Calibration of Micro Simulation with Heuristic Optimization Methods

Jingtao Ma
Ph.D Candidate
University of California, Davis
Dept. of Civil & Environmental Engineering
One Shields Avenue, Davis, CA 95616
Tel: 530-754-6429
Email: jtma@ucdavis.edu

Hu Dong
Perteet, Inc
2707 Colby Ave, Suite 900
Everett, WA 98201
Tel: (425)252-7700
Email: hdong@perteet.com

H. Michael Zhang*
Professor
University of California, Davis
Dept. of Civil & Environmental Engineering
One Shields Avenue, Davis, CA 95616
Tel: 530-754-9203
CKP Professor
Tongji University,
School of Transportation Engineering
1239 Siping Road
Shanghai, China 200092
Email: hmzhang@ucdavis.edu

Submitted to AHB45 (1), TRB Committee on Traffic Flow Theory and Characteristics, Simulation Sub-Committee

* Corresponding author

ABSTRACT

Model calibration is a crucial step in building a reliable microscopic traffic simulation application, because it serves as the additional check to ensure the model parameters accurately reflect the local driving environment, such that decisions made based on these results would not be misinformed decisions. Because of its stochastic nature and complexity, the calibration problem, usually formulated as an optimization problem, is often solved using heuristic methods. To-date, calibration is still a time-consuming task because many of the adopted methods require many simulation runs in search of an optimal solution. Moreover, many aspects of the calibration problem are not fully understood and need further investigation. In this study, we develop another heuristics calibration algorithm based on the simultaneous perturbation stochastic approximation (SPSA) scheme, and applied it to calibration several networks coded in Paramics. Our results indicate that the new heuristic algorithm can reach the same level of accuracy with considerably less iterations and CPU time than other heuristic algorithms such as the genetic algorithm (GA) and the trial-and-error iterative adjustment (IA) algorithm. Applications of all three heuristic methods in a northern California network also reveal that some model parameters affect the simulation results more significantly than others. These findings can help modelers better choose calibration methods and fine tune key parameters.

INTRODUCTION

Model calibration is the process that adjusts model parameters so that the simulated measurements (e.g., traffic volumes and speeds, path travel times) can match the field observed ones. This process is crucial to all simulation models, micro simulation in particular, because it serves as the check to ensure that the simulation models represent the real world traffic *correctly* and that misinformed decisions based on the results from an ill-calibrated model can be avoided. Unfortunately, calibration is also a time-consuming task, since micro simulation models are generally high fidelity and can easily have hundreds of parameters. It is not uncommon that even a medium size network [1] could take months to build, and most of the work time was consumed in calibrating these parameters [2]. Moreover, both the temporal and spatial scope of micro simulation has been considerably expanded to go beyond a freeway section [3, 4] or an isolated location such as an interchange or an intersection [5] to include corridor networks [6-8] and city and region wide transportation networks over 520 km^2 [9, 10]. In this trend, calibration task becomes an even more demanding task. Therefore, researchers begin to carry out systematic studies on how to calibrate the micro simulation models efficiently.

Documented calibration efforts have usually formulated this problem as an optimization problem, where the optimal set of model parameters is sought to minimize the objective function, i.e., the measurement of the "distance" between simulation results and real world traffic measurements. Because of the stochastic nature of micro simulation and the complex relations between its parameters and its outcomes, it is difficult to compute the gradient information. Mathematical programming methods are thus not applicable. Researchers resort to heuristic methods to search optimal parameter values. For example, genetic algorithm [3, 4, 11, 12 and 13], simulated annealing [14], Complex algorithm [15] and even trial-and-error enumeration [16] have been reported in literature. These methods automate the calibration process to a certain degree and it was generally reported that they improve simulation performance over the default model parameter values.

Yet several important questions remain unanswered. Micro simulation is a complex system that all parameters work together to influence its modeling results. In calibrating such a complex model, users could get trapped in a never-ending process of adjusting parameters at one place only to find problems popping up at others. To overcome this difficulty, Zhang et al [17] proposed a five-step procedure to divide the parameters into distinct groups that each can be calibrated individually with a certain order, so as to minimize the mutual interactions between different sets of parameters. For example, four major groups with relatively clear distinctions from each other, namely global driving behavior model parameters, local driving behavior model parameters, departure time and route choice model parameters and demand input [17] are identified. This calibration logic has also been implicitly applied in [18, 19]. Even within each group, however, it is still unknown whether there are key parameters that affect modeling results more significantly than others, and if the answer is yes, which are the key parameters. There are also lingering questions pertaining to optimization methods. Even though all heuristics methods were reported to obtain improved calibration results, did they get the "true" optimal set of parameter values? Which method can obtain similar or better calibration results with much less computational time? The latter is important to know as micro simulations are applied to larger and larger networks. Answers to these questions will help the analysts carry out model calibration more efficiently and reliably.

This study attempts to answer the above questions through calibrating the driving behavior model parameters in Paramics. We first introduce the simultaneous perturbation stochastic approximation (SPSA) method [20] that has been used successfully in other fields [21]. This heuristic method differs

from others in that it does not rely on evaluating a large pool of feasible solutions when updating the searching direction at every iteration. Meanwhile, the search direction is along the gradient approximated at each iteration. We then implement the genetic algorithm (GA) and trial-and-error iterative adjustment method to calibrate the same set of parameters and compare them with the SPSA results. Next, we apply these three methods to the calibration of a northern California network, which sheds lights on some of the questions raised above. Finally, we discuss the application guidelines of these heuristic calibration methods.

HEURISTIC CALIBRATION METHODS

Heuristic methods start with a feasible parameters' set(s) and evaluate the closeness between the field measurements and simulation under these feasible set(s). Based on rules unique to every method, unfit parameter sets will be replaced by better ones. This process is carried out iteratively until the gap is narrowed to an acceptable level. Three heuristics methods are introduced in this study, namely Simultaneous Perturbation Stochastic Approximation (SPSA), Genetic Algorithm (GA), and trial-and-error Iterative Adjustment (IA). Since the last two are widely applied, they are only briefly presented here. Our major coverage is on the SPSA method.

The Simultaneous Perturbation Stochastic Approximation (SPSA) Algorithm

The introduction and formulation of the SPSA method draws mainly on the theoretical work from [20-22].

The simultaneous perturbation stochastic approximation (SPSA) method works in the following way. For a system, the general objective function $L(\theta)$ is a scalar-valued performance measure, and θ is a continuous-valued *p*-dimensional vector of the control parameters that can be manipulated to achieve a better system performance. In the micro simulation calibration context, θ would be the vector of selected parameters to be calibrated. It is common that a noise ε could occur when observing $L(\theta)$, that is, the observation $z(\theta)$ would be:

$$z(\theta) = L(\theta) + \varepsilon \tag{1}$$

Assuming $L(\theta)$ is differentiable over θ and the minimum is obtained at a zero point of the gradient, i.e.,

$$g(\theta) = \left| \frac{\partial L(\theta)}{\partial \theta} \right|_{\theta=\theta^*} = 0 \tag{2}$$

With an initial guess θ_0 (e.g., the default parameter values in the simulation software), SPSA method applies a series of "simultaneous perturbation" over the successive steps until the approximation of the gradient $g(\theta)$ converges to zero *almost surely* (*a.s.*), under several regularity conditions. The readers are referred to [20] for the theoretical development of the regularity conditions. It is observed that for most engineering problems these conditions are almost automatically satisfied [20] with only one exception: the fitness value function is restricted from becoming excessively large in the calibration context. This would imply that the simulated results cannot represent the real world traffic at all when replacing the

default parameters in simulation with the estimated parameters (θ_k). Since it could be avoided by restricting certain feasibility ranges, this would be unlikely to occur in the calibration process.

Along the successive steps, θ_k is updated recursively in the standard form:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}(\hat{\theta}_k) \tag{3}$$

where the gain sequence $\{a_k\}$ also needs to satisfy the regularity conditions, and $\hat{g}(\bullet)$ is the derived gradient at each step.

The perturbation is performed upon deriving $\hat{g}(\hat{\theta}_k)$. First define a *p*-dimensional mutually independent mean-zero random variable vector $\Delta_k \in R^p = \{\Delta_{k1},..., \Delta_{kp}\}$ that is also independent of the estimated sequence of θ_k . The expectation of the inverse or higher-order inverse of each component of Δ_k must be bounded, i.e., $E(|\Delta_{ki}^{-2}|) \le \alpha_2$ for some constant α_2 . An optimal distribution of Δ_k is symmetric Bernoulli: $P(\Delta_{ki} = \pm 1) = \frac{1}{2}$ [23]. Let

$$z_k^{(+)}(\theta_k) = L(\hat{\theta}_k + c_k\Delta_k) + \varepsilon_k^{(+)} \tag{4}$$

$$z_k^{(-)}(\theta_k) = L(\hat{\theta}_k - c_k\Delta_k) + \varepsilon_k^{(-)} \tag{5}$$

where c_k is a positive scalar (A1), and $z_k^{(+)}(\theta_k)$, $z_k^{(-)}(\theta_k)$ are the measurements of the system under the perturbation $\hat{\theta}_k + c_k\Delta_k$, $\hat{\theta}_k - c_k\Delta_k$, respectively. The approximated gradient will read:

$$\hat{g}_k(\hat{\theta}_k) = \frac{z_k^{(+)} - z_k^{(-)}}{2c_k} \begin{bmatrix} \Delta_{k1}^{-1} \\ \vdots \\ \Delta_{kp}^{-1} \end{bmatrix} \tag{6}$$

Spall [20] shows that by recursively updating θ_k , the gradient will converge to zero that implies a local optimum, since it is unlikely that the approximation would settle down at a maximum or a saddle point because of the stochastic nature of the algorithm.

The gain sequences of a_k and c_k generally take the form of power functions:

$$a_k = \frac{a}{(1+A+k)^\alpha}, \qquad c_k = \frac{1}{(1+k)^\gamma} \tag{7}$$

where k is the iterator, and A is a constant introduced to stabilize the optimization process.

The above SPSA procedure is suitable for unconstrained optimization, and it has to be adapted to accommodate the constraints described above, i.e., imposing a lower and an upper bound for each component (the so-called "box" constraints). Sadegh [22] proposed a projection method to restrict $\theta_k \in R^p$ at iteration k to fall in the feasibility range. It simply replaces any violating $\hat{\theta}_k$ with the nearest $\theta_k \in G(\theta)$, where $G(\theta)$ is the feasibility set of the parameters to be calibrated:

$$\hat{\theta}_{k+1} = P(\hat{\theta}_k - a_k \hat{g}_k(\hat{\theta}_k)) \tag{8}$$

The perturbed vectors $\hat{\theta}_k \pm c_k \Delta_k$ in evaluating the system performances will also be projected into the feasibility region in the same manner. By enforcing an additional regularity condition (Proposition 1 in [22]) over the constraints, constrained SPSA is still able to converge to a Karash-Kuhn-Tucker point (zero-gradient as one of the necessary conditions) *a.s.*

Based on the above constrained SPSA method, we develop a SPSA calibration algorithm as follows:

SPSA Algorithm for Micro Simulation Calibration

Step 1: Initialization and Selection of Algorithmic Coefficients.

1.0 Set iterator $k = 0$*;*

1.1 Select the set of parameters to be calibrated as θ *and normalize it;*

1.2 Pick an initial feasible solution of θ_0 *(e.g., default values in the simulation software);*

1.3 Select nonnegative algorithmic parameters $\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{A}, \alpha$ *and* γ*.*

Step 2: Simultaneous Perturbation.

Generate a p-dimensional random perturbation vector Δ_k*, where each component is mutually independent Bernoulli* ± 1 *distributed with probability of ½ for each* ± 1 *outcome.*

Step 3: Loss Function Evaluation by Running Simulation with Perturbed Parameters.

3.1 Perturb the vector $\hat{\theta}_k$ *with* $\pm c_k \Delta_k$ *as in (4-5);*

3.2 Project the perturbed vectors onto $G(\theta_k)$ *from (7);*

3.3 Evaluate the calibration performance by running simulation with perturbed parameter sets obtained in (4-5).

Step 4: Compute the Approximate Gradient.

Calculate the approximated gradient from (6).

Step 5: Parameter Update.

Update $\hat{\theta}_k$ *with (8).*

Step 6: Check convergence.

Check if the maximum number of iterations has been reached or convergence criterion is met. If yes, stop. If not, set $k = k + 1$ *and go to step 2.*

Genetic Algorithm (GA)

Genetic algorithm (GA) is a popular calibration method for micro simulation, and has been shown to obtain near-global optima (e.g., [3-4, 11-13]). We refer the readers to [24] and [17] for an introduction of GA as well as detailed guidelines for calibration applications.

Trial-and-error Iterative Adjustments (IA)

The trial-and-error iterative adjustment method first enumerates the feasible solutions by dividing the feasible region into equal intervals and picking a value from each interval, then runs the simulation based on combinations of selected parameter values, often one parameter at a time. One can make the intervals smaller to increase the precision. This process continues until both precision requirements and the performance target are met. This method is simple and easy to apply. Thus, many calibration efforts ([15, 16, 25]) rely on trial-and-error to find a suitable set of model parameters. However, the choice of the feasible range and incremental steps of each parameter is quite *ad hoc*, often relying on the analyst's modeling experience and judgment.

DRIVING BEHAVIOR MODEL PARAMETER CALIBRATION

Global and Local Model Parameters

A reliable driving behavior model in a micro simulation ensures that local travelers' car-following, lane-changing and merging behavior can be realistically represented in the simulation. Following the framework in [17], the driving behavior model parameters are categorized into two related groups: 1) global parameters that affect driving behavior throughout the network, 2) local parameters that are peculiar to bottleneck locations, such as lane-drop locations or junctions where several roads meet, e.g., on-ramp merging sections or intersections.

Global model parameters and local model parameters are calibrated separately in this study. This idea is akin to highway capacity analysis, where one first identifies a set of ideal conditions and the ideal capacity under such conditions, then adjusts the ideal capacity for non-ideal conditions through discount factors to obtain the prevailing capacity. Similarly, we want to identify typical road sections for the calibration of global parameters, and road sections with special features (such as sharp curvatures, lane drops, on-ramps and intersections) for the calibration of local model parameters. Through such calibration, we want to obtain a set of parameters that can reproduce the flow capacities of various types of road sections.

Different simulation packages have their own underlying driving behavior models and corresponding parameters. The Paramics package (V5) has ten such parameters, as shown in Table 1[26].

(TABLE 1)

The Calibration Procedure

(Figure 1)

Figure 1 shows global and local parameter calibration in more detail. The process begins with extracting/creating a sub-network from the entire simulation network that has been coded and checked for coding errors. A typical road section is first selected at one location where no capacity affecting road features occur in the vicinity. That is, a sub-network consists of a typical road section with no special features such as i) on-ramps or off-ramps, without lane addition or lane drop; ii) sharp horizontal curves;

iii) significant vertical grade and iv) operational restrictions. One then performs calibration to obtain an optimal set of global parameter values. These values will be used as default throughout the network. Then for each bottleneck location, one extracts a sub-network and calibrates its local parameters against the bottleneck capacity and other selected traffic flow characteristics (such as the shape of the flow-density diagram). The process is complete when all bottleneck locations are calibrated.

Calibration target: Link Capacities

Various measurements have acted as calibration targets. A conveniently applicable measurement is the flow profile, where the 30-second or 1-minute flow rates at adjacent detector locations shall be matched in the simulation. This has been used in [3-4] for driving behavior model calibration. However, most networks may not have well-placed detector stations suitable for this approach, and finding such a place appropriate for global parameter calibration is even harder. Moreover, since the goal of driving behavior model calibration is to ensure the simulated roadway has the same maximum flow rates as in the field, a fundamental diagram (FD) approach is selected to calibrate the link capacities.

The FD approach is based on capacity and the shape of the flow-occupancy diagram (fundamental diagram) [27]. If the car-following model is able to reflect reality, field observed capacities of those sections as well as critical occupancies when the capacities occur should be closely replicated. As capacity and critical occupancy are not influenced by traffic volumes, an accurate O-D demand matrix is not necessary. Therefore, it is possible to use an artificial demand matrix in this approach.

In order to identify capacity and critical occupancy, one needs to increase traffic volume gradually so that a peak is clearly created. The goal is to obtain the top portion of the fundamental diagram for capacity and critical occupancy calculation (see Figure 6). Paramics provides a demand factor, which globally adjusts volume between each O-D pair by a certain percentage ranging from 0% to 200%. This factor is used to create the demand fluctuations for generating the shape of fundamental diagrams under different parameters in calibration. Then the maximum flow rate and the corresponding critical occupancy are estimated and compared with their counterparts obtained from field data. Readers can get the implementation details from [17].

The closeness between the simulated capacities and field observed ones is measured using the following fitness function:

$$F = \sum_{i=1}^{M} [GEH(Cap_i) + A \times GEH(Occ_i)] \quad (9)$$

Where:

M: number of data collection locations

Cap_i: capacity of all general purpose lanes in one direction on which the data collection location i is located;

Occ_i: critical occupancy of a link on which the data collection location i is located.

A: a weighting factor; in general, the GEH values of occupancy are found to be one magnitude lower than those of capacity, the value of A is chosen to be 10 in this work;

and *GEH* is a statistics by the British engineers [28] that reads:

$$GEH = \sqrt{\frac{(V_p - V_m)^2}{(V_p + V_m)/2}} \tag{10}$$

where V_p = value predicted by the model and V_m = value measured in the field. Note that the perfect match will result in a zero of the *GEH* value.

APPLICATION IN A NORTHERN CALIFORNIA NETWORK

State Route 99(SR-99) in the city of Sacramento, California is a congested corridor [17]. A network has been built to study the high occupancy vehicle (HOV) lanes along a 24-mile section of freeway in southbound, starting from MP (milepost) 24.35 to the county line. The study period spans one hour of PM peak, from 2:30 to 3:30. Because of the closely located interchanges and the existing HOV facilities, the driving behavior model is particularly important for the simulation performance. This network is then chosen to test various heuristics calibration methods; furthermore, a few PeMS loop detectors [29] provide the traffic measurements necessary for calibration needs, including flow rates and occupancies in a resolution as fine as 30 seconds.

The site for global calibration is selected based on those guidelines discussed above: a straight road section with no drastic changes in geometric features. The network is examined and a section between Florin Road and Mack Road meets the needs. This section is 2-mile long with two general purpose lanes and one HOV lane. A sub-network is then constructed with this section.

Global Model Parameter Calibration Results

The search space for global parameters calibration includes four dimensions: mean target headway (MTH), mean reaction time (MRT), driver aggressiveness (AGGR) and driver awareness (AWAR) (TABLE 1). Through an ordinary division of each dimension, e.g., a resolution of 0.02, the total number of feasible solutions can easily reach near 5 million. While enumerating all feasible solutions and then selecting the best set becomes impractical using the trial-and-error IA method, genetic algorithm (GA) and SPSA algorithm can generally obtain an optimal solution in much fewer number of iterations. For example, the genetic algorithm took only 600 simulation runs (population 30 times generation 20) to converge to a local optimal solution. Naturally one would wonder how good this solution is. Thus a trial-and-error IA process that searches exhaustively a reduced solution space (using a coarser division) is conducted to benchmark the calibration results.

Two separate exhaustive searches were carried out on the SR-99 network [17], one enumerating MTH and MRT with fixed mean driver aggressiveness and mean driver awareness, and the other, vice versa. For simplicity, herein we only report the first one. The first exhaustive search keeps aggressiveness and awareness unchanged, and enumerates MTH from 0.6 through 2.1 and MRT from 0.6 through 1.8, with the increment of 0.02 for each parameter. The results of IA exhaustive search are shown in Figure 2 and Figure 3.

(Figure 2)

Figure 2 visualizes the changes of the fitness value (FV) under various MTH-MRT combinations. First one can notice numerous local optima marked by small "valleys", denoting the lowest FV within the close vicinity of the local minimizers. Even though these local optimal could be caused by the stochasticity of micro simulation rather than the changes of MTH-MRT pairs, this feature adds to the difficulty of seeking the global optima. Second, one can also notice that a certain range of MRT values produced a similar level of FVs, marked by a big "valley". In the contour plot of Figure 3, correspondingly, a downward-bending band of low FVs is clearly shown, and the global optimal MTH-MRT combinations falls in the center of the band (the intersection of the two white dashed lines). These two figures indicate that in this network context, the simulation performance is more sensitive to the changes of MRT than those of MTH; and similar link capacities can be obtained from a certain range of the combinations.

(Figure 3)

(Figure 4)

Both GA and SPSA are run on the same reduced search space. Their FV convergence processes are shown in Figure 4 and the results are summarized in TABLE 2. Generally, both GA and SPSA can reach stable solutions that are very close to the global optimum obtained from the exhaustive enumeration. However, the fitness values in the SPSA calibration process show a quick drop during the first few dozens of iterations. Its FV drops to an acceptable value (lower than 2) in less than 50 simulation runs. Because the overwhelming proportion of computation time is spent on simulation runs, the SPSA calibration algorithm can obtain an acceptable solution in significantly less time (TABLE 2). But GA manages to reach a better solution in this context.

(TABLE 2)

Calibration of Local Driving Behavior Parameters

Based on the global parameters calibrated above, local driving behavior model parameters are calibrated subsequently. In the SR-99 network, the SR-99 southbound section between Fruitridge Road and Mack Road, including the Florin Road / SR-99 interchange, has three general purpose (GP) lanes to the north of the interchange but only two GP lanes to the south. The lane drop section, two on-ramps and two off-ramps, create frequent merging and weaving maneuvers. This section is thus selected for local parameter calibration [17]. The demand for this sub-network is obtained via the OD Estimator in the Paramics suite [30].

Calibration Results of Local Parameters

One bottleneck link (lane drop) and two onramps in the sub-network generate a total of 15 parameters to be calibrated, and the calibration target is the truncated capacity of the link upstream of the bottleneck. The target link has one PeMS detector group (VDS 312513) that provides flow and occupancy data in five-minute intervals.

The trial-and-error IA process becomes impractical in this context because of the large search space. Only GA and SPSA calibration algorithms are implemented and compared, shown in Figure 5 and TABLE 3. In Figure 5, the FVs from SPSA again show a quick drop during the first few evaluations. However, it

becomes very oscillatory during the remaining process. A close examination of the successive solutions indicate that only very minor changes occur from one iteration to the next, and generally it falls within the close region around the parameters' values in TABLE 3. It implies that the capacity of the target link is very sensitive to the changes of parameter values.

As to the GA method, a smoother convergence process is observed in terms of average FVs (Avg. Fitness in Figure 5). The ranges between the maximum and minimum fitness values within successive GA generation become smaller during the process, which implies that GA method manages to reach a more stable optimal solution.

(Figure 5)

TABLE 3 shows the optimized local parameters. The fitness value associated with the set of parameters from GA is 1.59, smaller than that from SPSA (4.94). According to some application guidelines of micro simulation [31], both are considered acceptable. But one can notice that these two sets are not close to each other. The fundamental diagrams under the set of best-matching parameters and field observations are plotted in Figure 6, where the field observed link capacity and critical occupancy are estimated to be 5,020 veh/hr and 0.10. The counterparts from SPSA and GA are (4,715 0.083) and (4,972 0.13), respectively. The GA results are better and thus selected in the application; but GA takes much longer time to produce those results than the SPSA method. Compared with the global parameter calibration (Figure 2, 3 and TABLE 2), various sets of local parameters can produce similar simulation performances. As reported elsewhere [17], the calibrated parameters can be verified to provide better simulation results against different data set. The verification implies that it is reliable to use such heuristics methods in micro simulation. It also came to our attention during our sessions' presentations at the TRB annual meeting that Balakrishna et al [32] also apply SPSA to calibrate the entire set of micro simulation model parameters including O-D matrix, driving behavior and route choice behavior model parameters. However, the mutual influence of different parameter groups makes it hard to examine their separate impacts upon the simulation performance as indicated in local parameter calibration.

(TABLE 3)

(Figure 6)

APPLICATION GUIDELINES OF SPSA IN CALIBRATION

SPSA is a method easy to implement, and generally shows good performance, especially when there are observation errors. Similar to other heuristics methods, selection of appropriate algorithmic parameters (including $\boldsymbol{a}$, $\boldsymbol{c}$, $\boldsymbol{A}$, α and γ in SPSA) is of crucial importance to its performance. For example, two different initial values of $\boldsymbol{a}$ generate diverse convergence performances when searching the best MTH-MRT pair (Figure 7), although they both start with the same fair guess (MTH = MRH = 1.60 seconds). Further analysis shows that in the first experimentation it takes only about 15 iterations to reach the band (see Figure 3), while the second experiment takes about 35 iterations to reach the same region. However, more experimentation indicates that choosing too large $\{a_k\}$ and $\{c_k\}$, i.e., larger $\boldsymbol{a}$ and $\boldsymbol{c}$ that aims at an even faster convergence could lead to drastic changes in $\{\hat{\theta}_k\}$ and the calibration process may not even converge. To assist the further application of this method, therefore, some general guidelines [21] and our experiences are summarized here.

(Figure 7)

For the value of c, Spall [21] recommended using the standard deviation of the performance measure (fitness function value) from a few runs under the initial value θ_0, so that the initial perturbation steps do not go excessively large. For $\{a_k\}$, a stabilization parameter A is introduced as $a_k = \frac{a}{(A+k+1)^{\alpha}}$, and 10% (or less) of the predefined maximum number of iterations proves practically effective for the value of *A*. and the value of *a* is chosen such that $\frac{a}{(A+1)^{\alpha}}$ (i.e., the initial step of change) times the magnitude of $\hat{g}_0(\hat{\theta}_0)$ would be equal to the smallest change in the magnitude of θ_0 during the early iterations. We recommend the changes of $\{\hat{\theta}_k\}$ during the first few iterations not exceeding 2-4%.

The other issue concerns the selection of stop criteria. Common criterion determines that a process converges when the percentage change of the objective function value is under a certain threshold, e.g., 2%. It performs well for problems with no observation errors ($\varepsilon_k^{(\pm)}$); but it is not the case in micro simulation calibration, where performance measures such as the FV defined in equation (9) can easily have a variation larger than 2% even under the same set of parameters. For example, in Figure 7, most FVs remain under 2 but some variations exist. Therefore, a predefined maximum number of iterations and an acceptance level of objective function values are better indicators to decide when to stop the calibration process.

CONCLUSIONS

The calibration a micro simulation is a complex problem that often defies mathematical programming based optimization methods and calls for reliable and more efficient heuristic optimization methods. In this study, we introduce another heuristic optimization method, simultaneous perturbation stochastic approximation (SPSA) to calibrate driving behavior model parameters. Compared to other heuristics methods such as the genetic algorithm, this method can generally obtain an acceptable set of parameters in much less time. Nonetheless, one cannot safely say one particular method would outperform all others in all cases; in the numerical experimentation, genetic algorithm can reach better and stable solutions. This finding can help the analyst determine the degree of tradeoff between calibration accuracy and computational time. Even though more powerful software and hardware advances such as parallel computing can nowadays reduce the computational burden, efficient algorithms such as SPSA are still able to significantly save the personnel and computational resources when calibrating micro simulation models.

Micro simulation responds more sensitively to the changes of some parameters but remains relatively invariant to others in certain ranges of parameter values. In the context of global driving parameters, for example, mean reaction time affects the link capacities more significantly than others in the tests. However, the formulated optimization problem could have numerous local optima; various calibration methods, or even different initial conditions for the same method, could produce different set of parameters. Besides using the performance measures such as fitness functions and fundamental diagrams, analysts should exercise their engineering judgment in evaluating their calibration results and choose the best set of parameter values.

ACKNOWLEDGEMENTS

This work is supported by a Caltrans grant TO 5508. We would like to thank Mr. Steve Hague, Caltrans' Division of Traffic Operations, for providing the data support and valuable comments. The revision of the paper benefits greatly from the anonymous reviewers. The views are the authors alone.

REFERENCES

1. Jeannotte, K. et al. 2004. Traffic Analysis Toolbox Volume II: Decision Support Methodology for Selecting Traffic Analysis Tools. Report FHWA-HRT-04-038, Federal Highway Administration.
2. Gardes, Y. and L. Bloomberg. 2003. Devore/I-15 Traffic Simulation Analysis Report-Level 1 Network Investigations. Institute of Transportation Studies, University of California, Berkeley.
3. Cheu, R.L., X. Jin, K.C. Ng, D. Srinivasan, *Calibration of FRESIM for Singapore expressway using genetic algorithm.* ASCE Journal of Transportation Engineering, 1998. 124(6): p. 526-535.
4. Lee, D.-H., Xu Yang, P. Chandrasekar. *Parameter calibration for paramics using genetic algorithm*. in *80th Transporation Research Board Annual Meeting CDROM*. 2001. Washington D.C.: Transportation Research Board.
5. ISAC, L. *Experiences with PARAMICS at ISAC in winter 98-99*. 1999.
6. Gomes, G., Adolf D. May, Roberto Horowitz. *A Microsimulation model of a congested freeway using VISSIM*. in *83rd Transportation Research Board Annual Meeting*. 2004. Washington, D.C.: Transportation Research Board.
7. Gardes, Y., Adolf D. May, *Simulation of IVHS on the Santa Monica freeway corridor using the integration model. Phase 2: Preliminary ATIS and ATMS experiments*. 1993, Institute of Transportation Studies, UC Berkeley: Berkeley, CA.
8. E.B. Lee, C.K. *Validation of a microscopic traffic simulation study on urban freeway reconstruction project with high-traffic volume*. in *84th Transportation Research Board Annual Meeting*. 2004. Washington D.C.: Transportation Research Board.
9. Smith, M.C., Adel W. Sadek. *Challenges in calibrating large-scale microscopic simulation model of diverse urban, suburban, and rural networks: practical guide*. in *86th TRB Annual Meeting*. 2006. Washington D.C.
10. Michael, Balmer, K. Nagel and R. Bryan. 2004. Large-scale multi-agent simulations for transportation applications. Journal of Intelligent Transportation Systems, Vol. 8 pp205-221.
11. Ma, T., B. Abdulhai 2002. Genetic algorithm-based optimization approach and generic tool for calibrating traffic microscopic simulation parameters. *Transportation Research Record* 1806: 6-15.
12. Kim, K. and L. R. Rilett. 2004. A genetic algorithm based approach to traffic micro-simulation calibration using ITS data. *83rd TRB Annual Meeting*, CD-ROM.
13. Ma, J., H. M. Zhang and H. Dong 2006. Calibration of Departure Time and Route Choice Parameters in Micro Simulation with Macro Measurements and Genetic Algorithm. *85th TRB Annual Meeting CD-ROM*, paper No. 2376.
14. Hoyer, R. M. Fellendorf. Parametrization of microscopic traffic flow models through image processing, *8th IFAC Symposium on Transport Proceedings*, Chania, Crete, June 1997.
15. Ben-Akiva, M. et al. 2004. Calibration of microscopic traffic simulation models with aggregate data. *83rd TRB Annual Meeting*, CD-ROM.
16. Gomes, G. et al. 2004. A Microsimulation model of a congested freeway using VISSIM. *83rd TRB Annual Meeting*, CD-ROM.
17. Zhang, Ma and Dong, 2006 *Development of a Calibration Procedure and Assistance Tools for Micro Simulation*. California PATH Research Report, in Review.
18. Chu, L. et al 2004. A calibration procedure for microscopic traffic simulation. *83th TRB Annual Meeting CD-ROM*, paper No. 04-4165.
19. Ben-Akiva, M. et al. 2002. Calibration and evaluation of MITSIMLab in Stockholm. *81st TRB Annual Meeting*, CD-ROM.
20. Spall. J. C. 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transcations on Automatic Control*, 37(3):332:341.

21. Spall, J. C. 1998. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transcations on Aerospace and Electronic Systems*, 34(3):817:823.
22. Sadegh, P. 1997. Constrained optimization via stochastic approximation with a simultaneous perturbation gradient approximation. *Automatica*, 33:889–892.
23. Sadegh, P., J. C. Spall 1998. Optimal random perturbations for multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transcations on Automatic Control*, 43(3):1480:1484.
24. Goldberg, D. E. 1989. *Genetic Algorithms in Search Optimization and Machine learning*. Addison-Wesley.
25. Gardes, Y. et al. 2001. Freeway calibration and application of the PARAMICS model. *80th TRB Annual Meeting*, CD-ROM.
26. Quadstone Ltd. 2004. *Paramics V5.1-Modeller Reference Manual*. Quadstone Ltd, Edinburgh, Scotland.
27. Banks, J. H. 2004. *Introduction to Transportation Engineering,* Second Edition, McGraw-Hill.
28. Highway Agency, Design Manual for Roads and Bridges, Volume 12, Traffic Appriasal of Road Schemes, Section 2, Part I, Traffic Appraisal in Urban Areas. 1996, The Stationery Office: London.
29. Choe, T., A. Skabardonis, P. Varaiya 2002. Freeway performance measurement system (PeMS): an operational analysis tool. *81st TRB Annual Meeting* CDROM.
30. Quadstone Ltd. 2004. Paramics V5.1 Estimator Reference Manual. Edinburgh, Scotland.
31. Dowling, R., J. Holland, and A. Huang 2002. *Guidelines for applying traffic simulation modeling software*. Dowling Associates Inc. Oakland, CA
32. Balakrishnan, R. et al. 2007. Calibration of microscopic traffic simulation models: methods and application. 86th TRB Annual Meeting, CD-ROM. Paper No. 07-2191.

LIST OF TABLES AND FIGURES

TABLE 1 Global and Local Parameters

Parameter Category	Parameter Name	Feasible Range	Unit	Description
Global Parameters	Mean Target Headway	0.6~2.4	Second	Mean headway between a vehicle and its following vehicle
	Mean Reaction Time	0.4~1.6	Second	Mean time lag between a change in speed in a leading vehicle and the following vehicle's reaction to this change
	Driver Aggressiveness	0.2~0.8 (mean)	N/A	A distribution that determines how long a headway is accepted by a DVU
	Driver Awareness	0.2~0.8 (mean)	N/A	A distribution that affects the use of a longer headway when a vehicle approaches a lane drop or a ramp
Local Parameters	Link Headway Factor	0.5~2.5	N/A	Adjustment factor for the mean headway on a link
	Link Reaction Factor	0.5~2.5	N/A	Adjustment factor for the mean reaction time on a link
	Ramp Headway Factor	0.5~2.5	N/A	Adjustment factor for the mean headway on a ramp
	Minimum Ramp Time	1~3	Second	Minimum time that a DVU remains on a ramp before considering merging into the freeway
	Ramp Awareness Distance	1~300	Meter	A distance at which a freeway DVU is aware of an approaching ramp
	Sign-posting	1~300	Meter	A distance from the hazard that the most aware vehicles become aware of the hazard ahead.

TABLE 2 Numbers of Performance Evaluations under Various Algorithms at Convergence

Algorithm	Total Number of FV Evaluations	Optimal Solution			CPU Time Taken (hour)
		Fitness Value	Mean Target Headway (second)	Mean Reaction Time (second)	
IA	2,400	0.15	0.96	1.24	12.30
GA	600	0.61	0.96	1.25	3.1
SPSA	150	0.70	0.87	1.27	0.9

TABLE 3 Best Optimized Local Parameters

	Bottleneck (mainline north of the interchange)		On-Ramp from Fruitridge Road WB		On-Ramp from Fruitridge Road EB	
	GA	SPSA	GA	SPSA	GA	SPSA
Link Headway Factor	0.63	0.93	1.96	0.91	0.53	1.18
Link Reaction Factor	1.58	0.99	0.93	1.13	0.80	1.06
Sign-posting	3043	801	924	802	759	805
Ramp Headway Factor	---	---	1.13	1.13	1.08	0.99
Minimum Ramp Time	---	---	2.94	1.44	1.28	1.64
Ramp Awareness Distance	---	---	212	484	8.72	245

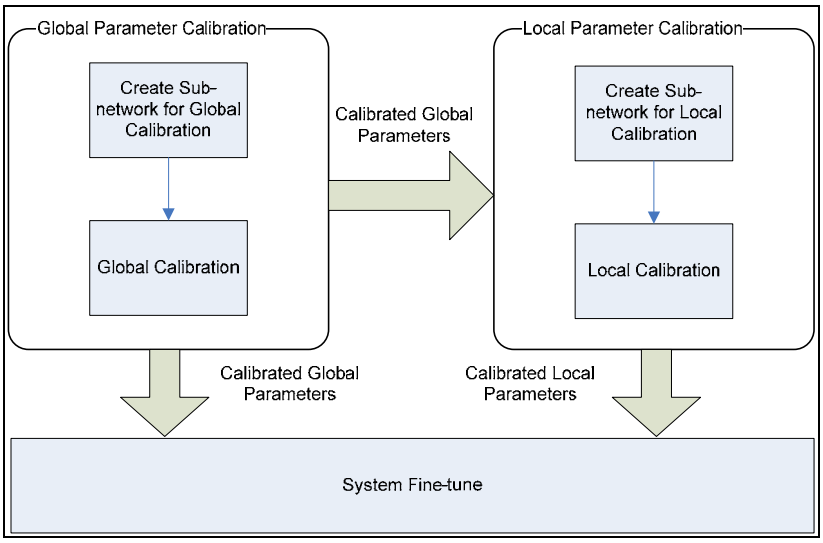

Figure 1 Detailed Global/Local Parameter Calibration Process

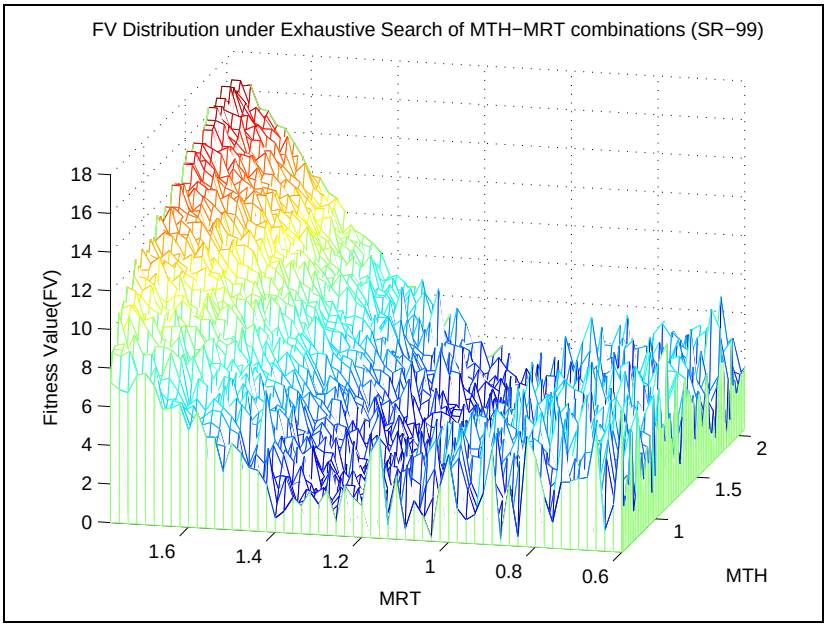

Figure 2 Mesh Figure from IA Exhaustive Search (Network SR-99) (with Fixed AGGR and AWAR)

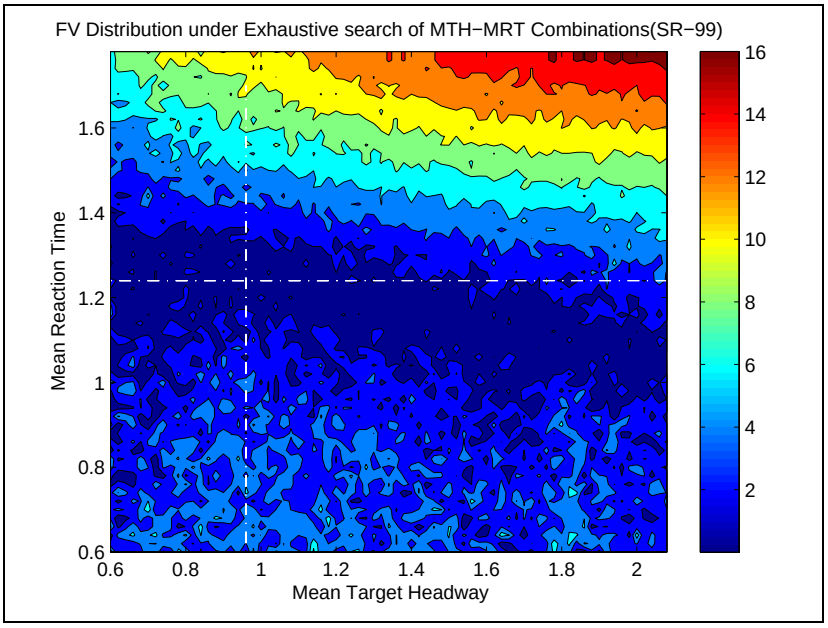

Figure 3 Contour Plot (MTH-MRT) from IA Exhaustive Search (SR-99) (with Fixed AGGR and AWAR)

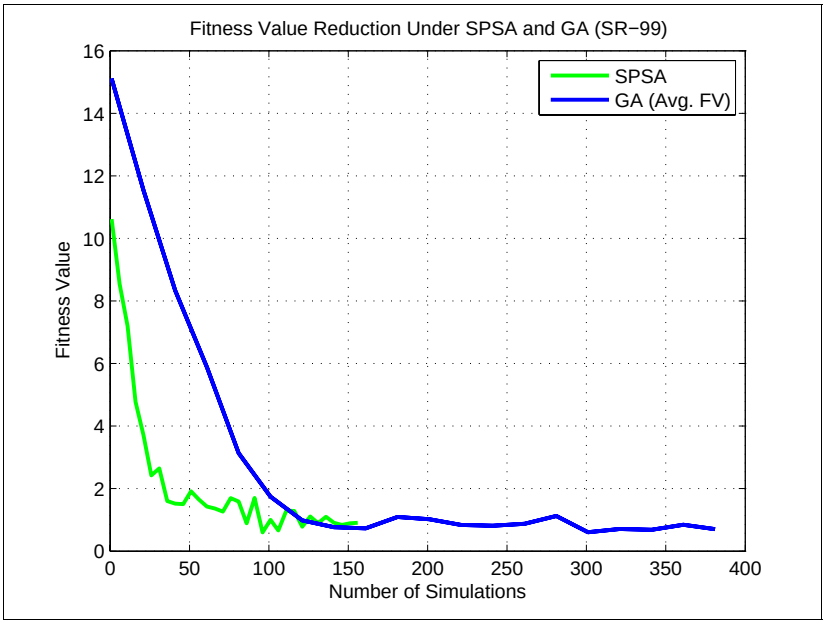

Figure 4 GA and SPSA based global parameter calibration

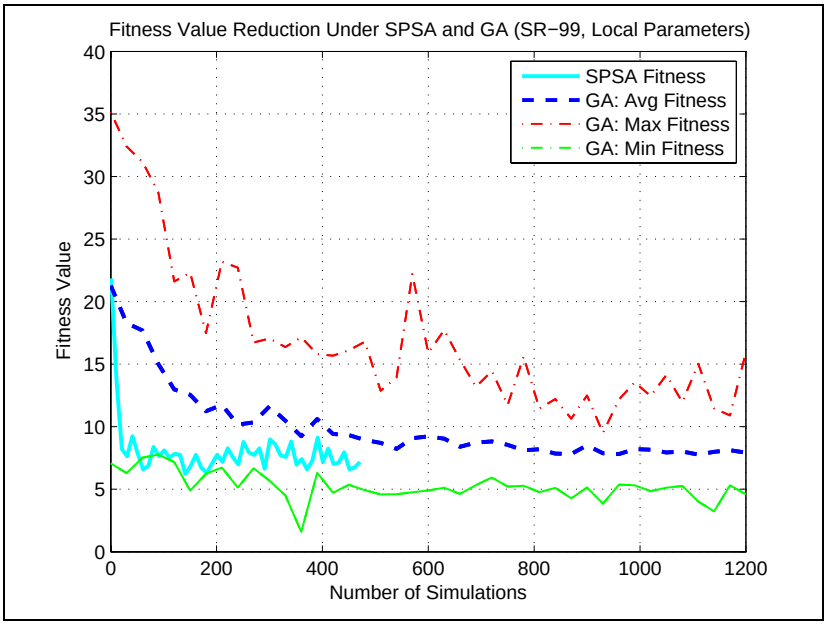

Figure 5 Local Parameter Calibration Convergence Diagram

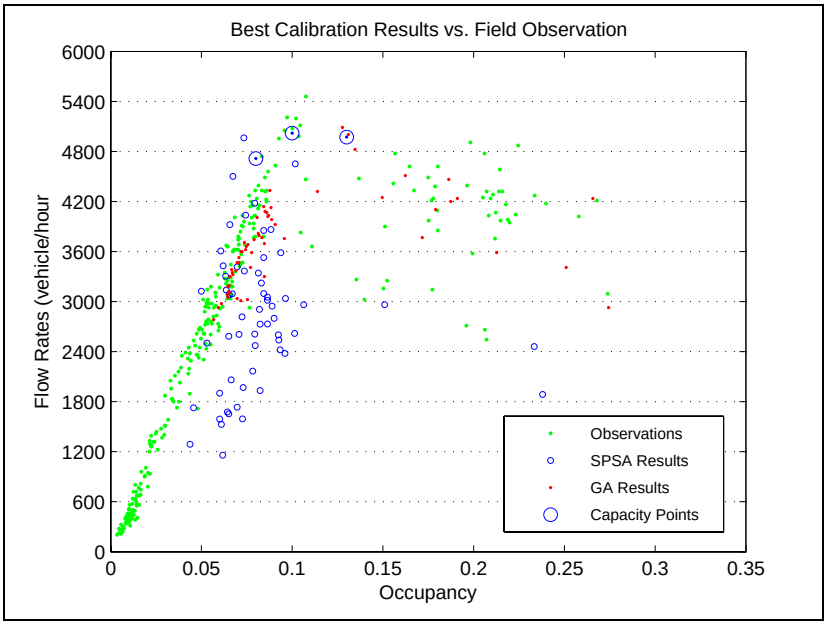

Figure 6 Best Simulation Results vs. Field Observation

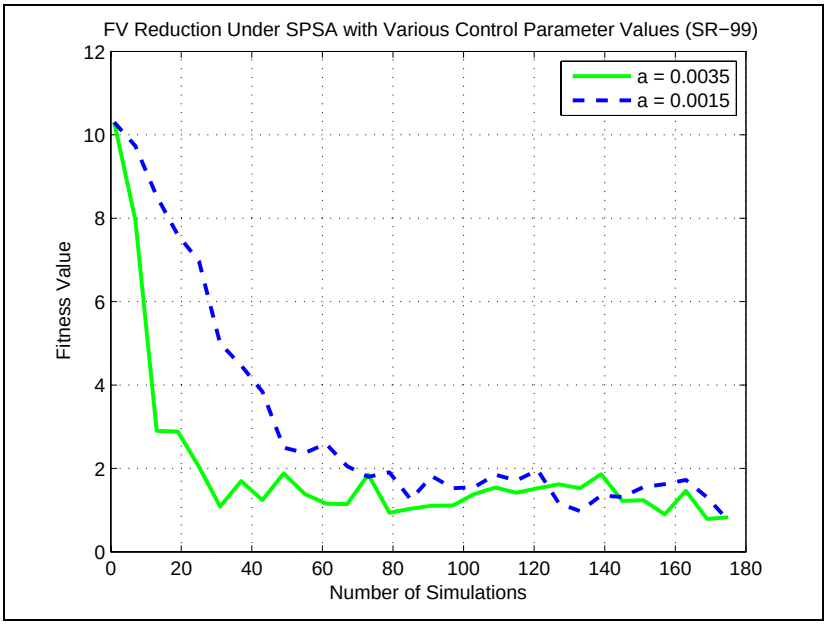

Figure 7 Experimentation with SPSA Algorithmic Parameters (SR-99)