Paper No. 03-2531

Microscopic Simulation Model Calibration and Validation

Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal System

Byungkyu (Brian) Park and J. D. Schneeberger

Microscopic simulation models have been widely used in both transportation operations and management analyses because simulation is safer, less expensive, and faster than field implementation and testing. While these simulation models can be advantageous to engineers, the models must be calibrated and validated before they can be used to provide meaningful results. However, the transportation profession has not established any formal or consistent guidelines for the development and application of these models. In practice, simulation model–based analyses have often been conducted under default parameter values or best-guessed values. This is mainly due to either difficulties in field data collection or lack of a readily available procedure for simulation model calibration and validation. A procedure was proposed for microscopic simulation model calibration and validation and an example case study is presented with real-world traffic data from Route 50 on Lee Jackson Highway in Fairfax, Virginia. The proposed procedure consisted of nine steps: (*a*) measure of effectiveness selection, (*b*) data collection, (*c*) calibration parameter identification, (*d*) experimental design, (*e*) run simulation, (*f*) surface function development, (*g*) candidate parameter set generations, (*h*) evaluation, and (*i*) validation through new data collection. The case study indicates that the proposed procedure appears to be properly calibrating and validating the VISSIM simulation model for the test-bed network.

Advances in computational technology along with the increased complexity of roadway design and management have created an environment in which microscopic simulation models have become useful tools for transportation engineers. Microscopic simulations can be used in several different transportation areas. They can be used to evaluate alternative timing plans and geometric changes before implementing the design in the field. They are also appealing for estimating certain quantities that are not easily estimated or observed in the field. These include air-quality impacts, fuel consumption rates, accident risk factors, and toll revenues. Finally, these simulation tools can also be used to evaluate emerging technologies such as intelligent transportation systems.

Microscopic simulation models contain numerous independent parameters to describe traffic control operation, traffic flow characteristics, and drivers' behavior. These models contain default values for each variable, but they also allow users to input a range of values for the parameters. Changes to these parameters during calibration should be based on field-measured conditions and should be justified and defensible by the user.

Unfortunately, many of the parameters used in simulation models are difficult to measure in the field, yet they can have a substantial impact on the model's performance. Examples of some of these variables in microscopic simulation models could include start-up lost time, queue discharge rate, car-following sensitivity factors, time to complete a lane change, acceptable gaps, and driver's familiarity with the network. This is why skeptics often view simulation modeling as an inexact science at best and an unreliable "black-box" technology at worst (*1*). This skepticism usually results from unrealistic expectations of the capabilities of simulation models and use of poorly calibrated and/or validated models (*1*).

It is understood that microscopic simulation model-based analyses often have been conducted under default parameter values or best-guessed values. This is mainly due to either difficulties in field data collection or lack of readily available procedures (or guidelines) on the simulation model calibration and validation. At times, simulation model outputs could result in unrealistic estimates of the impacts of new treatments if the simulation model is not properly calibrated and validated. Thus, calibration and validation of simulation models are crucial steps in assessing their value in transportation policy, planning, and operations. Sacks et al. (*2*) indicated that simulation model calibration and validation are often discussed and informally practiced among researchers but have not been formally proposed as a procedure or guideline.

To make simulation models look real, model calibration and validation are necessary. Model calibration is defined as the process by which the individual components of the simulation model are adjusted or tuned so the model will accurately represent field-measured or observed traffic conditions (*3*). The components or parameters of a simulation model requiring calibration include traffic control operations, traffic flow characteristics, and drivers' behavior. Model calibration is not to be confused with validation. Model validation tests the accuracy of the model by comparing traffic flow data generated by the model with that collected from the field (*3*). Validation is directly related to the calibration process because adjustments in calibration are necessary to improve the model's ability to replicate field-measured traffic conditions.

Hellinga (*1*) described a calibration process consisting of seven component steps: (*a*) defining study goals and objectives, (*b*) determining required field data, (*c*) choosing measures of performance, (*d*) establishing evaluation criteria, (*e*) network representation, (*f*) driver routing behavior, and (*g*) evaluation of model outputs.

B. Park, Department of Civil Engineering, University of Virginia, and Virginia Transportation Research Council, P.O. Box 400742, Charlottesville, VA 22904-4742. J. D. Schneeberger, Department of Civil Engineering, University of Virginia, P.O. Box 400742, Charlottesville, VA 22904-4742.

This process provides basic guidelines but does not give a direct procedure for conducting calibration and validation.

Sacks et al. (*2*) recognized four key issues on model validation. They are (*a*) identifying the explicit meaning of validation in a particular context, (*b*) acquiring relevant data, (*c*) quantifying uncertainties, and (*d*) predicting performance measures under new conditions. They demonstrated an informal validation process with the CORSIM (*4*) simulation model and emphasized the importance of data quality and visualization. The authors have not found any formal procedure for simulation model calibration and validation.

The objectives of this paper are to develop a procedure for simulation model calibration and validation and to demonstrate the proposed procedure through a case study. The scope of the work is limited to a coordinated actuated signal system with VISSIM (*5*), a microscopic simulation model.

TEST SITE AND SIMULATION MODEL

Test Site

An urban arterial street network in Fairfax, Virginia, was chosen for the test site. The site, presented in Figure 1, consists of an arterial, Lee Jackson Memorial Highway (U.S. Route 50), and 12 coordinated actuated signals between Sully Road and the Fairfax County Parkway. The 12 intersections contained coordinated actuated signal control. The site was also chosen because of the ease with which signal timing plans for the 12 intersections could be extracted from the Management Information System for Transportation (MIST) terminal located in the Smart Travel Laboratory at the University of Virginia. This system is directly linked to the timing plans used in the field test site and therefore provides access to real-time data.

Simulation Model: VISSIM

The simulation model used in this research was VISSIM (*5*), version 3.50. VISSIM is a microscopic, time step, and behavior-based simulation model. The model was developed at the University of Karlsruhe, Germany, during the early 1970s and the commercial distribution of VISSIM was launched in 1993 by PTV Transworld AG. In the United States, ITC Inc. distributes and supports the program.

Essential to the accuracy of a traffic simulation model is the quality of the actual modeling of vehicles or the methodology of moving vehicles through the network. VISSIM uses the psychophysical driver behavior model developed by Wiedemann (*5*). The basic concept of this model is that drivers of faster-moving vehicles start to decelerate as they reach their individual perception threshold to a slower-moving vehicle. Because they cannot exactly determine the speed of that vehicle, their speed will fall below that vehicle's speed until they start to slightly accelerate again after reaching another perception threshold. This results in an iterative process of acceleration and deceleration. Detailed descriptions of calibration parameters can be found in the Implementation of Proposed Procedure section.

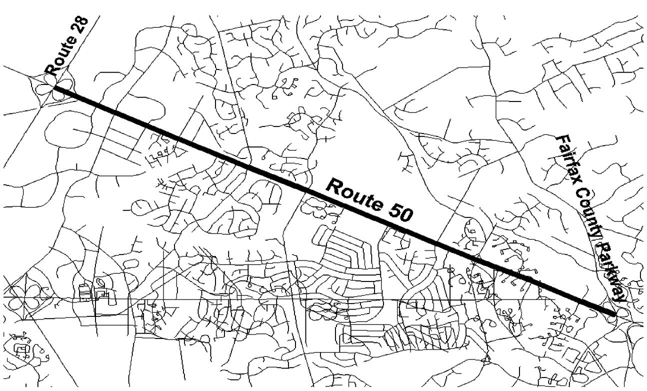

FIGURE 1 Test site: Lee Jackson Memorial Highway, Fairfax, Virginia.

PROPOSED PROCEDURE

The proposed procedure developed for the calibration and validation of microscopic simulation models is similar to those discussed in the literature review. While Hellinga's procedure set forth a general set of guidelines for the calibration and validation process, the proposed procedure is a more detailed step-by-step approach.

Determination of Measures of Effectiveness

The first step in the calibration and validation process is to determine measures of effectiveness appropriate for calibration and validation. In this step, one has to determine a performance measure and identify uncontrollable input parameters and controllable input parameters. The performance measure could be an average travel time between two data collection points in the network. Uncontrollable input parameters may include existing geometry, traffic counts, current signal timing plans, and so forth. Controllable input parameters in the simulation program may include lane-changing distances, waiting times before diffusion, minimum headways, minimum and maximum look-ahead distances, and so forth. It is important to clearly identify all measures of effectiveness before proceeding forward in the calibration and validation process.

Data Collection

Once the measures of effectiveness have been identified, the next step in the calibration and validation process is data collection from the field. Performance measures and uncontrollable input parameters should be collected from the field.

Identification of Calibration Parameters

All calibration parameters within the microscopic simulation model must be identified. Examples of the controllable calibration parameters are lane-change distance, desired speed, and minimum headway distances in the simulation model. Acceptable ranges for each of the calibration parameters should be determined.

Experimental Design

An experimental design is important because the number of combinations among feasible controllable parameters is so large that possible scenarios cannot be evaluated in a reasonable time. For example, in the case of 8 controllable input parameters with each parameter having a range of 5 values, there are 5^8 (i.e., 390,625) possible combinations of parameters. The number of simulation runs would be much larger if multiple simulation runs were required to reduce stochastic variability. An experimental design, one example being a Latin hypercube (LH) design algorithm, could reduce the number of combinations to 3^6 (i.e., 729) or 5^4 (i.e., 625) combinations. LH sampling is generally more precise for producing random samples than conventional Monte Carlo sampling, because the full range of the distribution is sampled more evenly and consistently. Thus, with LH sampling, a smaller number of trials achieve the same accuracy as a larger number of Monte Carlo trials. The purpose of a LH design is to minimize correlations among parameter dimensions and maximize the coverage of the entire parameter surface. Further detailed discussions can be found elsewhere (*6*).

Run the Simulation *N* Times

To reduce stochastic variability, multiple runs must be conducted for each scenario from the experimental design. The average performance measure and standard deviation should be recorded for each of the runs.

Development of Surface Function

A surface function, using the calibration parameters and measure of performance, should be created from the results of the multiple runs. Examples of surface functions could be linear and nonlinear regression models. In case within-group correlations affect model performance, a mixed-effects model or some other technique can be explored.

Determination of Parameter Sets Based on Surface Function

The purpose of this step is to find an optimal parameter set that provides a close match with the field performance measure. Because there could exist several parameter sets providing output close to the target (i.e., field performance measure), one should consider several parameter sets.

Evaluation of Parameter Sets

In this step, multiple runs are conducted to verify whether the parameter sets identified in the previous step generate statistically significant results. For each parameter set, a distribution of performance measure is developed and compared with the field measure. Statistical tests, including the *t*-test and the Kolmogorov–Smirnov test, can be conducted. For example, if the performance measure is an average travel time and individual travel times are available, field individual travel times will be compared with the individual travel times from the simulation model output of the 50th and 90th percentile cases. One of the most important factors in the simulation model calibration process is to watch the animations. Visualization is a very powerful tool that can easily identify any anomalies. For example, at times CORSIM creates an "intersection blockage" situation, which hardly occurs in the field. Therefore, on the basis of statistical tests and the animations, one can pick the best parameter set.

Collection of New Data Set for Validation

To perform validation of the microscopic simulation model, a new set of field data under untried conditions should be collected. One way of collecting validation data would be to collect data for different time periods or conditions. For example, if the calibration data were collected during the morning peak, then one should try to collect validation data from the evening peak or nonpeak period. One could also consider options for using different performance measures such as queue lengths or stopped delay at selected links.

IMPLEMENTATION OF PROPOSED PROCEDURE

The proposed procedure was applied in a case study with VISSIM. This section describes the procedure and the results from the case study of an urban arterial in Northern Virginia.

Measure of Performance Selection

Two measures of performance were selected for the calibration and validation process. The first measure of performance was eastbound left-lane travel times on Lee Jackson Memorial Highway, which was used for calibration. The maximum queue length between the intersections of Muirfield Lane and Intel Country Club Road was used as the performance measure for the validation process. These performance measures were chosen because of their ease of collection from the field and from VISSIM output files. Other performance measures such as speeds and delays are not as easily obtainable from the field, but they are from simulation models.

Data Collection

Data collection was required to provide simulation program input parameters and output measures of performance for calibration and validation of the microscopic simulation model. While some data were provided from Virginia Department of Transportation (VDOT) plans, other data were collected directly from the field on two weekdays.

Uncontrollable Input Parameters

The geometric characteristics of the test network were obtained from a SYNCHRO file used by VDOT. Distances between intersections, lengths of left- and right-turn lanes, intersection grades, detector locations, and speed limits were collected from this file. Signal timing plans for the 12 intersections were extracted from the MIST terminal located in the Smart Travel Laboratory at the University of Virginia. This system is directly linked to the timing plans used in the field at the test site and therefore gives real-time data. Phases, splits, minimum green times, offsets, and gap-out times were all taken from the system and used modeling in the microscopic simulation model VISSIM.

Although the MIST system terminal provided detector data, turning movement counts, especially for shared lane approaches, could not be collected without collecting data directly from the field. The MIST system terminal provides 15-min local and system detector data on most, but not all, approaches in the test network. Additionally, the unreliability of traffic volume counts from loop detectors made field collections of volumes a necessity. Field counts included both manual and video counts along the test network. Counts were collected on a normal weekday afternoon, Wednesday July 11, 2001, between 4:45 and 6:15 p.m.

A group of 16 undergraduate and graduate students from the University of Virginia performed simultaneous manual counts along the test network. Manual counts were conducted at locations along the test site where detector data from MIST were not suitable or obtainable. To use the data collected from the manual counts simultaneously with the MIST data, clocks used by the manual counters and the clock at the MIST system were synchronized before data collection. Shared lane approaches were counted manually because loop detectors cannot distinguish between turning and through movements. Manual counts also were conducted where there were no loop detectors or where the loop detectors were not working properly.

Four Partners for Advanced Transit and Highways (PATH) dual cameras were used to conduct the video counts. These cameras, provided by VDOT, are a pole-mounted video system. The cameras were positioned on Lee Jackson Memorial Highway at the intersections of Highland Oaks Drive, Intel Country Club Road, Stringfellow Road, and Lees Corner Road, where they recorded traffic conditions at the intersections.

The videotapes of these intersections were used to obtain accurate volume counts and turning percentages along the arterial and local streets. These video counts are more accurate than manual or loop detector-based MIST counts for several reasons. Manual counts involve some human error, especially when a counter is required to observe more than one movement at high volumes. Because counters are observing in real time, it is possible for them to miss some vehicles when conducting their counts. For example, it was often the case that manual counters missed right-on-red counts because they were busy watching other movements. Loop detector data on congested arterials do not always prove to be accurate either. Thresholds for monitoring a vehicle's presence are not always accurately set. As a result, volumes can be over- or underestimated depending on how the threshold for determining a vehicle's presence is set. Video counts are accurate mainly because the viewer may view the videotape more than once. Therefore, the viewer can concentrate on a single movement and then, when finished, review the tape and observe a different movement.

Measures of Performance

Eastbound leftmost lane travel times on Lee Jackson Memorial Highway were used as a measure of performance. Two video cameras were placed on top of bridges at the beginning and end of the network. The videotapes from two cameras with synchronized clocks positioned on the bridges of Sully Road and Fairfax County Parkway overlooking the test site were used to retrieve travel times. The cameras were focused on the license plates of vehicles traveling in the leftmost lane along Lee Jackson Memorial Highway. The cameras recorded a vehicle's license plate number at the time it entered and left the test network. License plate numbers and times were recorded for every vehicle at the beginning and end of the network and later matched. Subtracting the time the vehicle left the network from the time the vehicle entered gave the vehicle's eastbound travel time in the leftmost lane.

The University of Virginia's Smart Travel Laboratory Van, which uses an AUTOSCOPE video detection system, was used to measure and collect queue lengths, the validation measure of performance. Queue lengths were collected on a different day so they could be used as an untried data set for the validation process. The van was parked along the shoulder of Lee Jackson Memorial Highway between the intersections of Muirfield Lane and Intel Country Club Road. The van provided video of a small segment of Lee Jackson Memorial Highway between the two intersections. Queue lengths in the eastbound direction on Lee Jackson Memorial Highway were collected from the van's videotape by counting the number of vehicles in a queue at the end of red time for each cycle during the data collection period.

Identification of Calibration Parameters

The following sections describe the VISSIM parameters and their acceptable ranges used in the calibration process. These parameters include the emergency stopping distance, lane-change distance, desired speed, number of observed preceding vehicles, average standstill distance, additive part of desired safety distance, waiting time before diffusion, and minimum headway.

Emergency Stopping Distance

The emergency stopping distance defines the last possible position for a vehicle to change lanes. For example, if a vehicle cannot change lanes because of high traffic flows but needs to change lanes to stay on its route, it will stop at this position to wait for an opportunity to change lanes. The emergency stopping distance is assigned for each link in the network. The default emergency stopping distance is 5.0 m. Acceptable ranges for the emergency stopping distance were determined to be 2.0 to 7.0 m. These values were chosen because they enabled vehicles to make full use of the link for lane changes (up to 2.0 m, 3.0 m, etc., from the intersection). Larger values were not used because they would limit the amount of space a vehicle had on a link to attempt a lane change, thus forcing a vehicle to stop in the middle of the link and wait for an acceptable gap.

Lane-Change Distance

The lane-change distance parameter is used along with the emergency stopping distance parameter to model drivers' behavior to stay on their desired routes. The lane-change distance defines the distance at which vehicles will begin to attempt to change lanes. The default value for lane-change distance is 200.0 m. Acceptable values for lane-change distance were 150.0 to 300.0 m. These values were selected to ensure a vehicle had a reasonable distance to make a lane change before it reached the intersection. Too small values would force vehicles into the emergency stopped condition.

Desired Speed Distribution

The desired speed distribution is an important parameter with a significant influence on roadway capacity and achievable travel speeds. The desired speed is the speed a vehicle "desires" to travel at if it is not hindered by other vehicles. This is not necessarily the speed at which the vehicle travels in the simulation. If not hindered by another vehicle, a driver will travel at his or her desired speed (with a small stochastic variation called oscillation). The more vehicles differ in their desired speed, the more platoons are created. Any vehicle with a higher desired speed than its current travel speed will check for the opportunity to pass without endangering other vehicles. Minimum and maximum values can be entered in VISSIM for the desired speed distribution. The speed limit on Lee Jackson Memorial highway was 45 mph. Acceptable ranges of speed were 30 to 60 mph, 35 to 55 mph, and 40 to 50 mph. The desired speed distribution used in this exercise was 35 to 55 mph. This value was chosen based on prior experience with VISSIM. The 40- to 50-mph desired speed distribution was too tight. This distribution had all vehicles traveling at a similar speed with little interaction among them. The 30- to 60-mph distribution was not chosen because it did not appear to be reasonable for a vehicle to have a desired speed of 30 mph.

Number of Observed Preceding Vehicles

The number of observed preceding vehicles affects how well vehicles in the network can predict other vehicles' movements and react accordingly. The VISSIM default value for this parameter is two vehicles. One, two, three, and four vehicles were used in this study.

Average Standstill Distance

Average standstill distance defines the average desired distance between stopped cars and also between cars and stoplines, signal heads, and so forth. The default value for average standstill distance is 2.0 m. Acceptable ranges of values used for this parameter were 1.0 to 3.0 m. Larger or smaller values appeared to be unreasonable.

Waiting Time Before Diffusion

Waiting time before diffusion defines the maximum amount of time a vehicle can wait at the emergency stop position waiting for a gap to change lanes to stay on its route. When this time is reached, the vehicle is taken out of the network (diffusion). Sixty seconds is the default value. Other values used in the study were 20 and 40 s.

Minimum Headway

The minimum headway distance defines the minimum distance to the vehicle in front that must be available for a lane change. The default value is 0.5 m. The acceptable range used in the case study was between 0.5 and 7.0 m. The default value appeared to be too small of a distance for vehicles to attempt a lane change. It did not appear realistic that a vehicle would attempt a lane change given headway of 0.5 m. As a result, larger values were assumed to be more reasonable.

Experimental Design

A LH experimental design was used for the case study. LH sampling provides an orthogonal array that randomly samples the entire design space broken down into equal-probability regions. This type of sampling can be looked at as a stratified Monte Carlo sampling where the pairwise correlations can be minimized to a small value (which is essential for uncorrelated parameter estimates) or else set to a desired value. LH sampling is especially useful in exploring the interior of the parameter space and for limiting the experiment to a fixed (user specified) number of runs. The LH technique ensures that the entire range of each variable is sampled. A statistical summary of the model results will produce indices of sensitivity and uncertainty that relate the effects of heterogeneity of input variables to model predictions. The LH design consisted of 124 cases with the VISSIM parameters and three values per parameter.

Multiple Runs

Five random seeded runs were performed in VISSIM for each of the 124 cases for a total of 620 runs. The average eastbound left-lane travel time was recorded for each of the 620 runs. The results from the five multiple runs were then averaged to represent each of the 124 different parameter sets.

Development of Surface Function

A linear regression model was created in S-Plus program with the calibration parameters as independent variables and the eastbound left-lane travel time from VISSIM as the dependent variable Y. The regression model is specific only to this network. A new regression model must be created for each network being considered for calibration and validation. The linear regression model is as follows:

$$Y = 400.88 - 5.10X_1 - 0.68X_2 + 17.80X_3 + 28.63X_4 + 1.77X_5 + 30.20X_6$$

where

Y = eastbound left-lane travel time (s);
X_1 = emergency stopping distance (m) (t value, –2.31; p value, <.0212);
X_2 = lane change distance (m) (t value, –9.15; p value, <.0001);
X_3 = number of observed preceding vehicles (t value, 3.69; p value, .0002);
X_4 = standstill distance (m) (t value, 5.93; p value, <.0001);
X_5 = waiting time before diffusion (s) (t value, 7.32; p value, <.0001); and
X_6 = minimum headway (m) (t value, 11.37; p value, <.0001).

Candidate Parameter Sets

Candidate parameter sets were created with the linear regression model. Microsoft Excel's Solver was used to obtain candidate parameter combination sets. The eastbound left-lane travel time Y was set to a target value of 613.16 s, the travel time observed from the field. The Excel Solver was then used to determine combinations of parameters producing travel time values close to 613.16 s. Eight combinations of parameters were created and are presented in Table 1.

Evaluation of Candidate Parameter Sets with Multiple Runs

Fifty random seeded runs were made for each of the eight candidate parameter sets and evaluated based on two criteria. The first evaluation

TABLE 1 Candidate Parameter Sets

Parameter	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Emergency Stop Distance (m)	2.0	2.0	3.0	2.0	3.0	3.0	3.0	3.0
Lane Change Distance (m)	200	150	200	150	200	200	175	175
No. of Observed Vehicles	3	2	3	3	4	4	4	4
Standstill Distance (m)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Waiting Time Before Diffusion (s)	60	60	60	60	60	60	60	60
Minimum Headway (m)	2.5	2.5	3.0	2.0	3.0	3.5	3.0	2.5

criterion was distribution of eastbound left-lane travel times produced from VISSIM. The second criterion was visualization. Based on these criteria the best parameter set was selected.

Travel Time Distributions

Eastbound leftmost lane travel times were collected from 50 random seeded runs. Mean and median eastbound leftmost lane travel times for each of the eight cases are presented in Table 2. The results from these runs were compared with the travel times collected from the field. The average field travel time was 613.16 s with a standard deviation of 66.2. It should be noted that the field data were collected on a single day. It is not known whether the field data are an average representation of the travel times on Route 50. The data collected from the field may be average, but they may also be lower or higher than the true mean. Instead of comparing the average travel time from the 50 simulation runs with the field data, it is better to compare the field data with the distribution of the 50 runs. By comparing the field data with the distributions of travel times, variability is considered.

The *t*-test was used to determine whether the results from VISSIM were statistically equal to the field travel times. The *t*-test was applied with the VISSIM data from each case that was closest to the average field data. Table 2 also presents the results from the *t*-test. The "percentile field value" column indicates the percentage of runs that were less than the average field travel time. The *t*-test was conducted in this manner because of the variability in the runs. Because of this variability, the *t*-test is extremely sensitive. In most cases, the *t*-test indicates that the field distribution of leftmost lane travel time is statistically different from simulation-based results except for a few runs that are close to the field average travel time. Performing the *t*-test this way indicates that VISSIM represented the field conditions at least once in the 50 random seeded runs for Cases 2 to 7.

Visualization

The importance of visualization when using microscopic simulation models cannot be overemphasized. The purpose of the microscopic simulation model is to represent the field conditions as closely as possible. A model cannot be deemed calibrated if the animations are not realistic. For example, a parameter set may be statistically acceptable but the animations may not be realistic. Then, the model is not acceptable. Animations from the 50 multiple runs were viewed to identify at which percentile animations were not acceptable. An example of an unacceptable animation is presented in Figure 2.

In this VISSIM screenshot, unrealistic problems occur in the westbound direction that were not observed in the field. Vehicles in the figure are attempting to make lane changes at the stop bar. Vehicles in the leftmost lane are trying to make right turns and vehicles in the rightmost lane are attempting to make left turns. The vehicles were not able to make their desired lane change within the link and therefore are at an emergency stopped position. They will stay at this position, blocking other vehicles, until they are able to change lanes or they are kicked out of the system (due to waiting time before diffusion time). Regardless of the travel times produced from an animation like this, this parameter set cannot be chosen because its visual does not represent the real world.

Animations of each case were viewed to determine whether the animations were realistic or unrealistic. Each case was viewed at several different travel time percentiles to determine whether the

TABLE 2 Evaluation of Candidate Parameter Sets with Leftmost Lane Travel Time

Case	Mean (s)	Median (s)	Standard Deviation	Percentile Field Value	*t*-test (*p*-value)
1	449.6	439.5	57.5	100%	0.00
2	603.2	596.1	88.3	54%	0.42
3	473.7	465.3	64.3	98%	0.26
4	688.9	681.9	115.5	24%	0.40
5	467.9	455.9	59.3	98%	0.04
6	519.0	508.7	70.7	92%	0.85
7	523.2	523.2	70.6	90%	0.72
8	485.3	476.6	63.7	98%	0.00

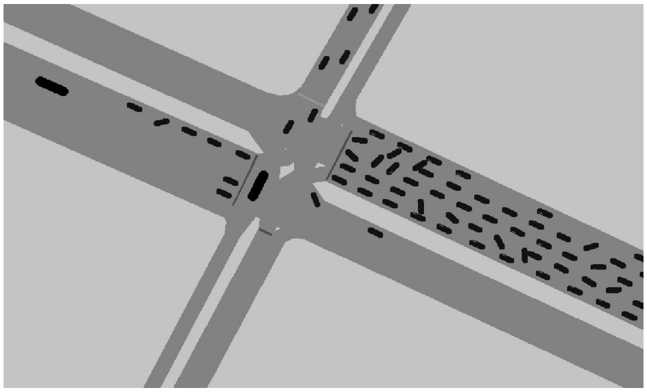

FIGURE 2 Example of unacceptable animation in VISSIM.

animations were realistic or not. It was found that Cases 2 and 4 were not acceptable.

Parameter Set Selection

Parameter Set 7 was chosen as the best parameter set based on its travel time distribution (Figure 3), statistical tests, and animations. Parameter Sets 2 and 4 were eliminated because their animations were unrealistic. Parameter Set 7 produced travel times closest to times in the field from the remaining six cases. Parameter Sets 6 and 7 produced similar results, but Parameter Set 7 was chosen because its animations were more realistic.

Validation with New Data

The eastbound maximum queue length between the intersections of Muirfield Lane and Intel Country Club Road was used for validation. It is noted that the maximum queue length data were collected on a different day and the input volumes used for the validation process were untried. The maximum queue length observed in the field was compared with the distribution of 100 runs in VISSIM. The field maximum queue length was about the top 90% of simulated distribution as indicated in Figure 4.

A comparison of the field data, the uncalibrated VISSIM model (default parameters), and the calibrated VISSIM model indicates the importance of calibration and validation for microscopic simulation models. Eastbound left-lane travel times on Lee Jackson Memorial Highway are compared in Figure 5. The median run was used for the calibrated and uncalibrated VISSIM histograms. As indicated in Figure 5, the calibrated model compares much better with the field data than the uncalibrated model. The uncalibrated model indicates shorter travel times than were observed in the field.

CONCLUSIONS AND RECOMMENDATIONS

This paper proposed a procedure for microscopic simulation model calibration and validation and demonstrated the procedure through a case study. The proposed procedure appears to be effective in the calibration and validation for VISSIM for signalized intersections.

Two important issues were encountered during implementation of the calibration and validation procedures. The first issue dealt with statistical testing when claiming the calibrated model was equal to the field data. The second issue was the importance of visualization in the calibration process.

The first issue is due to the variability in simulation runs. It was found that all multiple runs were not statistically equal to the field distribution. In other words, the simulation output has passed the statistical test at different percentiles for each parameter set. Given that the field data are just one realization of an infinite stochastic process, this appears natural. Thus, individual runs may not be applicable for statistical testing. Instead, the simulation results that are closest to the field data can be used. If the data pass this test, it ensures that the field data were represented at least once in the simulation model. In addition, the percentile of field average value at the distribution of simulation output can be used to determine how simulation represents the field condition.

The importance of visualization cannot be overstated. While obtaining measures of performance from the simulation close to those observed in the field is important, if the animations are not realistic the model cannot be claimed to be calibrated. The purpose of microscopic simulation models is to represent the real world as closely as possible. Simulation models with behavior not exhibited in the field are unrealistic. The parameter sets producing unrealistic simulations should not be considered. Visualization is a powerful tool.

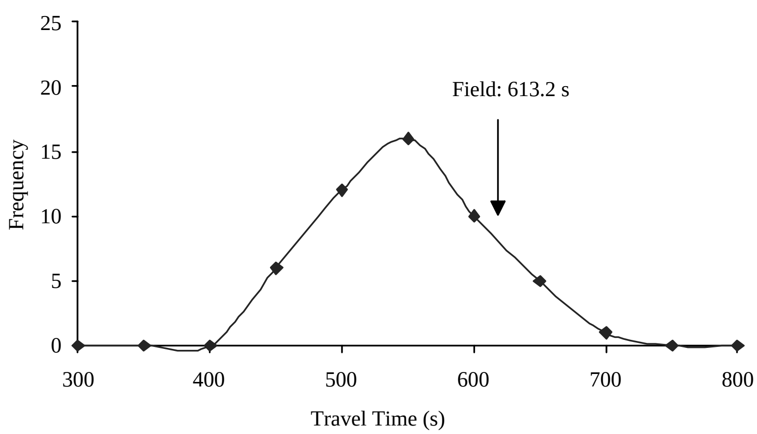

FIGURE 3 Eastbound left-lane travel time distribution of Parameter Set 7.

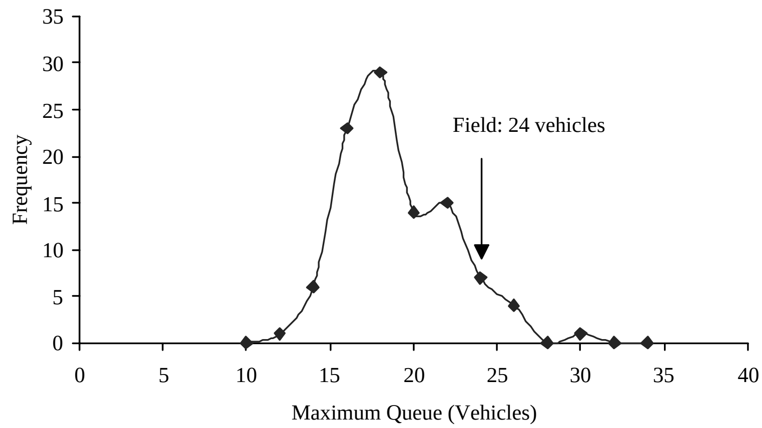

FIGURE 4 Maximum queue length histogram of Parameter Set 7.

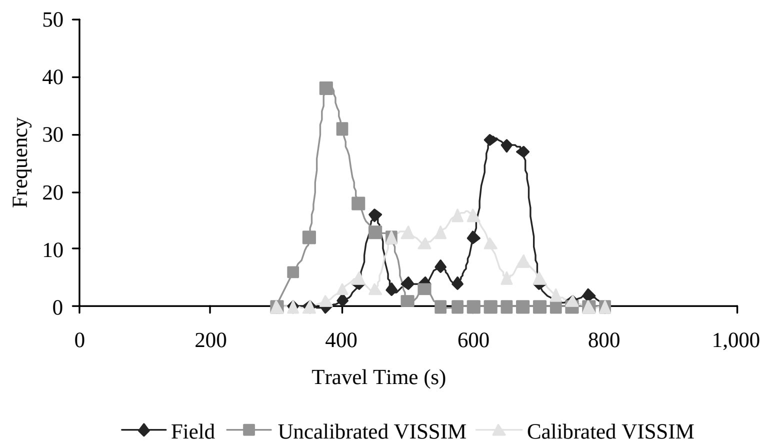

FIGURE 5 Comparison of field data, uncalibrated VISSIM, and calibrated VISSIM.

The study used only a single day of data collection and two measures of performance. Collecting multiple days of field data is recommended, if possible, to consider variability of field data. Further research should be conducted to determine whether the process is applicable to all networks or simply to this specific one. A range of flows and geometries would give more general conclusions. Using other performance measures such as number of stops, delays, fuel consumptions, or emissions is also recommended to see if they produce different variability.

ACKNOWLEDGMENTS

This research was made possible through funding from the Virginia Transportation Research Council. The authors thank staff members at Northern Virginia Smart Traffic Signal Systems for their support on this project.

REFERENCES

1. Hellinga, B. R. *Requirements for the Calibration of Traffic Simulation Models.* Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada. www.civil.uwaterloo.ca/b.PDF. Accessed March 2002.
2. Sacks, J., N. Rouphail, B. Park, and P. Thakuriah. Statistically Based Validation of Computer Simulation Models in Traffic Operations and Management. *Journal of Transportation and Statistics,* Dec. 2001.
3. Milam, R. T. *Recommended Guidelines for the Calibration and Validation of Traffic Simulation Models.* Fehr & Peers Associates, Inc., Roseville, Calif. www.fehrandpeers.com/publications/traff_simulation_m.html. Accessed March 2002.
4. *CORSIM User's Manual.* FHWA, McLean, Va., 1997.
5. *VISSIM Version 3.6 Manual.* Innovative Transportation Concepts. PTV Planung Transport Verkehr AG, Karlsruhe, Germany, Dec. 2001.
6. Ott, R. L. *An Introduction to Statistical Methods and Data Analysis,* 4th ed. Wadsworth, Inc., Belmont, Calif., 1993.

Publication of this paper sponsored by Committee on Traffic Signal Systems.