

A CASE FOR REAL-TIME CALIBRATION OF DATA-DRIVEN MICROSCOPIC TRAFFIC SIMULATION TOOLS

Dwayne Henclewood
Wonho Suh
Michael Rodgers
Michael Hunter

Richard Fujimoto

Civil and Environmental Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Computational Science and Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

Despite recent technological advancements in alleviating roadway congestion, there is still a considerable amount of time and fuel wasted by travelers. In searching for solutions to mitigate congestion, a number of research efforts have been geared toward developing simulation tools to provide real-time performance measures. One of the challenges of such tools is that the underlying simulation model does not always adequately reflect field conditions outside of the time period for which it was calibrated. In this paper, this is highlighted when calibrating a model for two different periods. During this exercise, 1000 model replicates were generated to explore the sensitivity of potential calibration parameter values. From this analysis only one replicate was found to be adequately calibrated for both periods. This paper suggests that a real-time calibration algorithm should be included in online, data-driven microscopic traffic simulation tools.

1 INTRODUCTION

Traffic congestion made drivers pay more than one hundred billion dollars in the United States in 2010 (Schrank, Lomax, and Eisele 2011). There has been a fundamental shift in the manner in which congestion issues are addressed. Previously, congested roadways were remedied by increasing roadway capacity. However, acquiring the right of way for roadway expansion has become increasingly difficult and expensive. Thus, there has been a significant movement by the government, private industry, and the research community to develop and implement alternate solutions to alleviate congestion. These solutions utilize advanced technologies to increase the efficiency of today's transportation network. Some of these solutions include Advanced Traffic Control System (ATCS), Advanced Traffic Management Systems (ATMS), Advanced Traveler information System (ATIS), and Ramp Metering and Managed Lane Strategies.

In these advanced traffic congestion mitigation solutions, microscopic simulation is being viewed as a tool that is able to increase the capabilities of the solutions. This is particularly the case for solutions that involve providing real-time estimates of performance measures to transportation system managers and travelers to facilitate more informed decisions as it relates to more efficient roadway operation and usage. A few of these solutions include those proposed in (Miska 2007) and (Henclewood et al. 2010). The proposed solutions rely on microscopic simulation, and as such, for results to better reflect field performance measures the associated microscopic models ought to be appropriately calibrated. To date, the predominant means of calibrating a microscopic simulation model is based on selecting a set of calibration parameters that allows the model to reflect the average field performance measures that were collected at a par-

ticular point in time. Although this method might be sufficient to provide the improvement in the model, given the intentions of the modeling effort, applying this method to a time-dependent modeling effort may be less the optimal. Toward this end this paper presents a case for real-time calibration of real-time / time-dependent model efforts while presenting an alternate calibration procedure.

The goal of employing a real-time calibration procedure when undertaking a real-time / time-dependent modeling effort is to allow the corresponding simulation model to better represent dynamic underpinnings that make up traffic related performance measures. Two of the more dominant reasons for the dynamicity in these underpinnings are the constant changes in driving behavior and driving environment. A real-time calibration procedure facilitates adjustments to the values of the various calibration parameters so that the effects of the behavior and environment changes can be captured appropriately.

In making the case for a real-time calibration procedure the following section presents a brief literature review of previous calibration efforts as it relates to the microscopic modeling of surface transportation. Afterwards, a real-time, data-driven, microscopic simulation experiment is described. The procedure that was used to calibrate the associated simulation model is also presented. The subsequent results are then presented and a real-time calibration procedure for time-dependent modeling efforts is proposed.

2 STATE OF THE PRACTICE - MICROSCOPIC TRAFFIC SIMULATION CALIBRATION

A variety of methods have been developed to calibrate microscopic traffic simulation models. Hollander and Liu presented a rather comprehensive review of many of the current calibration methods, while attempting to highlight the fundamental requirements for calibrating microscopic simulation models (Hollander and Liu 2008). Zhang and Ma also surveyed a number of calibration methods and grouped them in three categories, 1) trial-and-error heuristics, 2) genetic algorithms, and 3) simulated annealing (Zhang and Ma 2008). Among these three categories, a majority of calibration methods fall into the first two categories. Trial-and-error heuristic methods tend to be more frequently selected than genetic algorithm methods as they are generally less complex. As for the trial-and-error methods, Chu and Liu developed a four-step process that includes calibration of driver behavior models, route choice, origin-destination estimation and model fine tuning (Chu and Liu 2004). Also, Oketch and Carrick presented a method that first determines suitable values for model parameters such as, aggressiveness, awareness, target headways and reaction times that provided realistic results and second estimates representative origin-destination (OD) matrices (Oketch and Carrick 2005). Toledo et al. employed an iterative approach that calibrated a model by jointly estimating OD flows and values for behavioral parameters (Toledo et al. 2004). Efforts presented by (Dowling, Holland, and Huang 2002; Dowling, Skabardonis, and Alexiadis 2004; Dowling et al. 2004) illustrated a series of iterations to a method whose foundation employed a four step process to calibrate a model. These four steps include error checking, calibration for capacity, calibration for demand, and overall review. As for calibration methods that belong to the genetic algorithm category, two of the more noteworthy applications were presented in (Park and Won 2006) and (Zhang, Ma, and Dong 2008).

In addition to the various methods used to calibrate microscopic simulation models, numerous criteria have been established to determine when a model is calibrated. Hollander and Liu provide a tabulated summary of a number of these calibration criteria (Hollander and Liu 2008). In examining these criteria and those established by others one notices that many tend to be subjective due to their dependence on what is being modeled and the goals of the modeling effort (Hellinga 1998). To date, the criteria to determine whether or not a model is calibrated typically involve parametric, first moment statistical comparisons of field and simulated performance measures. For instance, Park and Schneeberger used the results from the *t*-test to compare simulated and field travel time means as the criterion to determine when a model is calibrated (Park and Schneeberger 2003). Park and Won developed a criterion that deemed a model as calibrated when the model's travel time distribution "include the entire field-measured values" (Park, Won, and Yun 2006). Although these criteria may be sufficient to evaluate general traffic performance it is questionable if such means can provide more accurate evaluation for time-dependent traffic performance.

Upon examining the various calibration methods and their associated calibration criteria, one recognizes that many of these works aim to find a set of calibration parameter values that results in a model producing performance measures that is similar to those collected from the field at a particular point in time. After finding this set of values, it is a common practice to use this calibrated parameter set to explore conditions outside of the time period from which the field data was collected. Although such an approach may be acceptable for some modeling efforts, the field conditions can be better represented with a set of parameters from the real-time calibration when undertaking real-time / time-dependent traffic analysis. This research effort is proposing that the values of the calibration parameters be allowed to change with respect to time. To support this proposal the calibration of a model that was a part of a data-driven simulation experiment is presented.

The calibration process presented here involves a comprehensive evaluation of the selection of calibration parameter values and a two-part criteria process. The primary goal for this calibration effort is to increase the quality and accuracy of information that is desired from some of today's more involved simulation models. This goal is a partial response to the desire to communicate both facility level and vehicle specific information to travelers and transportation system managers. It is hoped that this information will in turn arm these consumers with the knowledge to facilitate improved network usage and management. The accuracy of the information provided is therefore paramount (especially on the microscopic scale) hence the need for more robust calibration procedure and criteria to dictate when a model is adequately calibrated

3 THE DATA-DRIVEN SIMULATION EXPERIMENT

The primary goal of this experiment is to determine the ability to provide accurate estimates of field performance measures by driving an online simulation with pertinent traffic data. The experiment employed an extensive set traffic data that was collected as part of the FHWA Next Generation Simulation (NGSIM) program (NGSIM 2011). The NGSIM data was collected on November 8, 2006, between 12:45PM and 1:00PM (referred as Noon period in this paper) and 4:00PM and 4:15PM (referred as Evening period in this paper), along Peachtree Street in Atlanta, Georgia. The segment being studied extended just south of the intersection of Peachtree and 10th Street and north of the intersection of Peachtree and 14th Street. 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, OD data for each vehicle, turning movement distribution data at each intersection, and a series of other traffic related information were also collected. Video of the entire corridor is also available during data collection periods.

For the experiment, a detailed VISSIM model of the study area was created. VISSIM is a discrete, stochastic, time step based microscopic simulation model (PTV 2011). This behavior-based multi-purpose traffic simulation program has been developed to model a wide range of traffic conditions including freeway, arterial, and public transit operations. In this model all vehicles are modeled individually, based on a psycho-physical driver behavior model developed by Wiedemann (PTV 2011). Roadway geometry was based on existing conditions at the time of the experiment and additional information such as vehicle volume, turning movement distribution, routing decisions and signal timing plans were added based on the NGSIM data set. 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.

To simulate streaming detector data, trajectory and OD information were used to create a VISSIM trip-chain file, which approximates the process of streaming detector data into the data-driven online simulation process. A trip chain file is able to approximate a detector stream as each file's record consists of a time-stamp, indicating when a vehicle entered the network (i.e. crossed a boundary link detector), and a zone number indicating a vehicle's origin (i.e. the boundary detector crossed) and destination. This string of information is similar to that from a detector, except for a vehicle's destination.

3.1 Experimental Results

Upon appropriately inputting the necessary field data into the VISSIM model of the NGSIM study area, 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 difference between VISSIM and field travel time is approximately eight seconds, which occurred for the Noon-South time period. While the largest difference, 32.7 seconds, occurred for the Evening-Southbound period. The simulation does a slightly better job estimating travel times for the Noon period compared to the Evening. 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 variation travel time estimates is rather similar to that of the field. With dissimilar means and “similar” standard deviations it is anticipated that the observed discrepancies may be addressed through a more rigorous calibration effort. Density plots were examined to further corroborate this hypothesis, Figure 1.

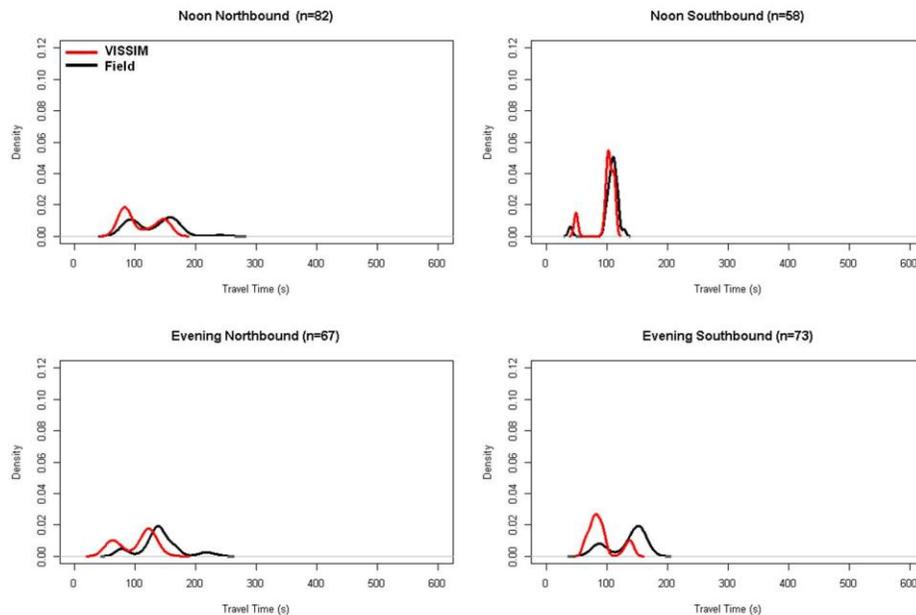


Figure 1 Density Plots of Field vs. VISSIM (single run) Travel Times

The density 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, a very encouraging finding, likely indicating many of the differences may be addressed through a more involved calibration procedure.

4 MODEL CALIBRATION

4.1 Calibration Procedure

The calibration procedure that was employed to calibrate the VISSIM model of the NGSIM corridor is a Monte Carlo inspired approach that was adapted from (D. Miller 2009). The foundation of this approach is the use of ten effective calibration parameters and the creation of unique sets of values for each of these pa-

rameters to sufficiently represent the available sample space (D. Miller 2009; D. Miller et al. 2012). This Monte Carlo approach produced 1000 unique parameter sets that were simulated in the NGSIM model. From these model runs / parameter set simulations, travel time and saturation flow measures were extracted. These measures were then analyzed to determine which combination of parameter values most closely reproduced the NGSIM results, and is therefore a calibrated parameter set.

Figure 2 presents travel time density plots for each NGSIM study period and travel direction. Each of the 1000 VISSIM simulation runs, each containing the described three-batches, produced these density plots. This set of results will be used in conjunction with saturation flow measures to aid in the NGSIM model calibration process.

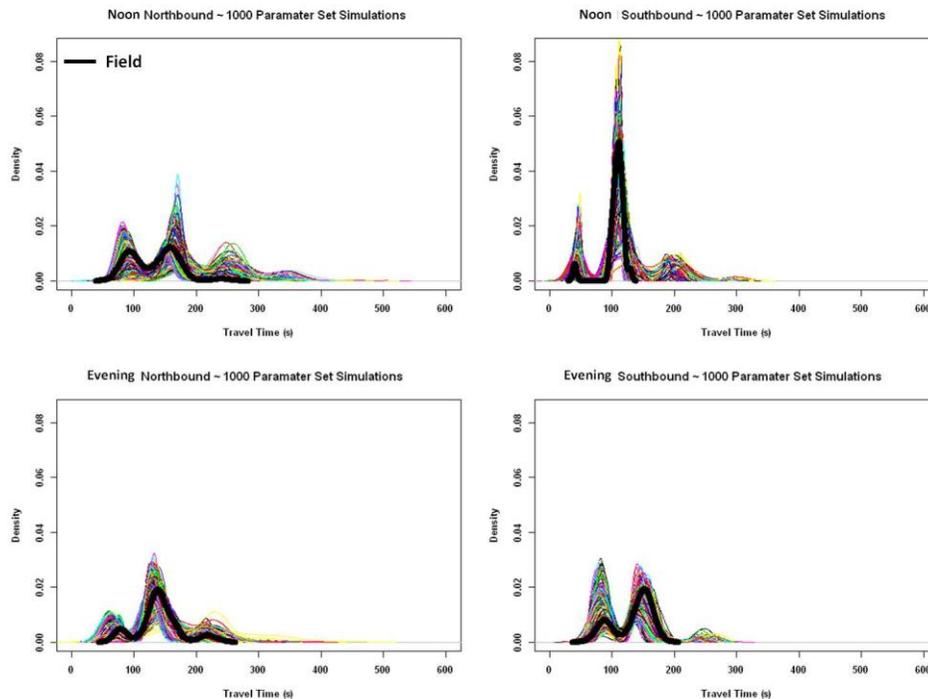


Figure 2 Travel Time Density Plots for Each Period and Direction

4.2 Calibration Criteria

This research effort formulates a two-part criteria process to select calibrated parameter sets. The first part compares field and simulated saturation flow rates to ensure that the models produce reasonable estimates. This is an imperative step as it is possible for models to produce accurate estimates of performance measures, such as travel time, while over or under estimating saturation flow rates. Failure to adequately model saturation flow could prove significant. For example, by over estimating saturation flow a simulation will over estimate capacity. In a scenario analysis 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 part of this process involves the statistical evaluation of the mean and the distribution of the performance measures being studied, travel time in this case. By evaluating travel time means and distributions, the calibrated model will consider reflecting both corridor level and individual vehicle level traffic information, which is necessary for real-time simulation.

4.2.1 Flow Criterion

Establishing a flow criteria not only facilitates greater confidence in the results from an calibrated model but it also provides some level of protection from the potential dangers in the application and implementation of the calibrated model.

Acceptable ranges for flow were established based on appropriate field measurements. Videos of the NGSIM corridor operation were observed and headway measurements collected. Headway measurements were collected and arranged in two groups. One group contained measurements from the second vehicle in the queue to fifth vehicle, while the next group contained measurements from the sixth to the ninth vehicle. These headway measurements were then averaged for each cycle and converted to startup and saturation flows, respectively. Similar measurements were obtained from the 1000 VISSIM models of the NGSIM corridor. Table 1 presents a comparison of the summary statistics for both the field and simulated flow measurements.

Table 1 Summary Statistics of Field and Simulated Startup and Saturation Flow Measurements

Statistic	Field Measures		Simulated Measures	
	Start-up Flow	Saturation Flow	Start-up Flow	Saturation Flow
Min	959	1229	960	798
Mean	1408	1673	1357	1433
Max	1732	2480	1886	2257
Standard Deviation	182	281	181	306

To create the startup and saturation flow criterion, a reasonable range was 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 was constructed by forming a 95% confidence interval around the mean flow values from the field. A challenge in constructing an appropriate confidence interval is the limited field data as only 28 cycles were observed and particularly for saturation flow estimates only 3 cycles had queue lengths that were greater of equal to nine vehicles. In an attempt to address the negative consequences that may occur because of limited data, such as incorrect variance estimation, a bootstrap approach was taken to bolster the field data that will be used to inferences about field startup and saturation flow estimates.

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. To carry out the bootstrap approach on the field flow measurements 10,000 re-samples were randomly drawn, with replacement, from the startup and saturation flow estimates. (Mooney and Duval 1993)

A 95% confidence interval was constructed using the bootstrapped flow measurements from the field. This confidence interval was produced by the percentile-*t* method and was selected as the reasonable range for simulated flow rates. 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 will now be examined via the statistical evaluation criteria to determine which models are calibrated.

4.2.2 Statistical Evaluation Criterion

For statistical evaluation criteria travel time distributions from each of 159 models were examined. Parametric and non-parametric statistical tools were explored to compare field and simulated travel time estimates. Parametric tools are more commonly used for these types of comparisons. However, the limita-

tion of parametric tools is that they require data to fit a known distribution, typically a normal distribution, and field data often do not fit known distributions. Therefore, to ensure proper comparisons of field and simulated data, non-parametric tools were used.

Two sets of non-parametric tools were used to establish the statistical calibration criterion, one is a general distribution comparison to primarily determine the homogeneity between field and simulated travel estimates. The other set of tools performed a more stringent comparison of travel time distributions. With these tools working in tandem, parameter set simulations that produce accurate estimates of field travel distributions will be considered as calibrated replicates.

4.2.2.1 General Distribution Comparison

The Wilcoxon rank sum and Mann-Whitney tests were identified as appropriate tools for such an evaluation. When comparing populations, the rank-sum test has as its null hypothesis that the “two population are equal – $F_x(x) = F_y(x)$. (Kvam and Vidakovic 2007). In other words, the rank sum test attempts to determine “whether the groups are homogeneous or one group is ‘better’ than the other” (Kvam and Vidakovic 2007). The Mann-Whitney test was also selected as its function is “to find differences in two populations” (Kvam and Vidakovic 2007). Given the similarity in intentions between this test and the rank sum test, and the equivalency of the Mann-Whitney and Wilcoxon test statistic; the Wilcoxon-Mann-Whitney (WMW) test was chosen to evaluate the general differences between field and simulated travel time distributions.

The Wilcoxon-Mann-Whitney test was implemented using R (Gentleman and Ihaka 1993). More specifically, the WMW test was applied using the *wilcox.exact()* command. One of the outputs from the *wilcox.exact()* command is a p -value that is used in the decision to accept or reject the null hypothesis, H_o , which is in essence states that the field and the simulated travel time distributions are likely equivalent. A calibration criterion based on the WMW p -value is the rejection of H_o , 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 travel time distribution.

The need for a more stringent comparison of travel time distributions is due to one of the fundamental principles of the WMW test. Because the WMW relies on ranks, the evaluation of travel time distributions is based on relative differences amongst the data points versus the magnitude of those differences. By not basing the comparison on the magnitudes of these differences, it weakens the distribution comparison between field and simulated measures.

4.2.2.2 Secondary Distribution Comparison

Given the multimodal travel time distributions, with hard to discern parametric descriptors, the Kolmogorov-Smirnov (KS) non-parametric test was selected to further compare the distribution of travel time datasets. The KS test was considered as a secondary distribution test as its test statistic “is the basis of many nonparametric goodness-of-fit for distributions” (Kvam and Vidakovic 2007). In addition, 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. However, given the data at hand, the KS test was applied in conjunction a less formal, non-parametric heuristic form fit (HFF) test to select replicate runs that produced travel time distributions that were more similar to those obtained from the field.

The HFF test was included to provide an alternative distribution comparison method whose assumption(s) were not violated by the data at hand, unlike the KS test. One of the assumptions of the KS test is that data being analyzed is continuous (Law and Kelton 1991; Kvam and Vidakovic 2007). The field travel times for time experiment is not continuous and it is expected that there will be other circumstances under which such measures will not be continuously distributed. The lack of continuous travel times is in part due to signalized intersections not affording the realization of some travel time values. Although the KS test can still be applied to discontinuous data, inferring from possibly incorrect results could prove

costly. Incorporating the HFF test is an attempt to mitigate potential incorrect inferences and subsequent elimination of possible calibrated parameters sets. It is in this vein that both the KS and the HFF test will be used to evaluate distribution fits.

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 KS test was also implemented in the R environment using the $ks.test()$ command that includes an associated p -value as a part of its output (Gentleman and Ihaka 1993).

To limit erroneous inferences from the KS test, particularly in the instances when the data being compared is discontinuous, the HFF test is devised to compare the rate of change of the CDFs 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;

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

There are no assumptions associated with the HFF test. To make inferences from the HFF test consider the magnitude of H . The smaller the magnitude of H , the more likely that the parameter set simulation being compared will be calibrated. A disadvantage of this method is that it does not take into account shifts along the x -axis, i.e. differences in central tendencies (mean/median). This is concerning as a 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.

The heuristic nature of the HFF test also aids in selecting a calibrated model. For a given travel time segment 1000 values of H will be calculated. 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 calibration model candidates.

The statistical evaluation criteria uses the WMW, KS and HFF tests to thoroughly examine the field and simulated travel time distributions. All three statistical tests were used to determine which parameter set satisfy the statistical evaluation criteria. These tests were only performed on the 159 parameter set simulations that satisfied the above saturation flow criteria. The results from these tests will further determine which parameter sets are calibrated. 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 to perform a general comparison of field and simulated travel time distributions.
2. Retain parameter set simulations whose WMW test yielded p -values ≥ 0.01 . This will be a set of parameter sets denoted by $M_n|U$.
3. Conduct the KS test to compare the distributions of field and simulated travel times.
4. Retain parameter set simulations whose KS test yielded p -values ≥ 0.01 . This will be a set of parameter sets denoted by $M_n|D$.
5. Conduct the HFF test to compare the shapes of the distributions of field and simulated travel times.
6. 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 $M_n|H$.

7. To obtain the set of parameter set simulations that satisfy the statistical evaluation criteria, carry out the following set operation

$$M_f = M_n|U \cap (M_n|D \cup M_n|H)$$

5 RESULTS AND DISCUSSION

The two-part criteria process was carried out for each time period and direction of travel in the NGSIM data set (12PM and 4PM, Northbound and Southbound). Using only the parameter set simulations that satisfy part one of the calibration criteria, the next set of parameter sets simulations were eliminated after applying the WMW test. The remaining parameter set simulations, $M_n|U$, had sufficiently similar distributions that there was insufficient evidence to reject H_0 . Similarly, the KS and HFF tests were conducted. The final set of parameter set simulations that were obtained after applying the above set operation. The remaining set of parameter simulations are deemed calibrated – according to the two-part criteria process. Table 2 presents a summary of the number of parameter set simulations that were retained as the various calibration criterion were applied.

Table 2 Number of Models Remaining After the Application of Each Calibration Criteria

Criteria	12 NB	12 SB	4 NB	4 SB
	Number of Models			
Initial	1000	1000	1000	1000
Saturation Flow	159	159	159	159
Statistical Evaluation				
WMW	44	93	35	2
KS	51	101	35	2
HFF	91	122	100	158
Satisfied Criteria	44	92	34	2

A final set of calibrated parameter sets must next be selected based on comparisons with four different field datasets – two periods with two travel directions. The preceding analysis resulted in a selection of parameter sets that are adequately calibrated for each time period – direction alternatives, $M_{f12/NB}$, $M_{f12/SB}$, $M_{f4/NB}$, and $M_{f4/SB}$, where the subscript represents time period and direction (i.e. $M_{f12/NB}$ represents the set of adequately calibrated replicates for the 12PM, northbound traffic). 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 3 presents the number of parameter sets that are the same for different time periods and travel direction. From the table below, one observes that there was 43 simulation models that were calibrated for Noon period, and two for the Evening period. And that there was only one model that was calibrated for both periods and travel directions.

Table 3 Number of Common Models for each Approach and Time Period

	12-NB	12-SB	4-NB	4-SB
12-NB	44	43	29	1
12-SB	43	92	31	2
4-NB	29	31	34	2
4-SB	1	2	2	2
Common to all Periods and Directions		1		

Given the various sets of parameter values that produced calibrated simulation models, the next step is deciding which parameter set(s) should be used for real world applications. Two types of objectives are presented to aid in this decision.

The first objective is associated with planning purposes, which typically involves the evaluation and communication of corridor level traffic information, such as average travel time and volume. If the intention of the model is to provide general information regarding the traffic's current and future conditions, especially as it relates to the effects of certain strategies or policies, or changes in demand, any of the calibrated parameter sets may be used. In the case of the NGSIM study corridor, any one of the 95 parameter sets may be used as they were deemed adequately calibrated for at least one period and direction. This recommendation does allow for different time periods to utilize different parameter sets. However, where a parameter set(s) is found to satisfy all time periods and directions, one as in this case, it is recommended that this parameter set be included in the analysis.

To obtain a more comprehensive view of the traffic's performance, it is reasonable to perform a study where not only the random seed is altered for each run but the parameter set may also be changed. In this instance, for each enhanced model run, a parameter set is randomly selected from the family of calibrated sets. This approach parallels, and improves upon, a multi-run simulation analysis. Instead of only altering the random seed of a calibrated model, the values of the entire set of effective calibration parameters are changed. Such a change will provide greater insight into the variability of the simulated output. As a precautionary measure, it is recommended that throughout this analysis, the modeler should also monitor saturation flow measures in addition to the other performance measures, to ensure that a simulation's output remains valid.

The second category of objectives is more temporal in nature. This refers to models that are created to communicate corridor and individual vehicle level traffic information in a time-base or real-time fashion. For such purposes, it is advised that the calibrated replicates that best reflect traffic conditions during the analysis period be used. If traffic information about the Noon period is requested, any number of the 136 parameter sets should be used to deliver the relevant information. Again, using more than one of these parameter sets is encouraged, making the information more robust.

Given the above temporal objective, one of the more immediate expansions of the effort is to have an array of time-of-day calibrated parameter sets which are intended to provide time-based traffic information for periods in which driver behavior maybe be deemed homogeneous. The use of these time-of-day based parameter sets is analogous to applying different signal timing plans, throughout the day, for an intersection. It is expected that there will be multiple calibrated parameter sets for each period. And similar to previous scenarios, these parameter sets should be used as a part of a multi-run simulation analysis – providing a comprehensive view of anticipated performance measures. A further expansion of the time-of-day parameter set construct is the development of a real-time calibration algorithm. The goal of such an algorithm is to allow a real-time data driven simulation model to adapt to changes beyond those of the physical transportation network. This algorithm will be responsible for intelligently adjusting the effective calibration parameters in order for the simulated environment to continue accurately estimating measures – even as drivers' behavior and environment changes.

6 CONCLUSION

As the transportation community continues to develop tools to alleviate congestion, traffic simulation is beginning to play a more significant role. This is particularly the case when attempting to employ real-time traffic information to manage and use the transportation facility more efficiently. In light of the growing prominence of traffic simulation, particular emphasis ought to be placed on the calibration of the associated models. Toward this end, a number of calibration methods that have been employed in practice was highlighted. These methods have been successful in producing calibrated models. However, they are often calibrated using data from a single time period and it is hypothesized that that the resulting calibrated model may not be able to sufficiently represent traffic conditions outside of that time period. This

research effort explored this hypothesis by employing a Monte Carlo based calibration procedure to the VISSIM model that is associated with a data-driven experiment.

The Monte Carlo based procedure produced 1000 unique candidate parameter sets to calibrate a VISSIM model. To determine which of the 1000 models were well calibrated, a two-part criteria process was applied to the models' output. There were 93 and 34 calibrated models belonging to Noon and Evening time periods respectively. This discrepancy in the number of calibrated models is one of the first indicators that different periods do require different set of values for pertinent calibration parameters. In light of these results it is recommended that for real-time or time-dependent simulation efforts, a corresponding calibration procedure ought to be developed and implemented to bolster the inferences that may be made via simulations' output, even with changing time-periods.

REFERENCES

- Chu, Lianyu, and Henry X Liu. 2004. A Calibration Procedure for Microscopic Traffic Simulation. In *83rd Transportation Research Board Annual Meeting*, 1-25. Washinton DC.
- Dowling, Richard, Joseph Holland, and Allen Huang. 2002. *Guidelines for Applying Traffic Microsimulation Modeling Software*. Oakland.
- Dowling, Richard, Alexander Skabardonis, and Vassili Alexiadis. 2004. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software*. Vol. III. Oakland.
- Dowling, Richard, Alexander Skabardonis, John Halkias, Gene McHale, and Grant Zammit. 2004. "Guidelines for Calibration of Microsimulation Models: Framework and Applications." *Transportation Research Record* 1876 (1) 1-9.
- Gentleman, Robert, and Ross Ihaka. 1993. The R Project for Statistical Computing. <http://www.r-project.org/>.
- Hellinga, BR. 1998. Requirements for the Calibration of Traffic Simulation Models. In *Canadian Society for Civil Engineering 1998 Annual Conference*, 211-222. Halifax, Nova Scotia. <http://www.civil.uwaterloo.ca/bhellinga/publications/Publications/CSCE-1998-Calibration.PDF>.
- Henklewood, Dwayne, Angshuman Guin, Randall Guensler, Richard Fujimoto, and Michael Hunter. 2010. Real-Time Data Driven Arterial Simulation for Performance Measures Estimation. In *Winter Simulation Conference*, ed. B. Johansson, S. Jain, J. Montoya-Torres, J. Hukan, and E. Yücesan, 13. Baltimore.
- Hollander, Yaron, and Ronghui Liu. 2008. "The principles of calibrating traffic microsimulation models." *Transportation* 35 (3) 347-362.
- Kvam, Paul H., and Brani Vidakovic. 2007. *Wiley Series in Probability and Statistics, in Nonparametric Statistics with Applications to Science and Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Miller, David. 2009. Developing a Procedure to Indentify Parameters for Calibratoin of a VISSIM Model. Georgia Institute of Technology.
- Miller, David M, Dwayne Henklewood, Michael Rodgers, and Michael Hunter. 2012. Parameter Selection Procedure for Signalized Arterial Simulation Calibration. In *TRB 91st Annual Meeting Compendium of Papers CD-ROM*. Vol. 7228. Washinton DC.
- Miska, Marc Philipp. 2007. Microscopic Online Simulation for Real time Traffic Management. Delft University of Technology.
- Mooney, Christopher Z., and Robert D. Duval. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage Unive. Newbury Park, CA: Sage.
- NGSIM. 2011. The NGSIM Community Website. <http://ngsim-community.org/>.
- Oketch, T, and M Carrick. 2005. Calibration and validation of a micro-simulation model in network analysis. In *84th Transportation Research Board Annual Meeting*, 17. Washinton DC: Transportation Research Board.

- Park, Byungkyu (Brian), and J D Schneeberger. 2003. "Calibration and Validation Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal System." *Transportation Research Record* 1856: 185-192.
- Park, Byungkyu (Brian), and Jongsun Won. 2006. *Microscopic Simulation Model Calibration and Validation Handbook. Network*. Chapel Hill, NC.
- Park, Byungkyu (Brian), Jongsun Won, and Ilsoo Yun. 2006. "Application of Microscopic Simulation Model Calibration and Validation Procedure: Case Study of Coordinated Actuated Signal System." *Transportation Research Record* 1978.
- PTV. 2011. VISSIM 5.10 User Manual. Karlsruhe, Germany.
- Schrank, David, Tim Lomax, and Bill Eisele. 2011. *Urban Mobility Report 2011. Report for the Texas Transportation Institute*. College Station.
- Toledo, Tomer, M Ben-Akiva, Deepak Darda, Mithilesh Jha, and Haris N. Koutsopoulos. 2004. "Calibration of microscopic traffic simulation models with aggregate data." *Transportation Research Record* 1876.
- Zhang, Michael, and Jingtao Ma. 2008. *Developing Calibration Tools for Microscopic Traffic Simulation Final Report Part 1: Overview Methods and Guidelines on Project Scoping and Data Collection*. Berkeley.
- Zhang, Michael, Jingtao Ma, and Hu Dong. 2008. *Developing Calibration Tools for Microscopic Traffic Simulation Final Report Part II: Calibration Framework and Calibration of Local / Global Driving Behavior and Departure / Route Choice Model Parameters*. Berkeley.

AUTHOR BIOGRAPHIES

DWAYNE HENCLEWOOD is a Ph.D. Candidate at the Georgia Institute of Technology. His current research interest is in arterial performance evaluation and microscopic traffic simulation. He holds a M.S. in Civil Engineering from the University of Massachusetts Amherst. His email address is <dahen-cle@gatech.edu>.

WONHO SUH is a Research Engineer in the School of Civil and Environmental Department at the Georgia Institute of Technology. His research interests include traffic simulation, intelligent transportation systems, and transportation safety. His e-mail address is <wonho.suh@ce.gatech.edu>.

MICHAEL RODGERS is the director of the Georgia Tech Air Quality Laboratory. His areas of interest are modeling and simulation of vehicle activity and emissions and statistical analysis of vehicle emissions data. He earned his Ph.D. at the Georgia Institute of Technology. His e-mail address is <michael.roddgers@ce.gatech.edu>.

MICHAEL HUNTER is an Associate Professor in the School of Civil and Environmental Department at the Georgia Institute of Technology. His primary areas of interest are traffic safety, operational performance and control, and simulation. He earned his Ph.D. at the University of Texas at Austin in August 2003. His e-mail address is <michael.hunter@ce.gatech.edu>.

RICHARD FUJIMOTO is a Regents' Professor and Chair of the School of Computational Science and Engineering at the Georgia Institute of Technology. He received his M.S. and Ph.D. degrees from the University of California (Berkeley) in 1980 and 1983 respectively. Among his past activities he lead the definition of the time management services for the DoD High Level Architecture (HLA), and has also been a member of the steering committee for the Workshop on Parallel and Distributed Simulation, (PADS) since 1990. His email address is <fujimoto@cc.gatech.edu>.