A MICROSIMULATION MODEL FOR RESIDENTIAL MOBILITY:

AN APPLICATION TO THE CITY OF HAMILTON

A MICROSIMULATION MODEL FOR RESIDENTIAL MOBILITY:

AN APPLICATION TO THE CITY OF HAMILTON

By

YIFEY WANG, B.E.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2009)	McMaster University
(School of Geography and Earth Sciences)	Hamilton, Ontario

TITLE:	A Microsimulation Model for Residential Mobility: An Application to the City of Hamilton
AUTHOR:	Yifei, Wang, B.E. (McMaster University)
SUPERVISOR:	Dr. Pavlos S. Kanaroglou
NUMBER OF PAGES:	VIII, 115

ABSTRACT

URM-MicroSim is a prototype system for a microsimulation model of urban residential mobility. It is developed for the city of Mytilene, Greece. However, it is only a prototype and fails to meet the requirements of practical use, especially with regard to execution time. Therefore, a comprehensive analysis is first presented to fully understand the required improvements to the existing system. These are divided into functional and non-functional requirements, which are discussed separately. On the basis of the analysis, several functions (such as user interface and logging system) have been implemented and the time consuming functions were indentified and revised without affecting the simulation results. The revised system was tested for consistency in performance, and the results were convincing.

Within this context, URM-MicroSim is calibrated for the city of Hamilton. The calibration methods include identifying the probabilities of demographic events and rebuilding the immigration sub-model. After URM-MicroSim is applied for Hamilton, simulation results from the system are validated against census data from Statistics Canada. Results from the validation provide evidence that URM-MicroSim is able to capture the overall trend of residential mobility at both aggregate and disaggregate levels.

Lastly, some directions for future research are indicated, that focus on reducing system execution time and broadening the scope of the model.

ACKNOWLEDGMENTS

I would like to thank my supervisor Pavlos Kanaroglou. Thank you for your limitless support and patience. Your helps, both the knowledge of modeling and statistics and the guidance of being a good researcher, will benefit me throughout my life.

I would like to thank Hanna Maoh and Leyden Martinez-Fonte for all the knowledge and helps to my research and this thesis whenever I needed them.

I would like to thank my colleague Justin Ryan for your assistance throughout my thesis as well as creating the synthetic population, which is critical to this thesis. I would like to thank my colleague Mari Svinterikou for creating the URM-MicroSim model. Without this model, my research in this thesis would never go this far.

I would like to thank all the members in the Centre for Spatial Analysis for their supports and friendships. Special thanks go to Manolis Koronios and Konstantinos Vassos for your sincere friendship, pleasant day-to-day company, and all the fun we had.

I would like to thank my girlfriend, Jia Liu. Thank you for your support and tireless revision to my thesis. Especially, your patience and encouragement helped me overcome difficulties and make it all the way to the submission of this thesis.

Finally, I would like to thank all my friends during the period of my graduated studies. All your friendships enriched my life in McMaster University.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 3

CHAPTER 4

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

Chapter 1: Introduction

1.1 Research Context

Today, urbanization has led to vehicle dependency and transportation-related problems in most countries. A few integrated land use and transportation models have been implemented that provide systematic support to urban planners in understanding and circumventing these problems. By aggregating individuals into groups, these models aim to simulate the interrelation between land use and transportation and present an overview of different policies. However, it is argued that these aggregate models do not generate reliable representations of the complex behavioral relationships (Svinterikou, 2007). The current state-of-the-art approach, microsimulation, operates at the level of individual behaviors and can capture the diversity among individual entities. It is regarded as one of the most useful modeling approaches available for the study of the potential futures of metropolitan areas (Citro and Hanushek, 1991; OECD, 1996; Brown and Harding, 2002).

In Canada, the city of Hamilton, Ontario, has severe problems, many of which are caused by urban sprawl, such as higher commuting time (car-based) and more vehicle emissions. Therefore, a group of researchers at the Centre for Spatial Analysis, McMaster University, have dedicated themselves to urban modeling so as to better understand urban issues. In particular, their work on the spatial behaviors of firms, individuals, households, and property developers has been taken forward by several researchers in the Centre.

Svinterikou (2007) developed a prototype microsimulation model, named URM-MicroSim, to understand the residential mobility and housing market changes in the city of Mytilene, Greece. URM-MicroSim is well structured and has all the necessary components, even though the system is only a prototype and fails to meet the requirements of practical use, especially with regard to execution time. To meet the data input needs of URM-MicroSim, Ryan (2008) studied population synthesis techniques and created synthetic populations specific to different cities. Maoh (2005) developed a framework for an agent-based microsimulation model for Hamilton, Ontario, to simulate the evolution of small- and medium-sized business establishments. Koronios (2009) studied the key spatial factors that affect the choices of land developers regarding the construction of certain housing types at specific locations in Hamilton. Based on all the listed research, Hamilton's severe urban problems have led to an increased demand for an integrated urban microsimulation model for the city.

1.2 Research Objective

Within this context, the purpose of this research is to implement and test the microsimulation model URM-MicroSim for the city of Hamilton. To this end, the specific objectives of this research are as follows:

1. The execution time of URM-MicroSim is to be minimized. Moreover, the system should run smoothly without returning any unexpected errors and it should produce consistent results.
2. URM-MicroSim must be calibrated to the city of Hamilton. The calibration method must be based on valid data. The system should be able to capture the overall trend of urban residential mobility for Hamilton at both aggregate and disaggregate levels.

1.3 Thesis Structure

This thesis consists of five chapters. In Chapter One, the research context and overall objectives are described.

Chapter Two starts with a review of available literature on the history of integrated urban models. It proceeds with an evaluation of literature on the study and exploration of the microsimulation model and the residential mobility process. Finally, it provides an overview of existing residential mobility microsimulation models.

Chapter Three focuses on why and how the existing URM-MicroSim model for the Greek city of Mytilene can be revised. A comprehensive analysis is first presented to fully understand the required improvements in the existing system. These are divided into functional and non-functional requirements, which are discussed separately. Then, methods of applying these improvements are presented. Finally, the revised URM-MicroSim for Mytilene is tested for consistency in performance.

In Chapter Four, the revised URM-MicroSim is adapted to the context of Hamilton. The probabilities concerning the different demographic events are collected and the immigration sub-model is rebuilt. Then, with the adapted model, different approaches are adopted to test the simulation results with census data from Statistics Canada. The results are discussed in the context of the comparison.

Chapter Five begins with an overview and the conclusions of the work. These are followed by a discussion of the limitations of this research and the possible directions for future research.

Chapter 2: Research Background and Literature Review

2.1 Introduction

The purpose of this research is to implement a residential mobility microsimulation model under the framework of an integrated land-use and transportation urban model for Hamilton, Ontario. We will discuss the historical development of integrated urban models in Section 2.2. Then, residential mobility models used to project the spatial distribution of individuals and households over time are examined in Section 2.3. Section 2.4 discusses the modeling method, microsimulation, which is considered as the state-of-the-art. Lastly, Section 2.5 reviews several microsimulation models, implemented to simulate residential mobility.

2.2 Integrated Land-use and Transportation Models

2.2.1 Introduction

As people are more dependent on vehicles nowadays, transportation-related issues are more prominent than ever before. Congestion, emissions, and increasing consumption of gasoline have a negative effect on national economies, the environment, and society as a whole. Increasing road capacity is found to be an insufficient solution. It is important to understand why people depend on vehicles in order to provide solutions to the resulting problems.

Researchers find that changes in land-use patterns in cities result in fairly predictable impacts on transport demand (Yunquan et al., 2006). The different land-use patterns in the city determine people's spatial activities, which generate the demand for travel. Meanwhile, spatial activities over the transportation network affect the accessibility of travel destinations. Land-use and transportation are mutually interdependent (Moore and Thorsnes, 1994). Figure 2-1 (Wegener, 1995) shows the circular interrelationship between transportation and land-use. In the cycle, land-use and transportation are interdependent through accessibility and activities. This provides a good understanding of the behavior of actors in the urban system while also taking into account the natural processes that take place in it (Kanaroglou and Scott, 2002).

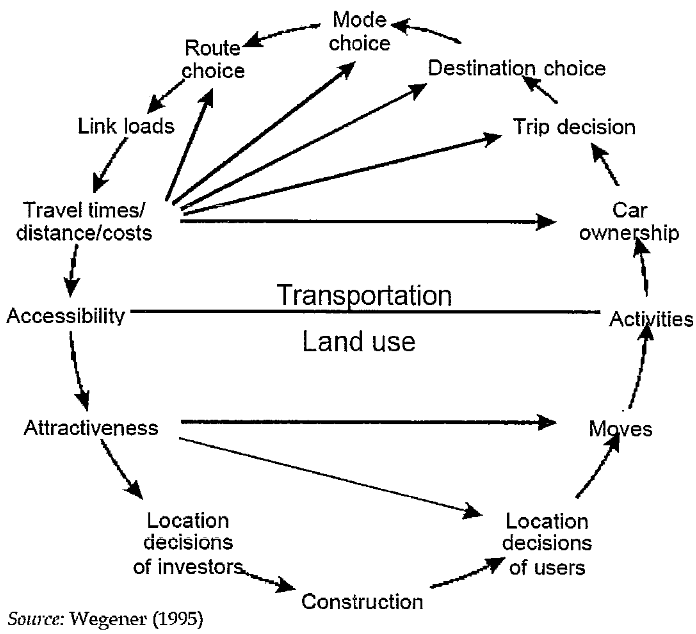

Figure 2-1: The land-use/transportation feedback cycle (Wegener, 1995)

Within this context, the history and future of integrated transportation and land-use urban models are discussed in this section, with a particular focus on the land-use component.

2.2.2 Integrated land-use and transportation Models

In 1960s, the model of land-use changes, as basis of transport planning process, was first studied. Lowry's model was developed in 1964 for the Pittsburgh region (Lowry, 1964). It is based on the principle that urban growth is an expansion of employment in the basic sector, having an impact on retail and residential sectors. The spatial distribution of residences is determined by the place of employment. At the same time, the place of retail service is determined by the resident population. The retail service produces additional employees, which, in turn, affect the distribution of residences. As the size of population in a residential neighborhood increases, more retail services are attracted. This iteration continues until equilibrium is reached. Lowry's concepts are elaborated upon on the basis of this principle in a series of reports (see Harris, 1965; Goldner, 1971). Since Lowry's model works at the zonal level, it is the first large-scale land-use simulation model to become operational (Weiner, 1999). It provides a foundational framework for integrated urban models.

However, a Lowry type model is not based on the economic theories of cities or regions; rather, it is physical in nature (Anas, 1987). From the 1970s, several models stand out as they are based on a number of well-established economic theories.

Putman's (1983) ITLUP (Integrated Transportation Land-Use Package) is considered to be the first fully operational integrated transportation and land-use package (Timmermans, 2003). The land-use model combines two sub-models: EMPAL (Employment Allocation model) and DRAM (Disaggregated Residential Allocation Model). The EMPAL sub-model forecasts the future spatial distribution of employment by different sectors, while the DRAM forecasts the distribution of households. However, ITLUP does not realistically reflect the role of households, business establishments, developers, or governments in making decisions about their respective spatial choices in the development process (Yen, 2003).

MEPLAN, an integrated modeling package, was developed through a series of studies in different countries by Echenique and Partners. It started with a model of stock and activities (Echenique et al., 1969), followed by the incorporation of a transport model developed for Santiago, Chile (de la Barra et al., 1975). It incorporated an economic evaluation system for Sao Paulo, Brazil (Flowerdew, 1977), and represents market mechanisms in the land-use model for Tehran, Iran (Hirton and Echenique, 1978). The core of the system is an input-output model to predict the change in demand over space. An input-output model created by Wassily Leontief is widely used in economic forecasting to predict flows between economic sectors. The economic input-output method simulates the flow of spatial activities, which is used to relocate the land uses. It has been applied to over 25 regions throughout the world, including Sacramento, California and the Cross- Cascades Corridor in the United States (see Echenique et al., 1990; Echenique and Williams, 1980; Echenique, 1985; Hunt and Echenique, 1993; Hunt and Simmonds, 1993; Hunt, 1994).

TRANUS (the transportation and land-use model), similar to MEPLAN, was implemented by de la Barra in the early 1980s. The land-use component outputs the location of activities, which consume the spaces, and land prices. A distinction is made between demand and supply elements. The interaction of activities represents the demand side for housing in the activities subsystem. The real-estate market represents the supply side to provide building space for activities. These two elements interact and generate a state of equilibrium. In TRANUS, the interaction between activities and transport takes place with the aid of origin-destination matrices with the economic flows by sector. These matrices are produced by activities being input into the transport model (see de la Barra, 1982, 1989; de la Barra et al., 1984). One distinctive innovation brought about by TRANUS is the introduction of random utility theory (McFadden, 1973). Although MEPLAN borrows this theory for use in its transport models, TRANUS implements this into all components of the urban/regional and transport system, from trip generation to mode choice, path choice, location choice, land-use choice, and others. TRANUS is a long chain of linked discrete choice models (Modelistica, 2007).

Random utility theory predicts choices between different alternatives as a utility function of characteristics of each alternative. Differences in taste between the decision makers, unobserved characteristics of the alternatives and uncertainty or lack of information could be captured by the stochastic nature of the model (Domencich and McFadden, 1975). After random utility theory is introduced to identify behaviors, it becomes the theoretical basis of the state of practice models. The California Urban Futures Model (CUF), initially known as the Bay Simulation Model (BASS), is developed by Landis (Landis, 1994, 1995). This model uses multinomial land-use logit models to simulate the land-use changes. It was later updated as the CUF II, the second generation California Urban Futures Model (Landis and Zhang, 1998).

MUSSA is a land-use model developed to interact with the Santiago four-stage transport model called ESTRAUS (Martinez, 1992, 1997). The model implements bid-rent theory to predict the location of households and firms. A multinomial logit model is used to estimate the bid choices. In the city of Chicago, CATLAS (Chicago Area Transportation and Land-Use Analysis System) was developed by Anas (Anas, 1983). It synthesizes the knowledge of "location rent analysis" from urban economics with the knowledge of "travel demand analysis" from transportation planning (Anas and Duann, 1986). Two sub-models comprise CATLAS: the demand sub-model of commuters employed over space and the housing supply sub-model. Both are derived as multinomial logit models. Similar models have been developed recently, such as IMREL (Anderstig and Matsson, 1991, 1998), TILT (Eliasson, 2000), UPLAN (Johnston et al., 2003), IMULATE (Kanaroglou and Anderson, 1997), and IMPACT (Maoh et al., 2009). Of these models, IMULATE and IMPACT are calibrated for the city of Hamilton, Ontario, which is the focus area of this research.

IMULATE (Kanaroglou and Anderson, 1997) is an operational integrated urban model calibrated to study urban issues in the Hamilton Census Metropolitan Area (CMA). IMULATE enables researchers to assess the impact of land-use changes and related environment problems. Working at the zonal level, IMULATE is a typical aggregate model that simulates urban changes over five-year intervals starting from 1986. The land-use sub-module consists of two components: POPMOB and EMPLOC. POPMOB, the population mobility model, relates to intra-urban population mobility and the operation of the housing market. It can predict the changes in the distribution of households based on the number of new dwelling units added in each tract. EMPLOC, the Employment Location Model, relates to intra-urban firm mobility and location choice behaviors in the city, which can predict the changes in the spatial distribution of firms by sector.

IMPACT is an integrated model for population aging consequences on transportation that was developed to assess the population aging problem, the effects of which are believed to be more pronounced in urban areas of developed countries (Maoh. et al., 2009). Although the model is still aggregate, based on zones, it has a powerful demographic model that can project the future spatial distribution of people by five-year age groups (etc. 21 to 25, 26 to 30). During the process, vital statistics such as fertility, mortality and migration rates are taken into account inside the model.

2.2.3 Future integrated urban models

Until recently, academics and policy makers developed urban models using conventional aggregate modeling techniques that aggregate individuals into zones to represent a population (Wegener, 1994). It is argued that aggregate-models do not generate reliable representations of the complex behavioral relationships (Svinterikou, 2007). Moreover, the field of modeling land-use and transportation has been consistently criticized for its complexity and black-box characteristics (Timmermans, 2003).

Model development depends heavily on the availability of data and computer technology. Nowadays, micro data have been widely collected and are available to researchers in developed countries. Meanwhile, computational speed is increasing sharply. Therefore, the discrete choice model, one of the disaggregate approaches, is used extensively to study the individual behaviors. A number of researchers have applied a modeling technique called microsimulation for the development of integrated urban models at the disaggregate level to simulate spatial activities that are specifically generated by land-use patterns. Inside Wegener's cycle, the land-use patterns are comprised of the decisions of two actors: investors and dwellers. Investors decide whether to build a building and where to build it, while dwellers decide whether to buy a new dwelling and where to live. The land-use model generally deals with describing spatial activities of land consumers and suppliers, including individuals, households, firms, and investors. These actors, considered to be agents in a microsimulation model, interact with each other and generate spatial activities, which could be captured in a microsimulation model. In the review of integrated urban models, Timmermans concludes that, after the 1990s, the microsimulation method is emerging and becoming the state of the art within the urban models (Timmermans et al., 2003). This technique has the potential to significantly improve the modeling performance of the integrated urban model and is discussed in Section 2.4.

2.3 Residential Mobility

2.3.1 Introduction

Residential mobility refers to the spatial movement of individuals and households between dwellings within an urban area. It has been estimated that each year almost one-sixth of individuals in the United States and Canada move to another place of residence, and two-thirds of them move within municipality (Short, 1978; Ryan, 2008). This considerable amount of intra-urban movement affects the urban structure and has significant repercussions for urban transportation. Geographers, sociologists, economists and psychologists have contributed extensively to the literature on residential mobility and its effects on the urban fabric (Dieleman, 2001).

Brown and Moore (1970) divide the mobility process into two stages: decision of move and the choice of location. Firstly, people become dissatisfied with the location of their present home because of changes to the housing environment or their life-cycles. This stresses the households and forces residents to make the decision to either relocate or stay at the present location. The studies of these stages are separately discussed in Sections 2.3.2 and 2.3.3 along with the introduction of residential mobility modeling in Section 2.3.3.

2.3.2 Why do families move?

Rossi (1955) conducted some fundamental research to explain why families move. He focused on the individual household and its motivation to migrate. He believed that the reason why households move has a strong relation with the stage of the family life-cycle. "Households change in a more or less regular way in response to demographic processes — births, deaths, marriages, and divorces — and that the time-related character of such processes constantly shifts the size and age composition of members of the household" (Rossi, 1980). McCarthy (1976) explains that there is a strong relationship between housing consumption and progression through the life-cycle. According to this life-cycle theory, a household will go through these vital stages in its lifetime. During these processes, changes will apply to the structure of the households. Other studies have shown that changes in family size and housing characteristics are major determinants of short-distance residential relocations (Clark and Huang, 2003; Feijten and Mulder, 2002; Mulder and Wagner, 1998). Some life events, like union formation and union dissolution, may lead to migration and residential mobility (Clark and Huang, 2003; Mulder and Wagner, 1993). For households, these changes are inter-dependent on each other. Changes occurring in one dimension of the household process are necessarily linked to changes in other dimensions (Clark and Dieleman, 1996). Finally, households decide to give up their current dwellings for more desirable alternatives. Overall, the spatial distribution of demographic characteristics is a major factor in differentiating urban mobility rates (Simmons, 1968).

However, changes in a family's life cycle do not fully account for household mobility (Kendig, 1984). The demographic changes are also connected with social and economic factors, such as changes to education and income. Research has shown that a job change in the local housing and labor market often acts as a trigger for a residential move (Clark and Withers, 1999). Long-distance moves are mostly provoked by education-related factors (Kulu and Billari, 2004).

2.3.3 Where do families go?

The decision to move to a new dwelling depends on both demand conditions and supply constraints (Simmons, 1968). The demand conditions refer to the first stage introduced in the previous section, which includes life-cycle, family size, and place of employment (Clark, 1980). The supply constraints refer to characteristics of the dwelling including its neighborhood conditions.

In the case of housing characteristics, tenure, physical attributes, and neighborhoods (Spear, 1974; Brown, 1975) are widely discussed in the literature as the three major components. The decision to own or rent a dwelling unit contributes a great deal of the variation in location choices (Boehm, 1980). It is also pointed out that renters are considerably more mobile than owners (Rossi, 1955). Physical attributes of housing are extensively studied in literature (McCarthy, 1976; Goodman, 1976; Adair et al., 2000; Clark et al., 2006). Among all of these housing characteristics, age of the head of the house and housing size are usually found to be the dominant factors (Rossi, 1955; Goodman, 1976). It is generally agreed that neighborhood attributes also have an influence on residential mobility. Comprehensive studies have been conducted on these attributes: neighborhood quality (Speare, 1974), social composition (Simmons, 1968), neighborhood relationship (Connerly, 1986), public service and accessibility (Clark, 1982).

Psychological aspects are also captured in the decision-making process. The urgency of the move is considered a factor for the willingness to substitute a current dwelling. The desire to move may be driven by an urgent event such as job change, marriage, divorce or death of family members. In a study of several regions, Hooimeijer and Oskamp (1999) conclude that the urgency of a move is important to the substitution of dwellings.

2.3.4 Residential mobility modeling

Residential mobility models have evolved over time and early models were mostly from two fields: economic models were first introduced by Alonso (1960, 1964) and sociological models were first introduced by Wolpert (1965, 1966). The development of economic models is mainly derived from the equilibrium model of the housing market (e.g., Goodman, 1976). This equilibrium is established by balancing the supply and demand. On the other hand, the sociological models relate more to the life-cycle of households. They tend to be interested in the spatial behavior of individuals. However, these behaviors are studied at a disaggregate level. Particularly for movement behaviors, isolating spatial patterns of movements and discovering regularities in them at an aggregate level poses a significant problem (Golledge, 1980).

More disaggregate census data are becoming available to researchers nowadays. As a result, residential mobility models are in the process of being further developed.

Microsimulation models have been applied by a number of researchers. This technique has the potential to significantly improve the modeling of the residential mobility process (Svinterikou, 2007). Also, Hooimeijer and Oskamp (1999) believe that microsimulation is especially appropriate to evaluate the housing decisions of individuals and households, which naturally constrain choice. Microsimulation residential mobility is discussed in Section 2.5.

2.4 Microsimulation

2.4.1 Introduction

Microsimulation models were first introduced by Orcutt (1957), who found that the socio-economic model failed to predict the distribution of individuals, households, or firms. He introduced a new type of socio-economic system that works at the disaggregate level. Orcutt divides the model into four parts: units, input, output and operating characteristics (Orcutt, 1957). Units, also known as agents, are identified by their characteristics such as age, gender, marital status of an individual; size and type of a household; and type of employment provided by a firm. These distinct characteristics of agents specify the probabilities of various outputs, such as birth, marriage, divorce and death. This makes it necessary to model them individually. Mitton summarizes the situation thus: "Microsimulation models use micro-data on persons (or households, or firms or other micro-units) and simulate the effect of changes in policy (or other changes) on each of these units. Differences before and after the changes can be analyzed at the micro-level; and also aggregated to show the overall effect of the changes. It is the dependence on individual information from the micro-data at every stage of the analysis that distinguishes microsimulation models from other sorts of economic, statistical or descriptive models" (Mitton et al, 2000).

2.4.2 Classification

In general, these microsimulation models can be divided into two types: static and dynamic. These two types are distinguished by the time sequence of changes. Micro-data, which usually comprises a cross-sectional table, is stored with certain behavioral relations and institutional conditions (Merz, 1991). A static microsimulation uses cross-sectional micro-data that has been updated to the required point of time and aging techniques are normally not used. Dynamic models are more complicated in nature. To accurately represent the real world, the temporal element is introduced into the model. Agents are aged and undergo transitions such as stochastically being subject to different sophisticated policy analysis. The agents within the original micro-data are progressively modeled. Their characteristics are recalculated over time.

Dynamic models have several advantages over static models. First, the ability to age the agents can be used to examine inter-temporal issues. Second, dynamic microsimulation models can simulate the effort of structural changing patterns, such as age, employment, and income. Both of these advantages make dynamic microsimulation a strong tool to project future development. Another main advantage is that dynamic models allow the integration of diffident processes, such as housing markets and demographic models (O'Donoghue, 2001).

However, with all these advantages, relatively few dynamic microsimulation models have been developed and employed worldwide for economic and social policy analyses (Merz, 1991). This is because such models are costly to develop and require significant effort because they are more complex than static models. For dynamic models, incorporating behavioral responses in a time sequence will alter the nature of the transition probabilities that are used to age the agents (Mitton, 2000). Therefore, incorporating such a model requires more data and effort. For static models, the aging of interactions among members of different agents (like marriage and market process) are omitted. This makes static approaches less expensive and less time-consuming. Over the past two decades, microsimulation models have become widely used in many governments in a routine manner to analyze the impact of policy on tax and cash-transfer programs (Brown and Harding, 2002). The complex demographic aging of interactions does not need to be simulated. As a consequence, even until 2001, Creedy (2001) summarizes that most microsimulation models are still static. No attempt is usually made to model the dynamic changes due to the complexity.

2.4.3 Applications

In recent years, microsimulation models have been increasingly applied to quantitative analysis and social policy problems. Such models have frequently played a decisive role in determining whether a particular policy is implemented (Brown and Harding, 2002). In 2005, the International Microsimulation Association (IMA) was founded by an international group of researchers for different research purposes. The IMA aims to provide a greater interchange of ideas and expertise. This conference brought together experts from around the world, such as governmental (policy makers and model developers) and intergovernmental organizations, academics and consultants. Their primary focus is on microsimulation models and their applications (International Microsimulation Association, 2007). Different applications of microsimulation models have been compiled, but research is largely devoted to three fields: population-based, traffic-based, and firm-based analysis.

Most of the models are implemented for population-based analysis. These applications are motivated by long-term governmental programs such as pensions, health, and long-term care and educational financing. A number of models such as DYNAMOD, the SfB3 cohort model (Hain and Helberger, 1986), and LIFEMOD (Harding, 1993) have been used to examine changes in education finance, allowing for the payment of education costs. Fölster (1997) used DYNASIM3 to analyze the long-term distributional consequences of retirement and aging issues (Favreault, 2004). EUROMOD (Immervoll et. ac., 1999) is a tax-benefit microsimulation model covering all 15 European Union countries. It is currently being developed by a team from 18 institutions, coordinated by the Microsimulation Unit of the Department of Applied Economics at the University of Cambridge. These tax-benefit models are recognized by urban planners and demographic researchers since they try to understand the behavior of individuals and households. Meanwhile, several residential mobility microsimulation models have also been implemented, which will be discussed in Section 2.5.

Most traffic-based models focus on transport planning purposes, which represent transport trip mechanisms. A traffic-based microsimulation simulates the choice of where and when to travel, the choice of mode, and the nature of the journey at a microscopic level on the basis of the available transportation network. Several models are listed here. Aimsun, which stands for advanced interactive microscopic simulator for urban and non-urban networks, was developed in 1986 as part of a research project carried out by LIOS, a research group at the Technical University of Catalonia (Barceló and Casas, 2005). A compatible traffic microsimulation model called DRACULA (Dynamic Route Assignment Combining User Learning and Microsimulation) has been developed at the Institute for Transport Studies, University of Leeds (Liu et al., 1995). For the same purpose, PARAMICS (SIAS) is a car-following and lane-changing model developed in

1992-1997. It can perform versatile and comprehensive analysis of road traffic flow (Duncan, 1997). Similar to PARAMICS (SIAS), PARAMICS (QUADSTONE) is another traffic microsimulation model with 3-D visualization (Quadstone Limited, 2000). SIMWALK is an agent-based pedestrian simulation software designed to analyze pedestrian traffic flow and improve public transport management in the real world. It is mainly focused on station design and related subjects (Siebers et al., 2007). TRANSIMS is a travel forecasting model system that was originally developed by the Researchers at the Los Alamos National Laboratory (LANL). It can furnish transportation planners with the complete and accurate information necessary to deal with the impacts of traffic, congestion, and pollution (Smith et al., 1995).

To deal with government tax policy, several firm-based microsimulation models have been designed to capture the life-cycle development of firms. Several economic indicators such as CPI and GDP have been interpreted in these models as well. DIECOFIS is aimed at microsimulation purposes, which focus on enterprises and economic activity. The model is based on a multi source, integrated database of cross-section and longitudinal microdata from enterprise administrative registers and surveys (V. Parisi, 2003). MSMNE-02 is a multi-agent microsimulation model designed by Danylo Kozub and developed as an MS Excel program (Danylo, 2003). XEcon (for eXperimental Economy) is an experimental economic theoretical growth model. In contrast with other models, XEcon is designed to analyze more from a theoretical aspect than as a practical application (Wolfson, 1995).

In conclusion, microsimulation models are applied to simulate the behaviors of individuals, individual vehicles, and individual firms. These three components could be naturally combined to portray the sophistication of urban policy under the framework of integrated land-use and transportation models.

2.4.4 Pros and cons of microsimulation model

Microsimulation models are considered state of art modeling techniques. O'Donoghue (2001), Burtless (1996), Harding (1993), Orcutt et al. (1980), and Zaidi and Rake (2001) outline several advantages of microsimulation models over the traditional aggregate approaches. These advantages are summarized as follows:

First of all, the principle of modeling directly the behavior of decision makers makes them intuitively appealing. Agents in the aggregate models are simply statistically averaged or represented by a small fraction of the population, whereas these individuals, households, and firms are actually very diverse in their age, size, income, and life-course events. As a result, the decision rules developed in behavior theories could be easily incorporated into the models.

Second, due to data aggregation, the macro approach tends to suffer from data loss and abstraction (Orcutt et al., 1986). Unlike the macro approach, the results produced from microsimulation models contain detailed information in a disaggregate form. This allows the models to capture the diversity of individuals. The researchers can analyze the urban process of the different policies at different scales (individuals, parcels, communities, census) based on various research purposes. Also, it allows the researchers to study the distribution of micro units over space or different groups.

The third advantage is associated with inferential errors. Within the discussion between macro and micro approaches, Zaidi and Rake (2001) conclude that the "microsimulation method involves less of a chance in making inferential errors due to heterogeneity. Aggregation tends to mask a lot of the variation that is present within data. As a result, this 'hidden' heterogeneity is often mistaken for substantive relationships between variables. In microsimulation, the results are in a disaggregate form and, therefore, the variation in the results is not masked. This situation reduces the chance of drawing spurious conclusions."

The final advantage of microsimulation models is that the behavior of individuals can be incorporated with corresponding behavior theories, which is studied in the level of the individual, solely focusing on the process of decision making. This incorporation can increase the validity of the model and better serve future projections.

However, microsimulation models have some limitations that restrict practical application. The major discussion concerns computational speed and data availability. In

1957, when Orcutt first introduced microsimulation models, he noticed that the speed and capacity of available computers would limit the simulation process. However, he believed that the fantastic rate at which the speed and capacity of computers was increasing could make microsimulation models practical in five to ten years (Orcutt, 1957). But this is hardly true in reality: "Although the costs of development declined over the 1980s, as computing power and data availability increased, resulting benefits from these models did not match what was expected. It is unsurprising given the data and computing resources available at the time that results were less useful than policy makers would have liked" (O'Donoghue, 2001). As discussed in Section 2.4.2, only static models have been developed extensively. It is still hard to implement dynamic models for a large sample. However, this problem will become less important as computer technology keeps improving. In addition, data availability is becoming less of a concern in microsimulation. Panel surveys conducted by census offices should be available in many developed countries (O'Donoghue, 2001). And nonetheless, even with limited data, synthetic populations remain an alternative.

2.5 Residential Mobility Microsimulation Models

2.5.1 Introduction

At the micro level, residential mobility models are meant to study the location behaviors of individuals who need to incorporate the decision-making process with the behavior theory (see Section 2.4.4). In the literature, several microsimulation models have been developed to capture residential mobility. After a comprehensive review to these models, Svinterikou (2007) summarizes them into two major categories: location-choice models and demographic and housing market models (Svinterikou, 2007).

2.5.2 Location-choice models

Location-choice models represent an important range of integrated urban models that describe the location behaviors of an individual or the behavior of a household in settling at a location. For these models, the demographic changes of individuals and households are stored and iterated in a transition model or are provided exogenously (Svinterikou, 2007).

UrbanSim (Waddell, 2001, 2004, Waddell et al., 2003) is an open source platform for urban models. (Source code available at http://www.urbansim.). Because it is open source, it has received a fair bit of attention and is implemented in a couple of regions (Friedman et al., 2008). In this model, residential mobility is divided into two parts: demographic transition model and household location-choice model. The demographic transition model is based on exogenous total population and households by type to provide a mechanism for users to approximate the net results of spatial population changes. On the basis of these changes, the location-choice model is specified as a multinomial logit model. With random sampling of alternatives from all vacant dwellings, decisions are made mainly on the basis of utility functions of housing characteristics, regional accessibility and neighborhood quality.

ILUMASS is a fully microscopic integrated urban model designed to study land-use, transportation, and the environment, and it is an integration of several existing models (Moeckel et al., 2002). Although it has been claimed that the prospective objective is not achieved, the land-use sub model performs well (Wagner and Wegener, 2007). The land-use component of the models is IRPUD (Wegener and Spiekermann, 1996), which is claimed to be a microsimulation model. However, the housing market and ageing sub-model, which predicts the distribution and characteristics of individuals, households, and employment, works at the zonal level. These individuals are aggregated together by types and characteristics, which are updated by household events such as household formation and household dissolution with the passage of time. After the aging takes place, the housing market model predicts the probabilities of movement and location choice using utilities theory.

Compared with IRPUD, ILUTE (Miller et al., 2005) is a more disaggregate integrated urban model. The ILUTE project aims to simulate the evolution of an entire urban region over an extended period of time using the dynamic microsimulation method. The land-use component of this model is operated on the entire population, which is synthesized from the census data with highly detailed information about individuals. Not only can the basic characteristics of individuals be projected over time, but life-cycle events such as fertility, mortality, and household formation and dissolution can also be captured. Apart from these dynamic demographic changes, the housing location model is introduced along with random utility theory. Combined with this theory, the housing alternatives are restricted on the basis of statistically reliable rule sets for heterogeneous groups of actors. A prototype has been put together to demonstrate the structure of this model in the literature (Miller et al., 2004). However, it has not been incorporated with real data. It is still under development and has proven to be a huge task because much more work needs to be done.

2.5.3 Demographic and housing market models

Demographic models capture residential mobility mainly triggered by the life-cycle events of individuals and households, and their inter-relationships.

SVERIGE (System for Visualising Economic and Regional Influences in Governing the Environment) is a dynamic spatial microsimulation model for economic-demographic effects. It describes the movement of individuals and households from one state to another in Sweden (Vencatasawmy et al., 1999; Holm et al., 2002). The foundation of SVERIGE is a model called CORSIM (Cornell Microsimulation Model) (Caldwell, 1997), which is widely used in the United States. SVERIGE calibrated the CORSIM model to the European context and, on that basis, introduced a spatial module into it. The model can specify the life-cycle event of each individual with pre-determined orders: fertility, education, employment and income, marriage, divorce, leaving home, migration, mortality, immigration, and outmigration. These life-cycles update the characteristics of individuals, such as their age, marriage status, and family structure. Then the migration of those individuals is determined by regression models and utility functions with labor market adjustments.

SMILE (Simulation Model for Irish Local Economy) is a dynamic spatial microsimulation model designed to capture the population changes in rural Ireland (Ballas et al., 2005). This model includes a static process to create the base disaggregate spatial population, known as synthetic population. After the synthetic process, the next step is to watch individuals go through three major processes: mortality, fertility, and internal migration. All of the three processes are determined by probabilities based on age, sex, and location, which are derived from statistical data. One exception is that the probability of a migrant moving to a particular region is determined by the population density of that region.

PUMA (Predicting Urbanization with Multi-Agents) is a multi-agent system working at disaggregate level to represent the land-use changes in a behaviorally way (Ettema et al., 2007). The model is comprised of three sub models: land conversion module, household module, and firm module. However, the operational model is currently only capable of the first two modules. The life-cycle events are embodied in this model to update the characteristics of each individual year by year. Similar to other models, random utility method is used to simulate the decision of location choices. This model is calibrated in the northern part of the Dutch Randstad and comprises 1.5 million households with 3.16 million inhabitants (Ettema et al., 2007). The population and household figures are synthesized using the Monte Carlo method with a 500 × 500 m grid. With a large relevant dataset, the model takes about 12 h to run a 30-years simulation with C++ as the implementation tool. This implies that a very detailed and behaviorally sound model of urban dynamics can potentially be developed (Ettema et al., 2007).

Most of the location choice modeling techniques used to simulate residential mobility is based on the utility of the decision makers. However, Oskamp believes that "housing search must not be seen as a process of utility maximization based on complete information, but as a process of satisfying behavior" (Oskamp, 1994). Because of incomplete information, the decision makers of a dwelling do not know whether or not the current offer is the best. Oskamp believes that the most satisfying behavior takes into account the fact that the housing market is more useful (Oskamp, 1994). On this basis, he composes demographic, housing supply, and housing market together into a microsimulation model, named LocSim, and calibrates the model in the Dutch municipality of Lelystad with 10% sample of the 57,368 total population in 1990 (Statistics Netherlands, 2009). In his demographic sub-model, the life-cycle event is simulated by the events similar to SVERIAGE. In the housing supply sub-model, exogenous changes such as renovation and demolition are updated by economic growth. Then the movement decisions are triggered by the results of demographic events and dwelling changes. Meanwhile, the hierarchical urgency of the movement is taken into consideration. This dynamic microsimulation model captures the inter-relationship between demographic structure and housing market (Hooimeijer and Oskamp, 2000). However, four weaknesses of this model are pointed out by Svinterikou (2007). First, the model cannot store the historical information of agents, including individuals, households, or dwellings. Second, the education and the employment of the individuals are not implemented in the model. Third, the housing prices should be updated within the model instead of being exogenous.

On the basis of the concept of Oskamp, Svinterikou (2007) implements a URM-MicroSim with a fully geographic representation and a Unified Modeling Language (UML) technique to capture the residential mobility within the spatiotemporal framework. More details are discussed in the next chapter.

2.5.4 URM-MicroSim

Within the land-use component of an integrated urban model, the dynamic life-cycle demographic events of the housing market seems to be a new approach to capture the spatial distribution of individuals and their demand to travel. As a result, UML based URM-MicroSim is introduced by Svinterikou (2007). The model consists of three sub-modules: housing demand, housing supply, and residential search and migration. The housing demand sub-model identifies the demand for new dwellings based on the simulation results from three major events: demographic change, income change, and demand change. Demographic change includes life-cycle events such as death, fertility, union formation, union dissolution, flatmate formation, flatmate leaving, nest leaving, and out-migration. Being close to reality, only one of these events is assumed to have taken place and the Monte Carlo sampling method is used to choose the undertaken event if more than one event occurs within a one-year simulation. In addition, immigration is simulated to generate immigrants and corresponding households into the city. Income change is subject to changes of job and education. Demand change refers to regular residences that have not experienced demographic or income changes. Meanwhile, the housing supply module simulates changes of dwellings and buildings within four events: new construction; structure conversion; demolition; and housing expenditure change. Within the four events, size, market value, rent value, and availability are maintained in the database. Then, individuals decide to move to new dwellings, triggered by events in housing demand and housing supply. Along with immigrants' housing choices, all these housing needs are simulated inside the Residential Search and Migration sub-module at different levels of intensity based on different reasons of movement. Households that cannot find satisfactory housing will be subject to the probability of out-migration. At the end, after all households have been simulated through the model, dwelling rent and market values are updated based on supply and demand in the housing market. The detail processes and the interrelationship between each sub-model are clearly demonstrated in Figure 2-2 (Svinterikou, 2007).

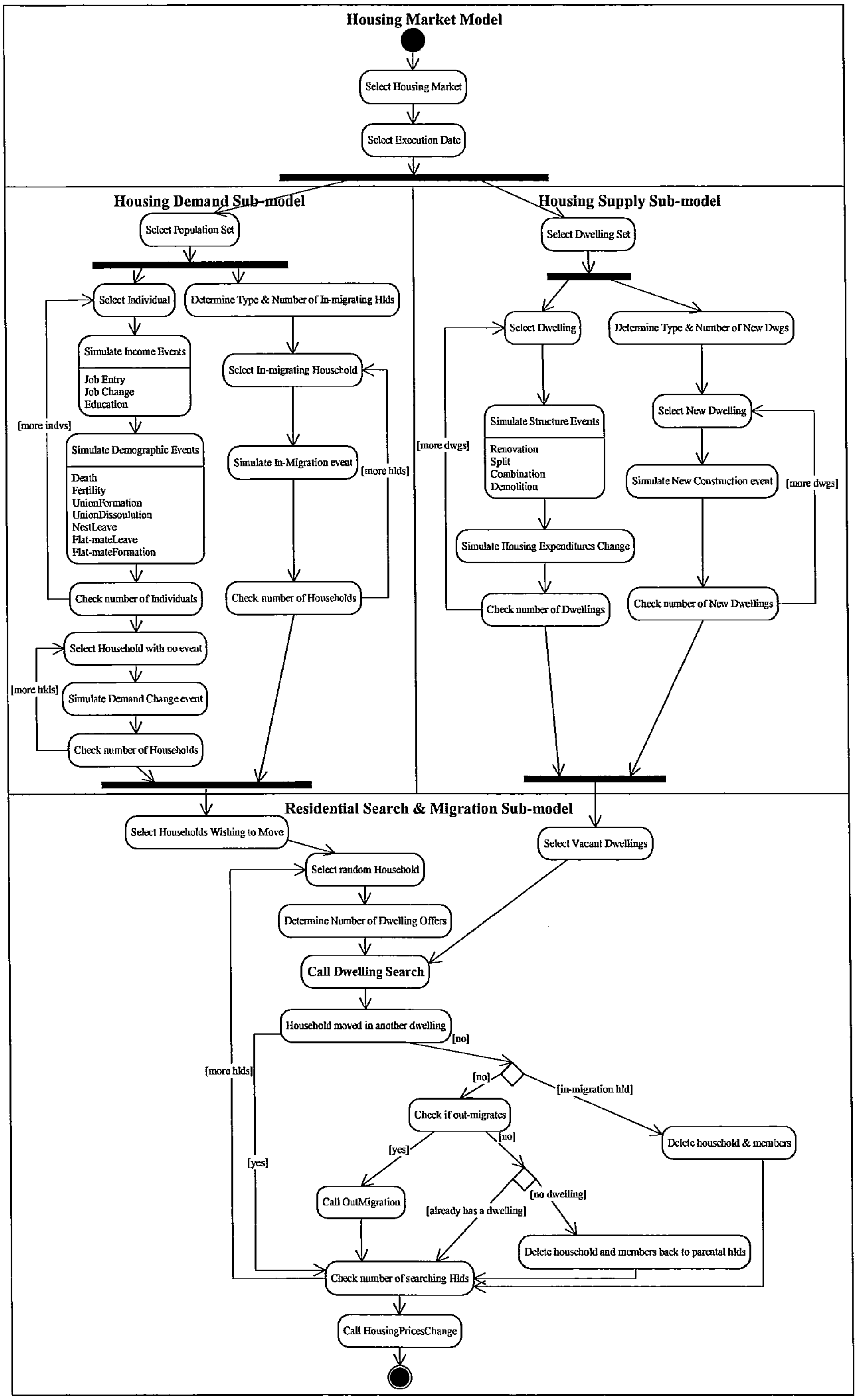

Figure 2-2: Design model of URM-MicroSim (Svinterikou, 2007)

This procedure is simulated on a one-year basis with detailed base-year population database. A prototype is designed with the Unified Modeling Language (UML) and calibrated for Mytilene, Greece, based on the synthetic population developed by Ryan (2008). However, the prototype is not complete. The housing supply is not fully implemented since data are not available. The housing market is static and the prices of dwellings are never updated within the model. The framework of the prototype is shown in Figure 2-3. The housing market sub-model is linked to two major components: housing demand, and search and migration sub-model. Housing demand includes immigration sub-model, demographic sub-model, and income change sub-model. However, the income change sub-model is not linked to the housing demand model since it has to be supported by the behavior of firms, which is not included in the model. Therefore, the income change sub-model cannot trigger the movements of households.

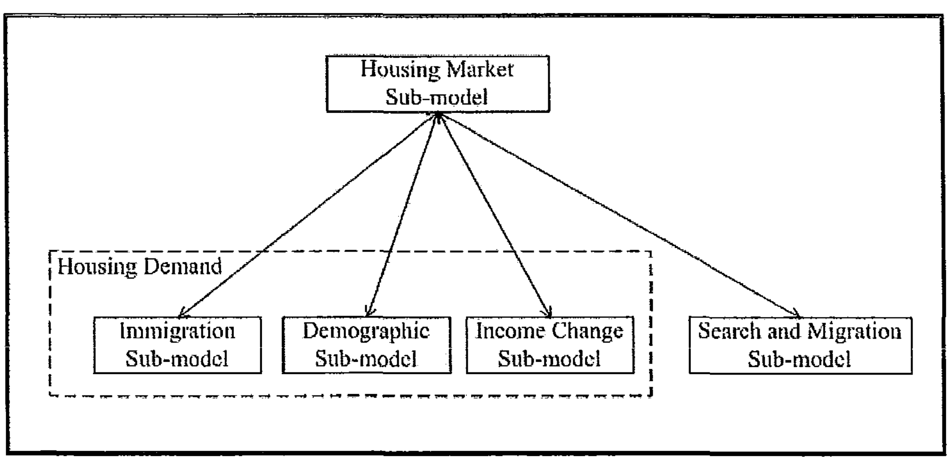

Figure 2-3: Framework of URM-MicroSim

One of the major challenges of microsimulation models is the micro-data. The basic component of microsimulation models is the micro-database, which is referred to as the input of the model in Orcutt's initial introduction (1957). The micro-data need to record the characteristic information of individual agents. Although the development techniques of census and sampling methods provide a great amount of available data about individuals, households, and firms, it is still impossible to collect the data for each individual. So, synthetic data are a common solution to compose the micro-data. Iterative Proportional Fitting (IPF), Synthetic Reconstruction technique (IPFSR), and the Combinatorial Optimization method (CO) are commonly used as the dominant techniques to create the synthetic population (Wilson and Pownall, 1976; Huang and Williamson, 2002). The micro-level data input for URM-MicroSim was produced by Ryan (2008) using the CO method. The data inputs for the synthetic population consist of a small sample from the population (generally with no spatial identifiers), as well as tabulations representing the spatial distribution of population characteristics. The synthetic micro-data are created in two major steps. Firstly, all of the individual agents associated with attribute data, including the geographic location of each agent, are synthesized based on the combination of two techniques: Combinatorial Optimization (CO) and Synthetic Construction (IPSFR). Second, the links between these agents are connected based on a series of rules that are intended to minimize unrealistic interactions.

On the basis of this theory, Ryan (2008) created the synthetic population for Mytilene, Greece, and Hamilton, Ontario.

However, URM-MicroSim for Mytilene is not fully operational; it is still a prototype in nature. Firstly, the system is not stable since system-errors randomly pop up during the simulation, but the reasons for the problems have still not been determined. Second, URM-MicroSim produces inconsistent results for the spatial population distribution. Thirdly, due to the complexity of the model, the system execution time of the model is so long that it could not be calibrated for a larger metropolitan area.

Chapter 3: URM-MicroSim for Mytilene

3.1 Introduction

The URM-MicroSim model for Mytilene, Greece, was introduced by Svinterikou (2007) and a prototype system was implemented. However, for several reasons the developed prototype could not simulate urban development for practical uses. In order to evaluate the performance of the system, we perform a comprehensive analysis in Section 3.2. Following that, we introduce a number of improvements, as described in Section 3.3. Results obtained from the revised system, along with an efficiency and consistency analysis are discussed in Section 3.4.

3.2 System Analysis

3.2.1 Introduction

Since the model is well structured and programmed in the existing system by Svinterikou (2007), URM-MicroSim for Mytilene is studied and improved from the viewpoint of systems engineering. First, the system architecture and user interface of the existing system is introduced. Then, the system is comprehensively analyzed in terms of its functional requirements and non-functional requirements.

3.2.2 Existing model

3.2.2.1 System architecture

System architecture refers to the platform on which a computer program is created. It always plays an important role in the system performance. URM-MicroSim is a Geographic Information System (GIS) tool aimed at modeling the spatial distribution of population in an urban area. It is based on the ArcGIS platform, in conjunction with Visual Basic for Applications (VBA), an object-oriented programming language, and a geodatabase capable of storing, querying, and manipulating geographic information.

The general framework of the system is displayed in Figure 3-1. The base of the framework is a geodatabase. Spatial data such as the locations of each individual, household, and dwelling, and aspatial data such as the characteristics of each agent are stored in Microsoft Access 2003, a relational database management system (RDBMS). These data can be retrieved and managed by ArcSDE, which provides a model that allows storing and managing spatial data with traditional RDBMS functionality. The integration of Microsoft Access 2003 and ArcSDE establishes an environment in which spatial data are managed as a traditional database that is accessible to applications. In this application, the logic of the model is programmed and compiled in ArcInfo Visual Basic for Applications (VBA), which is a combination of a programming language that is close to Visual Basic and ArcObjects — a technology that supports GIS applications. During the simulation process, VBA and ArcObjects work together and exchange information with the geodatabase through ArcSDE. After simulations are preformed, all the results can be accessed by ArcView for visualizing, managing, and analyzing.

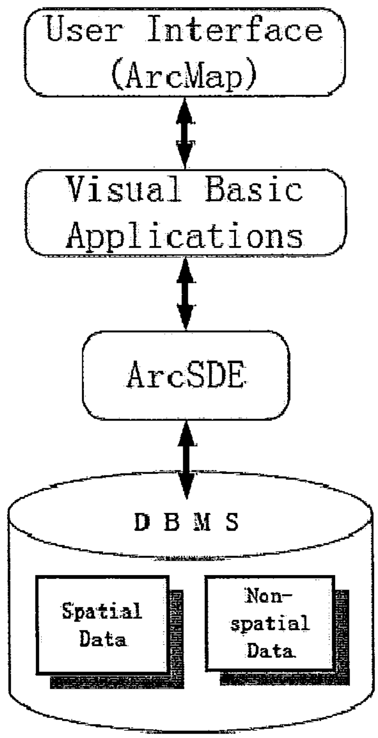

Figure 3-1: General system architecture of URM-MicroSim for Mytilene

Within this architecture, components are well structured and communicate with each other with relative ease. However, after the system was implemented, several disadvantages were discovered, as described in Section 3.2.3.

3.2.2.2 System user interface

As a prototype, the user interface is not implemented inside the existing model. Rather, the interface relies on ArcView to map the spatial distribution of agents. ArcView provides the "Editor" function to allow users to prepare the simulation scenario, i.e., to change the location and characteristics of each agent. The overall user interface of the existing system is shown in Figure 3-2. After the scenario has been set up, users can run the model simply by pressing the button "Housing Market" at the top of the toolbox. Then, the system starts to simulate the interrelations among the agents over time. After several hours of running, the result is displayed and future analysis can be conducted with the help of the spatial analysis function provided by ArcView.

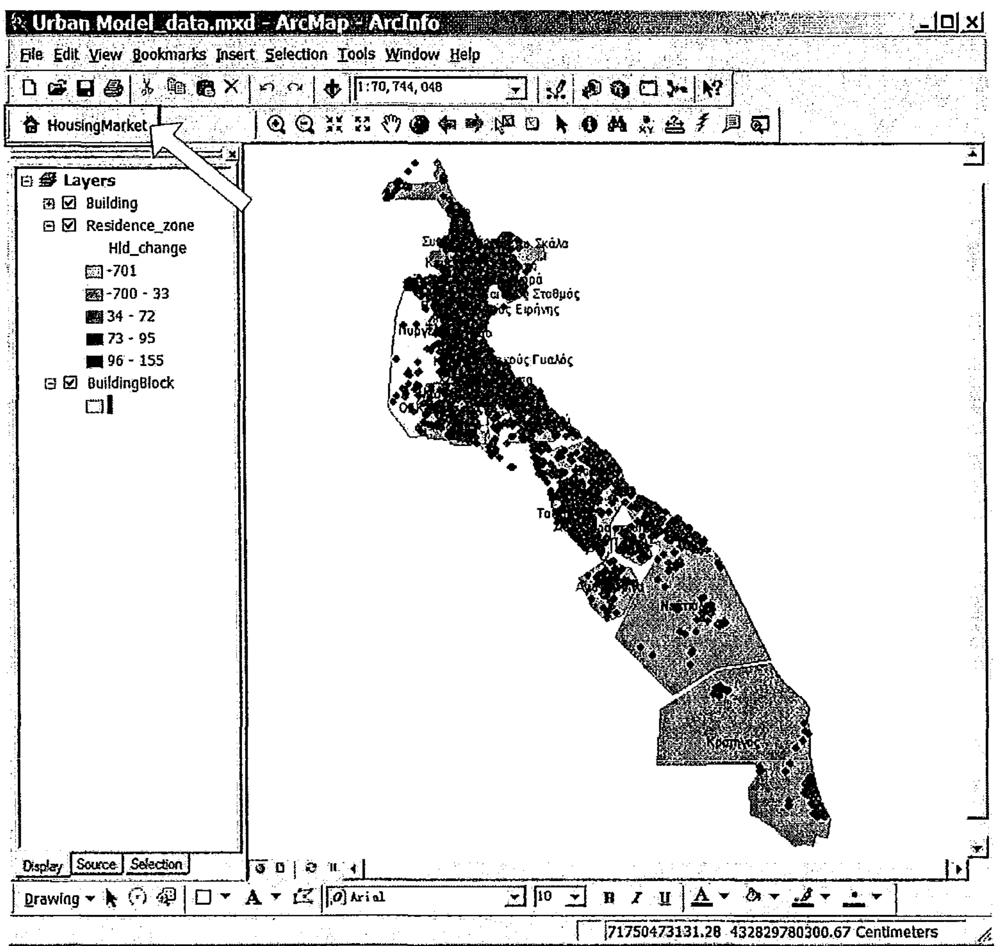

Figure 3-2: User interface of URM-MicroSim for Mytilene

However, the requirement is not just to identify the changed location of agents or their characteristics. Mostly, users want to know the outcome as a result of changes in demographic parameters, such as birth rates, death rates, and marriage rates. With the present set-up of the user interface, this is a rather complex task for the user. This is because these parameters are not represented in the interface of the system. The user is required to locate and change the appropriate lines of code in the computer program. Therefore, in order to transform the existing model into a more practical system, several changes and improvements need to be applied. These are introduced in the following section.

3.2.3 System Improvement Analysis

3.2.3.1 Introduction

When developing a practical system, it is important to perform an exhaustive analysis of the user requirements. The functional and non-functional requirements are discussed separately in the following sections.

3.2.3.2 Functional requirements

"Functional requirements capture the intended behavior of the system. This behavior may be expressed as services, tasks, or functions the system is required to perform" (Malan and Bredemeyer, 1999). In our case, the functional requirements include the modeling processes, user interfaces, and other system behaviors. Since the modeling process is comprehensively studied and analyzed by Svinterikou (2007), as introduced in the literature review, the major task of this section is to discuss other requirements beyond the basics.

User Interface

As discussed in Section 3.2.2.2, the system user interface does not meet the complex requirements of urban system users. In order to achieve this, a customized user interface is required that will allow users to change a large number of system parameters through a series of interface templates without the need to change the program code.

Logging System

The process that the system simulates is rather complex. Depending on the population simulated, millions of tasks are required to be performed, often leading to several hours of computational time. In order to understand the process and to keep track of system performance, a logging system is required. Such a system should be able to capture the running time, running status, and vital results for each finished sub-model, as well as different types of system errors. Implementation of the logging system could improve the ability of trouble-shooting problems and maintaining the system.

Multi-year Simulation

The overall purpose of the urban system is to simulate scenarios of different policies to analyze how cities evolve over a long time period. However, the existing model can only simulate one year. The simulation for the second year, based on the results of the first year, usually leads to a series of problems, such as dead loop or system error massage ("not enough space on temporary disk"). With a logging system and a comprehensive study of the system, these problems were discovered and rectified. Besides, the automatic running for multi-year simulation of the system is another function that is important to be implemented. Inside the existing system, the simulation of the second year needs to be initiated manually after preparing several parameters in the source code based on the results from the first year. Under these circumstances, developing and evaluating long-term scenarios would be extremely cumbersome. The automatic multi-year function needs to be programmed and implemented into the system as one part of the simulation process.

3.2.3.3 Non-functional requirements

Non-functional requirements are important aspects that are often neglected. Such requirements are used to judge the operation of a system, rather than specific behaviors. It usually refers to the maintainability, extensibility, and the execution time of the system. In fact, non-functional requirements not only determine the quality of software products, but they also affect the performance of functional requirements. If non-functional requirements are inadequately designed, the resulting products often fail to meet the system objectives, and could even produce inaccurate results.

URM-MicroSim for Mytilene is designed within the Unified Modeling Language (UML). Following the UML protocols, each component and sub-model in this system is well structured and recorded in the documents, which are easier to understand for the purposes of maintaining and extending the system. However, the system framework, introduced in Section 3.2.2.1, as well as the implementation algorithms for several components, render the system impractical because of its long execution time. It takes almost 1.5 h to run a one-year simulation for a city with 37,881 inhabitants. If implementing this system to a larger city, such as Hamilton with half a million inhabitants, the execution time would take months for a thirty-year simulation. As a result, the non-functional requirements here will mostly focus on the running time of the system.

An exhaustive analysis of the code is conducted to understand the time costs of each sub-model and key function. Two types of analysis are undertaken: runtime analysis and frequency analysis. The runtime analysis records how long each sub-model or key function takes in order to perform a one-year simulation. The frequency analysis records how many times each sub-model or key function is invoked within a one-year simulation. In this way, an overall and detailed understanding of the system and each sub-model is achieved and, at the same time, the parts of the system that are the most time consuming can be pinpointed.

As introduced in the second chapter, the existing system can be broken down into five major components: housing market sub-model, income sub-model, demographic sub-model, immigration sub-model, and search and migration sub-model. Concerning the runtime analysis, the time spent for each sub-model is listed in Table 3-1. From Table 3-1, it is clear that the two models with the longest running time inside the system are demographic sub-model, which takes 3,194 s to run, and search and migration sub-model, which takes 1,459 s to run.

Table 3-1: The runtime for each sub-model within a one-year simulation

Model	Runtime of URM-Mytilene (seconds)
Housing Market	2
Income	119
Demographic Change	3,194
Immigration	20
Search and Migration	1,459
Total	4,794

The "heavier" functions of the demographic sub-model are those that access the database. Execution time of these functions consumes almost 80% of the total runtime of the demographic sub-model. The related analysis results can be seen in Table 3-2. The analysis is broken down into six groups. The core components of the demographic model, including the logic of life-cycle events (such as death, fertility, union dissolution, and union formation), only takes 6.09% of the total running time. However, partner searching process (22.87% of total runtime), "StartEditing" functions, and "StopEditing" functions (37.98% of total runtime) consume most of the running time of the demographic model.

Table 3-2: The runtime for each key function of demographic sub-model

Functions	Proportion of Total Simulation Time
Life-cycle Event (fertility, death and others)	6.09%
Search Partners for Marriage Query	22.87%
Table Update (modify the data)	2.63%
Individual Insert + Individual Update + Event Experienced Insert	0.72%
Find in Table(get information from database)	9.67%
Stop Start Editing (edit session mechanism)	37.98%
Other Functions Irrelevant to the Database	20.04%

The most time-consuming system functions, "StarEditing" and "StopEditing," are part of the edit session mechanism provided by ArcSDE. All the changes to the geodatabase have to be included within an edit session that is bracketed between the "StartEditing" and "StopEditing" methods. Edit sessions usually match up with a long transaction. After the "StartEditing" method has been called, the required modification of the geodatabase is grouped to function. Meanwhile, the modified dataset and corresponding modification history are locked and stored in a temporary memory inside the database engine. Therefore, if errors occur, the dataset can roll back to the original file. In this way, this mechanism ensures the integrity of the dataset. However, this is designed especially for multiple-user cases. Usually, if multiple users are interacting with the dataset at the same time, it is important to make sure that only one of them at any given time can modify the data. However, this urban system is designed for single-user only. The edit session mechanism for this system is a burden because it requires large amounts of time and memory. Unfortunately, the edit session mechanism is mandatory if ArcSDE is used to interact with the database. There is no alternative but to cut off the running time of this function without changing the geodatabase engine.

With regards to the partner searching process, Figure 3-3 shows the algorithm of the procedure. It starts within the union formation event for each individual person. If a person meets the marriage criteria, such as single and in legal age to get married, he or she will be selected to look for a partner. Next, a list with all qualified potential partners is created, determined by the sex and the age of a person. From the list, a possible partner is randomly chosen. Then the individual is subjected to the probability of getting married, as determined by a Monte Carlo model based on the sex and the age-specific marriage probability of the individual. If the individuals fail to find a partner or refuse to get married, the simulation will pass to the next life-cycle event. Otherwise, these two persons will get married and their information will be updated in the database accordingly. Inside the algorithm, for the search for all qualified partners and the procedure of selecting one among them randomly are time consuming since the database management engine, ArcSDE, does not support a database automatic random selection. As an alternative, this task is achieved by reading all the required records from the database and writing on system memory, and then randomly choosing one from this huge group of records. In this way, the program consumes not only massive memory but also wastes time on transferring data between system memory and database.

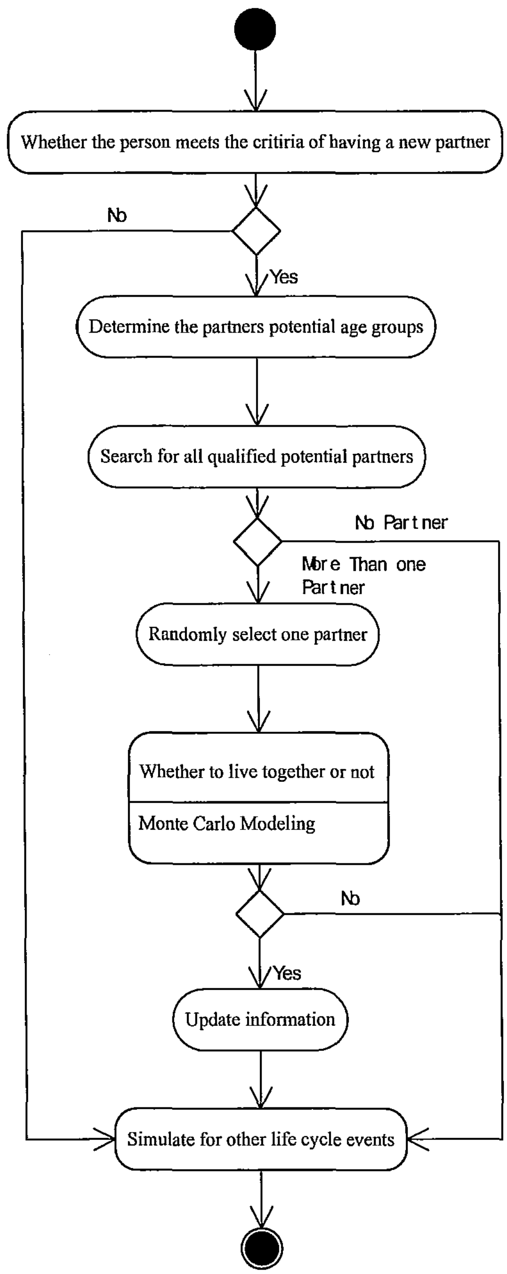

Figure 3-3: Partner searching algorithm for union formation

Meanwhile, frequency analysis indicates that a search for a partner is invoked much more than it should. Table 3-3 lists how many times each life-cycle event takes place in the city of Mytilene within a one-year simulation. It is noticeable that the process of partner searching is invoked almost 23 times more than a successful union formation. This means, after executing the time-consuming process of getting a partner, 95% of the individuals decide not to get married. It is a waste of time. Since the decision to get married is determined only by the age and the sex of the individual irrespective of the potential partner chosen, the individual should decide whether or not to get married before starting to search for a partner. The same issue appears in the dwelling search process. In the existing system, a household has to decide whether to get a new dwelling after it spends a long time searching for qualified dwellings. That results in an enormous increase in execution time. A household should make the decision to get a new dwelling before the dwelling searching process. The solution to these problems is introduced in the next chapter.

Table 3-3: Times of each life-cycle event in Mytilene within a one-year simulation

Life-cycle Event	Happened Per Year
Death	284
Fertility	354
Union Dissolution	113
Union Formation	433
Out Migration	1,027
Nest Leaving	55
Flat Mate Formation	46
Search for Partner	9,863
Flat Mate Leave	8

3. 3 Methodologies

3.3.1 Introduction

The system analysis introduced in the previous section has highlighted the problems with the existing system. In this section, methodologies to overcome these problems are explained in the same order as the problems were presented.

3.3.2 Functional improvement

User interface

The programming platform of ArcInfo VBA provides the tool to design and implement a customized user interface. The user interface of this system is designed under this platform. Figure 3-4 shows that all the system parameters are well organized. Users can set up the time period for a simulation. All the other parameters are categorized into four groups: income, demographic change, immigration, and search and migration. These parameters correspond to the sub-models of the system. The relating parameters of each group could be revised in the templates. By changing these numbers, the simulation scenario could be set up.

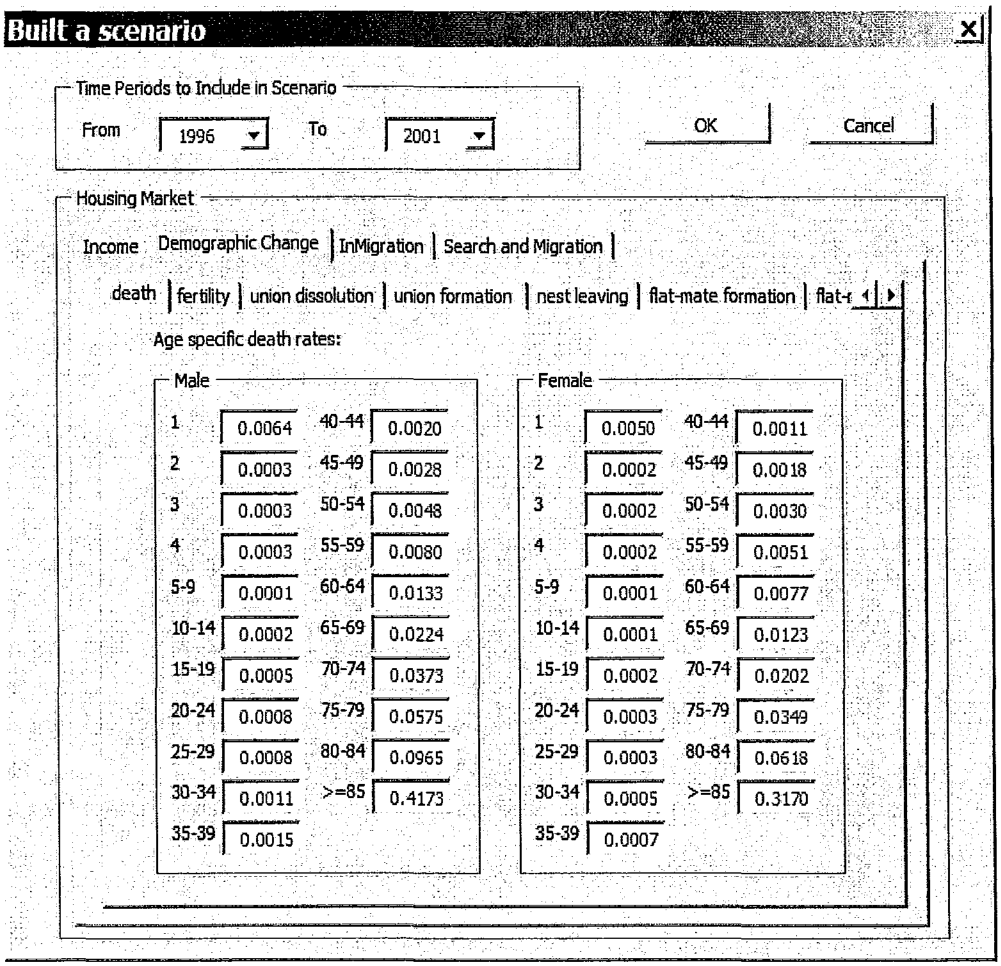

Figure 3-4: User interface of URM-MicroSim for Hamilton

Logging system

The logging system is encapsulated in the function object, which is called "TextStreamWriting". This function object can record all system messages into a text file as well as the date and the time when each message is posted. The logging system mainly captures two types of system massages. First, it records the history of system running process. If the program invokes any sub-model or key function, the logging system will record the corresponding start time and end time into the logging file. In this way, the system runtime index and efficiency index, presented in Section 3.2.3.3, can be obtained. Second, system errors are recorded. To maintain or expand the system, changes made to the data and codes are inevitable. These changes could easily lead to different unexpected system errors, such as losing the index of records. In order to trace, understand, and fix these errors, the logging system can record the sequences, locations, and the relevant context parameters of these errors.

Multi-year projection

As we discussed in Section 3.2.3.2, the system as developed by Svinterikou (2007) could only run a one-year simulation. Unexpected errors occurred within the second-year simulation. Therefore, debugging is used to solve these errors. Debugging is a methodical process of finding and reducing the number of bugs, or defects, in a computer program or a piece of electronic hardware, thus making it behave as expected (Myers, 1976). Since the existing system is tightly coupled (functions are closely dependent on each other), debugging tends to be harder, as changes in one may result in bugs to emerge in another. In general, it is a cumbersome and tiring task. Therefore, the debugger software tools play a significant role and make it possible to monitor the execution of a program, set breakpoints, pause the execution, record the context parameters, and revise them during system operations.

For the existing system, as introduced before, the application is implemented on the ArcInfo VBA platform, which offers an integrated development environment, including an integrated debugger system that provides access to all the ArcObjects technology. Combined with the logging system, errors inside the system are identified. These errors include shortage of variables validation (e.g. inspections on unexpected empty variables) that lead to an endless loop, incomplete logic that provides inconsistent results, incorrect sequence of programming codes that cause fatal system errors, unnecessary memory usage that results in more memory consumption, and increasing size of dataset and memory consumption that stops the simulation process.

Among these, the last is a major problem. During system operations, the amount of necessary information that must be temporarily stored in memory increases. In addition, keeping records of the dataset modification history further increases the size of the temporary dataset. Therefore, as the simulation continues with huge numbers of data operations, the system can easily crash. The solution to this problem is summarized in the following three steps.

The first step is to accurately declare variables. Generally, there are two types of variables: global variables that are accessible by every model and occupy the system memory throughout the whole simulation process; and local variables that are accessible only from the function or block in which they are declared. The local variables only exist within one part of the simulation. Therefore, declaring the variables that consume a lot of computer memory as local variables instead of global becomes the first principle of the solution.

Second, compacting the dataset within the simulation helps to reduce the occurrence of this problem. The major sub-models are interconnected. The output of one sub-model is always input to another. The dataset compact function provided by ArcSDE can compress the dataset before outputting it to another sub-model. In this way, the size of the dataset is reduced sharply. For the trade-off, each compacting process will take several minutes. Therefore, the compacting function must only be invoked after the demographic and search and migration sub-models have executed.

Last but not least, changing the corresponding database engine parameter can fix the problem temporarily. The Access Database is based on the Microsoft Jet engine. In order to prevent data from being corrupted or invalidated, the engine employs a locking policy by setting up a parameter, called MaxLocksPerFile, which prevents the number of transactions from exceeding the pre-specified number. In this system, enormous interactions between the dataset and application will exceed the default number and lead to a serious system error. The solution, provided by the Microsoft support center, is to manually change the value in the Windows registry to a bigger number. After different trials, the recommended setting is to change it from 9,500 to 50,000.

After fixing all the bugs in the system, the second-year simulation can proceed. Therefore, automatic running of the system for multiple years becomes feasible. A "Main" function is created to sequentially link different years of simulation inside a loop. Meanwhile, the system's time series simulation results over multiple years are stored in the database for policy analysis.

3.3.3 Non-functional improvement

On the basis of the non-functional analysis introduced in Section 3.2.3.3, the major issue is how to reduce the long execution time of the system. In order to solve this problem, the partners searching process is revised. Moreover, other program optimization methods are implemented.

In order to optimize the algorithm of searching for partners, we can simplify the procedure into three steps: creating the searching criteria, searching for a partner, and determining whether to get married. As discussed before, whether an individual gets married or not depends on their characteristics. The last step of partner searching, which is the determination of marriage, is independent of the other steps. As described by the Commutative Laws ($x \times y = y \times x$ and $x \cap y = y \cap x$), the order in which we take the conjunction (or disjunction) of two propositions does not affect the result. Therefore, changing the sequence of the steps of partner searching is unlikely to affect the result. Thus, the last step is moved to the front in order to avoid unnecessary system operations.

The whole procedure is described in Figure 3-5. Firstly, the system determines whether or not the individual is getting married, and then it creates the searching criteria. Finally, if an individual decides to get married, he/she starts to search for partners and select one from them.

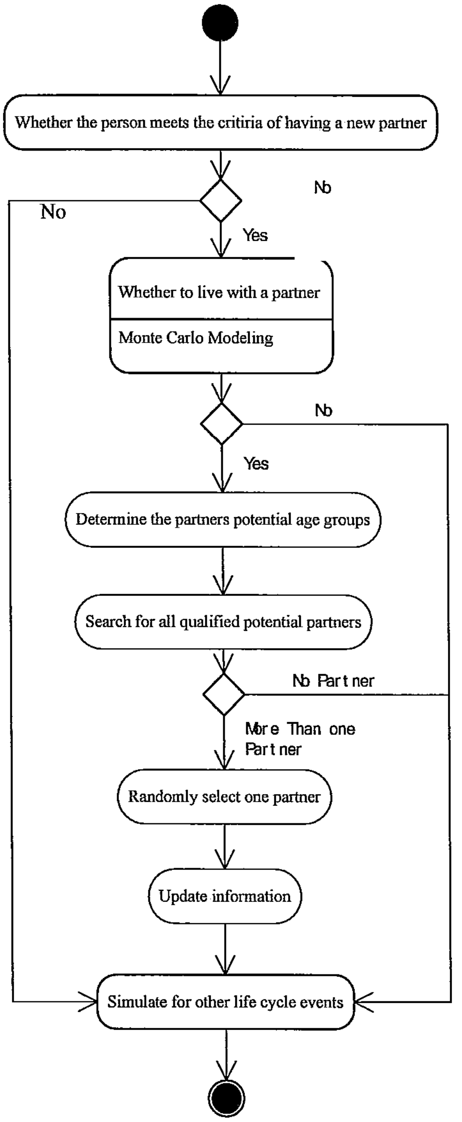

Figure 3-5: Improved partner searching algorithm for union formation

The same problem emerges in the search and migration sub-model, when individuals are searching for new dwellings. Similar improvements are applied to the sub-model. After it is revised, an individual has to determine whether or not to make a move before searching for all prospective dwellings.

As a result, the system running frequency is reduced sharply without affecting the simulation results. With the new scheme, the time consuming partner searching function is invoked 433 times, as opposed to 9,863 times. Similar improvement is estimated from the revision of the dwelling searching function.

Other methods related to improvements to the performance of the system are also implemented. Both the execution time and system memory have been considered.

Loops

A loop for programming is an iterative statement that allows code to be repeatedly executed. For the microsimulation model, since it is designed to simulate each individual agent respectively, complex loops are widely used. However, such a practice could lower the system speed, while increasing the system memory usage. Moreover, the endless loop, which is a continuous repetition of a program segment consuming all available processor time, is commonly found within complex loops. In order to alleviate the problem, it is important to clearly understand the scope of the loop and exit conditions. Irrelevant operations such as variables declaration and constant calculations should be put outside of the loops to avoid unnecessary calculations and system memory usage. Meanwhile, there is usually more than one way to end a complex loop. All the other additional terminating conditions, including the termination by unexpected system errors have to be checked and handled.

Data types

Using proper data types can also optimize the system. In VBA, the default data type is Variant. This means that if any type is not explicitly declared, it will be recognized as Variant. Other types should be claimed by the word Dim. Variant has its own advantages in flexibility, which can bring great convenience to the system. But at the same time, it has some major disadvantages. In VBA, each Variant type takes up 16 bytes of memory space. In contrast, other data types like Integer, Long, and Double- require 2 bytes, 4 bytes, and 8 bytes, respectively, and they save more memory space. Moreover, when the system operates a variable declared as Variant, it requires an additional step to exchange Variant type to the explicit data type. This extra operation will lower system efficiency and reduce computing speed. Therefore, if a certain data type can be determined, it should be used to decrease system memory and reduce the execution time.

Others

Other small tips from the programming side can be used to decrease execution time as well. For instance, the select case statement can be very handy if it is used in the proper way. The syntactic form of this statement is listed below:

Select selector

Case Value 1
  Return Statement 1
Case Value 2
  Return Statement 2
Case Value 3
  Return Statement 3
……
Case Value n
  Return Statement n
End Select

If the selector is equal to Value 1, Statement 1 will be executed. If the selector is equal to Value 2, Statement 2 will be executed, and so on. The program will begin by comparing the selector with Value 1. If they match, the program ends, otherwise it will go on to compare the selector with Value 2 until the right one is discovered. And the rest of the statements will not be operated at all. If the most possible value for the selection is placed at the end of the statement, the program will waste a lot of time on value comparisons. Therefore, the sequence of the values should be ordered according to their frequency of appearance.

Overall, with all of these non-functional improvements, the execution time of the system will be substantially decreased. This makes the system feasible to be calibrated to the city of Hamilton with its larger population size. The detailed results are discussed in the next section.

3.4 Results and Discussion

3.4.1 Efficiency test and analysis

After the introduction of the improvements described in Section 3.3.3, the runtime of the system was reduced significantly. The runtime (in seconds) before and after improvement of each model, function, and sub-model for the one-year simulation is recorded by the logging system, listed in Table 3-4. It is noticeable that the most significant improvement happened in the "search for partner query," which is reduced from 2933 to 343 s. Similar changes to the Search and Migration sub-model have a clear impact to the runtime as well. The total runtime saved from this sub-model is 475 s, from 1459 to 974 s. Other changes applied to the system are not recognized clearly in this analysis since other sub-models originally only consumed little time to run. While applying the system to larger urban areas, these changes could reduce runtime considerably.

Table 3-4: Runtime analysis for URM-MicroSim for Mytilene before and after improvement

Model	Function/Sub	Runtime of URM Mytilene(seconds)		Complexity Analysis T(n)	Projection Runtime for Hamilton (seconds)
		Before Improve-ment	After Improve-ment		
Housing Market	Total	2	2	c	2
Income	Total	119	123		
	Job Entry	8	7	n + c	106
	Job Change	1	1	n + c	15
Demographic Change	Total	3,194	604		19,609
	Death	40	39	n + c	593

	Fertility	38	37	n + c	562
	Union Formation	31	35	n + c	532
	Union Dissolution	14	14	n + c	213
	Nest Leaving	7	7	n + c	106
	Flatmate Leaving	1	1	n + c	15
	Flatmate Formation	6	6	n + c	91
	Outmigration	124	122	n + c	1,855
	Search Partners Query	2,933	343	3n + c	15,642
Immigration	Total	20	20	Model Changed	
Search and Migration	Total	1,459	974		88,835
	Main Model	45	45	6n + c	4,104
	Death	162	84	6n + c	7,661
	Fertility	268	152	6n + c	13,863
	Nest Leaving Leavers	19	11	6n + c	1,003
	Nest Leaving Stayers, Union Dissolve Stayers, Union Formation Stayers , Out-migration Member Stayers	522	390	6n + c	35,571
	Union Dissolve Leaver	38	27	6n + c	2,463
	Union Formation Leavers	14	6	6n + c	547
	Union Formation	195	114	6n + c	10,398
	Flatmate Formation	2	2	6n + c	182
	Flatmate Formation Stayers or Flatmate Leaving Stayers	19	15	6n + c	1,368
	Flatmate Formation (one flatmate already has a dwelling)	30	25	6n + c	2,280
	Flatmate Leaving Leaver	4	4	6n + c	365
	Immigration	141	99	6n + c	9,029
Total		4,794	1,723		108,568

In general, the improvements are encouraging. The one-year projection of simulation is reduced from 1.5 h to 0.5 h. Therefore, the model becomes more feasible for urban policy analysis. However, Mytilene is only a small city with a population of nearly 30,000. If applying the model to urban areas with a larger population size, the performance of the system is questioned. In order to project the performance, the computational complexity analysis is introduced.

The computational complexity is analyzed for each function. The projection of runtime could be estimated if implementing such a system to another urban context.

However, the runtime projection of URM- MicroSim Hamilton is used as an instance. The computational complexity refers to the number of steps that a sub-model or an algorithm needs to simulate. It denotes the runtime of the algorithm and is expressed as a function of the population size n. Usually, the value of *T(n)* estimates the runtime of the algorithm in the worst-case scenario. In this analysis, the estimation is based on the average case. After the improvement, almost all the functions grow linearly. Therefore, the complexity is described with two functions:

$$T(n) = c$$

$$T(n) = a \times n + c$$

The first function means it is determined by a constant number of steps. Increasing the size of the records will not influence the runtime of the model. The second function means that it is determined not only by a constant number of steps. With the increasing size of the records *(n)*, the number of running steps and the runtime of the function are linearly increasing correspondingly by the factor a. In most functions of the income and demographic sub-models, the system complexities of these models are $T(n) = n + c$. This means, if the size of the population increases by 100%, the runtime of the model will also increase by 100%. The function of search for partners query inside the demographic sub-model, however, appears to be more complex. If an individual decides to search for a partner, the average number of qualified partners searched from the dataset is around

1,000 people. However, for a city with a larger population like Hamilton, based on the projection of the model, a person who is looking for a partner can find a lot more qualified persons. In the system, the partners searching capacity is restricted to 3,000. As a result, adding one more individual to URM-MicroSim for Hamilton will make it subject to two times more runtime on partner searching query. The complexity of this sub-model is $T(n) = 3n + c$. In the search and migration model, all the system complexities of the functions are $T(n) = 6n + c$. This is because each household has to make several offers to qualified dwellings before they can find a new dwelling in which to live. The average number of dwelling offers is 6. Therefore, one more household added into the system may cause all additional 6 searches for dwellings.

With the computational complexity analysis, the projection runtime to URM-MicroSim for Hamilton could be achieved. The population size of Hamilton in 1996 was 15.2 times larger than the population size of Mytilene in 2001. On the basis of the complexity analysis, the runtimes of all functions increase accordingly. As listed in Table 3-4, the search and migration model increases substantially, from 974 to 88,835 s, almost 100 times more. The major reason for this is the high complexity derived from dwelling offers from the dwelling searching mechanism. Each individual first has to search for thousands of qualified dwellings from all potential dwellings in the housing market. Then, each individual will randomly select one of those dwellings. This whole procedure consumes a lot of runtime. However, this is not the case in reality. As discussed in the literature review, people tend to move close to the previous residential location. Therefore, the dwelling searching for a household can be confined only to the census tract in which this household resides. In this way, a smaller number of qualified dwellings will be found. Since the city of Hamilton is divided into 144 census tracts, the maximum number of searching results, 3000, can be accordingly reduced to 200. As a result, the total runtime of this model could be reduced substantially. The complexity function will be $T(n) = 200 \times 6 \times n/3000 + c$. Thus, the projection of the runtime for URM-MicroSim for Hamilton will be reduced from 88,835 s to 5,923 s.

Despite all of the existing improvements, the runtime for URM-MicroSim for Hamilton is still long. It takes more than 10 h to run the simulation for only one year. It is not convenient for practical policy analysis if long-term simulations (20 to 30 years) are required. To reduce the runtime, the system architecture improvement, introduced in Section 3.4.3, may play a significant role. Applying the architecture introduced in this work, the prospective runtime is estimated to be about 2 h per year of simulation, which will be more practical for the system users.

3.4.2 Consistency analysis

Once URM-MicroSim for Mytilene is calibrated, it is very important to understand the performance of the simulation results. Unfortunately, no empirical data are available for the Municipality of Mytilene. Therefore, no validation could be applied to the outcome of URM-MicroSim for Mytilene (Svinterikou, 2007). However, the consistency of the system can be analyzed.

To understand the consistency of the system, simulations of the basic scenario, meaning no changes or input to the system, have been tested for 100 times. For each time, the simulation is designed to run for 10 years, from 2001 to 2011. All the household migration patterns within 10 years are recorded by census tracts, which divide Mytilene into 29 zones. After the 10-year simulation, for each census tract, the net balance of total household changes, which is the difference between the number of total in-migrants and the number of total out-migrants, is used to test the system consistency. In short, the selected result from each simulation is a series of numbers representing the changes in household distribution by census tracts after 10 years.

From these results, at the aggregate level, the maximum, minimum, mean and standard deviation of the total changes of the households in Mytilene are presented in Table 3-5. The Coefficient of Variation is only 0.0185. Therefore, the difference in results between 100 simulations is relatively small. This proves that the total households migration simulated by the system produced consistent results.

Table 3-5: The summary statistic of total households changes in Mytilene based on 100 times of 10-year simulations

After 10-year Simulation	Max	Min	Mean	Std Dev	Coefficient of Variation
Total Household Changes	4,514	4,141	4,314	80	0.0185

At the disaggregate level, Table 3-6 shows the statistics summery of household changes distributed by census tracts, including the census tract number, the number of households before simulation, the mean, standard deviation, and coefficient of variation of household changes in the corresponding census tracts for 100 times of the 10-year simulations. In most census tracts, the simulations produce consistent results. The coefficients of variation, which are the normalized measurement of dispersion, are centered on 20% to 30%. However, in some census tracts, the coefficients of variation of household changes within 100 simulations are a little bit high. Note that, the census tracts that produce high values of coefficients of variation have fewer household migrations. This is reasonable. Because the model is based on a series of stochastic processes, residuals are expected. If the household migration is small, the residual will be magnified.

Table 3-6: The summary statistic of household-changes distributed by census tracts in Mytilene based on 100 times of 10-year simulations

		Total households changes, 2011			
	Number of Households, 2001	**Mean**	**Standard Deviation**	**Standard Deviation /Household**	**Coefficient of Variation**
Zone 1	1,167	564	108	0.09	0.191606
Zone 2	1,014	358	87	0.09	0.244354
Zone 3	745	357	90	0.12	0.251705
Zone 4	409	22	40	0.10	1.787627
Zone 5	197	63	40	0.20	0.629856
Zone 6	465	160	67	0.14	0.420475
Zone 7	817	318	82	0.10	0.257974
Zone 8	580	83	52	0.09	0.626869
Zone 9	198	146	49	0.25	0.333865
Zone 10	63	185	48	0.76	0.258237

Zone 11	297	19	22	0.07	1.117480
Zone 12	541	167	49	0.09	0.295957
Zone 13	1,117	190	76	0.07	0.397757
Zone 14	179	45	27	0.15	0.598767
Zone 15	209	47	32	0.15	0.684065
Zone 16	509	496	106	0.21	0.213408
Zone 17	294	-62	17	0.06	-0.276070
Zone 18	536	187	64	0.12	0.340365
Zone 19	193	29	21	0.11	0.705123
Zone 20	78	39	39	0.50	0.989252
Zone 21	339	250	65	0.19	0.260418
Zone 22	351	370	82	0.23	0.220613
Zone 23	411	142	39	0.10	0.274556
Zone 24	72	44	7	0.10	0.164861
Zone 25	61	32	6	0.09	0.175643
Zone 26	148	9	8	0.05	0.864946
Zone 27	44	12	4	0.10	0.355118
Zone 28	13	27	4	0.33	0.155267
Zone 29	122	10	8	0.06	0.760578
Average	385	149	46	0.16	0.458644

Meanwhile, it is more important to compare the simulation results among different times at the level of census tract. Derived from each simulation, the spatial distributions of household changes by census tract after 10 years are recorded in a vector. For each two simulation results, a Pearson correlation between the corresponding two vectors can be calculated to describe the linear relationship between them. In order to compare the 100 simulation results, the average correlation has been used. First, for each simulation result, 99 correlations are calculated based on their correlation with other simulated results. Then, the average correlation for this simulation result is calculated by getting the average of all 99 correlations. The calculation of average correlation is listed in the function.

$$Average\ Correlation\ (i) = \frac{\sum_j Correlation_{ij}}{n-1} \quad for\ i, j= 1, 2, 3..., n;\ n = 100;\ i \neq j$$

The summary statistics of the 100 average correlations are listed in Table 3-7, including mean, standard deviation, minimum, maximum, and coefficient of variations. The results among all simulations are highly correlated. The variation is relatively small, with a coefficient of variation of 0.02. This is evidence that the microsimulation model produces consistent results.

Table 3-7: Statistic summery of average correlations in Mytilene based on 100 times of 10-year simulations

Mean	Standard Deviation	Min	Max	Coefficient of Variation
0.891904	0.019928	0.825285	0.923381	0.02234

3.4.3 Conclusion and proposed future improvements

Overall, the research in this chapter improved the performance of URM-MicroSim Mytilene, which is the basis for URM-MicroSim Hamilton, introduced in Chapter 5. Functionally, users can automatically simulate the urban time evolution with a well designed user interface. Furthermore, the history of system running process and system errors are recorded by the logging system, which provides users with flexibility in handling errors and alarms. Non-functionally, the runtime of the system reduced substantially without affecting the results. On the basis of these achievements, the consistency analysis could be implemented based on 100 times of simulation. The results are encouraging with high correlations between these simulations. Finally, evidence is provided that URM-MicroSim performed consistently. This is significant to the research introduced in Chapter 5. On the basis of this evidence, the results, which are based on a single simulation by URM-MicroSim for Hamilton, are convincing.

Nonetheless, URM-MicroSim is still not practical for the city of Hamilton. As introduced in Section 3.4.1, the projection of system runtime for Hamilton is around 30 h for a one-year simulation. This means the system will take more than a month for a long-term simulation lasting 30 years. One of the major problems is the weaknesses of ArcSDE. These weak points, as described in the system analysis, include mandatory edit session mechanisms and users' inability to automatically random select a record from the database. These problems will inevitably lead to a longer execution time. Therefore, using ArcSDE as the database management engine is not suitable for a practical system. In order to sharply reduce the system runtime, the system architecture needs to be restructured.

As described in Section 3.2.2.1, the system architecture of URM-MicroSim is mainly composed of two parts: geodatabase and programming language. Geodatabase management system, known as spatial database engine (SDE), is an extension of the traditional relational database, which constructs the mapping relationship between the geodatabase model and the relational database model. As a result, spatial data can be added to the relational database system. The way SDE is designed and implemented affects the entire GIS performance. SDE can be implemented externally or internally. An external SDE is individually developed as a separate function model connecting with the Relational Database Management Systems (RDMBS). It is functionally rich and powerful in spatial data analysis. However, each data process has to firstly transfer the data from RDBMS to SDE and then perform its functions. An internal SDE is combined with an existing relational database management system (RDBMS). The spatial data management methods are the extension built on the relational database management methods. Although the method's ability to analyze and locate the spatial component is weak, this type of approach could be better integrated with the database and the aspatial data can be directly maintained by the RDBMS. This can lead to an improved performance of aspatial data management. Meanwhile, the data management process operates inside the RDBMS. Therefore, only the necessary data have to be transferred to the user for processing. This method is very efficient for massive datasets like our micro dataset. Moreover, the current system hardly has any complex spatial processing or analysis within the simulation. Most of spatial processes are preformed when users prepare the simulation scenarios. Meanwhile, for most of the simulations, especially in the demographic model, the spatial components of the data are not considered. It is more efficient to interact with the relational database directly, if non-spatial data are only required. Therefore, choosing the internal SDE can meet the system requirement.

The weakness of ArcInfo VBA is another reason why system architecture needs to be restructured. According to Rajesh (2008), VBA is functionally rich and extremely flexible but it does have some important limitations, including limited support for function pointers which are used as callback functions in the Windows API (Roman, 2002). The weakness of function pointers results in inefficient array list uses. An array list is an ordered collection of data stored in a variable. In the urban system, array lists are employed a lot. Individuals, households and dwellings are always stored inside the array lists by pre-determined order. Lack of function pointers in the array lists makes the system cumbersome. For example, the application is asked to select the hundredth individual from a group of individuals stored inside an array list. With a more powerful language, the function pointer can usually be moved to the hundredth individual and pick it out directly. However, VBA has to retrieve all the 99 individuals in front of it before selecting the hundredth individual. Clearly, this type of operation makes VBA unsuitable for large datasets.

Consequently, more advanced programming language and database management engines are required. Firstly, there are generally two types of programming languages: scripting languages, such as VBA, Perl, Python, and Ruby; and conventional languages such as C++ and Java. The scripting languages are more closely related to human languages, so they are of a higher level to allow for enhanced productivity. However, these types of languages are recognized as being time consuming when it comes to execution. The reason for this is that most scripting languages are interpreted from the source code or "semi-compiled" to byte code, which is interpreted, unlike the applications with which they are associated, which are traditionally compiled to native machine codes for the system on which they run (Morin and Brown, 1999). Unlike the scripting languages, the traditional programming languages are interpreted directly with the help of a compiler, which is often faster and subject to less restrictions to the implementation environment. Of all the conventional languages, C++ is widely recognized as the least taxing in terms of runtime and memory, and it is especially suitable for massive computations. Admittedly, a lot of criticisms have been referred to the complexity of the language. However, the advantages in terms of execution time clearly out-weigh the drawbacks.

The overall proposed framework is shown in Figure 3-6. For the SDE, the internal approach PostGIS is selected. It is an open source SDE built on the PostgresSQL relational database. Therefore, PostgresSQL is the database used for storing aspatial and spatial data. On top of that, C++ is the chosen programming language. It will directly retrieve aspatial data from the PostgresSQL. Spatial data and relevant analysis will be conducted with the aid of PostGIS. Finally, the user interface is built based on

MapObjects. It is a powerful collection of embeddable mapping and GIS components. Dynamic live maps and GIS capabilities can be easily created.

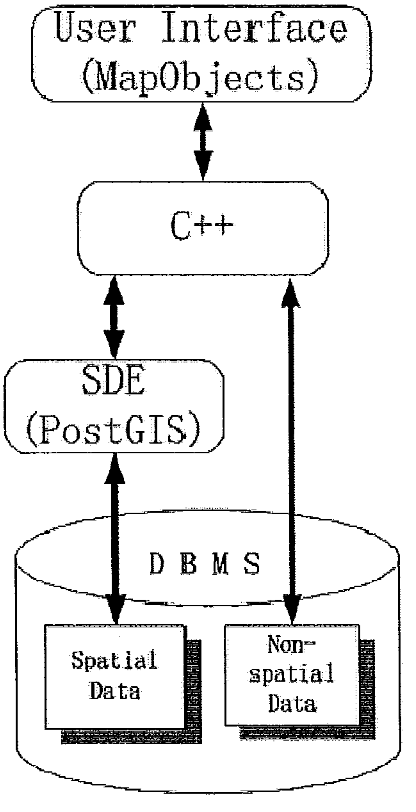

Figure 3-6: Proposed system architecture of URM-MicroSim for Mytilene

Chapter 4: URM-MicroSim for Hamilton

4.1 Introduction

As introduced in the literature review, urban issues in Hamilton can be studied more comprehensively with a microsimulation model, which can be made to run considerably faster. Thus, microsimulation can be applied to a city with a large population. URM-MicroSim is implemented for the city of Hamilton, Ontario, in this research. Detailed methodologies and corresponding results are introduced in the following sections.

4.2 Methodologies

4.2.1 Introduction

The implementation of the MicroSim model for the city of Hamilton entails two major steps. The first is to create a synthetic population for the city. The second is to adapt the model to Hamilton's urban characteristics. As stated previously, Ryan (2008) created a synthetic population for Hamilton that meets the input criteria of the URM-MicroSim model. The results of his research are encouraging: Not only were the distribution of individuals and dwellings in each census tract highly correlated with the actual distribution, but the characteristics that were added to the individuals were also highly correlated. Overall, validation carried out on the nature of synthetic family structures yielded, for the most part, positive results (Ryan, 2008). Therefore, this work focuses on the second step, which is to adapt the model to different urban contexts.

The existing URM-MicroSim is designed for a European city. To adapt it to the city of Hamilton, or other Canadian cities, the first step must be to study the differences between the two contexts. The major difference lies in resident living patterns. Although residents inevitably have the same demographic living patterns (birth, school, marriage, giving birth, and death), the differences that exist stem from the different probabilities of demographic events such as fertility and mortality. Consequently, the main part of the calibration process involves identify the probabilities of demographic events in the region of Hamilton. Furthermore, another vital difference between the two contexts is in the observed immigration patterns. Detailed discussions with all the calibration efforts are described in the following chapters.

4.2.2 Probabilities

In URM-MicroSim, all demographic events are represented by their probability of occurrence. Each individual in the population is sequentially subjected to the probability of experiencing each demographic event (death, fertility, union formation, union dissolution, flatmate formation, flatmate leaving and nest leaving) (Svinterikou, 2007). In order to identify all of the necessary probabilities required by URM-MicroSim for Hamilton, data were collected from Statistics Canada, including CANSIM (database), Canadian Vital Statistics and Census Hamilton. To correspond with the synthetic population of Hamilton, created for 1996, the collected data should be derived from the information for Hamilton in 1996. However, some of the data are not available. Therefore, the available statistical data for Ontario or all of Canada from 2000 to 2005 were used instead. Moreover, if some of the probabilities could not be found in any of the sources, the parameters for Mytilene were used. In the following segment, we provide a detailed description of the determination of the probabilities for all demographic events in question.

Death event

Mortality is simulated in URM-MicroSim using the death probabilities classified by sex and age. They are retrieved directly from the Vital Statistics, Death Database, based on 1996 for Ontario, as shown in Table 4-1.

Table 4-1: Age-specific death probabilities, Ontario

	Probabilities	
Age Group	**Males**	**Females**
Less Than 1	0.0064	0.0050
1	0.0003	0.0002
2	0.0003	0.0002
3	0.0003	0.0002
4	0.0003	0.0002
5 - 9	0.0001	0.0001
10 - 14	0.0002	0.0001
15 - 19	0.0005	0.0002
20 - 24	0.0008	0.0003
25 - 29	0.0008	0.0003
30 - 34	0.0011	0.0005

35 - 39	0.0015	0.0007
40 - 44	0.0020	0.0011
45 - 49	0.0028	0.0018
50 - 54	0.0048	0.0030
55 - 59	0.0080	0.0051
60 - 64	0.0133	0.0077
65 - 69	0.0224	0.0123
70 - 74	0.0373	0.0202
75 - 79	0.0575	0.0349
80 - 84	0.0965	0.0618
85 and Over	0.2087	0.1585

Age is determined as of the last birthday preceding death. All probabilities represent mortality rates, i.e., the number of deaths per 1,000 individuals of a particular age group in 1996.

Fertility event

In the simulation of childbirth, four sets of probabilities are used: fertility rate, sex ratio at birth, mortality rate of newborns, and multiple births ratio.

Within the context of this work, age-specific fertility rates for mothers are from Vital Statistics, Marriage Database in 2000–2005, Ontario, as shown in Table 4-2. The ages of the mothers represent the last birthday preceding birth. Each rate for 2000–2005 is calculated by dividing the number of new mothers who are in a specific age group for that year by the total number of married females in the corresponding age group for that year. Then, the average rates for 2000–2005 are calculated and applied to the model. In addition, Hamilton accounted for 7,443 newborn males and 7,129 newborn females in 1996. Therefore, the probability of a newborn female child is 0.4892 in Hamilton in 1996.

The mortality rates of infants are directly derived from Statistics Canada. Infant mortality (less than one year of age) rates per 1,000 live births are 6.4 and 5 for males and females, respectively. Lastly, during 2000 – 2005, Canada accounted for 1,946,161 single births and 58,717 multiple births. Therefore, the multiple birth ratio is 0.03.

Table 4-2: Age-specific death probabilities, Ontario, 2000 – 2005

Age of Mothers	2000	2001	2002	2003	2004	2005	Average
15 to 19	0.618140	0.605927	0.542684	0.553126	0.513427	0.529678	0.560497
20 to 24	0.235692	0.229582	0.215835	0.214695	0.211548	0.205849	0.218867
25 to 29	0.164647	0.171262	0.165950	0.170870	0.173452	0.174701	0.170147
30 to 34	0.123959	0.132326	0.128992	0.133647	0.138749	0.141960	0.133272
35 to 39	0.049283	0.051741	0.051981	0.054603	0.055980	0.059168	0.053793
40 to 44	0.008876	0.009236	0.009326	0.009734	0.010111	0.010259	0.009590
45 to 49	0.000431	0.000416	0.000373	0.000398	0.000502	0.000489	0.000435

Union formation event

In URM-MicroSim, three types of union formation are distinguished: marriage, cohabitation and marriage of cohabitating partners.

Sex- and age-specific marriage probabilities — males at age x marrying females at age y, and females at age x marrying males at age y — are from Vital Statistics, Marriage Database, based on 2000–2002 data for Canada as shown in Table 4-3 and Table 4-4, respectively. The dataset of Statistics Canada directly provides the total numbers of marriages between bridegrooms of a specified age group and brides of another specified age group from 2000 to 2002. For each year, the probabilities of males of age x marrying females of age y are calculated by dividing these numbers by the total number of unmarried males of the same age groups. Then, the average probabilities for 2000 – 2002 are calculated and applied to the model. In the same way, the probabilities of females of age x marrying males of age y are acquired. Moreover, the cohabitation probability is 0.05 according to Statistics Canada. In 2000, there were 279,180 married people and 14,420 people who were married under common law. Therefore, the cohabitation probability can be calculated. In the case of marriage of cohabitating partners, since no statistics exist, the probability that a cohabitating couple decides to get married is assumed to be 0.5 — the same assumption used for Mytilene.

Table 4-3: Probabilities ($\times$ $\mathbf{10^{-4}}$) of males of age x marrying females of age y, Canada, 2000 – 2002

Probabilities ($\times 10^{-4}$)	**Female Age Group**												
Male Age Group	**< 15**	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**>= 70**
< 15	0	0	0	0	0	0	0	0	0	0	0	0	0
15-19	0	1	6	2	1	0	0	0	0	0	0	0	0
20-24	0	3	136	137	34	8	2	0	0	0	0	0	0
25-29	0	0	35	236	128	38	10	3	1	0	0	0	0
30-34	0	0	4	41	95	58	21	7	3	1	0	0	0
35-39	0	0	1	7	22	39	26	12	5	2	1	0	0
40-44	0	0	0	2	5	14	22	17	9	4	1	0	0
45-49	0	0	0	0	1	4	10	16	14	7	3	1	1
50-54	0	0	0	0	0	1	3	7	12	10	5	2	1
55-59	0	0	0	0	0	0	1	2	5	7	7	4	3
60-64	0	0	0	0	0	0	0	0	1	3	5	5	5
65-69	0	0	0	0	0	0	0	0	0	1	2	4	6
>=70	0	0	0	0	0	0	0	0	0	0	1	2	14

Table 4-4: Probabilities ($\times \mathbf{10^{-4}}$) of females of age x marrying males of age y, Canada, 2000 – 2002

Probabilities ($\times 10^{-4}$)	**Male Age Group**												
Female Age Group	**< 15**	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**>= 70**
< 15	1	1	21	7	2	0	0	0	0	0	0	0	0
15-19	1	1	21	7	2	0	0	0	0	0	0	0	0
20-24	1	1	129	134	31	6	2	0	0	0	0	0	0
25-29	0	0	33	229	116	29	8	2	1	0	0	0	0
30-34	0	0	5	43	94	48	18	7	3	1	0	0	0
35-39	0	0	1	8	25	38	26	14	7	3	1	0	0
40-44	0	0	0	2	6	14	22	19	11	6	3	1	1
45-49	0	0	0	0	2	4	9	16	16	11	5	2	1
50-54	0	0	0	0	0	1	2	6	13	13	8	4	3
55-59	0	0	0	0	0	0	0	1	4	7	9	5	5
60-64	0	0	0	0	0	0	0	0	1	2	5	6	6
65-69	0	0	0	0	0	0	0	0	0	1	2	4	8
>=70	0	0	0	0	0	0	0	0	0	0	1	2	17

Union dissolution event

Table 4-5 shows the age-specific divorce rates in Canada calculated on the basis of Canadian Vital Statistics, Divorce Database Table, for 2004. The total divorce cases specified by the ages of the husbands at divorce are divided by the total numbers of married males grouped by age in 2004 to calculate the divorce probabilities. For each married couple, the chances of divorce are subject to the probabilities on the basis of the age of the husbands. In the case of de-habitation, since no statistics exist, the probability of union dissolution of cohabiters is assumed to be 0.1 — the same as that assumed in URM-MicroSim for Mytilene.

Table 4-5: Age-specific divorce probabilities based on the age of the husbands, Canada, 2004

Age of Husbands at Divorce	Divorce Cases	Number of Married Male	Divorce Probabilities
Under 20 Years	6	9,111	0.000659
20 to 24 Years	853	144,176	0.005916
25 to 29 Years	4,679	437,160	0.010703
30 to 34 Years	9,182	690,045	0.013306
35 to 39 Years	11,560	858,554	0.013464
40 to 44 Years	13,573	1,024,675	0.013246
45 to 49 Years	11,702	985,533	0.011874
50 to 54 Years	8,231	876,613	0.009390
55 to 59 Years	5,115	775,828	0.006593
60 to 64 Years	2,525	586,467	0.004305
65 Years and Over	2,094	1,364,720	0.001534

If a union dissolves, the couple should decide who keeps the children and whether the female or the male moves out of the dwelling. In most cases, the dwelling is granted to the mother of the children. Therefore, the model assumes that the mothers keep the dwellings and the children at home. Meanwhile, the leaving partners are subject to the probability of out-migration. In Table 4-6, the divorce-related out-migration probability is estimated from Statistics Canada for Hamilton on the basis of data for 2000. In URM-MicroSim, the probability of the male moving out is 0.23. From the statistical data for 2000, there were 16,270 divorced males and 26,230 divorced females. If the partners of these 9,960 additional divorced females are assumed to have migrated out of the city, the out-migration rate for males after divorce is 23% of the total 42,500 divorced people. It should be noted that the difference between divorced females and males could be due to the fact that a greater percentage of the divorced males remarry or that some of the divorced females lived in other municipalities before returning to Hamilton (in-migration) after their divorce. However, since these are the only data available, they have to be used to have a rough estimation of the divorce-related out-migration probability.

Table 4-6: Divorce-related out-migration probability, Hamilton, 2000

Divorced Out-migrate Probability, Hamilton, 2000	
Population	**Divorced people**
Total	42,500
Males	16,270
Females	26,230
The Difference Between Divorced Females & Males	9,960
Divorced Out-migrate Probability	0.234352941

Nest leaving event

In URM-MicroSim, nest leaving is defined as the process of leaving the parental home and living alone. Children who leave the parental home and then live together (cohabitation or marriage) are categorized under union formation events.

The nest leaving rate is not available from the statistical data, and hence must be estimated. Our basic assumption is that children start to leave their nest after their 18th birthday. Census data from Statistics Canada for Hamilton show that the city has 72,965 people aged 18–24, of whom 29,615 stay with their parents. Therefore, the proportion of individuals between 18 and 24 years of age who have never married and live with their parents is 0.4. Hence, for a single year, the nest leaving rate for individuals between 18 and 24 years of age is 0.09 (1 minus 0.4 and then divided by 6). In addition, for individuals beyond the age of 24 who still live with their parents, it is assumed that the probability is 0.02 since data are not available. It should be noted that nest leaving is a process that strongly depends on housing market opportunities: children who wish to leave the parental home may only do so if a house is available. This implies that the model requires the probability of the intention to leave the nest. Since the data are not available, the probability is estimated on the basis of actual behaviors.

4.2.3 Immigration model

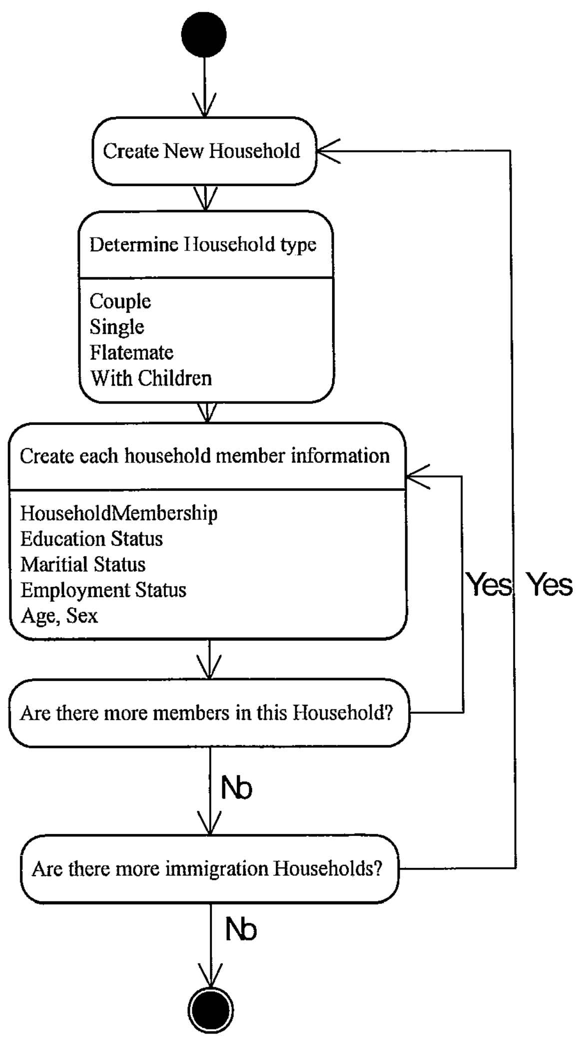

Figure 4-1: Immigration model of URM-MicroSim for Hamilton

Mytilene is the capital city of Lesbos, a Greek island in the Aegean Sea with a relatively small population. The city boasts one university, and most immigrants are students. Therefore, URM-MicroSim for Mytilene assumes that every year only 700 immigrants move into the city, all students. However, tens of thousands of people from diverse households immigrate to Hamilton every year. Hence, the immigration sub-model of URM-MicroSim needs to be rebuilt for Hamilton.

On the basis of the gravity model, the immigrant population of a city is usually determined by its population. Therefore, the number of immigrants in a year is calculated as a constant fraction of the number of existing households. This fraction is calculated by dividing the number of immigrant households by the total number of households in Hamilton in 1996; this is equal to 0.054. The required data are retrieved using census data from Statistics Canada for Hamilton for 1996. On the basis of the total number of immigrant households, the algorithm of the immigration pattern is shown in Figure 4-1. First, for each immigrant household, the household type is determined. There are four types of households: single person; flatmates, consisting of two or more people that share the same dwelling; couples, consisting of a husband and wife who are subject to the probability of living with their children; and single parents, who live only with their children. With the household type in hand, each in-migrating household is created in the model. Next, the number of members in each household is determined on the basis of the household type. Then, all necessary characteristics such as age and sex of all members of the household are determined and inserted into the population dataset. The household type and individual characteristics are simulated by the Monte Carlo model using probabilities distributions calculated directly from census data from Statistics Canada, as shownin Table 4-7 and Table 4-8.

Table 4-7: Proportion of household types for immigrants, Hamilton, 1996

Household Types	Household Size	Total Proportion	Probability of	Rate
Singles		0.228682171	Male	0.525424
Flatmates	2 persons	0.065891473	Male	0.651163
	3 persons	0.027131783		
	4 persons	0.007751938		
	5 persons	0.007751938		
Couples	Without children	0.279069767	Common Law	0.361111
	With children	0.271317829	Common Law	0.057143
	1 child	0.112403101	Male Children	0.577778
	2 children	0.089147287		
	3 children	0.050387597		
	4 children	0.019379845		
Lone Parents	1 child	0.065891473	Single Fathers	0.160000
	2 children	0.031007752	Male Children	0.577778
	3 children	0.011627907		
	4 children	0.003875969		

Table 4-8: Age-specific immigrants by household types, Hamilton, 1996

Age Groups	Singles	Flatmates	Couples with Children	Couples without Children	Lone Parents
0 to 4	0.000	0.000	0.000	0.000	0.000
5 to 9	0.000	0.000	0.000	0.000	0.000
10 to 14	0.000	0.000	0.000	0.000	0.000

15 to 19	0.032	0.080	0.000	0.000	0.032
20 to 24	0.097	0.240	0.104	0.014	0.129
25 to 29	0.355	0.187	0.239	0.096	0.194
30 to 34	0.194	0.120	0.090	0.192	0.226
35 to 39	0.097	0.107	0.075	0.342	0.097
40 to 44	0.097	0.027	0.075	0.123	0.194
45 to 49	0.032	0.053	0.060	0.123	0.097
50 to 54	0.032	0.027	0.060	0.068	0.032
55 to 59	0.032	0.027	0.060	0.027	0.000
60 and More	0.032	0.133	0.239	0.014	0.000

4.3 Results Validation and Discussion

On the basis of the model calibration introduced in the previous section, URM-MicroSim for Hamilton can capture the residential mobility at the disaggregate level. All the dwellings, households, and individuals are simulated and stored by the system. As shown in Figure 4-2, the locations and detailed characteristics of each actor can be identified by the system, which provides detailed information for further spatial analysis.

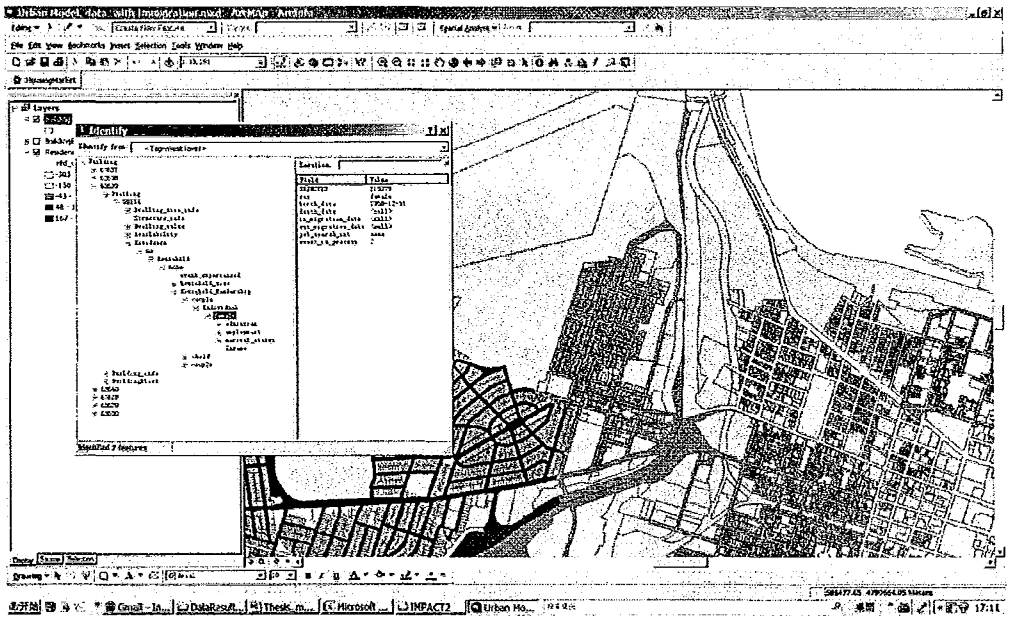

Figure 4-2: User interface of the system indicating information on individuals

In order to test the performance of the system, the results of a simulation should be validated against census data. Since a one-year-simulation for the system takes about 20 h of computational time, it is not feasible to repeat the simulation many times and this is not necessary to compare the consistency of the results. However, since URM-MicroSim for Mytilene shares the same principles and produces consistent results, a single simulation by URM-MicroSim for Hamilton can represent the model performance. Therefore, a five-year simulation from the base year 1996 to 2001 without other inputs or changes is executed by the system in order to validate the model. Still, the census data from Statistics Canada are used to represent the reality. However, the equivalent disaggregate data of Hamilton in 2001 are not available. So, the aggregate census data are used for the evaluation. These data divide the city of Hamilton into 132 census tracts and record the summary statistics of residences inside. In order to validate the model, the results produced from the simulation are aggregated into the same census tracts. The number of individuals inside each census tract is first used to test the results.

Table 4-9 in Appendix shows the comparison between simulation results and census data concerning the numbers of individuals distributed by census tracts. Overall, the results are encouraging. The simulation results are relatively close to the census data. The Pearson correlation is 0.95, which implies that the results from URM-MicroSim can explain the residential mobility at the aggregate level. To supplement correlation analysis, linear regression is used to study the relationship between observed value and modelled results. The scatter plot in Figure 4-3 visually shows that most of the points are distributed along the regression line. However, as identified in Table 4-10, the constant value is negative 503 and the regression coefficient is 1.3. The constant value is far from 0 and the regression coefficient is more than 1 because there are several outliers. It is noticeable that some of the points are far from the regression line. For the census tracts with smaller population sizes in particular, the simulation results are overestimated. And, for the census tracts with larger population sizes, the simulation results are underestimated.

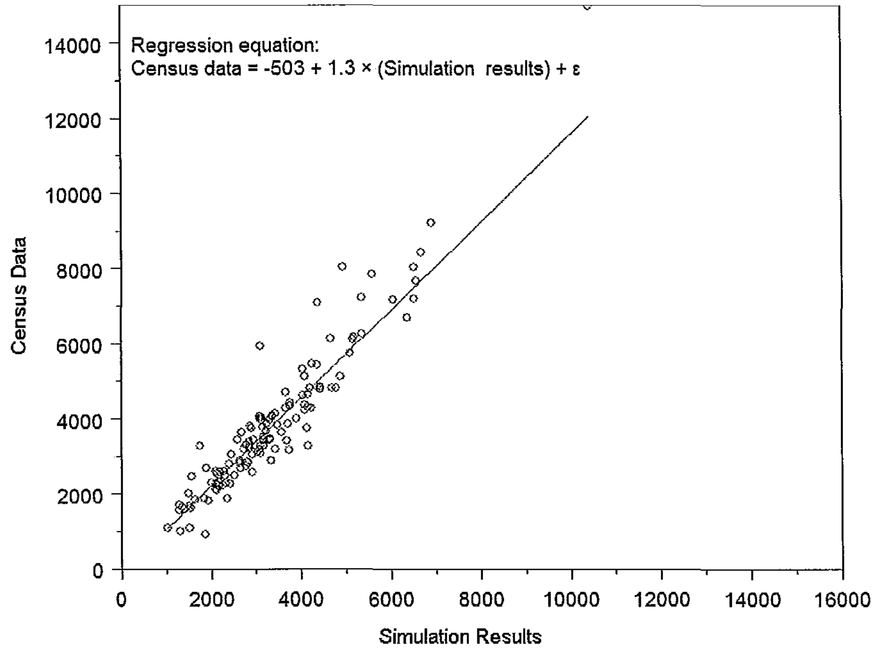

Figure 4-3: The household-changes correlation between simulation and census data, Hamilton, 2001

Table 4-10: Results of regressions with simulation results and census data, Hamilton, 2001

	Value	Standard Error	T Value	Pr(>|t|)
Intercept	-503.5059	145.1384	-3.4691	0.0007
Simulation Results	1.3033	0.0397	32.8535	0

In order to capture and understand these outliers, the measurement of precision for each census tract is used here by the function listed as follows:

$$The\ Measurement\ of\ Precision = \frac{Population\ from\ Census\ Data - Simulated\ Population}{Population\ from\ Census\ Data}$$

In this way, overestimations are recognized as negative values and underestimations are indentified as positive values. For a census tract, the further the measurement of precision is from 0, the higher bias the simulated population size is from the observed data.

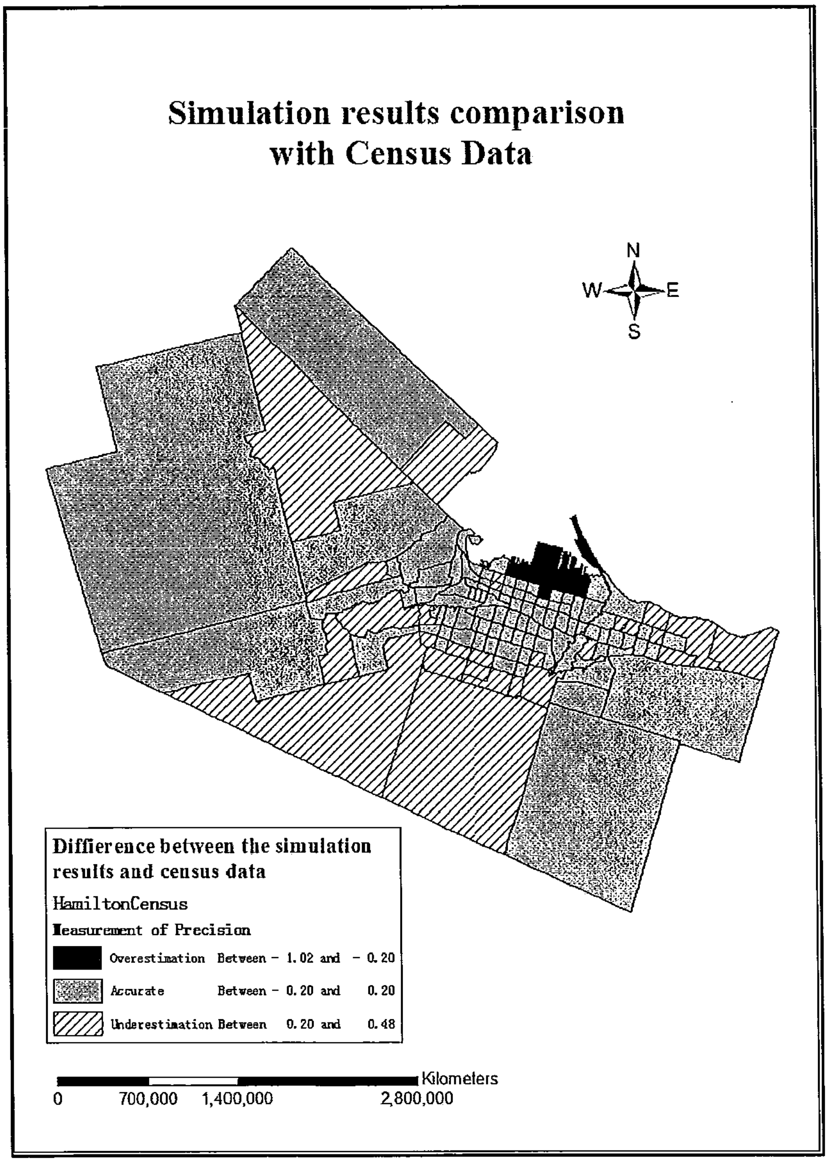

Figure 4-4: The measurement of precision distribution from the comparison between simulation results and census data, Hamilton, 2001

The overall distribution of the measurement of precision is displayed on the map, shown in Figure 4-4. It is noticeable that the overestimated census tracts, which are presented in solid black, are concentrated in industrial zones, where the biggest steel companies, Steel Company of Canada (Stelco) and Dominion Foundries and Steel (Dofasco) are located. These two companies began to pollute the environment one hundred years ago. They used to employ many workers and a lot of dwellings were located nearby to the east of Hamilton. However, the employment numbers have decreased since the late 1990s. At that time, both companies went down. Because of the layoffs and bankruptcies, the employees had to leave the area and moved closer to their new jobs. During this period, no more dwellings were built in the area and several old, cracked buildings were abandoned. As a result, the quality of the neighborhood environment is poor. In contrast, the simulation results from URM-MicroSim for Hamilton show that people are more likely to choose to live in these locations because the houses are cheaper and a large portion of the dwellings are empty. The reason for this overestimation is that the neighborhood condition is not considered inside the housing decision-making process.

On the other hand, the underestimated areas, which are represented with hatching, are distributed in suburban areas. This is due to the urban sprawl problem that has occurred in all Canadian cities. The reason why the model cannot capture this situation can be summarized into two points. First, the behavior of the urban developers is not captured by the housing market, since the housing market in the existing model is fixed and no new dwellings have been input into the market. However, in reality, developers built a lot of dwellings in the suburban areas outside the city. Second, the system cannot capture the neighborhood environment, and this also leads to the overestimation. In reality, people move out of the city not only because there are empty dwellings with lower prizes, but also because they can expect fresh air with quiet and agreeable neighbors. However, these physiological factors are not considered by the model. Therefore, the dwelling search sub-model needs to be redesigned.

Meanwhile, being a microsimulation model, URM-MicroSim for Hamilton can catch the mobility of the population by their demographic characteristics, such as sex, age, marriage, and even income and education. Since income and education need to be linked to the simulations of firm location, which is not currently complete in this model, the age and sex are chosen here to test the performance of the system.

At the aggregate level, the overall distributions by sex and age groups are tested. In Table 4-11 in Appendix, the numbers of population in each age group by males and females are listed. The correlation for females and males between simulation and census data are 0.94 and 0.91 accordingly. It is important to note that the original population dataset is synthetic. Therefore, the model is well preformed at the aggregate level.

At the disaggregate level, census data from Statistics Canada provide the population distribution by sex and age for each census tract. To compare with the simulation results, the mean age for each census tract is calculated to represent the spatial distributions of individuals. The calculation of mean age for each census tract is listed below:

$$Mean\ age = \frac{\sum_{each\ age\ group}(Middle\ value\ of\ the\ age\ group\ \times Population\ size\ in\ the\ age\ group)}{Total\ population\ size\ in\ the\ census\ tract}$$

On the basis of the simulation results, the sex-specific mean age of the population in each census tract is aggregate and shown in Table 4-12 as well as the calculated mean ages from the census data. On the basis of the Pearson correlation, the correlation coefficients of mean ages for females and males between the simulation results and census data are 0.88 and 0.86, respectively. This is a decent result at the disaggregate level. Thus, users of URM-MicroSim can identify the overall trends of the average age for each neighborhood area, which can assist city planners with the planning process.

On the basis of the census data from Statistics Canada for 1996, 6,366 individuals out of 17,162 did not live in the same location within five years. Given that one third of the individuals change their dwellings every five years, the statistical results demonstrated in this section are not bad. Therefore, it can be concluded that URM-MicroSim for Hamilton can capture overall trends of the residential mobility based on age and sex at both aggregate and disaggregate levels.

Chapter 5: Conclusion

5.1 Thesis Conclusion

URM-MicroSim is a microsimulation model capable of modeling urban residential mobility at the disaggregate level. The model has been improved for stable system performance and better user experience. This research has achieved significant reductions in the execution time of the model. Furthermore, strong evidence of consistent results from the model was found. The variation between the results of a simulation repeated 100 times was only 2%, showing that a one-time simulation is representative of the performance of the model.

Within this context, URM-MicroSim has been calibrated for the city of Hamilton. The results from a validation test show that the overall trend in residential mobility is captured by the model. Approximately 95% of individuals' migration patterns, collected from census data, are explained by the simulation results. Furthermore, the overall trend in the sex-specific average age in most of the census tracts is identified by the model. However, several outliers were seen in the simulation results. The main reason for this is that the neighborhood environmental conditions are not taken into consideration in the model when simulating the behavior of an individual searching for a suitable dwelling.

5.2 Thesis Limitation

Although the model produces impressive results, it should be noted that

URM-MicroSim is not complete. To guide future work, the limitations of the model are summarized as follows:

1. The runtime of URM-MicroSim for Hamilton is still too long. A one-year simulation takes almost 30 h, which cannot be tolerated in practice. Users cannot wait for more than one month for the results of a 30-year simulation. Therefore, the system architecture needs to be revised.
2. As introduced in the literature review, URM-MicroSim did not implement housing supply mode due to the lack of data. The current housing market for the system is static and based on synthetic data for 1996. Therefore, there is no interaction between housing demand and supply. Since demand and supply always act together to reflect the dynamic housing market, it is very important to incorporate the behavior of the developers into the model in order to incorporate housing supply into the model. Undoubtedly, this will lead to better performance.
3. The current income sub-model has not been applied and linked with housing demand because the behavior of firms is not included in the existing model. Therefore, the migration is mainly triggered by demographic events. This is not sufficient. Changes in a family's life cycle do not fully account for the housing considerations of a household (Kendig, 1984). Research has shown that employment changes in the local housing and labor market often acts as a trigger

for a residential move (Clark and Withers, 1999). Therefore, the behavior of firms needs to be applied and linked with housing demand.

4. In URM-MicroSim, if a household fails to find a dwelling, it is considered a candidate for out-migration. The out-migration probabilities largely depend on the type of events experienced by the household. However, no statistical data are available for these probabilities. Meanwhile, inside the "search and migration model", the searching criteria for new dwellings are determined by the probabilities of preference. These probabilities largely depend on current household tenure and on the events experienced by the households. These probabilities that determine the desired dwelling are also not available. Instead, for both cases, the same parameters of out-migration and dwelling preference probabilities used by URM-MicroSim Mytilene are also used for Hamilton. However, this may lead to inaccurate results since people in different contexts behave differently. Survey data on these probabilities are required for Hamilton.

5.3 Future Work

Some directions for future work along this line are listed as follows:

1. The system architecture needs to be revised. A faster programming language, C++, and a more suitable database engine, PostGIS, are proposed as introduced in Section 3.4.3.

2. When an individual is searching for a dwelling, the quality of neighborhood is an important factor that needs to be considered in the model. In addition, it is notable that people tend to move to places close to their former dwellings. Therefore, the distance to the current dwelling should be taken into consideration when individuals are looking for a new place to settle in.
3. For the out-migration and dwelling search processes, existing parameters are not accurate. Since no other statistics are available, a comprehensive survey of related behaviors should be designed and conducted for Hamilton.
4. The model should incorporate the behaviour of urban developers in building more dwellings, as studied by Koronios (2009), thus providing a dynamic housing market. In this way, the housing prices are updated within the model instead of being exogenous.
5. The behavior of firms, studied by Maoh (2005), should also be incorporated. This behavior should include failure, mobility, changes in employment, and location choice for establishments. This behavior can be linked to individuals via their jobs and incomes. Consequently, resident migration can be determined on the basis of the location of employment.

With all the proposed improvements, researchers should be able to use URM-MicroSim to study the city of Hamilton comprehensively with detailed urban policies on a large scale. In addition, the differences between microsimulation models and traditional aggregated models (IMULATE and IMPACT) can be fully studied, since all these models are designed and calibrated for the same city. This research may act as a guide for future urban modeling approaches.

References

Adair, A., McGreal, S. and Smyth, A. (2000) House Price and Accessibility: The Testing of Relationships within the Belfast Urban Area, *Housing Studies*, Vol. 15, No. 5, pp. 699-716.

Alonso, W. (1960) A Theory of the Urban Land Market, *Papers and Proceedings, Regional Science Association,* Vol. 6, pp. 149-158.

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

Anas, A. (1983) *The Chicago Area Transportation Land-Use Analysis System*, Final Report, US Department of Transportation, Washington, D.C.

Anas, A. (1987) *Modeling in Urban and Regional Economics*, Harwood Academic Publishers, GmbH, Chur, Switzerland.

Anas, A. and Duann, L.S. (1986) Dynamic Forecasting of Travel Demand, Residence Location, and Land Development: Policy Simulations with the Chicago Area Transportation/Land Use Analysis System. In: Hutchinson, B. and Batty, M. (eds), *Advances in Urban Systems Modeling*, North-Holland, Amsterdam, pp. 299-322.

Anderstig, C. and Mattsson, L.G. (1991) An Integrated Model of Residential and Employment Location in a Metropolitan Region, *Papers in Regional Science*, 70, 167-84.

Anderstig, C. and Mattsson, L.G. (1998) Modelling land-use and transport interaction: evaluationsand policy analysis, In: *Network Infrastructure and the Urban Environment: Recent Advances in Land-Use/Transportation Modelling* (Lundqvist. L., Mattsson, L.G. and Kim, T.J. eds.), 308-328, Springer, Berlin/Heidelberg/New York.

Ballas, D., Clarke, G.P. and Wiemers, E. (2005) Building a Dynamic Spatial Microsimulation Model for Ireland, Population,Space and Place, *Popul. Space Place* 11, 157-172.

Barceló, J. and Casas, J. (2005) Dynamic Network Simulation with AIMSUN, In *Simulation Approaches in Transportation: Recent Advances and Challenges*, Edited by Kitamura, R. and Kuwahara, M., ISBN: 0-387-24108-6, Springer.

Boehm (1981) Tenure Choice and Expected Mobility: A Synthesis, *Journal of Urban Economics,* Vol. 10, Issue 3, November 1981, Pages 375-389

Brown, H.J. (1975) Changes in Workplace and Residential Locations, *Journal of the American Institute of Planners,* Vol. 41, pp. 32-39.

Brown, L. and Harding, A. (2002) Social Modelling and Public Policy: Application of Microsimulation Modelling in Australia. *Journal of Artificial Societies and Social Simulation* 5 (4), 6.

Brown, L.A. and Moore, E.G. (1970) The Intra-Urban Migration Process: A Perspective, *Geografiska Annaler*, 52B, 368–381.

Brown, Laurie, and Harding, A. (2002) Social Modeling and Public Policy: Application of Microsimulation Modeling in Australia, *Journal of Artificial Societies and Social Simulation*, 5 (4).

Burtless, G. (1996) A Framework for Analyzing Future Retirement Income Security in Hanushek, E.A. and Maritato, N.L. (eds.) *Assessing Knowledge of Retirement Behaviour*, Washington D.C.: National Academy Press.

Caldwell, S. B. (1997) CORSIM 3.0 User and Technical Documentation, Ithaca, NY: Cornell University.

Citro and Hanushek (1991) Citro, C. and Hanushek, E.A. (eds.) (1991) *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, Washington, DC.

Clark, W.A.V. (1980) Preface: Looking back at Why Families Move, In: *Why Families Move* (Ed, Rossi, P.H.), SAGE Publications, Beverly Hills and London.

Clark, W.A.V. (1982) Recent Research on Migration and Mobility: A Review and Interpretation, *Progress in Planning,* 18, 1-56.

Clark, W.A.V. and Dieleman, F.M. (1996) Households and Housing: Choice and Outcomes in the Housing Market, *CUPR Press*, Rutgers University, New Jersey.

Clark, W.A.V. and Withers, S.D. (1999) Changing Jobs and Changing Houses: Mobility Outcomes of Employment Transitions, *Journal of Regional Science*, 39(4), 653–673.

Clark, W.A.V., and Huang, Y. (2003) The Life Course and Residential Mobility in British

Housing Markets, *Environment and Planning A*, Vol. 35, pp. 323-339.

Clark, W.A.V., Deurloo, M.C. and Dieleman F.M. (2006) Residential Mobility and Neighbourhood Outcomes, *Housing Studies*, Vol. 21, No. 3, pp. 323-342.

Connerly, C.E. (1986) The Impact of Neighbourhood Social Relations on Prospective Mobility, *Social Science Quarterly*, Vol. 67, pp. 186-194.

Creedy, J. (2001) Tax modelling, *Economic Record*, 77 (237), 189-202, 2001.

Danylo, K. (2003) Microsimulation Model of National Economy MSMNE-02, *Available online: http://dankozub.com/simulation/index.html.*

de la Barra, T. (1982) Modelling Regional Energy Use: A Land Use, Transport and Energy Evaluation Model, *Environment and Planning B: Planning and Design*, 9, 429-443.

de la Barra, T. (1989) *Integrated Land Use and Transport Modeling: Decision Chains and Hierarchies,* Cambridge University Press, Cambridge.

de la Barra, T., Pérez, B. Vera, N. (1984) TRANUS-J: Putting large models into small computers, *Environment and Planning B: Planning and Design*, 11, 87-101.

Dieleman (2001) Modelling Residential Mobility: A Review of Recent Trends in Research, *Journal of Housing and the Built Environment* 16: 249–265, 2001.

Domencich, T. and McFadden, D. (1975) *Urban Travel Demand: A Behavioural Analysis*, North-Holland, Amsterdam.

Duncan, G. and Mcarthur, D. (1997) Paramics technical report, *EPCC-SIAS.*

Echenique, M.H. (1985) The Use of Integrated Land Use Transportation Planning Models: The Cases of Sao Paolo, Brazil and Bilbao, Spain, in *The Practice of Transportation Planning*, M.Florian (ed.), The Hague: Elsevier.

Echenique, M.H. and Williams, I.N. (1980) Developing Theoretically Based Urban Models for Practical Planning Studies, *Sistemi Urbani*, 1, 13-23.

Echenique, M.H., Crowther, D. and Lindsay, W. (1969) A Spatial Model for Urban Stock and Activity, *Regional Studies*, 3:281–312.

Echenique, M.H., Flowerdew, A.D., Hunt, J.D., Mayo, T.R., Skidmore, I.J. and

Simmonds, D.C. (1990) The MEPLAN Models of Bilbao, Leeds and Dortmund, *Transportation Reviews*, 10, pp. 309-22.

Eliasson, J. (2000) Integrated Travel Pattern Modeling: Theory and Estimation of An Integrated Land Use-transportation Model, paper presented at the *IATBR Meeting*.

Ettema, D., de Jong, K., Timmermans, H. and Bakema, A. (2005) PUMA Multi agent Modelling of Urban Systems, In: *45th Congress of the European Regional Science Association*, Vrije Universiteit, Amsterdam.

Favreault, M. and Smith, K.E. (2004) A Primer on the Dynamic Simulation of Income Model (DYNASIM3), Discussion Paper, the Retirement Project, The Urban Institute.

Feijten, P. and Mulder, C.H. (2002) The Timing of Household Events and Housing Events in the Netherlands: A Longitudinal Perspective, *Housing Studies*, Vol. 17, pp. 773-792.

Flowerdew, A.D.J. (1977) An Evaluation Package for A Strategic Land Use/Transportation Plan. In: P. Bonsall, Q. Dalvi and P.Halls (eds.), *Urban Transportation Planning, Abacus Press*, Turnbridge Wells, pp. 241-258.

Fölster S. (1997) Social Insurance Based on Personal Savings Accounts, *European Economy*, No. 1997/4.

Friedman, B., Borning, A., Davis, J., Gill, B., Kahn, P., Kriplean, T. and Lin, P. (2008) Laying the Foundations for Public Participation and Value Advocacy: Interaction Design for a Large Scale Urban Simulation, *Proceedings of the 9th Annual International Conference on Digital Government Research*, Montréal, Canada.

Goldner, W. (1971) The Lowry Model Heritage, *Journal of the American Institute of Planners*, Vol. 37, 100-110.

Golledge, R.G. (1980) A Behavioural View of Mobility and Migration Research, *Professional Geographer*, Vol. 32, pp. 14-21.

Goodman, J.L. Jr. (1976) Housing Consumption Disequilibrium and Local Residential Mobility, *Environment and Planning A*, Vol. 8, pp. 855-874.

Hain, W. and Helberger, C. (1986) Longitudinal Microsimulation of Life Income, In (Orcutt, Merz and Quinke, eds.) *Microanalytic Simulation Models to Support*

Social and Financial Policy, pp. 251-270, North-Holland, Amsterdam.

Harding, A. (1993) Lifetime Income Distribution and Redistribution: Applications of A Microsimulation Model, *Contributions to Economic Analysis*, Vol. 221, Amsterdam: North Holland.

Harris, B. (ed.) (1965) Urban Development Models: New Tools for Planning, *Journal of the American Institute of Planners*, Vol. 31, 90-183.

Hirton, J.E. and Echenique (1979) An Operational Land Use and Transport Model for the Tehran Region, *Iran Transport Research Circular,* 199, 6-7.

Holm, E., Holme, K., Mäkilä, K., Kauppi, M.M. and Mörtvik, G. (2002) The SVERIGE Spatial Microsimulation Model: Content, Validation, and Example Applications, In: *GERUM kulturgeografi*. Umeå, Sweden: CERUM Kulturgeografi, Kulturgeografiska Institutionen/SMC, Umeå University.

Hooimeijer, P. and A. Oskamp (2000) Locsim: Microsimulation of Households and Housing Market, Paper delivered at *the 10th Biennial Conference of the Australian Population Association*, Melbourne, Australia.

Hooimeijer, P. and Oskamp, A. (1999) Advances in the Microsimulation of Demographic Behavior, In: *Population Issues and Interdisciplinary Focus*, Kluwer Academic/Plenum Publishers, New York.

Huang, Z. and Williamson, P. (2002) A Comparison of Synthetic Reconstruction and Combinatorial Optimization Approaches to the Creation of Small-area Microdata Working Paper, Department of Geography, University of Liverpool.

Hunt, J.D. (1994) Calibration the Naples Land Use and Transport Model, *Environment and planning B*, 21, 569-590.

Hunt, J.D. and Echenique, M.H. (1993) Experiences in the Application of the MEPLAN Framework for Land Use and Transportation Interaction Modeling, *Proc. 4th National Conference on the Application of Transportation Planning Methods,* Daytona Beach, FL, pp.723-754.

Hunt, J.D. and Simmonds, D.C. (1993) Theory and Application of an Integrated Land-Use and Transport Modeling Framework, *Environment and Planning 20B*, 221-244.

Immervoll, H., O'Donoghue, C. and Sutherland, H. (1999) An Introduction to EUROMOD, EUROMOD Working Paper No. 0/99, Department of Applied Economics, University of Cambridge.

International Microsimulation Association (2005) The International Journal of Microsimulation, *Available Online: http://www.microsimulation.org.*

Johnston, R.A., Gao, S. and Shabazian, D.R. (2003) Uplan: A Versatile Urban Growth Model for Transportation Planning, *Proceedings of the TRB Conference*, Washington, D.C.

Kanaroglou, P. and Scott, D. (2002) Integrated Urban Transportation and Land-use Models for Policy Analysis, In: XDijst, V., Schenkel, W. and Thomas, I. (eds.), *Governing Cities on the Move: Functional and Management Perspectives on Transformations of European Urban Infrastructures,* Aldershot (UK): Ashgate, pp. 43-73.

Kanaroglou, P.S. and Anderson, W.P. (1997), Emissions from Mobile Sources in Urban Areas: An Integrated Transportation and Land-use Approach, *Proceedings of the Fifth Conference of Environmental Science and Technology*, Vol. A, Molyvos, Lesvos, Greece, September 1-5, pp. 467-475.

Kendig, H.L. (1984) Housing Careers, Life Cycle and Residential Mobility: Implications for the Housing Market, *Urban Studies*, Vol. 21, pp. 271-283.

Koronios, M. (2009) Exploring and Modeling the Urban Land Development Process in Hamilton, Ontario, M.A. Thesis, McMaster University, Hamilton, Ontario, Canada.

Kulu, H. and Billari, F. (2004) Multilevel Analysis of Internal Migration in a Transitional Country: The Case of Estonia, *Regional Studies*, 38 (6): 679-696.

Landis, J. (1994) The California Urban Futures Model: A new Generation of Metropolitan Simulation Models, *Environment and Planning B*, 21, 399-420.

Landis, J. (1995) Imagining Land Use Futures: Applying the California Urban Futures Model, *Journal of the American Planning Association*, 61, 438-457.

Landis, J. and Zhang, M. (1998) The Second Generation of the California Urban Futures Model, Part 2: Specification and Calibration Results of the Land Use Change Submodel, *Environment and planning B*, 25, 758-824.

Liu, R., Vilet, D.V., and Watling, D.V. (1995) Dracula: Dynamic Route Assignment Combining User Learning and Microsimulation, *PTRC*, Vol. E, pp. 143-152.

Lowry, I. S. (1964) *A Model of Metropolis,* RM-4035-RC. Santa Monica (CA): The Rand Corporation.

Malan, R. and Bredemeyer, D. (1999) Functional Requirements and Use Cases, *Available Online: http://www.bredemeyer.com/pdf_files/functreq.pdf.*

Maoh, H. (2005) Firm Demography in Urban Areas: An Agent-based Approach, Ph.D. Thesis, McMaster University, Hamilton, Ontario, Canada.

Maoh, H., Kanaroglou, P., Scott, D., Páez, A. and Newbold, B. (2009) IMPACT: An Integrated GIS-based Model for Population Aging Consequences on Transportation, *Computers, Environment and Urban Systems*, 33: 200-210.

Martinez, F. (1992) The Bid Choice Land Use Model: An Integrated Economic Framework, *Environment and planning A*, 24, 871-885.

Martinez, F.J. (1996) MUSSA: A Land Use Model for Santiago City, *Transportation Research Record 1552*: 126-134.

McCarthy, K. F. (1976) The Household Life Cycle and Housing Choices, Papers of *the Regional Science Association*, Vol. 37, pp. 55-80.

McCarthy, K.F. (1976) The Household Life Cycle and Housing Choices, *Papers of the Regional Science Association*, Vol. 37, pp. 55-80.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behaviour, In: P. Zaremka (Ed.) *Frontiers in Econometrics*, pp. 105-142. New York: Academic Press.

Merz, J. (1991) Microsimulation - A Survey of Principles, Developments and Applications, *International Journal of Forecasting,* 7, 77-104.

Miller, E., Hunt, J., Abraham, J. and Salvini, P. (2004) Microsimulating Urban Systems, *Computers, Environment and Urban Systems*, Vol. 28, pp. 9-44.

Miller, E.J., Salvini, P. (2005) ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems, *Networks and Spatial Economics 5*, 2005 Springer Science & Business Media, Inc. Netherlands. pp 217-234.

Mitton, L., Sutherland, H. and Weeks, M. (2000) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*, Cambridge University Press.

Mitton, L., Sutherland, H., Weeks, M. (Eds) (2000) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*, Cambridge University Press, Cambridge.

Modelistica (2007) TRANUS: Integrated Land Use and Transport Modeling System , General Description of TRANUS, *Available Online: http://www.modelistica.com.*

Moeckel, R., Schurmann, C. and Wegener, M. (2002) Microsimulation of Urban Land Use, *Proceedings of the 42nd European Congress of the Regional Science Association*, Dortmund.

Moore, T. and Thorsnes, P. (1994) *The Transportation/Land Use Connection*, American Planning Association, Planning Advisory Service, Report Number 448/449, Chicago.

Morin, R. and Brown, V. (1999) Scripting Languages: A Cross-OS Perspective, *The Journal of Macintosh Technology*, Vol. 15, Issue Number: 9.

Mulder, C.H. and Wagner, M. (1993) Migration and Marriage in the Life Course: A Method for Studying Synchronized Events, *European Journal of Population*, Vol. 9, pp. 55-76.

Mulder, C.H. and Wagner, M. (1998) First-Time Home-Ownership in the Family Life Course: A West-German-Dutch Comparison, *Urban Studies*, 35(4), 687-713.

Myers, G. J. (1976) *Software Reliability: Principles and Practices*, Wiley, New York.

O'Donoghue C. (2001) Dynamic Microsimulation: A Methodological Survey, *Brazilian Economic Journal of Economics 4(2)* [on-line journal], Paper Available Online: http://www.beje.decon.ufpe.br/v4n2/cathal.htm.

O'Donoghue, C. (2001) Redistribution in the Irish Tax-Benefit System, Ph.D., London School of Economics.

OECD (Organisation for Economic Cooperation and Development) (1996) Policy Implications of Ageing Populations: Introduction and Overview, OECD Working Paper No. 33, Paris.

Orcutt, G., Merz, J. and Quinke H. (eds.) (1986) *Microanalytic Simulation Models to*

Support Social and Financial Policy, North Holland, Amsterdam.

Orcutt, G.H. (1957) A New Type of Socio-economic System, *Review of Economics and Statistics*, 39, 116-123.

Oskamp, A. (1994) LOCSIM: A Probabilistic Model of Choice Heuristics, *Netherlands Journal of Housing and the Built Environment*, vol. 9, pp. 285-309.

Parisi, V. (2003) A Cross Country Simulation Exercise Using the DIECOFIS Corporate Tax Model, *European Commission IST Programme DIECOFIS*, Work Package No. 7, Deliverable No. 7.2.

Putman, S.H. (1983) *Integrated Urban Models: Policy Analysis of Transportation and Land Use*. London: Pion.

Quadstone Limited (2000) *Paramics User Guide Version 3*, Edinburgh, UK.

Rajesh, M. (2008) Using Visual Studio Tools for Office in Visual Studio 2008, *Available Online: http://www.developer.com/net/vb/article.php/3743191.*

Roman, S. (2002) *Writing Excel Macros with VBA*, Sebastopol CA: O'Reilly.

Rossi, P.H. [1955] (1980) *Why Families Move,* Sage Publications, Beverly Hills and London.

Ryan, J. (2008) Population Synthesis Techniques: Creating Input Data for Microsimulation Models, M.A. Thesis, McMaster University, Hamilton, Ontario, Canada.

Short, J.R. (1978) Residential mobility, *Progress in Human Geography*, Vol. 2, pp. 419-447.

Siebers, P.O., Aickelin, U., Celia, H., Clegg, C.W. (forthcoming b) (2007) A Multi-Agent Simulation of Retail Management Practices, *Summer Computer Simulation Conference 2007 (SCSC 2007)*, San Diego, CA.

Simmons, J.W. (1968) Changing Residence in the City: A Review of Intra-urban Mobility, *Geographical Review*, Vol. 58, pp. 622-651.

Smith, L., Beckman, R., Anson, D., Nagel, K. and Williams, M. (1995) TRANSIMS: Transportation Analysis and Simulation System, *Proceedings of the 5th National Transportation Planning Methods Applications Conference*, Seattle.

Speare, A. Jr. (1974) Residential Satisfaction as an Intervening Variable in Residential Mobility, *Demography*, Vol. 11, pp. 173-188.

Svinterikou, M. (2007) Microsimulation Models in Geography Using Object-Oriented Programming: An Application to Residential Mobility, Ph.D. Thesis, University of the Aegean, Lesvos, Greece.

Timmermans, H. (2003) The Saga of Integrated Land Use-transport Modeling: How Many More Dreams before We Wake Up?, paper presented at *10th International Conference on Travel Behaviour Research*, Lucerne, International Association for Travel Behaviour Research, 10-15 August.

Vencatasawmy, C.P., Holm, E., Rephann, T., Esko, J., Swan, N., Öhman, M., Åström, M., Alfredsson, E., Holme, K. and Siikavaara, J. (1999) Building a Spatial Microsimulation Model, *11th European Colloquium on Quantitative and Theoretical Geography in Durham*, England.

Waddell, P. (2001) Between Politics and Planning: UrbanSim as a Decision Support System for Metropolitan Planning, in *Building urban planning support systems: combining information, models, visualization,* Klosterman, R. and R. Brail (eds), Center for Urban Policy Research.

Waddell, P. (2004) Urban Simulation at the Intra-Metropolitan Scale: Model Design and Application in the Central Puget Sound Region, *Paper presented at The State of the Art in Regional and Urban Modeling*, Seville, Spain.

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, Preprint of an article that appeared in *Networks and Spatial Economics*, Vol. 3, No. 1, pp. 43-67.

Wagner, P., Wegener, M. (2007) Urban Land Use, *Transport and Environment Models: Experiences with an Integrated Mircroscopic Approach*, disP 170, 45-56.

Wegener, M. (1995) Current and Future Land Use Models, *Proceedings of Land Use Modelling Conference*, Arlington (TX).

Wegener, M., Spiekermann, K. (1996) The Potential of Microsimulation for Urban Models, In: Clarke, G. (ed.): *Microsimulation for Urban and Regional Policy Analysis*, European Research in Regional Science 6, London: Pion, 146-163.

Weiner (1999) *Urban Transportation Planning in the United States: An Historical Overview*, Praeger; Rev Exp Su Edition.

Wilson, A.G. and Pownall, C.E. (1976) A New Representation of the Urban System for Modelling and for the Study of Micro-level Interdependence, *Area*, Vol. 8, pp. 246-254.

Wolfson, M. C. (1995) XEcon: An Experimental / Evolutionary Model of Economic Growth, *Statistics Canada, Analytical Studies Branch Research Paper 67*, ACE Evolutionary Micro-Macro Model with Endogenous Growth.

Wolpert, J. (1965) Behavioural Aspects of the Decision to Migrate, *Papers and Proceedings, Regional Science Association*, Vol. 15, pp. 159-169.

Wolpert, J. (1966) Migration as an Adjustment to Environmental Stress, *Journal of Social Issues*, Vol. 22, pp. 92-102.

Yen, Y. (2003) An Application of Transportation Land Use Model to Tainan Metropolitan Area, *Journal of the Eastern Asia Society for Transportation Studies*, Vol.5, October.

Zaidi, A. and Rake, K. (2001) Dynamic Microsimulation Models: A Review and Some Lessons for SAGE, *SAGE Discussion Paper no.2*, London School of Economics.

Zhang, Y., Xu, X. and Yan X. (1999) A Review of Literature on Urban Land Use and Rapid Rail Transit, *1999 International Conference on Review and Prospect of Chinese Cities for Half Century*, Zhongshan City, Guangdong, PRC.

Appendix

Table 4-9: The comparison of numbers of individuals by census tracts between simulation and census data for Hamilton, 2001

	Census	Simulation			Census	Simulation	
CTID96	2001	2001	MUNICIPAL	CTID96	2001	2001	MUNICIPAL
5370001.01	2,014	1,487	Hamilton	**5370045.00**	3,762	4,134	Hamilton
5370001.02	5,466	4,236	Hamilton	**5370046.00**	3,189	3,074	Hamilton
5370001.03	8,042	4,916	Hamilton	**5370047.00**	3,043	2,912	Hamilton
5370001.04	5,322	4,028	Hamilton	**5370048.00**	1,658	1,354	Hamilton
5370001.05	3,283	1,746	Hamilton	**5370049.00**	2,471	1,573	Hamilton
5370001.06	5,426	4,369	Hamilton	**5370050.00**	4,364	4,085	Hamilton
5370001.07	4,054	3,348	Hamilton	**5370051.00**	4,626	4,046	Hamilton
5370002.01	3,652	2,692	Hamilton	**5370052.00**	3,839	3,480	Hamilton
5370002.02	7,100	4,369	Hamilton	**5370053.00**	3,446	3,272	Hamilton
5370002.03	3,457	2,945	Hamilton	**5370054.00**	2,874	3,328	Hamilton
5370002.04	5,130	4,092	Hamilton	**5370055.00**	3,154	3,740	Hamilton
5370003.01	3,995	3,090	Hamilton	**5370056.00**	3,409	3,685	Hamilton
5370003.02	3,517	3,168	Hamilton	**5370057.00**	3,191	3,425	Hamilton
5370003.03	3,291	2,785	Hamilton	**5370058.00**	2,561	2,920	Hamilton
5370003.04	6,199	5,167	Hamilton	**5370059.00**	3,288	4,160	Hamilton
5370004.01	3,269	2,995	Hamilton	**5370060.00**	2,732	2,780	Hamilton
5370004.02	4,328	3,763	Hamilton	**5370061.00**	3,483	3,307	Hamilton
5370005.01	6,141	5,131	Hamilton	**5370062.00**	3,457	3,312	Hamilton
5370005.02	3,976	3,305	Hamilton	**5370063.00**	3,440	2,581	Hamilton
5370005.03	4,036	3,127	Hamilton	**5370064.00**	1,849	1,645	Hamilton
5370006.00	4,836	4,363	Hamilton	**5370065.00**	2,574	2,184	Hamilton
5370007.00	3,272	3,162	Hamilton	**5370066.00**	5,735	5,092	Hamilton
5370008.00	2,554	2,117	Hamilton	**5370067.00**	1,872	2,365	Hamilton
5370009.00	3,876	3,721	Hamilton	**5370068.00**	998	1,300	Hamilton
5370010.00	3,111	3,039	Hamilton	**5370069.00**	920	1,862	Hamilton
5370011.00	2,228	2,267	Hamilton	**5370070.00**	2,271	2,435	Hamilton
5370012.00	1,599	1,391	Hamilton	**5370071.00**	7,184	6,519	Hamilton
5370013.00	2,835	2,629	Hamilton	**5370072.02**	4,078	3,353	Hamilton
5370014.00	2,693	2,666	Hamilton	**5370072.03**	7,235	5,354	Hamilton
5370015.00	1,567	1,271	Hamilton	**5370072.04**	3,740	2,885	Hamilton
5370017.00	3,201	2,721	Hamilton	**5370073.00**	1,091	1,523	Hamilton
5370019.00	3,790	3,139	Hamilton	**5370080.01**	2,313	2,179	Stoney Creek

5370020.00	4,220	4,079	Hamilton	**5370080.02**	3,863	3,239	Stoney Creek
5370021.00	4,831	4,688	Hamilton	**5370081.00**	2,268	2,097	Hamilton
5370022.00	4,810	4,772	Hamilton	**5370082.00**	3,674	3,218	Hamilton
5370023.00	2,284	2,336	Hamilton	**5370083.00**	2,488	2,311	Stoney Creek
5370024.00	2,601	2,297	Hamilton	**5370084.01**	2,806	2,401	Stoney Creek
5370025.00	2,859	2,789	Hamilton	**5370084.02**	3,252	2,843	Stoney Creek
5370026.01	3,451	3,169	Hamilton	**5370084.03**	2,292	1,999	Stoney Creek
5370026.02	1,629	1,533	Hamilton	**5370084.04**	1,710	1,291	Stoney Creek
5370026.03	1,874	1,841	Hamilton	**5370084.05**	3,049	2,439	Stoney Creek
5370026.04	1,672	1,508	Hamilton	**5370085.01**	4,697	3,666	Stoney Creek
5370026.05	4,432	3,753	Hamilton	**5370085.02**	6,284	5,346	Stoney Creek
5370026.06	6,143	4,655	Hamilton	**5370085.03**	2,687	1,889	Stoney Creek
5370027.00	1,091	1,032	Hamilton	**5370086.00**	4,049	3,086	Stoney Creek
5370028.00	2,884	2,634	Hamilton	**5370100.00**	4,293	3,654	Glanbrook
5370029.00	4,301	4,145	Hamilton	**5370101.00**	7,852	5,573	Glanbrook
5370030.00	4,275	4,223	Hamilton	**5370120.00**	5,948	3,100	Ancaster
5370031.00	2,133	2,100	Hamilton	**5370121.00**	1,820	1,932	Ancaster
5370032.00	3,394	2,837	Hamilton	**5370122.00**	9,213	6,903	Ancaster
5370033.00	3,284	3,102	Hamilton	**5370123.00**	6,687	6,375	Ancaster
5370034.00	4,855	4,426	Hamilton	**5370124.00**	3,817	2,876	Ancaster
5370035.00	4,138	3,442	Hamilton	**5370130.00**	8,045	6,514	Ancaster
5370036.00	2,617	2,101	Hamilton	**5370131.00**	4,831	4,189	Dundas
5370037.00	2,503	2,515	Hamilton	**5370132.00**	3,077	3,097	Dundas
5370038.00	3,631	3,567	Hamilton	**5370133.00**	8,441	6,685	Dundas
5370039.00	5,134	4,886	Hamilton	**5370140.00**	14,988	10,420	Flambrough
5370040.00	2,111	2,135	Hamilton	**5370141.00**	4,010	3,896	Flambrough
5370041.00	2,210	2,167	Hamilton	**5370142.00**	7,680	6,545	Flambrough
5370042.00	2,822	2,832	Hamilton	**5370143.00**	3,961	3,134	Flambrough
5370043.00	4,790	4,441	Hamilton	**5370144.00**	7,157	6,040	Flambrough
5370044.00	4,642	4,158	Hamilton				

Table 4-11: Comparison of sex-and age-specific population distributions between simulation and census data, Hamilton, 2001

Males/ 2001	**Census Data**	**Simulation Results**	**Females/ 2001**	**Census Data**	**Simulation Results**
0-4	14,156	12,261	**0-4**	13,415	13,404
5-9	16,072	16,498	**5-9**	15,396	18,042

10-14	16,618	15,381	**10-14**	15,652	16,401
15-19	16,744	14,608	**15-19**	15,618	15,563
20-24	15,580	14,078	**20-24**	15,799	14,858
25-29	14,746	16,210	**25-29**	15,325	17,543
30-34	16,437	20,801	**30-34**	16,786	20,733
35-39	19,668	19,851	**35-39**	19,807	19,124
40-44	19,704	18,394	**40-44**	19,833	17,605
45-49	17,200	16,188	**45-49**	18,124	14,908
50-54	15,581	16,148	**50-54**	16,290	14,704
55-59	12,017	15,208	**55-59**	12,311	13,611
60-64	9,978	12,617	**60-64**	10,617	12,523
65-69	9,049	10,043	**65-69**	10,173	10,218
70-74	8,175	7,837	**70-74**	10,214	5,863
75-79	6,601	6,994	**75-79**	9,350	5,125
80-84	3,467	5,518	**80-84**	5,756	3,570
85 and Over	2,163	2,853	**85 and Over**	4,867	2,851

Table 4-12: Comparison of sex-specific mean age distributions by census tracts between simulation and census data, Hamilton, 2001

Mean Age/ CTID96	**Simulation (2001)**		**Census Data (2001)**		**Mean Age/ CTID96**	**Simulation (2001)**		**Census Data (2001)**	
	Female	**Male**	**Female**	**Male**		**Female**	**Male**	**Female**	**Male**
5370001.01	37.92	34.05	34.03	32.61	**5370030.00**	36.69	34.70	38.42	36.11
5370001.02	31.18	30.09	27.49	27.34	**5370031.00**	36.89	34.18	37.55	34.40
5370001.03	31.99	30.37	29.68	28.62	**5370032.00**	36.46	33.79	40.70	37.44
5370001.04	31.94	30.54	28.45	27.32	**5370033.00**	36.54	33.89	37.02	34.87
5370001.05	34.99	31.63	39.71	33.69	**5370034.00**	38.19	36.11	38.90	35.55
5370001.06	32.13	32.51	29.95	29.27	**5370035.00**	36.55	35.51	40.09	37.04
5370001.07	31.42	30.70	29.13	28.57	**5370036.00**	37.16	37.00	35.08	34.97
5370002.01	32.99	34.05	31.95	30.42	**5370037.00**	45.72	43.20	53.25	46.33
5370002.02	37.21	34.24	40.08	36.06	**5370038.00**	45.42	40.75	53.27	47.58
5370002.03	33.07	32.76	31.01	30.65	**5370039.00**	40.03	37.02	41.99	38.16
5370002.04	33.25	32.48	30.82	29.79	**5370040.00**	38.19	37.46	38.25	35.91
5370003.01	41.65	38.07	41.69	38.57	**5370041.00**	36.21	34.70	43.28	36.33
5370003.02	38.50	36.11	36.73	34.49	**5370042.00**	36.67	34.45	36.91	35.28
5370003.03	35.51	33.78	33.35	31.98	**5370043.00**	39.27	36.08	41.15	36.31
5370003.04	37.28	34.93	36.65	34.25	**5370044.00**	41.63	38.65	43.09	39.78

5370004.01	40.01	37.86	39.93	36.90	**5370045.00**	39.01	36.94	39.53	37.02
5370004.02	38.27	36.04	36.57	34.25	**5370046.00**	38.49	35.09	42.39	37.64
5370005.01	39.46	36.65	40.36	36.61	**5370047.00**	41.90	39.01	45.61	39.90
5370005.02	36.67	35.06	35.28	32.95	**5370048.00**	38.94	37.78	38.10	35.30
5370005.03	38.47	37.30	37.13	35.98	**5370049.00**	38.47	35.59	45.75	37.39
5370006.00	40.60	38.28	42.19	40.03	**5370050.00**	40.74	38.47	44.61	38.68
5370007.00	43.63	37.02	47.19	41.74	**5370051.00**	36.02	35.09	37.66	34.64
5370008.00	43.80	39.43	52.70	45.52	**5370052.00**	36.51	34.28	37.72	34.83
5370009.00	36.83	33.90	38.52	34.92	**5370053.00**	34.69	33.50	34.17	33.71
5370010.00	40.71	37.44	42.58	39.63	**5370054.00**	35.23	33.91	36.23	33.92
5370011.00	41.83	39.74	45.92	43.05	**5370055.00**	36.80	32.76	35.98	34.81
5370012.00	39.72	36.75	43.07	37.67	**5370056.00**	38.06	35.96	40.16	37.71
5370013.00	43.19	40.36	44.48	43.53	**5370057.00**	37.08	36.03	37.22	33.93
5370014.00	42.91	41.14	44.62	42.92	**5370058.00**	35.50	33.91	36.04	33.89
5370015.00	43.09	40.05	50.16	43.72	**5370059.00**	34.25	33.14	34.20	33.22
5370017.00	37.65	36.58	37.87	36.61	**5370060.00**	36.36	33.99	38.26	35.84
5370019.00	39.05	34.76	41.93	39.01	**5370061.00**	35.00	34.37	34.60	33.34
5370020.00	40.14	36.05	41.71	38.38	**5370062.00**	35.78	34.43	35.84	34.00
5370021.00	37.11	34.98	38.21	35.95	**5370063.00**	35.23	34.70	33.54	32.70
5370022.00	41.46	37.09	44.20	39.89	**5370064.00**	37.79	38.35	36.93	37.30
5370023.00	41.41	39.85	44.73	41.67	**5370065.00**	33.96	32.94	31.08	29.58
5370024.00	43.79	38.62	43.39	41.97	**5370066.00**	35.30	34.11	33.69	32.93
5370025.00	39.73	37.46	41.19	39.82	**5370067.00**	33.31	31.18	34.24	33.60
5370026.01	44.64	40.91	46.21	43.39	**5370068.00**	32.59	36.02	34.46	32.28
5370026.02	46.12	42.17	47.23	45.19	**5370069.00**	33.59	29.79	39.44	35.81
5370026.03	38.25	36.74	37.21	37.12	**5370070.00**	36.56	34.94	35.98	34.64
5370026.04	39.45	38.95	40.59	40.23	**5370071.00**	37.31	34.28	35.47	32.81
5370026.05	38.80	37.37	38.37	36.19	**5370072.02**	42.91	37.98	43.27	38.23
5370026.06	33.82	32.30	30.64	29.47	**5370072.03**	37.58	35.39	35.74	33.91
5370027.00	41.70	38.45	45.30	41.32	**5370072.04**	36.87	35.92	34.39	34.18
5370028.00	41.27	38.70	44.73	40.97	**5370073.00**	38.41	36.65	42.14	38.41
5370029.00	41.82	37.97	44.78	41.27	**Total**	38.05	35.87	39.07	36.38