Statistical Calibration for Data-Driven Microscopic Simulation Model

Call Title: Analytical Methods for Traffic Models for Calibration and Validation

Sponsoring Committee: AHB45: Traffic Flow Theory and Characteristics
Co-Sponsoring Committee: ABJ80: Statistical Methods

Authors:

Dwayne Henclewood, Ph.D.
Senior Consultant, Booz Allen Hamilton
henclewood_dwayne@bah.com

Wonho Suh, Ph.D.*
Research Engineer, Georgia Institute of Technology
wonho.suh@ce.gatech.edu

Michael O. Rodgers, Ph.D.
Adjunct Professor and Principal Research Scientist, Georgia Institute of Technology
michael.rodgers@ce.gatech.edu

Michael Hunter, Ph.D.
Associate Professor, Georgia Institute of Technology
michael.hunter@ce.gatech.edu

* Corresponding author
Phone: 404-894-2201 / Fax: 404-894-2278
School of Civil and Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive, Atlanta, GA 30332

Word Count: 7425 (text: 5175; tables: 1 x 250: 250; figures: 8 x 250: 2000)
Submission Date: November 15, 2012
ABSTRACT
For many decades efforts have been made to solve transportation problems. A number of research efforts have been geared toward developing accurate traffic simulation models. One of the challenges is that the model does not always adequately reflect field conditions without proper calibrations. This paper aims to highlight the importance of proper calibrations by providing a statistical calibration procedure based on the NGSIM Next Generation Simulation (NGSIM) dataset. First, a Monte Carlo approach is employed to generate candidate parameter sets for calibration. Simulations with these parameter sets are evaluated against a robust set of calibration criteria including startup and saturation flow characteristics and travel time distributions. The parameter sets that satisfy these criteria are considered as adequately calibrated. The results suggest that parameters determining distance between cars under various conditions are dominant meeting the evaluation criteria. The results suggest that this approach offers a robust and effective method of calibrating simulation models where disaggregate level vehicle data are available.
INTRODUCTION

Traffic congestion incurs a cost of an estimated one hundred billion dollars each year in the United States [1]. Addressing this high cost of congestion is a primary issue facing many of today’s urban and suburban areas. However, as resources have become increasingly constrained, and right-of-way and construction costs have increased, there has been a fundamental shift in the manner in which congestion issues are addressed. This shift is reflected in the substantial efforts to develop and implement alternate means for alleviating congestion. These solutions utilize advanced technologies to increase the efficiency of today’s transportation network.

Microscopic simulation is being viewed as a tool that is able to increase the capabilities of these advanced transportation solutions. However, microscopic models must be appropriately calibrated to provide results that reflect field performance measures [2]. To date, the predominant means of calibrating a microscopic simulation model is based on selecting a set of calibration parameters that allow the model to reflect the average field performance measures, not the field performance measure distribution [2]. To highlight the importance of proper calibrations, this paper presents an argument for a statistical calibration method that may offer a more effective method of calibrating these models where significant quantities of vehicle trajectory or individual vehicle travel time data (e.g. from GPS probe vehicles or video-based techniques) are available. Statistical calibration methods coupled with individual vehicle data have the potential to improve the assignment of model parameter values and increase the likelihood of producing robust and precise traffic simulation models.

The following section presents a brief literature review of previous calibration efforts related to microscopic modeling of surface transportation. Next, the proposed statistical calibration method and the results of its application to the Next Generation Simulation (NGSIM) dataset are presented.

LITERATURE REVIEW

A variety of methods have been developed to calibrate traffic simulation models. Hollander and Liu presented a comprehensive review of the current calibration methods and highlighted the fundamental requirements for calibrating simulation models [3]. Zhang and Ma also reviewed the current calibration methods and grouped them in three categories, 1) trial-and-error heuristics, 2) genetic algorithms, and 3) simulated annealing [4]. They found that a majority of calibration methods were either trial-and-error heuristics or genetic algorithms methods, with the majority being trial-and-error heuristic methods.

A few notable trial-and-error heuristic methods are highlight below. These include Chu et al.’s four-step calibration process [5]. These four steps include determine a suitable driver behavior model(s), select an appropriate route choice mechanism, estimate origin-destination pairs accurately and then fine tune the model to reflect field performance measures [5]. Oketch and Carrick also developed trial-and-error heuristic method that is based on 1) determining proper model parameter values such as, aggressiveness, awareness, target headways and reaction times and 2) estimating representative origin-destination matrices [6]. The trial-and-error method that Toledo et al. proposed is an iterative calibration method to jointly estimate origin-destination flows and values for behavioral parameters [7]. Dowling et al.’s method included a series of a four step processes to calibrate a model, which included error checking, calibration for capacity, calibration for demand, and overall review [8, 9, 10].

As for calibration methods based on the genetic algorithms, two of the more noteworthy applications were presented by Park and Won.[11], and Zhang et al. [12].

TRB 2013 Annual Meeting  
Paper revised from original submittal.
In addition to the numerous calibration methods, various criteria have been developed to determine when a model is properly calibrated. Hellinga found that the criteria tend to be subjective due to their dependence on what is being modeled and the goals of the study [13]. Hollander and Liu presented a summary of a number of these calibration criteria in [3]. To date, one of the main criteria to determine whether or not a model is calibrated involves a parametric, first moment statistical comparisons of field and simulated performance measures [2]. For instance, Park and Schneeberger used the results from the t-test to compare simulated and field travel time averages as the criterion to determine when a model is calibrated [14]. Park and Won developed a criterion which determined a model as calibrated when the model’s travel time distribution “includes the entire field-measured values” [15]. Although these criteria may be sufficient to evaluate general traffic performance at an aggregated level, it is questionable if such means can represent traffic performance at an individual vehicle level. This paper highlights the importance of proper calibrations by providing a statistical calibration procedure which allows the corresponding simulation to accurately represent underlying distributions that make up traffic related performance measures at both aggregated level and individual vehicle level.

**STATISTICAL CALIBRATION METHOD**

The proposed calibration method includes a comprehensive evaluation of the selection of calibration parameter values and a two-part criteria process (Figure 1). It is assumed at the start of this effort that the model is well constructed. This calibration is not intended to address issues related to model construction but instead underlying model parameter value selection.

![Figure 1 Statistical calibration procedure.](image)

First, a Monte Carlo approach is employed to generate potential parameter sets for calibrating a traffic simulation model. After identifying effective parameter sets for calibration, simulations with these parameter sets are evaluated against a robust set of calibration criteria to determine which are calibrated. There are two steps to the calibration parameter value selection process: 1) evaluation of startup and saturation flow and 2) statistical evaluation of the
performance measure distributions (travel time in this example) including Wilcoxon-Mann-Whitney test, Kolmogorov-Smirnov test, and a heuristic form test. The parameter sets that satisfy both these steps are considered as adequately calibrated.

**Monte Carlo Process – Identifying Effective Calibration Parameters**

This Monte Carlo process is expanded from [16, 17]. This process consists of three steps: 1) performance measure and initial parameter selection, 2) Monte Carlo simulations, and 3) parameter elimination. Steps 2 and 3 are repeated until a pre-defined stopping condition is satisfied.

In the first step, the performance measure for the model is selected. The Performance measure can be any number of, and any combination of, travel times, flows, delay, and queue lengths. For the initial parameter selection, certain parameters can be eliminated immediately, for example, parameters which are not used in the model due to the facility type being modeled (i.e. arterial versus freeway). Miller [16] demonstrated that approximately half of VISSIM’s 50 calibration parameters could be eliminated in this step. For each of the selected parameters, a reasonableness range is determined.

Step two involves a Monte Carlo experiment that generates a series of simulation runs used to isolate the sensitivity of the performance measure to the parameter values. One thousand parameters sets are created in a Monte Carlo fashion. The value for each of these parameters is limited by its reasonableness range, determined as part of step 1. These parameter sets are then used as inputs to generate 1000 unique simulation runs where only these parameter values vary between runs. From each simulation run, the selected performance measures are extracted and analyzed. Note, replicate runs were not conducted for each model as employing the Monte Carlo method to create these models appropriately approximates the effect of executing replicate runs of each model – on an aggregate level. Step three compares the performance measures with the values for each parameter and eliminates parameters whose values have little or no effect on the measures. Default values are assigned to the eliminated parameters and step two is repeated until no parameters can be eliminated due to a lack of influence on the performance measures. Figure 2 illustrates the execution of the process.

**Calibration Criteria (Flow & Statistical Evaluation Criteria)**

The proposed method includes two steps to select calibrated parameter sets. The first step compares field and simulated saturation flow rates to ensure that the models produce reasonable estimates. This step is necessary as it is possible for models to produce accurate estimates of performance measures, such as travel time, while producing over or under estimated saturation flow rates. Failure to adequately model saturation flow could prove significant. For example, by overestimating saturation flow a simulation will overestimate capacity. In a scenario where base volumes are increased to a higher level to represent future conditions it is possible they could exceed field capacity (implying significant congestion) but the simulation would continue to show uncongested operations. As a result, the simulation would no longer reflect field conditions. The second step involves the statistical evaluation of the mean and the distribution of the performance measures being studied, travel time in this study. This step includes Wilcoxon-Mann-Whitney (WMW), Kolmogorov-Smirnov (KS), and heuristic form fit tests (HFF). By evaluating travel time means and distributions, the calibrated model will consider reflecting both aggregated level and individual vehicle level traffic information, which is necessary for a properly calibrated model.
Startup and Saturation Flow Criteria

Startup and saturation flow criteria not only facilitate greater confidence in the results from a calibrated model but also provide some level of protection from the potential dangers in the application and implementation of the calibrated model. Startup and saturation flow are measured from headway measurements at an intersection. Startup flow is measured from the second vehicle in the queue to fifth vehicle, while saturation flow is estimated based on measurements from the sixth to the ninth vehicle. Headway measurements are then averaged for each cycle and converted to startup ($s_{su}$) and saturation ($s$) flows via Equation 1 [18]. The startup and saturation flow rates (per cycle) are then used to create frequency density plots of the observed flow rates.

$$s_{su} = \frac{3600}{h_{su}} \quad s = \frac{3600}{h}$$

where: $s_{su}$ = startup flow (vehicles/hour)
$h_{su}$ = startup headway (second)
$s$ = saturation flow (vehicles/hour)
$h$ = saturation headway (second)

To create the startup and saturation flow criterion, a reasonable range is chosen to aid in the evaluation of whether or not a parameter set may be retained for further consideration as a calibrated model. The reasonable range is constructed by forming a 95% confidence interval around the mean flow values from the field. If field data are limited to construct an appropriate confidence interval, it is recommended to use a bootstrap approach to bolster the field data that will be used to make inferences about field startup and saturation flow estimates [2].

Figure 2 Monte Carlo calibration parameter selections.
Bootstrapping is a means of making statistical inferences in the presence of limited data. The bootstrap method involves the re-sampling of data, with replacement, in order to generate an empirical estimate of the entire sampling distribution of a statistic [19].

Statistical Evaluation Criterion

For the models producing startup and saturation flow measurements within 95% confidence interval, a statistical evaluation criterion is examined. Since field data often do not fit known distributions, non-parametric tools are used to ensure proper comparisons of field and simulated data. Two sets of non-parametric tools were used to establish the statistical calibration criterion. The first set of tools employed the use of WMW test to conduct a general distribution comparison to primarily determine the homogeneity between field and simulated performance measures (travel time). The second set of tools involves the use of a more stringent comparison of the distribution of performance measures. This set of tools employ the use of the KS and HFF tests. Both sets of tests that were selected here complement each other. Each test attempts to satisfy the shortcomings of another which were in part due to some the fundamental principles in their original development. For a full discussion regarding the selection of each of tests and how they are able to complement each other, readers are encouraged to review the works presented in [2] and [20].

The WMW test was chosen to evaluate the general differences between field and simulated travel time distributions. One of the outputs is a p-value that is used in the decision to accept or reject the null hypothesis, $H_0$, (for this case that the field and the simulated travel time distributions are equivalent). A calibration criterion based on the WMW p-value is the rejection of $H_0$, and subsequently a model parameter set, when the p-value is ≤0.01.

This criterion is then paired with another that is based on a more stringent comparison of the travel time distributions. The need for a more stringent comparison of travel time distributions is due to the nature of the WMW test that does not explicitly consider the absolute magnitude of the differences in travel time. Given the multimodal nature of the actual travel time distributions, the KS non-parametric test was selected to further compare the distribution of travel time datasets. Unlike the WMW test, the KS test does take into account the magnitude of the differences between the data points of the samples being compared. To formalize the calibration criterion based on the KS test, a model comparison whose test statistic corresponds to a p-value of ≤0.01 will result in a rejection of $H_0$. The null hypothesis in this case states that there is insufficient evidence to suggest that a parameter set’s simulated distribution of travel time estimates is different from same distribution obtained from the field. The rejection of $H_0$ for a particular travel time segment removes that model from being considered as a possible calibrated model for a particular period and travel direction.

The HFF test (developed for this effort) was also included to provide an alternative distribution comparison method. The HFF test is devised to compare the rate of change of the CDFs (instantaneous density) of the two distributions. A test statistic, $H$, is created and is defined as the sum of squares of the difference between the rate of change between the CDF of the field and simulated data, $F_n(x)$ and $\hat{F}_n(x)$, respectively. Mathematically;

$$H = \sum \left( \frac{dF_n(x)}{dx} - \frac{d\hat{F}_n(x)}{dx} \right)^2$$

While there are no special assumptions associated with the HFF test, the method does not
take into account shifts along the x-axis, i.e. differences in central tendencies (mean/median). This is concerning as for a given comparison, \( H \) having a value equal to or close to zero does not necessarily mean that the results from a parameter set fits field data. The only definitive statement that may be made is that the shapes, or forms, of the two distributions, including its modal characteristics, are similar. Since there is currently no \( p \)-value for this test statistic, parameter set simulations that produce \( H \) values in the bottom half of the range of \( H \)-values are considered as potential calibration model candidates.

The statistical evaluation criteria use the WMW, KS and HFF tests to examine the field and simulated travel time distributions. All three statistical tests are used to determine which parameter sets satisfy the statistical evaluation criteria. The following steps outline the application of these tests and how parameter sets that satisfy the statistical evaluation criteria are selected:

1. Conduct the WMW test and retain parameter set simulations whose WMW test yielded \( p \)-values \( \geq 0.01 \). This will be a set of parameter sets denoted by \( M_n|U \).
2. Conduct the KS test and retain parameter set simulations whose KS test yielded \( p \)-values \( \geq 0.01 \). This will be a set of parameter sets denoted by \( M_n|D \).
3. Conduct the HFF test and retain parameter set simulations whose \( H \) value is in the bottom half of the range of \( H \)-value. This will be a set of parameter sets denoted by \( M_n|H \).
4. To obtain the set of parameter set simulations that satisfy the statistical evaluation criteria, carry out the following set operation

\[
M_f = M_n|U \cap M_n|D \cup M_n|H
\]

The parameter set simulations that satisfy both the KS and HFF tests are combined as they evaluate the same characteristic of the data – the shape of its distribution. The union of these two sets facilitated the inclusion of simulated data that may have been otherwise excluded. The exclusion of such datasets may have been the result of a violation of a KS test assumption and/or the inability to provide a large enough \( p \)-value to be included in \( M_n|D \), despite having similarly shaped distributions. In other words, \( M_n|H \) and \( M_n|D \) are combined to further minimize the probability of committing a Type I error, regarding distribution shapes. With these tools working in tandem, parameter set simulations that produce accurate estimates of field performance measure distributions will be considered as calibrated replicates.

**MODEL APPLICATION**

The proposed statistical calibration method is applied to the data collected as part of the NGSIM program [21]. This application is intended to investigate the feasibility of the proposed statistical calibration method when individual vehicle data are available. The NGSIM data were collected on November 8, 2006, between 12:45PM and 1:00PM (referred to as Noon period in this paper) and 4:00PM and 4:15PM (referred to as Evening period in this paper), along Peachtree Street in Atlanta, Georgia. This data set consists of trajectory information (with a resolution of a tenth of a second) for all vehicles traversing the corridor during the study period. In addition, signal phase information at each intersection, origin/destination (OD) data for each vehicle, turning movement distribution data at each intersection, and a series of other traffic related information were also collected.

For the model application, a detailed VISSIM model of the study area was created.
VISSIM is a discrete, stochastic, time step based microscopic simulation model. In this model all vehicles are modeled individually, based on a psycho-physical driver behavior model developed by Wiedemann [22]. Several verification iterations were completed to ensure that the model correctly represented the study area, as well as the traffic operations during the study period. A VISSIM trip-chain file was created to provide a time-stamp, indicating when a vehicle entered the network and OD information.

**Uncalibrated Model**

After inputting the necessary field data from the NGSIM study area into the VISSIM model, ten replicate runs were conducted. The results from these runs were then used for the comparison between field and simulated performance measures. In examining the average of the 10 replicate runs there were some discrepancies between the simulated and field travel time estimates. A common discrepancy is that VISSIM tends to under-estimate field travel times. The smallest and largest differences between VISSIM and field travel time are approximately 8 seconds (Noon-Southbound) and 33 seconds (Evening-Southbound). When comparing standard deviations, it is seen that the values produced by VISSIM are similar to those from the field, which indicates that VISSIM’s approximation of the travel time variation estimates is rather similar to that of the field. Travel time frequency density plots are provided in Figure 3. The plots of the simulated travel times generally capture the bi-modal or tri-modal form of the field travel times. The differences between the plots tend to be a shifting of the centroid of the modes or proportionality between the different modes. However, in all cases the general form of the distribution is reflected, likely indicating many of the differences may be addressed through a calibration procedure.

![Travel time frequency density plots](image)

**FIGURE 3 Travel time frequency density plots of NGSIM Field vs. VISSIM (single run).**
Model Application - Monte Carlo Process

Based on the procedure described earlier, a Monte Carlo approach was applied to create candidate parameter sets that sufficiently represented the sample space for each parameter. Ten final parameters were selected to calibrate the NGSIM model. The selected parameters are average standstill distance, additive part of safety distance, multiplicative part of safety distance, maximum deceleration (own), maximum deceleration (trailing), minimum headway (front/rear), safety distance reduction factor, maximum deceleration for cooperative braking, lane change distance and desired speed distribution range [16]. The Monte Carlo process produced 1000 unique parameter sets and each set was simulated in the VISSIM model. Travel time (Figure 4) and startup and saturation flow measures were extracted from the simulation runs.

![Travel time frequency density plots of NGSIM field vs. 1000 VISSIM runs.](image)

Model Application - Startup and Saturation Flow Criteria

Travel time and saturation flow measures were extracted from the simulation runs with the 1000 unique parameter sets in the Monte Carlo process. These measures were then analyzed to determine which combination of parameter values most closely reproduced the NGSIM results. From the NGSIM video, a 95% confidence interval was constructed using the bootstrapped flow measurements. This confidence interval was produced by the percentile-t method and was selected as the reasonable range for simulated flow rates. The green horizontal lines in Figure 5 represent the percentile-t 95% confidence interval for the startup flow (left) and saturation flow.
(right), while the red line represents the respective mean value. The upper and lower bounds for startup flow is 1342 and 1488 veh/hr/ln respectively; while for saturation flow the bounds are 1499 and 1796 veh/hr/ln. In applying the flow criterion, simulated models that produced average startup and saturation flow measurements that are within the respective confidence intervals are retained for further consideration as calibrate models. After applying this criterion, of the 1000 VISSIM models only 159 produce flow measurements that are within the 95% confidence intervals. These 159 models are examined against the statistical evaluation criteria to determine which models most closely simulate the field results.

![FIGURE 5 Field (red) and simulated startup and saturation flows.](image-url)

**Model Application – Statistical Evaluation Criteria**

For statistical evaluation criteria the WMW, the KS and a HFF tests were applied on the 159 parameter set simulations that satisfied the above saturation flow criteria. These tests were performed for each time period and direction of travel in the NGSIM data set (Noon and Evening, Northbound and Southbound). Ideally, the parameter sets for each of these time periods and directions should be the same. However, the above analysis yielded a number of different calibrated parameter sets across periods and directions. This means, for each time period and direction, different parameter sets were able to produce a calibrated model. Table 1 presents the number of parameter sets that are the same for different time periods and travel direction. Forty three parameter sets were considered candidate calibration parameter sets for Noon period, and two for the Evening period. And there was only one model that was calibrated for both periods.
and travel directions.

Figure 6a presents travel time frequency density plots from this single calibrated parameter set (red line), as well as plots from the field data (black line) and the original VISSIM model with default parameter values (yellow line). For comparison, Figure 6b shows the startup and saturation flow plots for the same parameter set. The selected parameter set is deemed properly calibrated.

FIGURE 6 NGSIM field, VISSIM default parameter, and calibrated model comparison: (a) travel times, (b) startup flow rate, (c) saturation flow rate.
Effective calibration parameter analysis

One of the opportunities that the calibration process affords is the exploration of the values for each effective calibration parameter and its ability to produce a calibrated simulation. The objective of this analysis is to provide some guidance to the modelers regarding how to choose and combine values for effective calibration parameters in order to increase the likelihood of producing an adequately calibrated model. To do this, the parameter values for all replicates are compared to those that produced an adequately calibrated simulation.

Figure 7 presents a parallel coordinate, similar to a cobweb plot shown in [23]. This plot is intended to identify the effective parameters among the Monte Carlo simulations, while Punzo et al. [23] used this plot to explore the existence and the nature of the local minima in the calibration experiments. The y-axis represents the normalized calibration parameter values ranging from zero to one. The x-axis represents each of the effective calibration parameter in addition to the “distance between cars (d)” parameter. This “d” parameter is a function of “average standstill distance (ax)”, “additive part of safety distance (bx_add)”, and “multiplicative part of safety distance (bx_mult)”, each of which are included in the set of effective calibration parameters. Each colored line in Figure 7 represents a single parameter set and its intersection with a vertical axis reflects the value of that particular parameter for that set parameters.

Figure 7a demonstrates that all the values, within the reasonable range, for each calibration parameter were sufficiently covered by all the candidate parameter sets that were produced by the initial Monte Carlo process. To help inform the selection of parameter values that increase the likelihood of producing a calibrated model, the difference between Figure 7a and 7b were explored. The difference between these two figures is that Figure 7b only presents parameter sets which satisfy both aforementioned saturation and statistical criteria to calibrate a model, for at least one of the four time periods being analyzed. It was anticipated that Figure 7b would present distinct patterns in the values for each calibration parameter, and that these patterns would inform modelers as to how to select values for these parameters to increase the likelihood of producing a calibrated model. However, distinct patterns were only observed for a few parameters. To further explore these patterns and derive possible insights, Figure 8, which presents the density plots of the values for each parameter, will be analyzed in parallel with Figure 7b.

In examining Figure 7b and Figure 8 strong patterns for parameters “ax”, “bx_add”, “bx_mult”, and “d” are observed. Typically, for a given calibrated parameter set, as the value of “ax” approaches its maximum value, the value of “bx_add” approaches its minimum and “bx_mult” approaches its maximum value; and vice versa. The strongest pattern was observed in the values for “d”. All of parameter sets that satisfied the aforementioned calibration criteria produced a “d” value between 26% and 53% of its maximum value. What this observation suggests is that the values for parameters, “ax”, “bx_add”, and “bx_mult” are acting in a combined manner to produce a “d” value between 26% and 53% of its maximum value.
FIGURE 7 Parameter values: (a) 1000 uncalibrated sets, (b) calibrated sets.
FIGURE 8 Frequency density plots of uncalibrated and calibrated parameter values.
The VISSIM manual states that “ax”, “bx_add”, “bx_mult”, and “d” are the main parameters that dictate flow [22]. Therefore, an accurate estimate of field saturation flow, paired with a reasonable confidence interval, is imperative to increase the likelihood obtaining a parameter set that is able to calibrate a model. In the larger context of model calibration, the dominant patterns in the above parameters indicate that accurately estimating saturation flow is critical when developing a calibrated model.

As for the other effective parameters, patterns in their assigned values for calibrated replicates are less distinguishable. Therefore, it is seemingly just as likely for the default values of these parameters to produce calibrated replicates, as it would be if any other values were assigned. Thus, a future study could explore a reduced calibration effort in which only the four discussed parameters are considered and the other parameters are set to their default value. Based on the given results it is highly likely that this approach would result in calibrated models of nearly the same quality with a significantly reduced computational effort.

CONCLUSION AND DISCUSSIONS
Transportation impacts every aspect of daily life. For many decades efforts to improve transportation and alleviate congestion have been made and traffic simulation has played a significant role in these efforts. In light of the growing prominence of traffic simulation, additional emphasis is needed on the calibration of the associated models. Toward this end, a statistical calibration procedure for data-driven traffic simulation model was proposed. This procedure was applied to calibrate the VISSIM model of the NGSIM study area. One thousand potential parameter sets were generated and these parameter set simulations were evaluated against a robust set of calibration criteria to determine which were calibrated. Two calibration criteria were applied: 1) evaluation of startup and saturation flow and 2) statistical evaluation of travel time distributions. The parameter sets that satisfy both these criteria are considered as adequately calibrated. As this procedure requires disaggregate individual data for calibration its large-scale implementation remains challenging. However, with the continuing expansion of individual vehicle based data collection technologies (e.g. GPS probe vehicles, video-based techniques, Bluetooth, etc), it is anticipated that disaggregate level vehicle data (similar to the NGSIM data) will be readily obtainable in near future.

Also, the structure of the proposed calibration procedure provided insight into the assignment of parameter values in order to increase the likelihood of producing a calibrated parameter set. Inferences regarding parameter value assignment were made about four of the ten effective calibration parameters. As for the values of other parameters, initial evidence seems to indicate that default values are as likely to produce calibrated replicates as any other reasonable values. Therefore, for future implementation of a more efficient calibration process it may be possible to further limit the number of parameters requiring calibration, potentially significantly reducing the computational time required to develop a reasonably calibrated model.

There is, however, a need for additional future research. First, the calibrated model based on the NGSIM data was not validated in this paper, since the validation of the calibrated parameters requires additional independent data. Validating the calibrated model is an important step to determine if the proposed calibration method is effective. If travel time distribution is the validation criterion, additional individual vehicle travel time data are necessary for the validation. Also, this effort is particular to a specific corridor at set times. Although it is expected that different driving behaviors (for example, urban versus rural) may lead to different startup and saturation flow parameter selections, it is not known if the developed parameters are
transferable to other time periods or locations in neighboring areas. Identifying the calibration parameters in other time periods or nearby locations are topics that warrant further research. Understanding the sensitivity of the parameter selections under different traffic environments is important to applying the calibrated model.

It should be noted that this Monte Carlo method requires a significant computational cost which can make the approach less effective. Each VISSIM run used in this study required approximately 3 minutes of computer runtime. In total, it took approximately 48 hours to run 1000 VISSIM runs on a computer with 3 GHz Intel Pentium 4 PC with 2 GB RAM. However, it is expected in the automated approach that multiple techniques would be used to shorten the length of analysis time required. For instance, there is no requirement that the VISSIM simulations be executed sequentially. An efficient automated approach should take advantage of the increasing presence of multi-core computers and VM’s allowing for parallel execution of multiple VISSIM instances. (It is worth noting that this is not a significant leap in technology or programming and that as a separate distributed simulation research effort the authors are currently running 10 VISSIM model instances in parallel.) Also, the authors believe that it may be possible to gain efficiency by eliminating non-important parameters at the beginning stage of the Monte Carlo process to reduce computational loads. It is important to note the current paper is intended to explore the potential for utilizing a Monte Carlo approach for a statistical calibration method in data-driven microscopic simulation models. The authors’ main attention was not on optimizing the efficiency of the approach. Future effort will consider the approaches to increase efficiencies and look into trade-offs in those approaches. Beyond these limitations, research into the vehicular conflicts with other transportation modes is needed. For example, pedestrian and bicycle activity was minimal in the NGSIM data. However, these could create inaccurate parameter selections even though travel time distribution and flow parameter are well calibrated, particularly in locations with high pedestrian and bicycle activities.

ACKNOWLEDGEMENTS

This work was sponsored in part by the National Science Foundation “Embedded Distributed Simulation for Transportation System Management” and the Georgia Department of Transportation research project “Real Time Estimation of Arterial Travel Time and Operational Measures through Integration of Real Time Fixed Sensor Data and Simulation”. The findings and conclusions presented herein represent the opinion of the authors and not necessarily that of the National Science Foundation and the Georgia Department of Transportation.

REFERENCES


