A PRACTICAL PROCEDURE FOR CALIBRATING MICROSCOPIC TRAFFIC SIMULATION MODELS

By

John Hourdakis[1]

Panos G. Michalopoulos[2]

Jiji Kottommannil[3]

[1] Department of Civil Engineering
University of Minnesota
500 Pillsbury Dr. SE
Minneapolis, MN 55455
Phone: (612) 625-8832,
Fax: (612) 626-7750
Email: hourd001@tc.umn.edu

[2] .Department of Civil Engineering
University of Minnesota
500 Pillsbury Dr. SE
Minneapolis, MN 55455
Phone: (612) 625-1509,
Fax: (612) 626-7750
Email: micha001@tc.umn.edu

[3] Department of Civil Engineering
University of Minnesota
500 Pillsbury Dr. SE
Minneapolis, MN 55455
Phone: (612) 626-1648,
Fax: (612) 626-7750
Email: kott0048@umn.edu

Submitted for presentation and publication
Transportation Research Board 2003 Annual Meeting
January 2003
Washington, D.C.

WORDS: 7490

July 2002

ABSTRACT

As employment of simulation is becoming wide spread in traffic engineering practice, questions about the accuracy and reliability of its results need to be addressed convincingly. A major criticism related to this is proper calibration of the simulation parameters as well as validation which is often not performed, or dealt with in an ad-hoc fashion. This paper presents a complete, systematic and general calibration methodology for obtaining accuracy needed in high performance situations. A technique for automating a significant part of the calibration process through an optimization process is also presented. The methodology is general and is implemented on a selected simulator to demonstrate its applicability. The results of the implementation in two freeway sections of reasonable size and complexity in which detailed data were collected and compared to simulated results, demonstrate the effectiveness of the manual calibration methodology. For instance, through calibration the average volume correlation coefficient on 21 detecting stations improved from 0.78 to 0.96. Comparable results were obtained with the automated calibration procedure with significant time savings and reduced effort.

INTRODUCTION

Traffic simulation is increasingly being used in practice as the sophistication and requirements prior to deployment of ATMS systems increases along with the complexity of problems engineers are faced with in daily practice. The effectiveness of a traffic simulator in evaluating traffic management strategies lies in its ability to accurately replicate actual traffic conditions; this requires proper calibration of its parameters rather than using default values. Calibration is the process in which the model parameters of the simulator are optimized to the extent possible for obtaining a close match between the simulated and the actual traffic measurements, which primarily include volume, speed and occupancy. Generally, calibration is an iterative process in which the engineer adjusts the simulation model parameters until the results produced by the simulator match field measurements; the comparison part is often referred to as validation.

There are two main issues related to calibration. First, a systematic procedure for the calibration process is lacking. Typically, a high performance simulator has numerous parameters that must be calibrated to obtain accurate results. In the absence of earlier calibrations at a site, the best-suited values of the simulator parameters are currently determined iteratively by trial-and-error and in an ad hoc fashion; this makes calibration a time-consuming and inefficient process, and as a result it is usually not performed or treated only superficially in most practical applications. The second issue related to calibration is that the goodness-of-fit tests usually employed to assess the effectiveness of calibration do not provide sufficient information for assisting the user to identify weaknesses during the course of the calibration. Existing tests measure only the magnitude of the percentage error and assess trends i.e. mean square error or regression coefficient. For a more rigorous accuracy assessment of the simulator there is a need

to use appropriate test statistics that can measure linear bias, as well as systematic and unsystematic error therefore providing the user with more information about the nature of the error.

This paper presents a complete and systematic general calibration methodology that addresses these issues and was implemented and tested at several sites for assessing freeway ramp metering performance. Its implementation was greatly simplified by an optimization technique also presented here along with results from a 20-Km freeway section in Minneapolis, Minnesota. The latter is related to a recently completed ramp metering evaluation study [1] in which very accurate results were obtained following the calibration methodology presented. The optimization technique produced comparable results faster, i.e. while the manual calibration required about 2 months for implementation of the first stage, the similar stage through automated calibration required only 6 hours.

BACKGROUND

The general requirements of a simulation calibration procedure have been discussed in only a few publications [2,3,4,5,6] along with Goodness-of-fit tests for validating traffic simulators. However, rigorous traffic simulator calibration methodologies are still lacking. Most of the published methodologies are not general but rather applicable only to a particular simulator; in addition their statistical analysis and verification of goodness-of-fit is not sufficiently detailed as mentioned earlier. This section provides a review of the most widely known calibration procedures.

INTRAS [7], the microscopic traffic simulator developed by FHWA, was employed to simulate traffic operation on Southern California freeways for evaluating incident detection

algorithms and for training artificial neural network models to detect freeway incidents [8]. The calibration procedure adopted in that study was performed in two stages. First, through trial-and-error, the parameters that influence vehicle movement during incident-free conditions were calibrated with incident-free data. The parameters were calibrated sequentially, i.e. while the best value of any particular parameter was being calibrated, the remaining parameters were treated as constants. When a parameter was being calibrated, the objective was to increase the slope and r^2 (regression coefficient) of the simulated vs. actual station volume and occupancy plots with greater emphasis on volume. After a suitable combination of the non-incident parameters was found, through trial-and-error, parameters related to incidents were calibrated against incident data sets. The RMS percent error between the simulated and actual occupancy during and after the incident was used as a performance measure during this stage. Even though, according to the authors, this calibration methodology produced satisfactory results, its shortcoming is that it failed to seek the optimum combination of parameters in a systematic way.

MITSIM [9], the microscopic simulator developed by the Massachusetts Institute of Technology, was recently employed to evaluate traffic management schemes involving coordinated traffic control systems, bus priority at signals and bus-lane operations in Sweden [10]. The calibration process was performed in 2 stages; in the first stage the driver behavior parameters were calibrated while in the second the travel behavior parameters were calibrated. The objective during driving behavior parameter calibration was to minimize the sum of squares of errors between the simulated and actual sensor speeds. Calibration of the travel behavior parameters involved the calibration of the route choice model parameters followed by OD estimation. The objective during route choice model calibration was to match the split of trips between two sensors through either one of which all trips pass. The objective during OD

estimation was to minimize the deviation between the estimated and the observed sensor counts and also minimize the deviation between the estimated OD and the seed matrix. The effectiveness of the calibration was evaluated by comparing 3 types of observed and simulated measurements: traffic flows, travel times and queue lengths. The two goodness-of-fit measures used for this purpose were the root mean square percent error and the mean percent error. The calibration methodology adopted in this study is quite complicated and laborious. In addition, all three measurements used are general and do not assist in calibrating local model parameters. Calibration of the driving behavior parameters and the OD estimation process can be automated by the incorporation of appropriate optimization techniques.

PARAMICS [11], is a microscopic simulator developed by Quadstone Limited in Edinburgh and was recently employed to evaluate freeway improvement strategies (ramp metering strategy, auxiliary lane addition, HOV lane addition) in the San Francisco Bay Area [12]. According to the calibration methodology followed in this study, several parameters such as link speed, vehicle top speed, simulation time step and signposting distances were calibrated based on engineering judgment or experience. In order to calibrate the mean headway and mean reaction time, simulations were performed with multiple combinations of these parameters using the average network speed and maximum vehicle throughput as performance indicators. Appropriate target values of these performance measures were determined for the particular test site selected. As a final measure of the effectiveness of calibration, a chi-square test was performed to compare the simulated vs. actual speed contour graphs. Similarly to the previous study, the measurements used are too general to allow detailed calibration of local parameters. As discussed later such a high level calibration might prove misleading.

FRESIM [13], another simulator developed by the FHWA, was employed to simulate expressway traffic operation in Singapore [14]. An automated calibration procedure was followed in this case. A Genetic Algorithm [15] optimization technique was used to search for the best combination of 12 simulator parameters. The objective function used in the optimization algorithm was a combination of the average absolute error (AAE) between the simulated and the actual 30 sec volume and speed averaged across all lanes. The AAE between the simulated and actual mainline average volume and speeds were used to measure the effectiveness of calibration. The calibration methodology adopted in this study, involving an optimization technique, is efficient as it is searching systematically for the best combination of simulator parameter values and as such it represents a significant improvement over conventional calibration methodologies. However, the calibration process involved only global parameter optimization i.e. no attempt was made to calibrate local parameters.

METHODOLOGY FOR PRACTICAL CALIBRATION AND VALIDATION

As mentioned earlier, the reliability of any simulator depends on its ability to produce results close to reality. The process of determining whether the simulation model is close enough to the situation being simulated is generally achieved through an iterative trial-and-error process involving calibration of the model parameters, comparing the model to the actual system behavior and using the discrepancies between the two to improve the results until the accuracy is judged to be acceptable. The behavior of the actual system is usually defined in terms of measurable traffic variables such as volumes, speeds, occupancies, queue lengths, etc., which for practical purposes are measured by detectors or observed manually. To validate the simulation

model, the simulator should be able to emulate actual measurements and produce a series of matching simulated values.

Goodness-of-fit measures used for validation

Regardless of the exact calibration procedure employed its success and efficiency depends on the measurements used during the validation as well as the goodness-of-fit measures employed. The measurements used to compare reality with simulation can not be easily defined because they depend on the given site to be modeled and the available instrumentation. In freeways the most common measurements are volume, speed or occupancy, and rather infrequently density which can be derived from occupancy. In some cases where entrance ramps are metered, an important validation measure is the queue size. The methodology described in this paper deals primarily with freeway sections where volume and speed are the primary validation parameters as in most cases in practice. However, further refinements are also presented for cases where demanding objectives, such as ramp metering, need to be evaluated.

In order for the calibration methodology to be efficient and robust the goodness-of-fit tests used should not just provide a metric describing the fit but they should include information as to what is the nature of the discrepancy between reality and simulation so the user can target specific model parameters for calibration. A typical statistical procedure for comparing two sets of data for a close match is through a hypothesis test such as the t-test. The null hypothesis in this context could be that the mean of the simulated traffic measurements is equal to that of the actual traffic measurements. However, there is a limitation of applying the t-test to traffic measurements. To apply this test, the observations should be identically and independently distributed (i.i.d.) but simulated and actual traffic measurements are time series that are not

necessarily i.i.d. Therefore the validation of a simulator cannot be based on such a hypothesis test.

A widely used error measure that can provide a fairly good initial estimate of the degree of fit between the simulated and the actual traffic measurements is the Root Mean Squared Percent Error (RMSP), defined in Eq. 1. This error measure gives an estimate of the total percentage error and is defined as:

$$\text{RMSP} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - y_i}{y_i}\right)^2} \qquad (1)$$

where RMSP is the root mean squared percentage error

x_i is the simulated traffic measurement value at time i

y_i is the actual traffic measurement value at time i

The correlation coefficient (***r***) is another popular goodness-of-fit measure used to measure the strength of the linear association between the simulated and the actual traffic measurements and is defined as:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \qquad (2)$$

where r is the correlation coefficient

$\bar{x}$ is the mean of the simulated traffic measurement values

$\bar{y}$ is the mean of the actual traffic measurement values

σ_x is the standard deviation of the simulated traffic measurement values

σ_y is the standard deviation of the actual traffic measurement values

n is the number of traffic measurement observations

The RMSP has an inherent deficiency in considering the disproportional weight of large errors while ***r*** although being a good measure does not provide any additional information to the modeler as to the nature of the error (difference) between real measurements and simulation. Theil, in his work on economic forecasting [16] developed a goodness-of-fit measure called "Theil's Inequality Coefficient", this is more sensitive and accurate than the RMSP or ***r*** and it can also be decomposed into three other metrics that provide specific information about the nature of the error. Theil's Inequality Coefficient [16] is defined as:

$$\mathrm{U} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}} \qquad (3)$$

The square of the numerator in Eq. 3 can be decomposed into the three components of the equation:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 = (\bar{y} - \bar{x})^2 + (\sigma_x - \sigma_y)^2 + 2(1-r)\sigma_y\sigma_x$$

Based on this and Eq. 3, three components of U can be derived, namely Um, Us and Uc, which can be used to measure different aspects of the error between the simulated and the actual traffic measurements. These components are defined as:

$$U_m = \frac{n(\bar{y} - \bar{x})^2}{\sum_{i=1}^{n}(y_i - x_i)^2}$$

$$U_s = \frac{n(\sigma_y - \sigma_x)^2}{\sum_{i=1}^{n}(y_i - x_i)^2}$$

$$U_c = \frac{2(1-r)n\sigma_y\sigma_x}{\sum_{i=1}^{n}(y_i - x_i)^2}$$

where

U_m is the bias proportion, which is a measure of systematic error that can be used to determine consistent over-counting or undercounting caused by an excess/loss of vehicles.

U_s is the variance proportion, which can be used to measure the simulated measurements' ability to replicate the degree of variability (fluctuations) in the actual measurements.

U_c is the covariance proportion, which is a measure of unsystematic error.

r is the correlation coefficient of the simulated and actual data

The other variables are as defined earlier.

Methodology for Practical Calibration and Validation

The calibration methodology presented in this paper was primarily developed for freeway simulation. Since the most common freeway measurements are volumes, occupancies and speeds, the methodology is illustrated by using 5-minute measurements collected from detector stations[1] but can be easily applied to any set of measurements or time slices. Calibration is enabled by using mainline station simulated and actual measurements and attempting to obtain the best match between the two by adjusting the simulator parameters through trial-and-error in

[1] Each detector station agregates counts from all its lane detectors and reports the total volume and average occupancy.

the manual process or through optimization. The simulator parameters to be calibrated for this objective fall into two main categories: global (those that affect the performance of the entire model) and local (those that affect only specific sections of the roadway). Examples of global parameters are the vehicle characteristics (Length, Width, Desired speed, Max Acceleration/Deceleration, and minimum headway). Speed limits of sections of the freeway model are local parameters. During the calibration process, the global parameters are calibrated first followed by local parameter calibration.

The calibration process is performed in two main stages based on volume and speed, followed by an optional stage in which the control variable depends on the specific purpose for which the simulation is performed. For example, if the objective of the simulation is to test the effectiveness of an adaptive ramp-metering algorithm, ramp queues could be used as the appropriate validation variable in the third stage. Similarly, if the objective is to simulate accidents, the congestion backup can be used as an appropriate variable in a similar way as presented here. Volume-based calibration is performed first as it is less complicated. Speed is a more sensitive measure to the fluctuations of traffic and progresses the calibration further. The optional 3rd stage is used to fine-tune the simulation model for the specific purpose of the simulation. The step-by-step procedure to be followed in each of the 3 stages of the calibration process is described below.

Stage 1: Volume-based Calibration.

The objective during this stage is to obtain simulated mainline station volumes as close as possible to the actual mainline station volumes which are used for ground truth as they are not input to the simulator and are routinely measured in freeway surveillance systems. The global simulator parameters to be modified in this stage are those related to vehicle characteristics like

speed, acceleration, deceleration rates and other parameters related to interaction between vehicles. Prior to beginning the calibration an initial run should be performed using default or best estimates of the simulator parameters and the results checked for general reasonableness and resemblance to mainline detector station data. In some instances [1] discrepancies are not caused by the model parameters but rather erroneous demand patterns either in the data collection stage or in the data entry. Once this possibility is ruled out, the user should check if the demand patterns randomly generated by the simulator are close to the specified ones. In the case of freeways, demand patterns include entrance and exit volumes or percentage of mainline volumes exiting. Small deviations from their specified values can easily be accounted for by replicating the simulation runs a number of times; usually 10 replications were found to be sufficient by the authors. Following this preliminary checking, the aforementioned model parameters are adjusted through trial-and-error over several iterations. For manual calibration the systematic procedure presented next is suggested as a guideline.

Starting from the first upstream station the goodness-of-fit statistics for that station are calculated and the vehicle model parameters starting with the desired speed are sequentially adjusted. It is recommended not to proceed with subsequent stations unless good fit is reached with the one in hand. The first couple of upstream stations in a freeway section usually have little interference from input traffic patterns (just one or two exit/entrances) therefore the response to changes in the model variables should be strong. The calibration exercise proceeds from upstream to downstream until the end of the freeway section is reached. If acceptable accuracy has not been reached in all the stations the cycle starts again from upstream. The goal is to gradually change and fine-tune the simulation parameters until all stations have acceptable fit. It has been observed that the change in the global parameters diminishes as the process moves

downstream. For example, for a 10-mile section with about 15 detector stations, the global parameters change in magnitude reduces beyond the first 4 stations. Therefore, after the simulated volumes of the first few stations attain a close match with the actual ones, the global parameters may have little effect and the engineer should focus on local parameters like speed limits and lane changing parameters. In the rest of this section, guidelines on how to interpret and use the information given by the statistics are presented.

While calibrating the global parameters, the focus should be to attain satisfactory values for the three statistics, namely RMSP, ***r*** and U. At the first stage generally the possibility of a high accuracy score is small but one can aim for numbers like RMSP below 15%, r above 0.8, and U lower than 0.3. An unsatisfactory value of RMSP or r can be attributed to inappropriate values of global simulator parameters like vehicle speeds, acceleration and deceleration rate, which require calibration.

The more sensitive statistics namely Us, Uc, and Um provide hints as to the nature of the discrepancy between the simulated and actual station volumes. An unsatisfactory value of Um along with satisfactory values of Uc and Um (Fig.1a) indicates a consistent loss/excess of vehicles that could be a result of erroneous demand data at the previous entrance or due to an error in the number of vehicles exiting before that station. The latter can be caused by improper weaving that might prevent vehicles from reaching the proper lane on time for the exit or due to wrong turning percentages. A valuable observation in a number of simulations, where the demand is entered as input and exit flows updated over short periods (5 to 15 minutes) is that the exit volumes might not be the ones expected simply because the correct amount of vehicles does not reach the exit at the proper time during the simulation. In such a case, although the correct percentage of vehicles exits the actual resulting volume is wrong. This problem is not observed

in simulations where the demand is described through time dependent origin-destination information.

An unsatisfactory value of Uc (less than 0.9) (which is often accompanied by an unsatisfactory value of Um (more than 0.1)) at a particular station (Fig.1b) with the value of Uc at the station downstream being satisfactory indicates the existence of a bottleneck between the two stations either in reality or generated by the model. Through calibration of the local parameters, this bottleneck should be either generated or suppressed accordingly. If unsatisfactory values of Uc and Um are observed (Fig.1c), it is indicative either of an error in the vehicle behavior, which can be attributed to the acceleration / deceleration rates or due to incorrect exiting volumes at the previous exit, for the reasons mentioned earlier. An unsatisfactory value of Us (more than 0.1), often accompanied by unsatisfactory values of Um and Uc, (Fig.1d) reflects large variability in either the simulated or the actual volumes which may be caused by vehicles driving close to each other; in such instances the acceleration/deceleration rates need to be appropriately adjusted.

Stage 2: Speed-based calibration

The objective during this stage is to obtain simulated mainline speeds as close as possible to the actual mainline speeds, and to match the actual breakdown conditions of known bottleneck locations. The simulator parameters to be calibrated during this stage comprise mostly of local speed limits though global parameters related to vehicle characteristics such as desired speed, acceleration and deceleration rates, might need some further fine tuning.

To compare the simulated and the actual mainline speeds, contour/3-D graphs can be used wherein bottleneck locations can be visualized easily. If paired loop detectors are not

available to measure actual mainline speeds, the speeds can be derived from the volume and occupancy measurements from single loop detectors using the following equation:

$$S = (0.11*Vol*(Lv+Ld)) /(Occ)$$

where S is the actual speed in kmph

Vol is the volume over the detector in 5 minutes

Lv is the average vehicle length in m

Ld is the detector effective length in m

Occ is the detector occupancy in %.

The process of adjusting the speed limits, as with the volume, should be performed starting upstream and proceeding downstream. In order to suppress a false bottleneck, often generated in an uncalibrated simulation, the speed limits should be increased in the region after the bottleneck location so that vehicles leave the location faster. Conversely, in order to generate a missed bottleneck, the speed limits before the region should be lowered; an increase in the grades will also produce a reduction in available gaps therefore vehicles will have to create gaps and subsequently cause congestion. Modification of the acceleration and deceleration rates also tends to affect the speed of vehicles to a certain extent. Calibration of the appropriate vehicle speed (global parameter) along with the local speed limits produce variability in speeds that are closer to those observed in reality.

Through extended observations of freeway traffic, we found two major categories of bottlenecks. The first category describes the bottlenecks generated through weaving, be that from a nearby entrance/exit or simply due to a lane drop. The second category describes bottlenecks attributed to driver behavior at complex geometries. For example, a well known bottleneck location in Minneapolis, MN is on freeway I-94 eastbound before the exit to I-35W north. This

four lane location has substantial distance between successive entrances/exits to allow for weaving to be completed well before the observed bottleneck location. Regardless of this, a sharp speed drop is observed on all lanes. When this location was visually observed we noticed that it contains two turns in close proximity, the second of which is located under an overpass and all of this during a considerable downhill grade. The lack of good visibility in conjunction with the constrictive environment of the underpass prompts the drivers to sharply reduce their speed hence creating the ground for the generation of a bottleneck.

Although bottlenecks due to weaving are relatively straight forward to model since most simulators have enough parameters available to control the lane changing, gap generation and acceptance behavior, the second category of bottlenecks, attributed solely to driver behavior, needs a lot of observations and familiarity with the site, in order to be modeled accurately.

Stage 3 (Optional): Objective-based calibration

While the first two stages are well defined, this third stage depends on the objective of the simulation. The need for this stage was revealed during a real project. Specifically, in this project the objective was to evaluate an adaptive ramp metering algorithm. Although the calibration results from the first two stages resulted in high accuracy without control, when ramp control was implemented the accuracy dropped to unacceptable levels on the mainline and especially in reproducing the ramp queue sizes. The reason was that small initial discrepancies in the model gradually amplified due to the adaptive nature of the control algorithm. The algorithm gradually forced ramp rates a lot different than the ones observed in reality. Therefore, a third calibration stage was devised and implemented. In this stage the Queue sizes were the validation variable and very specific local variables like speed limits and mainline section grades where further adjusted.

Even though the details of the third stage calibration may vary based on the objective of simulation, similar procedures can be followed as pointed out at the beginning of this section. One should be careful not to be misled by the good results of the first two stages but validate the simulation with at least one additional measurement.

IMPLEMENTATION

The calibration methodology presented was used in a real life project for a number of freeway sections in Minneapolis, Minnesota. The objective of the simulations was to test the effectiveness of adaptive ramp metering [1, 17] following a period of public controversy. This section describes one of the test sites followed by results of the implementation; the simulator employed in this case was a well respected microscopic one called AIMSUN [18]

Test site and data

The test site is a 20 km (12 miles) long section of TH 169 northbound starting from the interchange with I-494 and ending at I-94. The site is a circumferential freeway of average geometric complexity and carries moderate traffic volumes i.e. approximately 50,000 vehicles daily. The site consists of mainly two lanes with 10 weaving sections, 24 entrance ramps and 25 exit ramps.

The detector data used for calibration comprised of 5-minute volumes and occupancies for March 23rd from 14:00 hrs to 20:00 hrs. 5-minute volume and occupancy from 14:00 hrs to 20:00 hrs for March 21st, 2000 and March 22nd, 2000 was used for validating the calibrated simulation model. It is important to note that in order to replicate real-time ramp control all

demand patterns and boundary conditions had to be collected simultaneously for each day of simulation i.e. average values would not realistically emulate the ramp metering strategy employed.

Results

Stage 1:

The volume-based calibration process required about 300 simulation iterations in which the parameters were successively changed as described earlier. The calibration process proved to be very effective as indicated by the test statistics shown in Table 1. As can be observed from the table, there is significant improvement in the values of the test statistics compared to those obtained using the best initial estimates of the simulation parameter values. The simulator parameter values prior to and after each stage are shown in Table 2. As can be observed from the table, the simulator parameters that had to be modified during the first stage were the parameters related to vehicle characteristics and the local speed limits.

During this stage, a number of irregularities in the input data were observed. Specifically, in two locations the placement of the entrance ramp loop detector was not the one reported in the plans. Because of the sensor misplacement, the simulation results deviated substantially from the actual measurements in spite of careful screening of the data entry process prompting an investigation. After some analysis and visits to the field the true location of the detectors and the nature of the problem was revealed. Mn/DOT was not aware of these discrepancies until that time.

Stage 2:

At the beginning of this stage, the actual mainline speeds were compared with the simulated mainline ones through contour graphs as shown in figures 2(a) and 2(b) respectively. As can be seen from the contours, there was significant discrepancy between the speeds in spite the stage 1 calibration; moreover the bottleneck locations did not match the observed ones. Hence, the second stage of the calibration process was performed which required approximately 100 additional simulation iterations during which the local speed limits, mainline section grades, and lane changing parameters were adjusted through trial-and-error. Following this, the matching of the bottleneck locations improved considerably as can be seen from the speed contour graphs in Figure 2(c). The simulator parameter values obtained at the end of this stage are shown in Table 2. As can be observed from the table, the parameters related to vehicle characteristics did not have to be altered much, but the local speed limits had to be modified substantially.

Stage 3:

After stage 2, when the simulation was performed with ramp metering, it was observed that the simulated and the actual merge detector counts, at almost all the entrance ramps did not match during the ramp-metering period. As a result, the simulated and the real ramp queues did not match. An example of this on a single ramp is depicted in Figure 3. This problem was considered major and prompted development and implementation of the 3rd stage in the proposed calibration methodology. This stage required approximately 100 additional simulation iterations. At the end of this stage, the simulated ramp queues on all the entrance ramps had a close match with the actual ones as shown for the TH-55WB example ramp of Figure 3. The simulator parameter values did not generally change in this stage, as can be seen in Table 2.

What changed were the speed limits and grades on specific sections (less than 10% of the total site).

After the 3 stages of the calibration process the simulation model was validated based on the remaining two days, i.e. March 21st and March 22nd, 2000. The results were very satisfactory as indicated by the values of the test statistics shown in Table 1. It can therefore be seen that the adopted calibration methodology is very effective.

Since the calibration process involved modification of the simulator parameters iteratively by trial-and-error, it was a very time-consuming procedure. The volume, speed and queue-based calibrations required a total of about 4 months to complete, while the volume alone required 2 months.

AUTOMATION OF THE CALIBRATION PROCESS

As the previous section suggests, the number of iterations, effort and time involved in a rigorous calibration can be substantial. Clearly, there is a need to automate the iterative process, of manually modifying the simulator parameters, to the extent possible. Typically this is achieved through optimization techniques which seek the best-suited values of the model parameters through efficient search procedures. Such an approach was followed and presented here. The optimization problem in the context of the problem at hand is to calibrate the simulator parameters so that an objective function is minimized. The sum of squared errors of the mainline station volumes is defined here as the objective function to be minimized, subject to bounds on the simulator parameters. Mathematically, the optimization problem can be stated as:

$$\text{Minimize} \quad \mathrm{F} = \sum_{j=1}^{st} \sum_{i=1}^{m} (v_{si}^{j} - v_{ai}^{j})^2$$

Subject to

$$L_{xp} < x_p < U_{xp}, p = 1, 2, \ldots, n$$

where F is the objective function to be minimized

v_{si}^{j} is the simulated traffic measurement of station j during time interval i

v_{ai}^{j} is the actual traffic measurement of station j during time interval i

L_{xp} is the lower limit of simulator parameter x_p

U_{xp} is the upper limit of simulator parameter x_p

n is the number of simulator parameters to be optimized

st is the number of detector stations on the freeway section

m is the number of time intervals.

It should be noted here that the objective function is not an explicit function of the simulator parameters. Therefore, the optimization problem cannot be solved by the usual approach of differentiating the objective function and setting it to zero to obtain a solution corresponding to the minimum. The alternative is to employ an appropriate optimization technique. The problem defined above is a non-linear unconstrained one; such problems can be solved using nonlinear programming techniques like the quasi-Newton methods, the method of steepest descent, Newton-Raphson method, Fletcher-Powell method, etc.

Several computer programs are available that incorporate such non-linear programming techniques. The well-proven and state-of-the-art optimization program 'MINOS' [19,20,21,22] was selected here because of its widespread use and ability to solve a variety of large-scale optimization problems. These include problems that are linear, non-linear, bounded, unbounded, constrained or unconstrained based on the form of the objective function. If the objective

function is non-linear and is unconstrained as in the problem at hand, the quasi-Newton algorithm is used in MINOS.

Implementation

A computer program was written that integrates MINOS as a subroutine with AIMSUN and facilitates the data transmission between these two modules. From the manual calibration methodology only stage one has been automated in order to demonstrate the technique; in addition, automation of the first stage alone significantly reduces the calibration effort as this stage is the most time consuming one. This section describes the results of the implementation of the automated calibration in the selected simulator.

Sensitivity analysis and preliminary experimentation

Prior to implementation of the automated calibration process a sensitivity analysis of the critical simulator parameters was conducted in order to determine the behavior of the objective function and the solution that can be expected from the program. Space limitations do not allow presentation of the details here; suffice it to say that the objective function resulting from changes to all the simulator parameters was non-smooth due to interaction effects of these parameters. The objective function which considers the interaction of all the parameters can therefore be expected to be highly non-smooth. Hence, the gradient-based optimization technique can be expected to provide a solution that lies in one of the local minima of the objective function.

In any optimization algorithm, specification of the appropriate step size is a required step. In MINOS, the step size is represented by the ‘difference interval’ denoted here as ‘h’. In order to

estimate the gradient of the objective function with respect to a variable x, the variable is perturbed by h(1+|x|). Several values for the difference interval were tried; the one that provided the best value of the objective function was 0.03 and was therefore used subsequently.

The effect of using different initial simulator parameter values was also analyzed. It was found that even though the final values of the simulator parameters obtained through the optimization depend on their initial estimates, the final objective function values were comparable. This further simplifies the calibration task as good initial simulator parameter values need not be specified in order to obtain a satisfactory solution. Moreover, this suggests the existence of multiple solutions, at least for the selected simulator, all of which might be equally acceptable.

Results

The implementation results of the optimization technique are shown in Tables 1 and 2. The automated calibration process required about 9 iterations of the simulator. In each iteration, approximately 100 different combinations of the simulator parameters were tried. It can be seen from the Table 1 that the values of all the goodness-of-fit measures obtained using the automated calibration process are very close to the manual process with significant savings in time and effort; for instance, not only the number of iterations was reduced substantially but also the time to obtain the desired results was reduced to 6 hours, compared to 2 months of the manual process for stage one of the calibration.

CONCLUSIONS

In this paper, a three-stage general and systematic methodology for manually calibrating microscopic traffic simulators was presented. Its implementation on a selected simulator proved

very effective. For example, an average correlation coefficient of 0.961 between the simulated and the actual mainline station volumes was obtained when all calibration stages were completed. This is an unusual high fit which can be explained by the detailed data collected as input and the quality of the simulator which resulted in high accuracy even prior to calibration (***r*** = 0.78). The Theil's goodness-of-fit statistics presented here were effective in identifying discrepancies between simulated and actual volumes that would not have been accounted for by commonly used tests. Furthermore, the correct bottleneck location identification was enabled in the second stage of the calibration. Finally, the third stage of the calibration process also proved very effective in obtaining a close match between the simulated and the actual entrance ramp queues.

The procedure for automating a significant part of the calibration process through optimization yielded comparable results as the manual calibration process with substantial savings in time. For instance, the automated volume-based calibration process required about 6 hours for 9 iterations (plus 2 months for manual stage 2 & 3 calibration) resulting in a final average correlation coefficient of 0.946 whereas the corresponding fully manual calibration required about 4 months for 300 iterations to obtain an average correlation coefficient of 0.961. This suggests that even though the gradient-based optimization procedure employed here does not ensure attainment of the global optimum, it is sufficient for practical purposes. It is worth mentioning that although optimum simulator parameter estimation depends on their initial values, the final objective function values obtained using rough initial parameter values through the unconstrained non-linear optimization proved to be satisfactory. This suggests that unlike the manual procedure, the automated calibration technique does not rely on very good initial parameter estimates, which further simplifies the calibration task. It also indicates the existence

of multiple solutions, all of which are equally acceptable at least for the simulator employed and the example at hand. Finally, the automated calibration procedure used here is general and allows employment of any optimization technique one wishes to use; such techniques include genetic algorithms, simulated annealing, Nelder-Mead, and others that could possibly result in better parameter estimates.

Before concluding it is worth mentioning that the proposed methodology is not restricted to freeways only but it can be used for arterial streets as well. This was recently demonstrated in another study [23] in which this calibration was implemented in a freeway corridor that included 5 major arterial streets and 250 intersections in addition to the freeway and its ramps.

ACKNOWLEDGMENTS

This research was supported jointly by Minnesota Department of Transportation and The Center for Transportation Studies. The authors wish to thank Mr. Frank Lilja, Mr. James Aswegan, and Mr. Richard Lau at the Mn/DOT Metro Division for their invaluable help in obtaining data that was critical for performing this research. The authors also wish to thank Dr. Jaime Barceló for his guidance and support in the development of the calibration methodology.

REFERENCES

1. Hourdakis, J., Michalopoulos, P.G. (2002), "Evaluation of ramp control effectiveness in two twin cities freeways", Paper presented at the 81st Transportation Research Board Meeting.
2. Kleijnen, J.P.C. (1995), "Verification and validation of simulation models; Theory and methodology", European Journal of Operational Research 82, 145-162.
3. Hellinga, B.R.(1998), "Requirements for the calibration of traffic simulation models", Proceedings of the Canadian Society for Civil Engineering, Vol IVB, 211-222.
4. Rao, L., Owen, L. (1998), "Development and application of a validation framework for traffic simulation models", Proceedings of the 1998 Winter Simulation Conference, IEEE. 1079-1086.
5. Rakha, H., Hellinga, B., Van Aerde, M., Perez, W. (1996), "Systematic verification, validation and calibration of traffic simulation models", Paper presented at the 1996 Transportation Research Board Annual Meeting.
6. Jayakrishnan, R., Oh, J., Sahraoui, A.(2001), "Calibration and path dynamic issues in microscopic simulation for advanced traffic management and information systems", Paper presented at the 80th Transportation Research Board Meeting.
7. Wicks, D.A. and Andrews, B.J. (1980) Development and testing of INTRAS, a microscopic freeway simulation model, Vol. 2: User's Manual, Report FHWA/RD-80/107.
8. Cheu, R.L., Recker W.W., Ritchie, S. G. (1994), "Calibration of INTRAS for Simulation of 30-sec loop detector output", Transportation Research Record 1457, 208-215.

9. Yang, Q. and Koutsopoulos, H.N. (1996) A microscopic traffic simulator for evaluation of dynamic traffic assignment systems, Transportation Research, 4C, pp. 113-129.
10. Ben-Akiva, M., Davol, A., Toledo, T., Koustopoulos, H.N. (2002), "Calibration and evaluation of MITSIMLab in Stockolm", Paper presented at the 81st Transportation Research Board Meeting, January 2002.
11. Smith, M., Druitt, S., Cameron, G. and MacArthur, D. (1994) Paramics Final Report, Technical Report EPCC-PARAMICS-FINAL, University of Edinburg.
12. Gardes, Y., May, A.D., Dahlgren, J., Skabardonis, A. (2002), "Freeway calibration and application of the Paramics model", Paper presented at the 81st Transportation Research Board Meeting.
13. FHWA (1995) The FRESIM User's Guide, Office of Traffic Safety and Operations R&D, Federal Highway Administration
14. Cheu, R.L. (1998), "Genetic Algorithm approach in FRESIM calibration", Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering, ASCE. 191-198.
15. Goldberg, D.E. (1989) Genetic algorithms in search optimization and machine learning, Addison-Wesley.
16. Theil, H. (1961) Economic forecasts and policy, North-Holland Publishing Company-Amsterdam, 1961.
17. "Employment of the Traffic Management Laboratory (TRAMLAB) for the evaluation of Ramp Metering in the Twin Cities". Project Report. Mn/DOT, 2002.
18. Transport Simulation Systems (2001) Aimsun version 4.0 – User manual, TSS. April 2001.

19. Murtagh, B. and Saunders, M. (1983) MINOS 5.5 User's Guide, Stanford University.
20. Murray, W. (1972), "Numerical methods for unconstrained optimization", Academic Press.
21. Wolfe, M.A.(1978), "Numerical methods for unconstrained optimization", Van Nostrand Reinhold Company.
22. Murtagh, B.A, Saunders, M.A.(1978), "Large Scale Linearly constrained optimization", Mathematical Programming, v14, 41-72.
23. Shekar, A. "Issues in freeway corridor simulation", Thesis. University of Minnesota. 2002

LIST OF TABLES

TABLE 1: Volume statistical measures prior to and after calibration

TABLE 2: Simulator parameter values prior to and after calibration

LIST OF FIGURES

FIGURE 1: Examples of unsatisfactory U_m, U_s and U_c

FIGURE 2: Average mainline speed contours

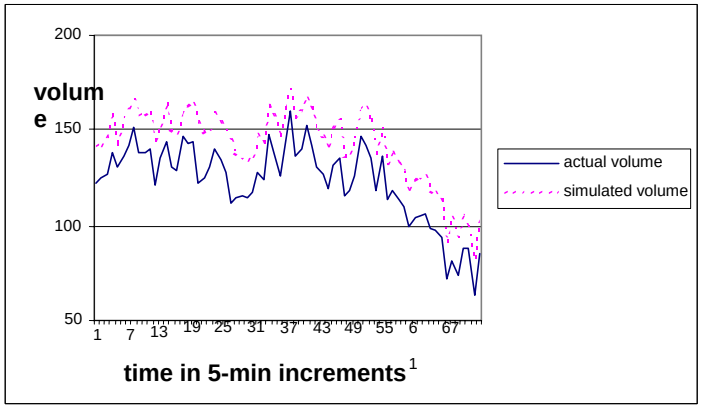
FIGURE 3: Simulated and actual queues at TH-55WB ramp

	Root Mean Square Error %	Correlation coefficient	Theil's Inequality Coefficient	Theil's Bias Proportion	Theil's Variance Proportion	Theil's Covariance Proportion
Initial	40	0.78	0.3	0.40	0.31	0.29
Mar 21st	10.62	0.98	0.00426	0. 30877	0.01052	0. 68070
Mar 22nd	6.42	0.97	0.00154	0.12352	0.05365	0.82281
Mar 23rd	7.39	0.96	0.00238	0.08826	0.03098	0.88075
Automated	8.84	0.95	0.004	0.078	0.011	0.91

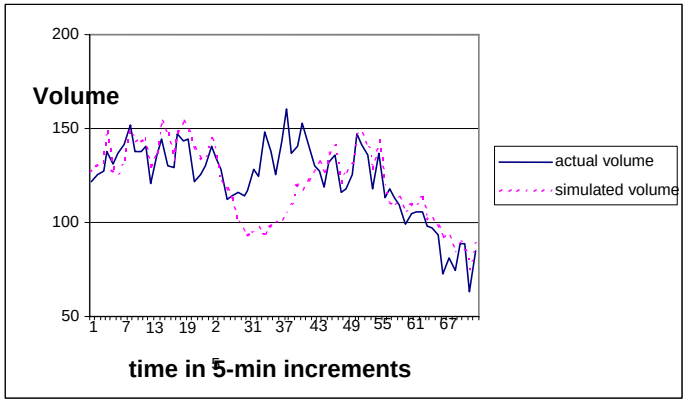
Table 1: Volume statistical measures prior to and after calibration.

Parameter	Initial	After stage 1 (manual)	After stage 2 (manual)	After stage 3 (manual)	After stage 1 (automated)
Max. desired speed (kmph)	100.000	110.000	110.000	110.000	104.249
Max. acc. rate (m/s2)	2.800	3.000	3.000	3.000	2.838
Normal dec. rate (m/s2)	4.000	4.000	4.000	4.000	3.983
Max. dec. rate (m/s2)	8.000	7.000	7.000	7.000	6.901
Reaction time (sec)	0.750	0.590	0.590	0.590	0.512
Percent overtake	0.950	0.950	0.940	0.940	0.950
Percent recover	1.000	1.000	0.990	0.990	1.000
Max. speed difference (kmph)	40	40	60	60	40
Max. speed difference on-ramp (kmph)	50	50	70	70	50
Av. section speed (regular section, kmph)	110	100	105	105	110
Av. section speed (weaving section, kmph)	90	75	70	72	90
Av. section speed (ramp section, kmph)	60	60	55	55	60

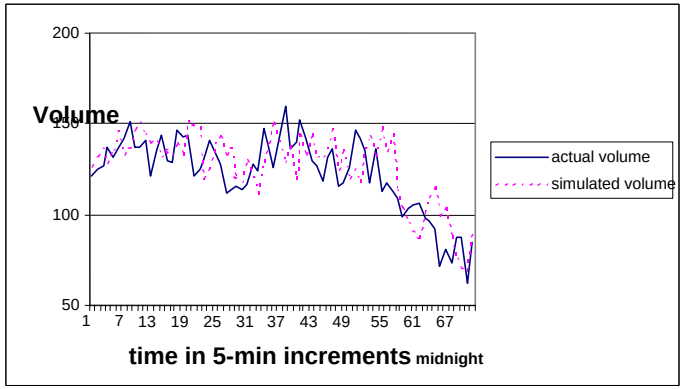
Table 2: Simulator parameter values prior to and after calibration



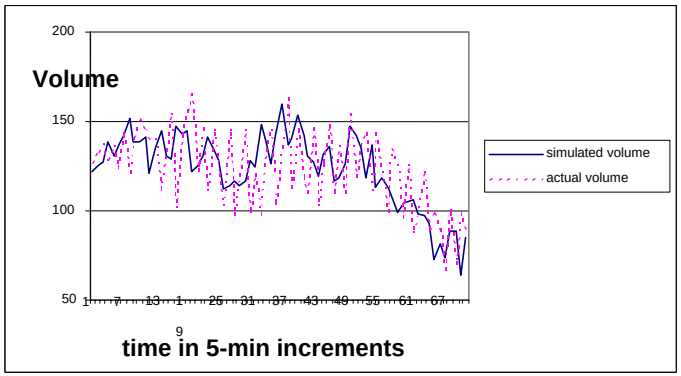
(a): Illustration of unsatisfactory Um



(b): Illustration of unsatisfactory Uc



(c): Illustration of unsatisfactory Uc and Um



(d): Illustration of unsatisfactory Us

Figure 1: Examples of Unsatisfactory Um, Us, and Uc.: Lane Volume vs. Time

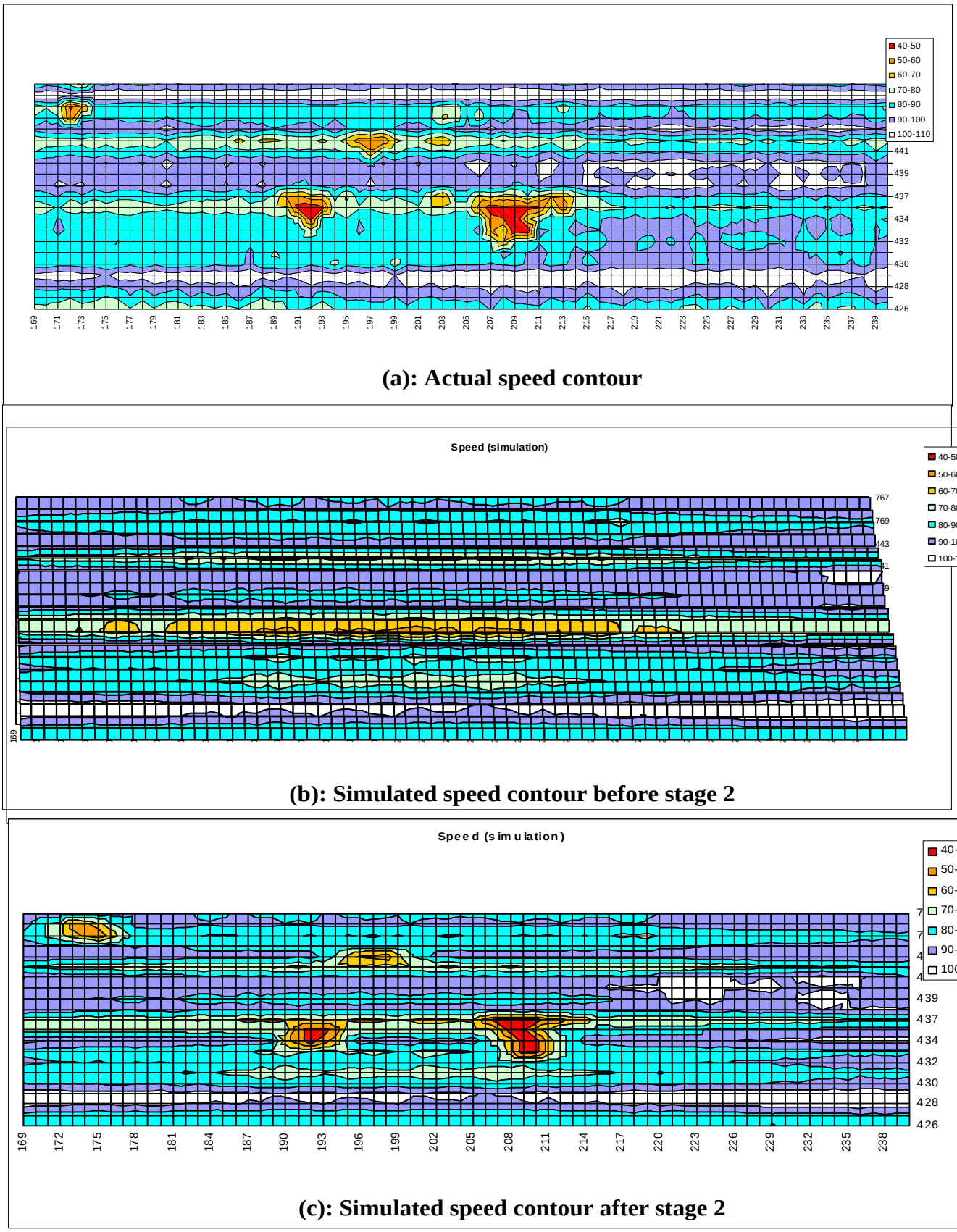


Figure 2: Average Mainline speed contours used in calibration.

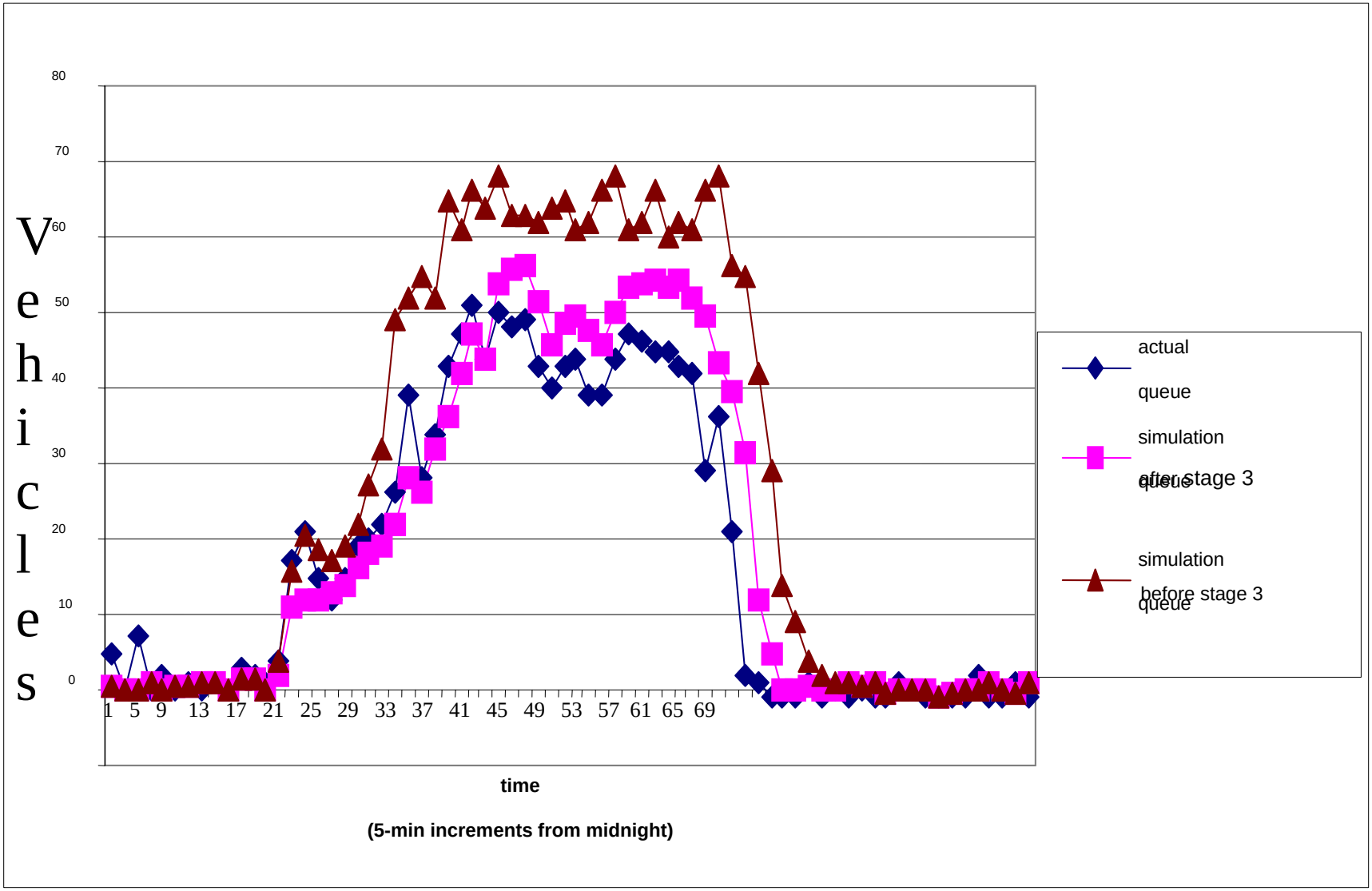


Figure 3: Simulated and actual queue at TH-55WB ramp.