Calibration of Microscopic Traffic Simulation Models

Methods and Application

Ramachandran Balakrishna, Constantinos Antoniou, Moshe Ben-Akiva, Haris N. Koutsopoulos, and Yang Wen

A mathematical framework and a solution approach are presented for the simultaneous calibration of the demand and supply parameters and inputs to microscopic traffic simulation models as well as a large-scale application emphasizing practical issues. Microscopic traffic simulation models provide detailed estimates of evolving network conditions by modeling time-varying demand patterns and individual drivers' detailed behavioral decisions. Such models are composed of elements that simulate different demand and supply processes and their complex interactions. Several model inputs (such as origin–destination flows) and parameters (car-following and lane-changing coefficients) must be specified before these simulation tools can be applied, and their values must be determined so that the simulation output accurately replicates the reality reflected in traffic measurements. A methodology is presented here for simultaneously estimating all microscopic simulation model parameters by using general traffic measurements. A large-scale case study for the calibration of the MITSimLab microscopic traffic simulation model by using the network of Lower Westchester County, New York, is employed to demonstrate the feasibility, application, and benefits of the proposed methodology.

Microscopic traffic simulation tools find wide-ranging applications in network design and evaluation, planning and analysis, and the evaluation of intelligent transportation systems. Such tools possess the flexibility to represent rich spatial and temporal demand patterns, model complex driving behavior phenomena and the interactions between individual vehicles, mimic drivers' travel behavior decisions such as route choice and response to information, and simulate the operations of traffic management strategies, technologies, and infrastructure. However, owing to the complex nature of their constituent model components with their dependence on a large number of inputs and parameters, their use requires that the various inputs be determined to best represent observed local traffic data. Model calibration, which matches simulator output to real measurements from the deployment site, is thus a vital step in the application of complex traffic simulation tools.

Several papers present frameworks for the calibration of microscopic traffic simulation models. Ben-Akiva et al. (*1*), for example, divide the process into two steps. First, individual model components are calibrated with disaggregate data such as vehicle trajectories and surveys. Such detailed data are well suited for the estimation of driver and travel behavior phenomena such as car following, lane changing, and route choice. This step is followed by the fine-tuning of all parameters so that the simulator's output matches aggregate data (such as time headways, speeds, and flows).

Disaggregate data, although ideal for microscopic model calibration, are extremely costly and difficult to collect. In most practical applications, only aggregate traffic measurements will be available. Model calibration in such cases must be performed by using aggregate data alone, so as to minimize the deviation between observed and simulated measurements. However, because of the large set of unknown parameters and the significant computational burden associated with large-scale traffic simulation runs, such calibration has largely been restricted to trial-and-error approaches (*2–4*) and simpler subproblems that estimate subsets of parameters while fixing the remaining variables at some predetermined values. Often, the focus is on only the set of origin–destination (O-D) flows to be assigned onto the network or just a few key behavioral parameters selected through past experience or a sensitivity analysis. For example, the O-D flows (the primary inputs to a simulator's demand models) may be estimated while the driving and travel behavior model parameters are fixed (*5–7*). Alternatively, a few coefficients in the car-following and lane-changing models (affecting supply-side processes and traffic dynamics) may be estimated while fixing the O-D flows at their best-known values (*5, 8–13*).

Aggregate calibration is based on the output of the simulation model, which is a result of the interaction among all the components of the simulation. Therefore, it is generally not possible to isolate the effects of individual models on various traffic measurements. For example, common O-D estimation methods require an assignment matrix or its equivalent as an input. The time-dependent O-D matrix $\boldsymbol{x}_h$ defines the number of trips between each O-D pair departing the respective origin during time interval h. The assignment matrix maps these O-D flows to link flows (in both current and future intervals). Such a mapping is a complex function of drivers' route choice and driving behavior decisions, as well as the travel times on the network, and has to be generated by the simulation itself. Similarly, the most critical element in route choice models are the link travel times, which would be affected by the simulated flows, which in turn are a

R. Balakrishna, Caliper Corporation, 1172 Beacon Street, Newton, MA 02461. C. Antoniou, Department of Transportation Planning and Engineering, National Technical University of Athens, 5 Iroon Polytechniou St., Zografou 15773, Athens, Greece. M. Ben-Akiva and Y. Wen, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-181, Cambridge, MA 02139. H. N. Koutsopoulos, Northeastern University, 437 Snell Engineering Center, Boston, MA 02115. Corresponding author: R. Balakrishna, rama@caliper.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 1999, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 198–207.
DOI: 10.3141/1999-21

function of the O-D flow matrices. Effectively, the values of each of these parameters influence the values of the other parameters that are being calibrated. The calibration approach therefore needs to take these interactions into consideration. Often, this step is done by iteratively calibrating the different components—that is, the behavior models, route choice model, and O-D matrices—while keeping the other parameters in the simulation model fixed at each step.

Ben-Akiva et al. (*1*) provide an iterative approach that reestimates each parameter subset in turn until convergence. However, the computational effort involved is usually found to be too great to be practical. Further, an approach that simultaneously solves for all variables at once is expected to yield the most accurate estimates and is thus desirable.

Previous work (*14–17*) has clearly demonstrated the benefits of calibrating O-D flows and all other simulation parameters together in the context of dynamic traffic assignment (DTA) systems. Here the methodology is adapted for the simultaneous calibration of all demand and supply models within a microscopic traffic simulation model using aggregate, time-varying traffic measurements. The method accurately captures the complex relationships between various model parameters and data without using any linear approximations such as the assignment matrix. The approach also provides the flexibility to use any general traffic data to estimate all parameters and facilitates the simultaneous estimation of O-D flows from multiple time intervals. This latter capability is a requirement on large and congested networks (when a majority of the counts measured in a particular interval are caused by O-D flows departing in previous intervals), yet computational limitations restrict a similar approach based on assignment matrices.

The detailed nature of microscopic traffic simulation models generally results in higher running times and greater stochasticity than in more aggregate (mesoscopic or macroscopic) models. The calibration of such models has therefore typically been performed by focusing on a small subset of parameters and a data set suited to that subset. For example, car-following and lane-changing model parameters are generally estimated with detailed trajectory data available for just a few locations on the network, whereas O-D flows are estimated with count data. An efficient methodology and solution approach to simultaneously estimate all parameters of interest using readily available sensor data, and in a reasonable time, will therefore be invaluable to both practitioners and researchers. The main contribution of this study is a demonstration of the feasibility and benefits of such an approach when applied to a large-scale traffic microsimulation model.

METHODOLOGY

A systematic calibration methodology [reported elsewhere (*14–17*)] is presented that possesses several advantages over existing methods. The time period of interest is divided into intervals $h = 1, 2, \ldots, H$, $\mathbf{x}_h$ denotes the vector of O-D flows departing their respective origins during time interval h, and $\boldsymbol{\beta}_h$ is the vector of simulation model parameters that must be calibrated together with the O-D flows. The calibration problem may then be formulated mathematically in the following optimization framework:

Minimize

$$z(\mathbf{x}_1, \ldots, \mathbf{x}_H, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_H) = \sum_{h=1}^{H} \left[z_1(\mathbf{M}_h, \mathbf{M}_h) + z_2(\mathbf{x}_h, \mathbf{x}_h^a) + z_3(\boldsymbol{\beta}_h, \boldsymbol{\beta}_h^a) \right] \tag{1}$$

subject to the following constraints:

$$\left.\begin{aligned} &\mathbf{M}_h = f(\mathbf{x}_1, \ldots, \mathbf{x}_h, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_h, G_1, \ldots, G_h) \\ &l_h^x \le \mathbf{x}_h \le u_h^x \\ &l_h^\beta \le \boldsymbol{\beta}_h \le u_h^\beta \end{aligned}\right\} \forall h \in \{1, 2, \ldots, H\} \tag{2}$$

where

$\mathbf{M}_h$, $\mathbf{M}_h$ = observed and fitted sensor measurements for interval h,
$\mathbf{x}_h^a$, $\boldsymbol{\beta}_h^a$ = a priori values corresponding to $\mathbf{x}_h$ and $\boldsymbol{\beta}_h$, and
z_1, z_2, z_3 = goodness-of-fit functions.

The simulation model, $f()$, is a function of the O-D flows, the network G_h, and model parameters $\boldsymbol{\beta}_h$, up to interval h. The terms l_h^x, l_h^β, u_h^x, and u_h^β represent lower and upper bounds on the O-D flows and model parameters.

A priori parameter values, when available, can be used to ensure reasonable calibrated estimates. Such values can be based on the modeler's experience and judgment, may be obtained from past studies, or can be transferred appropriately from similar studies at other locations. Individual parameters may also be subject to lower and upper bounds. For example, O-D flow variables must be nonnegative, and the travel time coefficient in a route choice model can be constrained to be negative.

The direct use of the simulation model's output accurately captures the complex nonlinear dependencies between the variables ($\mathbf{x}_h$ and $\boldsymbol{\beta}_h$) and the data ($\mathbf{M}_h$). This approach is superior to existing methods that approximate the complex relationships between the calibration variables and traffic measurements. In traditional O-D estimation approaches, for example, the mapping between the unknown O-D flows and the observed link counts is replaced by a linear measurement equation:

$$\mathbf{y}_h = \sum_{p=h-p'}^{h} \mathbf{a}_h^p \mathbf{x}_p + \mathbf{v}_h \tag{3}$$

where

$\mathbf{y}_h$ = vector of link sensor counts for interval h,
$\mathbf{a}_h^p$ = assignment matrix,
$\mathbf{v}_h$ = error vector, and
p' = number of intervals spanning longest trip.

The resulting problem is typically solved with a generalized least-squares approach (*18, 19*).

Assignment matrices may be computed within the traffic simulation model by tracking vehicle trajectories or by measuring the simulated travel times from each demand origin node to every sensor location (*20*). However, since the simulated assignment matrices depend on the starting demand profile, this O-D estimation formulation requires the solution of a fixed-point problem [see paper by Cascetta and Postorino (*21*), for example]. In the approach proposed here, the assignment mappings are captured directly by treating the simulation model as a black box.

The proposed approach introduces the flexibility to easily incorporate any general traffic measurements, beyond the standard link counts. For example, speeds and densities, also recorded by loop detectors, may be added to the data set. More sophisticated data such as point-to-point counts and travel times from automatic vehicle identification (AVI) systems and probe vehicles may also be incorporated. Further, all model inputs and parameters of interest may be calibrated simultaneously, fully using all information contained in the available measurements and without the need to iterate between various parameter subsets.

SOLUTION APPROACHES

Equations 1 and 2 together represent a complex, nonlinear, nonanalytical optimization problem, owing to the use of a sophisticated simulator to obtain the fitted measurements. The high degree of nonlinearity introduces an objective function with potentially many local optima. The local minimum closest to the starting solution may thus be far from a global optimum. The nonanalytical nature is attributed to the lack of an explicit form for $f(\ )$ as a function of the calibration variables. Consequently, classical algorithms that rely on the knowledge of exact analytical gradients are not suitable for calibration. Methods that work directly with function values are more appropriate in the context of the current application. When the simulator is stochastic (as is often the case), the optimization problem must also account for the inherent noise in model outputs. Further, the problem is large in scale, with the number of O-D pairs and time intervals increasing rapidly with the size of the network and the desired temporal modeling resolution. Consequently, appropriate solution algorithms must be able to

- Perform a global search by overcoming local hills and valleys,
- Work without analytical derivatives (which would generally be unavailable), and
- Converge in a reasonable time frame that does not grow rapidly with problem size.

The proposed calibration framework can be solved with several alternative approaches. Population-based simulation optimization methods such as Box–complex (*22*) and stable noisy optimization by branch and fit (SNOBFIT) (*23*) are ideally suited for such applications, since they work directly with function evaluations and do not require the calculation of derivatives. Such methods maintain a set of points (and their function values) and systematically traverse the feasible search space to potentially identify global optima. Although their capabilities have been demonstrated in other areas, the application of these algorithms to complex and large-scale transportation problems is limited.

The calibration problem has been established as a large-scale iterative problem optimizing an output from a stochastic simulator. Therefore stochastic approximation techniques can also be used in the calibration context. The finite difference stochastic approximation (FDSA) method computes a search direction by comparing the objective function after perturbing each variable individually. Although FDSA is an option with proven performance, it is associated with considerable computational complexity: each gradient calculation involves at least $n + 1$ function evaluations, where n is the number of parameters to be calibrated.

An optimization algorithm developed by Spall (*24; 25,* pp. 529–542) specifically addresses stochastic problems. Simultaneous perturbation stochastic approximation (SPSA) provides significant computational improvements over FDSA by perturbing all variables at once. SPSA requires only two computations of the objective function at a given iteration for estimating the gradient vector, irrespective of the number n of parameters to be calibrated.

Although FDSA is expected to give a more accurate and reliable estimate of the gradient vector at each iteration, the associated computational cost is high. Studies have shown that SPSA and FDSA require a comparable number of iterations to reach the global optimum. The computational savings of SPSA per iteration thus result in a significantly more efficient algorithm for large-scale applications. Spall shows, through several standard stochastic problems, that SPSA outperforms FDSA in terms of overall convergence speed. The performance of SPSA for the calibration of large-scale traffic simulation models has also been demonstrated in studies by Balakrishna (*14*) and by Balakrishna et al. (*15, 16*); more details about the algorithm are given therein.

A case study comparing genetic algorithms (GAs) and SPSA for calibrating microscopic models is presented by Ma et al. (*13*). However, their application is limited to a 2-mi freeway corridor over a 1-h study period. A small set (up to 15) of supply-side parameters is calibrated assuming that O-D demand and other model parameters are calibrated elsewhere. The authors report very long running times for the GA despite the small problem instance, which can affect its scalability to large networks. Further, the simulation output generated with SPSA-calibrated parameters shows flow–occupancy values lower than the observed values.

CASE STUDY

The presented methodology is applied to the calibration of the MITSimLab microscopic traffic simulator (*26, 27*). The case study is based on the freeway and parkway network of Lower Westchester County (LWC), New York. Drivers there experience heavy traffic conditions, especially during commute periods. The main arteries in the network include the New York State Thruway (I-87), New England Thruway (I-95), Cross Westchester Expressway (I-287), Cross County Parkway, Hutchinson River Parkway, Sprain Brook Parkway, Saw Mill River Parkway, Bronx River Parkway, and Taconic State Parkway. Four adjoining arterials (Tuckahoe Road, Ardsley Road, Hartsdale Road, and Weaver Street) and Routes 9, 22, 100, and 119 provide alternate routes (Figure 1).

Truck traffic is prohibited from parkways. Given the significant truck percentage in the network traffic, it was decided that passenger cars and trucks should be treated independently; that is, multiclass calibration would be needed.

The network representation of the study area comprises 1,767 directed links, further subdivided into 2,564 segments that capture changing link characteristics. The data that were available for the calibration process included counts from 33 sensors and an all-day static, planning-level, O-D matrix, provided by the New York State Department of Transportation. Dynamic O-D profiles, derived from this static matrix in a previous study (*28*), were used as starting values. Disaggregate data from toll plazas were also available, consisting of individual vehicle observations recorded at the toll plazas. These data therefore contained records of each vehicle's type, which were used for the estimation of the vehicle mix (and subsequently employed in the development of multiclass O-D matrices). There are 482 O-D pairs in the network. The sensor locations are shown in Figure 1.

Data Processing and Analysis

The available surveillance data are limited and include counts at several locations in the network. Because of the different technologies, sensor setup, and data-collection infrastructure, data are available at different levels of aggregation. Data from most sensors are available

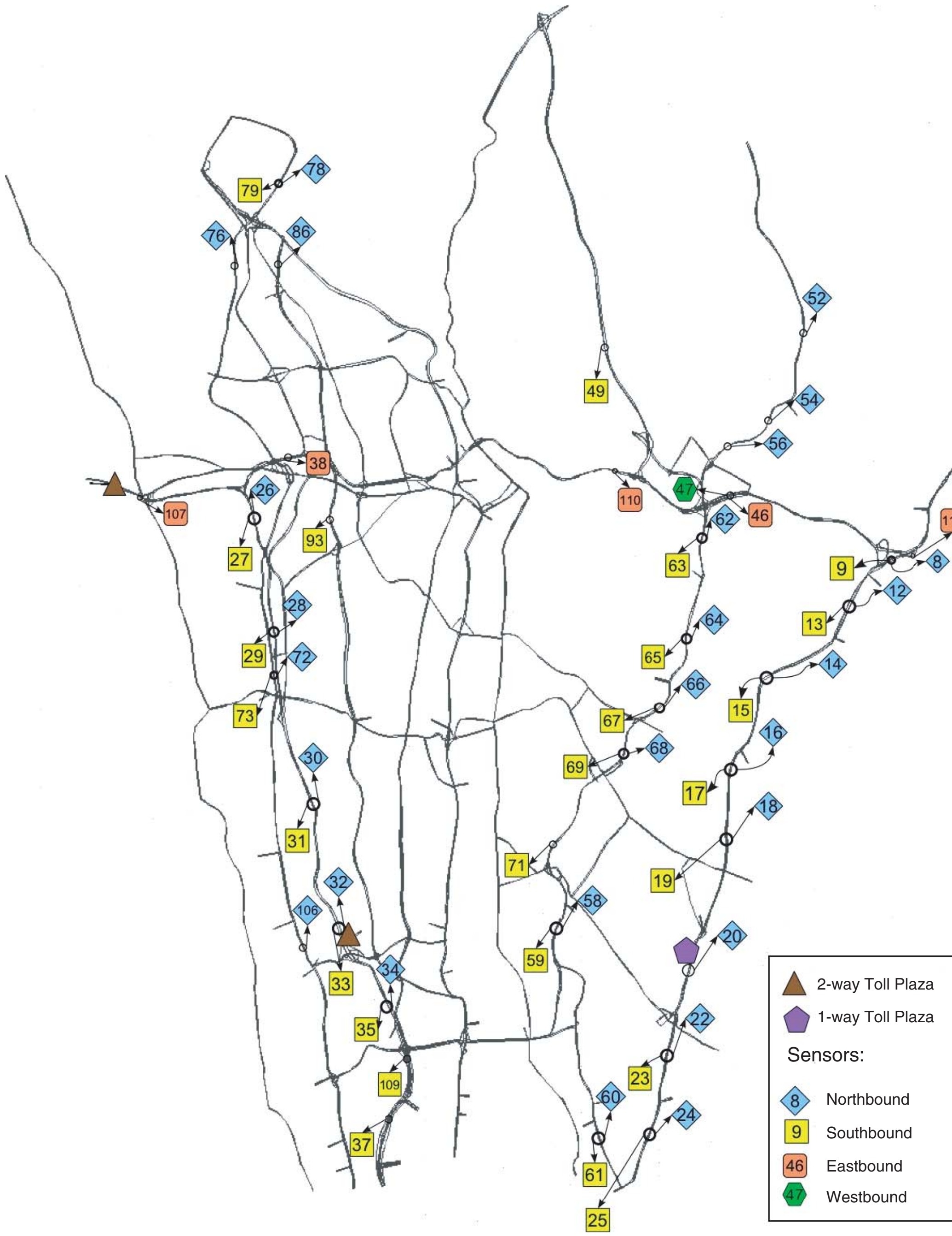

FIGURE 1 Sensor locations used for MITSimLab calibration.

in hourly intervals, whereas some sensors' data are available in 15-min intervals. Because of these limitations, extensive data processing and analysis was undertaken, consisting of several steps. Figure 2 summarizes the data preparation approach. Hourly data were first disaggregated and then smoothed. The data from individual days were analyzed and weekdays were pooled together and separated from weekend data, since a clear distinction was observed. Vehicle mix information was incorporated in order to separate the data into two classes: passenger cars and heavy vehicles.

The disaggregation process involved breaking down the hourly counts into 15-min counts and smoothing the results in order to ensure continuity. Because of different traffic patterns, sensors were grouped geographically, and smoothing was performed for each group. For each 15-min interval, the Euclidean norm of the vector of the counts in each sensor group was calculated. These norms were then fed into a seven-point Savitzky–Golay filter with degree 2, which generated smoothed norms (*29*). This approach preserves the features of the distribution such as relative maxima, minima, and width. The smoothed norm of each interval was then divided by the original norm to obtain the corresponding smoothing coefficient. Smoothed counts were obtained by multiplying the sensor counts in the same interval with this coefficient. Figure 3 (*top*) shows the sensor count data for one

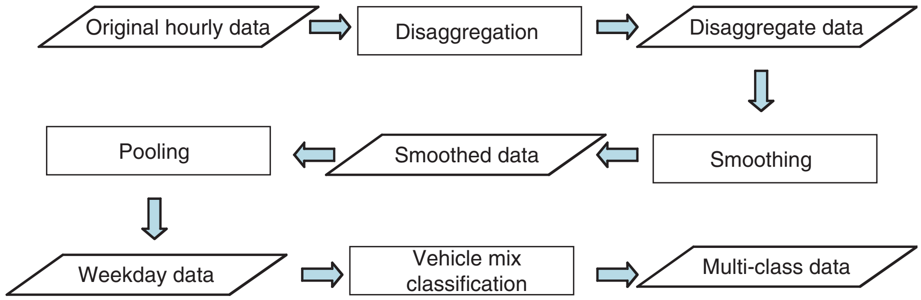

FIGURE 2 Data processing methodology.

of the sensors in the original disaggregated form (3*a*), as well as the result after smoothing (3*b*). It was verified that the total flow per hour was preserved after smoothing. Data from the toll plazas were of a much finer interval and were aggregated into 15-min counts.

It was important to segregate the sensor counts on weekdays from those on weekends or other holidays, since the prevalent demand patterns over these two kinds of days can be very different. Figure 3*c* and *d* demonstrates that the weekdays show similar traffic patterns. Hence, all weekday data were combined for calibrating a typical weekday.

The LWC network consists of predominantly freeway and parkway traffic that contains automobile commuters and heavy vehicles (trucks and commercial vehicles). Vehicle classification counts (the proportion of each vehicle type) are important inputs to O-D estimation, since the route choice process for various vehicle types could differ. For example, heavy vehicles are not permitted to use the parkways. The availability and implementation of this information are critical for O-D estimation, or the model would otherwise not reflect actual field situations: if all vehicle types on the network followed a similar route choice process, it would become theoretically impossible to match the observed sensor counts. Hence, the ratio of commercial traffic to passenger-car traffic must be specified for each O-D pair and can be used to form two separate O-D matrices (one for each of the two vehicle types). These O-D matrices can

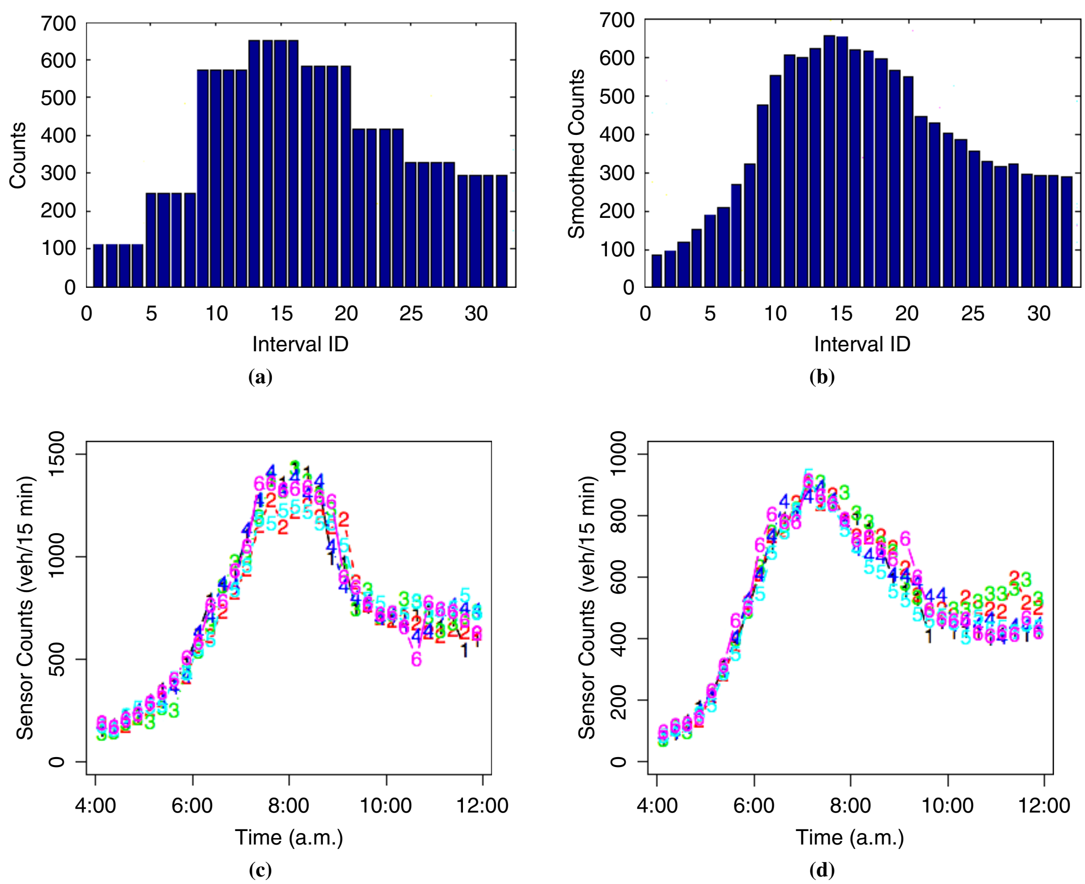

FIGURE 3 Sensor count smoothing sample for (*a*) disaggregated form and (*b*) after smoothing; indicative weekday sensor count distribution for (*c*) sensor E0115N and (*d*) sensor M0036S.

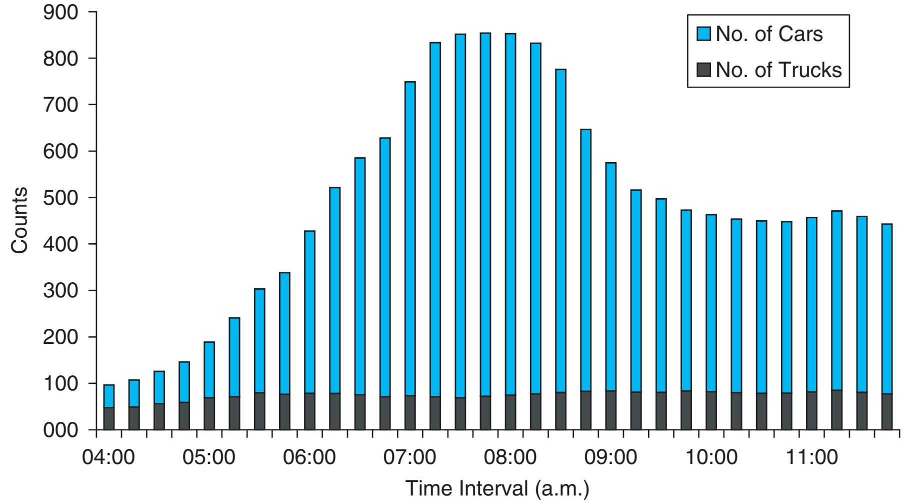

FIGURE 4 Vehicle mix: number of trucks versus number of cars by time interval.

then be loaded onto the network, with different individual route choice models.

Data from toll plazas were used to assign vehicles to two classes (Figure 4). The toll plazas, though few in number compared with the total number of sensors in the network, were the only source of data on vehicle classification counts. There were no other plausible ways of obtaining reliable classification counts by O-D pair. The average heavy-vehicle proportion among all toll sensors for each time interval was therefore used as an approximation of the vehicle mix ratio for every O-D pair for that corresponding time interval.

Measures of Goodness of Fit

The following statistics were used (*30–33*):

- Normalized root-mean-square error (RMSN) (*32, 33*);
- Root-mean-square percent error (RMSPE) (*34*);
- Mean percent error (MPE) (*34*);
- Theil's *U*-coefficient, with its bias, variance, and covariance components (*35*); and
- The GEH statistic, used in traffic engineering, traffic forecasting, and traffic modeling to compare sets of traffic volumes (*36*).

Multiple statistics are used to capture different aspects of the obtained results. For example, two sets of values are considered: 5 and 8, and 50 and 72. A statistic measuring the percent change in these two sets of values would suggest that a bigger change has been obtained in the first case (60% versus 44%). However, a statistic focusing on the actual numerical increase would suggest that a bigger change is obtained in the second case (22 versus 3).

The RMSN and RMSPE quantify the overall error of the simulator. These measures penalize large errors at a higher rate than small errors. RMSN is calculated as follows:

$$\text{RMSN} = \frac{\sqrt{N\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2}}{\sum_{n=1}^{N} Y_n^o} \tag{4}$$

where

N = number of observations,
Y_n^o = observation, and
Y_n^s = simulated value at time n.

RMSPE is calculated as follows:

$$\text{RMSPE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\frac{Y_n^s - Y_n^o}{Y_n^o}\right]^2} \tag{5}$$

The MPE statistic indicates the existence of systematic under- or overprediction in the simulated measurements and is calculated as follows:

$$\text{MPE} = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{Y_n^s - Y_n^o}{Y_n^o}\right] \tag{6}$$

Percent error measures are often preferred to their absolute error counterparts because they provide information on the magnitude of the errors relative to the average measurement. Another measure that provides information on the relative error is Theil's inequality coefficient:

$$U = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s\right)^2} + \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^o\right)^2}} \tag{7}$$

U is bounded between zero and 1 ($U = 0$ implies perfect fit between observed and simulated measurements). Theil's inequality coefficient may be decomposed into three proportions of inequality: the bias (U^M), the variance (U^S), and the covariance (U^C) proportions:

$$U^M = \frac{\left(\bar{Y}^s - \bar{Y}^o\right)^2}{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2} \tag{8}$$

$$U^S = \frac{\left(s^s - s^o\right)^2}{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2} \tag{9}$$

$$U^C = \frac{2\left(1-\rho\right)s^s s^o}{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2} \tag{10}$$

where ρ is the correlation between the two sets of measurements, and s^s and s^o are the standard deviations of the average simulated and observed measurements, respectively. By definition, the three proportions sum to 1. The bias proportion reflects the systematic error. The variance proportion indicates how well the simulation model is able to replicate the variability in the observed data. These two proportions should be kept as close to zero as possible. The covariance proportion measures the remaining error and therefore should be close to 1. Since the various measurements are taken from nonstationary processes, the proportions can only be viewed as rough indicators to the sources of error.

The statistics listed, while widely used, are sometimes perceived to be ineffective if the network consists of several roadway functional classes. The GEH statistic overcomes this potential problem by computing percent errors with respect to the mean value of the observed and the simulated counts. The formula for the GEH statistic is (*36*)

$$\text{GEH} = \sqrt{\frac{2\ \left(Y_n^s - Y_n^o\right)^2}{\left(Y_n^s + Y_n^o\right)}} \tag{11}$$

where Y_n^s is the simulated traffic volume and Y_n^o is the observed traffic count. GEH values below 5 are considered a good match between estimated and observed counts. According to FHWA guidelines (*36*), at least 85% of the observed links in a traffic model should have a GEH less than 5.0. If the GEH is greater than 10.0, there is a probability of error or errors with either the travel demand model or the data.

CALIBRATION RESULTS

Owing to the limited amount of data available, the methodology presented earlier was applied for the calibration of time-dependent O-D matrices. Flows between 482 O-D pairs were estimated for each 15-min departure time interval. The SPSA algorithm was selected for its proven performance and computational properties in large-scale problems (*14–16*). For the driving-behavior model parameters, values obtained in earlier studies (*1*) were used. Sensitivity analyses of these parameters confirmed that the selected values were appropriate for this network. A dynamic O-D matrix calibrated in another study (*28*) in the same network was used as a seed. The O-D calibration estimated a single flow per time interval for each O-D pair, which was further decomposed into two components (passenger cars and trucks) before being loaded onto MITSimLab. Time-dependent truck percentages from the observed toll-plaza data were used for this purpose.

The results for each 15-min interval in the 6:00 to 9:00 a.m. analysis period are summarized in Table 1. The difference between the counts obtained by loading the seed O-D matrices to MITSimLab (without performing any calibration iterations) and the smoothed observed counts are reported as the before case in this table. The after case corresponds to the difference between the counts obtained by loading the final calibrated O-D matrices to MITSimLab from the smoothed observed counts. The reported percent change between the fit of the simulated counts by using the seed matrix and the simulated counts by using the calibrated matrix is meant to demonstrate that the calibration process performs as expected and has the capability to steer the seed O-D matrix toward the values that will produce simulated counts that will best match the observed measurements. RMSN and RMSPE capture the deviation of the simulated counts from their observed counterparts in a manner that penalizes higher errors more than smaller errors. The percent change in terms of RMSN ranges between 25% and 71% for the various 15-min time intervals, with overall RMSN improvement equal to 51%. Similarly, the percent change in terms of RMSPE ranges between 22% and 65%, with an overall improvement of 47%.

Besides capturing the deviation around the observed values, MPE provides a measure of the direction of these errors, namely, whether there is a distinct bias toward under- or overestimation. A clear negative bias was evident in the results obtained by the starting values, which has been effectively eliminated as a result of the calibration. In absolute terms, the MPE value has been reduced between 55% and 99% for the 15-min intervals, whereas for the entire calibration period the improvement is on the order of 82%.

Theil's U should be as close to zero as possible, and indeed the calibration procedure decreases it by 57% overall for the entire calibration period. Improvement in the 15-min intervals ranges between 29% and 75%. The first two components of the coefficient are presented; the third one can be obtained as $U^C = 1 - U^M - U^S$. The overall improvement of the bias component (U^M) is 75% for the entire time period, and its improvement in the 15-min periods ranges between 36% and 100%. The overall improvement of the variance component (U^S) is 27%. The improvement in all but two 15-min periods ranges between 35% and 100%. However, it should be noted that for two time periods (6:15 to 6:30 and 7:15 to 7:30) the value of the variance component decreases considerably (highlighted values).

Traffic problems involving general networks present an additional difference that often challenges the appropriateness of the foregoing statistics. For example, traffic flows may have different magnitudes, with freeway flows in the thousands and arterial flows in the low hundreds or less. The GEH statistic provides an attractive and practical "self-scaling" alternative. Table 1 presents the GEH statistic both before and after calibration. The overall improvement is consistent with the improvement of the previous statistics; it ranges between 30% and 74%, with an overall improvement of 56%. The overall GEH value for the analysis period is less than the threshold of 5.0 recommended by FHWA guidelines (*36*), although individual GEH values are also close to this value.

Figure 5 presents scatterplots of an indicative subset of the same results. Each point in these plots represents the counts for one sensor, with the observed value plotted on the x-axis and the corresponding simulated value plotted on the y-axis. A diagonal (45-degree) line indicates the location of the points in the case of perfect fit. Ideally, all points would lie on that line. The goal of successful calibration is to bring these points as close as possible to this line. Furthermore, the deviations of the points should be as balanced as possible above and below the line, to indicate that no bias is evident in the calibration.

TABLE 1 Calibration Results Statistics

Time Period	RMSN			RMSPE			MPE		
	Before	After	% Change	Before	After	% change	Before	After	% Change[a]
6:00–6:15	0.537	0.179	67	0.535	0.216	60	−0.474	0.004	99
6:15–6:30	0.219	0.158	28	0.224	0.172	23	−0.128	−0.006	95
6:30–6:45	0.208	0.155	26	0.190	0.148	22	−0.137	−0.044	68
6:45–7:00	0.263	0.198	25	0.253	0.181	29	−0.171	−0.052	70
7:00–7:15	0.405	0.248	39	0.381	0.225	41	−0.271	−0.121	55
7:15–7:30	0.528	0.245	54	0.443	0.237	47	−0.335	−0.118	65
7:30–7:45	0.553	0.236	57	0.454	0.217	52	−0.353	−0.120	66
7:45–8:00	0.546	0.160	71	0.453	0.157	65	−0.337	−0.062	82
8:00–8:15	0.656	0.236	64	0.638	0.230	64	−0.614	−0.070	89
8:15–8:30	0.435	0.249	43	0.390	0.242	38	−0.262	−0.065	75
8:30–8:45	0.427	0.285	33	0.409	0.294	28	−0.199	−0.011	94
8:45–9:00	0.420	0.268	36	0.403	0.281	30	−0.192	0.030	84
6:00–9:00	0.476	0.231	51	0.416	0.221	47	−0.289	−0.053	82

Time Period	U			U^M			U^S		
	Before	After	% Change	Before	After	% Change	Before	After	% Change
6:00–6:15	0.316	0.084	73	0.769	0.003	100	0.008	0.001	91
6:15–6:30	0.111	0.076	32	0.462	0.064	86	0.083	0.140	−68
6:30–6:45	0.106	0.075	29	0.463	0.127	72	0.045	0.000	100
6:45–7:00	0.136	0.097	29	0.451	0.115	75	0.004	0.000	99
7:00–7:15	0.225	0.127	44	0.526	0.315	40	0.025	0.016	35
7:15–7:30	0.304	0.126	59	0.488	0.312	36	0.007	0.024	−240
7:30–7:45	0.325	0.119	63	0.509	0.301	41	0.016	0.000	100
7:45–8:00	0.317	0.079	75	0.480	0.195	59	0.011	0.001	91
8:00–8:15	0.436	0.118	73	0.862	0.161	81	0.029	0.015	50
8:15–8:30	0.245	0.124	49	0.511	0.143	72	0.087	0.023	74
8:30–8:45	0.233	0.139	40	0.373	0.037	90	0.038	0.007	82
8:45–9:00	0.227	0.127	44	0.355	0.000	100	0.016	0.005	72
6:00–9:00	0.264	0.113	57	0.461	0.116	75	0.020	0.015	27

Time Period	GEH		
	Before	After	% Improved
6:00–6:15	13.27	3.51	74
6:15–6:30	4.71	3.19	32
6:30–6:45	4.47	3.15	30
6:45–7:00	6.24	3.92	37
7:00–7:15	10.57	5.82	45
7:15–7:30	13.43	5.98	55
7:30–7:45	13.89	5.39	61
7:45–8:00	13.89	3.67	74
8:00–8:15	21.88	5.67	74
8:15–8:30	11.47	6.08	47
8:30–8:45	11.28	6.67	41
8:45–9:00	10.68	6.24	42
6:00–9:00	11.31	4.94	56

[a]In absolute values

CONCLUSION

Microscopic traffic simulation models are a mature and proven technology with applications in research and practice. Calibration of such models is a key aspect of any application. The detailed models in microscopic traffic simulators make calibration difficult, especially in the absence of adequate and reliable surveillance data.

The application of a systematic traffic simulation model calibration methodology is presented here that does not rely on the traditional assignment matrix approximation. The complex transformations that map O-D flows and other model parameters to traffic measurements are instead captured implicitly through the black-box simulation-based assignment model itself. This approach allows the estimation procedure to utilize richer data other than link counts, thus improving the accuracy and efficiency of the resulting flow and parameter estimates. The formulation is flexible to accommodate the simultaneous estimation of O-D flows and model parameters that affect O-D estimation, such as route choice and supply parameters. Algorithms to solve the highly nonlinear and large-scale problem are outlined, and their feasibility is demonstrated through a real case

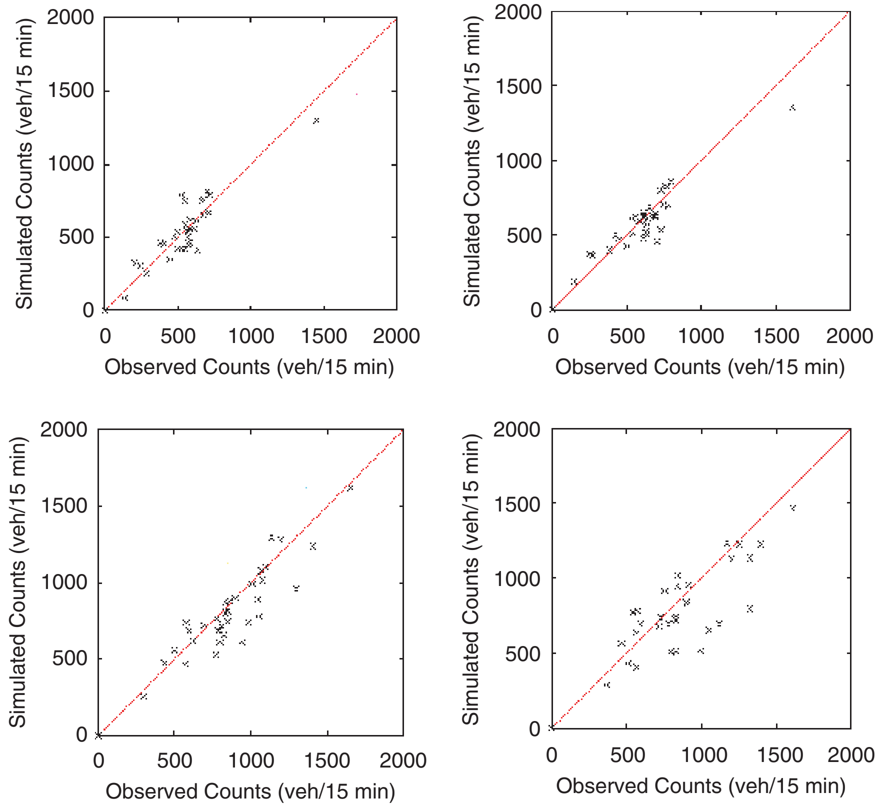

FIGURE 5 Indicative calibration results.

study involving a detailed microscopic traffic simulation model and a large, complex freeway network.

The calibration methodology is applied in a large-scale network in Lower Westchester County, New York, and practical aspects and difficulties of model calibration are identified. A multistep data processing and analysis approach is presented that overcomes the limitations associated with the available data. An interesting aspect of this application is the use of a multiclass O-D matrix to model parkway access restrictions for heavy vehicles. The calibration results show that this process is successful in matching prevailing traffic conditions.

The available sensor data for Westchester County are sparse relative to the number of links in the network. However, such sensor coverage levels are not uncommon in real-world situations, underlining the need for a calibration methodology that can extract maximum information from limited data. In this context, the proposed methodology efficiently estimates all model parameters of interest simultaneously. More data will be available for calibration as sensor coverage improves and is expected to result in increased accuracy. An important advantage of the proposed method is its ability to handle any type and amount of sensor data.

Future research should consider the impact of additional and more reliable surveillance data and test the sensitivity of the calibration process to such information. Such an analysis for DTA models is available elsewhere (*14*). In particular, the impact of AVI and probe-vehicle information should be considered. AVI and probe-vehicle data have already been successfully used for O-D estimation (*37, 38*) and DTA models (*39*).

REFERENCES

1. Toledo, T., H. N. Koutsopoulos, A. Davol, M. E. Ben-Akiva, W. Burghout, I. Andreasson, T. Johansson, and C. Lundin. Calibration and Validation of Microscopic Traffic Simulation Tools: Stockholm Case Study. In *Transportation Research Record: Journal of the Transportation Research Board, No.* 1831, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 65–75.
2. Ozbay, K., B. O. Bartin, and S. Mudigonda. Microscopic Simulation and Calibration of an Integrated Freeway and Toll Plaza Model. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
3. Liu, H. X., P. Chootinan, A. Chen, J. X. Ban, and L. Ding. Streamlined Network Calibration Procedure for California SR-41 Corridor Traffic Simulation Study. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
4. Smith, M. C., and A. W. Sadek. Challenges Calibrating a Large-Scale Microscopic Simulation Model of a Diverse Urban, Suburban, and Rural Network: A Practical Guide. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
5. Darda, D. *Joint Calibration of a Microscopic Traffic Simulator and Estimation of Origin-Destination Flows.* Master's thesis. Massachusetts Institute of Technology, 2002.
6. Mahanti, B. *Aggregate Calibration of Microscopic Traffic Simulation Models.* Master's thesis. Massachusetts Institute of Technology, Cambridge, 2004.
7. Zhang, M. H., J. Ma, and H. Dong. Calibration of Departure Time and Route Choice Parameters in Microsimulation with Macro Measurements and Genetic Algorithm. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
8. Abdulhai, B., J. B. Sheu, and W. Recker. *Simulation of ITS on the Irvine FOT Area Using the 'PARAMICS 1.5' Scalable Microscopic Traffic Simulator. Phase I: Model Calibration and Validation.* PATH Research Report UCB-ITS-PRR-99-12. University of California, Berkeley, 1999.

9. Kurian, M. *Calibration of a Microscopic Traffic Simulator.* Master's thesis. Massachusetts Institute of Technology, Cambridge, 2000.
10. Lee, D.-H., X. Yang, and P. Chandrasekr. Parameter Calibration for PARAMICS Using Genetic Algorithm. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
11. Kim, K.-O., and L. R. Rilett. Genetic-Algorithm-Based Approach to Traffic Microsimulation Calibration Using ITS Data. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
12. Brockfeld, E., R. D. Kuhne, and P. Wagner. Calibration and Validation of Microscopic Traffic Flow Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 62–70.
13. Ma, J., H. Dong and H. M. Zhang. Calibration of Microsimulation with Heuristic Optimization Methods. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 208–217.
14. Balakrishna, R. *Off-line Calibration of Dynamic Traffic Assignment Models.* PhD thesis. Massachusetts Institute of Technology, 2006.
15. Balakrishna, R., M. Ben-Akiva, and H. N. Koutsopoulos. Time-Dependent Origin-Destination Estimation Without Assignment Matrices. Presented at Second International Symposium on Transport Simulation, Lausanne, Switzerland, 2006.
16. Balakrishna, R., M. Ben-Akiva, and H. N. Koutsopoulos. Offline Calibration of Dynamic Traffic Assignment: Simultaneous Demand-and-Supply Estimation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2003,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 50–58.
17. Balakrishna, R., H. N. Koutsopoulos, and M. Ben-Akiva. Simultaneous Offline Demand and Supply Calibration of Dynamic Traffic Assignment Systems. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
18. Cascetta, E., and S. Nguyen. A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts. *Transportation Research,* Vol. 22B, No. 6, 1988, pp. 437–455.
19. Cascetta, E., D. Inaudi, and G. Marquis. Dynamic Estimators of Origin–Destination Matrices Using Traffic Counts. *Transportation Science,* Vol. 27, No. 4, 1993, pp. 363–373.
20. Ashok, K. *Estimation and Prediction of Time-Dependent Origin-Destination Flows.* PhD thesis. Massachusetts Institute of Technology, Cambridge, 1996.
21. Cascetta, E., and M. N. Postorino. Fixed Point Approaches to the Estimation of O/D Matrices from Traffic Counts on Congested Networks. *Transportation Science,* Vol. 35, No. 2, 2001, pp. 134–147.
22. Box, M. J. A New Method of Constrained Optimization and a Comparison with Other Methods. *The Computer Journal,* Vol. 8, No. 1, 1965, pp. 42–52.
23. Huyer, W., and A. Neumaier. SNOBFIT—Stable Noisy Optimization by Branch and Fit. *ACM Transactions on Mathematical Software* (submitted 2004).
24. Spall, J. C. Implementation of the Simultaneous Perturbation Algorithm for Stochastic Approximation. *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 34, No. 3, 1998.
25. Spall, J. C. Stochastic Optimization, Stochastic Approximation and Simulated Annealing. In *Wiley Encyclopedia of Electrical and Electronics Engineering* (J. G. Webster, ed.), Wiley-Interscience, New York, 1999.
26. Yang, Q., and H. N. Koutsopoulos. A Microscopic Traffic Simulator for Evaluation of Dynamic Traffic Management Systems. *Transportation Research,* Vol. 4C, 1996, pp. 113–129.
27. Yang, Q., H. N. Koutsopoulos, and M. E. Ben-Akiva. Simulation Laboratory for Evaluating Dynamic Traffic Management Systems. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1710,* TRB, National Research Council, Washington, D.C., 2000, pp. 122–130.
28. Akhil, C. *Development and Evaluation of Diversion Strategies Under Incident Response Using Dynamic Traffic Assignment System.* Master's thesis. Massachusetts Institute of Technology, Cambridge, 2003.
29. Savitzky, A., and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry,* Vol. 36, 1964, pp. 1627–1639.
30. Toledo, T. *Integrated Driving Behavior Modeling.* PhD thesis. Massachusetts Institute of Technology, Cambridge, 2003.
31. Toledo, T., and H. N. Koutsopoulos. Statistical Validation of Traffic Simulation Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 142–150.
32. Ashok, K., and M. Ben-Akiva. Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin–Destination Flows. *Transportation Science,* Vol. 34, No. 1, 2000, pp. 21–36.
33. Ashok, K., and M. Ben-Akiva. Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows. *Transportation Science,* No. 36, No. 2, 2002, pp. 184–198.
34. Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts,* 4th ed. McGraw-Hill, New York, 1997.
35. Theil, H. *Economic Forecasts and Policy.* North-Holland, Amsterdam, Netherlands, 1961.
36. *Traffic Analysis Toolbox,* Volume III: *Guidelines for Applying Traffic Microsimulation Modeling Software.* Publication FHWA-HRT-04-040. FHWA, U.S. Department of Transportation, 2004. http://ops.fhwa.dot.gov/trafficanalysistools/tat_vol3/index.htm. Accessed Nov. 15, 2006.
37. Antoniou, C., M. Ben-Akiva, and H. N. Koutsopoulos. Incorporating Automated Vehicle Identification Data into Origin-Destination Estimation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1882,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 37–44.
38. Antoniou, C., M. Ben-Akiva, and H. N. Koutsopoulos. Dynamic Traffic Demand Prediction Using Conventional and Emerging Data Sources. *IEE Proceedings Intelligent Transport Systems,* Vol. 153, No. 1, 2006, pp. 97–104.
39. Antoniou, C., B. Nowotny, A. Rechbauer, and M. Linauer. Calibration of DTA Models Using Floating Car Data: An Application of DynaMIT in Vienna. Presented at Second International Symposium in Transportation Simulation, Lausanne, Switzerland, 2006.

The Traffic Flow Theory and Characteristics Committee sponsored publication of this paper.