Calibration of a Macroscopic Traffic Simulation Model Using Enhanced Simultaneous Perturbation Stochastic Approximation Methodology

Jung-Beom Lee, MS.c. (Corresponding Author**)**
Graduate Student,
Department of Civil and Environmental Engineering,
Rutgers, The State University of New Jersey
623 Bowser Road,
Piscataway, NJ 08854-8014 USA
Tel: (732) 445-4012
Fax: (732) 445-0577
E-mail: jungbeom@eden.rutgers.edu

Kaan Ozbay, Ph.D.
Associate Professor,
Department of Civil and Environmental Engineering,
Rutgers, The State University of New Jersey,
623 Bowser Road,
Piscataway, NJ 08854-8014 USA
Tel: (732) 445-2792
Fax: (732) 445-0577
E-mail: kaan@rci.rutgers.edu

Abstract: 262
Word count: 4,729 Text + 4 Tables + 7 Figures = 7,479
Submission Date: August 1st, 2007
Revised version submitted Date: November 15th, 2007

Paper submitted for Presentation and Publication in the

TRANSPORTATION RESEARCH RECORD, JOURNAL OF TRANSPORTATION RESEARCH BOARD AFTER BEING PRESENTED

***TRANSPORTATION RESEARCH BOARD'S 87TH ANNUAL MEETING*, WASHINGTON, D.C., 2008**

ABSTRACT

Simulation, a popular and widely used method for studying stochastic, complex real-world systems, can accurately represent those conditions when its parameters are effectively calibrated. Previous studies on calibration generally focused on minimizing the sum of relative error between the observed data from a certain period of time in a typical day and the simulation output for the same period. This static approach can be explained as calibration with data obtained at one point in time. However, this type of calibration approach cannot capture a realistic distribution of all possible traffic conditions and may produce inaccurate calibration results. This paper proposes a calibration methodology based on the Bayesian sampling approach. Instead of a single demand matrix and corresponding observed traffic conditions that represent a specific point in time, this calibration methodology uses randomly generated demand matrices and corresponding traffic conditions from an observed statistical distribution of these variables. The goal of using input values generated from an observed distribution of demands is to accurately represent a wide range of all likely demand conditions observed at a facility. Moreover, at each iteration, the proposed calibration methodology reestimates optimal parameters by using a stochastic optimization algorithm known as the simultaneous perturbation stochastic approximation (SPSA) algorithm. A macroscopic simulation model of a portion of I-880 in California based on the cell transmission approach is calibrated with the proposed methodology. The proposed enhanced SPSA algorithm outperforms a simple SPSA algorithm based on several case scenarios studied as part of this paper. Future work will focus on calibrating larger networks and microscopic simulation models with the same methodology.

INTRODUCTION AND MOTIVATION

Traffic simulation models are increasingly used to evaluate complex real-world traffic problems. However, to estimate traffic conditions accurately, an effective calibration of the simulation model is required. In most real-world cases, simulation output obtained through default parameters might not always represent observed traffic conditions. Thus, proper selection of input parameters is required to enable an accurate representation of the prevailing traffic conditions of a modeled transportation facility. A number of studies that address the calibration and validation of microscopic and macroscopic simulation models have been conducted in the past. Ding (2003) [1], Ma and Abdulhai (2002) [2], Gardes et al. (2002) [3], and Lee et al. (2001) [4] selected mean target headway and mean reaction time as the major calibration parameters for PARAMICS-based microscopic simulation models. Kundé (2002) [5] used speed–density relationship and capacity to calibrate DynaMIT-P. The default values of selected parameters were changed by the simultaneous perturbation stochastic approximation (SPSA) algorithm and the box complex algorithm until various observations such as simulated flow and density matched the field observations. The SPSA algorithm is better than the box complex algorithm in terms of level-of-fit statistics. Ma and Abdulhai (2002) [2] and Lee et al. (2001) [4] used genetic algorithms to determine “best” values for the calibration parameters of mean target headway and mean reaction time as their overall methodology. Kim and Rilett (2003) [6] used the simplex algorithm to optimize the degree-of-fit for their models in CORSIM and TRANSIMS. **Table 1** presents a comprehensive summary of some of the most recent calibration studies performed.

Table 1: Summary of Previous Calibration Studies

Authors	Simulation Tool	Calibrated Parameters	Optimization Methodology	Type of Roadway Section	Objective Function	Validation Measure
Ding [1]	PARAMICS	Mean target headway, mean reaction time	SPSA algorithm	Freeway	Flow, density	MAE(1)
Ma and Abdulhai [2]	PARAMICS	Mean headway, mean reaction time, feedback, perturbation, familiarity	Genetic algorithm	Roadway	Traffic counts	MAE(1)
Gardes et al. [3]	PARAMICS	Mean target headway, mean reaction time	N/A	Freeway	Speed, Volume	N/A
Lee et al. [4]	PARAMICS	Mean target headway, mean reaction time	Genetic algorithm	Freeway	Occupancy, Flow	N/A
Kundé [5]	DynaMIT-P	Speed–density relationship, Capacity	Box complex, SPSA algorithm	Network	Free-flow, Minimum speed	RMSE(1)
Kim and Rilett [6]	CORSIM, TRANSIMS	CORSIM: Car-following factors, driver’s aggressiveness factor TRANSIMS: O-D matrix, PT1 parameters	Simplex algorithm	Freeway	Volume	MAER(1)
Schultz and Rilett [7]	CORSIM	Driver behavior parameters, vehicle performance parameters	Automated Genetic algorithm	Freeway	Volume, Travel time	MAE(1)
Jha et al. [8]	MITSIMLab	Parameters of the driving behavior models and route choice model, O-D flows, habitual travel times	Iterative approach	Urban Network	Travel time	N/A
Balakrishna et al. [9]	MITSIMLab	Driver behavior model parameters	SPSA algorithm	Freeway, Parkway	Traffic Counts	RMSN(1) RMSPE(1) MPE(1)
Toledo et al.	MITSIMLab	O-D flow,	Complex	Freeway	Speed,	RMSE(1),

[10]		behavioral parameters	algorithm	and arterial	Density	RMSP(1), MAE(1), MAPE(1)
Hourdakis et al. [11]	AIMSUN	Global, local parameters	Trial and error	Freeway	Volume	RMSP(1)
Park and Qi [12]	VISSIM	Eight parameters	Genetic algorithm	Intersection	Average travel time	N/A
Ma et al [13]	Microscopic simulation	Global parameters (Mean target head, mean reaction time etc) Local parameters (link headway factor, link reaction factor, etc.)	SPSA algorithm	Freeway	Capacity	N/A
Kim et al [14]	Microscopic Simulation	Various microscopic simulation parameters in VISSIM	Genetic Algorithm with Non-parametric statistical test	Freeway	Travel Time Distribution	N/A

[1] **RMSP:** root-mean-square percent error **RMSE:** root-mean-square error **MAPE:** mean-absolute-percent-error **MAE:** mean absolute error **MAER:** mean absolute error ratio **RMSN:** root mean square error **RMSPE:** root mean square percent error **MPE:** mean percent error.

Most previous calibration studies focused on minimizing the sum of relative error between observed data and simulation output obtained at one or a few points in time. Although this approach might achieve a good fit at these specific points, it might not fully capture all possible occurrences of time-dependent traffic conditions. For example, the same facility does not have identical traffic characteristics for the morning peak on different days of the same week, let alone in different weeks and months. Thus, this paper proposes a new and enhanced calibration methodology based on applying the SPSA stochastic optimization method in conjunction with Bayesian sampling techniques to more effectively calibrate traffic simulation models.

Bayesian techniques are typically adopted to capture time-dependent traffic conditions in a better fashion. Molina et al. (2005) [15] were the first researchers to propose the calibration of a microscopic simulation model developed in CORSIM with a Bayesian approach. The authors applied the Bayesian theory to overcome over- and under tuning of parameters and inaccurate data.

This study proposes an enhanced simulation calibration methodology that combines the stochastic optimization algorithm and Bayesian sampling techniques. The main purpose of using sampling methodology is to incorporate time-varying conditions from a distribution of demand. In our problem, the output of Daganzo's cell transmission traffic simulation model (CTM) [16] is generated as a result of the randomly sampled origin–destination (O-D) demands and input parameters calibrated in the previous iteration. The objective of the proposed enhanced SPSA (E-SPSA) analysis is to find the best simulation parameters given randomly sampled demand values. This study, in part, replicates the methodology of Molina et al. [15] in that the initial distribution (prior distribution in the case of Molina et al. [15]) of demands is obtained by the Bayesian sampling technique. Including a full Bayesian methodology will be part of future work.

This paper consists of (*a*) a description of the macroscopic simulation model development and proposed methodology of calibration, (*b*) an explanation of the underlying process in the calibration procedure, and (*c*) a presentation of the implementation of the proposed methodology, through a case study of the calibration and validation procedure using the CTM.

METHODOLOGY FOR CALIBRATING CTM

Calibrating the parameters of a traffic simulation model can be formulated as an optimization problem in which the analytical form of the objective function is unknown. This simulation–optimization problem can be more formally stated as the minimization of the sum of measures of accuracy for various inputs to the simulation model. This function is calculated at the j^{th} iteration as shown in equation [1].

$$L(D_1^j, D_2^j, ..., D_I^j, \Theta_1^j, \Theta_2^j, ..., \Theta_I^j) = \sum_{i=1}^{I} \left\{ g_1(Q_i^{Ob}, Q_i^S) + g_2(D_i^0, D_i^j) + g_3(\Theta_i^0, \Theta_i^j) \right\} \quad [1]$$

subject to constraints,

$$Q_k^S = A(D_1^j, D_2^j, ..., D_k^j, \Theta_1^j, \Theta_2^j, ..., \Theta_k^j) \text{ for } k \in (1, ..., I)$$

$$\underline{\Theta}_i \leq \Gamma_i^j \leq \overline{\Theta}_i$$

where

D_i^j is the O-D demand vector for the i^{th} time interval,

D_i^0 is the starting O-D demand vector for the i^{th} time interval

Θ_i^j is the parameter vector for the i^{th} time interval,

Θ_i^0 is the starting parameter vector for the i^{th} time interval,

$\underline{\Theta}_i, \overline{\Theta}_i$ are the lower and upper bounds for the parameter vectors, respectively,

Q_i^{Ob} is the vector of the observed set of flows for the i^{th} time interval,

Q_i^S is the vector of the simulated set of flows for the i^{th} time interval,

$A(D_1^j, D_2^j, ..., D_k^j, \Theta_1^j, \Theta_2^j, ..., \Theta_k^j)$ is the autoregressive assignment matrix for the k^{th} period,

L is the objective function in the optimization process, and

g_1, g_2, g_3 are measures of accuracy for traffic flows, O-D demands, and the parameter set, respectively.

In addition to the objective function that does not have a closed-form representation, the calibration problem is a "stochastic optimization" problem. Variables of many traffic simulation models have a stochastic component to reflect random variations in real-world observations. Thus, the traffic simulation calibration problem has to be approached as a multivariable stochastic optimization problem that does not have a closed form of objective function.

The functional form of the objective function in stochastic approximation algorithms is probabilistic. Each algorithm can be divided into two general categories: gradient and gradient-free settings. The steepest descent method [17] and the Newton-Raphson method [17] are gradient-based deterministic algorithms. Nelder and Mead (1965) [18] proposed the nonlinear simplex algorithm, a deterministic method that is based on a gradient-free multivariate optimization method.

The stochastic approximation of the objective function (usually termed as loss function L) can be used in the presence or absence of the gradient function $g(\theta)$. The stochastic root-finding algorithm of Robbins-Monro [17] was generally used for nonlinear problems when the gradient function is available. When the measurement of gradient is impossible, such as in the case of simulation, a gradient-free approach is applied. The finite-difference (FD) approximation is the most well-known gradient approximation method. However, FD approximation is performed only when the noise measurements of the loss function are available. SPSA, one of the well-known stochastic approximation algorithms that can be applied in both the stochastic gradient and gradient-free settings, can also be applied to solve optimization problems with a large number of variables.

Applying SPSA in the optimization of multivariable loss functions described by Spall [17], [19], [20], and in equation [1] led to the basic form below for each parameter in Θ_i^j:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}_k(\hat{\theta}_k) \quad [2]$$

Here, $\hat{g}_k(\hat{\theta}_k)$ is the gradient $g(\theta) = \dfrac{\partial L(\theta)}{\partial \theta}$ estimated, based on the loss function measurements, at $\theta = \hat{\theta}_k$ at the k^{th} iteration.

The new value of θ, obtained for every iteration, is calculated by subtracting the product of step size and the gradient at the present value from the previous value of θ, as shown in equation [3]. With simultaneous perturbation, loss measurements are obtained by randomly perturbing the elements of $\hat{\theta}_k$. Assuming that θ is p-dimensional, the SP gradient approximation step can be shown in the following form:

$$\hat{g}_k(\hat{\theta}_k) = \begin{bmatrix} \dfrac{\hat{L}(\theta + c_k\Delta_k) - \hat{L}(\theta - c_k\Delta_k)}{2c_k\Delta_{k1}} \\ \bullet \\ \bullet \\ \bullet \\ \dfrac{\hat{L}(\theta + c_k\Delta_k) - \hat{L}(\theta - c_k\Delta_k)}{2c_k\Delta_{pk1}} \end{bmatrix}$$

$$= \frac{\hat{L}(\theta + c_k\Delta_k) - \hat{L}(\theta - c_k\Delta_k)}{2c_k}\left[\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, ..., \Delta_{kp}^{-1}\right]^T \quad [3]$$

Here, the p-dimensional random perturbation vector $\Delta_k = \left[\Delta_{k1}, \Delta_{k2}, ..., \Delta_{kp}^{-1}\right]^T$ is a user-specified vector for which the components of Δ_k are usually ± 1 Bernoulli variables. The gain sequence is used to balance the algorithm stability and the desired form of the gain sequence is shown below:

$$a_k = \frac{a}{(k+1+A)^{\alpha}}, \qquad c_k = \frac{c}{(k+1)^{\gamma}} \quad [4]$$

If a is small, the calculations are stable initially; however, this may result in sluggish performance for large calculations. On the other hand, a large numerator, $a_k > 0$, which is used to produce nonnegligible step sizes, leads to instability early in the calculation. It is most effective to set the numerator c to a small positive number.

In this study, flows and densities were obtained with a macroscopic simulation model based on the CTM proposed by Daganzo (1994) [16]. Since the goal in this paper is to introduce a new calibration methodology, a well-accepted but easy-to-implement approach was chosen for the simulation component of the study. However, it is clear that the proposed calibration methodology can be implemented in conjunction with any other simulation tool. The CTM is a simple and accurate representation of traffic situations such as acceleration–deceleration, stop and go, and shockwaves. CTM limits the flow to the minimum value between the upstream capacity and downstream capacity of the cell [16]. The maximum number of vehicles in the undercongested condition is the product of the jam density (K_j) and cell length at cell i, and the maximum number of vehicles in the overcongested condition is the capacity of the cell $i-1$. **Figure 1** depicts a representative flow–density relationship for the basic CTM. It is defined as a trapezoid, where V_f and K_j indicate free-flow speed and jam density, respectively.

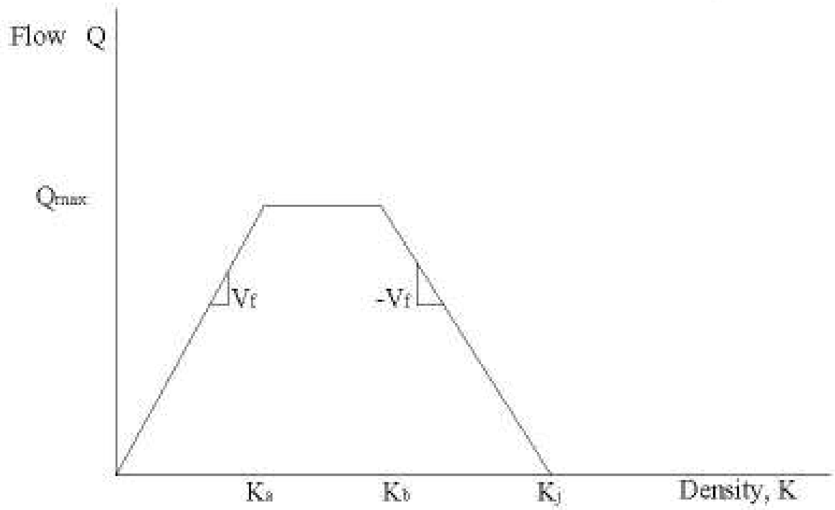


Figure 1: Flow-density relationship for the basic cell transmission model [16]

Previously, Munoz et al. (2004)(2006) [21], [22] calibrated a cell-transmission-based model of a portion of westbound I-210 in Southern California. Free-flow speed and congestion parameters were calibrated using a constrained least-squares fit on the flow–density relationship. The difference between the simulated and the observed total travel times is the objective function used to evaluate the performance of the simulation. However, this model focused only on the deterministic aspects of the calibration problem.

This paper proposes using the SPSA approach proposed by Spall (1992) in conjunction with a Bayesian sampling methodology. Figure 2 shows a flowchart of the proposed combined E-SPSA calibration and validation methodology.

The basic steps of the proposed methodology can be summarized as follows:

1. Increment iteration: iteration = iteration + 1

   Iteration:

   a. Generate the O-D demand matrix from a probability distribution function of demands developed with real-world data by using the Bayesian sampling methodology [23], [24].
   b. Use the SPSA algorithm to determine the optimal parameters given the demand matrix generated.

2. Compare the output of simulation for the given demand matrix in the current iteration—namely, flows and densities—with the observed distribution of flows and densities to determine the correlation between the two distributions. If it is satisfactory, terminate the iterative process and proceed to the validation step. If unsatisfactory, return to Step 1.

3. If verification and validation tests are satisfactory, then stop. If not, return to Step 1.

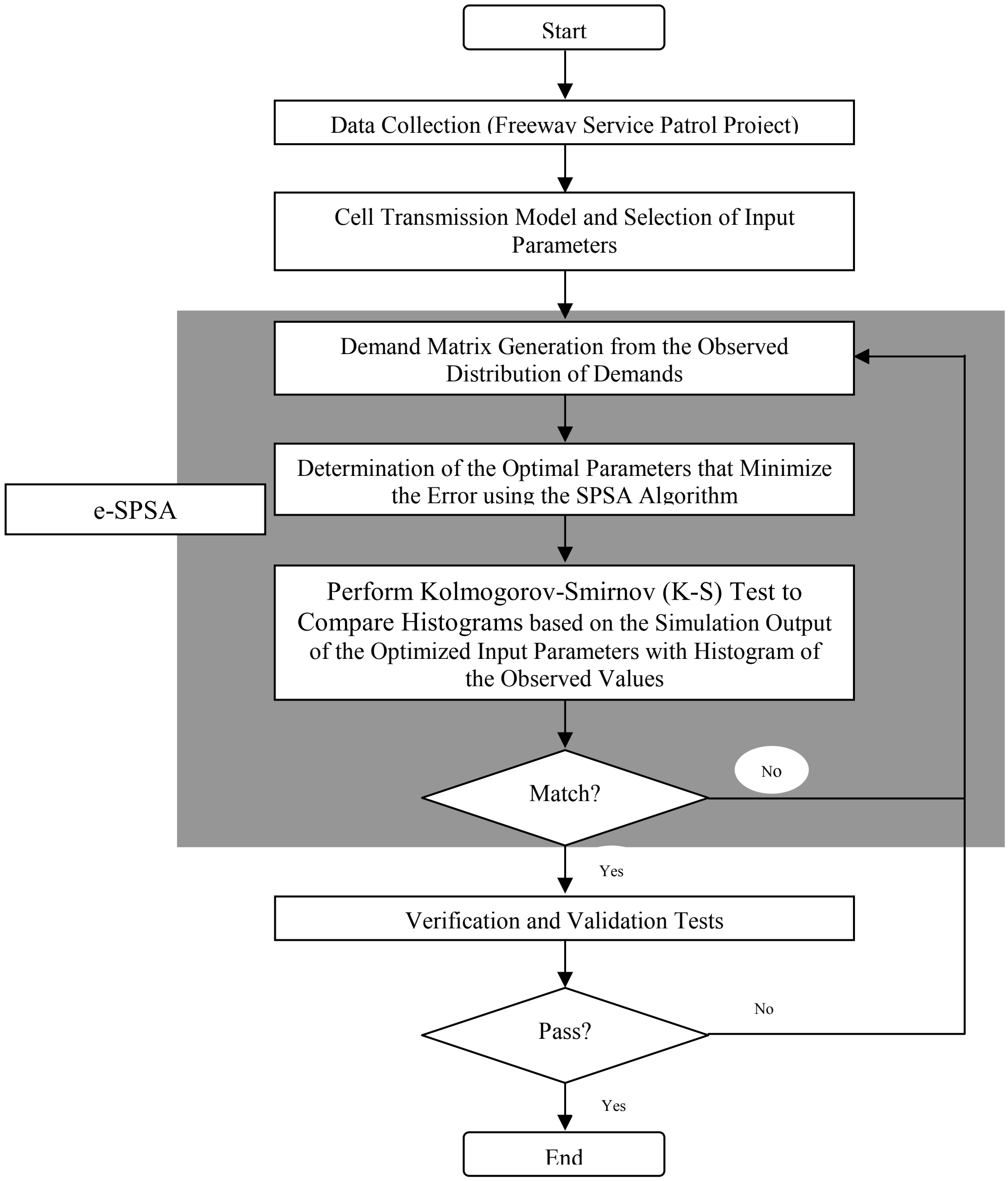


Figure 2 Proposed combined enhanced SPSA simulation calibration methodology.

Description of Sampling Methodology

The initial distribution consists of random samples generated from the observed demand matrix, using WinBUGS, a software package that performs Markov chain Monte Carlo simulations to randomly generate time-dependent O-D demand matrices. Each randomly generated demand (d_k) ($k \in (1,...,I)$, n indicates the number of demand matrix) is a random sample of the initial distribution. This initial distribution is used to find the optimal input parameter values of the free-flow speed and the jam density. At each iteration, these input parameters are determined by using the SPSA algorithm, as discussed in the preceding section.

IMPLEMENTATION DETAILS OF E-SPSA CALIBRATION METHODOLOGY FOR A FREEWAY SEGMENT

To test the effectiveness of the E-SPSA calibration methodology, a freeway segment with and without ramps was considered.

Demand Matrix Generation from Observed Distribution of Demands

In traffic simulation models, the vehicular demand is modeled by using traffic zones. In the case of freeway segments, these zones are usually placed at a segment upstream of the segment under study. If the study segment is relatively short and the flow is observed on all sections (before and after ramps), the demand generated from the zone will, on an average, be equal to the flow observed at the segment under study. So, it can be assumed that the flow observed at a point on the freeway is the resulting effect of the same amount of demand generated at a zone upstream of the segment. In this study, the demand matrices were generated by using the traffic counts (i.e., flows from loop detectors). Hence the autoregressive assignment function A as described in equation [1] will depend only on the current O-D demand D_1^j and parameters Θ_1^j. In the case study presented (shown in Figure 3), flows are observed on all sections. Hence, the demands could be determined uniquely.
At each iteration, a new demand matrix is formed from the flows, which are randomly selected from the distribution of observed flows, which is the prior distribution. Thus, each generated demand matrix is effectively a sample from the initial distribution. Depending on the existence of intermediate ramps, a random demand matrix is formed from the distribution of observed flows for the two distinct geometries. The basic calibration procedures of demand originating from a single zone and demand originating from two zones may differ slightly because of the sampling methodology. For the on-ramp scenario, the prior distribution is obtained from the relationship between the flow on the mainline and that on a ramp. For generating the demands from two zones, the demands for the mainline and ramp (the two components of D_1^j) are sampled simultaneously formed from the flows Q_i^{Ob}. To take into account the existence or lack of correlation between the two demands, the demand matrix is generated by using conditional or independent samples, respectively. The existence of the autoregressive assignment function shows that there can be a correlation (through Q_i^{Ob}) between the ramp demands and mainline demands D_1^j.

Determination of Optimal Parameters That Minimize the Error Using the SPSA Algorithm

Based on the randomly generated demand matrix, the calibration parameters are reestimated at each iteration. To select the optimum values of the parameters, a stochastic optimization algorithm, the SPSA algorithm, is used. The format of the objective function for the given simple freeway section is of the form shown in equation [5].

$$L = \sum_i \sum_{lane} \left[\frac{|Q_i^{real} - Q_i^{sim}|}{Q_i^{real}} + \frac{|K_i^{real} - K_i^{sim}|}{K_i^{real}} \right] \quad [5]$$

where

L : objective function,

Q_i^{real} : observed flows for a given time period *i* and lane,

Q_i^{sim} : simulated flows for a given time period *i* and lane,

K_i^{real} : observed density for a given time period *i* and lane,

K_i^{sim} : simulated density for a given time period *i* and lane, and

i: time period $i \in (1,...,I)$.

It is apparent from the explanation in the section on demand matrix generation that the demands D_1^j and flows Q_i^S are directly correlated. If the function for the measure of accuracy can be assumed to be the same—namely, relative error—then the objective function will have same terms repeating twice. Therefore, the objective function to be optimized reduces to that shown in equation [5]. Because CTM is a macroscopic simulation model, there is an additional term K_i^{real} density that can be measured and calibrated easily.

CTM simulation is then executed multiple times with different random seeds to consider the variability in the simulation at each step of the SPSA algorithm. The random seeds are an important factor that is studied to determine their effect on the simulation results. In the stochastic analysis, the random seeds produce more accurate results by representing measurement noise. In the CTM methodology used, the random seeds directly affect the demand and hence the flow. Thus, to minimize the variance of errors, it is important to run the simulation with several different random seeds. At the end of each simulation run, the relative error between the observed and the simulated is calculated with equation [5]. The SPSA algorithm is implemented to optimized loss function L. Ding [1] suggested that the values of a, c, α, and γ in equation [4] be set to 0.2, 0.5, 0.602, and 0.01, respectively; those values are used in this study.

Perform Kolmogorov–Smirnov (K-S) Test

Random samples from the prior distribution are generated from the observed demand matrix. The SPSA algorithm for each generated demand matrix calibrates these parameters. When one iteration of the calibration procedure is completed and the optimized parameter values are found by the SPSA algorithm, the evaluation process is performed. The distribution of the flows and densities from the cell transmission is compared with the distribution of the observed values by using the K-S test. If the two distributions pass the K-S test, the procedure moves to the validation step. If they do not pass the test, the sampling process is repeated. The advantage of using a nonparametric statistical test such as the K-S test is that it does not depend on the distribution of the parameter in question.

CASE STUDY

The calibration of a single zonal demand case was performed to test the effectiveness of the proposed E-SPSA approach. Additionally, the calibration was done with two demand zones over an extended road segment (mainline and on-ramp).

Data Collection

Data were obtained from the database of the Freeway Service Patrol project for a portion of the I-880 freeway in Hayward, California [25]. Data were collected from September 27 to October 29, 1993, from 05:00 to 10:00 and from 14:00 to 20:00 for weekdays and were aggregated into 5-minute counts. The data were used differently depending on the existence or nonexistence of an intermediate ramp. Flow data for all the time periods for 17 different days were estimated in the cases with no intermediate ramp. When intermediate ramps were present, data were divided into four intervals: depending on morning or evening and on peak (reach the capacity) or nonpeak (uncongested) period (05:00 to 07:30, 07:30 to 10:00, 14:00 17:00, and 17:00 to 20:00). In the case of a freeway with a single zone (no intermediate ramp), the selected freeway segment is a two-lane one-way road and the length of the section is 1 mile. In the case of a freeway with an intermediate ramp, one lane for the mainline and on–ramp, respectively, is used and the length of the section is 1.2 miles.

CTM and Selection of Input Parameters

In the case of single zonal demand (no intermediate ramp), the road segments in CTM are initially modeled as a simple road segment to test the effectiveness of the E-SPSA approach. The selected freeway segment is a two-lane one-way road and the length of the section is 1 mile. The case of an intermediate ramp is modeled as an extended road segment as described by Daganzo [26] (Figure 3). The selected freeway segment is one lane for the mainline and on-ramp, respectively, and the length of the mainline is 1.2 miles. Because the zonal demand, one of the major input variables, is sensitive to the flow–density relationship, free-flow speed (V_f) and jam density (K_j) were selected as the input parameters.

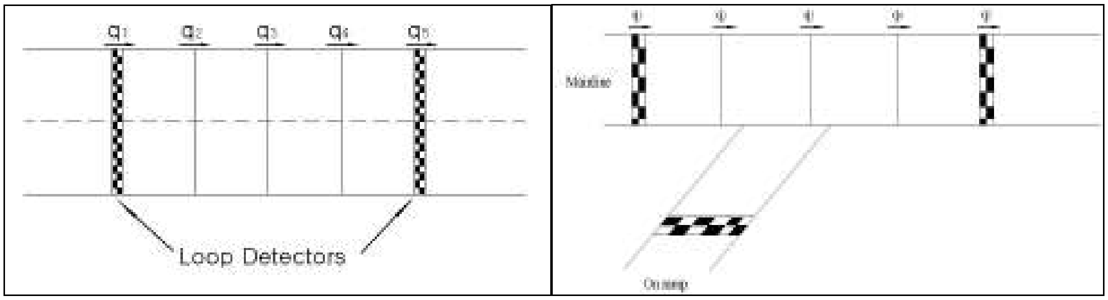
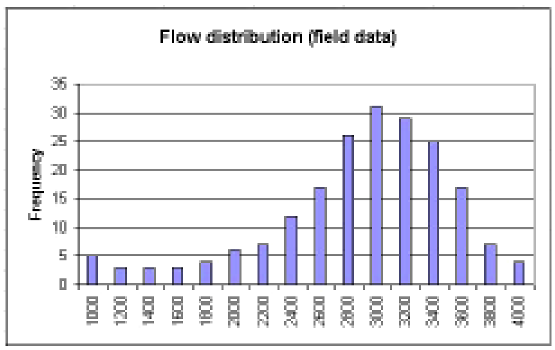


Figure 3 The network geometry

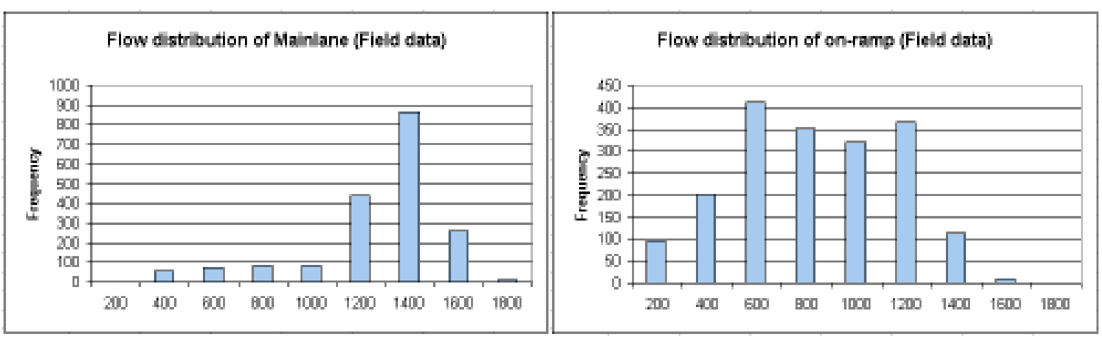
Demand Matrix Generation from Observed Distribution of Demands

When a ramp exists, the demand matrix (formed from flows) is generated depending on whether the flows from the two zones are correlated. The main assumption is that there is no correlation between these flows, which makes it possible to sample the mainline and ramp flows independently.

Figure 4 shows the histogram of the observed demand matrix formed from the flows.



(a)



(b)

Figure 4 Histogram based on distribution of the 17 days of observed demand matrix: (a) no intermediate ramp and (b) intermediate ramp.

Determination of Optimal Parameters That Minimize Error Using the SPSA Algorithm

No Intermediate Ramp For single zonal demand (no intermediate ramp), the initial values of free-flow speed and jam density are taken as 55 mph and 110 vehicles per mile for two lanes, respectively. The SPSA algorithm is used to determine the "optimal" values of the calibrated simulation parameters for each iteration. The same procedure is repeated until the sum of the relative error between observed and simulated values is less than the acceptable error of 5%. Each iteration is performed with three different random seeds when using the SPSA algorithm.

The input parameters converge when there is only a small difference between the simulated and observed values of flow and density. These are the optimized input parameter values for each generated demand matrix. The best fit for flow and density is observed for optimized input parameters of 59.6 mph and 107.0 vehicles per mile for two lanes, respectively. The sum of relative error is 4.6 %.

Intermediate Ramp For two zonal demand (mainline and on ramp), the resulting output of the calibrated cell transmission model is compared with the observed data. The initial parameters of free-flow speed and jam density are set to 60 mph and 60 vehicles per mile for four time periods. The sum of relative errors for four time periods (05:00 to 07:30, 07:30 to 10:00, 14:00 to 17:00, and 17:00 to 20:00) were found to be 4.71, 4.06, 3.97, and 4.43, respectively. Since a trapezoidal flow-density curve is assumed, $Q_{max} \leq K_j V_f/2$. $K_j V_f/2 \approx 1800$ veh/h. which is a reasonable result.

Results of K-S Test [27]

The obtained flow and density distributions of the macroscopic simulation output based on the optimized input parameters were compared with the observed distributions, using the K-S test [27]. The null hypothesis states that simulated flow and density distributions are not statistically different from the distributions of observed flow and density values, respectively. If the null hypothesis is rejected, new demand is generated from the observed distribution of demands. For the single zonal demand case, the K-S test values for flow and density distributions are 0.019 and 0.139, respectively. These values are less than the critical values of 0.247 and 0.340 obtained from the K-S table at the 95% confidence level. For the two zonal demand cases, the critical K-S values from the K-S table are greater than the K-S values for flow and density distributions, as shown in Table 2. For the scenario with two simple road segments, the null hypothesis could not be rejected, so there is no reason to doubt its validity; the null hypothesis states that the simulation flow and density distributed are not different from the observed distributions. Figure 5 shows the distribution of simulated flow and density values when the optimized values of input parameters are used.

Table 2 Kolmogorov-Smirnov Values for Each Time Period (Intermediate Ramp Case)

Time	K-S Value of Flow	K-S Value of Density	Critical Value from K-S Table for Flow (95 %)	Critical Value from K-S for Density (95 %)
05:00~07:30	0.096	0.089	0.340	0.544
07:30~10:00	0.038	0.063	0.389	0.453
14:00~17:00	0.040	0.032	0.453	0.544
17:00~20:00	0.019	0.025	0.340	0.453

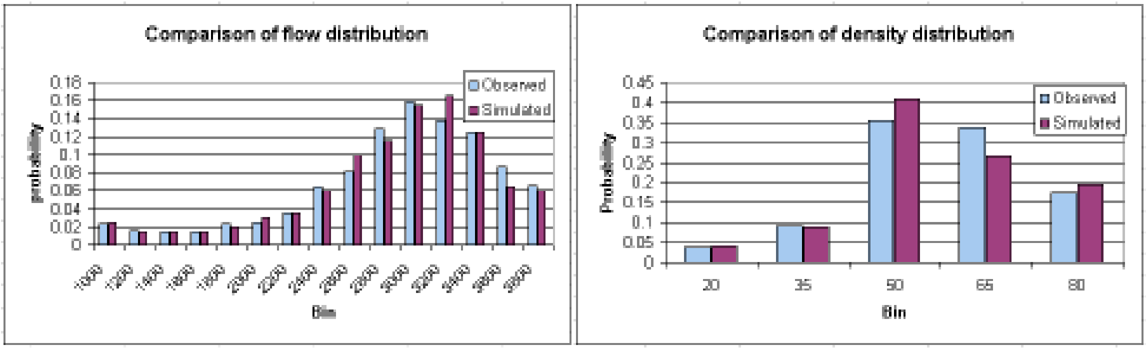


Figure 5(a)

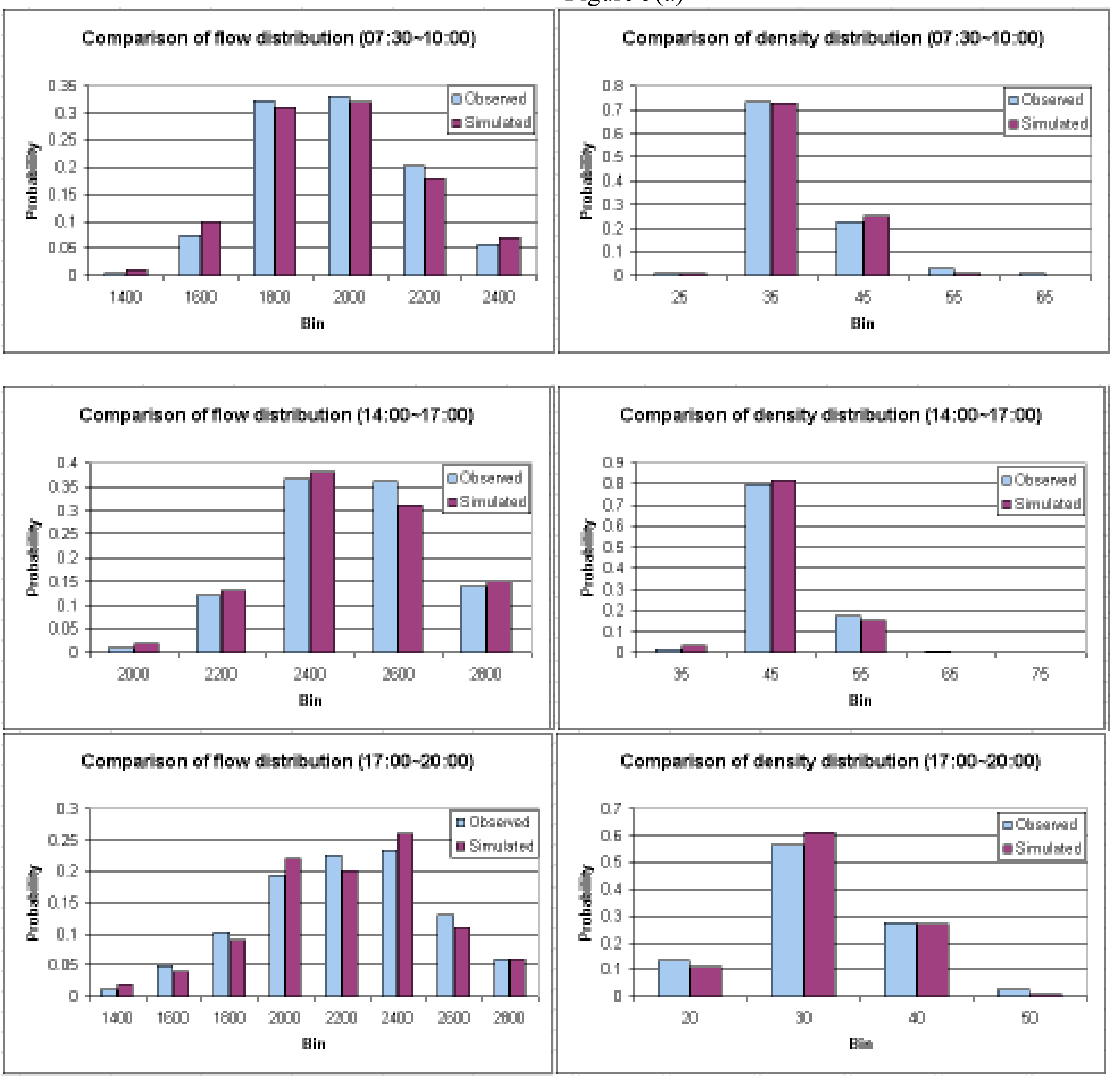


Figure (5b)

Figure 5 Comparison between simulated and observed distributions: (a) no intermediate ramp, and (b) intermediate ramp.

Verification and Validation Tests

The verification test was performed to ensure that the determined optimal input parameters (the free-flow speed and jam density) represent realistic and accurate values under real traffic conditions. This verification analysis tested whether the objective function could satisfy the predetermined stopping criteria (5%). The result of this verification test and the performance of the simulation tool show that real-world conditions can be fairly captured by the calibrated simulation model.

When the verification processes were satisfied, the validation of calibrated parameter values was performed for different days in the same time period and for the same network. It is difficult to compare the calibrated model performance with the similar performance results found in other calibration papers. However, the mean square variation (MSV)[2] that Sanwal et al. (1996) [28] used is a good method to compare the degree of deviations with the observed values. MSV is calculated by subtracting the mean square error from one. If the model's estimations are close to real-world measurements then the MSV should be close to one.

The optimized parameter values were used to simulate randomly selected days as a part of the validation process.

The results of the validation for the case of the two geometric configurations (with and without intermediate ramp) are discussed below.

No Intermediate Ramp The optimized parameter values were used to simulate two randomly selected days. The distributions of flow and density based on the optimized parameter values are compared with the observed data distributions (Figure 6). Based on the K-S test, the values of the flow distributions for September 30, 1993, and October 13, 1993, were 0.019 and 0.018, both less than the critical K-S value from the table (0.194). The values of density distributions 0.037 and 0.012 are also less than the critical K-S value of 0.453 at the 95% confidence level. According to the K-S test, observed and simulated flow and density distributions are proven to have an acceptable level of similarity (fit) with respect to each other.

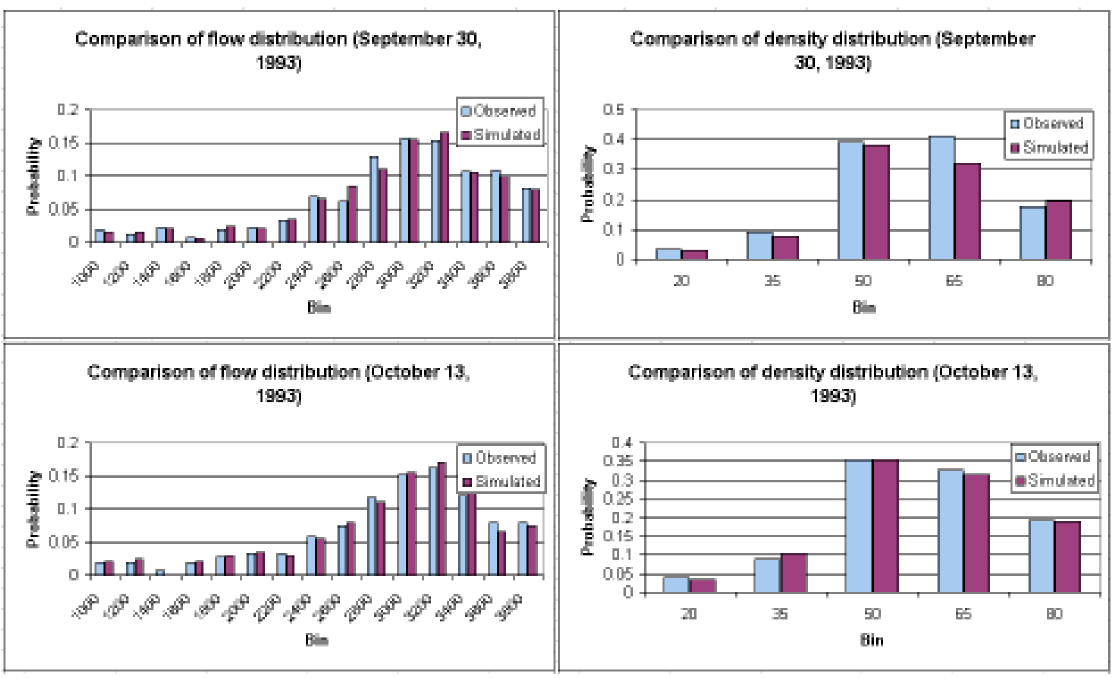


Figure 6 Comparison between simulated and observed flow and density distributions (September 30, 1993, October 13, 1993).

Intermediate Ramp The validation test was performed for three scenarios. The first scenario used the E-SPSA approach for four randomly selected days. The second and third scenarios tested the SPSA algorithm without the sampling methodology for single and multiple days. These tests were performed to determine the effectiveness of the E-SPSA approach compared with the SPSA-only approach. Finally, the MSV measure used by Sanwal et al. (1996) [28] was applied as an evaluation criterion to compare the degree of deviations from the observed values.

---

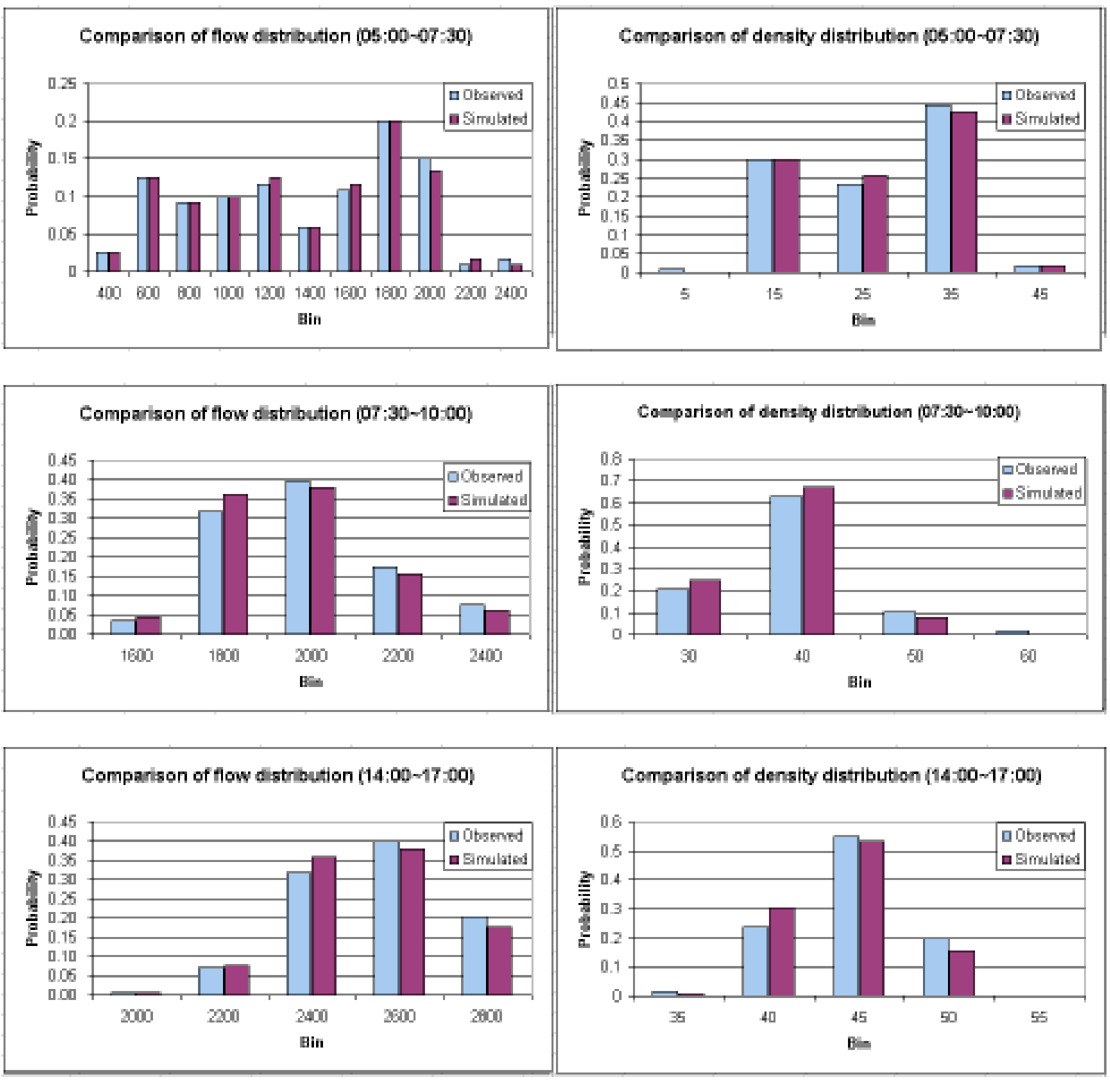
[2] Fitness criteria for MSV are $R_{fit}^{v} = 1 - E\left[(v - v_m)^2\right] / E\left[v_m^2\right]$ (here, v_m and v denote measured speeds and model speeds, respectively) and $R_{fit}^{T} = 1 - E\left[(T - T_m)^2\right] / E\left[T_m^2\right]$ (here, T_m and T denote measured travel times and computed travel times, respectively.) [28].

First Scenario Optimized parameter values of four time periods match well with the distributions of observed values shown in Table 2. The optimized parameter values were used for four randomly selected days that were not used in the calibration process. The statistical test results of this process are shown in Table 3.

Table 3 Kolmogorov-Smirnov Values for Each Time Period

Time	K-S Value of Flow	K-S Value of Density	Critical Value from K-S Table for Flow (95 %)	Critical Value from K-S for Density (95 %)
05:00~07:30	0.042	0.034	0.272	0.389
07:30~10:00	0.116	0.026	0.453	0.453
14:00~17:00	0.028	0.036	0.247	0.453
17:00~20:00	0.087	0.080	0.194	0.389

For all periods of the day, each K-S value was less than the critical K-S values given by the table. This result also shows that two distributions between the observed and the simulated flow and density fit closely. Figure 7 shows the fitness between the observed and the simulated distributions of flow and density for each time period.



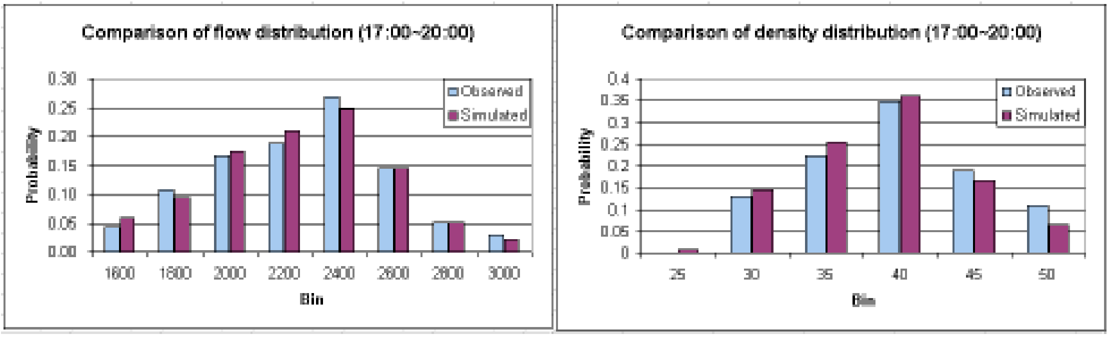


Figure 7 Comparison between simulated and observed distributions.

Second Scenario In the second scenario, the validation test was performed for the calibration process based on the SPSA-only algorithm. Calibration was performed with a single day's data until the error was within the bounds. September 30, 1993, data were tested to determine optimal parameter values and the resulting relative errors were calculated as 2.13, 2.66, 3.51, and 4.56. For three randomly selected days, the error was calculated using these optimized parameter values. In the case of the third time period, 14:00 to 17:00, the relative errors range from a low of 3.53 on October 13, 1993, to a high of 13.53 on October 5, 1993. The resulting relative errors are summarized in Table 4. On the basis of the SPSA-only algorithm result, flow and density distributions for the time period 07:00 to 10:00 on September 13, 1993, were compared with observed data distributions using the K-S test. The critical value of the K-S test for flow and density distributions is 0.272, which is more than the value calculated for simulated flow distribution (0.207) and less than the calculated value for the simulated density distribution (0.448). Thus, the null hypothesis is rejected for density distribution.

Table 4 Sum of Relative Errors for the Validation Test

	Sum of relative Error of Flow and Density			
Days	05:00~07:30	07:30~10:00	14:00~17:00	17:00~20:00
09/30/93	2.13	2.66	3.51	4.57
10/05/93	8.25	11.17	13.53	5.79
10/13/93	7.36	13.49	3.53	4.06
10/18/93	5.73	5.51	6.55	6.41
Average	**5.87**	**8.21**	**6.78**	**5.21**

Third Scenario In the third scenario, the validation test was performed for the same time period on multiple days and relative errors of 4.58, 4.82, 4.19, and 4.88 were obtained. The optimized parameter values based on the SPSA algorithm were used for four randomly selected days and for four different time periods on those days. The relative error increased to 10.42, 16.21, 17.58, and 15.52. With this type of calibration approach (SPSA), the two independent data sets for the same facility can be different, even if the mean of the observed data distribution closely matches the simulated data. Hence, it is concluded that the optimal parameter values from the SPSA algorithm are not always transferable to data sets from the same time periods on different days.

On the basis of the variation approach proposed by Sanwal et al. (1996) [28], the MSV of flow of four randomly selected days was compared with the simulation results obtained from the simulation model calibrated using the E-SPSA algorithm. From the morning period between 05:00 and 07:30 to the evening period between 17:00 and 20:00, the MSVs of flow are 0.968, 0.979, 0.981, and 0.950. Sanwal et al. (1996) [28] obtained a value of 0.945 for the variation of speed and 0.968 for the variation of travel time when they applied the optimized parameters from one day to another. The results of degree of deviation are comparable to measured data when the method of Sanwal et al. (1996) [28] was applied.

CONCLUSIONS AND FUTURE RESEARCH

Careful calibration of traffic simulation models is necessary to accurately represent prevailing traffic conditions. In this paper, a new calibration methodology based on the Bayesian sampling approach is proposed. Instead of using a single demand matrix and corresponding observed traffic conditions that represent one point in time, the new methodology uses randomly generated demand matrices and corresponding traffic conditions from an observed distribution of these variables. The goal of using input values, generated from the observed distribution of demands, is to enable an accurate wide-range representation of all likely demand conditions observed at a facility. Observed demand values were used to determine a distribution of observed demands for 17 days.

At each iteration of the proposed Bayesian sampling framework, a new demand matrix, which is randomly sampled from this distribution, is loaded into the macroscopic simulation model. Moreover, at each iteration, the proposed calibration methodology reestimates optimal parameters by using a stochastic optimization algorithm, SPSA, and distributions of flow and density from the macroscopic simulation are compared with the distribution of the observed flow and density [24].

A cell transmission-based macroscopic simulation model of a portion of I-880 in California was calibrated with the proposed methodology. Two relatively simple road segments, shown in Figure 3, one with a single zonal demand (no intermediate ramp) and the other with two zonal demands (mainline and on-ramp), were used as case studies. The distribution of output from CTM was obtained by loading demand matrices randomly sampled from the distribution of observed demand and from the calibrated input parameters, which are the free-flow speed and jam density. The distribution of simulation output was compared with the observed data distribution by using the K-S test. The null hypothesis for the K-S test stated that simulated flow and density distributions are not different than their observed counterparts. For all scenarios, the null hypothesis could not be rejected at the 95% confidence level. Thus, it can be concluded that the differences between the distributions of observed and simulated flow and density values are not statistically significant.

The validation test with a single zonal demand was also performed over two randomly selected cases using the same time period and network. On the basis of optimized input parameters, the flow and density distributions were compared. The differences between the observed and the simulated flow and density distributions were found to be statistically insignificant at the 95% confidence level. Thus, these parameters can be considered to have been validated. As an extended simulation model, a road segment with an intermediate ramp was modeled using the CTM. The sum of relative errors for four time periods (05:00 to 07:30, 07:30 to 10:00, 14:00 to 17:00, and 17:00 to 20:00) were found to be 4.71, 4.06, 3.97, and 4.43, respectively. For three different scenarios, the model parameters were tested for validation and for the effectiveness of the E-SPSA approach. From Table 3, the results of four different time periods were found to satisfy the statistical test for accepting the null hypothesis, and it is concluded that parameter values calibrated by using E-SPSA are applicable to data from a different day. On the other hand, in the second and third scenarios, where SPSA was used without the sampling approach as the main calibration algorithm, predetermined validation constraints were not always satisfied. According to Table 4, none of the average values of the sum of relative error of flow and density values satisfied the constraint of an acceptable relative error of 5%. In addition, in the scenario in which parameters were calibrated with the SPSA-only methodology, the difference between the distributions of simulated flow and density values and the distributions of observed values could not be shown to be statistically insignificant when the K-S test was performed. Therefore, the E-SPSA method was found to improve the results of simulation calibration by accurately capturing a wide range of real-world conditions.

Bayesian analysis was applied as part of the calibration methodology to represent the distribution of the observed traffic characteristics. The main purposes of using Bayesian analysis were to overcome the over- and underestimation of calibration parameters and to acquire a realistic distribution of all possible traffic conditions. Previously, the Bayesian analysis method was not extensively used because of its complex computational requirements. However, improved performance of modern computers, coupled with efficient computational techniques, allow the Bayesian method to be widely used in a number of areas.

Testing of E-SPSA for larger networks, as well as for microscopic traffic simulations such as PARAMICS, are future research tasks. In the future, other simulation parameters and more extensive data sets will be used to test the strengths and weaknesses of the proposed E-SPSA calibration methodology. Also, a full Bayesian methodology (B-SPSA) can be developed along the lines of Molina et al. [15]. An enhancement to the approach they followed could be that, instead of assuming a general prior distribution for the parameters Θ_1^j, the prior can be estimated by using the E-SPSA algorithm described in this study and performing a Markov chain Monte Carlo simulation to obtain the posterior distribution of parameters given the observed data Q_i^{Ob}.

REFERENCES

1. Ding, L., "Evaluation of ITS Technologies In South Jersey Priority Corridor Project", Master Thesis, Rutgers, the State University of New Jersey, 2003.
2. Ma, T., and Abdulhai, B., "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.
3. Gardes, Y., May, A.D., Dahlgren, J., and Skabardonis, A., "Freeway Calibration and Application of the PARAMICS Model", Paper presented at the 81st Transportation Research Board Meeting, 2002.
4. Lee, D. H., Yang, X., and Chandrasekar, P., "Parameter Calibration for PARAMICS Using Genetic Algorithm", 80th Annual Meeting Preprint CD-ROM, Transportation Research Board 01-2399, 2001.
5. Kundé, K. K., "Calibration of Mesoscopic Traffic Simulation Models for Dynamic Traffic Assignment", Master Thesis, Massachusetts Institute of Technology, 2002.
6. Kim, K.O., and Rilett, L.R., "Simplex-based Calibration of Traffic Microsimulation Models with Intelligent Transportation Systems Data", Transportation Research Record: Journal of Transportation Research Board, No. 1855, 2003, pp.80-89.
7. Schultz, G.G., and Rilett L.R., "Analysis of Distribution and Calibration of Car-Following Sensitivity Parameters in Microscopic Traffic Simulation Models", Transportation Research Record: Journal of Transportation Research Board, No. 1876, 2004, pp.41-51.
8. 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.
9. Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H. N., and Wen, Y., " Calibration of Microscopic Traffic Simulation Models: Method and Application", Transportation Research Record: Journal of Transportation Research Board, No. 1999, 2007.
10. 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.
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. Park, B., and Qi, H., "Development and Evaluation of Simulation Model Calibration Procedure", Transportation Research Record: Journal of Transportation Research Board, No. 1934, 2005, pp.208-217.
13. Ma, J., Dong, H., and Zhang, H. M., "Calibration of Micro Simulation with Heuristic Optimization Methods", Transportation Research Record: Journal of Transportation Research Board, No. 1999, 2007.
14. Kim, S.J., Kim, W., and Rilett, L., "Calibration of MicroSimulation Models using NonParametric Statistical Techniques", Transportation Research Record: Journal of Transportation Research Board, No. 1935, 2005, pp.111-119.
15. Molina, G., Bayarri, M. J., and Berger, J. O., "Statistical Inverse Analysis for a Network Microsimulator", 2005 American Statistical Association and the American Society for Quality Technometrics, Vol. 47, No. 4 November 2005.
16. Daganzo, C. F., "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory", Transportation Research Part B. Vol. 28, No. 4, 1994, pp.269-287.
17. Spall, J. C., "Introduction to Stochastic Search and Optimization", Willy-Interscience, 2003.
18. Nelder, J. A., and Mead, R., "A Simplex Method for Function Minimization", The Computer Journal, vol. 7, 1965, pp.308-313.
19. Spall, J. C., "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation", IEEE Transactions on Automatic Control, Vol. 37, No. 3, 1992, pp.332-341.
20. Spall, J. C., "A One-measurement Form of Simultaneous Perturbation Stochastic Approximation", Elsevier Science Ltd. Vol.33, No.1, 1997, pp.109-112.
21. Munoz, L., Sun, X., Sun, D., Gomes, G., and Horowitz, R., "Methodological Calibration of the Cell transmission Model", Proceeding of the 2004 American Control Conference, Boston, Massachusetts, June 30-July 2, 2004.
22. Munoz, L., Sun, X., Horowitz, R., and Alvarez, L., "A Piecewise-Linearized Cell Transmission Model and Parameter Calibration Methodology", Transportation Research Record: Journal of Transportation Research Board, Vol. 1965, 2006, pp. 183-191

23. http://www.mrc-bsu.cam.ac.uk/
24. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., "Bayesian Data Analysis", Second edition, 2004.
25. Skabardonis, A., Petty, K., Varaiya, P., and Bertini, R., "Evaluation of the Freeway Service Patrol (FSP) in Los Angels", California PATH Research Report. UCB-ITS-PRR-98-31. 1998.
26. Daganzo, C. F., "The Cell Transmission Model, Part 2:Network Traffic", Transportation Research Part B. Vol. 29, No. 2, 1995, pp. 79-93.
27. Miller, I., Freund, J. E., and Johnson, R. A., "Probability and Statistics for Engineers" Fourth Edition, 1990.
28. Sanwal, K. K., Petty, K., Walrand, J., and Fawaz, Y., "An Extended Macroscopic Model for Traffic Flow", Transportation Research Part B. Vol. 30, No. 1, 1996, pp.1-9.