access to population, service employments, highway nearness, mid income households and jobs in the same sector. The probability decreases with land value, access to employment, high income households and travel time to train station.

6.3 Real Estate Development Models

Three different real estate models were calibrated (for commercial, industrial and residential developments). The Utah model considered a much wider dis-

Variable	**Coefficient**	**Std Error**	**T-value**
Is near highway	0.487	0.042	11.664
Log service employment w.w.d.	0.264	0.012	21.511
Log total land value	-0.627	0.014	-43.426
Log work access to employment	-3.512	0.688	-5.103
Log work access to population	2.476	0.701	3.530
High income households w.w.d.	0.000	0.000	-3.599
number of jobs same sector	0.013	0.000	76.140
Mid income households w.w.d.	0.000	0.000	9.317
Travel time to train station	-0.057	0.006	-8.785
Initial Log-Likelihood	-2.689E+04		
Log-Likelihood	-2.285E+04		
Likelihood ratio test	8.08E+03		
Log-Likelihood ratio index	0.150		
Number of observations	7907		
Convergence statistic	0.000		

Table 7: Commercial Location Choice Model (Public Services)

aggregation for development categories, which makes comparison between the two models difficult. However, all models performed well and all of the variables are significant with coefficient signs that are reasonable. Results are shown in tables 8 to 10. An analysis for each model is presented next.

The commercial development location model shows that real estate development for this sector is more probable in areas with high access to employment and that already have commercial jobs located in them. Development is less probable when the gridcell is of development type "Mixed 1" or "Mixed 5", when the gridcell has high land value, or when there is high access to population. This last result doesn't seem very reasonable since commerce should prefer to be near potential customers; however it can be explained as a compensating effect because commercial development is already attracted to zones with high home access to employment.

The industrial development models indicate that the probability of development for this sector increases with the number of industrial jobs and with the accessibility to employment for households; this complies with expectations due to the existence of agglomeration economies. The probability decreases with the access to population for jobs.

The residential development model shows that residential real estate development is more likely in areas which already have a large number of residential units, with high access to employment and near to highways. On the other hand, land value decreases the attractiveness of an area for real estate development; this is very reasonable and complies with expectations. The negative value for the "home access to population" coefficient seems counterintuitive since the Household Location Choice Model shows that households prefer locations with high access to population. This is likely due to omitted variable bias and should

Variable	**Coefficient**	**Std Error**	**T-value**
Log home access to employment	3.205	0.745	4.305
Log total land value	0.136	0.038	3.596
Log work access to population	-2.295	0.821	-2.794
retail jobs	0.006	0.001	8.903
Hotel jobs	0.007	0.002	3.505
Financial Services jobs	0.003	0.001	3.577
Real Estate jobs	0.004	0.001	7.291
Public services jobs	0.007	0.003	2.555
Gridcell is plan type Mixed 1	-1.154	0.392	-2.947
Gridcell is plan type Mixed 5	-1.492	0.309	-4.831
Initial Log-Likelihood	-2139		
Log-Likelihood	-1956		
Likelihood ratio test	3.67E+02		
Log-Likelihood ratio index	0.086		
Number of observations	629		
Convergence statistic	0.001		

Table 8: Commercial Development Location Choice Model

Variable	**Coefficient**	**Std Error**	**T-value**
Log basic employment w.w.d.	0.271	0.062	4.377
Log home access to employment	3.030	0.973	3.115
Log work access to population	-2.959	0.933	-3.171
Metalurgy jobs	0.032	0.009	3.655
Machine manufacture jobs	0.008	0.003	3.023
Electric manufacture jobs	0.008	0.002	4.409
Hydro jobs	0.007	0.003	2.636
Construction jobs	0.022	0.002	10.826
Transport and comunic jobs	0.003	0.001	5.019
Food products jobs	0.007	0.002	3.759
Wood products jobs	0.025	0.013	1.927
Paper and printing jobs	0.005	0.002	3.080
Chemical jobs	0.011	0.004	2.755
Gridcell is plan type Industrial 4	0.443	0.132	3.360
Gridcell is plan type Industrial 1	0.492	0.202	2.435
Initial Log-Likelihood	-1592		
Log-Likelihood	-1388		
Likelihood ratio test	4.08E+02		
Log-Likelihood ratio index	0.128		
Number of observations	468		
Convergence statistic	0.001		

Table 9: Industrial Development Location Choice Model

Variable	Coefficient	Std Error	T-value
Log home access to employment	0.390	1.029	1.808
Log home access to population	-0.565	1.181	-2.125
Log total land value	-0.682	0.072	-12.603
Gridcel is near highway	0.212	0.064	3.896
Log residential units	0.794	0.036	20.666
Gridcell is plan type Mixed 3	-1.507	0.425	-2.757
Gridcell is plan type Mixed 4	1.395	0.606	2.271
Gridcell is plan type Residential 1	0.501	0.057	3.240
Gridcell is plan type Residential 3	-0.782	0.113	-8.822
Initial Log-Likelihood	-5775		
Log-Likelihood	-5064		
Likelihood ratio test	1.42E+03		
Log-Likelihood ratio index	0.123		
Number of observations	1698		
Convergence statistic	0.001		

Table 10: Residential Development Location Choice Model

be improved in further model development.

6.4 Land Price Model

The land price model performs well; every variable considered is statistically significant. However there are some significant differences with the Utah model. First of all, the number of variables included for Lausanne-Morges (7 variables) is small in comparison with the Utah Land Price Model specification (57 variables). Secondly, some of the coefficients have different signs: the Utah model indicates that access to employment and having a residential plan type are attributes that affect the price negatively. In the Lausanne model all the variables considered have a positive effect on land price. Of course this difference could be the result result of the smaller number of variables available for the Lausanne-Morges specification and consequent omitted variable bias. It could also be explained by the fact that population and job density were used as proxies for land-value (the dependent variable). Despite these weaknesses, the Lausanne-Morges land-price model seems reasonable.

7 Simulation Results

Every simulation with UrbanSim requires a predefined "scenario," which is a set of border conditions that create bounds for the model's predictions. For each scenario a set of Urbansim simulations must be done (one for each simulation year); the interaction with the transport model (EMME) is done every five years. For the present simulations there were four subsets of UrbanSim simulation

Variable	Coefficient	Std Error	T-value
Log retail employment w.w.d.	0.139	0.012	11.515
Log home acces to employment	1.017	0.094	10.796
Total population w.w.d.	0.077	0.017	4.465
Gridcell is plan type Ind 1	0.356	0.068	5.202
Gridcell is plan type Mixed 2	0.756	0.043	17.630
Gridcell is plan type Residential 2	0.501	0.032	15.754
Gridcell is plan type Residential 3	0.809	0.039	20.937
Constant	-3.415	0.861	-3.968
Number of observations	4488		
R-Squared	0.442		
Adjusted R-Squared	0.441		

Table 11: Land Price Model

		Zoning Variants	
		Current Zoning	**Less Restrictive**
Transport Variants	**M2**	Base Case	Land-use scenario
	No M2	Transport Scenario	N/A

Table 12: Simulation scenarios

(2001-2005, 2006-2010, 2011-2015 and 2016-2020), each one of them interacting with an EMME simulation (2005, 2010, 2015 and 2020). Figure 5 provides a diagram that depicts the interactions between UrbanSim and the EMME transportation model.

The results of three scenarios are presented here. Scenarios are distinguished by two characteristics - transport and zoning characteristics. There were two transport variants (refer to Table 12). Both used the 2005 road network for all simulation years. They differed in the public transport network that was used. One scenario included the 2008 changes to the public transport network (M2, etc.) whereas the other assumed the same network as in 2005 for all simulation years. These scenarios are identified as having the M2, or not having the M2. For "M2" scenarios, it is not only the M2 that is included in the transport network, but all modifications to the public transport network that accompany the introduction of the M2.

There were also two land-use variants. One variant used current zoning and the other a less restrictive zoning. The "current zoning" variant used the development constraints as described in Section 5.1.6. That is, for plan types considered as undevelopable, the maximum number of residential units and non-residential surface area were set to zero. For "developable" plan types, maximum residential units and non-residential surface area was set as the average + two standard deviations for the commune in which the hectare was found. Table 13 shows which plan types were considered to be developable and undevelopable. In the "less restrictive" variant, two plan types considered undevelopable in

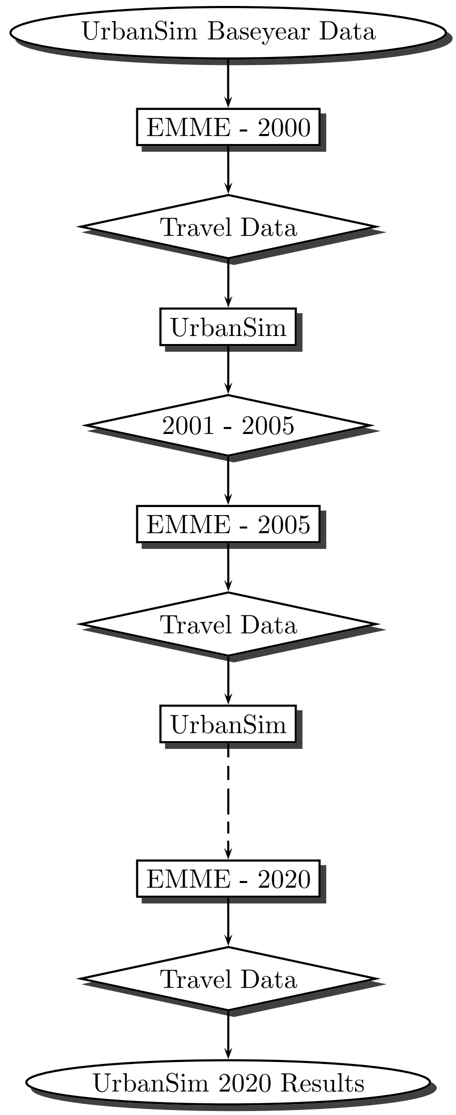

Figure 5: UrbanSim integration with the EMME travel model for Lausanne

Plan Type	**Current Zoning**	**Less Restrictive**
A plan spécial	Developable	Developable
Activités artisan. ou tert.	Developable	Developable
Agricole	Undevelopable	**Developable**
Agricole protégée	Undevelopable	Undevelopable
Agricole spéciale	Undevelopable	Undevelopable
Aire de communication	Developable	Developable
Aire forestière	Undevelopable	Undevelopable
Camping	Undevelopable	Undevelopable
Centre localité	Developable	Developable
Divers	Developable	Developable
Eau	Undevelopable	Undevelopable
Exploitation des matériaux	Developable	Developable
Habitation faible densité	Developable	Developable
Habitation moyenne densité	Developable	Developable
Habitation forte densité	Developable	Developable
Hameau	Developable	Developable
Hospitalière	Developable	Developable
Hôtelière	Developable	Developable
Industrie	Developable	Developable
Intermédiaire	Undevelopable	**Developable**
Non bâtir	Undevelopable	Undevelopable
Site protégé	Undevelopable	Undevelopable
Sport et détente	Undevelopable	Undevelopable
Tourisme	Undevelopable	Undevelopable
Utilité publique	Undevelopable	Undevelopable
Verdure	Undevelopable	Undevelopable
Viticole	Undevelopable	Undevelopable
Viticole protégée	Undevelopable	Undevelopable

Table 13: Developable and undevelopable plan types for the two zoning variants

the "current zoning" varying were considered to be developable. These newly developable plan types were the agricultural (agricole) and intermediary (intermédiaire) plan types.

The simulation scenarios tested were defined as a combination of the two variants. In the "Base Case" scenario, the "current zoning" and "M2" variants were used. This represents what is believed to be the most realistic or likely scenario. The other two scenarios were developed to demonstrate the types of simulations that can be done with UrbanSim. In particular, two types of scenarios were tested: one which asks the question of what the effect of transportation infrastructure on urban development is; the second which asks the question of what the effect of land-use policies on transportation system performance is. The "Transport" scenario asks what would happen if the modifications to the

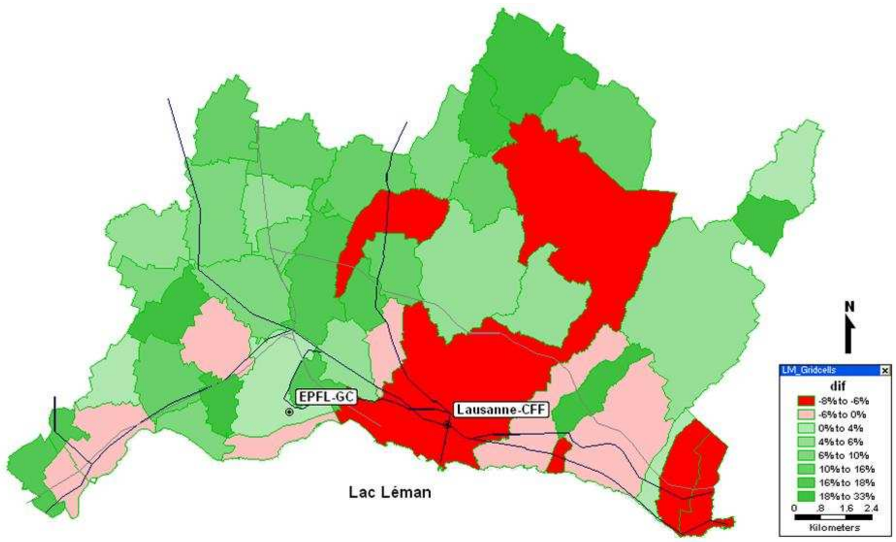

Figure 6: Difference in household growth rate by commune (2000 -2007)

public transport network, including the new M2, were not implemented. The "land-use" scenario asks what the effect of a less restrictive zoning policy would be on transportation system performance. The following sections describe and analyze the results obtained for each one of these scenarios.

7.1 Base Case Scenario (model validation)

Estimation results for the period 2000-2007 were compared with observed data for the same period at the commune level. This comparison allows evaluation of the quality of UrbanSim's predictions at an aggregate level. Figure 6 shows the difference between observed and predicted population growth rates by commune. The difference was calculated as actual growth minus predicted growth, so a negative value implies overprediction in the number of located households while a positive value indicates an underpredicted number of households by commune.

There are some considerable differences between the predicted and observed growth rates. However results are surprisingly good, considering the available data and resources. Most of the bigger communes in terms of population (Lausanne, Morges, Pully, Prilly, Ecublens, Renens) have a difference not higher than 10% between predicted and observed location. Also, there are 242354 persons located in communes with a difference not higher than 10%, which represents 85% of the total population in the modeled area. Most dramatic differences like the one seen in Lausanne (which is the bigger commune and shows 8% overprediction) can be explained by the use of fictitious development constraints, which were built from observed data and which do not necessarily comply with "real"

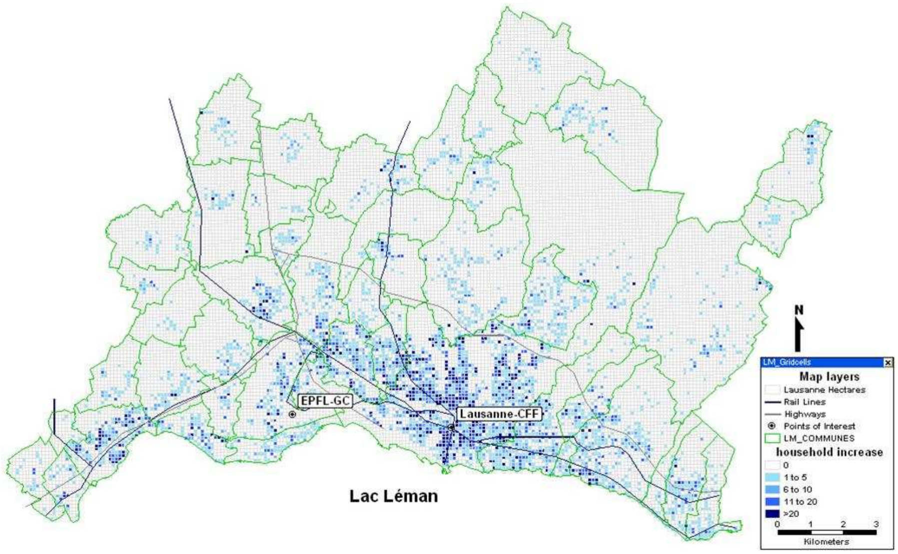

Figure 7: Household location increase between 2000 and 2020

development constraints (see 5.1.6). The difference can be also explained by the lack of information; better results should be expected with models including more variables and more accurate data.

The base scenario considered actual conditions for the transport system (including the M2 after 2008) and development plan types; simulations were run until year 2020. Figure 7 shows the increase in household location between years 2000 and 2020 while figure 8 does the same for job location.

New households are concentrated mostly in the communes of Lausanne, Prilly, Renens, Crissier and Bussigny. These are areas that already presented considerable development in the base year (2000). In general the predicted location of new households shows a densification in already developed areas.

New jobs are located mostly in the communes of Lausanne, Pully, Prilly, Renens, Morges and Lutry. These result, in general, shows a densification of jobs in already developed areas. Both new jobs and new households tend to locate in the axis that joins Lausanne's central area with Renens.

The observed densification in forecasted results is product of the parameters obtained for calibrated submodels (see section 6) and the restrictions considered in the development constraints (see 5.1.6). The Household Location Choice Model (see 6.1) and the Employment Location Choice Model (see 6.2) indicate that both households and jobs generally tend to agglomerate, choosing locations near other households and jobs, and preferring zones near the city centre. In addition, the Real Estate Development Models (see 6.3) also tend to locate new developments in already developed and central areas. While the behavior predicted by these models is correct, the lack of more restrictive constraints (such

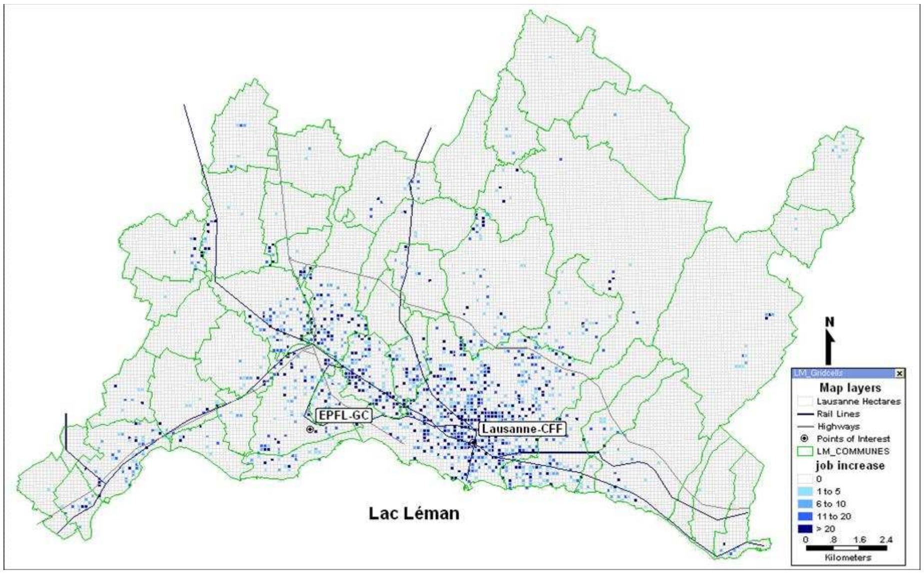

Figure 8: Job location increase between 2000 and 2020

as capacity constraints) is allowing both agents and development to agglomerate excessively. In reality we see that both location and development are strongly bounded by political or physical restrictions (despite how attractive a location is). In order to get better results is necessary to gather more comprehensive data on development constraints, to model a more realistic set of restrictions. A wider spread location may be possible if development constraints were to change, this will be addressed in the land-use scenario.

Regarding the transport system performance, the total travel time for the base scenario is 459788 minutes.

7.2 Transport Scenario

The "transport scenario" considered current zoning for the land use and transport system but no change in the transport network after 2008 (the transport system network remains as the 2005 network for every simulation year). This difference in the transport system between the two scenarios should generate different system performances as well. Figure 9 shows the percentage difference in the "access to population" variable between the two scenarios at the traffic analysis zone level. The difference was calculated as access in the base scenario (with M2) minus access in the transport scenario (no M2), so a positive value implies better accessibility in the base scenario.

Figure 9 also includes the location of the M2. The map clearly shows that the construction of the M2 line improves the level of access in zones which are

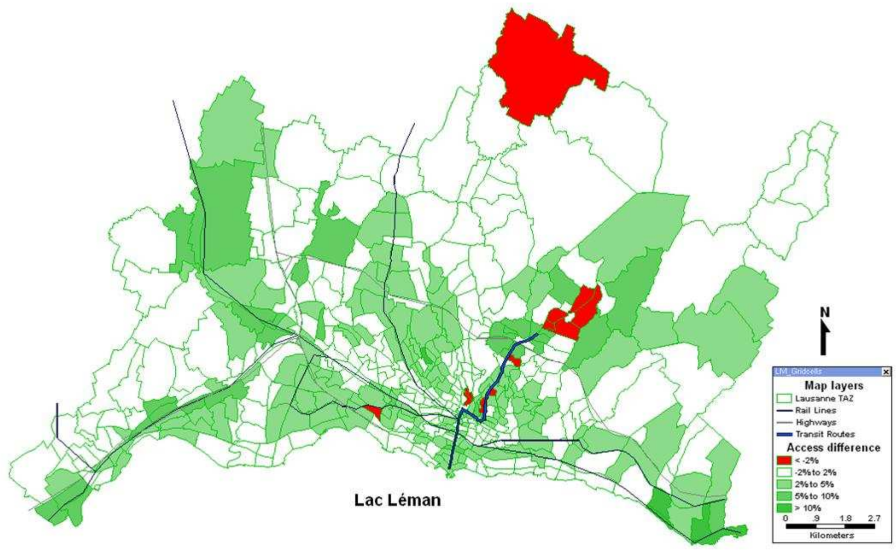

Figure 9: Access to population difference between base and transport scenario

near the new metro. However, there are some exceptions, that can be explained by the disappearance of some bus lines in the M2 corridor. The stops of a metro line are more distant than the stops of a bus line; this implies that some zones which had a bus stop within them when they were only served by buses might have lost the bus stop without gaining a metro stop when the M2 was built. Differences in accessibility measures can be also explained by changes in activity location patterns (see Equation 1) which explain some differences in areas that are not directly influenced by the new M2 line.

Regarding overall transport-system performance, the total travel time for the transport scenario in year 2020 is 474879 minutes which is considerably more than the total travel time for the same year in the base scenario (459788 minutes). This is a very logical result since the M2 line (included in the base scenario) is expected to have a strong positive effect on the transport system performance.

The different conditions in the transport system should have an effect on household and job location. Figure 10 shows the difference in household location between the two scenarios. The difference is calculated as the number of households by gridcell in the base scenario minus number of households in the transport scenario. Only positive differences are shown, so the figure shows the increase in the number of households by gridcell.

The map shows that there is an important difference in household location between the base and the transport scenario. There is, however, only a weak pattern of more households locating near the new M2 line.

Figure 11 shows the positive difference in job location between the transport

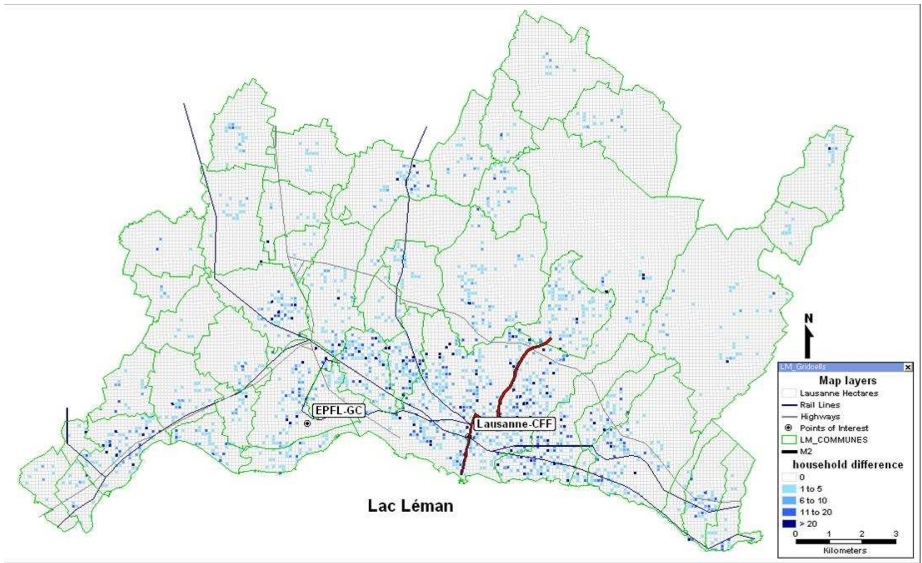

Figure 10: Difference in household location (base scenario vs transport scenario)

scenario and the base scenario. It was calculated in the same way as the difference in the household location, so it shows the increase in the number of jobs by gridcell if the M2 metro line is built.

Again there is not a clear pattern showing an increase of job location only in the surroundings of the M2 line. However the difference is considerable and a big part of it it's because higher job location near the M2 line.

It's interesting to analyze if location by income level is affected by the construction of the M2 line. Figure 12 shows the difference in the number of high income households located by gridcell. Even though the effect of the M2 is not obvious in the general household location, it's very clear that the new metro line attracts a considerable amount of high income households.

7.2.1 Land-Use Scenario

The "land-use scenario" consisted in a relaxation of development constraints for some plan types, leaving the rest of the system conditions untouched (both for the land-use and the transport system). In the base scenario the "agricultural" (agricole) and the "intermediary" (intermediare) types were considered as "undevelopable", so their maximum number of residential units and maximum non-residential surface was set to zero. In this scenario both agricultural and intermediary plan-types were considered as allowing real-estate development, having the same development constraints than the medium-density residential type in each of the communes.

The first expected effect of this scenario's conditions is a change in the

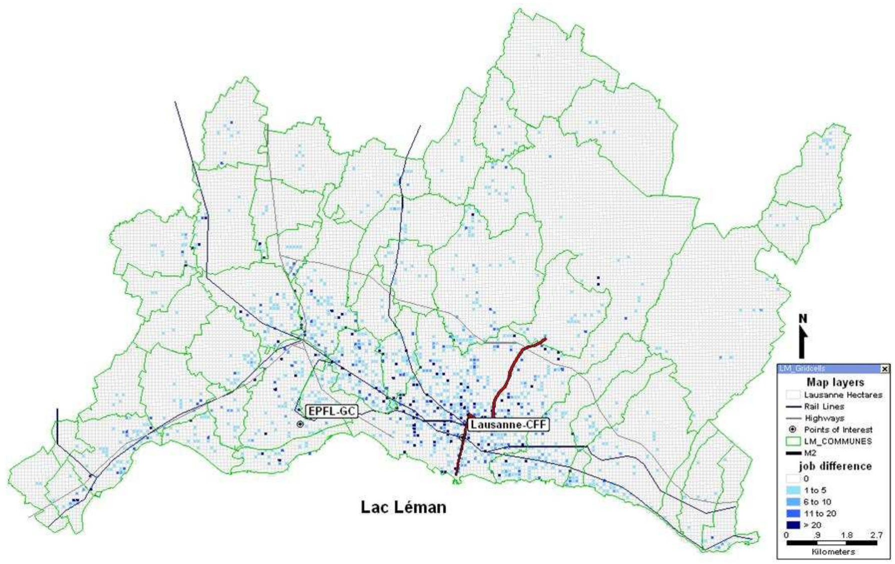

Figure 11: Difference in job location (base scenario vs transport scenario)

distribution of located households. Figure 13 shows new located households, only in agricultural and intermediary zones; figure 14 does the same for new located jobs.

The maps show that, for both households and jobs, there is new location (and therefore new real-estate development) in gridcells that forbid development in the base scenario. There are 1338 new households and 780 new jobs located in these types of gridcells, which represents 0.8% and 0.4% of the total number of households and jobs respectively. This is certainly not a massive redistribution of real-estate development or activity location, but the results show that changes in the constraints have effects in location and development. Also, for the same reason, this land-use scenario show a little more spread in activity location than the base scenario.

It's relevant to analyze if these changes in development constraints (and the following changes in location) have any considerable effect on the transport system performance. As a general indicator, the total travel time for the land-use scenario in year 2020 is 458092 minutes, which is a little less than the total travel time for the base scenario in the same year (459788 minutes). It's possible to conclude then, that the change in regulations modeled in the land-use scenario generates a better overall performance for the transport system (having the same network in both scenarios). It is hard to say if this is a correct result or not; usually a more spread-out city has higher travel times. However, lower travel costs are possible possible under certain conditions, specially if the spread implies agglomeration of activities in "suburban centers", reducing travel

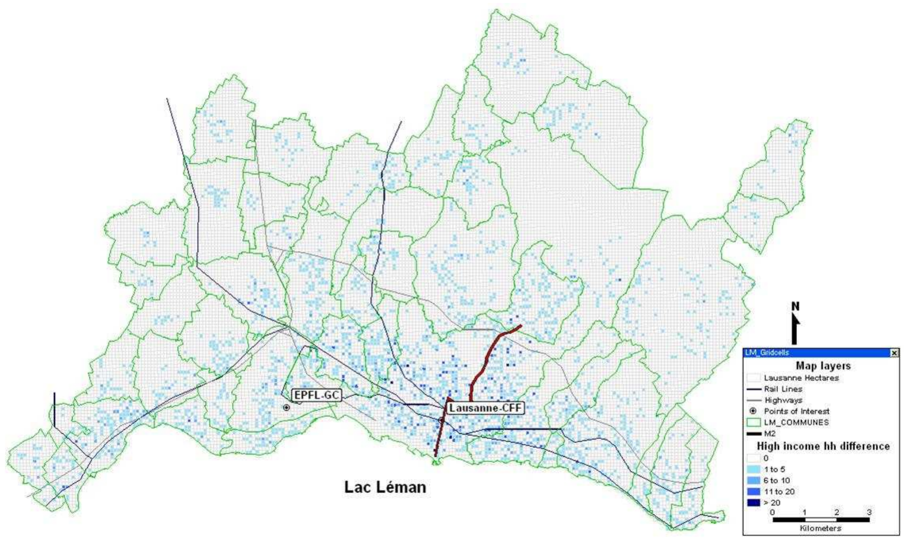

Figure 12: Difference in high income household location (base scenario vs transport scenario)

distances for a part of the population. In any case, a deeper analysis is required for this kind of results.

It's worth mentioning that previous UrbanSim applications (Waddell et al. 2007) showed that there is a big difference in forecasted transport system performance when using integrated simulation or just transport-system simulation. The results presented here show that changes in the land use system have an effect in the transport system, but it would be good to test if there is any relevant differences between straight EMME forecasts for year 2020 and integrated UrbanSim-EMME forecasts for the same year.

8 Evaluating the Lausanne UrbanSim Prototype

This document reports on the development of the Lausanne prototype UrbanSim model. It was designed to ask two questions:

1. Is it possible to develop an UrbanSim model for evaluation purposes for the Lausanne-Morges region given readily-available data and limited human resources?
2. If so:
   (a) how does the model perform?
   (b) what would be required to develop a fully-operational model that could be used for planning purposes?

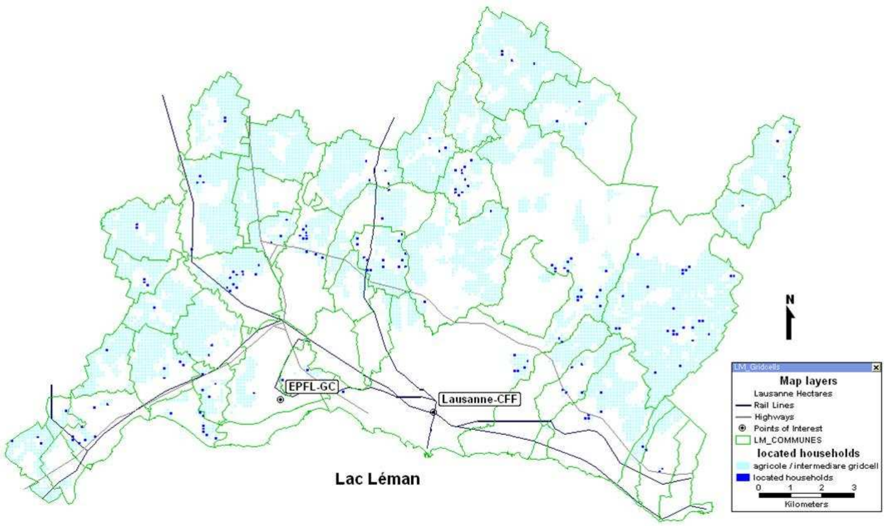

Figure 13: Location of households in agricultural and intermediary zones

This section is structured in such a way so as to answer these questions consecutively.

8.1 Feasibility

This work has demonstrated that it has indeed been feasible to develop an UrbanSim model for evaluation purposes given the use of readily-available data and limited human resources. Preparation for development began in January. These initial stages involved coordination with the project partners (TRANSP-OR, CHOROS and LaSIG) to establish data availability. Base data was delivered in March. Model development began in earnest mid-April.

Preparation of the data for the baseyear database (not including submodel coefficients) took until the first week of June when the first simulations were run. This work was done primarily by one postdoctoral researcher with around one week's time from a PhD student for the development and preparation of the basic elements (gridcell definitions, basic geographical characteristics) of the gridcells table. Three weeks were used to estimate initial location choice and land-use models and to organize integration with the transport model. The first simulations with the submodels estimated for the Lausanne region and integration with the transport model were done in the last week of June.

In the first weeks of July, the submodels were estimated and improved by a PhD student who also undertook more detailed analysis of development constraints. Two more person-weeks of time were required to undertake various simulations that incorporated the refined models and development constraint scenarios and finally produce the results reported here. Taking the work of the

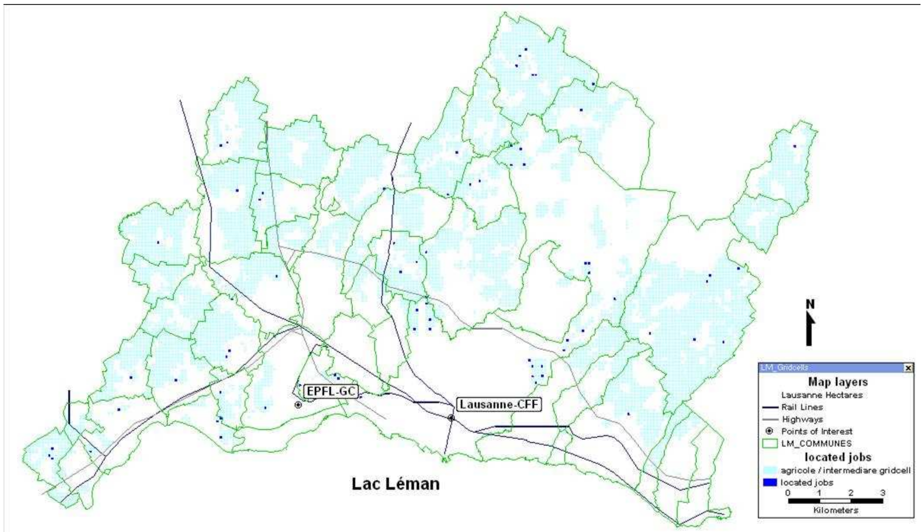

Figure 14: Location of jobs in agricultural and intermediary zones

various people who worked on the development of the project together (one postdoctoral researcher, senior researcher (who operated the transport model), two doctoral students, 4 person-months of effort were required to arrive at the model used to produce the results reported here.

8.2 Evaluation of Performance

The following data were not readily-available and thereby not used in the prototype model:

1. Land-price data,
2. non-residential building data,
3. residential and non-residential improvement value,
4. space requirements per job,
5. various geographical characteristics of gridcells (e.g. slope, inside stream buffer, etc.),
6. household car ownership,
7. household income.

Despite of the lack of this data that would need to be incorporated into an operational UrbanSim model, results were pleasantly surprising. The individual submodels had fewer variables than those in a typical operational model (see Waddell et al. (2007)). However, the variables that were included in the final models had signs that were almost always consistent with *a priori* expectations and were statistically significant.

Simulation results were also pleasantly surprising. From a qualitative perspective, the prototype model produces results that, albeit not perfect, are not absurd. It places jobs and population in locations that are believable and not in locations where development would be impossible given zoning regulations even though at first glance it seems to over predict population increases in already dense areas. In particular, the model over-predicts growth rates in the most densely populated areas of the region (e.g. in Lausanne itself) at the expense of too-small growth rates in surrounding communes. At the same time, the observed differences are remarkably small given the lack of some particulary critical data (e.g. land-prices). All in all, it is fair to say that the model performs amazingly well given the data and human resources that were invested in the development of the model.

8.3 Further Work Required for an Operational Model

Were development on an operational model to be continued, it would be required in two broad areas: data collection and incorporation, and submodel improvement and calibration.

In terms of data collection, the various data mentioned in Section 8.2 would need to be obtained for an operational model. The most important of these is land-price data. While not publicly available, land-price information is collected by private entities in Switzerland (e.g. Wuest & Partner) and from whom it would be possible to purchase it.

Better data on buildings (particularly non-residential buildings) would also be necessary. This would include data on building type (commercial, vs. industrial, etc.) as well as on size (surface area). It seems that such data exists and could be obtained from the *Registre cantonal de bâtiments* from the Canton of Vaud. The *Office d'information sur le territoire* would need to be contacted to obtain such data.

Information that is lacking for residential as well as non-residential buildings is improvement value. It is unclear where this data could be obtained, but it might be available from the *Section de statistique économique, financière et de l'environnement* of the *Service cantonal de recherche et d'information statistiques (SCRIS)* of the Canton of Vaud. Some data (improvement value per m^2) is also available from private sources such as Wuest & Partner and could be purchased from them.

Data on space requirements of different types of jobs per gridcell (e.g. m^2 per job) could be derived by a combination of non-residential building data (see above) and the already available data on enterprises. The enterprise data includes the number of employees by enterprise. If enterprises could be assigned to buildings, the calculation of space requirements would be elementary.

The actual gridcell table could be enriched by the inclusion of more geographic data. Such data could include average gridcell slope, proportion of the gridcell covered by roads, etc. This information could prove helpful in the improvement of models. For the most part, the geographic layers required for these analyses are already available in the LaSIG laboratory and thus it is just a question of performing the geographic queries necessary to obtain them.

With respect to household data: household income data, or at least estimates by occupation might be obtainable from public sources (e.g. *Départment fédéral du finance*), or even private sources such as Kelly Employment Services. Automobile ownership data is normally collected in the census, but was unavailable for the data at our disposition. Perhaps this data could be obtained from the *Office fédéral des statistiques* to complement the data on households currently available at the EPFL.

Statistical methods to estimate some of the missing data would be possible. Hedonic methods (e.g. Baranzini and Ramirez, 2005) for land-prices and count data models (e.g. Milton and Mannering, 1998) for household automobile ownership (using data from the Swiss microcensus on transportation (see Section 5.1.2)) could be used. While it is always possible to "model" data, it should be recognized that experience with modeled data with UrbanSim has been mixed, and that some strongly discourage the use of modeled data as much as possible Nguyen-Luong, 2008.

Finally, one other issue related to data has not to do with obtaining data *per se*, but rather with making it consistent. As described in Section 5.1.7, UrbanSim baseyear data were not entirely consistent with the jobs and population data per zone used in the original EMME transport model. This difference appears to come primarily from the attribution of hectares to TAZ, but also partly from the subset of households from the census included in the UrbanSim data. Since UrbanSim models the location choices of households, "collective households" (prisons, psychiatric facilities, etc.) were not included in the baseyear data whereas they were in the original EMME transport model. As a result, some work would need to be done to either make the population figures consistent with the transport model figures, or the EMME generation step will need to be recalibrated with the UrbanSim data.

Moreover, effort will need to be spent on improving development constraint estimates. Currently, maximum residential units and non-residential surface area estimates have been derived as being +2 standard deviations of the average, per plan type, per commune. These estimates do not seem to be constraining enough given the exaggerated densification predictions from the model. Future work would need to be done to find better values for these development constraints.

Inclusion of missing data will go a long way to helping to improve the location choice and land-price models. It will allow the inclusion of more explanatory variables (e.g. non-residential improvement value within walking distance). The inclusion of new variables will probably not be sufficient to improve the submodels. In addition to the testing and addition of new variables, more disaggregated models should be developed. In particular, separate models should be developed for industrial sector subgroups to replace the single model currently used for industrial jobs. The household location choice model should be further refined to include more detailed information on households (e.g. income interaction effects for various explanatory variables). Finally, time should be spent in formally calibrating the models. That is, the models should be estimated using 80% of the observations and then used to compare model predictions with actual outcomes from the remaining 20% of the data.

Naturally, there will need to be interaction with experts (either at the EPFL or in local administrations) to evaluate the quality and realism of model predictions. Observations of these experts will likely result in identification of model weaknesses and result in re-evaluation and improvement of data and model weaknesses in a non-linear, iterative process. It would also be judicious to validate the overall UrbanSim model against actual data. One possibility would be to use 1990 census data to create a new 1990 baseyear. With the 2000 data, it would then be possible to compare model predictions with actual land-use manifestations. Such an endeavor, however could add considerable time to the project, probably 4-6 person months.

Based on the experience developing this model, as well as with the knowledge of data requirements for a fully fledged model, it is estimated that the prototype model could be improved to a state that it could be used for planning purposes with 1-1.5 person years of work. This of course does not include the amount of time that may pass in waiting for access to the various data required for an operational model.

9 Conclusion

The purpose of this document was to report on the development of the prototype UrbanSim model for Lausanne. Its purpose was to evaluate whether a prototype UrbanSim model for the Lausanne-Morges region could be developed given available data and limited human resources. The main conclusion of the report is that it was indeed possible to develop such a model given the data and human resource constraints. In fact, the entire prototype model was developed using roughly four person-months of work.

The prototype model was put together despite the fact that a fair bit of required data was not available for use in the model. Despite this lack of data, the location choice and land-price submodels, and the overall model itself perform remarkably well when compared with actual population growth rates by commune until 2007. It has been shown how the model can be used to test different transportation and land-use scenarios into the future and that the current model tends to overpredict densification in the central part of the Lausanne-Morges region.

An operational model would require a significant amount of data that is not readily-available including among others, land-price, improvement value and non-residential buildings characteristics. As well, effort would need to be placed in analyzing already available data (e.g. geographical characteristics of hectares, development constraints, etc.). In addition to improved data, more analysis is required to fine-tune the location choice and land-price submodels. Before being ready for planning purposes, the model would need to be evaluated by local experts to ensure the realism of its model results. Naturally, this process will not be linear but would involve iterative improvement as weaknesses are discovered and corrected along the way. Given these requirements and the experience of having developed this prototype model, it is expected that 1-1.5 person-years of effort would be required to improve the current model so that it could be used for planning purposes.

References

Administration Cantonale Vaudoise (2008). Zones d´affectation du sol du canton de Vaud, *Website*, Administration Cantonale Vaudoise. Information available at: `http://www.asit.vd.ch/geoportal/geocatalog/public.asp`.

Alonso, W. (1964). *Location and Land Use*, Harvard University Press, Cambridge.

Baranzini, A. and Ramirez, J. V. (2005). Paying for quietness: The impact of noise on rents in Geneva, *Urban Studies* **42**(4): 1–14.

Bettex, L. (2008). Analyse géographique pour l'implémentation d'un prototype de modèle intégré de transport et occupation du sol pour la région lausannoise, *Technical report*, EPFL - Transport and Mobility Laboratory.

Hunt, J., Kriger, D. and Miller, E. (2005). Current operational urban land-use transport modelling frameworks: A review, *Transport Reviews* **25**(3): 329–376.

Milton, J. and Mannering, F. L. (1998). The relationship among highway geometrics, traffic-related elements and motor vehicle accident frequencies, *Transportation* (25): 395–413.

Muth, R. (1969). *Cities and Housing*, University of Chicago Press, Chicago.

Nguyen-Luong, D. (2008). An integrated land use-transport model for the Paris region (SIMAURIF): Ten lessons learned after four years of development, *Technical report*, Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France (IAURIF).

Office fédéral de la statistique Suisse (2008a). Recensement fédéral de la population, *Website*, Office fédéral de la statistique. Information available at: `http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/erhebungen__quellen/blank/blank/vz/uebersicht.html`.

Office fédéral de la statistique Suisse (2008b). Recensement fédéral des entreprises, *Website*, Office fédéral de la statistique. Information available at: `http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/erhebungen__quellen/blank/blank/bz/01.html`.

Office fédéral de la statistique Suisse OFS (2008). La mobilité en suisse: Résultats du microrecensement 2005 sur le comportement de la population en matière de transports, *Website*, Office fédéral de la statistique, Neuchâtel.

Ortuzar, J. and Willumsen, L. (2001). *Modelling Transport*, third edn, John Wiley, Chichester.

Patterson, Z. and Bierlaire, M. (2007). An UrbanSim model of Brussels within a short timeline, *Proceedings of the 7th Swiss Transport Research Conference*, Monte Verità, Switzerland.

Patterson, Z. and Bierlaire, M. (2008). Development of prototype UrbanSim models, *Technical Report TRANSP-OR 080814*, Transport and Mobility Laboratory, EPFL, Lausanne, Switzerland. Available at: http://transp-or2.epfl.ch/publications.php#techrep.

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. and Nourzad, F. (2002). Incorporating nonmotorized mode and neighborhood accessibility in an integrated land use and transportation model system, *Transportation Research Record* **1805**(1): 119–127.

Waddell, P. and Ulfarsson, G. (2003a). Accessibility and agglomeration: Discrete-choice models of employment location by industry sector, *82nd Annual Meeting of the Transportation Research Board, January 12-16, 2003, Washington, D.C.* .

Waddell, P. and Ulfarsson, G. (2003b). Dynamic simulation of real estate development and land prices within an integrated land use and transportation model system, *82nd Annual Meeting of the Transportation Research Board, January 12-16, 2003, Washington, D.C.* .

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.