DEVELOPMENT AND EVALUATION AN ALGORITHM TO PRODUCE THE POPULATION IN REGIONAL LEVEL AND DISSEMINATION AREA LEVEL

Mohammad Hesam Hafezi, Department of Civil and Resource
Engineering, Dalhousie University, Halifax, NS, Canada
Muhammad Ahsanul Habib, School of Planning, and Department of
Civil and Resource Engineering, Dalhousie University, Canada

Introduction

In the past decade, transportation researchers' interest in developing the agent-based microsimulation modeling has increased in response to the growing importance of complex policy measures, such as travel demand management and road pricing. Advanced agent-based micro-simulation models utilize disaggregate data for a higher level of accuracy and reliability of the models. TASHA [1], AMOS [2, 3], FAMOS [4], ALBATROSS [5] and CEMDAP [6] are examples of agent-based micro-simulation models that simulate the behaviour of individuals and households rather than aggregate accounting. As such, the model produced is able to address the impacts of flexible work hours, pricing-based strategies, greenhouse gas (GHG) emissions, interfaces between micro-scale land use changes and travel activity [7]. However, one of the major challenges of disaggregate modeling is the requirement for a large amount of individual-level data [8]. UrbanSim [9], ILUTE [10], TRANSIMS [11], TRANUS [12], TRESIS [13], MUSSA [14], IRPUD [15] and Metrosim [16] are some examples of micro-simulation platforms that apply the agent-based model and use individual-data as input for model estimation. The availability of micro-data is essential for the operation of micro-simulation models. In particular, it is necessary to input details of the individual or household characteristics, as well as home and work locations, for the entire population of the study area to achieve the most fine-grained model results. This information is typically collected in a population census, however, the disaggregate data is not accessible to the public due to privacy concerns [8, 17]. Population synthesis aims to produce virtual individuals by expanding the public disaggregate data to mirror known aggregate sample data with the same demographics as the real population. The population synthesis methodology described in this paper is comparable to work done by Ma and Srinivasan [18] in that it produces a list of households to match several multilevel controls without the need for a joint multi-way distribution. However, in this paper we presented new functional form for the fitness calculation in the population synthesis generator algorithm that resulted to improve the efficiency of the algorithm in both of computational time and distribution of seed data presented in the final synthetic population. We call this the multi-dimensional population synthesis (MPS) approach.

The synthetic algorithm is examined by three models: first, using only the household level control tables (HL model); second, using both individual and household level control tables (HPL model); and third, with weights added to both individual and household level control tables (WHPL model). Both the HL and the HPL models were examined in the work done by Ma and Srinivasan (18), where, the WHPL model is a new weighted model, tested in this paper. In the new WHPL model, the distribution of household attributes in seed data based on control table categories, is calculated. Additionally, through the comparison of the results of the synthesized population between the three different models, the WHPL model’s repetition of larger households selections has declined. Furthermore, it can be noted that the distribution of selected households is more uniform and homogeneous. The performance of the algorithm for each model is validated using error-percentages.

Literature review

Due to the lack of accessibility and completeness of the micro-data, population synthesis is an essential step for developing a disaggregate travel demand forecasting model. Population synthesis aims to produce virtual individuals by expanding the disaggregate public data to mirror known aggregate sample data with the same demographics as the real population. Table 1 is shows a summary of some population synthesizer methods by reviewing their pros and cons. The key differences between Iterative Proportional Fitting (IPF), Combinatorial Optimization (CO), Iterative Proportional Updating (IPU) and Fitness Based Synthesis (FBS) is their varied ability to simultaneously control for both the household and individual attributes. Traditionally, the IPF method is the only population synthesis procedure that is incapable of controlling for individual attributes in the process. With the exception of the FBS method, other methods, namely CO and IPU, use the IPF procedure as a part of their process. The original IPF methods were developed by Beckman et al. [11] to generate synthetic baseline populations as input data for the TRANSIMS model using sample and census aggregate data in the United States. The traditional IPF method contains two steps; fitting unadjusted cell data (called seed data), and generating the synthesized households. The seed data must be adjusted to a known margin for both the rows and columns of the table (called as control or marginal tables). However, the traditional IPF method has some limitations; high-dimensional (memory) problems, control for individual attributes, zero cell problem and rounding cell values of the joint distribution. These limitations cause a reduction in the accuracy and validity of the synthesized population.

Of the examined previous literature each study introduced a new modification in addressing a limitation to the original IPF model. Pritchard and Miller [19-21] configured the synthesis of large-scale attributes per agent by using a list-based method in which household and individual level attributes were fitted simultaneously. The use of sparse matrices increased the capability of controls and categories resulting in decreasing memory use and computational time. Guo and Bhat [22] studied the ‘zero cell’ issue and, along with Arentze et al. [5], controlling individual level attributes. They defined a certain tolerance for individual attribute levels in the initial steps involving the seed data from which household data is ensuring that individual level restrictions were not disrupted. This method also allows manipulation of data from different data set sources, which solves another limitation of the traditional IPF method.

Table 1 Pros and cons of four population synthesizer methods

Population synthesizer method	**Pros**	**Cons**
Traditional Iterative Proportional Fitting (IPF)	- Combination of probability table - Zone-by-zone versus multi-zone - The format of the seed is remembered	- High-dimensional problems - Zero cell problems - Rounding of the cell values in joint distribution - Association structure for household and individuals - The method is unable to control for individual attributes. - The method ignores differences in household type among households grouped within a cell
Combinatorial Optimization (CO)	- Variance value is small - Continuous/discrete characteristics - Proficient memory - Changing the number of layers is insignificant	- PUMA level data is required to be applied to the procedure due to correlation structure within the zonal population - Search planetary is very large
Iterative Proportional Updating (IPU)	- Controls for both household and individual attributes simultaneously - Matches seed level data and marginal controls characteristics on several analysis levels - Less running time - Extra corrections to match individual and household level data are not required	- Discrepancies in matching of seed level data to marginal data - Rounding of values
Fitness Based Synthesis (FBS)	- Produces a list of households to match several multilevel controls - Determining a joint multi-way distribution is not required	- Creates the same fitness value iteration in the case of few-control-tables

Another population synthesizer method that addresses some of the limitations of the IPF method is the combinatorial optimization (CO) method. Similar to the IPF method, CO also requires information on population characteristics both at the sample (seed data) and marginal (control table) levels. The CO method uses the integer reweighting technique. There are two initial weights of 0 and 1 in the method, which are given to the sample. A weight of 0 is assigned to all households at the beginning of the procedure, then, after selecting each household, a weight of 1 is assigned. An exhaustive review of the combinatorial optimization (CO) method can be found in the Voas and Williamson [23, 24]; and, Huang and Williamson [25]. Iterative Population Updating (IPU) is another population synthesis method that can simultaneously generate synthetic populations with both household and individual attribute level data. Xin et al. [7] introduced a new algorithm that can synthesize populations at a high performance level by matching household and individual level distributions. The IPU method consists of three main steps. The first step constrains household and individual level attributes by selecting 5% of household and individual level attributes from PUMS as seed data to create marginal tables from the census file. Second, the weights of the household and individual level joint distributions are estimated. These weights are assigned so that both household and individual level distributions can be closely matched. Finally, household data is drawn from the procedure in the previous step, and subsequently the synthetic population for the region is generated [7, 17].

In the all of the three methods of population synthesis, a joint multi-way distribution table is created in the beginning of the procedure. Srinivasan et al. (2008) introduces a method that can generate synthetic populations with the capability of matching several multilevel controls without necessitating a joint multi-way distribution. The Fitness Based Synthesis (FBS) method is based on the calculation of fitness value for each sample household. This value addresses the match for both household and individual level distributions. In each iteration, the highest fitness value will be chosen and added to the synthetic population result. Additionally, termination criterion for iterative methods is determined when there is an absence of positive fitness values. The FBS technique, in comparison, performs well in terms of simultaneously controlling both household and individual attributes of interest as well as run time [18, 26, 27] Through a trade-off between different population synthesizer methods, it is highlighted that still there are some issues such as high-dimensional problems (association structure for household and individuals), discrepancies in matching of seed level data to marginal data, creates the same fitness value iteration in the case of few-control-tables and long running time (search planetary is very large) which need to be considered in synthesize population in the disaggregate level. In this paper we produced a multi-dimensional population synthesis (MPS) approach that solved and optimized the above mentioned issues. In this paper, a population is synthesized for individuals and households in both RL and DA levels using the multi-dimensional population synthesis (MPS) method. Details of the MPS approach are explained further in the methodology section in this paper.

Data used in the generating a synthetic population

Two main data sources are used in this study. The first source is the 2006 Canadian Census Hierarchical Public Use Microdata File (PUMF). The second source used is the 2006 Canadian Census (Statistics Canada) to synthesis population for base year. Initially, this study aimed to generate synthetic population for the Halifax region only. The 2006 PUMF hierarchical file included large Census Metropolitan Areas (CMAs), namely, Montréal, Toronto, Calgary, Edmonton, and Vancouver; and did not include Halifax. The Halifax CMA is only available in the 2006 PUMF individual file. The individual PUMF does not give adequate information concerning household-level data. It focused more closely on the attributes of individuals within the household. Therefore, this study used the Atlantic Canada (New Brunswick, Newfoundland and Labrador, Nova Scotia and Prince Edward Island) region (RL) for this empirical application. Control tables for synthesizing population in regional level (RL) model are extracted from four regions for entire of the Atlantic Canada. Synthesized population for Atlantic Canada has been used as a seed data to synthesis population for Halifax region (DA level). Control tables for synthesizing population in the dissemination area (DA) level are extracted from Halifax Census.

Attributes Considered for Synthesizing Population

Individual and household attributes are the two groups of attributes considered for synthesizing population used in this study. Households and individual attributes have six and five variables, respectively. Household variables include: household size, type of household, tenure, household income, structural type of dwelling and labour force activity. Household size is categorized for 1, 2, 3, 4, 5 and 6+ residents. The type of household is categorized into four: married, single parent, parents with children and single occupant. Tenure has two categories: owned or rented properties. Household income is divided as follows: under $ 19,999; between $ 20,000 and $ 24,999; between $ 25,000 and $ 29,999; between $ 30,000 and $ 34,999; between $ 35,000 and $ 39,999; between $ 40,000 and $ 44,999; between $ 45,000 and $ 49,999; between $ 50,000 and $ 59,999; and, $ 60,000+. Dwelling type has eight divisions: single-detached house, semi-detached or double house, row house, apartment/flat in a duplex, apartment in a building that has five or more storeys, apartment in a building that has fewer than five storeys, other single-attached house and mobile home, and, other movable dwelling. Labour force activity has two categories, employed and un-employed.

Variables associated with specific individuals included: age, education level, ethnicity, legal marital status and gender. Age was defined in the following 11 categories: ≤19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-64, 65-74 and 75+. Education level is defined by the subsequent measures: high school graduation diploma or equivalent certificate; apprenticeship or trades certificate or diploma; college, CEGEP or other non-university certificate or diploma; university certificate or diploma below bachelor level; bachelor's degree; university certificate or diploma above bachelor level; degree in medicine, dentistry, veterinary medicine or optometry; master's degree; and, earned doctorate degree. Ethnicity has five categories: British Isles origins, French origins, Aboriginal origins, Canadian, European origins, Asian origins and other origins. Legal marital status is defined by divorced; legally married; single and widowed. Gender may be identified by either female or male.

Methodology

The MPS procedure involves selecting a set of households in each iteration from the seed data. This ensures that the count table closely replicates the control table. The number of count tables is equivalent to the number of control tables in the method. Initially, the count tables for all household attributes are given an initial value of zero. Then in each iteration, households are added or removed to the count table singly according to the corresponding fitness value. The population is then synthesized in an iterative fashion. This iterative fashion is continued until the count table replicates the control table as closely as possible. However, the values may not be an exact match for all count tables. Therefore, the iterative process is terminated when there is an absence of positive fitness values. Ma and Srinivasan [18] defined two ways to calculate the fitness value. This is required because of the need to determine whether to add or removing the selected household from the method. The first fitness value (type 1) corresponds with the error if the selected household is added to the count table in the current iteration. The second fitness value (type 2) corresponds with the error if the selected household is removed to the count table in the current iteration. The average distribution of the control table in the seed data is used as a new weight to calculate the fitness values. These measures ensure that the performance and accuracy of the synthetic population produced has been improved. Type one and type two fitness values are given by the following:

$$F_1^{xy} = \sum_{q=1}^{Q} \sum_{w=1}^{W_q} n_q * [\left(G_{qw}^{y-1}\right)^2 - \left(G_{qw}^{y-1} - MH_{qw}^{x}\right)^2] \qquad (1)$$

$$F_2^{xy} = \sum_{q=1}^{Q} \sum_{w=1}^{W_q} n_q * [\left(G_{qw}^{y-1}\right)^2 - \left(G_{qw}^{y-1} + MH_{qw}^{x}\right)^2] \qquad (2)$$

$$G_{qw}^{y-1} = H_{qw} - AH_{qw}^{y-1} \qquad (3)$$

$$m^q = find\left(F_1^{xy} > 0\right) \qquad (4)$$

$if\ isempty\ (m^q)$

$m^q = find(F_2^{xy} > 0)$

$if\ isempty\ (m^q)$ $\qquad \%\ Terminate\ the\ algorithm$

$$rm_y^1 = find(m1(:,:,y) >= 0) \qquad (5)$$

$\ldots$

$rm_y^{qw} = find(m1(:,:,y) >= 0)$

$$mm_y = intersect(intersect(\ldots(intersect(rm_y^1, rm_y^2, rm_y^3, \cdots, rm_y^{qw})) \quad (6)$$

Where:

F_1^{xy} is fitness value type 1. F_2^{xy} is fitness value type 2. x is selected household. y is iteration number. q is index representing for the both control and count tables. Q is the total number of both control/count tables. w is index representing the different cells in the count table. W is index representing the different cells in the control table. H_{qw} is represents the value of cell w in control table q. n_q is represents the average distribution of control table q in the seed data. AH_{qw}^{y-1} is represents the value of cell w in the count table q. G_{qw}^{y-1} is difference value between control and count tables for cell w in control table q. MH_{qw}^{x} is the contribution of the x^{th} household in the seed data to the w^{th} cell in control table q. $m^q$is selected household type 1 or 2 according to the fitness value. rm_y^{qw} is selected household for the cell w in the count table q in the iteration y. mm_y is a set of selected households for adding into the count tables and synthesized population list.

$G_{qw}^{y-1} - MH_{qw}^{x}$ and $G_{qw}^{y-1} + MH_{qw}^{x}$ are the number of households which need to added or removed, respectively for the cell w in the count table q in the iteration y. The MPS procedure continues by calculating the fitness value for each household within the seed data for each iteration (according to the formula 1 and 2). Positive type 1 or 2 fitness values are considered as the nominees for addition or removal from the list of updated synthetic population (according to the formula 4). Ma and Srinivasan [18] presented a random selection of one of the household among of all candidate sets of household. Here, a new selection approach than can cause increase computational time and efficiency of the algorithm is presented. Instead of one selection per each iteration, we will select all of the positive fitness value for each of the variables with some conditions. First, all of the positive fitness for the cell w in the count table q in the iteration y is selected (according to formula 5). Then, the same household that is repeated in all of selected sets, will consider as a set of households for adding into the count tables and synthesized population list (according to formula 6). Note that according to increase the number of iterations, the number of selected household in the formula 6 will be decreased due to the defined condition in the formula 3. This practice is argued by discussing that as long as algorithm is trying to expend the seed data according to the defined control tables and as well as sequence of household selection is not important in the procedure, therefore, adding all of the potential repetition number of households at once can be reasonable and logical. Following the selection process, all the values of all dimensions in the count tables are updated. The next iteration repeats the previous steps by calculating fitness value, selecting households and updating tables in the algorithm. The termination criteria occurs when there are no positive fitness values left in the procedure (according to formula 4).

Discussion of results

The synthetic algorithm is examined by three models: first, only using the household level control tables (HL model); second, using both individual and household level control tables (HPL model); and third, with weights added to both individual and household level control tables (WHPL model). n_q value is considered as 1 for the HL and HPL model. In the WHPL model the distribution of household attributes in the seed data based on the control table categories is calculated. Then, n_q is calculated by the average of the all household attributes in one control table category. Error-percentage measures are calculated as follows:

$$EP_q = \frac{\sum_{q=1}^{Q} \left| H_{qw} - AH_{qw} \right|}{\sum_{q=1}^{Q} H_{qw}}$$

EP_q is represents the error-percentage for control table q. H_{qw} is represents the value of cell w in control table q. AH_{qw} is represents the value of cell w in the count table q. The error percentages for regional level (RL) model are calculated in Table 4. For each of the control tables, in each of the three models, error percentages are summarized in Table 4. Additionally, error percentages were calculated for the dissemination area (DA) model with just WHPL sub model. Table 5 is representative of results of error percentages for 10 DA models.

Table 4 Error percentages of three models (regional level)

Explanatory variables	**HL model**	**HPL model**	**WHPL model**
Individual level			
Age	6.4%	0.9%	0.3%
Education level	8.3%	1.4%	1.1%
Ethnicity	18.9%	0.8%	0.1%
Legal marital status	7.9%	0.2%	0.1%
Gender	8.8%	0.1%	0.1%
Household level			
Household size	0.9%	1.1%	1.1%
Type of household	0.5%	0.8%	0.8%
Tenure	0.9%	1.9%	1.4%
Household income	0.4%	0.0%	0.9%
Structural type of dwelling	0.9%	1.6%	1.3%
Labour force activity	2.4%	1.2%	1.1%

Table 5 Error percentages of synthesized population (DA level)

Explanatory variables	**DA ID**									
	12090751	12090672	12090873	12090871	12090869	12090660	12090879	12090876	12090875	12090874
Individual level										
Age	0.2%	0.3%	0.3%	0.4%	0.4%	0.2%	0.4%	0.2%	0.3%	0.2%
Legal marital status	0.2%	0.3%	0.3%	0.4%	0.4%	0.2%	0.4%	0.2%	0.3%	0.2%
Gender	0.1%	0.2%	0.4%	0.5%	0.3%	0.0%	0.0%	0.2%	0.4%	0.3%
Household level										
Household size	0.2%	0.1%	0.3%	0.3%	0.3%	0.2%	0.1%	0.2%	0.4%	0.2%
Type of household	0.1%	0.3%	0.2%	0.2%	0.3%	0.1%	0.3%	0.1%	0.1%	0.2%
Tenure	0.2%	0.3%	0.2%	0.2%	0.2%	0.3%	0.1%	0.1%	0.2%	0.2%
Structural type of dwelling	0.2%	0.3%	0.2%	0.2%	0.2%	0.3%	0.1%	0.1%	0.2%	0.2%
Labour force activity	0.0%	0.1%	0.2%	0.3%	0.3%	0.1%	0.4%	0.1%	0.2%	0.2%

There are two main differences in the error percentages of the three models in Table 4. Error percentages for individual level control tables in the HPL and WHPL models are significantly smaller than in the HL model. This difference demonstrates that using both individual and household level control tables results in improvements to the accuracy of synthetic population. This is due to the increased precision of the fitness value calculation by involving additional control tables in each iteration. Secondly, the values of the error percentages for the household level control tables in the HPL and WHPL models are slightly larger than in those of HL model. This results is acclaimed in the table 5 for the dissemination area level models. Likewise, this is due to an increase in the total number of control tables in the method. When the number of control tables is increased the ability to replicate each count table during the household selection in the procedure is decreased.

Comparison between numbers of households' selection in the RL model is shown in Figure 5. To better illustrate of the difference in the two households' selection modes, Figure 5 is shows the only first 100 iterations in the procedure. According to figure 5, comparison between two models is showed that in the first 100 iteration of the procedure approximately 582,839 individual and 380,658 households' has been selected and added to the final synthesized population list while on the other side just 100 individuals and 39 households' are selected during first 100 iterations. Changes percentage is over than 99.98% that it has directly effect on the improvement of computational time.

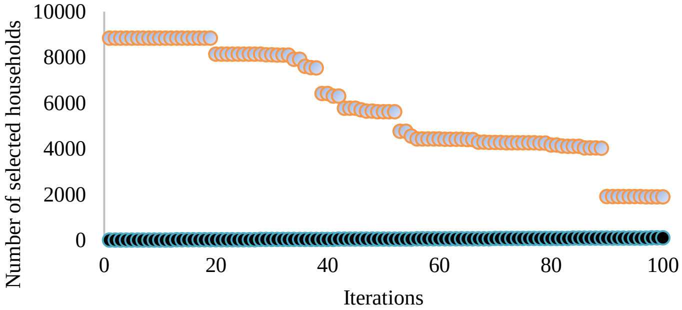

Figure 6. Comparison between numbers of households' selection

Conclusions

Agent-based micro-simulation models attempt to forecast the behaviour of individuals and households through simulating the behaviour of a disaggregate sample of individuals and households. Population synthesis is the primary step in micro-simulation modeling. In this paper, a population is synthesized by individuals and by household at both the regional level (RL) and the dissemination area (DA) levels using the multi-dimensional population synthesis (MPS). The algorithm used is implemented in MATLAB using the sparse matrix technique. The sparse matrix technique performs well with a reasonable runtime for 11 attributes (at both dimensions of household and individual levels) and represented realistic outcomes for the study area. Atlantic Canada as a RL case study include New Brunswick, Newfoundland and Labrador, Nova Scotia and Prince Edward Island has a population of over 2,284,779. The Halifax Regional Municipality, in Nova Scotia, as a DA level case study with a population of over 372,680. These two study areas are used to synthesis their population through MPS approach. Furthermore, three different models, the HL model (only using the household level control tables), the HPL model (using both individual and household level control tables) and the WHPL model (weighted individual and household level control tables) are tested to asses the performance of the algorithm. As validated by the error percentages, the MPS can efficiently obtain a satisfactory result using both individual and household level control tables. The results suggest that the distribution of selected households in the synthesized population through the WHPL model is more homogeneous in comparison with the HPL model. Future works are including examining the performance of the MPS approach with back-casting test and evaluate the performance of the algorithm in the case of uncontrolled attributes. This is an important step forward towards developing an activity-based microsimulation for travel demand forecasting modeling in Halifax, Nova Scotia, Canada.

Acknowledgments

This research is funded by Natural Science and Engineering Research Council (NSERC), the Discovery Grant and the Nova Scotia Department of Energy. The authors would like to thank Cara Wittich, Bachelor of Community Design, for proof reading this article.

References

1. Deng, Y., E.J. Miller, and J.A. Vaughan. *Application of Travel Activity Scheduler for Household Agents in a Chinese City*. in *Transportation Research Board 93rd Annual Meeting*. 2014.
2. Bowman, J.L. and M.E. Ben-Akiva, *Activity-based disaggregate travel demand model system with activity schedules.* Transportation Research Part A: Policy and Practice, 2001. **35**(1): p. 1-28.
3. Ben-Akiva, M.E. and J.L. Bowman, *Activity based travel demand model systems*, in *Equilibrium and advanced transportation modelling*. 1998, Springer. p. 27-46.
4. Pendyala, R.M., et al., *Florida activity mobility simulator: overview and preliminary validation results.* Transportation Research Record: Journal of the Transportation Research Board, 2005. **1921**(1): p. 123-130.
5. Arentze, T.A. and H.J. Timmermans, *A learning-based transportation oriented simulation system.* Transportation Research Part B: Methodological, 2004. **38**(7): p. 613-633.
6. Lin, D.-Y., et al., *Integration of activity-based modeling and dynamic traffic assignment.* Transportation Research Record: Journal of the Transportation Research Board, 2008. **2076**(1): p. 52-61.
7. Ye, X., et al. *A methodology to match distributions of both household and person attributes in the generation of synthetic populations*. in *88th Annual Meeting of the Transportation Research Board, Washington, DC*. 2009.
8. Anderson, P., et al. *Associations Generation in Synthetic Population for Transportation Applications. A Graph-Theoretic Solution*. in *Transportation Research Board 93rd Annual Meeting*. 2014.
9. Waddell, P., *Integrated land use and transportation planning and modelling: addressing challenges in research and practice.* Transport Reviews, 2011. **31**(2): p. 209-229.
10. Salvini, P. and E.J. Miller, *ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems.* Networks and Spatial Economics, 2005. **5**(2): p. 217-234.
11. Beckman, R.J., K.A. Baggerly, and M.D. McKay, *Creating synthetic baseline populations.* Transportation Research Part A: Policy and Practice, 1996. **30**(6): p. 415-429.
12. De la Barra, T., *Integrated land use and transport modelling. Decision chains and hierarchies*. 1989.

13. Hensher, D.A., *A systematic assessment of the environmental impacts of transport policy.* Environmental and Resource Economics, 2002. **22**(1-2): p. 185-217.
14. Martínez, F. and P. Donoso. *MUSSA: a land use equilibrium model with location externalities, planning regulations and pricing policies*. in *7th International Conference on Computers in Urban Planning and Urban Management (CUPUM 2001), Hawaii*. 2001.
15. Wegener, M., *The IRPUD model.* Dortmund. Online available at http://www.raumplanung.uni-dortmund.de/irpud/, 1998.
16. Anas, A., *Capitalization of urban travel improvements into residential and commercial real estate: simulations with a unified model of housing, travel mode and shopping choices*.* Journal of Regional Science, 1995. **35**(3): p. 351-376.
17. Lim, P.P. and D. Gargett. *Population Synthesis for Travel Demand Forecasting*. in *Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia*. 2013.
18. Ma, L. and S. Srinivasan, *Synthetic Population Generation with Multilevel Controls: A Fitness-Based Synthesis Approach and Validations.* Computer-Aided Civil and Infrastructure Engineering, 2014.
19. Pritchard, D.R. and E.J. Miller. *Advances in agent population synthesis and application in an integrated land use and transportation model*. in *Transportation Research Board 88th Annual Meeting*. 2009.
20. Pritchard, D.R., *Synthesizing agents and relationships for land use/transportation modelling*. 2008, University of Toronto.
21. Pritchard, D.R. and E.J. Miller, *Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously.* Transportation, 2012. **39**(3): p. 685-704.
22. Guo, J.Y. and C.R. Bhat, *Population synthesis for microsimulating travel behavior.* Transportation Research Record: Journal of the Transportation Research Board, 2007. **2014**(1): p. 92-101.
23. Voas, D. and P. Williamson, *An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata.* International Journal of Population Geography, 2000. **6**(5): p. 349-366.
24. Voas, D. and P. Williamson, *Evaluating goodness-of-fit measures for synthetic microdata.* Geographical and Environmental Modelling, 2001. **5**(2): p. 177-200.
25. Huang, Z. and P. Williamson, *A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata.* Population Microdata Unit, Department of Geography, University of Liverpool, Liverpool L69 3BX. URL http://pcwww.liv.ac.uk/~william/microdata, 2001.
26. Srinivasan, S., L. Ma, and K. Yathindra, *Procedure for forecasting household characteristics for input to travel-demand models*. 2008.
27. Ma, L., *Generating disaggregate population characteristics for input to travel-demand models*. 2011, University of Florida.