Calibration of a Macroscopic Traffic Simulation Model using Bayesian 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
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 USA,
Tel: (732) 445-2792
Fax: (732) 445-0577
E-mail: kaan@rci.rutgers.edu

Abstract: 267
Word count: 4996 Text + 6 Tables + 7 Figures = 8,246
Submission Date: August 1, 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 real-world conditions when its parameters are effectively calibrated. Previous studies on calibration generally focused on minimizing the sum of the relative error between observed data from a certain period of time during a typical day and simulation output for the same period. This static approach can be explained as calibration using 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. In this paper, a new calibration methodology, which is based on the Bayesian approach, is proposed. This new calibration methodology, instead of using a single demand matrix and corresponding observed traffic conditions that represent one specific point in time, uses randomly generated demand matrices and corresponding traffic conditions form an observed distribution of these variables. The goal of using input values generated from the observed distribution of demands is to be able to accurately represent wide-range of all the likely demand conditions observed at a facility. Moreover, at each iteration, the proposed calibration methodology re-estimates optimal parameters using a stochastic optimization algorithm known as Simultaneous Perturbation Stochastic Algorithm (SPSA). A cell transmission based macroscopic simulation model of a portion of the I-880 freeway section in California is calibrated using the proposed methodology. Proposed Bayesian-SPSA algorithm is shown to outperform simple SPSA algorithm based on several case scenarios studied as part of this paper. Future work will focus on the calibration of larger networks and microscopic simulation models using the same methodology.
INTRODUCTION AND MOTIVATION
Traffic simulation models are increasingly being used for the evaluation of complex real-world traffic problems. However, accurately estimating of traffic conditions requires an effective calibration of the simulation model. In most real-world cases, simulation output obtained using default parameters might not always represent observed traffic conditions. Thus, proper selection of input parameters is required to be able to accurately represent 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. Gardes et al. (2002) [3], Lee et al. (2001) [4], Ding (2003) [1] and Ma and Abdulhai (2001) [5] selected mean target headway and mean reaction time as the major calibration parameters for PARAMICS-based microscopic simulation models. Kundé (2002) [2] used speed-density relationship and capacity for the calibration of 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 was found to be better than the Box Complex algorithm in terms of level-of-fit statistics. Lee et al. (2001) [4] and Ma and Abdulhai (2001) [5] 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 (2002) [6] employed the simplex algorithm for the optimizing the degree-of-fit for their models in CORSIM and TRANSIMS. Table 1 consists of a comprehensive summary of some of the most recent calibration studies performed.

Table 1 Summary of Previous Calibration Studies

<table>
<thead>
<tr>
<th>Authors</th>
<th>Simulation Tool</th>
<th>Calibrated Parameters</th>
<th>Optimization Methodology</th>
<th>Type of Roadway Section</th>
<th>Objective Function</th>
<th>Validation Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourdakis et al. [15]</td>
<td>AIMSUN</td>
<td>Global, local parameters</td>
<td>Trial and error</td>
<td>Freeway</td>
<td>Volume</td>
<td>RMSP</td>
</tr>
<tr>
<td>Kim and Rilett [6]</td>
<td>CORSIM, TRANSIMS</td>
<td>CORSIM: Car-following factors, driver’s aggressiveness factor TRANSIMS: O-D matrix, PT1 parameters</td>
<td>Simplex algorithm</td>
<td>Freeway</td>
<td>Volume</td>
<td>MAER</td>
</tr>
<tr>
<td>Park and Qi [18]</td>
<td>VISSIM</td>
<td>Eight parameters</td>
<td>Genetic algorithm</td>
<td>Intersection</td>
<td>Average travel time</td>
<td>N/A</td>
</tr>
<tr>
<td>Ding [1]</td>
<td>PARAMICS</td>
<td>Mean target headway, mean reaction time</td>
<td>SPSA algorithm</td>
<td>Freeway</td>
<td>Flow, density</td>
<td>MAE</td>
</tr>
<tr>
<td>Gardes et al. [3]</td>
<td>PARAMICS</td>
<td>Mean target headway, mean reaction time</td>
<td>N/A</td>
<td>Freeway</td>
<td>Speed, Volume</td>
<td>N/A</td>
</tr>
<tr>
<td>Ma and Abdulhai [5]</td>
<td>PARAMICS</td>
<td>Mean headway, mean reaction time, feedback, perturbation, familiarity</td>
<td>Genetic algorithm</td>
<td>Roadway</td>
<td>Traffic counts</td>
<td>MAE</td>
</tr>
<tr>
<td>Lee et al. [4]</td>
<td>PARAMICS</td>
<td>Mean target headway, mean reaction time</td>
<td>Genetic algorithm</td>
<td>Freeway</td>
<td>Occupancy, Flow</td>
<td>N/A</td>
</tr>
<tr>
<td>Schultz and Rilett [22]</td>
<td>CORSIM</td>
<td>Driver behavior parameters, vehicle performance parameters</td>
<td>Automated Genetic algorithm</td>
<td>Freeway</td>
<td>Volume, Travel time</td>
<td>MAE</td>
</tr>
<tr>
<td>Toledo et al. [19]</td>
<td>MITSIMLab</td>
<td>O-D flow, behavioral parameters</td>
<td>Complex algorithm</td>
<td>Freeway and arterial</td>
<td>Speed, Density</td>
<td>RMSE, RMSP, MAE, MAPE</td>
</tr>
</tbody>
</table>

TRB 2008 Annual Meeting CD-ROM

Original paper submittal - not revised by author.
The majority of previous calibration studies focused on minimizing the sum of the relative error between observed data and simulation output, obtained at one or 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, it is obvious that the same facility will not have identical traffic characteristics for the morning peak of different days of the same week let alone of different weeks and months. Thus, this paper proposes a new and enhanced calibration methodology based on the application of the SPSA stochastic optimization method, in conjunction with Bayesian techniques to produce more effective calibration of traffic simulation models.

Bayesian theory is adopted to capture time-dependent traffic conditions in a better fashion. Molina et al. (2005) [12] was the first researcher who proposed the calibration of a microscopic simulation model developed in CORSIM using Bayesian approach. The authors applied the Bayesian theory to overcome over or under-tuning of parameters and inaccurate data. However, they did not use a stochastic optimization method to find “best” values for calibration parameters.

Thus, this study is the first one to propose an enhanced simulation methodology that combines both the stochastic optimization algorithm and Bayesian analysis techniques. The main purpose of using a Bayesian framework is to use the knowledge of the network conditions (prior distribution) and the likely or representative traffic characteristics (likelihood) to obtain the actual traffic characteristics such as flows, speeds, etc. (posterior distribution). In our problem, the posterior distribution is the output of the Cell Transmission traffic simulation model (CTM), generated as a result of the randomly sampled origin-destination demands and input parameters calibrated in the previous iteration. The objective of the Bayesian analysis is to find the best simulation parameters given the randomly sampled demand values.

This paper consists of (1) a description of the macroscopic simulation model development and proposed methodology of calibration, (2) an explanation of the underlying process calibration procedure and (3) a presentation of the implementation of the proposed methodology, through a case study of calibration and validation procedure using the CTM.

PROPOSED METHODOLOGY FOR THE CALIBRATION OF A MACROSCOPIC TRAFFIC SIMULATION MODEL

Calibration of the parameters of a traffic simulation model can be formulated as an optimization problem where the analytical form of the objective function is unknown. This simulation-optimization problem can be more formally stated as the minimization of the value of the function that is the relative error between the observed and simulated parameters. This function is calculated at each iteration shown in equation [1].

\[
F = \sum_{lane} \sum_{time} \left[ \frac{Q_{real} - Q_{sim}}{Q_{real}} \right]^2 + \left[ \frac{K_{real} - K_{sim}}{K_{real}} \right]^2
\]  

\[\text{RMSE: Root-Mean-Square Error} \quad \text{MAE: Mean Absolute Error} \quad \text{MAER: Mean Absolute Error Ratio} \quad \text{MSA: Method of Successive Averages}\]
Where,

\[ F \]: The objective function

\[ Q_{\text{real}} \]: Observed flows for a given time period

\[ Q_{\text{sim}} \]: Simulated flows for a given time period

\[ K_{\text{real}} \]: Observed density for a given time period

\[ K_{\text{sim}} \]: Simulated density for a given time period

In this study, observed flows and densities are obtained using a macroscopic simulation model based on the cell transmission model proposed by Daganzo (1994) \[7\]. Since our goal in this paper is to introduce a new calibration methodology, we chose a well-accepted yet simple to implement approach for the simulation component of our study. However, it is clear that our 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, or shock waves. CTM limits the flow to the minimum value between the upstream capacity and downstream capacity of the cell \[7\]. The maximum number of vehicles in the under-congested condition is the product of the jam density \( K_j \) and cell length at the cell \( i \), and the maximum number of vehicles in the over-congested 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.

In addition to the objective function not having 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 objective function.

Optimization algorithms can be classified into deterministic and stochastic approximation algorithms, where in the latter case the functional form of the objective is probabilistic. Each algorithm can be further divided into two general categories namely, gradient and gradient free settings. The steepest descent method \[9\] and Newton-Raphson method \[9\] are gradient-based deterministic algorithms. Nelder and Mead (1965) \[14\] proposed the nonlinear simplex algorithm, a deterministic method that is based on a gradient-free multivariate optimization method.

The stochastic approximation of the loss minimization problem can be employed in the presence or the absence of the gradient function \( g(\theta) \). The stochastic root-finding algorithm by Robbins-Monro \[9\] 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, 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 (Simultaneous Perturbation Stochastic Approximation), one of the well-known stochastic approximation algorithms that can be applied in both the stochastic gradient and gradient-free settings, is also applicable to solve optimization problems with a large number of variables.

Previously, Munoz et al. (2005)(2004) \[23\], \[24\] performed the calibration of a cell transmission based model of a portion of Interstate 210 Westbound in Southern California. Free-flow speed and congestion parameters were calibrated using a constrained least-square fit on the flow-density relationship. The difference between simulate and observed total travel times is the objective function used to evaluate the performance of the simulation. However, this model only focused on the deterministic aspects of this calibration problem.
In this paper, we propose to use the Simultaneous Perturbation Stochastic Approximation (SPSA) approach proposed by Spall (1992) in conjunction with a Bayesian framework. Figure 2 shows the flow chart of the proposed combined Bayesian-SPSA (B-SPSA) calibration and validation methodology.

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

1. Increment iteration: iteration=iteration+1
   Iteration:
   a. Generate the OD demand matrix from a probability distribution function of demands developed using real-world data.
   b. Use SPSA algorithm to determine the optimal parameters given the demand matrix generated

2. Compare the likelihood, the output of simulation for the given demand matrix in the current iteration namely flows and densities with the posterior 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 transferability tests are satisfactory then stop. If not, return to Step 1.
Description of the Bayesian Analysis Method

The basic Bayesian relationship is shown in equation [2], where the posterior distribution is acquired from the prior density function and a likelihood function.

\[
\text{posterior} \propto \text{prior} \times \text{likelihood} \tag{2}
\]

Here, the prior 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 OD demand matrices. Each randomly generated demand (\(d_n\), where \(n = 1, 2, \cdots, n\), where \(n\) indicates number of demand matrix), is a random sample of the prior distribution. The likelihood in our case is the optimal input parameter values of free-flow speed and jam density given the specific demand matrix. At each iteration, these input parameters are determined using the SPSA (Simultaneous Perturbation Stochastic Approximation) algorithm.
discussed in the previous section. Based on the product of a prior distribution and the likelihood, the posterior
distribution is obtained from the Cell Transmission based simulation model. This theoretical relationship can be
represented by equation [3].

\[ p(\text{Flow}, \text{Density} \mid V_f, K_{\text{jam}}) \propto p(V_f, K_{\text{jam}} \mid \text{Flow}, \text{Density}) p^*(\text{Flow}, \text{Density}) \]  

We apply Bayesian analysis as part of our calibration methodology to represent the distribution of the observed
traffic characteristics. The main purpose of using Bayesian analysis is to overcome the over- or under-estimation
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,
increased performance of modern computers coupled with efficient computational techniques allow the Bayesian
method to be widely used in a number of areas.

IMPLEMENTATION DETAILS OF THE PROPOSED B-SPSA CALIBRATION METHODOLOGY

Demand Matrix Generation from the Observed Distribution of Demands (Prior Distribution)

In traffic simulation models the vehicular demand is modeled using from traffic zones. In the case of freeway
sections, these zones are usually placed at a section upstream of the section under study. If the study section is
relatively short, the demand generated from the zone will, on an average, be equal to the flow that is observed at the
section 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 section. So, given the flow data at a various
sections on the freeway, the corresponding demand matrices can be generated. These demands can be uniquely
generated if the number of sections is the same as the number of zones. In this study, the demand matrices are
generated using the traffic counts i.e. flows from loop detectors.

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 prior 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 due to 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 are sampled simultaneously formed from the flows. To take into account the existence or lack of correlation
between the two demands, the demand matrix is generated using either conditional or independent samples. Table 2
classifies the possible demand sampling methods depending on the existence or nonexistence of an intermediate
ramp.

<table>
<thead>
<tr>
<th>Demand Sampling</th>
<th>Single Zone</th>
<th>Two Zones (On-ramp scenario)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Correlated</td>
<td>Independent sampling</td>
<td>Independent sampling</td>
</tr>
<tr>
<td>Correlated</td>
<td>Not applicable</td>
<td>Conditional sampling</td>
</tr>
</tbody>
</table>

Determination of the Optimal Parameters that Minimize the Error using the SPSA Algorithm (Likelihood)

Based on the randomly generated demand matrix, the calibration parameters are re-estimated at each iteration. In
order to select the optimum values of the parameters, a stochastic optimization algorithm known as the Simultaneous
Perturbation Stochastic Approximation (SPSA) algorithm is used.

Application of SPSA to the optimization of multivariable equations described in Spall [8], [9], [10] and equation [4]
led to the basic form below:

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

[4]
Where, \( g_k(\hat{\theta}_k) \) is the Simultaneous Perturbation (SP) of the gradient \( g(\theta) = \frac{\partial L(\theta)}{\partial \theta} \) estimated, based on the loss function measurements, at \( \theta = \hat{\theta}_k \) at the \( k^{th} \) iteration.

The new value of \( \theta \) 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 \( \theta \) as shown in equation [4]. With simultaneous perturbation, loss measurements are obtained by randomly perturbing the elements of \( \hat{\theta}_k \). Assuming that \( \hat{\theta}_k \) is \( p \)-dimensional, the SP gradient approximation step can be shown in the following form.

\[
\hat{g}_k(\hat{\theta}_k) = \begin{bmatrix}
\frac{\hat{L}(\theta + c_k \Delta_k) - \hat{L}(\theta - c_k \Delta_k)}{2c_k \Delta_{k1}} \\
\cdot \\
\cdot \\
\frac{\hat{L}(\theta + c_k \Delta_k) - \hat{L}(\theta - c_k \Delta_k)}{2c_k \Delta_{pk}}
\end{bmatrix}
\]

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

[5]

Where, the \( p \)-dimensional random perturbation vector, \( \Delta_k = [\Delta_{k1}, \Delta_{k2}, ..., \Delta_{kp}]^T \) is a user-specified vector for which the components of \( \Delta \) are usually \( \pm 1 \) Bernoulli variables. Here, \( c_k \) is a positive scalar.

CTM simulation is then executed multiple times with different random seeds because of the consideration of the variability in the simulation at each step of the SPSA algorithm. The random seeds are an important factor that is studied to determine its effect on the simulation results. In the stochastic analysis, the random seeds produce more accurate results by representing the measurement noise. Thus, it is important to run the simulation with several different random seeds in order to minimize the variance of the error. At the end of each simulation run the relative error between the observed and simulated is calculated using equation [1].

Perform Kolmogorov-Smirnov (K-S) Test to Compare Histograms based on the Simulation Output of the Optimized Input Parameters (Posterior Distribution) with Histogram of the Observed Values

Random samples from the prior distribution are generated from the observed demand matrix. The SPSA algorithm for each generated demand matrix calibrates these parameters. If one set of the calibration procedure is terminated and the optimized parameter values are found by the SPSA algorithm, the evaluation process is then performed. The distribution of the cell transmission output is compared with the distribution of the observed values using the Kolmogorov-Smirnov test. If the two distributions are comparable, the procedure moves to the validation step. However if they do not match, the sampling process restarts. At this point, the distribution of each of the calibrated parameters becomes the likelihood of posterior distribution.

CASE STUDY

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

The data is obtained from the database of the Freeway Service Patrol project for a portion of the I-880 freeway on Hayward, California [11]. The data is collected from September 27 to October 29, 1993 from 5:00 AM to 10:00 AM and from 2:00 PM to 8:00 PM. The data is used differently depending on the existence or nonexistence of intermediate ramps. When intermediate ramps are present, the data is divided into 4 sections: depending on morning or evening, and peak or non-peak period (5:00 AM~7:30 AM, 7:30 AM~10:00 AM, 2:00 PM~5:00 PM, and 5:00 PM~8:00 PM).

Cell Transmission Model and Selection of Input Parameters

In the case of one zonal demand (no intermediate ramp), the road sections in the cell transmission are initially modeled as a simple road section to test the effectiveness of the Bayesian approach. Secondly, the case of considering intermediate ramp is modeled as an extended road section (Figure 3). Because the zonal demand, one of the major input variables, is very sensitive to the flow-density relationship, we selected the Free-flow speed ($V_f$) and the jam density ($K_j$) as the input parameters.

Demand Matrix Generation from the Observed Distribution of Demands

In case of existence of a ramp, the demand matrix (formed from flows) is generated depending on whether or not the flows from the two zones are correlated to each other (Table 2). Our main assumption is that there is no correlation between these flows, which allows us to sample the mainline and ramp flows independently. Figure 4 shows the histogram based on the distribution of the observed demand matrix formed from the flows.
Determination of the Optimal Parameters that Minimize the Error using the SPSA Algorithm

No Intermediate Ramp - For the single zonal demand case (no intermediate ramp), the initial values of free-flow speed and jam density are taken as 55 miles per hour and 110 vehicles per mile, respectively. The SPSA algorithm is used as the stochastic optimization approach to determine the “optimal” values of the calibrated simulation parameters for this iteration. This 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 to one point where there is 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. Best fit of the flow and density is observed for optimized input parameters of 59.6 miles per hour and 107.0 vehicles per mile respectively. The sum of relative error is determined to be 4.6%.

Intermediate Ramp - For the two zonal demand case (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 miles per hour and 60 vehicles per mile for four different time periods. The best fit of the flow and density is observed when free flow speed and jam density are 58.6 miles per hour and 59.1 vehicles per mile during the time period 14:00 PM ~ 17:00 PM. Table 3 summarizes the results for optimal calibration parameters.

Table 3 Optimal Calibration Parameters and other statistics

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Optimal Free flow speed (mph)</th>
<th>Optimal Jam density (vpm)</th>
<th>Sum of Square Error of Flow</th>
<th>Sum of Square Error of Density</th>
<th>Sum of Total Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:00 ~ 7:30</td>
<td>59.4</td>
<td>60.7</td>
<td>2.01</td>
<td>2.70</td>
<td>4.71</td>
</tr>
<tr>
<td>7:30 ~ 10:00</td>
<td>58.7</td>
<td>61.3</td>
<td>2.02</td>
<td>2.04</td>
<td>4.06</td>
</tr>
<tr>
<td>14:00 ~ 17:00</td>
<td>58.6</td>
<td>59.1</td>
<td>1.83</td>
<td>2.14</td>
<td>3.97</td>
</tr>
<tr>
<td>17:00 ~ 20:00</td>
<td>59.3</td>
<td>60.7</td>
<td>2.53</td>
<td>1.90</td>
<td>4.43</td>
</tr>
</tbody>
</table>

Perform Kolmogorov-Smirnov (K-S) Test to Compare Histograms based on the Simulation Output of the Optimized Input Parameters with Histogram of the Observed Values

The posterior distributions of flow and density are obtained using the cell transmission model as the product of the prior distribution (demand distribution) and the likelihood (optimal input parameter values of free-flow speed and jam density). The obtained flow and density distributions of the macroscopic simulation output based on the optimized input parameters are then compared with the observed distributions using the Kolmogorov-Smirnov (K-S) test. The K-S test can be applied for any of shape of distributions [25]. The null hypothesis states that simulated flow and density distributions are not statistically different from the distributions of observed flow and density values respectively. 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 K-S values from the K-S table are greater than the K-S values for flow and density distributions in Table 4. For the two simple road sections scenario, the null hypothesis could not be rejected so there is no reason to doubt the validity of the null hypothesis, which states that the simulation flow and density distributed are not different from the actual observed distributions. Figure 5 shows the distribution of simulated flow and density values when the optimized values of input parameters are used.

<table>
<thead>
<tr>
<th>Time</th>
<th>K-S Value of Flow</th>
<th>K-S Value of Density</th>
<th>Critical Value from K-S Table for Flow (95 %)</th>
<th>Critical Value from K-S Table for Density (95 %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:00~07:30</td>
<td>0.096</td>
<td>0.089</td>
<td>0.340</td>
<td>0.544</td>
</tr>
<tr>
<td>07:30~10:00</td>
<td>0.038</td>
<td>0.063</td>
<td>0.389</td>
<td>0.453</td>
</tr>
<tr>
<td>14:00~17:00</td>
<td>0.040</td>
<td>0.032</td>
<td>0.340</td>
<td>0.453</td>
</tr>
<tr>
<td>17:00~20:00</td>
<td>0.019</td>
<td>0.025</td>
<td>0.340</td>
<td>0.453</td>
</tr>
</tbody>
</table>
Verification and Transferability Tests

The verification test is 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 test is based on whether the objective function can satisfy the pre-determined stopping criteria. The result of this verification test and the performance of the simulation tool also both show that the real world conditions can be fairly captured by the calibrated simulation model.

When the validation and verification processes were satisfied, the transferability test of calibrated parameter values was performed for different days during the same time period and for the same network. It is difficult to compare the calibrated model performance with similar performance results found in other calibration papers. However, the Mean Square Variation (MSV) that Sanwal et al. (1996) [17] used is a good method to compare the degree of deviations from observed values. Sanwal et al. (1996) [17] obtained 0.945 for the variation of speed and 0.968 for the variation of travel time when he applied the optimized parameters from one day to another.

The optimized parameter values were used for simulating 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 for simulating 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 Kolmogorov-Smirnov test, the values of the flow distributions for September 30, 1993 and October 13, 1993 were 0.019 and 0.018 that are 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 shown to have an acceptable level of similarity (fit) with respect to each other.
Intermediate Ramp - The transferability test is performed for three different scenarios. The first scenario used the B-SPSA approach for four randomly selected days. The second and third scenarios tested the SPSA algorithm without the Bayesian approach for single and multiple days. These tests are performed to determine the effectiveness of the B-SPSA approach as compared with the SPSA-only approach. Finally, the Mean Square Variation (MSV) measure used from Sanwal et al. (1996) [17] paper is applied as an evaluation criterion to compare the degree of deviations from the observed values. The mean square variation is calculated by subtracting the mean square error from one. If the model’s estimations are close to the real-world measurements then the MSV should have been close to one.

First Scenario - Table 3 shows the optimized parameter values of four different time periods which match well with the distributions of observed values shown in Table 4. The optimized parameter values were used for 4 randomly selected days that were not used in the calibration process.

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

For all the periods of the day, each K-S value is less than the K-S critical values given by the table. This result also shows that two distributions between observed and simulated flow and density fit closely. Figure 7 shows the fitness between observed and simulated distributions of flow and density for each time period.
Second Scenario—In the second scenario, the transferability test is performed for the calibration process based on the SPSA-only algorithm. Calibration is performed using a single day’s data until the error was within the bounds. September 30, 1993 data is tested to determine optimal parameter values and the resulting relative errors are calculated as 2.13, 2.66, 3.51, and 4.56. For three randomly selected days, the error is calculated using these optimized parameter values. In the case of the third time period, 14:00 PM ~ 17:00 PM, the relative error ranges 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 6. Based on the result of the SPSA-only algorithm, flow and density distributions for the time
period of 07:00 AM ~ 10:00 AM on September 13, 1993 are compared with observed data distributions using the K-S test. The critical value of K-S test for both flow and density distributions are 0.272 that 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>
<thead>
<tr>
<th>Days</th>
<th>Sum of the square Error of Flow and Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/30/93</td>
<td>2.13</td>
</tr>
<tr>
<td>10/05/93</td>
<td>8.25</td>
</tr>
<tr>
<td>10/13/93</td>
<td>7.36</td>
</tr>
<tr>
<td>10/18/93</td>
<td>5.73</td>
</tr>
<tr>
<td>Average</td>
<td>5.87</td>
</tr>
</tbody>
</table>

Third Scenario - In the third scenario, the transferability test is performed for multiple days for the same time period in different days and the relative errors of 4.58, 4.82, 4.19, and 4.88 are obtained. The optimized parameter values based on SPSA algorithm are used for four randomly selected days and for four different time periods during these days, the relative error increased to 10.42, 16.21, 17.58, and 15.52. Using 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, we conclude that the optimal parameter values from the SPSA algorithm are not always transferable to data sets from the same time periods of different days.

Based on the variation approach proposed in Sanwal et al. (1996) [17], the mean square variation of flow of 4 randomly selected days is compared with the simulation results obtained from the simulation model calibrated using the Bayesian-SPSA algorithm. From the morning period between 05:00 AM ~ 07:30 AM to evening period between 17:00 PM ~ 20:00 PM, the mean square variations of flow are 0.968, 0.979, 0.981, and 0.950. The results of degree of deviation are quite comparable with measured data when Sanwal et al. (1996) [17] method is 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, which is based on the Bayesian approach, is proposed. This new calibration methodology, instead of using a single demand matrix and corresponding observed traffic conditions that represent one specific point in time, uses randomly generated demand matrices and corresponding traffic conditions form an observed distribution of these variables. The goal of using input values generated from the observed distribution of demands is to be able to accurately represent wide-range of all the likely demand conditions observed at a facility. Observed demand values are used to determine a distribution of observed demands for a time period of 17 days.

At each iteration of the proposed Bayesian 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 re-estimates optimal parameters using a stochastic optimization algorithm known as Simultaneous Perturbation Stochastic Algorithm (SPSA) and distributions of macroscopic simulation’s output are compared with the distribution of the observed flow and density [13].

A cell transmission based macroscopic simulation model of a portion of the I-880 freeway section in California is calibrated using the proposed methodology. Two relatively simple road sections 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 cases studies. The cell transmission output i.e., the posterior distribution is obtained by loading demand matrices randomly sampled from the distribution of observed demand called a prior distribution and from the calibrated input parameters called likelihood that are the free-flow speed and jam density. The posterior distribution namely, simulation output, is compared with the observed data distribution using the Kolmogorov-Smirnov test (K-S). The null hypothesis for the K-S states 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 transferability test with a single zonal demand was performed over two randomly selected using the same time period and network. Based on the optimized input parameters, the flow and density distributions were compared. The differences between the observed and simulated flow and density distributions are found to be statistically insignificant at the 95% confidence level. Thus, these parameters are determined to be transferable. As an extended simulation model, a road segment with an intermediate ramp is modeled using the cell transmission approach. The sum of the square errors for four different time periods (5:00~7:30, 7:30~10:00, 14:00~17:00, and 17:00~20:00) is found to be 4.71, 4.06, 3.97, and 4.43, respectively (Table 3). For three different scenarios, the model parameters are tested for transferability and for the effectiveness of the distribution based calibration approach, rather than using single or few points in time when performing the calibration. From Table 5, results of four different time periods are found to satisfy the statistical test for accepting the null hypothesis, and it is concluded that the Bayesian approach satisfied constraints when the parameter values calibrated by using Bayesian-SPSA are applied to data from a different day.

On the other hand, in the second and third scenarios where SPSA is employed without the Bayesian approach as the main calibration algorithm, pre-determined transferability constraints are not always satisfied. According to Table 6, none of the average values of the sum of the square error of flow and density values satisfied the constraint of an acceptable relative error of 5%. In addition, in the scenario where parameters are calibrated using the SPSA-only methodology, the difference between the distributions of simulated flow and density values and distributions of observed values could not be shown to be statistically insignificant when the K-S test is performed. Therefore, B-SPSA method is found to improve the results of simulation calibration by accurately capturing a wide range of real-world conditions. Testing of B-SPSA for larger networks, as well as for microscopic traffic simulation 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 B-SPSA calibration methodology. Proposed Bayesian-SPSA algorithm is shown to outperform simple SPSA algorithm based on several case scenarios studied as part of this paper.

REFERENCES