Using Big Data and Efficient Methods to Capture Stochasticity for Calibration of Macroscopic Traffic Simulation Models

Sandeep Mudigonda,
Graduate Assistant, Ph.D. Candidate,
Rutgers Intelligent Transportation Systems (RITS) Laboratory,
Department of Civil and Environmental Engineering,
Rutgers, The State University of New Jersey;
96 Frelinghuysen Road, Piscataway, NJ 08854, USA;
Phone: (732) 445-2792, Fax: (732) 445-0577;
Email: sandeep.mudigonda@rutgers.edu

Kaan Ozbay,
Professor, Ph.D.,
Center for Urban Science + Progress (CUSP);
Department of Civil & Urban Engineering,
New York University (NYU);
One MetroTech Center, 19th Floor, Brooklyn, NY 11201, USA;
Tel: +1-(646)-997-0552, Fax: +1-(646)-997-0560;
Email: kaan.ozbay@nyu.edu

ABSTRACT

The predictions of a well-calibrated traffic simulation model are much more valid if made for various conditions. Variation in traffic can arise due to many factors such as time of day, weather, accidents, etc. Calibration of traffic simulation models for traffic conditions requires larger datasets to capture the stochasticity in traffic conditions. Big Data collected using RFID sensors, smart cellphones and GPS devices provide extensive traffic data. In this study we show the utility of Big Data to incorporate variability in traffic flow, speed for various time periods. However, Big Data poses a challenge in terms of computational effort. With the increase in number of stochastic factors, the numerical methods suffer from the curse of dimensionality. In this study, we propose a novel methodology to address the computational complexity due to the need for the calibration of simulation models under highly stochastic traffic conditions. This methodology is based on sparse grid stochastic collocation, which, treats each stochastic factor as a different dimension and uses a limited number of points where simulation and calibration are performed. A computationally efficient interpolant is constructed to generate the full distribution of the simulated flow output. We show that this methodology is much more efficient that the traditional Monte Carlo-type sampling.

INTRODUCTION

Simulation models are mathematical models in which output is derived from a particular model given the input. The input consists of two main groups of data: physical input data I_s (e.g., volume counts, capacity and physical features of roadway sections) and calibration parameters C_s (i.e., adjustable components of driver behavior). Outputs from a simulation model can be expressed as,

$$
\begin{aligned}
& O_{obs} : f(I_s, C_s) \rightarrow O_{sim} \mid I_s, C_s + \varepsilon \\
& f(I_s, C_s) = \text{functional specification of the internal models in a simulation system} \\
& O_{sim} = \text{ simulation output data given the input data and calibrations,} \qquad (1) \\
& \varepsilon = \text{margin of error between simulation output and observed field data, and,} \\
& O_{obs} = \text{observed field data.}
\end{aligned}
$$

The process of calibration entails adjusting the calibration parameters (C_s) so that the error between the output from simulation and field conditions is acceptable.

Calibration parameters estimated using limited samples are not always representative of all possible conditions of the simulated system and will thus result in inaccurate predictions. In other words, models that are not adequately calibrated cannot accurately capture time-varying conditions of traffic. Traditional sources of traffic data used in the calibration of traffic models are either limited by the availability of the data that only cover typical conditions or may not be reliable enough. However, with the advent of new technologies, information is on the fingertips of users by means of smart phones, GPS-equipped devices, RFID readers. The rapid rise in information technology has also resulted in innovative ways to obtain space- and time-sensitive information in real-time. This, in turn, has led to massive amount of passively collected location and event data for various time periods – also called "Big Data". Big Data provides an opportunity to validate and calibrate traffic simulation models for a variety of conditions.

Variability can be incorporated within *inputs (demands)* I_s and *calibration parameter set (supply)* C_s during different periods of the day, weather conditions, driver population composition, highway geometry, etc. There were studies that captured traffic variability (*1-5*) to name a few. However, the increase in the number of factors affecting stochasticity increases the dimensionality of the calibration process. This in turn results in increased computational effort required in calibrating traffic simulation models for different conditions such as variability within weekday/weekend, and seasonal variability, and special situations including adverse weather, workzones, etc.

In this study, we propose a novel methodology to address the computational complexity due to the need for the calibration of simulation models under highly stochastic traffic conditions. We show the utility of Big Data to incorporate variability in traffic flow, speed for various time periods.

LITERATURE & MOTIVATION

There are myriad of studies that deal with calibration of traffic simulation models (*5-11,12-19*). Due to space constraint, we show a sample of them in Table 1.

Table 1 also shows the data used in these studies for the calibrating process. It can be seen that in most studies data used for calibration is limited to AM and PM peak periods no more than a few days. Thus, the data captures only a few specific conditions, or is a dilute sample of different conditions. Hence, it is expected that the model predictions will only be accurate for those specific conditions.

Table 1 Summary of Literature on the Calibration of Traffic Simulation Models

Authors	**Complexity; Simulation Tool**	**Type of Roadway Section**	**Performance outputs**	**Data Used in Calibration**
Ma and Abdulhai (*7*)	Micro; PARAMICS	Urban	Traffic counts	Detector data for 1 hr during AM peak
Kim and Rilett (*6*)	CORSIM, TRANSIMS	Freeway	Volume	Data 5 loop detector stations for 13.9-mile section of freeway for 1 hr. during AM, PM and off peak
Hourdakis et al. (*11*)	Micro; AIMSUN	Freeway	Volume	5-min. data from 21 detector stations for a 12-mile freeway section during PM peak for 3 days
Jha et al. (*12*)	Micro; MITSIMLab	Urban Network	Travel time	Detector data for 15 days for AM and PM peaks on a large urban network
Toledo et al. (*13*)	Micro; MITSIMLab	Freeway, arterial	Speed, Density	Data from 68 detector stations on 3 freeways for 5 weekdays
Qin and Mahmassani (*14*)	Macro; DYNASMA RT-X	Freeway Network	Speed	Data from 7 detector stations on 3 freeways during AM peak for 5 weekdays
Balakrishna et al. (*15*)	Micro; MITSIMLab	Freeway, Parkway	Traffic Counts	15 min. data from 33 detector stations
Zhang et al. (*8*)	Micro; PARAMICS	Urban Freeway network	Flow, Occupancy	5-min detector count during PM peak for 7 days
Li et al. (*1*)	Macro	Freeway	Flow	Loop detector data
Lee and Ozbay (*5*)	Micro; PARAMICS	Freeway	Speed, Counts	5-min detector count during AM peak for 16 days
Zhong et al. (*3*)	Macro	Freeway	Flow	Loop detector data for 7 hours on three days in two years
Yang and Ozbay (*9*)	Micro; PARAMICS	Freeway	traffic conflict, lane change, volume & speed	NGSIM trajectory data for US-101 for 15 min.
Henclewood et al. (*18*)	Micro	Freeway	Travel time distribution	NGSIM trajectory data for Peachtree Street in Atlanta, Georgia for 30 min.
Punzo et al. (*19*)	Micro	Freeway	Speed	NGSIM trajectory data for I-180

The effect of data and parameter uncertainty in traffic simulation models has received considerable attention recently (*18*, *19*). Studies from other fields indicate that bias and variance in simulation output results are due to the bias and variance in the input models used, after simulation error is eliminated; the input models consist of simulation model inputs and parameters

Federal Highway Administration (FHWA)'s Traffic Analysis Toolbox (*20*) recommends that if GEH < 4 ($GEH = \sqrt{\dfrac{\frac{1}{T}\sum_{i=1}^{T}(O_{sim,i} - O_{obs,i})^2}{\frac{1}{2T}\sum_{i=1}^{T}(O_{sim,i} + O_{obs,i})}}$) for link volumes for 85% of the links and average travel times are within 15% of observed values, then it is considered as a satisfactorily calibrated model (*20*). In order to achieve this level of calibration for various conditions (peak, off-peak, weekends, normal and inclement weather, under accident, and other events), detailed level of data is required.

As illustrated in Table 1, most of the studies in traffic simulation used limited amount of data focusing on a small set of conditions and/or time periods. As depicted in Figure 1, using only smaller samples of data will not accurately capture variation in traffic data (Additionally, the sources of field data in most of these studies, except (*5-9*, *19*), have been traditional sources such as loop detectors or manually-acquired data from captured videos which can be cumbersome and not always accurate enough). Using these models for conditions other than the ones for which calibration data was available for would not yield accurate results.

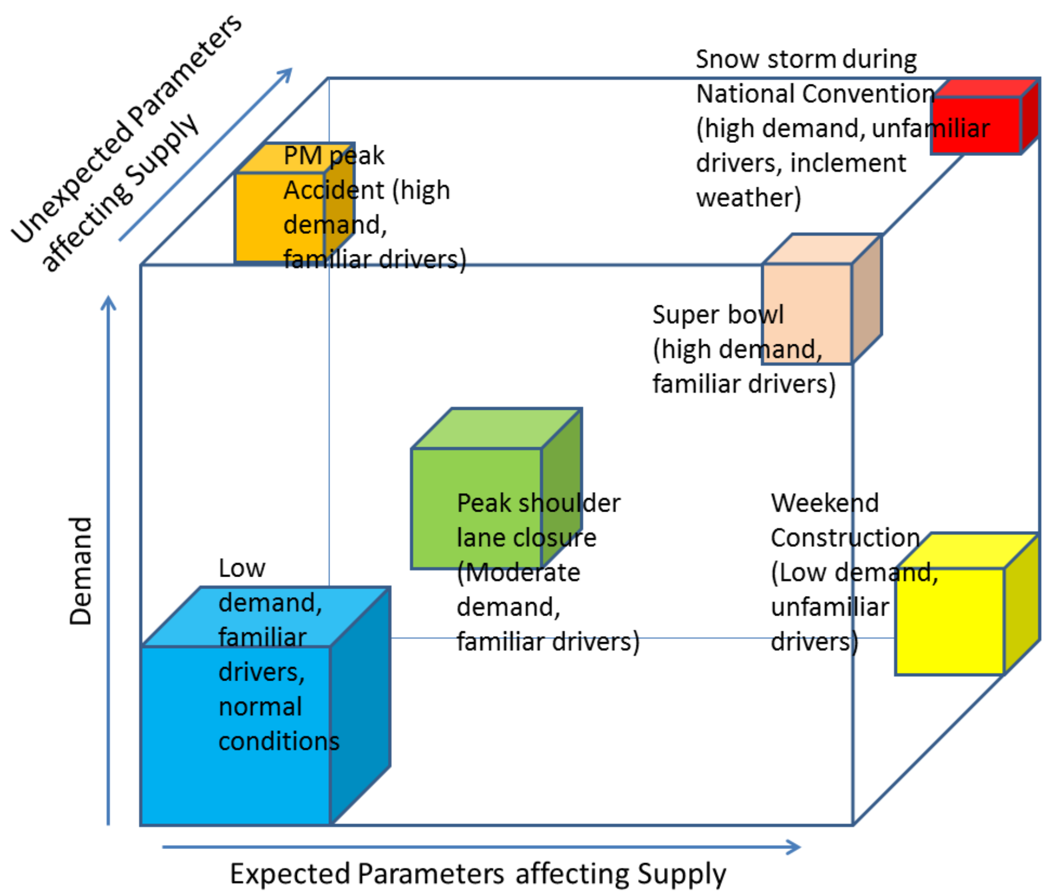

Figure 1. Illustration of various traffic conditions for which data is required for calibration (adapted from (*21*))

Ozbay et al. (*22*) showed that the existence of a "typical" day in traffic demand is not always likely. Hence, to obtain accurate predictions from a traffic simulation model, it is important to consider not only the demand from various clusters, but also the variation of demand within each cluster.

Computational complexity

In cases where large sources of data spanning different conditions are available, to capture the stochasticity in traffic conditions, there is an increase in number of factors of stochasticity. This in turn increases the dimensionality of the calibration process. This then results in increased computational effort required in calibrating traffic simulation models for various conditions.

Most studies capturing stochasticity in computational traffic models use a Monte Carlo (MC)-type independent sampling of M simulation runs for various traffic conditions. The number of replications needed to be at a level of precision γ is given by $\left(\frac{t_{M-1,1-\alpha/2} * S}{\gamma}\right)^2$. However, the convergence rate for MC-type method or Latin hypercube sampling is slow, $O(1/\sqrt{M})$ (*23*). All of these numerical methods suffer from the curse of dimensionality. Thus depending on the size of the network and number of stochastic dimensions, these approaches can become prohibitive in terms of computational effort. It may not be possible at all to simulate the output for each and every possible realization of parameter and input. Also, all possible points in the stochastic space of simulation output may not have the corresponding observed data. Thus it is important to obtain an effective interpolation methodology for predicting output for points with no data.

METHODOLOGY

Use of Big Data

Based on the discussions above, making accurate predictions using traffic simulation models requires calibration data from many sources and in great detail. This data need can be effectively addressed by the advent of new technologies such as GPS, cellular phones, RFIDs, etc. The ubiquity of these technologies ensures that data of great detail and variety are available. Various demand, speed, flow and event data can be obtained from the Big Data sources.

In this study, we use a hybrid of ETC data for demand and traffic sensor data for speed and flow. The electronic toll collection (ETC) data is collected for all toll ways in the U.S. and In New Jersey. Taking toll facilities in New Jersey as an example, New Jersey Turnpike (NJTPK) is spread over 150 miles with 28 interchanges and 366 toll lanes. Garden State Parkway (GSP) is about 170 miles long with 50 toll plazas and 236 toll lanes. Each freeway carries up to 400,000 vehicles per day (*24*). The ETC data is collected at toll plazas on these freeways. (*24*) The ETC dataset consists of the individual vehicle-by-vehicle entry and exit time data. It also consists of the information regarding the lane through which each vehicle was processed (both E-ZPass and Cash users), vehicle types, number of axles, etc.

The demand from the ETC data is divided into clusters so that different demand patterns can be analyzed separately. The corresponding sensor data is also obtained into clusters.

The simulation is performed using the clustered demand data distribution and simulation output of flow and density is compared to the observed distribution from sensor data.

Stochastic Collocation

As mentioned earlier, there is a need to combat the issue of high number of replications and dimensionality when using MC-type methods in capturing stochasticity using computational traffic models. Stochastic spectral methods provide an efficient alternative to MC-type methods. In these methods, stochasticity is treated as another dimension and the stochastic solution space Ω is discretized and approximated (Γ). One such stochastic spectral method is stochastic collocation (SC), where the approximation is performed using deterministic solutions at a set of prescribed nodes (collocation points) and an interpolation function. For a set of *Q* points in the stochastic space where the simulation output is calibrated, $\left\{\boldsymbol{\Theta}_j\right\}_{j=1}^{Q} \in \Gamma$, the polynomial interpolation can be shown as, (*25*)

$$\boldsymbol{\rho}(x,t,\boldsymbol{\xi}) \approx \int_{\Gamma} \boldsymbol{\rho}(x,t,\boldsymbol{\xi})\, p(\boldsymbol{\xi})\, d\boldsymbol{\xi} \approx \hat{\boldsymbol{\rho}}(x,\boldsymbol{\Theta}) = \sum_{j=1}^{Q} \tilde{\rho}_j(x)\Phi_k(\boldsymbol{\Theta}_j), \forall x \in D \tag{2}$$

$\Phi_k(\boldsymbol{\Theta}_j)$ = interpolating basis polynomials

$\tilde{\rho}_j(x)$ = the deterministic solution at $\boldsymbol{\Theta}_j$

The choice of weights, i.e., the basis functions Φ_k, in SC depends on the interpolation technique. It is important to note that, unlike MC-type sampling methods, the weights assigned to each deterministic output is different in stochastic collocation methods.

Computationally efficient schemes such as the Smolyak algorithm are used to reduce the number of collocation points (*26,27*) at higher dimensions of stochasticity. Smolyak algorithm, developed originally for multi-dimensional integration, entails evaluating deterministic solutions at the nodes of sparse sampling grids and building the interpolation function. One dimensional interpolant, with the Smolyak algorithm, is given by, $U^i(f) = \sum_{j=1}^{m_i} f(\Theta_j)L_j, m_i =$ # nodes at level *i*. Sparse interpolant in N-dimensions and q-N interpolation order is given by,

$$\Delta^i = U^i - U^{i-1}; U^0 = 0$$
$$A_{q,N}(f) = \sum_{|\mathbf{i}|\le q}(\Delta^{i_1} \otimes \ldots \otimes \Delta^{i_N})(f) = A_{q-1,N}(f) + \sum_{|\mathbf{i}|=q}(\Delta^{i_1} \otimes \ldots \otimes \Delta^{i_N})(f) \tag{3}$$

The advantage of this recursive/nested structure is that to increase the order of interpolation (accuracy) we can use all the deterministic solutions from the previous steps: $A_{q-1,N}$, by adding a few more deterministic solutions: Δ_{iN}. When new data is available, additional deterministic solutions can be evaluated and accuracy of interpolant is improved.

Convergence rate of the interpolant is of the order, $O(Q^{-2}|\log Q|^{3(N-1)})$ (for piecewise linear basis), $O(Q^{-k}|\log Q|^{(k+2)(N-1)})$ (for *k*-polynomial basis). This rate can be controlled by the polynomial order k (*26,27*). Thus, we show, empirically, that convergence of this interpolant is better than Monte Carlo method.

An illustration of the discretization of the stochastic space is shown in Figure 2.

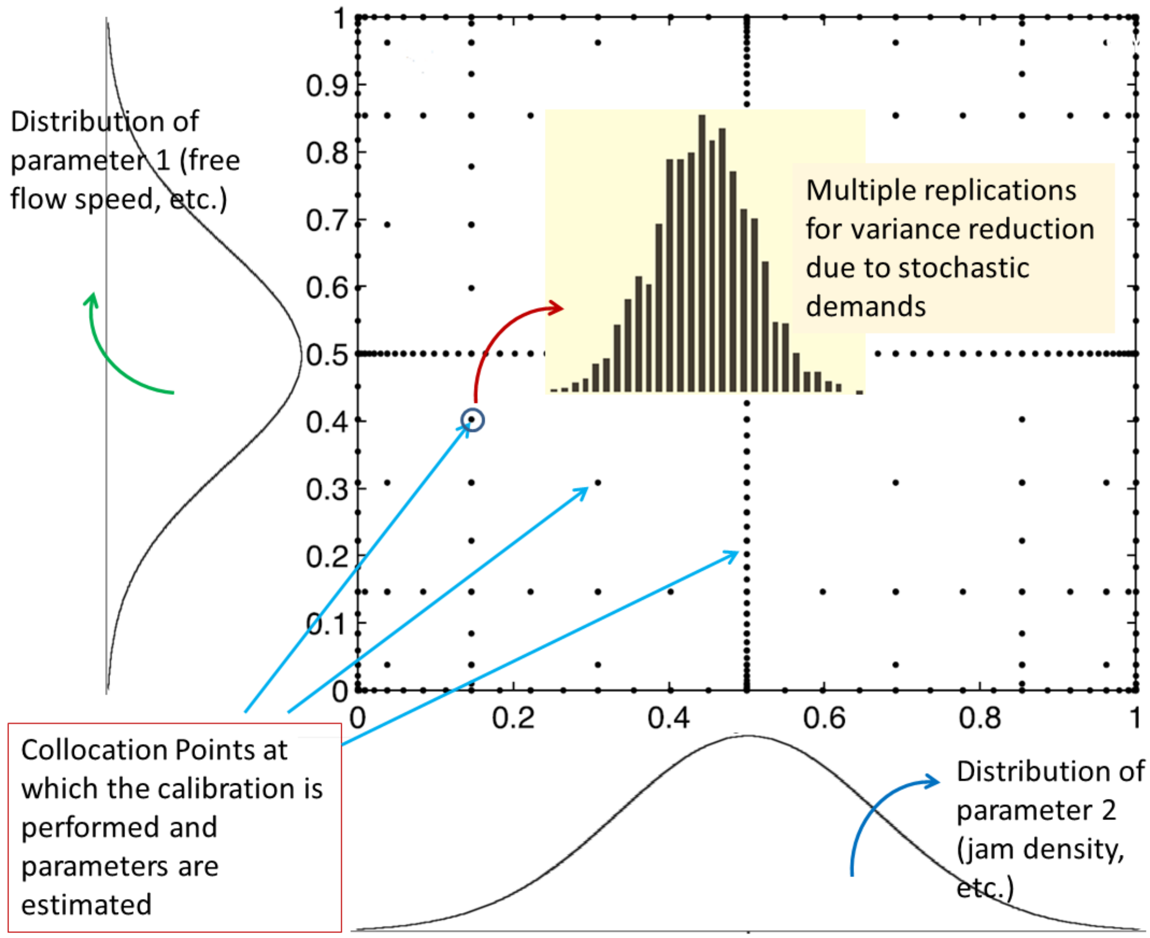

Figure 2 Example of approximation of stochastic space by collocation points

From each realization of the parameter set, using the demand distribution as an input, the simulation output distribution (e.g., flow or density distribution) is generated. This distribution is compared with the observed output distribution and using a test statistic (such as the test statistic from the KS test), the error is estimated. This error is used as an objective function and is minimized as part of the multi-objective parameter optimization (shown in equation (4)) using the simultaneous perturbation stochastic approximation (SPSA) algorithm. In our study the weights (w) used are the coefficients of variation of each output measure from the observed data.

$$\min_{\Theta_t} \sum_{i=1}^{N} \left\{ w_1 U_1(q_i^{Ob}, q_i^S(\Theta_t^k)) + w_2 U_2(\rho_i^{Ob}, \rho_i^S(\Theta_t^k)) \right\} \tag{4}$$

where,

q_i^{Ob}, q_i^S - observed and simulated flows at location i

ρ_i^{Ob}, ρ_i^S - observed and simulated densities for location i

Θ_t^k - parameter set for time period t and iteration k

w_1, w_2 - weights for the error measures

U_1, U_2 - functions representing the error in flow and density

The main advantages of using this methodology are the following:

1. Flexibility in applying to any type of traffic simulation (1^{st} order, 2^{nd} order, meso/microscopic, etc.),
2. Computationally more efficient than MC-type exhaustive sampling methods with effective interpolant to generate full distribution of simulation output.
3. Time consumed by the collocation approach can be further reduced by parallelizing the simulation under each condition,
4. Nested form of the algorithm is useful in refining the interpolant as and when there is new data available.

RESULTS

In order to illustrate the stochastic variation in traffic conditions, a three-lane section of the NJTPK turnpike at interchange 7 is chosen. Although, microscopic traffic simulation tools such as Paramics provide a detailed and relatively accurate platform for modeling, the model building, calibration and execution can be very time consuming. When studying the effects of various stochasticities, we are going to focus on a first order macroscopic traffic simulation model to model the traffic flow in the section. We discretize the time and space for the model using the cell transmission model. A schematic representation of the simulated section, the stochastic input and model parameters is shown in Figure 3.

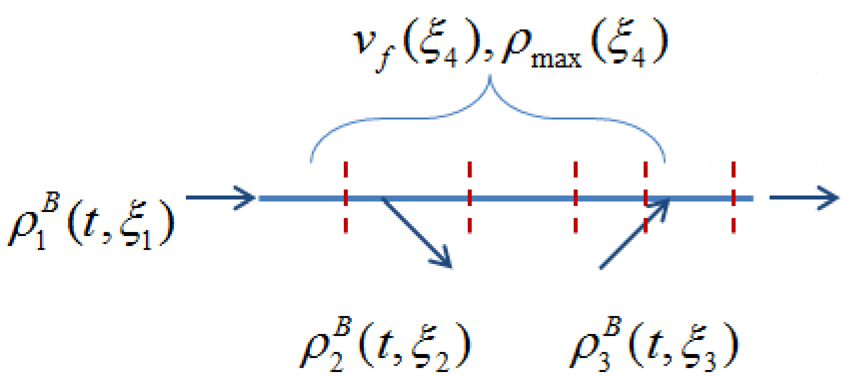

Figure 3 Schematic representation of the study section

The variation in demand at this section is captured using the ETC data for every 5 minutes between January 1, 2011 and August 31, 2011. The demand is divided into clusters using k-means algorithm. The optimal clusters for the demand distribution are shown in Figure 4. These data illustrate that primarily there are three clusters of data, weekday and weekend high and low.

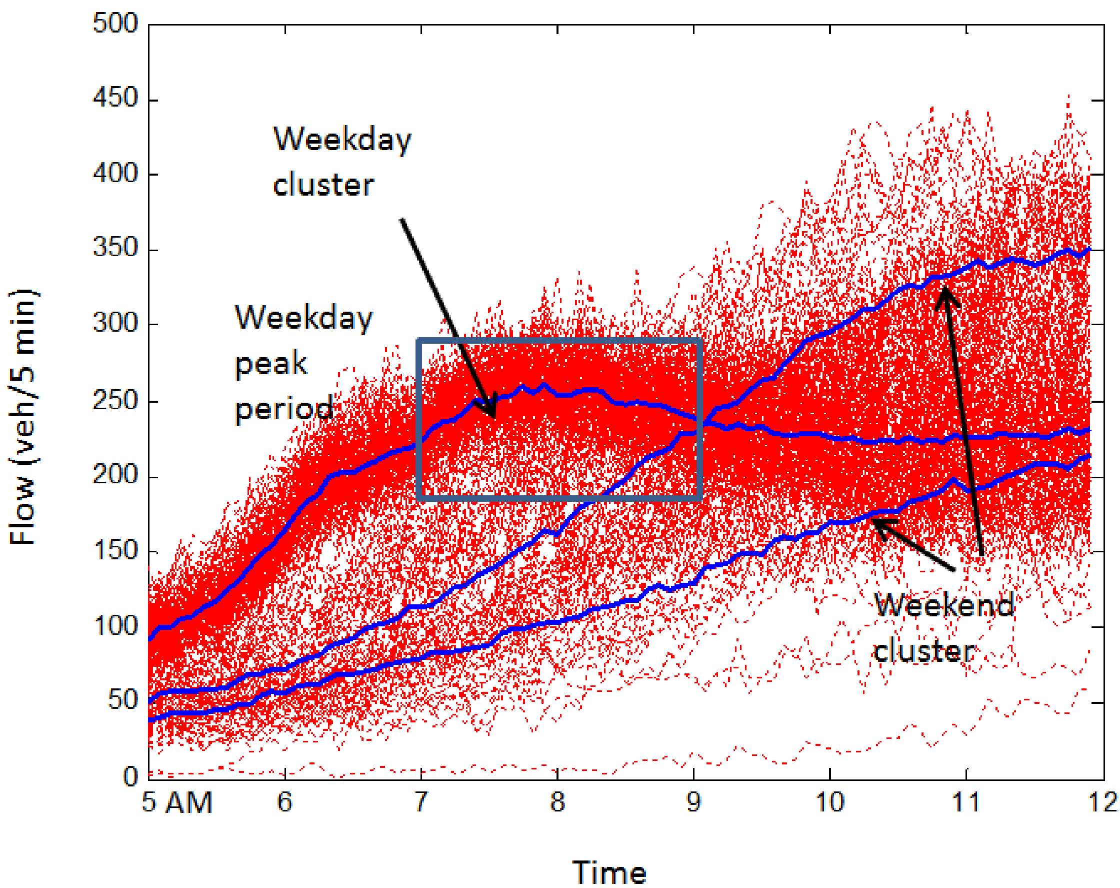

Figure 4 Clusters and distribution of demand for the mainline section at interchange 7 of NJTPK

Each line depicted (in Figure 4) by the mean of the cluster for each of the demand cluster. Using the actual demand values from each of the mean does not take into account the variability in traffic conditions and may not be useful. Thus for each cluster, the distribution of demand during each 5 minute time period is generated as illustrated in Figure 5. The simulation is performed for AM peak (7-9AM) and off-peak (10AM-12PM) period for weekday demand cluster and peak period (10AM-12PM) of the weekend demand cluster. It is to be noted that, especially during the peak period, the demand distribution for each sampling period (15min.) is different from the other period. Additionally, since this input parameter is demand as opposed to flow, each sampling period can be treated as independent of the other time period. So for simulating 7-9 AM with 15-minute demands, there will be eight separate demand distributions, each of which has to be treated as an independent stochastic dimension.

Traffic sensor data available between January 1, 2011 and August 31, 2011 for every 5 minutes is used to generate the speed and flow data required for calibration. The sensor data is separated into the same clusters for the same time periods as the demand clusters. Thus similar flow distribution data is generated.

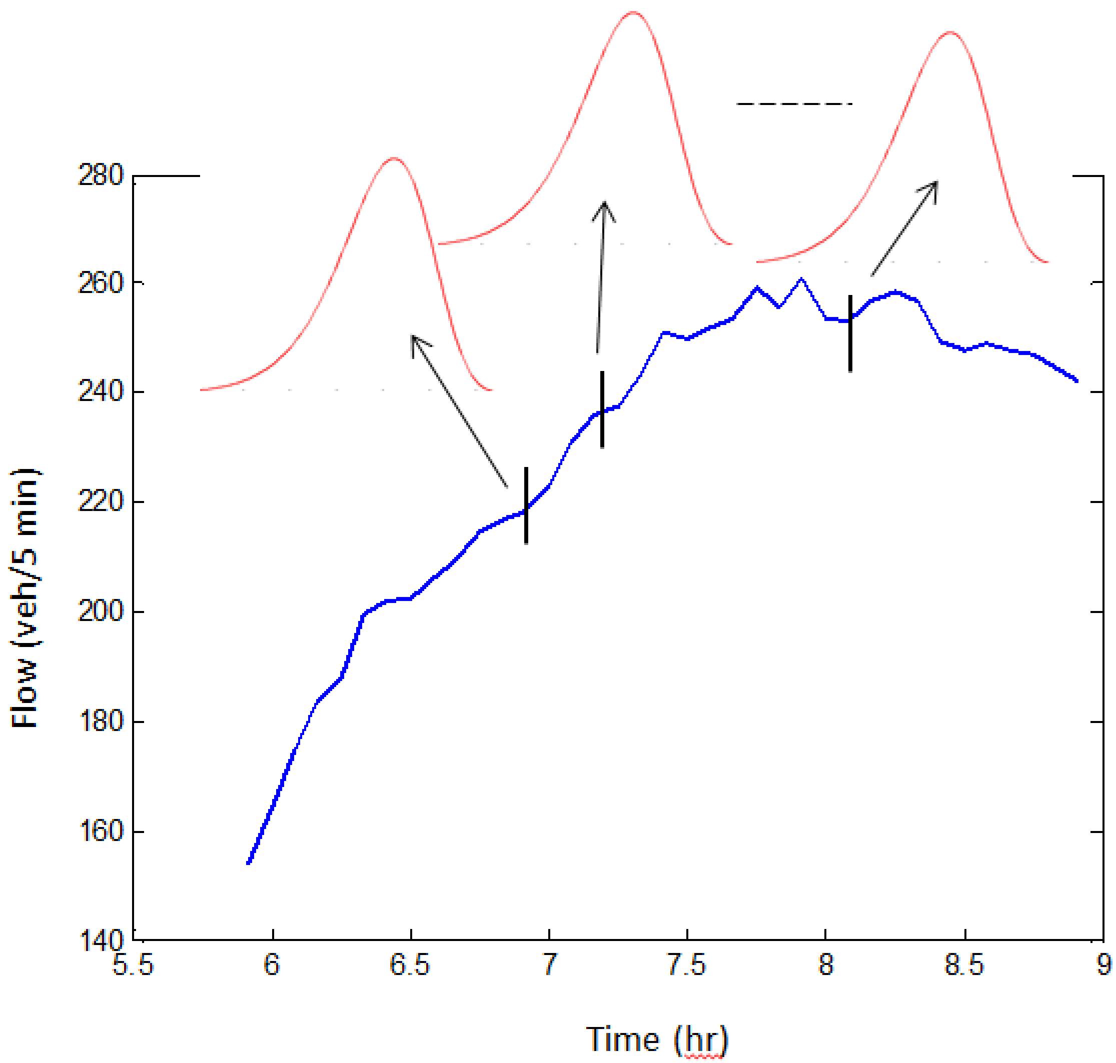

Figure 5 Illustration of demand distribution at various sampled times during AM peak period

For each of the three time periods, AM peak (7-9AM) and off-peak (10AM-12PM) period for weekday demand cluster and peak period (10AM-12PM) of the weekend demand cluster, we calibrate the simulation model. As mentioned in the methodology, the with the demand distribution as an input, for each realization of the parameter set, the simulation output flow distribution is generated. This distribution is compared with the observed flow distribution and using a test statistic (such as the test statistic from the KS test), the error is estimated. This error is used as an objective function and is minimized using the SPSA algorithm. The result of calibration is demonstrated using the comparison of simulated and observed flow.

For this study, the Clemshaw-Curtis grid (two-dimensional version of which can be seen in Figure 2) is the appropriate sparse grid to discretize the stochastic demand. The simulation is calibrated using the demand values at each of these grid points. The objective function for calibration is the test statistic used in the Kolmogorov-Smirnov test at 90% significance, maximum separation between two distributions. As mentioned in equation (3), a sparse grid interpolation is performed for the output of the simulation and a Smolyak interpolant is constructed. Distribution of simulated flows is obtained by repeated evaluation of the Smolyak interpolation function. The simulated flow distribution is compared to the observed distribution from the sensor data.

The comparison of observed and simulated flow distributions from the calibrated model for AM peak period is shown in Figure 6.

In order to compare the efficiency of the stochastic collocation approach, the distribution of simulated flow after model calibration is also generated using Monte Carlo sampling method. In order to achieve the flow distribution, the SC approach required 2433 evaluations for various stochastic demand combinations. However, using a MC-type sampling required 240,000 runs of the simulation model. The reason, as mentioned earlier, is due to the ability to construct efficient Smolyak interpolant that uses the simulation output from much fewer runs.

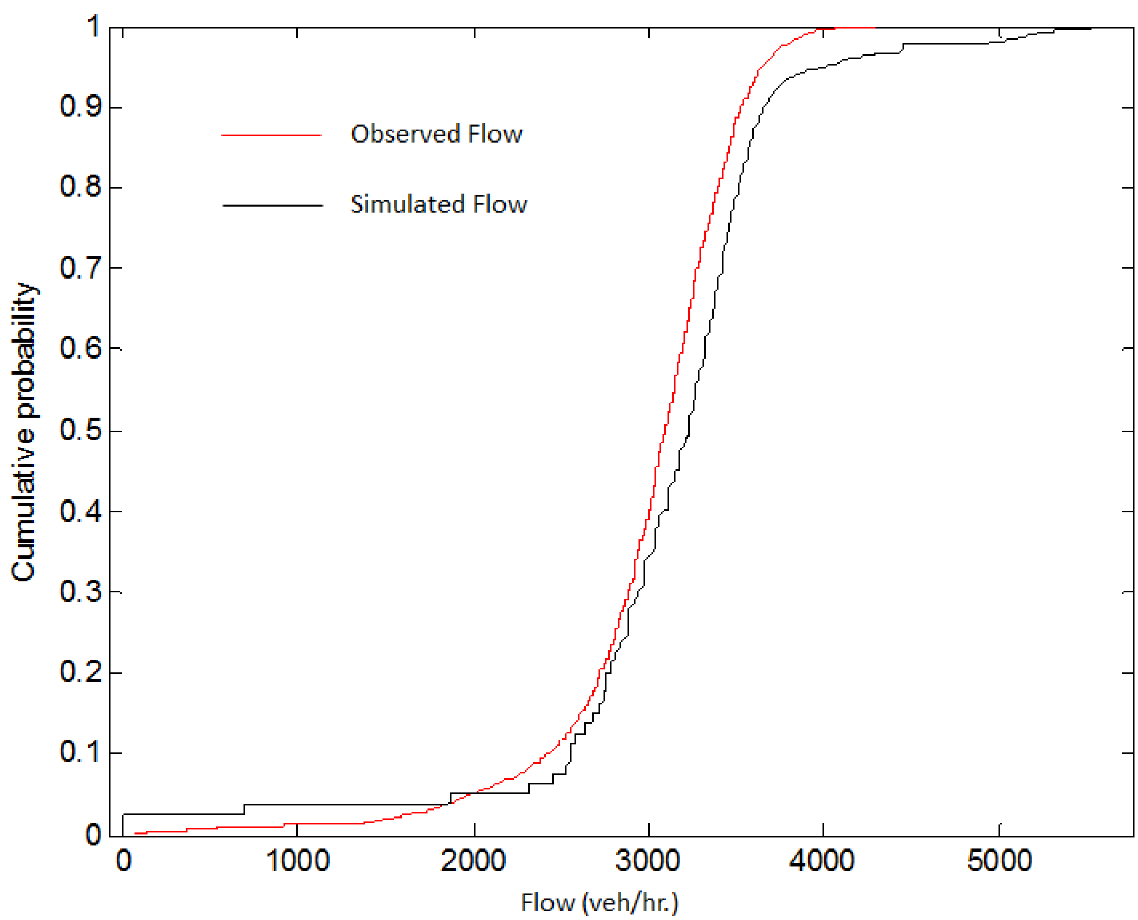

Figure 6 Comparison of observed and simulated link flow distributions during AM peak period

As the second case study, the flow distribution during the off-peak period is also generated using the sparse grid interpolation. The comparison of observed and simulated flow distribution from the calibrated model for weekday off-peak period is shown in Figure 7. It is observed that there was no flow breakdown during this time period. As in the first case study, the distribution of simulated flow after model calibration is also generated using Monte Carlo sampling method. In order to achieve the flow distribution, the SC approach required 441 evaluations for various stochastic demand combinations. However, when using a MC-type simulation required 5420 samples.

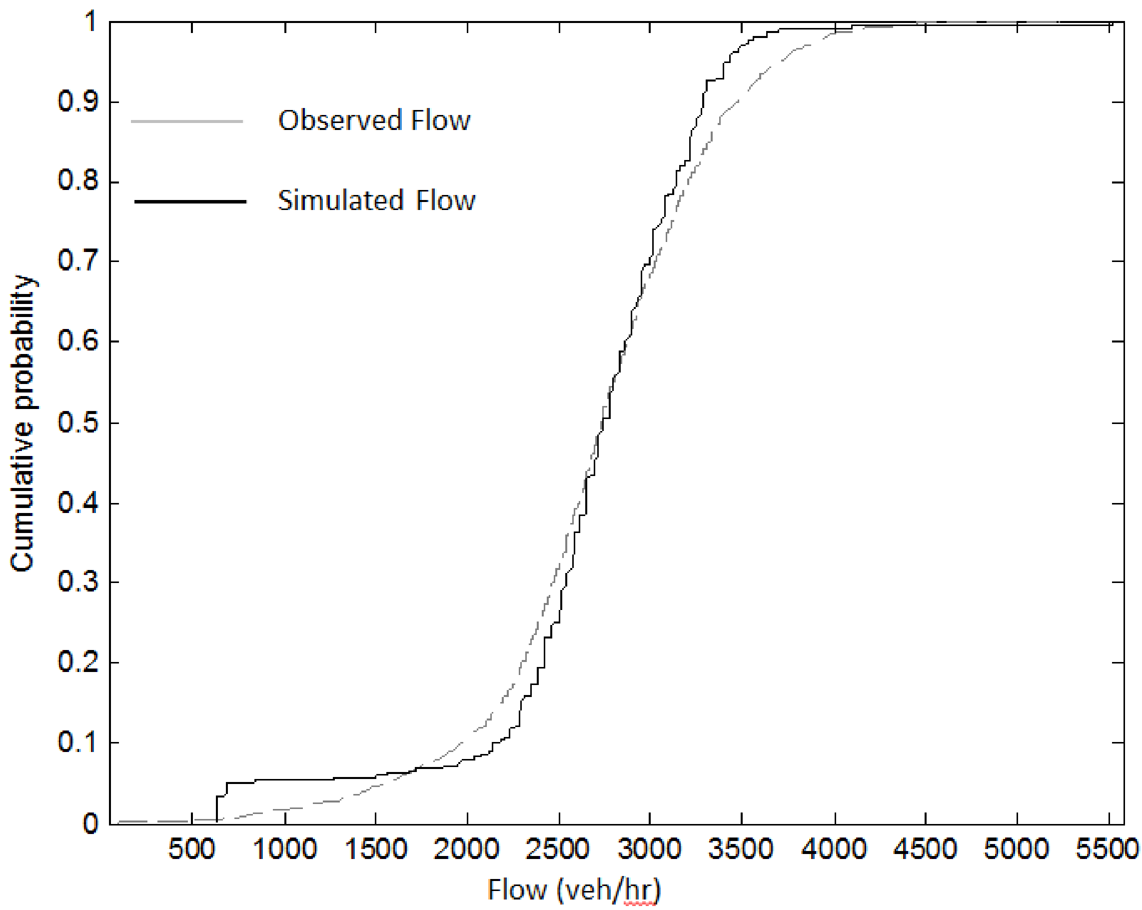

Figure 7 Comparison of observed and simulated link flow distributions during off-peak period

As the third case study, the flow distribution during the high and low demand weekend peak period is also generated using the sparse grid interpolation. The comparison of observed and simulated flow distribution for weekend peak period is shown in Figure 8. It is observed that there was no flow breakdown during this time period. Distribution of simulated flow is also generated using Monte Carlo sampling method. In order to achieve the flow distribution, the SC approach required 441 evaluations for various stochastic demand combinations. However, when using a MC-type simulation required 8000 samples.

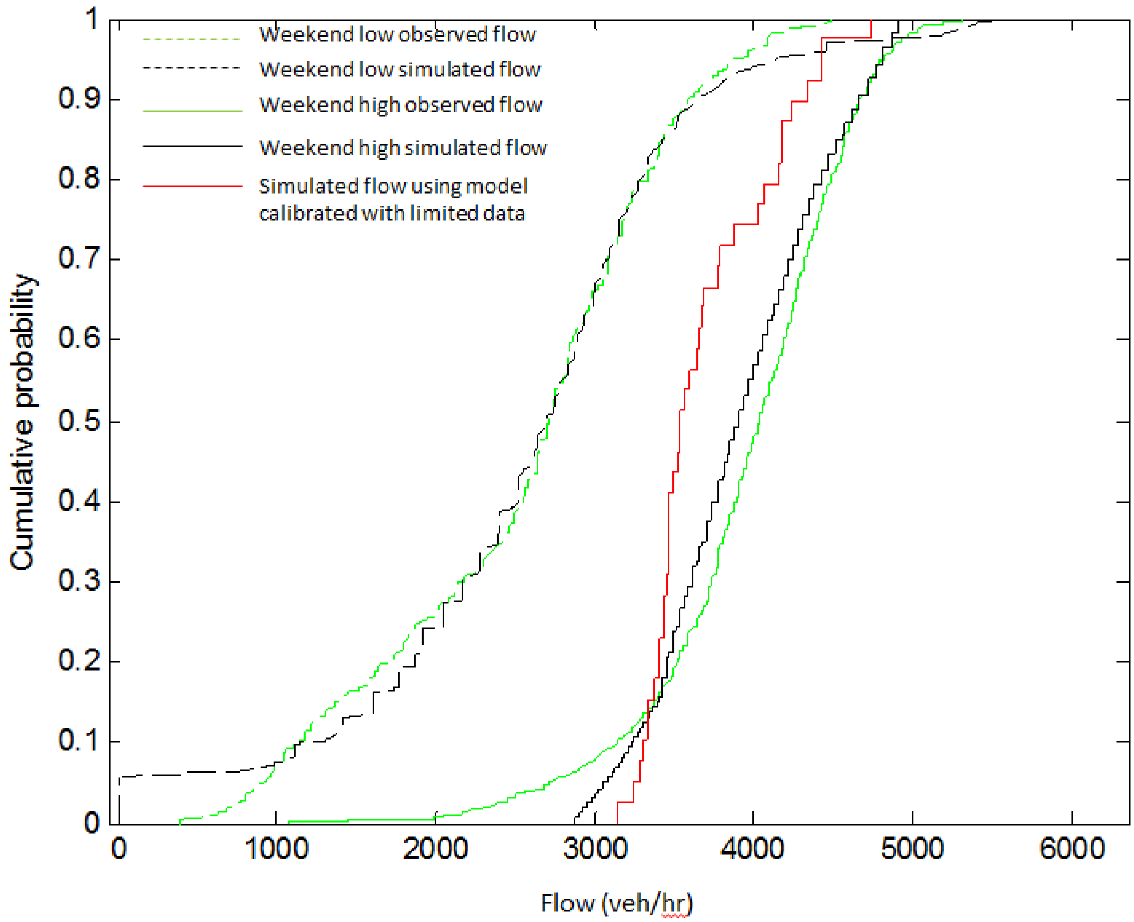

Figure 8 Comparison of observed and simulated link flow distributions for low and high clusters during weekend peak period

The motivation behind using data from a variety of conditions is to capture the stochasticity in traffic conditions. To illustrate the drawback of using limited data, we compare the distribution of flow for high and low weekend demands with the case where only three weekend days of flow and demand are used to calibrate the weekend model. The simulated flow distributions (shown in Figure 8) from limited data model does not match, not only the weekend model with high demand but also the weekend model with low demand. This illustrates the drawback in using limited data for model calibration and the importance of considering stochasticity in traffic conditions when calibrating traffic simulation models.

CONCLUSIONS AND FUTURE WORK

The predictions of well-calibrated traffic simulation model are robust if the predictions made for various conditions are accurate. Variations in traffic can arise due to many factors such as time of day, weather, accidents, etc. Calibration of simulation models for traffic conditions requires larger than traditionally adopted datasets capturing the stochasticity in traffic conditions. The advent of new technologies such as RFID sensors, smart cellphones and GPS devices provide extensive traffic data, an example of Big Data. Big Data provides a viable alternative to traditional traffic data sources in providing larger and richer datasets. Although Big Data provides greater variation in data, it poses a challenge in terms of computational effort. With the increase in number of stochastic factors, the numerical methods suffer from the curse of dimensionality. If traditional MC-type sampling is used, the computational effort required to simulate and calibrate traffic simulation models for various conditions could become intractable.

In this study, we use electronic toll collection data for which period from January to August, 2011 to capture various traffic conditions. Also, we propose a novel methodology to encapsulate stochasticity into macroscopic traffic simulation models and their calibration with much lower computational effort. We use stochastic collocation, a type of stochastic spectral method, to capture stochasticity in traffic. This method treats each stochastic factor as a separate dimension. Each dimension is discretized using a set of collocation points and an interpolant for the output is constructed using the simulation output at these points. In particular, we use the Smolyak sparse grid interpolation method due to the high number of stochastic dimensions.

The main advantages of using this methodology are the following:

1. Flexibility in applying to any type of traffic simulation (1^{st} order, 2^{nd} order, meso/microscopic, etc.),
2. Computationally more efficient than MC-type exhaustive sampling methods with effective interpolant,
3. Time consumed by the collocation approach can be further reduced by parallelizing the simulation under each condition,
4. Nested form of the algorithm is useful in refining the interpolant as and when there is new data available.

To demonstrate the usefulness of our methodology, we test it for a short on-ramp-off-ramp section of NJTPK in the vicinity of interchange 7. The variation in demand at this section is captured using the ETC data for every 5 minutes between January 1, 2011 and August 31, 2011. The demand is divided into clusters using k-means algorithm into weekday, weekend high and low clusters. The methodology is applied to calibrate the macroscopic first order traffic simulation model for AM peak (7-9AM) and off peak (10AM-12PM) period for weekday demand cluster and the peak period (10AM-12PM) for high and low weekend clusters. For calibrating the model, we use the test statistic from the KS test for flow distributions on the link as the objective function. This measure is minimized using the SPSA algorithm. We show that the comparison of simulated and observed flow distributions for the peak and off-peak periods match well. We also compare the simulated and observed flow distributions for weekend peak demand and found that they also match well. Also we illustrate the advantage of use to varied traffic conditions for model calibration by comparing the simulated flow distributions generated using models calibrated with large set of data and model calibrated using limited days of demand and flow data.

As a part of future work, the probability density function of flows and densities at various sections will be validated using a hold out dataset. Also we will extend this methodology to larger highway sections for larger freeway sections with stochastic flow density relationship.

REFERENCES

1. Li, J., Chen, Q., Wang, H., Ni, D., Analysis of LWR model with fundamental diagram subject to uncertainties. In: The 88th Transportation Research Board (TRB) Annual Meeting, Washington, DC, 2009 (Pre-print CD-ROM).
2. Ngoduy, D., in press. Multiclass first order model using stochastic fundamental diagrams. Transportmetrica. Vol. 7, No. 2, March 2011, pp. 111-125
3. Zhong, R., Sumalee, A., Stochastic cell transmission model for traffic network with demand and supply uncertainties. Proceedings of the Sixth International Conference of Traffic and Transportation Studies Congress 2008.
4. Jabbari, S.E. and Liu, H.X., A stochastic model of traffic flow: Theoretical foundations, Transportation Research Part B 46 (2012) 156–174
5. Lee J-B. and Ozbay, K., “New Calibration Methodology for Microscopic Traffic Simulation Using Enhanced Simultaneous Perturbation Stochastic Approximation Approach”, Transportation Research Record: Journal of the Transportation Research Board, No. 2124, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 233–240.
6. Kim, K. O., and Rilett, L. R., “Simplex-based Calibration of Traffic Micro-Simulation Models with Intelligent Transportation Systems Data”, Transportation Research Record: Journal of Transportation Research Board, No. 1855, 2003, pp.80-89.
7. Ma, T., B. Abdulhai, "Genetic Algorithm-Based Optimization Approach and Generic Tool for Calibrating Traffic Microscopic Simulation Parameters", Transportation Research Record: Journal of Transportation Research Board, No. 1800, 2002, pp.6-15.
8. Zhang, M., Ma, J., Singh, S.P., Chu, L., Developing Calibration Tools for Microscopic Traffic Simulation Final Report Part III: Global Calibration - O-D Estimation, Traffic Signal Enhancements and a Case Study, California PATH Research Report, UCB-ITS-PRR-2008-8, June 2008.
9. Yang, H., and Ozbay, K., Calibration of Micro-simulation models to account for Safety and Operation Factors for Traffic Conflict Risk Analysis, 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA.
10. Menneni, S., Sun, C., Vortisch, P., “Microsimulation Calibration Using Speed-Flow Relationships. Making NGSIM Data Usable for Studies on Traffic Flow Theory.” Transportation Research Records: Journal of Transportation Research Board, in printing, 2088, 2008, pp. 1-9.
11. Hourdakis, J., Michalopoulos, P. G., and Kottommannil, J., “A Practical Procedure for Calibrating Microscopic Traffic Simulation Models”, Transportation Research Record: Journal of Transportation Research Board, No. 1852, 2003, pp.130-139.
12. Jha, M., Gopalan, G., Garms, A., Mahanti, B.P., Toledo, T., and Ben-Akiva, M.E., “Development and Calibration of a Large-Scale Microscopic Traffic Simulation Model”, Transportation Research Record: Journal of Transportation Research Board, No. 1876, 2004, pp.121-131.

13. Toledo, T., Ben-Akiva, M. E., Darda, D., Jha, M., and Koutsopoulos, H. N., "Calibration of Microscopic Traffic Simulation Models with Aggregate Data", Transportation Research Record: Journal of Transportation Research Board, No.1876, 2004, pp.10-19.
14. Qin, X., and Mahmassani, H. S., "Adaptive Calibration of Dynamic Speed-Density Relations for Online Network Traffic Estimation and Prediction Applications", Transportation Research Record: Journal of Transportation Research Board, No.1876, 2004, pp.82-89.
15. Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H. N., and Wen, Y., "Calibration of Microscopic Traffic Simulation Models: Methods and Application", Transportation Research Record: Journal of Transportation Research Board, No. 1999, 2007, pp.198-207.
16. Li, J., Chen, Q., Wang, H., Ni, D., 2009. Analysis of LWR model with fundamental diagram subject to uncertainties. In: The 88th Transportation Research Board (TRB) Annual Meeting, Washington, DC (Pre-print CD-ROM).
17. Sumalee, A., R.X. Zhong, T.L. Pan, W.Y. Szeto. Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. Transportation Research Part B (2010).
18. Henclewood, D., Suh, W., Rodgers, M.O., Hunter, M., Statistical Calibration for Data-Driven Microscopic Simulation Model, 92nd Annual TRB Conference held Jan. 13 – 17, 2013, Washington, DC.
19. Punzo, V., Montanino, M., Ciuffo, B., Goodness of fit function in the frequency domain for robust calibration of microscopic traffic flow models, 92nd Annual TRB Conference held Jan. 13 – 17, 2013, Washington, DC.
20. Luttrell, T., W. Sampson, D. Ismart, and D. Matherly. Traffic Analysis Toolbox. Volume VII: Predicting Performance with Traffic Analysis Tools. Publication FHWA-HOP-08-055. FHWA, U.S. Department of Transportation, 2008.
21. Wunderlich, K., "Seattle 2020 Case Study, PRUEVIN Methodology", Mitretek Systems.
22. Ozbay, K., Mudigonda, S., Morgul, E., Yang, H., Bartin, B., "Big Data and the Calibration and Validation of Traffic Simulation Models", Presented at the 93rd Annual TRB Conference, Washington D.C., Jan. 10-15, 2014.
23. Loh, W-L. On Latin Hypercube Sampling, The Annals of Statistics, Vol. 24, No. 5 (Oct., 1996), pp. 2058-2080.
24. NJTA. http://www.state.nj.us/turnpike/. Accessed December 25, 2013.
25. D. Xiu, J.S. Hesthaven, High order collocation methods for the differential equation with random inputs. SIAM J. of Scientific Computing, 27 (2005), pp. 1118–1139
26. B. Ganapathysubramaniam, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems. JCP, 225 (2007), pp. 652–685
27. A. Klimke, Uncertainty Modeling using Fuzzy Arithmetic and Sparse Grids, Ph.D. thesis, Universitat Stuttgart, Shaker Verlag, Aachen, 2006.