Microscopic trip generation: Adding fidelity to trip-based travel demand models

Rolf Moeckel, Leta Huntsinger and Rick Donnelly

Abstract

In four-step travel demand models, average trip generation rates are traditionally applied to static household type definitions. In reality, however, trip generation is more heterogeneous with some households making no trips and other households making more than two dozen trips. Two improvements for trip generation are presented in this paper. First, the household type definition, which traditionally is based on experience or habituality rather than science, is revised to optimally reflect trip generation differences between household types. For this purpose, over 67 Million household type definitions were analyzed econometrically in a Big-Data exercise. Secondly, a microscopic trip generation module has been developed that specifies trip generation individually for every household. This tool allows representing the heterogeneity in trip generation found in reality, and it adds flexibility if additional household attributes are added in the future.

Rolf Moeckel (Corresponding author)
National Center for Smart Growth, University of Maryland
Email: rolf.moeckel@udo.edu

Leta Huntsinger, Parsons Brinckerhoff, Raleigh, North Carolina
Email: Huntsinger@pbworld.com

Rick Donnelly, Parsons Brinckerhoff, Albuquerque, New Mexico
Email: DonnellyR@pbworld.com

Introduction

A travel demand model is a series of mathematical equations used to describe travel and travel choices. In its most basic form, this series of models is broken into a 4-step process: trip generation, trip distribution, mode choice, and trip assignment. Best practice models use locally collected travel survey data to estimate and calibrate the models. The trip generation model provides an estimate of the number of trips generated and attracted to each traffic analysis zone (TAZ) in the study area.

Modeled transportation volumes are driven by the first step: trip generation. Conventionally, trip rates are calculated either by cross-classification or –though less common nowadays– by regression analysis. Cross-classification models condense the diversity in trip making into one single average trip rate per household type and trip purpose. As an example, **Table 1** shows the observed work trips for households with 1 worker. While the observed number of trips ranges from 0 to 5, the cross-classification model uses the average of 1.24 trips for all households of this type.

Table 1. Observed work trips of households with one worker[1]

Number of trips	Number of records	Expanded number of records
0	150	28,564
1	137	23,916
2	239	45,724
3	5	1,216
4	10	1,481
5	1	142
	Average Trip Rate:	1.24

In reality, the number of work trips may be influenced by many other factors, such as income, home and work location, auto-availability, presence of children, occupation, or education, among others. These diverse household attributes cannot be represented in aggregate trip-based approaches.

This paper describes a new approach in which the aggregate trip generation step was replaced with a microscopic trip generation module. The microscopic approach allows representing the full range of observed trip-

[1] Source: 2007/2008 TPB Household Travel Survey for the Baltimore/Washington region

making behavior. Using the example shown in **Table 1**, microscopic trip generation allows the model to generate anything between 0 to 5 work trips for households with one worker, rather than applying one static average trip rate to each of these households.

2. State of the art

Trip generation models have taken many forms over the years, including zonal regression models, household regression models, and cross-classification models. Early travel forecasts consisted primarily of the extrapolation of "desire lines" developed from origin-destination (OD) surveys (Federal Highway Administration, 1975). This practice advanced in the early 1950s to consider land use and socio-economic factors in quantifying urban trip volumes, providing an analytical approach for using future land use plans to estimate future travel demand (Federal Highway Administration, 1975). Regression models of trip generation became commonplace in the late 1950s and early 1960s opening the door for a greater insight into travel and the factors influencing it (Federal Highway Administration, 1975). Regression models have the advantage of allowing the analyst to consider multiple independent variables, but the disadvantage of treating trip rates as continuous rather than discrete.

The 1970s marked a shift away from aggregate zonal level regression analysis to disaggregate household cross-classification procedures. Cross-classification models estimate an average number of trips as a function of two or more household attributes (Ortuzar and Willumsen, 2011). This method has long been the most established model for estimating trips in a travel demand model. Cross-classification models overcome the limitations of regression models, but introduce another shortcoming with respect to the number of variables and stratifications considered before violating the minimum sample size requirements (about 30 samples per stratification), or conversely making the survey sample size prohibitively expensive. Another disadvantage of cross-classification is the lack of goodness of fit measures.

New model forms are becoming more common in the toolbox of models considered for trip generation. These disaggregate models based on discrete choice analysis are considered by some to be a major innovation in the field (Ben-Akiva and Lerman, 1985). While commonly used for mode choice modeling, recent applications have also considered destination choice, and even more recently generation choice (Golob, 2000). Generation choice models estimate the frequency of daily person trips or tours.

Models that estimate person trips are an improvement over household based models as they allow for a greater use of important variables and are more compatible with other components of the modeling system (Ortuzar and Willumsen, 2011).

Disaggregate trip generation models offer several advantages over the commonly used cross-classification model, including the flexibility to consider more independent variables, the ability to include continuous variables in addition to classification variables, and statistical measures for evaluating the significance of the independent variables. Also, unlike the cross-classification model, where sample size quickly limits the number of stratifications due to the requirement that any given cell has at least 30 observations, a disaggregate model can capture multiple variables, making it possible to capture relationships that are not possible with the standard cross-classification approach (Parsons Brinckerhoff, 2007). While disaggregate trip and tour generation has been accomplished for activity-based models, no comparable approach has been published for trip-based models. This paper aims at filling this gap.

3. Study area

This microscopic trip generation module was developed for the Maryland Statewide Transportation Model (MSTM). The MSTM is a state-of-practice four-step travel demand model that covers the state of Maryland and a buffer region around the state. An additional geographic layer for long-distance trips covers North America from Canada to Mexico (Mishra et al., 2011).

The MSTM trip generation module is designed as a traditional cross-classification trip generation model that distinguishes 20 household types for work trips (households by number of workers [4 stratifications] and by income [5]) and 25 household types for non-work trips (households by size [5] and by income [5]). This household stratification was simply copied from the travel demand model of the Baltimore Metropolitan Council and not further questioned for implementation in the MSTM. Trip generation rates were calculated using the 2007-2008 TPB/BMC Household Travel Survey, a survey conducted jointly by the Baltimore (BMC) and Washington (MWCOG) metropolitan planning organizations. For this survey, 14,365 households reported their travel behavior.

In particular, there were three reasons why the current household type segmentation had to be overhauled. For one, some household types did not have enough survey records and had to be aggregated with neighboring household types for trip rate estimations. Secondly, several trip rates were almost identical for the current household type segmentation, indicating that resources were not allocated very efficiently. Especially, households of income groups 3, 4 and 5 had very similar trip rates for most trip purposes. Thirdly, a recently implemented auto ownership model now allows using auto ownership or auto availability for household type segmentation. Including auto ownership in trip generation was expected to improve trip rate estimations, as households with higher auto availability tend to travel more (Giuliano and Dargay, 2006: 118).

4. Econometrically driven household type segmentation

Traditionally, households are segmented into household types, and trip rates are calculated separately for each household type. Household types are defined based on experience at best, but more often than not simply based on preconceived notions of which household type segmentation "seems to make sense" for the task at hand.

Aggregate trip generation models sometimes distinguish work and non-work trips, but often they use the same household type segmentation for all purposes. Microsimulation, in contrast, allows capturing more household attributes than aggregate approaches, and household types may be defined differently for different purposes and modeling tasks. Therefore, special attention was given to defining household types specifically for each trip purpose.

In this research, household types were defined using a Big-Data approach to optimally represent trip-making behavior. Big Data is defined as a research approach that uses volumes data that are too large to process using traditional database and software techniques (Mayer-Schönberger and Cukier, 2014). Big Data research is an exploratory approach, in which it becomes irrelevant *why* a certain household type segmentation is found, only *what* segmentation is found matters. However, the revealed segmentation is reviewed for reasonability, as shown at the end of this section. Rather than predefining household types, the household travel survey is analyzed to identify household types that ideally distinguish trip-making behavior. Five household attributes were taken into account for defining household types:

- household size (1-7+),
- number of workers (0-4+),
- income category (1-12),
- auto-ownership (0-3+) and

• region (urban, inner suburbs and outer suburbs).

In this Big Data approach, all possible household type definitions were tested using these attributes. Without further aggregation of these attributes, 5,040 household types (= 7 x 5 x 12 x 4 x 3; this not to be considered Big Data yet) would be created. Many household types would be rare types (such as households with income category 1 with 3 or more autos) that would be underrepresented in the survey or not be represented at all. As discussed in section 0, it is state of practice to mandate that every household type definition is supported by at least 30 household records in the survey. To ensure that sufficient survey records are available for each defined household type, household attributes need to be aggregated.

Aggregations may happen across several attributes, and within each attribute two or more categories can be lumped together. **Fig. 1** shows the potential aggregations of a single attribute with 4 categories (which could be, for example, 4 income categories). All values could be kept separate (shown in row 1), two categories could be aggregated (rows 2 through 5), three categories could be aggregated (rows 6 and 7), or all categories could be lumped together (row 8). With four categories, this attribute can be aggregated in eight different ways.

1	Aggregate 1	1-1	2-2	3-3	4-4
2	Aggregate 2	1-2	3-3	4-4	
3		1-1	2-3	4-4	
4		1-1	2-2	3-4	
5		1-2	3-4		
6	Aggregate 3	1-3	4-4		
7		1-1	2-4		
8	Aggregate 4	1-4			

Fig. 1. Eight aggregation options for a variable with four categories

A computer algorithm was written to identify all possible aggregations for any number of categories. **Fig. 2** shows that the number of aggregation options increases exponentially. While **Fig. 1** could be derived easily in a manual way, the same aggregation sets would be labor-intensive and error-prone to create manually for attributes with 10 or more categories. A computer algorithm was written to create possible aggregation shown in **Fig. 2**.

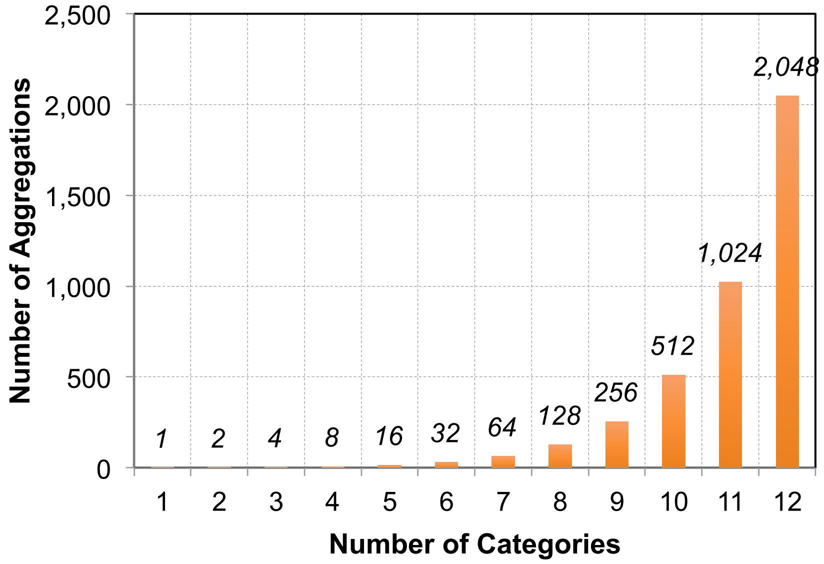

Fig. 2. Number of categories and number of aggregation options

Table 2 lists all attributes of the household travel survey that were considered relevant for trip generation. The number of categories for each attribute is provided in the second column. The column "No. of aggregations" lists the possible number of aggregations of categories for each attribute. The final row shows that the raw number of categories would lead to 5,040 combinations of attribute definitions. When possible aggregations are taken into account, the number of possible household type definitions increases to over 67.1 Million.

Table 2. Attributes and their number of categories and aggregations

Attribute	No. of categories	No. of aggregations
Household size	7	64
Number of workers	5	16
Income category	12	2,048
Auto-ownership	4	8
Region	3	4
Product	5,040	67,108,864

All 67.1 million household type definitions were generated and analyzed econometrically. For every household type definition, the number of rec-

ords per household type was counted. If one household type had fewer than 30 records, this household type definition was dismissed right away. This reduced the number of household type definitions to be further examined from 67.1 million to 51,401.

For the remaining 51,401 household type definitions, trip rate frequencies observed in the above-mentioned household travel survey were calculated. Within each household type definition, between 1 and 72 household types were distinguished (no household type definition with more than 72 types was found, as more types violated the rule of having at least 30 survey records per household type).

The standard deviation of trip frequencies were calculated for each household type within a given household type definition. For the household type shown in **Table 1**, the standard deviation would be 0.94. As will be shown below, much better segmentations can be found for work trips. If the standard deviation is small, a household type definition is assumed to represent well differences in trip-making behavior. Conversely, a large standard deviation suggests that the household type definition does not represent well differences in trip making. The coefficient of variance was calculated as well. It was found, however, that the standard deviation and the coefficient of variance correlated closely. Only the standard deviation was used subsequently.

In line with the state-of-practice in trip-based modeling, six trip purposes were distinguished:

- Home-based work (HBW)
- Home-based shop (HBS)
- Home-based other (HBO)
- Home-based education (HBE)
- Non-home-based work (NHBW)
- Non-home-based other (NHBO)

Traditionally, the same household type definition is used for every trip purposes. This research, however, revealed that varying household type definitions by trip purpose represent much better differences in trip making. In a microsimulation environment, it is almost effortless to distinguish household type definitions by trip purpose, an undertaking almost impossible in an aggregate approach.

Table 3 shows the household type definitions found to be ideal for every trip purpose after analyzing 51,401 possible segmentations. The ideal household type definitions were chosen based on the average standard deviation across trip rates within one household type. Statisticians generally advise not to average standard deviations. To calculate the average standard deviation, variances were calculated and averaged, and the square root of their average was taken.

Table 3. Household type definitions identified to distinguish trip-making behavior

Purpose	Definition ID	No. HH Types	min No. Records	ave No. Records	max No. Records	min StdDev	ave StdDev	max StdDev	min CoeffOfVar	ave CoeffOfVar	max CoeffOfVar
HBW	3329446	24	48	598.5	1,936	0.27	0.45	2.27	39.60	32.95	710.64
HBS	73434	30	30	478.8	2,514	0.86	0.71	4.07	112.69	23.15	175.93
HBO	45065	30	30	478.8	2,514	1.19	1.06	6.49	77.16	18.62	136.78
HBE	73434	30	30	478.8	2,514	0.06	0.39	2.97	74.18	93.68	1774.82
NHBW	3358727	42	30	342.0	1,544	0.11	0.50	2.01	100.94	50.74	1115.79
NHBO	309325	18	39	798.1	3,410	0.85	0.88	3.31	122.36	27.50	233.41

Table 3 also shows the smallest and highest standard deviations found. It was found, however, that minimum, average and maximum standard deviations correlate closely, which is why the selection of household type definitions could be reduced to review the average standard deviations. A small average standard deviation of trip frequencies was taken as evidence that a given household type segmentation represents well the observed trip-making behavior.**Error! Reference source not found.Table 4** reveals the household type definitions behind each Definition ID selected in **Table 3**. For HBW trip-making behavior, for example, household size was not found to be as relevant. Therefore, all seven household size categories were lumped together. Number of workers in each household, on the other hand, was identified to be very relevant for HBW trip-making behavior, and households with 0, 1, 2, and 3+ workers were distinguished. Income was fairly relevant, particularly at the high end, which is why categories 1-5, 6-7, 8, 9-10, 11 and 12 were kept separate. Little surprising, auto ownership was not found to be relevant, as most workers need to make work trips, regardless of auto availability. Regions (urban, suburban and rural) were not found to make a significant difference either, at least not if each household type needs to be covered by at least 30 survey records.

Table 4. Household type definition by size, workers, income, autos and region

Purpose	Definition ID	No. HH Types	sizeToken	workerToken	incomeToken	auto-ownership Token	regionToken
HBW	3329446	24	1-7	0-0.1-1 2-2.4	1-5.6-7.8-8 9-10.11-11.12-12	0-3	1-3
HBS	73434	30	1-1.2-2.3-3 4-4.5-7	0-4	1-6.7-12	0-1.2-2.3-3	1-3
HBO	45065	30	1-1.2-2.3-3 4-4.5-7	0-4	1-6.7-12	0-1.2-2.3-3	1-3
HBE	73434	30	1-1.2-2.3-3 4-4.5-7	0-4	1-6.7-12	0-1.2-2.3-3	1-3
NHBW	3358727	42	1-7	0-0.1-1 2-4	1-3.4-4.5-5.6-6 7-8.9-10.11-12	0-3	1-1 2-3
NHBO	309325	18	1-1.2-2.3-7	0-0.1-4	1-12	0-0.1-1.2-3	1-3

For non-work trip purposes, household size was found to be important and number of workers turned out to be irrelevant for most part. Auto ownership turned out to be important for all non-work purposes. This is in line with expectations, as many non-work trips are discretionary trips, and owning a car makes it easier, and therefore, more likely that discretionary trips are made. After these household types have been defined, a microscopic trip generation module was developed to replace an aggregate trip generation module.

5. Microscopic trip-generation methodology

In microscopic trip generation, trips by purpose are generated individually for each household. While aggregate travel demand models commonly work with aggregate socio-economic data, a microsimulation trip-generation module requires microscopic socio-economic data.

The land use model SILO (Moeckel, 2015) was used to create a synthetic population for the study area of the Maryland Statewide Transportation

Model (MSTM). SILO uses PUMS[2] micro data and expands these data to county-level control totals. PUMS data provide all household and person attributes necessary for microscopic trip generation, including household size, household income, number of workers and auto ownership. The microscopic households are updated for future years using the SILO land use model.

For every household, the number of trip is generated individually. Separately for every purpose, the household type definition shown in **Table 4** is used to define the household type. Using the household travel survey, the trip frequency distribution for a given household type and a given purpose (an example was shown in **Table 1**) is used to randomly select the number of trips generated by this household.

The flow diagram in **Fig. 3** shows the microscopic trip generation procedure. For every household, the number of trips generated for each purpose is chosen by Monte Carlo simulation based on the observed trip rate frequencies for this household type. Instead of selecting an average number of trips for each household of the same type, some households will be chosen to have unusually many trips, while others may be chosen to have no trips. This variety in trip generation is more realistic than assigning the same average number of trips to each household.

[2] Public Use Micro Data (PUMS) are anonymized microscopic census data of individual households and their household members, available for download at http://www.census.gov/acs/www/data_documentation/public_use_microdata_sample/

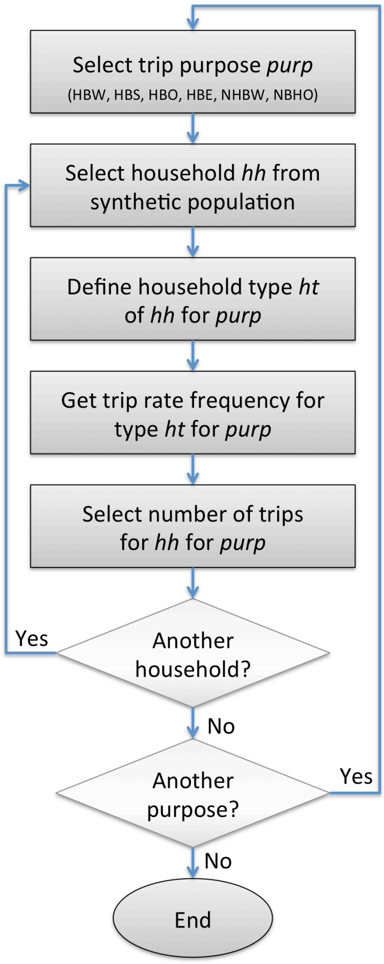

Fig. 3. Flow diagram for microscopic trip generation

The procedure continues until trips were generated for all households and all six trip purposes. The result of this step is a long list of trips by household and by purpose. As long as the following steps of the MSTM treat travel behavior in the aggregate form, trips are aggregated to trips generated by zone and purpose. As discussed below, however, this aggregation is only a placeholder until further steps of the MSTM are converted into microscopic modules as well.

6. Conclusions

Microsimulating trips offers several benefits. Foremost, the actual trip frequencies observed in travel demand surveys are preserved. Instead of forcing every household to generate an average number of trips –usually a fractional number that is not completed by any household– observed integer number of trips are generated for every household. The variety of trip frequencies is preserved as it can be found in reality.

Furthermore, household types do not need to be defined identically for every trip purpose. As shown in **Table 4**, household size is irrelevant for work-related trips, but highly relevant for non-work trips. For number of workers, the opposite is true. Income turned out to be very relevant for work-related trips, while auto ownership did not affect work trips. The distinction of regions (urban, suburban and rural) only turned out to be relevant for the trip purpose non-home-based work, for which urban received a different trip generation frequency than suburban and rural. This nicely aligns with the observation that most non-home-based work happen in urban centers where the highest job densities can be found.

A limitation of this approach is that travel behavior is still represented in trips and not in activities that result in tour generation. Activity-based models (Vovsha et al., 2005) care for the actual purpose of making a trip (that is doing an activity). By individually representing travelers, activity-based models usually create tours rather than single trips. In aggregate 4-step models, tours are represented less thoroughly by introducing non-home-based trips. Activity-based models are generally considered to better represent travel behavior because non-home-based trips are replaced by tour-based travel. However, implementing an activity-based model is a significant undertaking, and the proposed approach offers several benefits of microscopic modeling while keeping a working model operational. This development paradigm pursued here is called agile development in computer science (Donnelly, 2010). In agile approaches, single modules are updated while an operational model is preserved at any time. Resources are focused on modules that deserve most attention for improvements. Agile model development promises to implement advanced models while keeping an existing model operational. **Fig. 4** shows the vision of this development path for the MSTM.

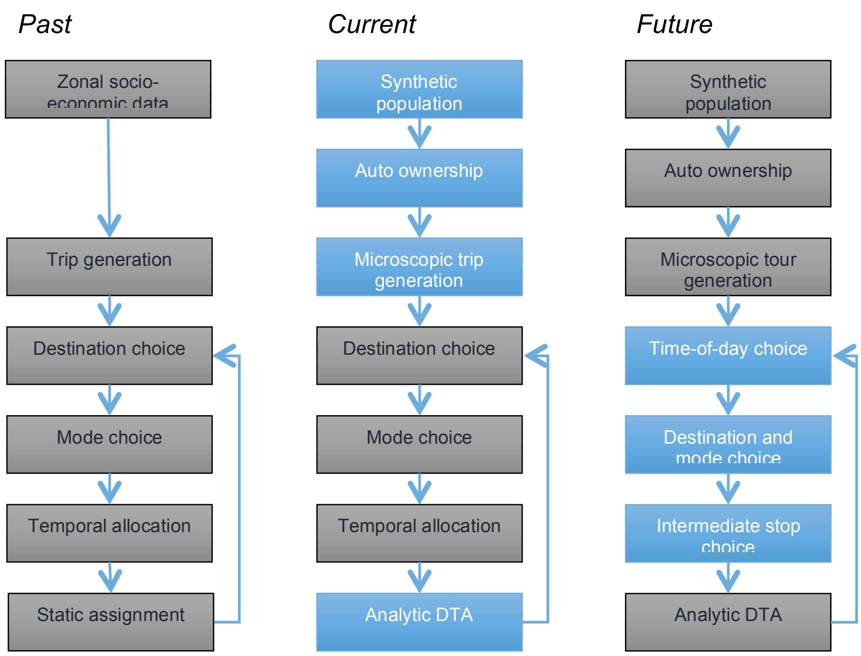

Fig. 4. Agile development of the MSTM towards an activity-based model

Following this envisioned path, it is planned in the near future to add a microscopic time-of-day choice model and a microscopic destination choice model. A side benefit of modeling destination choice microscopically will be to preserve the regular workplace defined in the synthetic population. In traditional aggregate models, the workplace is chosen every iteration anew, and different travel times will trigger households to choose a new workplace instantaneously. Obviously, this behavior is rather unrealistic. Furthermore, aggregate models are unable to respect a household's travel budget, both in terms of time and money. Zahavi (1979) suggested that travel budgets are fairly constant and change at most slowly over time. Trip-based models, however, by definition to not respect travel budgets. In trip-based models, worsening congestion will drive households to spend more time traveling, which is a violation of Zahavi's empirical findings. A microscopic destination choice model can be used to ensure that average travel budgets are respected, both temporally and monetarily. Microscopic trip generation is a required first step our models towards this goal.

References

Ben-Akiva M E, Lerman S, 1985 *Discrete Choice Analysis: Theory and Application to Travel Demand* (MIT Press, Cambridge, Massachusetts)

Donnelly R, 2010, "Are your models agile, or fragile", in *TRB Innovations in Travel Modeling*, Tempe, AZ

Federal Highway Administration, 1975, "Trip Generation Analysis", (USDOT, Washington, D.C.)

Giuliano G, Dargay J, 2006, "Car ownership, travel and land use: a comparison of the US and Great Britain" *Transportation Research Part A: Policy and Practice* **40** 106-124

Golob T F, 2000, "A simultaneous model of household activity participation and trip chain generation" *Transportation Research Part B: Methodological* **34** 355-376

Mayer-Schönberger V, Cukier K, 2014 *Big Data* (Houghton Mifflin Harcourt, New York)

Mishra S, Ye X, Ducca F, Knaap G-J, 2011, "A functional integrated land use-transportation model for analyzing transportation impacts in the Maryland-Washington, DC Region" *Sustainability: Science, Practice, & Policy* **7** 60-69

Moeckel R, 2015, "Modeling constraints versus modeling utility maximization: Improving policy sensitivity for integrated land-use/transportation models ", in *Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ortuzar J d D, Willumsen L, 2011 *Modelling Transport* (John Wiley & Sons Ltd., Chichester)

Parsons Brinckerhoff, 2007, "Triangle Regional Model 2006 Documentation", (Triangle Regional Model Service Bureau, Raleigh, North Carolina)

Vovsha P, Bradley M, Bowman J L, 2005, "Activity-Based Travel Forecasting Models in the United States: Progress since 1995 and Prospects for the Future", in *Progress in Activity-Based Analysis* Ed H Timmermans (Elsevier, Amsterdam, San Diego CA, Oxford) pp 389-414

Zahavi Y, 1979 *The "UMOT" Project* (U.S. Department of Transportation, Washington D.C.)