Testing Uncertainty in ILUTE - An Integrated Land Use -Transportation Micro-simulation Model of Demographic Updating

Seyed Ahad Beykaei1* and Eric J Miller2
Department of Civil Engineering, University of Toronto, Canada

Abstract
In microsimulation models, the behaviours of agents as intelligent objects are modeled through various behavioral modeling techniques, which give rise to uncertainties in the modeling workflow and predicted outputs. ILUTE (Integrated Land Use Transportation Environment model) is an agent-object-based microsimulation model system designed to simulate different activities in a city, in which many agents (persons, families, households, firms, etc.) intelligently and/or randomly act and interact in a complex way. A 150,000 (150K) synthetic base year (1986) household sample of the Greater Toronto and Hamilton Area (GTHA) population is used as the starting point for undertaking twenty-year history and land use system in a comprehensive solution space. Such integrated urban system models have been under development since the 1960s. A variety of model systems have been developed over time, including PECAS [1], MEPLAN [2], TRANUS [3], ITLUP [4], ILUTE [5], MUSSA [6], Delta [7], and UrbanSim [8].

LUTMs vary in terms of their level of spatial (and associated socio-economic) aggregation, as well as the extent to which they are deterministic or stochastic in their determination of process outcomes. The focus of this paper is on agent-based microsimulation LUTMs, in which system behavior is the emergent outcome of the decisions of individual agents (persons, households, firms), where these decisions usually are probabilistically determined through Monte Carlo simulation. Such models are motivated by the intrinsically non-linear nature of the behavior of significantly heterogeneous populations [9]. Delta [7], UrbanSim [8], and ILUTE [10] are three examples of highly disaggregated LUTM frameworks. This paper focuses on the ILUTE model system.

Of course, no LUTM model system is able to simulate an urban system with a perfect accuracy. This is due to the complexity of urban region's system on the one hand, and uncertain and changeable (stochastic) behavior of agents in the system on the other. In addition, human beings are the intelligent objects of an urban system, making the simulation of their behaviors and decision making processes over time is not a straightforward task for modeling. Given this, it is essential to analyze and test the uncertainties and stochastic properties associated with the microsimulation modeling processes. This paper tests the stochastic behavior of ILUTE's demographic updating module by running and comparing the model system forecasts using many random seeds for the same base scenario. ILUTE is an integrated-disaggregated land use transportation modeling framework designed to simulate dynamic evolution of urban spatial form, economic structure, demographics, and travel behavior over time for the Greater Toronto and Hamilton Area (GTHA). Several outputs simulated by the tested module are statistically scrutinized to understand the stochastic behavior of the ILUTE model.

An overview of ILUTE model system
ILUTE (Integrated Land Use Transportation Environment model) is an agent-object-based microsimulation model system designed to dynamically simulate the evolution of demographics, land use, and travel within urban areas [11]. It is a comprehensive, integrated solution space containing both transportation and land use components intertwined to explore or project the impact of different policies on an urban region. ILUTE has been under development at University of Toronto for several years [9,10,12-14].

Figure 1 illustrates a high-level flowchart of ILUTE simulation engine consisting of nine main modules: base year synthesis, demographic updating of agents, labour market, housing market, auto ownership model, commercial vehicle movement model, activity-based daily travel (TASHA), road and transit network assignment model, and environmental model [14]. The simulation starts from a base year representing the initial state of an urban system towards a target end year to simulate the evolution of the whole urban system in a

*Corresponding author: Seyed Ahad Beykaei, Assistant Professor, University of Mazandaran, Iran, Tel.: +98 11 3530 3565; E-mail: beykaei@gmail.com , miller@ecf.utoronto.ca

Received January 17, 2017; Accepted February 21, 2017; Published February 24, 2017

Citation: Beykaei SA, Miller EJ (2017) Testing Uncertainty in ILUTE - An Integrated Land Use -Transportation Micro-simulation Model of Demographic Updating. J Civil Environ Eng 7: 264. doi: 10.4172/2165-784X.1000264

Copyright: © 2017 Beykaei SA, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
monthly or a yearly time-step depending on the prediction model used. An extensive time-series database is assembled from a wide variety of sources to support the model development activity of ILUTE including census, a household travel survey, real estate, demographic, and special data. More details about input data are summarized by Miller et al. Several exogenous inputs such as interest rates, energy prices, in- and out-migration rates, and zoning are fed to the model system at each simulation time-step. In ILUTE, the evolution of system state over time is defined in terms of several agents including individual persons, families, households, dwelling units, firms, etc. that collectively define the urban system being modelled [15]. Agents play an important role in ILUTE's evolutionary engine as the decision making units, in which their behaviours and attributes are simulated over time. First, lists of aggregate distributions of agents are synthesized for the base year to generate disaggregated sets of statistically representative agents. In this regard, the interconnections of agents and objects, and their attributes and descriptions in the urban system are simulated. At the next time-step, the population demographics such as aging and mortality are updated. Family/household composition (such as marriage, divorce, births, adult children leaving/returning home) and school participation are also updated in the demographic updating module. The rest of key processes that are explicitly modelled within the ILUTE evolutionary engine that operate upon and change over time are listed below [16]:

1. Supply of new housing and commercial floorspace.
2. Firm growth/decline, location/relocation resulting in changes in the amount, type and location of employment.
3. Changes in labour force and school participation.
4. Changes in household residential location.
5. Changes in household auto ownership levels (also persons' possession of driver's licences and transit passes).
6. Commercial vehicle movements.
7. Person-based activity and travel.

The focus of this paper is testing the stochastic behaviour of demographic updating by examining the variability of predicted outputs in response to many random seed starting points. In the next section, a brief overview of the ILUTE's demographic updating module is presented.

**Demographic updating module**

The demographic update component of ILUTE updates socio-demographic attributes of person, family, and household agents throughout the simulation run. It is concerned with how life-changing events and major life decisions affect the demographic characteristics of the simulated population over time [17]. Individuals, families, and households are the three primary agents in ILUTE responsible for all demographic decisions. It updates residential population demographics throughout the simulation in yearly time steps. Several demographic attributes such as births, deaths, divorces, marriages, in-migration, out-migration, and dwellings are updated at each simulation year. New agents (persons, families, and households) are introduced in the model through birth and in-migration, while agents exit from the model through death and out-migration events. Unions between agents are formed through a marriage market, while a divorce model dissolves existing ones [18]. Figure 2 illustrates the model structure of the demographic updating component in ILUTE.

Historical validation of demographic updating module for the twenty-year of simulations (1986-2006) have been run, with model outputs being compared to Canadian Census data and Transportation Tomorrow Survey (TTS) data for 1991, 1996, 2001 and 2006. Figure 3 presents the simulated vs. historical family and household type distribution for four years (based on Census Data). The results demonstrate that ILUTE overestimates households and it seems that the errors are propagated slightly over the last three years. On the other
hand, ILUTE tends to produce families with a high accuracy over the twenty-year simulation, but a slightly overproduction of families has been occurred in the year 2006.

Figure 4 plots the simulation results of birth, death, out-migration and the total population from ILUTE vs. the corresponding historical data over time. The results show ILUTE simulates the birth and death rate with a high accuracy (with a slight under prediction of death), but out-migration rates start off slightly too high. As ILUTE simulates the
evolution of urban system incrementally over time in a yearly time steps, the out-migration results are corrected, this may be due to the population levels increasing faster than they should have [19].

**Stochastic Analysis of ILUTE Model System**

In micro-simulation modeling, the behaviors of agents (as intelligent objects) are modeled through a variety of behavioral modeling techniques. ILUTE is a multi-agent micro-simulation model system, in which a variety of agents such as persons, families, households, firms act and interact in complex ways. In ILUTE, a behavior is implemented by objects called "decision-making units" [10]. In order to project and predict real-world interactions between agents, a variety of behavioral models and events were designed in ILUTE to simulate the random behaviours of different agents (known as intelligent objects) through a set of random seed starting points. Specifically, the Mersenne Twister pseudo-random number generator is used [20]. However, utilizing such random numbers produces uncertainties to the outputs of ILUTE model system, in which each independent simulation run results a set of different outputs. Because ILUTE is an incremental model system, it is very important to examine the behavior of each individual component of ILUTE in a series of different independent runs and to investigate the distribution of results and its association errors. This examination shows how errors propagate incrementally over the simulation of urban system evolution and also whether different components of the system interact correctly with each other. This is the uncertainty of the model system in which the error term for the natural phenomena (most components of ILUTE simulate natural phenomena) is expected to have a normal distribution. Otherwise, the distribution of simulation runs is skewed to the right or left from the normality nature, therefore, there is a chance that uncertainties or errors in the model system propagate over simulation years. In this regards, the uncertainty associated with predicting models due to the random (Monte Carlo) simulation nature of the ILUTE model system is investigated in this paper.

Studying the variability in ILUTE outputs due to the random effects of behavioral models is essential to test its predictive performance. To do that, 1,000 independent simulation runs over a twenty-year (1986-2006) historical simulation period where undertaken for a sample of 150K households (around 10% of full population). Each simulation run starts with over 686K agents (including 423K persons, 113K families, 150K households), and the overall number of agents grow past 1.1 million.

![Figure 5: Distribution of simulated number of births.](chart)

![Figure 6: Normal distribution curves of ILUTE Multi-run over simulation years for (a) births (b) deaths (c) families (d) households.](chart)
normal frequency distributions for the predicted number of births, year and then overlaid for all simulation years. Figure 6 plots fitted curves are fitted to each of simulated demographic outputs for each are resulted from 1000 runs), the normal frequency distribution of births, deaths, families, and households for all simulated years (1986-2006) indicates their normal behaviour across the distributions of births, deaths, families, and households generated each year by the ILUTE demographic updating module. Each line in the graphs represents a normal frequency distribution for a specific simulation year. As can be seen from these plots, births and deaths each have generally similar probability distributions (variances) across the years, while the variances in the number of families and households increase over time following relatively the same pattern. Although the variance grows over time, in numerical terms, the run-to-run variability is small which is concluded that the overall confidence in simulating number of births, deaths, families, and households is high. To explore the uncertainties in the demographic outputs in more detail, the spread in predicted values and coefficient of variance (COV) for each simulation year for each variable are examined in the next two sections.

Testing distribution of demographic outputs

In ILUTE's simulation engine, demographic outputs are simulated in yearly time-steps. The distributions of each demographic output under examination for all simulation years are plotted across simulation runs. Figure 5 plots the histograms of the simulated number of births for the years of 1987, 2003, and 2006 and presents their associated means and standard deviations as an example. The results illustrate that the overall confidence in simulating number of births, deaths, families, and households generated each year by the ILUTE demographic updating module. Each line in the graphs represents a normal frequency distribution for a specific simulation year. As can be seen from these plots, births and deaths each have generally similar probability distributions (variances) across the years, while the variances in the number of families and households increase over time following relatively the same pattern. Although the variance grows over time, in numerical terms, the run-to-run variability is small which is concluded that the overall confidence in simulating number of births, deaths, families, and households is high. To explore the uncertainties in the demographic outputs in more detail, the spread in predicted values and coefficient of variance (COV) for each simulation year for each variable are examined in the next two sections.

Births and deaths

Figure 7 plots the spread in number of births and deaths by year across the simulation runs (plots a and c) and the corresponding COV's for these two outputs (plots b and d). Different colors in Figures 7a and 7c represent different runs of ILUTE. The red line in each graph is the average trend of all runs by year for each output. The result shows the number of deaths increases monotonically over time, which is to be expected since these basically track the overall growth in GTHA population during the simulation time period. The COV, however, declines over time, indicating that the increasing rate of spread in predicted values is not higher than the overall mean rate. They are also small in magnitude.

Births display a more complex pattern Figure 7a. The long-term average birth rate appears to be more or less constant, but the year-to-year changes in average birth rates follow a somewhat sinusoidal pattern. This presumably reflects the trend in exogenous birth rates that are inputs to the model, as well as possibly trends in the relative number of females of prime child-bearing ages, the number of marriages in the modeled system, etc. The birth COV displays approximately the inverse pattern, rising when the birth rate falls and vice versa, overall displaying a relatively constant (and small) variance, as was qualitatively observed in Figure 7a. In both the births and deaths, the COVs are constantly small in value indicating the stability of ILUTE in predicting the two variables.
In order to test the normality of the simulation runs for the birth and death components of the demographic updating module, the Shapiro-Wilk test is utilized at 0.05 confidence level. The Shapiro-Wilk test utilizes the null hypothesis principle to check whether a sample come from a normally distributed population [21] is calculated as below:

$$SW = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Where $SW$ is the Shapiro-Wilk value, $x_i$ is the simulation run for the variable of interest, $\overline{x}$ is the mean of all simulation run, and $a_i$ is the constant which depends on expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution and the covariance matrix of those order statistics.

The results of the test for the simulation results of birth and death are summarized for all years in Table 1. For an alpha level of 0.05, the simulation data set of birth and death with the p-value of greater than 0.05 (except for year 1999 of Death) rejects the alternative hypothesis, and concludes that the data comes from a normal distribution. It can be also concluded that the error term also has a normal distribution and there is a little chance that errors propagate over the simulation years (Figure 4).

Examining the normal Q-Q plots for each simulation year confirms the observation and the statistical test. As an example, normal Q-Q plots and detrended normal Q-Q plots for the simulation years of 1991, 2001, and 2006 are graphed comparing the observed birth and death simulated in ILUTE on the X-axis with the expected values (assuming a normal distribution) on the Y-axis Figures 8 and 9. If the distribution of ILUTE-generated birth and death for a simulation year is distributed exactly like a normal distribution, the points should fall on a straight line. The Q-Q plots indicate that the generated birth and death distribution is very close to the normal distribution at each simulation year but slightly deviates at the tails of the normal curve for a few number of ILUTE runs. The detrended Q-Q plots illustrate how much the observed simulated values deviate from the expected values of normal distribution. It also demonstrates that the concentration of birth and death results for 1000 runs is very close to the mean implying the stability of ILUTE in simulating dwellings.

### Families and households

Figures 10a and 10c plot the spread in predicted number of families and households by year respectively, with the red line indicating the mean value for each year. As can be seen from the two figures, while some growth over the twenty-year simulation occurs in the spread in the range of number of families and households generated, the spreads for the two variables remain small and tight around their means. Predicted families and households follow relatively the same pattern with monotonically increasing rate over time.

Figures 10b and 10d plot the coefficient of variation (CV) in predicted families and households for each year. As it is seen the CV for the two variables follows a similar pattern increasing for the first couple of simulation years and then tending to be more stable over time. In particular, the values of CV across the simulation years are very small, which indicates that the deviation in predicted families and households from their mean across simulation runs of ILUTE is negligible. It thus appears that ILUTE is very stable in simulating families and households across independent runs with different random seed starting points.

The normality test of Shapiro-Wilk is performed for the 1000 simulation runs of families and households. The results of the test are summarized for all years in Table 2. For an alpha level of 0.05, the simulation data set of families with the p-value of greater than 0.05 rejects the alternative hypothesis, and concludes that the data comes from a normal distribution. However, the results for the households are different in the way that the simulation outputs have normal distribution for the years 1987 to 2003, but not for the last three years. Comparing the household historical data with the ILUTE results also demonstrates that the errors tend to be increased for the last three years of simulation. This indicates that the error term of the model does not have a normal distribution and there is more chance that the errors propagate over simulation years as expected from the observation of the historical data and ILUTE results (Figure 3). This condition may also affect the accuracy of other ILUTE components which interact directly with the components of the household simulation module.

| Year | Statistic | df | Sig. | Statistic | Births | df | Sig. |
|------|-----------|----|------|-----------|--------|----|------|
| 1990 | 0.998     | 1000 | 0.319 | 0.998     | 1000   | 0.500 |      |
| 1991 | 0.999     | 1000 | 0.742 | 0.999     | 1000   | 0.666 |      |
| 1992 | 0.999     | 1000 | 0.750 | 0.999     | 1000   | 0.669 |      |
| 1993 | 0.997     | 1000 | 0.108 | 0.998     | 1000   | 0.423 |      |
| 1994 | 0.997     | 1000 | 0.031 | 0.999     | 1000   | 0.689 |      |
| 1995 | 0.997     | 1000 | 0.104 | 0.997     | 1000   | 0.115 |      |
| 1996 | 0.996     | 1000 | 0.332 | 0.998     | 1000   | 0.258 |      |
| 1997 | 0.998     | 1000 | 0.155 | 0.998     | 1000   | 0.483 |      |
| 1998 | 0.999     | 1000 | 0.706 | 0.998     | 1000   | 0.357 |      |
| 1999 | 0.996     | 1000 | 0.009 | 0.997     | 1000   | 0.051 |      |
| 2000 | 0.998     | 1000 | 0.481 | 0.998     | 1000   | 0.271 |      |
| 2001 | 0.999     | 1000 | 0.788 | 0.999     | 1000   | 0.714 |      |
| 2002 | 0.999     | 1000 | 0.963 | 0.999     | 1000   | 0.733 |      |
| 2003 | 0.998     | 1000 | 0.256 | 0.999     | 1000   | 0.583 |      |
| 2004 | 0.999     | 1000 | 0.689 | 0.998     | 1000   | 0.495 |      |
| 2005 | 0.997     | 1000 | 0.107 | 0.999     | 1000   | 0.947 |      |
| 2006 | 0.999     | 1000 | 0.891 | 0.997     | 1000   | 0.079 |      |

Table 1: Test of normality (Shapiro-Wilk test) for births and deaths simulations.
Figure 8: Normal Q-Q plots of number of births.

Figure 9: Normal Q-Q plots of number of deaths.
In order to compare the stochastic behavior of all tested demographic updating variables, their associated COVs are overlaid in a graph shown in Figure 11. The result shows simulated births and deaths are less stable than other attributes. It is seen families and households follow a similar behavior across simulation years and their associated values of COV are close to zero.

### Conclusion

Behavioural modeling techniques are the key components in a microsimulation ILUTM system, through which the behaviours of different agents in the urban system are modeled. Behavioural models impose different level of uncertainties to the model system due to their random simulation nature. ILUTE is an agent-based microsimulation model system using various behavioral models developed to simulate the evolution of an urban region’s transportation and land use systems over time. Its utility as a forecasting and policy analysis tool depends on it being able to generate robust forecasts within reasonable confidence limits. This paper has presented the results of an investigation into the stochastic properties of a number of key ILUTE demographic outputs.

To test the stochastic properties of ILUTE due to the random effects within the model, the variability of demographic outputs was examined through 1000 independent simulation runs of the same base scenario over a twenty-year (1986-2006) historical simulation period for the Greater Toronto-Hamilton Area (GTHA). Several variables including births, deaths, families, and households were tested in this study. For each variable for each year of the simulation, average values, range of values, standard deviations and coefficient of variation were computed across the 1000 runs. The hypothesis that the simulated distributions were normally distributed was also tested. In general, it was found that simulated distributions were generally normally, or very close to normally, distributed for all variables. Findings from normal curves of the simulated outputs indicates that the run-to-run variability is small, which is concluded that the overall confidence in simulating number of births, deaths, families, and households is high. The COVs for births, deaths, families, and households are all stable or declining over the time and are also universally small. In addition, no evidence was found that errors were propagated and magnified through the chain of models within the ILUTE model system.
In general, these results are very encouraging for the use of large-scale stochastic LUTMs for practical policy analysis. They are, however, still partial in that other system components (notably the labour market and transportation demand models) still need to be added into the tests. The results also need to be extended to the 100% population case. Based on the current results, however, there is no reasons to expect that a (properly validated) more complete model system or larger sample size would result in significantly more chaotic or unreliable results.

Acknowledgment

This project is funded by Sothern Ontario Smart Computing Innovation Platform (SOSCIP) and IBM Canada.

References

1. Hunt JD, Abraham JE (2005) Design and implementation of PECAS: A generalized system for allocating economic production, exchange and consumption quantities. Integrated land-use and transportation models. Behavioural Foundations 1: 253-273.
2. Abraham JE, Eng P (1998) A review of the MEPLAN modelling framework from a perspective of urban economics. University of Calgary. Department of Civil Engineering.
3. De-la Barra T (2005) Integrated land use and transport modelling. Decision chains and hierarchies. Paperback Edition, Cambridge University Press, NY, U.S.
4. Putman SH (1995) EMPAL and DRAM location and land use models: A technical overview. Land Use Modeling Conference, Dallas, TX. 19-21.
5. Miller EJ, Salvini PA (2001) The integrated land use, transportation, environment (ILUTE) microsimulation modelling system: Description and current status. Travel Behaviour Research: The Leading Edge. 7:11-24.
6. Martinez F (1996) MUSSA: Land use model for Santiago city. Transportation Research Record: Journal of the Transportation Research Board. 1552:126-34.
7. Simmonds DC (1999) The design of the DELTA land-use modelling package. Environment and Planning B: Planning and Design. 26: 665-684.
8. Waddell P (2002) UrbanSim: Modeling urban development for land use, transportation, and environmental planning. Journal of the American planning association. 68(3): 297-314.
9. Miller EJ, Hunt JD, Abraham JE, Salvini PA (2004) Microsimulating urban systems. Computers, environment and urban systems. 28(1): 9-44.
10. Salvini P, Miller EJ (2005) ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. Networks and Spatial Economics. 5(2): 217-234.
11. Miller EJ, Farooq B, Chingcuanco F, Wang D (2010) Microsimulating urban spatial dynamics: historical validation tests using the ILUTE model system. In: Workshop on Urban Dynamics, Marbella, Chile.
12. Miller EJ, Salvini PA (1998) The integrated land use, transportation, environment (ILUTE) modelling system: A framework. In: Proceedings 77th Annual Meeting of the Transportation Research Board, Washington, DC.
13. Miller EJ, Salvini PA (2001) The integrated land use, transportation, environment (ILUTE) microsimulation modelling system: Description and current status. Travel Behaviour Research: The Leading Edge. p. 711-724.
14. Miller EJ, Farooq B, Habib MA (2008) Development of an operational integrated urban model system, Volume IV: Housing market model, Toronto: Urban Transportation Research & Advancement Centre, University of Toronto.
15. Miller EJ (2008) Integrated urban models: Progress to date: Future prospects. Invited resource paper, Kitamura, R. (ed.), Proceedings of the 11th International Association for Travel Behaviour Research Conference, Kyoto.
16. Miller EJ, Chingcuanco F, Farooq B, Habib KM, Habib MA (2008) Development of an operational integrated urban model system. Urban Transportation Research & Advancement Centre, University of Toronto.
17. Miller EJ (2008) Development of an operational integrated urban model system, Volume I: Project Final Report, Toronto: Urban Transportation Research & Advancement Centre, University of Toronto.
18. Chingcuanco F, Miller EJ (2013) Demographic microsimulation model for integrated land use, transportation, and environment model system. In: Transportation Research Board 92nd Annual Meeting.
19. Chingcuanco F, Miller EJ (2016) The ILUTE demographic microsimulation model for the Greater Toronto-Hamilton Area: Current operational status and historical validation, Forthcoming, J.C. Thill (ed), Gecomputation, Springer.
20. Matsumoto M, Nishimura T (1998) Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation (TOMACS). 8:3-30.
21. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika. 52: 591–611.