CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

A BLENDING APPROACH FOR TRIP GENERATION RATES ESTIMATION: MINIMUM SAMPLE SIZE FOR SPECIAL GENERATORS

Feras Elsaid[1,2], Luis Amador-Jimenez[1]
[1] Concordia University, Canada
[2] f_elsaid@live.concordia.ca

Abstract: Sometimes governments embark in expensive data collection campaigns to estimate local trip generation rates, given a suspicion that rates from well-established manuals are not applicable to local circumstances. This leads to inaccurate evaluation of the impacts that new developments introduce in the local transportation system. The estimation of trip rates requires collecting data from several sites with common characteristics; similar land use subcategories. However, certain land use subcategories may have limited number of local sites, such is the case with special generators. This paper presents an approach that blends rates from the few available local observations with rates from other databases or manuals. Moreover, this papers addresses the question of which manual to use to blend the rates and how many local observations are required to obtain an accurate estimation of local trip rate. An incremental approach, based on Full Bayesian, is followed to estimate the minimum required number of observations. Raw trip count data from previous studies are used to test the effect of blending rates from other manuals with rates estimated from local sites. The analysis in this study departs from the knowledge of the true rate, however, this rate is reserved for validation. The results show that at least three sites are required when the prior rate is borrowed from another manual of a region with similar characteristics and travel habits. However, up to 18 sites are required to obtain an acceptable estimate when the initial rate comes from a database of a dissimilar location.

1 INTRODUCTION

Trip generation can be traced back to the 1960's, when the Institute of Transportation Engineers (ITE), commenced to count trips from developments in suburban areas in the United States of America. The original aim was to use these rates in conjunction with traffic impact assessment to estimate the impact of new developments on the road network. Since then, the understanding of impact has been extended to multimodal alternatives including transit and active transportation (TDB 2010). Trip generation rates are amply used today by transportation consultants to estimate travel demand for new developments and the corresponding impacts on the transportation network when preparing Transportation Impact Assessments.

Two main approaches exist to produce trip generation and parking rates: the one centered on a database containing the raw data from which the analyst can estimate the rates and the one based upon the ITE manual in which the consultant is provided directly with the rates or equations to estimate them (used by ITE). There are several databases worldwide such as TRICS for United Kingdom and Trips Database Bureau (TDB) for New Zealand and Australia.

Several transportation agencies and regional authorities voiced concerns of the applicability of ITE rates on their jurisdictions (e.g. Arizona, British Columbia, Vermont, Florida, California) (TRANS Committee

2009). These concerns are due to the different considerations taken in ITE rates, and different lifestyle and travel patterns. Even for cities in North America, ITE national average rates should be used as a first approximation. This is particularly true for regions outside North America, in which the PM peak hour of commercial land uses could be shifted by late (Dubai, UAE) or early closure (Canada), the dominance of the automobile is challenged by public transportation (Singapore and Hong Kong), the culture for active transportation affects modal split (Amsterdam), or simply people are forced to travel by car because of extreme seasons and environments (Iceland, Finland, Northern Russia, Northern Canada and Gulf countries).

Hence many authorities attempt to produce local calibrations of the rates (i.e., Dubai, Abu Dhabi, Qatar). However, the main question is: how many sites is required to estimate reliable rates for a given land use subcategory?. The common approach is to initially start collecting data from a small number of sites of the same land use class (subcategory). Then, testing for the statistical fitness of the obtained rate. According to ITE (2017), a low coefficient of determination (R^2) triggers reiterations with more sites (incremental approach) until the R^2 complies with the minimum level desired. For TRICS the target is to observe low mean to medium trip rates; in order to consider that the rates do not have a significant weighting factor (bias). However, TRICS advises that such estimation can only proceed when at least 20 sites are being used. Securing sufficient sites in some categories which satisfy the selection criteria is challenging. Even if a site satisfies all such screening criteria, still the estimation of the trip generation and parking rates, of such development, requires data relating to the site itself. If such data is unavailable, the site cannot be used. In addition, there is also the need to collect data from sites of different orders; sites which have various magnitudes of the independent variable. Hence, being able to reduce the number of the required sites to sample and gather data is a key.

Several transportation agencies are still modeling transportation demand using the sequential four-step procedure. The first step in this modelling procedure involves estimating the number of trips produced or attracted among different origin and destination zones using trip rates from surveys, cross-classification analysis or regression analysis. However, a special attention should be paid to facilities which shall be classified as special generators. Special generators are facilities with trip generation characteristics differing from those of the facilities related to residential, commercial and industrial territories (Levashev, Mikhailov, and Sharov 2018). Trip generation characteristics of special generators are not fully captured by typical trip generation models. These facilities include universities, military bases, educational institutions, hospitals, and major shopping center (Levashev, Mikhailov, and Sharov 2018). The unique characteristics of special generators have a major impact on the transportation system; therefore, treating them like other generators such as a state capital or a major downtown employer does not realistically represent the performance of the transportation system and the behavior of its users (Desai Monicaba Vala 2017).

The literature points out at the need to survey a significant number of sites. ITE (2017) advices that at a bare minimum three sites might be used to estimate a rate, but at least six sites are required, and preferably 20 sites should be used (also suggested by TRICS 2017). In this regard, it is important to recall the possibility to blend rates. In fact, ITE (2017) suggests, in section 9.5, the possibility to blend ITE rates with locally estimated rates. The Ottawa-Gatineau region utilized such a procedure to combine locally estimated rates with those of ITE (TRANS Committee 2009). In this study, we propose taking advantage of the possibility to blend rates to answer the minimum sample size conundrum.

2 OBJECTIVE

The objective of this paper is proposing an approach that estimates blended trip rates, investigates the influence of the trip rate selected to be used as the point of departure, and determines an advisable sample size to accomplish an accurate estimation.

3 LITERATURE REVIEW

Classical statistical analysis uses simple linear regression with the mean trip rate (μ) as the dependent variable (in and out, during the AM, MD and PM periods) and at least one independent variable (x_i) to estimate trip generation and attraction rates. The independent variable is often related to the floor space, number of employees, number of students, or number of rooms. The traditional formulation for the calculation of trip generation rate (β_1) contains a correction factor for the observed level of occupancy (α), as shown in Equation 1.

[1] $\mu = \alpha \beta_1 x_1$

Bayesian statistics can also be used to estimate the trip rate. In general, it requires two elements: (1) a prior value for the trip rate, (2) the likelihood. The prior value consists of previous trip rates from other manuals, databases, or based on an expert's opinion (in the worst case). The likelihood is obtained by the locally collected data. The prior value and locally collected data are mixed to obtain a blended trip rate.

Full Bayesian is the process by which the classical Bayesian statistics are used through an iterative process; a walk algorithm moves over the space of trip rate parameters via a Markov Chain Monte Carlo simulation(Gamerman and Lopes 2006). This allows the construction of the probabilistic distribution of the trip rate. The algorithm uses the locally collected data and the prior to estimate a blended trip rate, called posterior of the trip rate.

One of the most important concerns in running MCMC simulations for the posterior inference is whether or not iterations are stable and convergence is reached. In fact, by running more than one chain the convergence can be verified graphically in a conventional program such as OpenBUGS. OpenBUGS is used for running Markov Chain Monte Carlo (MCMC) simulations for Bayesian inference (Lunn et al. 2009).

Bayesian statistics has some advantages in respect to Maximum Likelihood Estimation (MLE) such as (1) interesting probabilistic interpretative properties, (2) superiority in dealing with uncertainty and randomness, and (3) the ability to analyze complex data and data comprising small number of observations (Amador-Jiménez and Mrawira 2011; Gelman and Hill 2007; Mitra and Washington 2007). Moreover, in the Bayesian approach, hierarchical models can be introduced in the analysis resulting in more flexible model (Congdon 2010). Hierarchical models are those in which one or more parameters of the model are in turn dependent on a series of other parameters (called hyper-parameters) based on certain probability density functions (hyper-priors). In this case, also hyper-parameters follow a particular prior distribution. Different levels of hierarchy can be set up in the analysis. The Bayesian paradigm is widely used in some fields such as reliability engineering and epidemiology. In road safety, also, several researchers have applied Bayesian methods for hotspot identification, evaluation of countermeasure effectiveness, and parameter estimation in developing Safety Performance Functions (El-Basyouny and Sayed 2010; Lord and Miranda-Moreno 2008; Washington and Oh 2006).

4 Methodology

4.1 Blending Trip Generation Rates

The blending mechanism proposed in this study is based on the Full Bayesian MCMC simulation, as shown in Figure 1. The procedure starts with the selection of a trip rate (called prior trip rate) for a given Land Use Subcategory or Class (LUC), which can be obtained from external Trip Generation manuals or databases. This prior trip rate is used as a departure point in the estimation process. Then, a dataset containing data collected for a given independent variable (Gross Floor Area or Dwelling units) and the observed access counts (entry/exits) at sites with the same LUC is used as the likelihood in the estimation process. The minimum number of sites starts at two sites and follows an incremental process. Then the estimation of trip rates is done by blending locally collected data with the prior trip rate through a simulation process. The simulation is run with two chains to monitor the convergence. If the convergence is not verified, the number of locations (i.e. the sample size) is increased and a new trial is taken. The process continues until the convergence is verified by monitoring the two chains. Once the convergence is verified, the mean calculated based on the posterior distribution is the blended trip rate.

This study uses the data from 1989 at Montgomery County (Douglas and Douglas 1989) as the likelihood. The observed trip rates in Montgomery County (i.e. the true values) are treated as true values to which the model is expected to arrive. Trip rates obtained from the Dubai Trip Generation and Parking Rates Manual were used as prior trip rates, and the mechanism is expected to reject them and arrive to the true trip rates (from Montgomery County). Dubai's trip rates represent sites and information from very dissimilar circumstances, hence being ideal to test the approach to identify the true values. The experiment is done by incrementally increasing the number of sites in each trial. The estimation started with two sites, and an additional site was added in each new trial. Besides monitoring the convergence of the two chains, the difference between the estimated trip rate by the model and the observed trip rate is calculated for each trial. The sample size that corresponds to the trial at which the convergence is verified is considered to be the minimum sample size. The blended trip rate can be calculated as the mean of the posterior distribution.

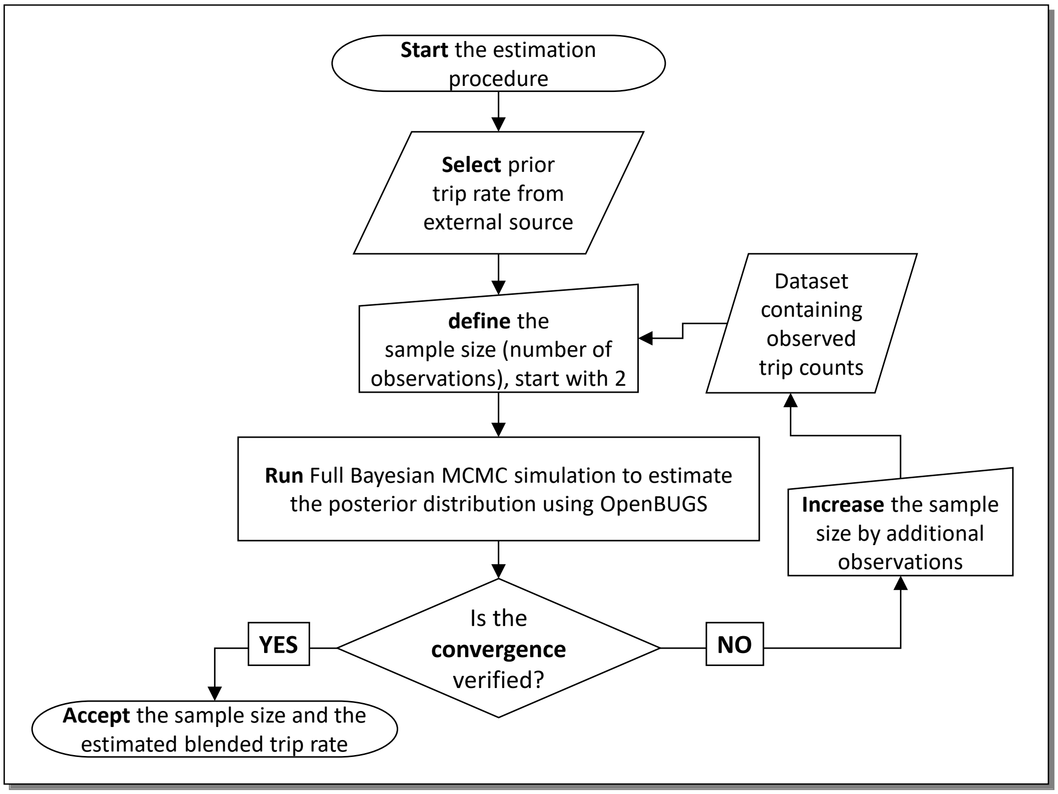

Figure 1: The proposed methodology for blending

The Full Bayesian MCMC analysis requires the definition of a functional form that captures the relationships between the response and the independent variable(s). Equation 2 shows the functional form used in this study; which follows a single variable response, widely accepted for estimating trip generation rates. The main contributing factor explaining trip generation is the independent variable (x_1), which varies according to the land use category. Two land use categories were considered in this study; residential and institutional (i.e. office buildings). The independent variable for residential was the number of Dwelling Units (DU), and for offices was the Gross Floor Area (GFA).

[2] $\mu = \alpha \beta_1 x_1$

where:

μ = the mean of the number of generated trips;

β_1 = the trip generation rate, to be estimated using the Full Bayesian MCMC analysis;

x_1= the independent variable; and

α = the occupancy factor, assumed to be 1 in this study.

A summary of the observed dataset can be found in Table 1. The trip counts include two major land use categories; residential and office buildings. For the residential land use category, three subcategories were considered; single-family detached houses, and apartments (high-rise and mid-rise buildings). For the office buildings category, there were no subcategories. The number of dwell units and gross floor area with their corresponding observed number of trips during the morning peak hour (AM) were used to estimate blended trip rates. In this study, the midday (MD) and afternoon (PM) trips were excluded since they simply can be estimated following the same steps. Table 2 shows trip rates for stand-alone houses, two-bedroom apartments and multi-user office buildings. The standard deviation values for these subcategories were not provided, hence a value of 0.5 was assumed in this study.

Table 1: Observed trip generation rates for three Land Use Subcategories

Subcategory	**Single-family detached houses**	**Rise apartments**	**General offices**
Average trip rate	0.73	0.38	1.55
Standard deviation	0.21	0.14	0.49
Total number of sites	18	18	25
Source	Douglas & Douglas (1989)		

Table 2: Trip generation rates used as departure points (prior trip rates)

Category	**Average trip rate**
Stand-alone houses (trips/DU)	5.86
Two bedrooms apartment (trips/DU)	0.70
Multi-user office building (trips/ 100 GSM)	2.08
Source	Dubai Trip Generation and Parking Rates Manual (2013)

4.2 The effect of the departure point (prior trip rate)

This is done to investigate the effect of the trip rate used as the point of departure, which can be obtained from an external manual, on the number of locations (sample size) required to estimate the blended trip rate. The standard deviation (σ) can be used as an indicator of the degree of similarity between the two locations: the location where the prior trip rate has been estimated (i.e. trip rates found in manuals), and the location where the blended trip rate is needed to be estimated. The analysis was done based on the trip counts obtained from Douglas and Douglas (1989). Two samples of 15 and 25 locations under the category of office buildings were used. The observed trip generation rate for office buildings category is 1.55 trip/1000 GSF and the standard deviation is 0.49. In this analysis, different arbitrary mean values of prior were calculated using Equation 3. Each calculated prior value corresponds to a certain range of standard deviation. This represents various scenarios of having locations with different travel behavior. The standard deviation of prior is assumed to be 0.5 in this analysis. For example, by setting c_p and c_p to be 3, the arbitrary value of prior trip rate is 4.55.

[3] $$\mu_p = \mu_b + c_p\sigma_P + c_b\sigma_b$$

where

μ_p = prior mean trip rate

μ_b = blended mean trip rate

σ_p = prior standard deviation

σ_b = blended standard deviation

c_p and c_b: factors defining the standard deviation range

5 Results

5.1 Blending Trip Rates

The analysis was run using 100,000 iterations for each trial (a sample size of 200,000 for both chains in each trial). Initial values for chains were set to 0 and 10. As the sample size increases, the estimated blended trip rates, as depicted by the two chains, get closer to the observed trip rate. This is because of the increasing effect of the likelihood (dataset containing locally observed trip counts) and its ability to overcome the influence of the prior trip rate. The analysis was done considering three cases corresponding to three land use subcategories in Dubai Trip Generation and Parking Rates Manual (2013): stand-alone houses, two-bedroom apartments and multi-user office building.

The first case was stand-alone houses. For this case, the percent of error suddenly dropped to approximately zero at a sample size of 18 locations. Figure 2 reveals that estimated blended trip rates, based on 17 locations, have not converged (multinomial distribution and separate chains). Figure 3 shows the convergence of the two chains as well as the probabilistic distribution of the independent parameter β_1 (i.e. the estimated blended trip rate). Table 3 shows mean, median, quantiles, observed trip rate, and calculated error for each trial. The need for such large number of sites might be due to the relatively large difference between the prior trip rate and the observed trip rate; 5.86 and 0.73 respectively, as shown in Table 5.

Table 3: Case 1: stand-alone houses

Number of locations (sample size)	Blended trip rate (mean)	Value at 2.5 percentile	Median	Value at 97.5 percentile	Observed trip rate	Calculated error (%)
5	5.587	4.574	5.589	6.591	0.896	524
6	5.535	4.514	5.538	6.544	0.842	558
7	5.476	4.446	5.479	6.490	0.850	544
8	5.422	4.384	5.424	6.442	0.814	566
9	5.360	4.309	5.362	6.395	0.803	567
10	5.299	4.243	5.303	6.337	0.797	565
11	5.236	4.168	5.241	6.282	0.786	566
12	5.180	4.100	5.185	6.233	0.757	585
13	5.112	4.016	5.118	6.176	0.750	582
14	5.044	3.932	5.051	6.116	0.741	581
15	4.506	0.681	4.896	6.029	0.728	519
16	2.170	0.621	0.734	5.726	0.721	201
17	3.645	0.731	4.487	5.823	0.740	393
18	0.729	0.629	0.725	0.852	0.730	0

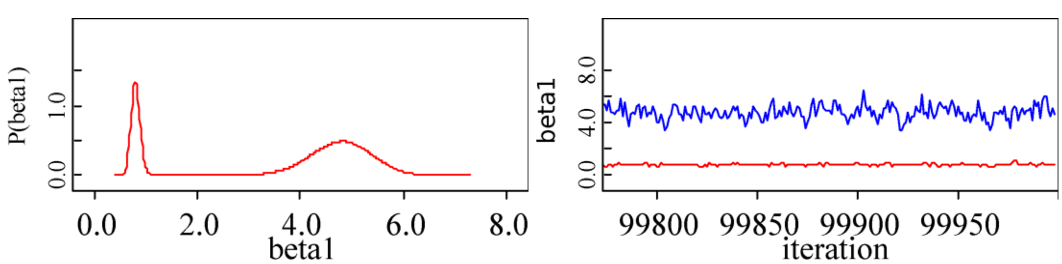

Figure 2: Probabilistic distribution of estimated blended trip rates (left) and trace of two chains in OpenBUGS (right), sample size of 17 sites, stand-alone houses

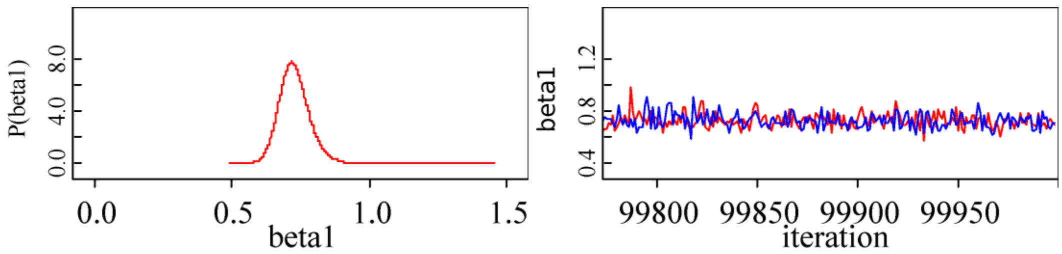

Figure 3. Probabilistic distribution of estimated blended trip rates (left) and trace of two chains in OpenBUGS (right), sample size of 18 sites, stand-alone houses

For the other two cases, the results show that sample sizes as low as 2 and 3 can be sufficient for subcategories of two-bedroom apartments and multi-user office buildings, respectively. This can be explained as the prior and observed trip rates are close indicating very similar characteristics (Table 4).

Table 4: Summary of results of the analysis of the three cases

Case	Subcategory	Sample size	Estimated mean	Value at 2.5 percentile	Median	Value at 97.5 percentile	Error (%)
1	Stand-alone houses	18	0.729	0.629	0.725	0.852	0
2	Two-bedroom apartment	2	0.230	0.150	0.220	0.370	4.5
3	Multi-user office building	3	1.784	1.312	1.758	2.408	3

5.2 The Effect of the departure point (prior trip rate)

The effect of the number of locations (i.e. sample size) can be understood as follows; as the sample size increases the influence of collected dataset increases, so that it easily contradicts the prior trip rate obtained from other sources. This is obvious when analyzing larger sample sizes (i.e., 25 locations). The standard deviation range defined by c_p and c_p represent the degree of similarity in terms of trip generation characteristics between two locations. For simplicity, the two factors were set to be equal (c). The results of investigating the influence of prior trip rate are shown in Table 5. The influence of the prior trip rate diminishes as the sample size increases. As noticed from Table 5, the prior trip rate has insignificant effect when the sample size is 25 sites. This indicates that the estimation is reliable even when the two locations have very different characteristics (i.e. c=5). When the sample size is 15 locations, the percent of error is acceptable up to c=3. However, the percent of error is relatively high for larger values of c, as shown in Table 6.

Table 5: Results of analysis of a sample of 25 locations of office buildings

Standard deviation range factor (c)	Arbitrary prior trip rate	Estimated blended trip rate	Observed trip rate	Error (%)
1	2.55	1.70	1.55	9.7
2	3.55	1.73	1.55	11.6
3	4.55	1.75	1.55	12.9
4	5.55	1.78	1.55	14.8
5	6.55	1.82	1.55	17.4

Table 6: Results of analysis of sample of 15 locations of office offices

Standard deviation range factor (*c*)	Arbitrary prior trip rate	Estimated blended trip rate	Observed trip rate	Error (%)
1	2.55	1.52	1.55	1.9
2	3.55	1.63	1.55	5.2
3	4.55	1.90	1.55	22.6
4	5.55	4.14	1.55	167.1
5	6.55	5.67	1.55	265.8

6 Conclusion

This study proposes an approach that is able to blend trip rates from the few available local observations with the international trip rates, which can be obtained from other databases or manuals. Moreover, it addresses the question of which database or manual is better to be used in estimating a blended trip rate. This study also proposes an approach to determine the sufficient number of locations to obtain an accurate estimation of blended trip rates. A case study is constructed for three land use subcategories: single-family detached houses, rise apartments and general offices. The trip rate that is used as the departure point in the analysis is obtained from external sources, a prior trip rate. A dataset that contains trip counts from a previous study is used to investigate the effect of trip rates obtained from other manuals. The analysis takes advantage of the knowledge of the true value of the blended trip rate; however, this value is reserved for validation and is not used in the estimation. An incremental approach is followed to define the minimum required number of sites. The results reveal that at least three locations are required to accomplish accurate estimation of blended rates when the prior rate is obtained from another manual of a region with similar characteristics and travel behavior. However, up to 18 sites are required to obtain a good estimate when the initial rate comes from a database of a dissimilar circumstances. This confirms the recommended minimum three sites by ITE and the 20 sites recommended by ITE (2017) and TRICS (2017). In other words, obtaining trip rates from other studies that were conducted at locations with similar travel habits, site characteristics and the same land use subcategory, is a key to reduce the minimum sample size required for estimating blended trip rates of the sites under study.

It is recommendable to borrow trip rates from other countries or cities with similar characteristics (cultural attitudes, climate, income, ethnicity, transportation system and geography). The Full Bayesian approach takes advantage of the available information from previous studies and uses this to orient the estimation of the trip rates, therefore, reducing the required sample size.

References

Amador-Jiménez, Luis, and Donath Mrawira. 2011. Reliability-Based Initial Pavement Performance Deterioration Modelling. *International Journal of Pavement Engineering* 12 (2): 177–86. https://doi.org/10.1080/10298436.2010.535538.

Congdon, P. 2010. *Applied Bayesian Hierarchical Methods*. CRC Press.

Desai Monicaba Vala, Vaibhav V. 2017. Trip Generation Characteristics of Special Generators. *IJSTE-International Journal of Science Technology & Engineering* | 3 (10). www.ceramicindia.com.

Dubai Trip Generation and Parking Rates Manual. 2013. Road and Transport Authority, Dubai, U.A.E.

El-Basyouny, Karim, and Tarek Sayed. 2010. Full Bayes Approach to Before-and-After Safety Evaluation with Matched Comparisons. *Transportation Research Record: Journal of the Transportation Research Board* 2148 (December): 1–8. https://doi.org/10.3141/2148-01.

Gamerman, Dani., and Hedibert Freitas. Lopes. 2006. *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference.* Taylor & Francis.

Gelman, Andrew., and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Levashev, Alexey, Alexander Mikhailov, and Maksim Sharov. 2018. Special Generators in Tasks of Transportation Demand Assessment. *Transportation Research Procedia* 36 (January): 434–39. https://doi.org/10.1016/J.TRPRO.2018.12.119.

Lord, Dominique, and Luis F. Miranda-Moreno. 2008. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-Gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective. *Safety Science* 46 (5): 751–70. https://doi.org/10.1016/J.SSCI.2007.03.005.

Lunn, David, David Spiegelhalter, Andrew Thomas, and Nicky Best. 2009. The BUGS Project: Evolution, Critique and Future Directions. *Statistics in Medicine* 28 (25): 3049–67. https://doi.org/10.1002/sim.3680.

Mitra, Sudeshna, and Simon Washington. 2007. On the Nature of Over-Dispersion in Motor Vehicle Crash Prediction Models. *Accident Analysis & Prevention* 39 (3): 459–68. https://doi.org/10.1016/J.AAP.2006.08.002.

Montgomery County Trip Generation Rate Study. 1989. Douglas and Douglas Inc., Maryland, U.S.

TRANS Trip Generation Residential Trip Rates: Study Report. (2009). TRANS Committee.

TRICS. 2017. Trip generation and land use categories.

Trip Generation and Parking Rates Manual. 2012. Department of Transportation, Emirate of Abu Dhabi, U.A.E.

Trip Generation Manual. 2017. Institute of Transportation Engineers (ITE). 10th ed., Washington, DC, USA.

Trips Database Bureau: TDB Database User manual. 2010. TDB.

Washington, Simon, and Jutaek Oh. 2006. Bayesian Methodology Incorporating Expert Judgment for Ranking Countermeasure Effectiveness under Uncertainty: Example Applied to at Grade Railroad Crossings in Korea. *Accident Analysis & Prevention* 38 (2): 234–47. https://doi.org/10.1016/J.AAP.2005.08.005.