FINAL CONTRACT REPORT

DEVELOPMENT AND EVALUATION OF A CALIBRATION AND VALIDATION PROCEDURE FOR MICROSCOPIC SIMULATION MODELS

B. Brian Park, Ph.D.
Faculty Research Scientist
and
Research Assistant Professor of Civil Engineering

Hongtu Qi
Graduate Research Assistant

Project Manager
Michael A. Perfater, Virginia Transportation Research Council

ABSTRACT

Microscopic traffic simulation models have been widely accepted and applied in transportation engineering and planning practice for the past decades because simulation is cost-effective, safe, and fast. To achieve high fidelity and credibility for a traffic simulation model, calibration and validation are of utmost importance. Most calibration efforts reported in the literature have focused on the informal practice with a specific simulation model, but seldom did they propose a systematic procedure or guideline for simulation model calibration and validation.

The purpose of this study was to develop and evaluate a procedure for microscopic simulation model calibration and validation. Three widely used microscopic traffic simulation models, VISSIM, PARAMICS, and CORSIM, were selected for model review and practice of model calibration and validation. The validity of the proposed procedure was evaluated and demonstrated via two case studies including an actuated signalized intersection and a 5-mile freeway segment with a lane-closure work zone.

The simulation results were compared against the field data to determine the performance of the calibrated models. The proposed procedure yielded acceptable results for all applications, thus confirming that it was effective for the different networks and simulation models used in the study. Although the calibrated parameters generated the performance measures that were representative of the field conditions, the simulation results of the default parameters were significantly different from the field data.

FINAL CONTRACT REPORT

DEVELOPMENT AND EVALUATION OF A CALIBRATION AND VALIDATION PROCEDURE FOR MICROSCOPIC SIMULATION MODELS

B. Brian Park, Ph.D.
Research Assistant Professor of Civil Engineering
and
Faculty Research Scientist

Hongtu Qi
Graduate Research Assistant

INTRODUCTION

Microscopic traffic simulation models have been widely accepted and applied in transportation system design, traffic operations and management alternatives evaluation for the past decades because simulation is cost effective, safe, and fast. As recognized by the Highway Capacity Manual (HCM), simulation becomes a valuable aid in assessing the system performance of traffic flows and networks. Furthermore, traffic simulation models sometimes provide significant advantages over traditional planning or analytical models such as Highway Capacity Software (HCS). For example, delay estimation from HCS is often inappropriate when the impacts of a large-scale system are considered or oversaturated conditions are prevalent (White, 2002). Other common reasons for the popularity of simulation models include their attractive animations, stochastic variability to capture real-world traffic conditions and capabilities to model complex roadway geometries such as combined systems of urban streets and freeways. The most popular existing microscopic traffic simulation models include CORSIM (Federal Highway Administration [FHWA], 1997), VISSIM (PTV Planug Transport Verkehr AG [PTV], 2001), PARAMICS (Quadstone Limited, 2002), AIMSUN (TSS-Transportation Simulation System [TSS], 2002), WATSIM (KLD Associates, Inc. [KLD], 2002), TRANSIMS (Smith, 1995), MITSIM (Yang, 1996), SimTraffic (Trafficware, 2003), and INTEGRATION (Van Aerde, 2002). Each model has strengths and weaknesses. All models have achieved some degree of success in particular types of applications, including traffic operation, transportation planning, and intelligent transportation systems (ITS) strategies.

As more and more traffic engineers and researchers begin relying on simulation as an evaluation tool, credibility will become an essential concern.. Potential issues are whether these models accurately represent the real-world system and/or whether one can trust the decisions based on the simulations. Microscopic simulation models contain numerous independent parameters to describe traffic flow characteristics, driver behavior, and traffic control operations. These models provide a default value for each parameter, but also allow users to change the values to represent local traffic conditions. However, simulation model-based analyses often are performed under default parameter values or manually adjusted values. Rigorous calibration

procedures often are omitted because doing so requires a great deal of time and effort, as well as a huge amount of field data. The findings and conclusions based on the uncalibrated or poorly calibrated models could be misleading, even erroneous. Thus, proper calibration and validation of simulation models are crucial steps in simulation applications.

To achieve adequate reliability of the simulation models, it is important that a rigorous calibration and validation procedure be applied before any further study and analysis are conducted. Changes to the parameters during calibration should be justified and defensible by the users. Most of the calibration efforts were to achieve reasonable correspondence between field data and simulation model output. Recently, more and more transportation researchers and practitioners have realized the importance of model calibration and validation and spent significant time and efforts to demonstrate the validity of their models. However, it was indicated that simulation model calibration and validation often were discussed and informally practiced among researchers, but seldom have been formally proposed as a procedure or guideline (Sacks et al., 2002). Therefore, proposing a general procedure for simulation model calibration and validation is an urgent task.

PURPOSE AND SCOPE

The purpose of this project was to develop a procedure to calibrate and validate microscopic simulation models and evaluate the procedure through case studies with different simulation models. Specific objectives were as follows:

1. Examine the existing literature with regard to microscopic simulation models and the simulation model calibration and validation studies.

2. Evaluate existing simulation models and provide model selection criteria.

3. Propose a procedure for simulation model calibration and validation.

4. Evaluate the proposed procedure via three simulation models using two case studies.

Three microscopic simulation models that are widely used in the traffic operations and management analyses were selected for model review and used in the case studies of model calibration and validation. They are CORSIM, VISSIM, and PARAMICS. The SimTraffic simulation model was not selected as it does not provide command line execution capability needed to achieve an automation of calibration and validation optimization process. The validity of the proposed procedure was evaluated and demonstrated via two case studies including an actuated signalized intersection and a 5-mile freeway segment with a lane-closure work zone. This study considered relatively simple networks such that extensive data collections and experiments on calibration and validation could be relatively easy to conduct. Large-scale complex networks and testing for generalization and transferability of the proposed procedure were not considered in this study.

METHODOLOGY

The research approach in this study involved (1) a literature review, (2) evaluation of simulation models, (3) a proposed procedure for microscopic simulation model calibration and validation, (4) development of test beds and (5) calibration and validation of the simulation models.

Literature Review

Literature was reviewed on the current practices in the microscopic simulation model calibration and validation including general studies and genetic algorithm (GA) based studies. The Virginia Transportation Research Council (VTRC) library and the University of Virginia libraries were used for this purpose.

Evaluation of Simulation Models

Three widely used stochastic and microscopic simulation models were reviewed and evaluated. The models considered were VISSIM, PARAMICS, and CORSIM. Each model has pros and cons for simulating various types of traffic studies. A detailed description of these models in terms of available features such as user interface, traffic generation and assignment, traffic control, multi-modal transportation, ITS capability, measures of effectiveness (MOEs), and model weakness are provided.

Proposed Procedure for Microscopic Simulation Model
Calibration and Validation

A systematic procedure was proposed to calibrate and validate microscopic simulation models. The proposed approach consisted of simulation model setup, initial evaluation, initial calibration, feasibility test, parameter calibration using GA, evaluation of parameter sets, model validation, and feedback.

Development of Test Beds

Two test sites selected to evaluate the proposed calibration and validation procedure include an actuated signalized intersection and a 5-mile freeway segment with a lane-closure work zone. These two networks have very different traffic conditions in terms of geometric characteristics, traffic controls and driver behavior. It is desirable to examine the performance of the proposed model calibration and validation procedure with different networks.

The data needed for calibration and validation purposes are simulation input data and performance measures. The general input data include geometric characteristics, traffic volume, percentage of passenger cars and trucks, traffic controls, posted speed limit, and so forth. For

particular networks, such as signalized intersections, additional data such as signal timing and detector layout are required. Performance measures selected are usually those critical to the study network and representative of local driver behavior. In order to consider day-to-day variability, multiple days of field data were collected from two test sites, separately, in this study.

Calibration and Validation of the Simulation Models

The three microscopic simulation models were calibrated and validated using the proposed procedure via two case studies, for a total of six applications. The main steps were implemented in each application and simulation results using different parameter sets were compared with field data to assess calibration performance.

RESULTS

Literature Review on Simulation Model Calibration and Validation

General Studies

Sacks et al. (2002) identified four key issues on model validation: (1) identifying explicit meaning of validation in particular context, (2) acquiring relevant data, (3) quantifying uncertainties, and (4) predicting performance measures under new conditions. They demonstrated an informal validation process using CORSIM simulation model and emphasized the importance of data quality and visualization. However, the authors have not established any formal procedure for simulation model calibration and validation.

Hellinga (2003) proposed general requirements for the calibration of traffic simulation models. The proposed calibration process consists of three main phases: study definition, initial calibration, and evaluation of model outputs. The first step involves the tasks and activities prior to the modeling, such as identification of study goal, required field data, desired simulation performance measures, and so forth. The second step is to make sure that network coding is as accurate as possible. The third step is to run the simulation and compare the field data with simulation output. This process provides basic guidelines but does not give a direct procedure for conducting calibration and validation.

A set of guidelines for the calibration and validation of traffic simulation models was recommended by Milam et al. (2003). The guidelines include conduction of field-data collection, calibration of models to match field conditions, validation of models under certain criterion, such as 95 to 105 percent of observed value, and estimation of the minimum number of simulations with desired confidence interval. In addition, simulation outputs from CORSIM such as traffic counts, travel time, and queue were illustrated in the validation procedure. These guidelines could be viewed as a critical point for the application of traffic simulation models.

Park and Schneeberger (2003) proposed a calibration procedure consisting of nine steps and demonstrated the proposed procedure through a case study. The main components include (1) preparation before formal calibration such as data collection and selection of parameters and MOEs; (2) calibration effort such as experimental design, simulation runs, surface function development and candidate parameter sets generations; and (3) evaluation and validation. The case study results showed that the proposed procedure was effective in the calibration and validation for VISSIM mainly for signalized intersections. The study only made use of a single day of data collection and generated the parameter sets based on a linear regression model, which did not account for the correlations among the parameters.

A description was given for the application of calibration and validation techniques to traffic simulation models by Cohen (2004). The author recommended adjusting the major parameters close to the field observations during the calibration process and comparing the model performance measures with field data during the validation process. The MOEs used for validation should be independent of the measurement of the calibration parameters. In particular, Cohen asserted that calibration must be performed either by measuring the calibration parameters directly or by measuring a surrogate. Finally, the paper provided guidance to the model users on how one should proceed if the validation of the models fails. The procedure only applies to a few major parameters, which requires the knowledge and judgment of the model users. In addition, some parameters, especially those related to driver behavior, are not easily observed in the field.

Dowling et al. (2004) proposed a practical, top-down approach that consisted of three steps. First, capacity at the key bottlenecks in the system was calibrated. Second, traffic flow at non-bottleneck locations was calibrated. Finally, the overall model performance was calibrated against field performance measures. The authors divided the model parameters into categories and started with the most important parameters, usually global parameters. Then further fine-tuning with link-specific parameters was conducted if necessary. However, the procedure also focuses on a few selected key parameters, which are not easy to identify. As this approach calibrates model parameters one by one, the result may be trapped into a local optimal.

A systematic validation approach of a microscopic simulation model was described by Zhang et al. (2004). The procedure includes animation comparison and quantitative/statistical analysis at both macroscopic and microscopic levels. Data at the macroscopic level include the averages and other statistics of traffic variables and fundamental relationships of traffic flow parameters. Data at the microscopic level include the speed change pattern, vehicle trajectory plot, and headway distributions. Animation comparison was supplemented to examine the model validity. The procedure emphasized the importance of real-world datasets to model validation. However, the datasets used in this study were truly microscopic and were very expensive to obtain.

Genetic Algorithm-Based Studies

GA is a heuristic optimization technique based on the mechanics of natural selection and evolution (Goldberg, 1989). It works with a population of individuals, each representing a possible solution to a given problem. A fitness value is assigned to each individual according to

how good its represented solution is. Selecting the best individuals from the current generation and mating them to produce a new set of individuals produce a new population of possible solutions. Schema theorem and building blocks hypothesis are rigorous explanations of how the highly fit individuals survive to the next generations (Beasley et al., 1993).

GA has been considered for the past decades to be one of the most effective search and optimization techniques in developing artificial intelligence systems. GA has been applied to various transportation engineering problems. Recent works can be found in traffic signal design and optimization (Park et al., 2004; Sun et al., 2004), network design (Chen et al., 2004), vehicle location and assignment (Yang et al., 2004), pavement design (Tsai et al., 2004), and bridge management system (Hegazy et al., 2004). Successful applications of GA for simulation model calibration are summarized as follows.

GA was first introduced into model calibration and validation in late 1990s. Cheu et al. (1998) presented the application of GA as an optimization method for finding a suitable combination of FRESIM parameter values. Their objective was to calibrate the simulation model based on observed traffic conditions so that the model can accurately represent driver behavior on a 5.8 km freeway section in Singapore. Twelve traffic flow model parameters were considered in their calibration. The results showed that the GA was able to search for the optimal parameter values that enabled FRESIM to produce 30-second loop detector volume and speed closely matching the field data under peak and off-peak traffic conditions. The calibrated parameters were found to be quite reasonable and indicative of local driver behavior in Singapore.

Liu and Mahmassani (2000) applied GA to the calibration of discrete choice models. They recognized the fact that the structure of the model did not lend itself to traditional non-linear programming (NLP) techniques due to the likely abundance of local optima in the search space, and, consequently, they introduced the idea of calibrating multinomial probit (MNP) models using GA. The developed method extensively calibrated a MNP model with three alternatives and four model parameters using a synthetic data set. The results indicated a very small discrepancy between the genetically generated parameters and the optimal values (<1%).

GA was explored to determine a suitable combination of parameter values for PARAMICS by Lee et al. (2001). Two key parameters – mean target headway and mean reaction time – were calibrated against field data based on 30-second loop detector volume and occupancy. The results showed that the calibrated PARAMICS model by GA was able to realistically represent the freeway traffic flow in California—a significant improvement over default parameter values.

In 2004, Kim and Rilett illustrated a GA-based approach to traffic simulation model calibration using ITS data. Two microscopic simulation models, CORSIM and TRANSIMS, were calibrated for two freeway segments in Houston, Texas. It was found for both test beds that the GA had better results than either the default values or a simple, manual calibration. The analysis showed the benefit of using a GA for the automatic calibration of simulation models.

Evaluation of Simulation Models

CORSIM

CORSIM (CORridor SIMulation) is a detailed stochastic microscopic simulation model. The early version was developed in the mid-1970s by the FHWA. It is one of the most widely accepted models in the United States. CORSIM is a core component in the TSIS (Traffic Software Integrated System) suite, which is an integrated development environment that enables users to conduct traffic operations analysis.

CORSIM is designed for the analysis of freeways, arterial road links, and basic transit operations. It combines two microscopic simulation models: the arterial network model NETSIM and the freeway model FRESIM. As a microscopic model, CORSIM simulates the interactions of individual vehicles using traffic flow algorithms and provides many of the details of the refined modules, such as car-following module, traffic control module, traffic assignment module, and so forth. A Run-Time Extension (RTE) is provided to modify CORSIM's default behavior. The version of CORSIM used in this research was version 5.1.

User Interface

CORSIM's input files were constructed using a text editor because the input file has a defined structure. Users can code the network by providing an appropriate input file or build network nodes and links through a GUI-based network and simulation input editor, TRAFED. TRAFED can import bitmap background images. In addition, two signal optimization models, Synchro and TEAPAC, can export street networks to CORSIM. The animation is realized through a post-processor, TRAFVU (TRAFfic simulation Visualization Utility).

Traffic Generation and Assignment

Traffic volumes may be entered in terms of vehicle counts for certain time period or flow rate for up to 19 time periods. Vehicles can be generated from Uniform, Normal, and Erlang distributions.

Routing logic is based on link-based turning movement volumes or percentages for NETSIM and FRESIM models. NETSIM deploys a traffic assignment model, which combines static traffic assignment techniques with equilibrium and optimization capabilities. The criterion for determining when user equilibrium has been reached is that no driver can reduce his/her journey time (or impedance) by choosing a new route. The criterion for the system optimization is the minimum total cost of the entire network. The routing logic will convert the O-D demand into turning percentages and entry volumes for the NETSIM model. A gravity model is implemented in FRESIM to calibrate the turning percentages to an Origin Destination (OD) table (FHWA, 1997).

Traffic Control

CORSIM can model yield signs, stop signs, and traffic signals. CORSIM provides a user-friendly interface for editing actuated controller properties and is capable of simulating pretimed and actuated signals controllers allowing one signal plan to be implemented per simulation period. The program can simulate the NEMA or Type 170 controller logics, which are widely adopted in the field signal systems. Actuated controller operation may include semi-actuated, fully actuated, and volume-density functions.   In addition, three pedestrian actuation modes are provided for actuated control. Coordinated and uncoordinated signal operation may be used in the same network. CORSIM also can model freeway ramp metering with four types of on-ramp signal control strategies. Roundabouts are not modeled explicitly by CORSIM, but can be modeled using basic CORSIM elements.

Multi-modal Transportation

Traffic stream may consist of nine different types of vehicles having various operating and performance characteristics. The default four different fleet components include passenger car, truck, bus, and carpool, consisting of several default vehicle types. The user can specify any vehicle type with specific performance characteristics if he/she is not satisfied with the default values for vehicles. However, in general, CORSIM is not a multi-modal simulation tool because it is limited in transit capabilities, such as transit signal priority and little or no pedestrian or bicycle functionality.

ITS Capabilities

CORSIM can interface with a run time extension (RTE) to test adaptive signal control algorithms for feasibility before they are implemented in the field. FRESIM can simulate warning signs, similar to variable message signs (VMS), in the upstream of a long-term incident or a work zone. Therefore, travelers who are aware of these activities may consider lane change earlier to avoid the blocked lane(s).

Measures of Effectiveness

CORSIM provides a variety of numerical MOEs, which can be link specific, aggregated for multiple links or network wide. The main MOEs that NETSIM provides include vehicle trips; vehicle miles; travel time; delay time including queue delay, control delay, and total delay; stopped time and percentage; average volume/speed/occupancy; phase failure; and average/maximum queue length. The main MOEs that the FRESIM provides include vehicle throughput, vehicle miles, travel time, delay time, volume, density, and speed.

Weaknesses

As mentioned before, CORSIM is not a multi-modal tool because of limited transit capabilities and pedestrian or bicycle functionality. It has very limited capabilities to assess ITS technologies such as route guidance systems. It does not provide 3D graphical output, hence is not as attractive as other simulation models such as VISSIM and PARAMICS, which have

superior graphical output. Since CORSIM generates link-based outputs, it is difficult to obtain certain route-based measures or individual measures such as travel time unless RTE is used.

VISSIM

The Verkehr In Staedten SIMulation (VISSIM) was the second microscopic simulation model selected in this research. The model was developed at the University of Karlsruhe, Germany, during the early 1970s. Commercial distribution of VISSIM began in 1993 by PTV Transworld AG, which continues to distribute and maintain VISSIM today.

VISSIM is capable of simulating traffic operations in urban streets and freeways with special emphasis on public transportations and/or multi-modal transportation. VISSIM consists of two different programs: the traffic simulator and signal state generator. The traffic simulator is a microscopic simulation model comprising of car-following logic and lane-changing logic. Instead of using constant speeds and deterministic car-following models, VISSIM uses the psychophysical driver-behavior model developed by Wiedemann (1991), which is closer to field-driving behavior. It is noted that Wiedemann developed two different models: one for freeway and the other for surface street. The simulator is capable of simulating at a resolution of up to one-tenth seconds. The version of VISSIM used in this research was version 3.7.

User Interface

The VISSIM model provides a graphical user interface to construct the networks and view animation. It has the interface to Signal97/TEAPAC to evaluate and optimize traffic operations. For users' convenience, VISSIM can import background images only in a bitmap format. Other vector graphics, such as CAD files in DXF format, must be converted before they can be used in VISSIM. The interface for VISSIM provides the user with guidance regarding coding errors before the simulation starts. In addition, VISSIM is capable of displaying 3D animation.

Different from typical simulation models based on link-node structures, VISSIM uses link-connector topology for network construction. Its flexible network structure allows the detailed representation and modeling of any geometry from freeway corridors, signalized intersections, and roundabouts to entire bus stations and even airports.

Traffic Generation and Assignment

Users may enter traffic volume in terms of hourly rates or O-D demands for multiple time periods. The traffic volumes may change during the simulation at specified times for a particular link or for the whole network. Vehicles are generated using a Poisson distribution. VISSIM simulates traffic flow in terms of driver-vehicle units. Each vehicle in the network is processed up to 10 times per second.

There are basically two routing logics in VISSIM: static traffic assignment and dynamic traffic assignment (DTA), which is available only with optional VISSIM module "Dynamic Assignment." With static traffic assignment the path of vehicles traveling through the network

can be statically determined by either routing decisions or direct decisions. Routes are the recommended practice to assign vehicle paths because they are much easier to handle and the vehicle flows can be defined more precisely. Dynamic traffic assignment uses O-D demands to assign vehicles to the network based on the iterated simulations. That means a network is simulated not only once but also repetitively, and the drivers choose their routes through the network based on the travel cost they have experienced during the preceding simulations. The iteration of the simulation runs is continued until a stable situation is reached (PTV, 2001).

Traffic Control

VISSIM can model yield signs, stop signs and traffic signals, up to 128 phases. At signalized intersections, the signal controllers include built-in pretimed control and actuated control by the external signal state generator or NEMA graphical user interface (GUI), a new feature of version 3.7. VISSIM provides vehicle actuated program (VAP) macro language, which is similar to C language and allows users to completely define their specific signal control logic including any type of special features such as transit priority, railroad preemption, and emergency vehicle preemption. VISSIM also can act as a "virtual controller" using signal control logics such as VS-PLUS (Vehrkers System – Plus). VISSIM can import the signal timing plans of three TEAPAC programs: PRETRANSYT, TRANSYT-7F, and SIGNAL97 (PTV, 2001). Furthermore, pedestrians are shown in the simulation when pedestrian actuation is permitted.

Multi-modal Transportation

VISSIM uses a hierarchical concept to define and provide vehicle information at different levels throughout the application. There are six default vehicle classes in VISSIM: car, truck, bus, tram, bike, and pedestrian. Users are allowed to define additional vehicle classes with unique characteristics, such as length, fuel mileage, speed, and so forth. The NEMA GUI used to enter actuated signal timings also has custom menus to allow users to model bus, light rail transit (LRT) priority, and railroad preemption. Many successful VISSIM applications with complex transit alternatives exist in the literature.

ITS Capabilities

VISSIM is capable of testing adaptive signal control algorithms prior to field implementation. VISSIM also can model various transit priority strategies and train preemption strategies. VAP can be applied for ITS measures like dynamic speed control, dynamic lane allocation, and dynamic route guidance systems.

Measures of Performance

VISSIM provides a comprehensive set of output files that are user-customized. The numerical data are categorized as follows:

- *By link* – average speed, volume, travel times, total delay, stopped delay, average queue length, maximum queue length, number of vehicle stops within the queue, vehicle density, emissions, fuel consumption;
- *By path* (only with dynamic traffic assignment) – travel times, volume, path number;
- *By point* – speed, acceleration, occupancy;
- *By vehicle* – speed, acceleration, occupancy, travel time, delay, number of stops, queue time, emissions, fuel consumption, bus/tram waiting time.

Weaknesses

In VISSIM, the priority rule allows users to code the intersections in fine detail. However, entering conflicting movements at an intersection is a rather cumbersome and time-consuming procedure because priority rules require detailed adjustment on locations and parameters such as minimum gap and headway.

PARAMICS

Developed by Quadstone Limited, the PARAMICS (Parallel Microscopic Simulation) is a suite of high performance software tools consisting of PARAMICS Modeler, Processor, Programmer, Analyzer and Monitor. PARAMICS modeler is the core simulation tool, which is designed to model the movement and behavior of vehicles in urban and highway road networks (Quadstone, 2002). The three basic models – car-following, gap acceptance, and lane changing models – are implemented within PARAMICS to control the movement of individual vehicles. An application programming interface (API) is provided to allow the advanced user to customize many features of the underlying simulation model. The version of PARAMICS used in this research was version 3.0. The latest version, 4.2, is now available.

User Interface

PARAMICS Modeler provides a graphical user interface to construct the networks and to view animation. Users can build networks with the help of background image in DXF, BMP, and OS/NTF Strategi format (a UK standard format for map data) (Park et al., 2003). PARAMICS network coding is based on a node-link structure. Since travel demand is modeled as zone-to-zone movements in PARAMICS, zones and zone connectors need to be defined before simulation occurs.

Traffic Generation and Assignment

Travel demand in PARAMICS is represented by an origin/destination matrix of trips and can have multiple-demand time periods. Based on the time period and corresponding demand, vehicles are generated randomly. The three main assignment methods in PARAMICS are all-or-nothing assignment, stochastic assignment, and dynamic feedback assignment. If the network is simple and no alternatives exist, the all-or-nothing assignment usually is selected. A brief description of the three assignment methods follows (Quadstone, 2002):

- All-or-nothing assignment assumes that all drivers traveling between two zones choose the same route and that link costs do not depend on flow levels.
- Stochastic assignment tries to account for variability in travel costs (or drivers perception of those costs). These methods assume that the perceived cost of travel on each network link varies randomly, with predefined limits.
- Dynamic feedback assignment assumes that drivers who are familiar with the road network will re-route if information on the present traffic conditions is fed back to them. This is achieved by taking real-time information from the PARAMICS simulation and using these data to update the routing calculations.

Traffic Control

PARAMICS can model yield signs, stop signs and traffic signals up to 15 phases. Yield and stop driving behaviors are realized by coding priority junction. A hierarchy of priorities exists in the order of major, medium, minor, and barred for each movement in the priority junction. At signalized intersections, PARAMICS is capable of simulating pretimed control and actuated control. However, the API with Visual C++ or C language needs to be developed by the user to achieve actuated control logic.

Multi-modal Transportation

The default types are car; truck including lgv (light goods vehicle), ogv1 (ordinary goods vehicle class 1), and ogv2 (ordinary goods vehicle class 2); and service buses (large or minibus). Users can override any physical characteristic by changing the specification in the "vehicle" file. Users can add pedestrians to the model by creating a pedestrian vehicle (small, slow speeds).

ITS Capabilities

Though PARAMICS can model simple ITS technologies, API is eventually required to implement more complicated ITS strategies. With advanced plug-in modules using API, PARAMICS can accurately model the interface between ITS and the affected traffic under traffic management and control applications to identify the best location for information signs at the design stage and to determine the optimal strategy for the use of ITS in on-line traffic management and control at the operational stage. In addition, some fully integrated simulations of ITS elements can be realized, for example ramp meters, VMS signing strategies, and in-vehicle re-routing suggestions (Chu et al., 2004).

Measures of Effectiveness

PARAMICS provides many statistics through the output files, and the "measurement" file defines a list of statistics to be collected. The data include:

- *General network data* – snapshot, turn counts, queue counts, release counts (denied vehicles), number of vehicles on the network, distance traveled, mean speed, mean stopped time and bus-specific data;

- *By point data (loop detector)* – flow, speed, headway, occupancy, acceleration and count [light/heavy] vehicles only;
- *By link (loop detector)* – link flow, link speed, link density, lane change (the number of lane-changing events), lane usage (the proportion of vehicles in each lane), link delay and link counts;
- *By path/trip* – trip information, travel time, trip counts, total delay (total journey time) and bus delay;
- *Others* – gather distribution, group loop data, saturation flow, incidents, trace, feedback costs and car park information.

Furthermore, users can obtain more desired output by developing an API code.

Weaknesses

Although PARAMICS is a powerful simulation model that provides simulations in fine detail, it takes some time to become familiar with the tool and master the preconfigured functions. In addition, API is not easy for beginners to develop, especially those who do not have any programming experience on Visual C/C++. In addition, PARAMICS does not allow the dual-ring or NEMA control concept. Therefore, it may be cumbersome for users to develop actuated signals.

Comparison

Table 1 provides a general comparison of the models, highlighting the aspects important to most users. The price of these models varies widely. CORSIM has the lowest price, while VISSIM with high level and PARAMICS are much higher. The price of VISSIM is based on the overall functionality of the model. It is most expensive when transit and dynamic traffic assignment are included.

Among these three simulation models, VISSIM and PARAMICS do not have network limits (except as memory limited by a computer's memory), while CORSIM has different levels of limits on NETSIM and FRESIM. To account for stochastic variability of daily traffic, multiple simulation runs need to be conducted for each scenario. All simulation models allow command line execution of simulation model runs and can conduct simulation without animation, which are time efficient. In terms of traffic control, all models can simulate yield sign, stop sign, and traffic signals including pretimed and certain types of actuated signals. PARAMICS requires extra effort, however, such as an API code, for actuated signals.

All models provide a variety of numerical outputs that are link-based, and/or route-based, and/or network-based, and/or individual vehicle-based. However, users need to check carefully the definition and calculation method of some performance measure outputs, such as delay and queue data, as they vary by programs. In addition, none of the models provides output in the HCM format, so conversions must be created in other programs to compare results to traditional analytical models.

Table 1. General Comparison

Model Comparison		CORSIM	VISSIM	PARAMICS
General	Version used	5.1	3.7	3.0 & 4.0
	Price	$500	Basic (Level I)- $500	$13,310
			w/o transit (Level II) - $1,500	
			w/ transit (Level III) - $12,000	
			w/ dynamic assignment (Level IV) - $15,000	
	Developing tool	Yes - RTE	Yes - VAP	Yes - API
User Interface		Text editor; Graphical user interface (TRAFED); Network can be imported from Synchro and TEAPAC	Text editor; Graphical user interface (main option); TEAPAC can export singal to VISSIM	Text editor; Graphical user interface (main option)
Network Limit		NETSIM - 9000 nodes, unlimited links and vehicles, 7000 detectors, 1000 actuated signals, 1000 pedestrian phases, maximum link length 9999 feet	None, except for memory limits on a computer	None, except for memory limits on a computer
		FRESIM - 9000 nodes, unlimited links, segments, and vehicles, 6000 detectors, 5000 connectors, 2000 detector stations, 4000 ramp meters		

Table 1. General Comparison (Continued)

Model Comparison	CORSIM	VISSIM	PARAMICS
Traffic Control	Yield signs, stop signs, pretimed/actuated signal, ramp-metering control, roundabout (some)	Priority rules (similar to yield signs), stop signs, pretimed signal, actuated signal (VAP code or NEMA interface), roundabout	Priority junction (similar to yield signs), stop signs, pretimed signal, actuated signal (API code needs to be developed by the user), roundabout
Multi-model Transportation	Car, truck, bus, pedestrian, and user-friendly modification	Car, truck, bus, rail, tram, bike, pedestrian, and user-friendly modification	Car, truck, bus, pedestrian, and user-friendly modification
Traffic Assignment	Static traffic assignment with equilibrium and optimization capabilities	Static traffic assignment, dynamic traffic assignment	Static traffic assignment, dynamic traffic assignment
Measure of Performance	Traffic volume, delay time, travel time, control delay by turn movement, stopped delay, queue time, queue length, vehicle speed, vehicle fuel consumption, vehicle emission by link	Traffic volume, vehicle speed, mean speed, travel time, total delay, stopped delay, average queue length, maximum queue length, vehicle stops within the queue, bus/tram wait time, vehicle emission by vehicle/link/path	Point/link flow, point/link speed, headway, occupancy, acceleration, density, link/bus/total delay, turn/queue/link counts
Graphic ouput	2D animation	2D &3D animation	2D &3D animation
Multi-Run	Corsim driver interface window, command line, scripts	MULTI interafce, command line	Command line

Given individual strengths and weaknesses of the different simulation models, users need to choose appropriate simulation model and performance measures with caution for different simulation-based applications. Blind application of an inappropriate simulation model or an improper application of the right model can result in inaccurate and unreliable results.

A Proposed Procedure for Microscopic Simulation Model Calibration and Validation

The following sections describe the main steps in the proposed calibration and validation procedure, as well as the relevant techniques used in this study. Figure 1 depicts the proposed methodology using a flowchart.

Simulation Model Setup

Simulation model setup, the first step, comprises those tasks and activities that are conducted prior to the commencement of model calibration. These tasks consist of the definition of study scope and purpose, site selection, determination of MOEs, field data collection, and network coding.

Initial Evaluation

Once a simulation model is correctly set up, the simulation performance based on the default parameter is obtained and compared to the field data. If a close match between the two is found, the default model could be deemed appropriate to be used for further analysis. Otherwise, the next steps are to be practiced.

Initial Calibration

Identification of Calibration Parameters

All calibration parameters within the simulation model are to be identified, except for those that do not have any relevant impacts on the results. For example, for a simple network without route diversion possibility, the parameters related to the route changes could be eliminated from the calibration parameter list. For selected calibration parameters, their acceptable ranges initially are to be determined by the user.

Experimental Design for Calibration

The number of combinations among feasible controllable parameters is extremely large such that all possible scenarios cannot be evaluated in a reasonable time. A Latin Hypercube Design (LHD) algorithm, an experimental design method, is used to reduce the number of combinations to a practical amount while still reasonably covering the entire parameter surface. The LHD provides an orthogonal array that randomly samples the entire design space broken into equal-probability regions. This type of sampling can be regarded as a stratified
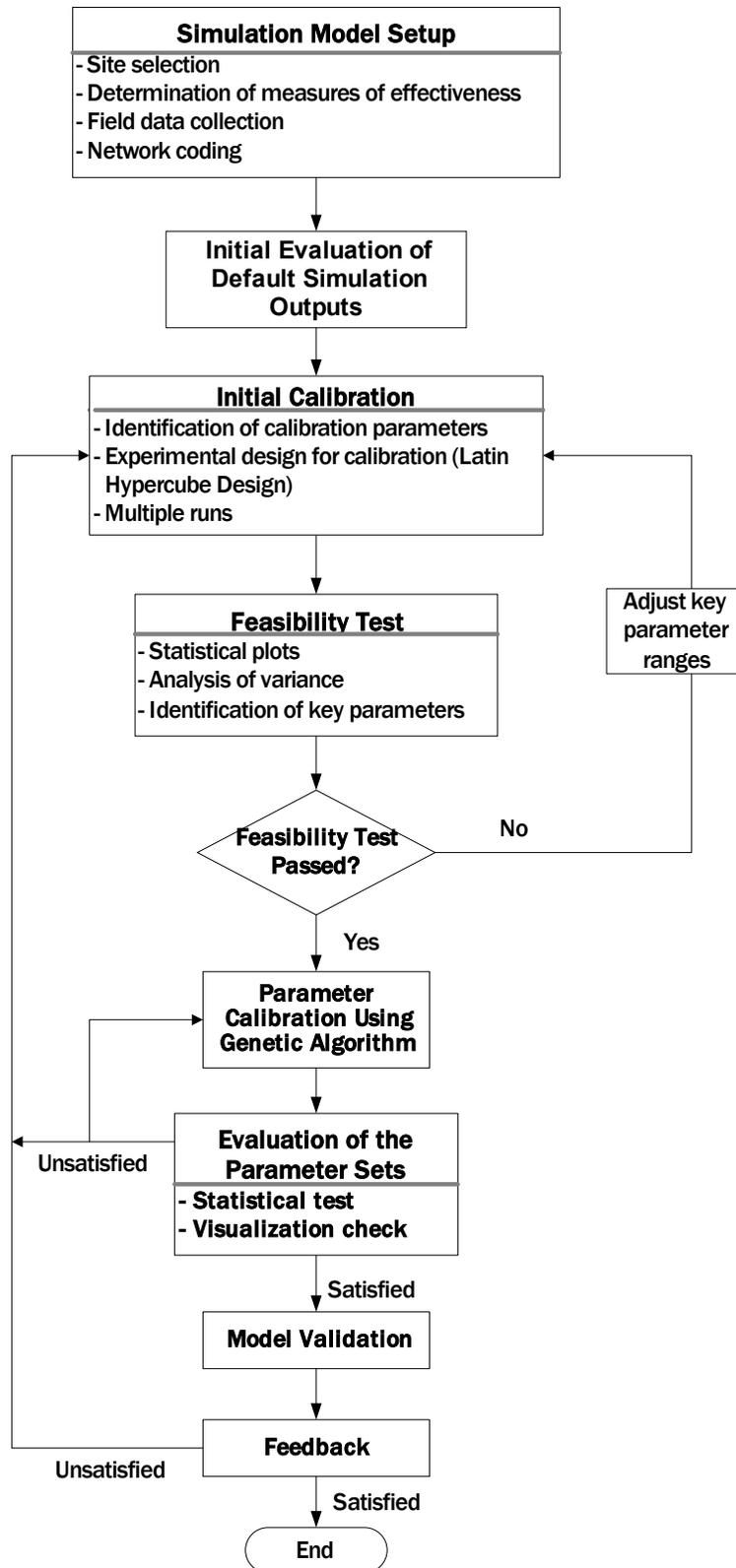
Figure 1. Methodology Flow Chart

Monte Carlo sampling, where the pair-wise correlations are minimized to a small value, which is essential for uncorrelated parameter estimates (McKay, 1979). The LHD is especially useful for exploring the interior of the parameter space and for limiting the experiment to a fixed or user-defined number of combinations, say N. This technique ensures that the entire range of each parameter is sampled. In this study Latin Hypercube Sampling toolbox in MATLAB (The Mathworks, 1999) was used to generate N scenarios using calibration parameters of each simulation model.

Multiple Runs

In order to reduce stochastic variability, multiple runs are to be conducted for each scenario from the experimental design. The number of replicates should be a function of system variability and computation time constraints. The average performance measure and standard deviation are to be recorded for each of the runs. The simulation results are to be used in the feasibility test.

Feasibility Test

The purpose of the feasibility test is to check whether the field data are reasonably covered by the distribution of simulation results, hence to determine whether the current parameter ranges are feasible. As shown in Figure 2, if field data fall within the middle 90 percent of the distribution, defined as the acceptable range, the current parameter ranges are deemed sufficient to generate the field condition. Otherwise, the ranges of key parameters are to be adjusted in order to achieve the desired results. However, the expanded value should be reasonable and realistic. With the adjusted parameters another set of N scenarios is generated and evaluated. This process is repeated until the field data fall within the acceptable range. Furthermore, the scenario containing the parameter set that provided the closest match to the field performance is recorded for final comparison, which is noted as LHD-based parameters.
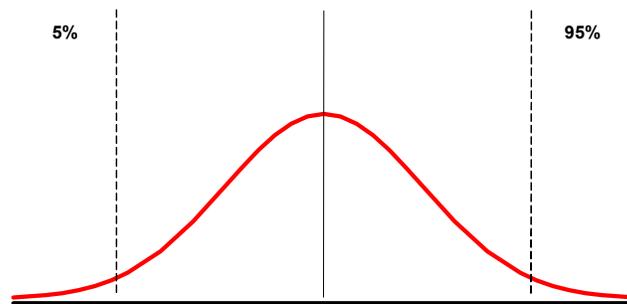

Figure 2. Feasibility Test Concept

Statistical Plots

With the performance measure recorded in each of the runs, statistical plots such as scatter plots and histograms are drawn to help understand the results. The plots include the distribution of performance measures from multiple runs and performance measure versus each calibration parameter. For the key parameters, usually the trend is easy to observe.

Analysis of Variance

Analysis of variance (ANOVA), another rigorous statistical method, is applied to identify key parameters based on statistical analysis result. ANOVA tests the null hypothesis that several group means are equal in the population by comparing the sample variance estimated from the group means to that estimated within the groups. This technique is an extension of the two-sample t test.

In this study each pair of calibration parameter and simulation results was examined by one-way ANOVA in a statistical package SPSS (SPSS, 2001). It can identify statistically significant parameters based on the significance value of the *F* test and can quantify the degree to which each parameter explains the performance measure. When the p value is smaller than the user-defined confidence level, the null hypothesis is rejected, which indicates that group means are statistically different. Therefore, the parameters with small p values and big SSRs are identified as key parameters. Furthermore, in order to see the main and interaction effects of parameters on the simulation results, multiple analyses of variance (MANOVA) is to be applied to parameters that may have interaction effects.

Parameter Calibration Using Genetic Algorithm

After identification of appropriate parameters and the acceptable ranges, GA is applied to find the optimal parameter values. Three basic operators employed in GA are reproduction, crossover, and mutation. The reproduction operator selects individuals with higher fitness. In this study, each individual was represented by a series of unsigned decimal numbers (i.e., 0,1,2,…,9). Then, the actual value of each parameter was linearly mapped into an integer range permitted by the given number of decimal digits. For example, the first parameter with the range of 0.8 sec to 1.8 sec was represented by the first two unsigned decimal digits in a chromosome, which offered a numeric range from 00 to 99. The lower limit of 0.8 sec was set to 00 while the upper limit of 1.8 sec was set to 99. Therefore, a value of 20 on these two digits would produce the parameter value of 1.0 sec by a linear mapping. Then, the actual parameter values were entered into the traffic simulator, and the fitness value was calculated based on the simulation output and field data. The crossover operator creates the next population from the intermediate population, whereas the mutation operator is used to explore some areas that have not been searched.

The procedure of applying GA for calibration is depicted in Figure 3. First, parameters that control the size of GA search problem and the function to calculate the fitness value are set. The GA is then started by generating a number of individuals in the population; each representing a feasible solution. These individuals form the initial solution set of the problem. Second, each individual is decoded into actual parameter values and inserted into the simulation model. A simulation is then carried out for each set of decoded values, with simulation results channeled into a fitness function. Third, the fitness value is calculated by comparing the simulation results and field data. If the comparison result does not meet the stopping criterion, then a new population is generated after implementations of crossover, mutation, and elitism options. The process is repeated until the stopping criterion is met or the maximum number of generations is reached.
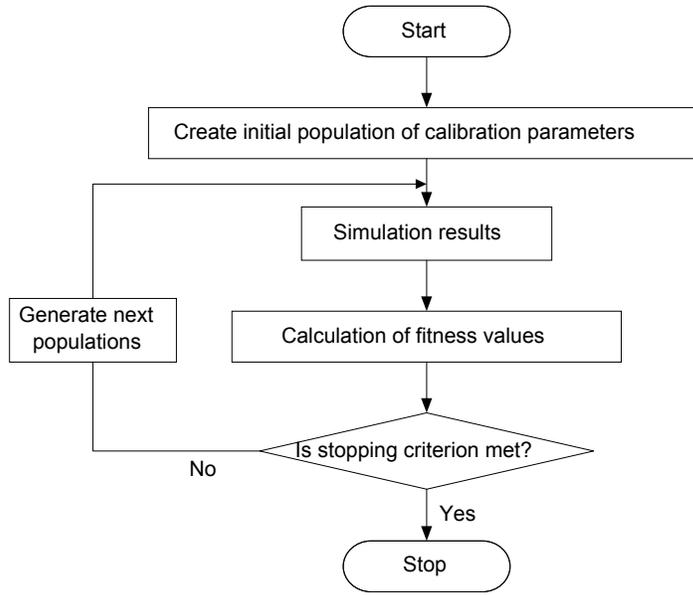
Figure 3. The Genetic Algorithm Process

GA was integrated into each simulation model to calibrate parameters. In this study the relative error of average travel time between simulation output and field data was used as the fitness value for the GA. The fitness function takes the form of:

$$\text{Fitness Value} = \frac{\left| TravelTime_{Field} - TravelTime_{Sim} \right|}{TravelTime_{Field}} \qquad (1)$$

where travel time refers to average value of individual travel times.

Evaluation of the Parameter Sets

Once GA finishes parameter calibration, multiple runs are conducted for default parameters and GA-based parameters, separately, for comparison. For each parameter set, a distribution of performance measure is to be developed and compared with the field measure. Statistical testing, such as t test, and visualization checking will be conducted. The t test is to be used to verify whether the calibrated parameter set from GA can generate statistically significant results and perform better than the default parameter set. In this study the performance measure used in the t test was average travel time within the study period. Visualization is to be used to evaluate the calibrated models. A model cannot be considered calibrated if the animations are not realistic. Animations from multiple runs of each model are to be viewed to ensure the animations are acceptable. The model will be regarded well calibrated only if the evaluation of travel time distributions, statistical tests, and animations are satisfactory. Otherwise, either parameter calibration using GA or the initial calibration stage must be repeated.

Other parameter sets that could be included in the comparison are the best-guessed parameter set and the best parameter set identified in the LHD approach, which has simulation results closest to the field. Here best-guess means to assign the most accurate value to each parameter based on engineering judgment and knowledge of local traffic condition. The inclusion of LHD-based parameter is to check whether it can achieve comparable performance with the calibrated parameters using GA and thus, provide an alternative to GA to obtain acceptable parameters.

Model Validation

In order to validate the model, it is desirable to collect another set of data under different conditions than in the calibration. Different conditions could be untried conditions such as after implementing new signal timing or different days with similar transportation facilities. In this study field data under untried conditions were not used due to some limitations. Instead, data that were collected on a different day were used for validation. The validation involves comparison of the prediction of system performance from the calibrated model with the newly collected field data. T test and visualization checking are to be conducted to evaluate if the predictions are acceptable.

Feedback

Feedback will be necessary if the results are not satisfactory. This feedback will go to the calibration stage and repeat the remaining steps.

<div align="center">Development of Test Beds</div>

Site Selection

The first test site was an actuated signalized intersection. It is the junction of U.S. Route 15 and U.S. Route 250 in Virginia, which has four legs with single lane approach from all directions. The test site is referred to as "Site 15" throughout the remaining report. The location and alignment of the test site is shown in Figure 4. There are exclusive right-turn bays of 110 feet long on each approach. The southbound approach carries relatively heavy traffic volume and a large proportion of left-turn vehicles during peak hour.

The second test site was a lane-closure work zone located on I-64 near Covington, Alleghany County, Virginia. The selected 5-mile highway segment plies longitudinally in the east-west direction, and the westbound direction is of interest in this study. The test site is referred to as "Covington network" throughout the remaining report. This site was selected because (1) it is an interstate highway, which has different driving behavior from the signalized intersection in the first case study; (2) it involves a relatively large work zone, which has specific driving phenomenon such as slowing down before the merge area and speeding up after the work zone; and (3) the work zone consists of a bridge reconstruction, which takes a fairly long time and hence is favorable for site visits and multiple-day data collections. There are two lanes at the beginning of the network but only one lane at the work zone area. Moreover, there are three different posted speed limits along the network. The normal posted speed limit for interstate

<div align="center">21</div>

highways is 65 mph. At the beginning of the network, there is a posted speed limit sign of 55 mph, and further downstream the posted speed limit is 45 mph. Figure 5 is an aerial photo showing the Covington network alignment and location.
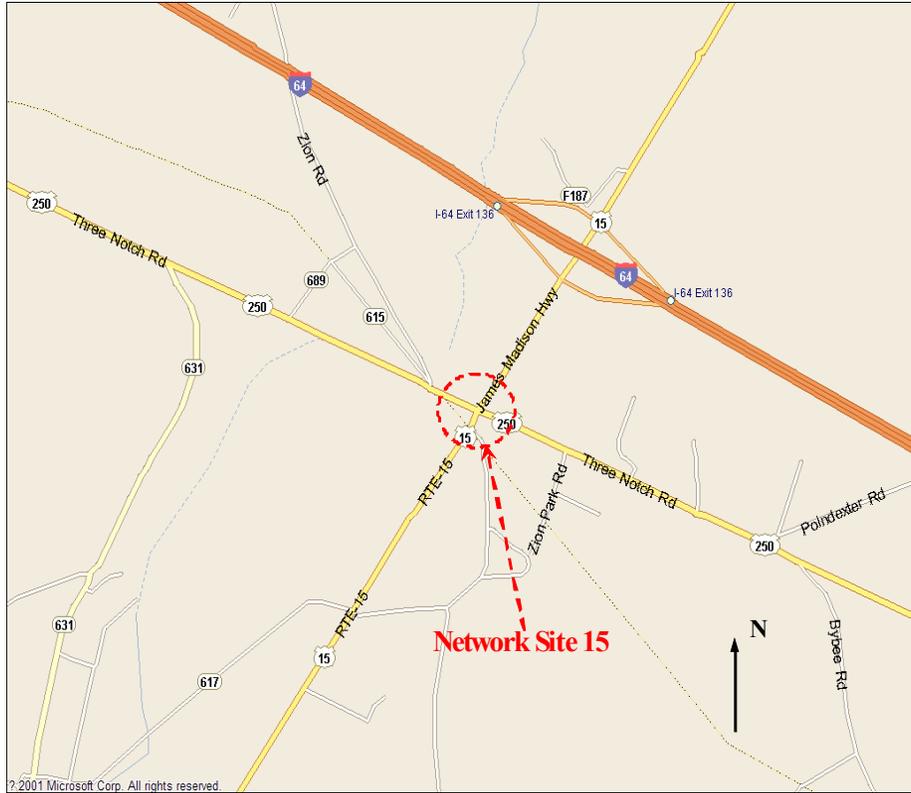

Figure 4. Schematic of Site 15

Figure 5. Schematic of Covington Network (From Geostat Center at University of Virginia)

Data Collection and Reduction at Site 15

Data Category

To code Site 15 in the simulation environment, input data such as traffic counts, length of each approach lane, intersection geometric characteristics, detector locations, and posted speed limit were required. The signal timing designed for the signal system was obtained from VDOT personnel. The MOE used in this network was travel time on the southbound approach, and it was selected because it easy to collect both from the field and the simulations. Other measures, such as delay and queue data, are calculated in different ways in each simulation model. For example, queue data usually are collected in the field at the beginning of green time for each cycle. In simulation environment, CORSIM provides average and maximum queue by lane from the beginning of the simulation, while VISSIM identifies average and maximum queue by link at certain time intervals according to a user-provided queue definition (PTV, 2001). Therefore, comparison of these MOEs to the field value may not be meaningful.

Data Collection and Reduction

Before the data collection the research team made site visits to determine the data collection strategy. The data collection plan included the following: one smart travel van (STV), two video cameras, and three persons, as shown in Figure 6. Person 1 recorded actual signal timing changes from the actuated signal system during the data collection period. Persons 2 and 3 were responsible for the two video cameras, which were located in the southbound approach, to obtain the license plates and times of each vehicle at the upstream and downstream point. The distance between these two cameras was 960 feet. Another two cameras attached to STV collected detailed data in two areas corresponding to the dashed circle and rectangle in Figure 6. The camera for circular area provided traffic counts and vehicle classification; the camera for rectangular area provided detailed traffic movements on the southbound approach.

To account for day-to-day variability, data were collected during evening peak hour between 5 and 6 P.M. on multiple days, April 15, April 22, May 13, May 20, and June 5 in 2003. Synchronization between clocks of all equipment was performed before data collection. Several things occurred during data collection that should be mentioned. On April 15 some of the traffic count data were missing due to the improper angle of the camera. Only one STV camera was working properly on April 22, May 13, and May 20, which resulted in the loss of detailed data on the southbound approach. Despite the missing data, the research group still collected a comprehensive range of data.

Videotapes were reduced to obtain traffic counts, individual travel time on the southbound segment, and real signal timing. Traffic counts and heavy vehicle counts for each approach were obtained through viewing the video attached to the STV. The signal statistics such as maximum/minimum green, red and yellow were extracted from the recorded signal data. The posted speed limit of 45 mph was observed at the site. Travel time was collected by the license plate matching method through watching the two videotapes. The procedure was to note the license plate numbers and times of each vehicle as it passed the observation points, then
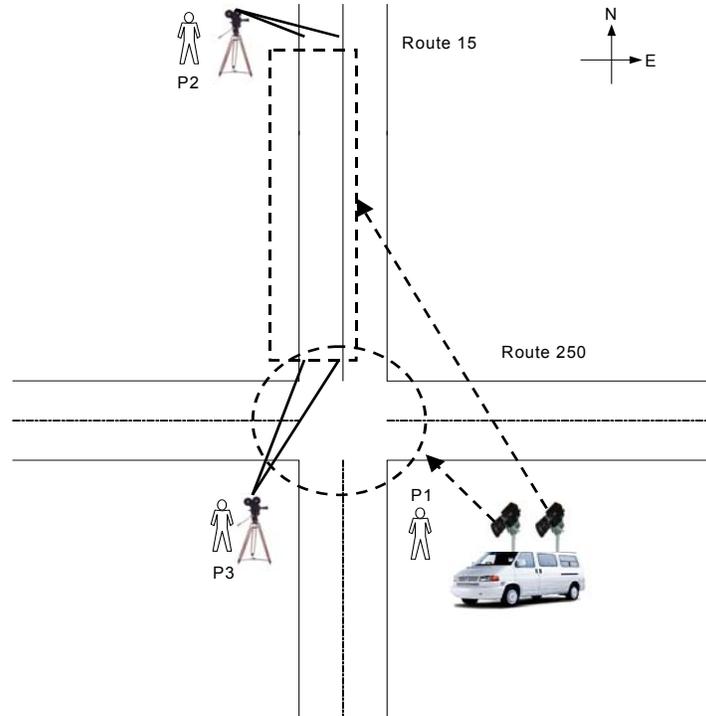
Figure 6. Site 15 Data Collection Schedule

match the license plate numbers one by one, and finally calculate the time differences to obtain the travel times of vehicles traveling along the segment.

Data Summary

Since there were only two phases in this signal system, real-time signal data were recorded only for major directions, southbound and northbound, as listed in Table 2. The signal timing for minor approaches could be inferred from Table 2.

Table 2. Actual Signal Timing at Site 15

Major Signal	Green (sec)			Red (sec)		
	Mean	Min	Max	Mean	Min	Max
5/13/03	30.9	12.6	58.2	24.1	12.6	36.9
5/20/03	29.6	12.0	48.1	23.0	12.1	37.2
6/5/03	30.6	13.5	70.6	25.0	6.1	37.3
Average	30.4	12.7	59.0	24.0	10.3	37.1

Table 3 lists traffic counts on four individual days and average traffic count of three days: April 22, May 13, and May 20, 2003. Data from these three days were used in the calibration process, while data on June 5, 2003, were reserved for validation. Trucks and other multi-axle large vehicles were identified as heavy vehicles (HV).

24

Table 3. Field Traffic Counts at Site 15

Site 15 Traffic Counts (vph)			Left Turn	Through	Right Turn	HV%
Calibration Data	4/22/2003	Southbound	217	312	122	3
		Northbound	36	188	40	2
		Eastbound	80	86	33	1
		Westbound	40	64	80	3
	5/13/2003	Southbound	195	270	111	2
		Northbound	22	193	40	5
		Eastbound	75	111	42	3
		Westbound	34	82	104	4
	5/20/2003	Southbound	208	289	99	1
		Northbound	32	171	33	4
		Eastbound	83	94	53	2
		Westbound	33	74	90	2
	Average Counts	Southbound	207	290	111	2
		Northbound	30	184	38	4
		Eastbound	79	97	43	2
		Westbound	36	73	91	3
Validation Data	6/5/2003	Southbound	192	300	93	5
		Northbound	32	172	29	6
		Eastbound	88	98	44	4
		Westbound	45	77	84	2

Table 4 shows statistics of travel time, including mean, median, standard deviation, and minimum and maximum values. The data indicate a huge variation of mean travel time among different days, ranging from 46.51 to 70.43 seconds. Even on the same day, travel time of individual vehicles varied dramatically as indicated in standard deviation, minimum and maximum values. The distributions of travel time on individual days are illustrated in Figure 7.

Table 4. Field Travel Time at Site 15

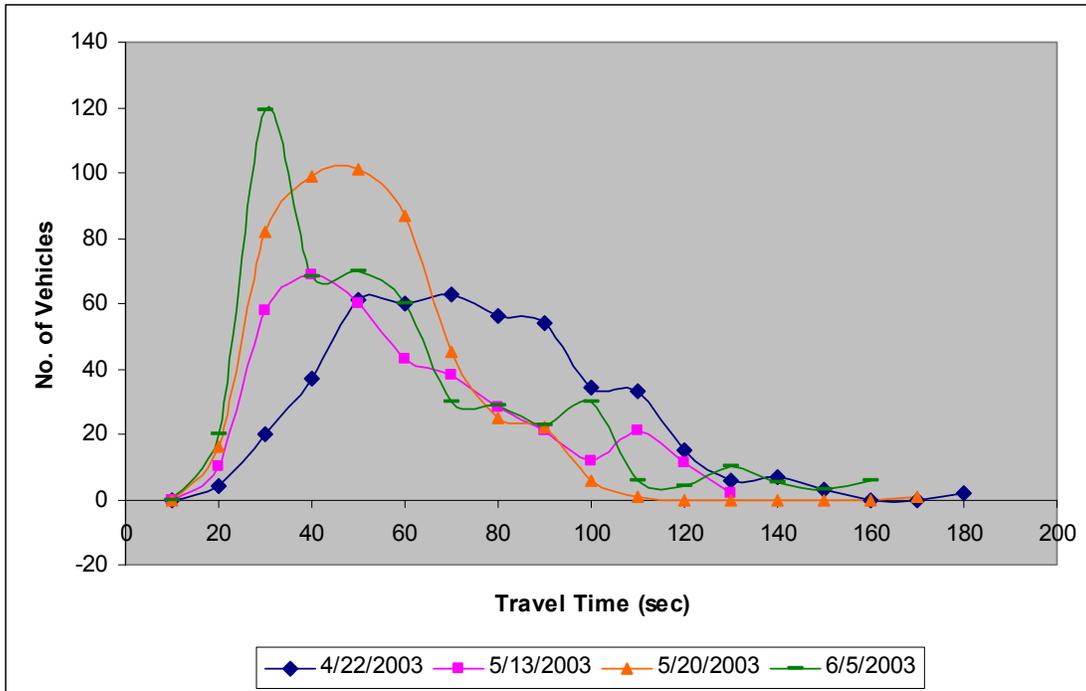| Travel Time (sec)	Mean	Median	Stdev	Min.	Max.
4/22/03	70.43	68.00	27.98	15.00	180.00
5/13/03	53.32	46.00	25.64	16.00	121.00
5/20/03	46.51	44.00	18.91	17.00	165.00
6/5/03	51.53	42.00	30.89	17.00	157.00
Ave.	55.45	50.00	25.86	16.25	155.75

Figure 7. Site 15 Field Travel Time

Data Collection and Reduction at Covington Network

Data Category

To code the second network in the simulation models, several required input data, including the number of lanes, the length of work zone, speed limits, and traffic volume, were collected. The MOEs used for the calibration and validation were travel time from the beginning to the end of the network and travel speed at the merge area. These specific MOEs were selected because of their straightforward collection from both the field and through simulation tools.

Data Collection and Reduction

Before the data collection the research team made two site visits to determine the locations for collecting data and setting up the equipment. To collect the desired traffic data, three video cameras were positioned separately at the beginning of the network, merge area and the end of the network. The distance between the first two cameras was approximately 3 miles, while the distance between the second and the third cameras was approximately 2 miles. To account for day-to-day variability, multiple days of data collection were conducted. A group of five people collected data during the evening peak hour between 5 and 6 P.M. on June 10, 24, 25, and 26, 2003. Three people were responsible for the video cameras, and the other two collected speeds using a LIDAR gun at the merging area. Synchronization between clocks of all the equipment was performed before data collection.

Data reduction through videotapes consisted of obtaining traffic counts, travel times to pass through the network and travel speeds at the merge area. Traffic counts and heavy vehicle counts were obtained through watching the videotape of the upstream camera because the segment does not contain a ramp. However, an on-ramp exists near the beginning of the network, and the traffic from that entrance also was included.

For the Covington network, the vehicle description matching method was used to extract travel times rather than the license plate matching method because the travel speed for this highway was relatively high, making it too difficult to obtain the license plate number for every vehicle. The process was to note the vehicle description and record the times it appeared at the beginning and end of the network. Each vehicle was then matched one by one, and the travel time passing through the network was determined by the time differences. Travel speed collection was relatively easy and time efficient as the software attached to the LIDAR gun automatically recorded the speed and time during the data collection.

Data Summary

Table 5 summarizes traffic counts on 4 days. Data on June 10, 2003, were reserved for validation. The compliance rate was calculated as the percentage of the drivers following the posted speed limits according to the collected speed data on each day. Table 6 shows the statistics of travel times, including mean, standard deviation, and number of records. In contrast to the travel time data of Site 15, the Covington network data show a small variation of mean travel time on different days. As seen in Figure 8, the travel time distributions on different days are quite similar.

Table 5. Field Traffic Volume and Composition for Covington Network

Covington	Traffic Counts (veh)	HV%	Compliance Rate(%)
06/10/03*	601	17	30
06/24/03	744	15	18
06/25/03	700	14	23
06/26/03	851	13	30

*The data were collected between 5:10 to 6:00 pm.

Table 6. Field Travel Time for Covington Network

Covington	Mean Travel Time (sec)	Stdev	Size
06/10/03*	328.21	19.53	574
06/24/03	330.04	20.91	690
06/25/03	332.73	19.26	700
06/26/03	332.19	21.83	790

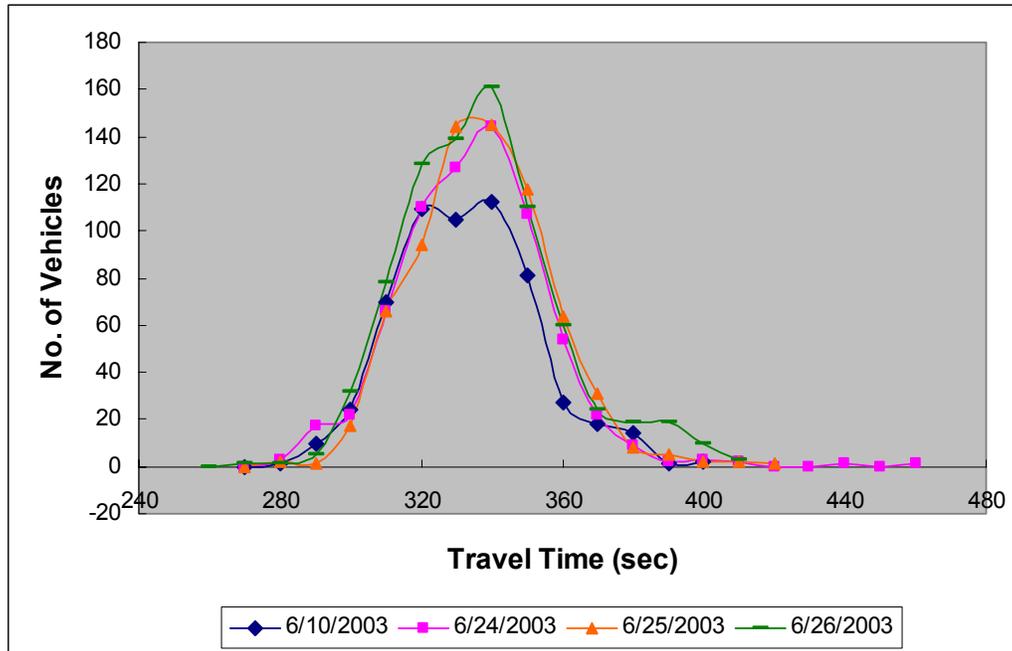*The data were collected between 5:10 to 6:00 pm.

Figure 8. Covington Network Field Travel Time

Table 7 summarizes the statistics of travel speeds, including mean, standard deviation, and number of records. The mean speed also shows small variation over four days. Table 8 lists the statistics of vehicle free flow speed, which were filtered from the raw data of 4 days only when the headway between two continuous vehicles was greater than 4 seconds. The distributions of speed data on different days are shown in Figure 9, and the histogram of free flow speed is shown in Figure 10. For speed data, the distribution of the collected speeds underestimated field conditions as some high speeds from the fast lane were not recorded during data collection.

Table 7. Field Travel Speed for Covington Network

Covington	Mean Speed (mph)	Stdev.	Size
6/10/03	51.54	6.84	234*
6/24/03	52.9	5.65	405
6/25/03	52.33	5.76	634
6/26/03	51.42	6.83	695

* The data were collected between 5:10 to 6:00 and the operator was interrupted by the police for about 10 minutes.

Table 8. Field Free Flow Speed for Covington Network

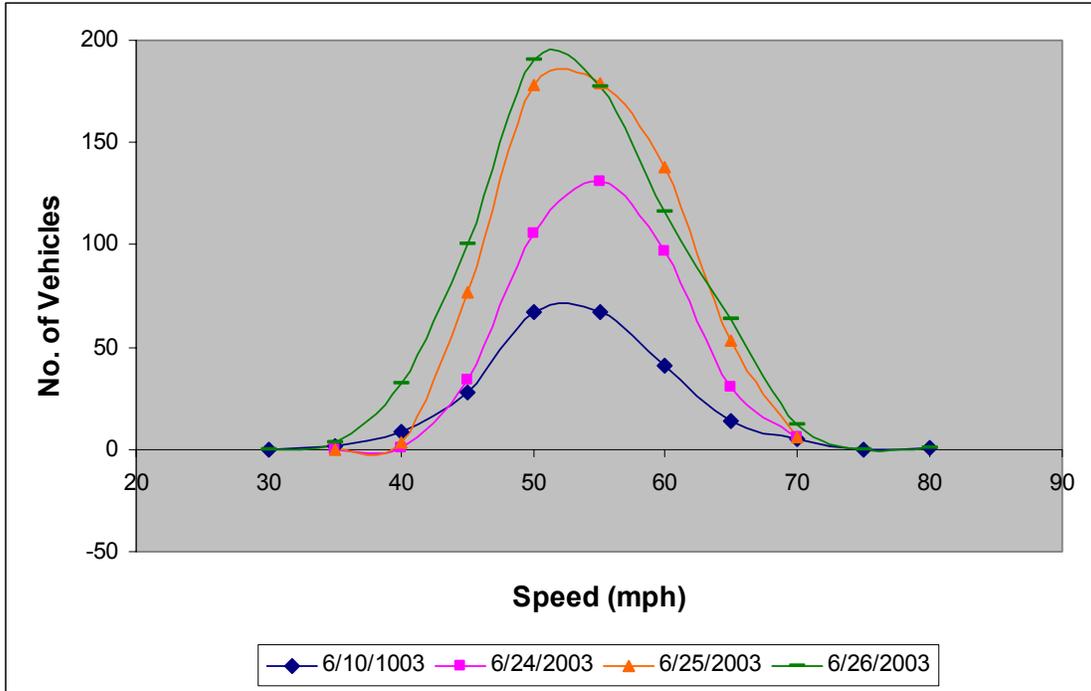| Covington	Mean Free Flow Speed (mph)	Stdev.	Size
6/10/03	51.67	6.71	109
6/24/03	53.30	5.97	240
6/25/03	52.98	5.96	282
6/26/03	52.12	6.60	272

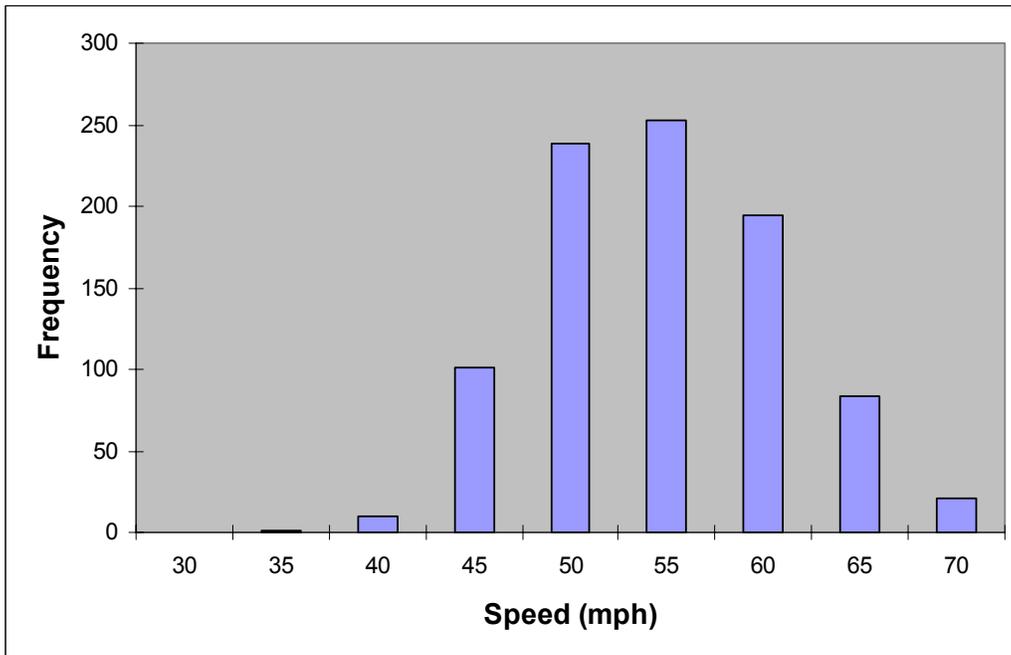**Figure 9. Covington Network Field Speed**

Figure 10. Covington Network Field Free Flow Speed

Network Coding

Site 15 and the Covington network were coded in VISSIM, PARAMICS, and CORSIM using the data collected from the field and VDOT information, such as signal-timing plan. For the Covington network, the aerial photograph was imported into the three models to help network coding. For Site 15, coding of the actuated signal controller was carefully performed in each model to include most features, including gapping out, vehicle recall, vehicle extension, and so forth. However, in some cases they could not function exactly as the field because the capability of modeling actuated signal varied among different simulation models. For example, NEMA graphical user interface (GUI) in VISSIM does not provide functions such as added initial and maximum initial while the actuated signal controller embedded in CORSIM provides a comprehensive range of settings. In PARAMICS, users need to write an API to include all the features, which requires additional time and effort.

Calibration and Validation of the Simulation Models

Site 15 – VISSIM

Identification of Calibration Parameters

In the initial phase users may not be able to correctly identify all the critical parameters or the reasonable range for each parameter. Two criteria could help users select the parameter candidates to be calibrated:

1. Conduct trial and error test for each parameter. If the parameter does not affect the simulation result at all, it can be excluded from the parameter candidate list;
2. Judge from a traffic engineer's point of view. For instance, in this study network, there was only one lane for each approach and Origin-Destination (OD) was simple without any diversion possibility. Therefore, some parameters in VISSIM related to lane change or route change, such as minimum headway (front/rear) and waiting time before diffusion, should not have any impact on the simulation result and hence, were eliminated from the parameter list.

Regarding the acceptable range for each selected parameter, users could determine the range based on literature review, including manuals of simulation model and/or other existing research findings. In order not to mislead the calibration direction, some unrealistic value should be avoided in the range. The feasibility of selected parameters and their ranges was verified later through a feasibility test.

The following is the list of parameters and acceptable ranges determined initially. The detailed description of each parameter can be found in the VISSIM manual (PTV, 2001).

1. Simulation resolution (Time steps/Sim. sec.): 1 – 9
2. Number of observed preceding vehicles: 1 – 4
3. Average standstill distance (meter): 1 – 5

4. Saturation flow rate (1756, 1800, 1846, 1895, and 1946 veh/h)
   - Additive part of desired safety distance: 2.0 – 2.5
   - Multiple part of desired safety distance: 3.0 – 3.7
5. Priority rules – minimum headway (meter): 5 – 20
6. Priority rules – minimum gap time (second): 3 – 6
7. Desired speed distribution (mph): 40-50, 30-40, 20-30

Note that the saturation flow rate values were based on the table (Flow rate, Throughs, 25 second green) in the VISSIM Manual 5.2.6.

Experimental Design for Calibration

The Latin Hypercube Sampling toolbox in MatLab was used to generate 200 scenarios using the first set of parameters and their ranges defined above.

Multiple Runs

Five random seeded runs were conducted in VISSIM for each of the 200 cases, for a total of 1000 runs. The average travel time was recorded for each of the 1000 runs. The results from the five multiple runs were then averaged to represent each of the 200 parameter sets.

Feasibility Test

With the simulation results based on the initial parameter set, it is important to answer the following key questions by applying statistical methods and analyzing plots.

- Test if the simulation results based on the current parameters and their ranges include the field conditions. If not, some parameter ranges need to be adjusted or new factors should be explored. Histogram analysis is appropriate for this purpose.

- Figure 11 shows that the average field travel time, 56.75 sec, falls outside of the simulated distribution generated from initial ranges of parameters. It indicates that current parameters and their ranges were not satisfactory and needed adjustment.

- Test if the simulation result is sensitive to the different values of each parameter and which parameters are critical to the simulation results. Plots and ANOVA are appropriate here.

Scatter plots of each parameter versus travel time from simulations were plotted. As seen in these figures, minimum gap and desired speed distribution were two important parameters to the results. Travel time became higher when minimum gap increased and/or mean desired speed decreased. Two example scatter plots are shown in Figures 12 and 13. Mean desired speed demonstrates an apparent trend, while simulation resolution has no effect on simulation result. However, it is more rigorous to identify the critical parameters through statistical analysis. The values of each parameter were made discrete, separated into several groups and indexed. Then one-way ANOVA was used to test the null hypothesis that the means of two or more groups were not significantly different. If the parameter was sensitive to the result, the means of

31

different groups should be statistically different. Table 9 shows ANOVA results listing the significance value of the F test, the mean difference between groups, and sum of squares (SSR) between groups. Small p value and big SSR indicate that the null hypothesis is rejected and that this parameter is important to the result. The group pairs were listed if the mean difference was significant at the 0.5 level and 0.05 level.
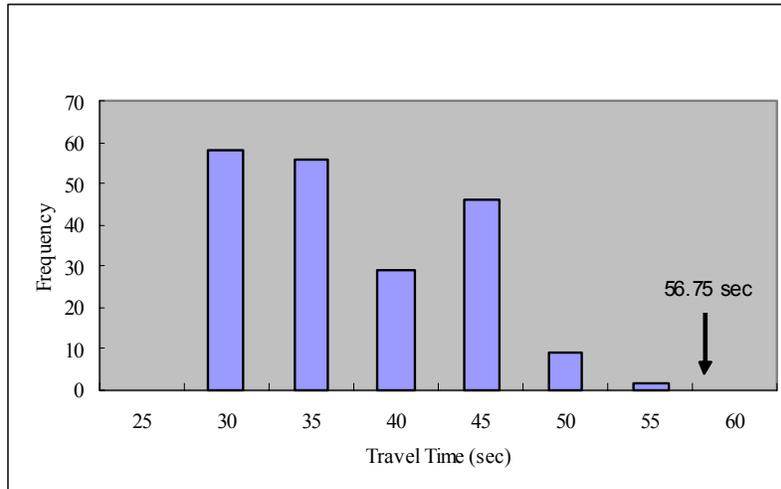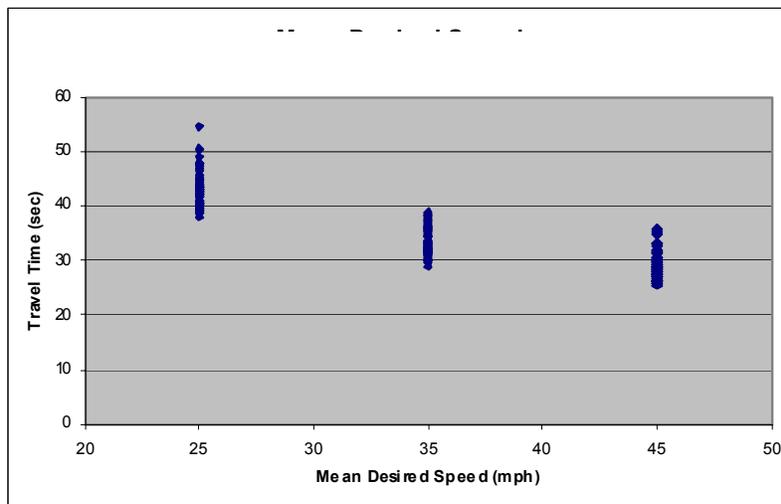

Figure 11. Feasibility Test Results for Site 15 with VISSIM

Figure 12. Mean Desired Speed versus VISSIM Travel Time
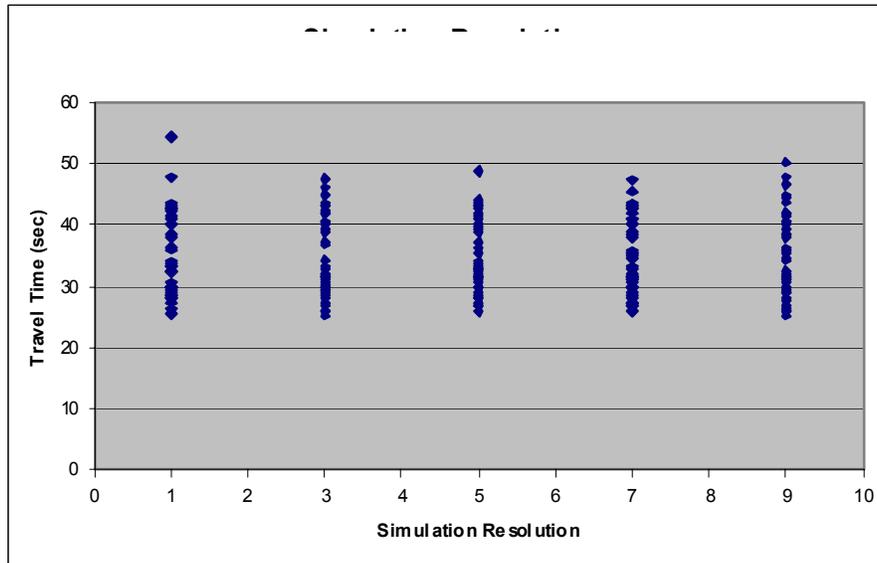
Figure 13. Simulation Resolution versus VISSIM Travel Time

Table 9. ANOVA Results for Site 15 with VISSIM

Site15 – VISSIM	Significance value (p value)	Significant mean differences (Sig. < 0.5)	SSR between groups
Simulation Resolution	0.989	NA	12.8
No. of Preceding Vehicles	0.336	1-3, 1-4	137.0
Ave. Standstill Distance	0.369	1-3	80.8
Saturation Flow	0.910	NA	40.9
Min. Headway (meter)	0.220	1-4, 2-4	178.2
Min. Gap Time (sec)	0.000	1-3, 1-4*, 2-4*, 3-4	834.8
Desired Speed Distribution	0.000	All	6655.9

* Significance value is less than 0.05.

Based on the ANOVA table, the desired speed distribution and the minimum gap time were identified as critical parameters and considered expansion from current ranges. Desired speed distribution is an important parameter that has a significant influence on roadway capacity and achievable speeds. If not hindered by other vehicles, a driver will travel at his desired speed with a small stochastic variation. Minimum gap is an important parameter to model conflicting traffic movements and determine when the movement with lower priority could proceed. To achieve higher simulation travel time, the mean desired speed should be decreased or the minimum gap should be increased. However, the upper bound of the minimum gap already had reached the possible maximum in this case. There was a possibility that some other factors that might impact the result were not yet identified. Therefore, additional tools such as HCM or field data were employed to help clarify the ranges of some parameters, such as saturation flow on the southbound approach and field speed within the intersection. Since the simulation travel time tends to be shorter than the field travel time, the saturation flow in VISSIM or the mean of desired speed distribution could possibly have been higher than those of true field values.

1. Check speed within the intersection.

Although the posted speed limit of 45 mph was observed in the field, the actual speed within the intersection could be lower than 45 mph due to poor sight distance, permitted left-turns, narrow intersection, and other factors such as gas stations located around 500 ft upstream of the stop line on the southbound approach. It is more reasonable to use the field average speed to set the link speed and turning speed. Therefore by viewing the videotapes, the speed data were retrieved as the tail of the vehicle crossed the stop bar on the southbound approach. The first few vehicles in queue were not considered because they were still accelerating. Figure 14 shows the histogram of field speed within intersection. This confirms that lower desired speed distributions should be used in VISSIM. The modified ranges were: 30-40 mph, 32.5-37.5 mph, and 27.5-42.5 mph.
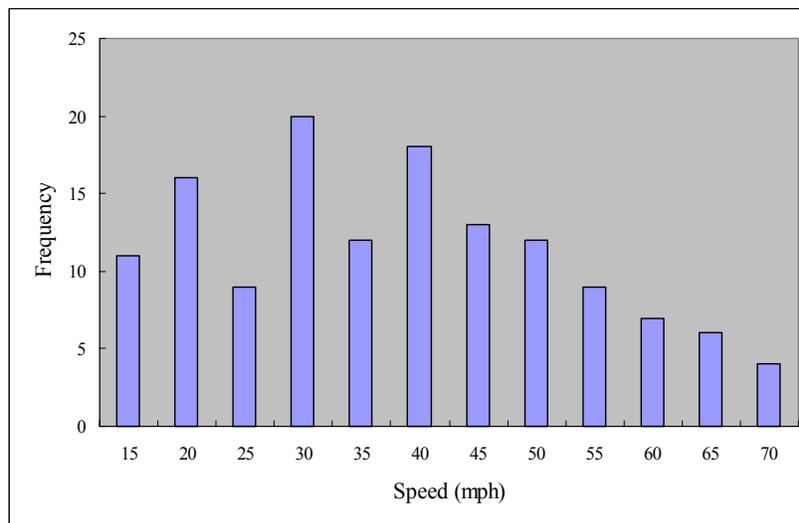

Figure 14. Field Speed within Intersection

2. Check saturation flow.

Saturation flow rate is an important factor to determine the intersection capability, which in turn affects the vehicle travel time. It can be obtained from field data, HCM procedure, and VISSIM simulation. The following shows how saturation flow rate was collected or calculated using these approaches:

- *Calculation of queue discharge headway from the field data*

The field saturation flow rate on the southbound approach was obtained based on the videotape on June 5. First, the time stamps of the third and the last vehicle in the queue were recorded as they passed the stop line in each cycle. Then the average discharge headway and the saturation flow rate were calculated using Equation (2). The average headway was 3.13 sec, which resulted in a saturation flow of 1149 vehicle/hour/lane. Generally, field data showed that the southbound approach has a relatively low saturation flow rate, which accounts for the longer travel time in the previous simulations.

$$Saturation\ Flow = \frac{3600}{(t_3 - t_n)/(n-3)} \qquad (2)$$

- *Calculation using HCM procedure*

A saturation flow rate can be computed using HCM procedure. The ideal saturation flow rate, which is usually 1900 vphpl, can be adjusted for the prevailing conditions to obtain the saturation flow for the lane group being considered. The adjustment is made by introducing factors that correct for number of lanes, lane width, heavy vehicle percentage, right and left turns, and so forth. The adjusted saturation flow rates were computed and summarized in Table 10. The result also confirms that the southbound approach has a low saturation flow.

Table 10. Saturation Flow Rate Using HCM Procedure

6/5	EB			WB			NB			SB		
	L	T	R	L	T	R	L	T	R	L	T	R
Volume	88	98	44	45	77	84	32	172	29	192	300	93
Proportion of LT or RT	0.4		0.2	0.2		0.4	0.1			0.1	0.33		0.2
Baisc saturation flow		1900			1900			1900			1900	
Number of lanes		1			1			1			1	
Land width adjustment		1			1			1			1	
Heavy vehicle adjustment		0.962			0.98			0.943			0.952	
Grade adjustment		1			1			1			1	
Parking adjustment		1			1			1			1	
Bus blockage adjustment		1			1			1			1	
Area type adjustment		1			1			1			1	
Lane utilization adjustment		1			1			1			1	
Left-turn adjustment factor		0.791			0.807			0.787			0.726	
Right-turn adjustment factor		1			1			1			1	
Adjusted saturation flow		1445.79			1502.634			1410.068			1313.189	

- *Calculation of queue discharge headway from VISSIM animation*

In VISSIM, the saturation flow rate is determined by two parameters: additive part of desired safety distance (bx_add) and multiple part of desired safety distance (bx_mult). The VISSIM manual provides four tables to roughly estimate saturation flow rate based on these two parameters under certain signal timings. However, the relationship between these two parameters and saturation flow rate in VISSIM is not clear. An accurate way is to record vehicle discharge headways in VISSIM and calculate saturation flow rate. Because of the low saturation flow rate obtained from field data and HCM procedure, it was envisioned that the current ranges for bx_add and bx_mult were not appropriate. Therefore, the ranges of these two parameters were expanded. The new ranges of bx_add and bx_mult were 1.0-5.0 and 1.0-6.0, respectively.

To test the capability of VISSIM to achieve the desired saturation flow rate, two cases with simulation results close to the field data were selected. The VISSIM animations were watched, and queue discharge headways were recorded for each cycle. Table 11 shows the observed queue discharge headways and saturation flow rates in VISSIM. The result shows that VISSIM can achieve comparable saturation flow rate to the field condition.

Table 11. Comparison of Saturation Flow

Comparison	Queue Discharge Headway		Saturation Flow (veh/h/l)	Travel Time (sec)
	Ave. (sec)	Stdev (sec)		
Field	3.13	0.82	1149	55.45
VISSIM (Case 1)	2.96	0.69	1216	53.04
VISSIM (Case2)	3.03	0.83	1190	54.79

Based on the speed and saturation flow rate conditions, the parameter set is modified as follows.

1. Simulation resolution (Time steps/Sim. Sec): 1 – 9
2. Number of observed preceding vehicles: 1 – 4
3. Maximum look ahead distance (meter): 200 – 300
4. Average standstill distance (meter): 1 – 5
5. Saturation flow rate
   - Additive part of desired safety distance: 1.0 – 5.0
   - Multiple part of desired safety distance: 1.0 – 6.0
6. Priority rules – minimum headway (meter): 5 – 20
7. Priority rules – minimum gap time (second): 3 – 6
8. Desired speed distribution (mph): 30-40, 32.5-37.5, 27.5-42.5

With the new parameters, another 200 cases were generated using the LHD method. The new simulated distribution is shown in Figure 15, which covers the field data and indicates that the new parameter ranges were able to capture the field condition. The parameter set with best performance was selected as LHD-based parameters in the following evaluation.
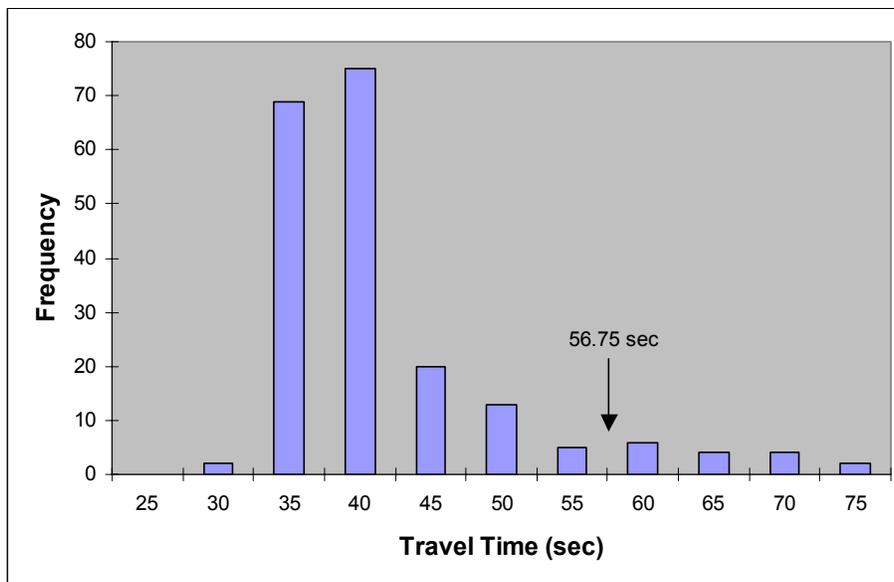
Figure 15. New Feasibility Test Results for Site 15 with VISSIM

Parameter Calibration Using Genetic Algorithm

A genetic algorithm was integrated with the VISSIM model to calibrate parameters. Figure 16 shows the convergence of the GA. The results were based on seven generations of GA evolution and a population size of 20 in each generation. Both best fitness value and average fitness value decrease as the number of generation increases, which indicates the simulation result more closely approaches the field value. With these parameters, VISSIM was multiple run 100 times, and average travel time was recorded for each run. The resulting distribution of travel time is shown in Figure 17. Field travel times from three days are also shown in the figure, and they all fall within the simulated distribution.

To test the consistency of results from GA optimization, the GA process was repeated two more times with 10 generation and 20 populations. Naturally increasing the number of generations and/or populations would help to reach the better fitness value, but doing so requires more time. The convergence results are shown in Figures 18 and 19. The parameter set with the best fitness value was selected to represent the local traffic characteristics.
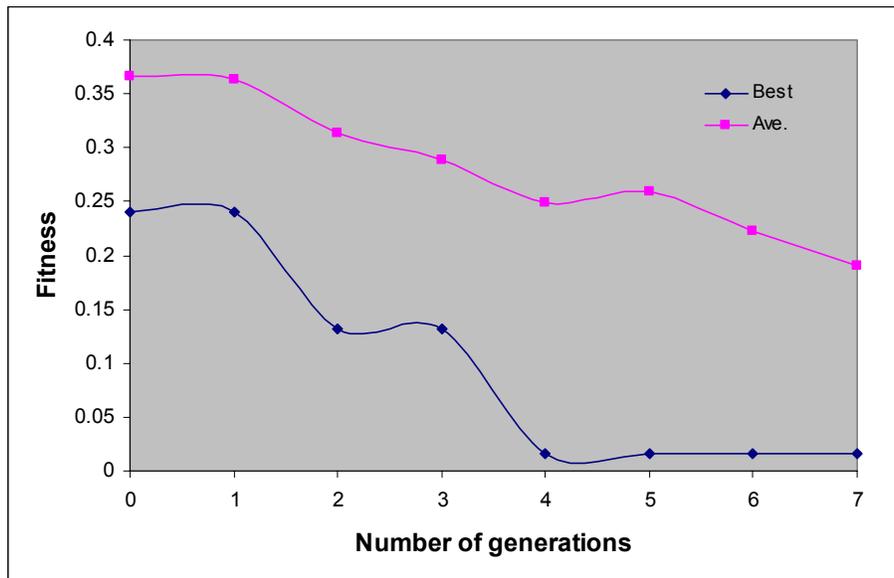
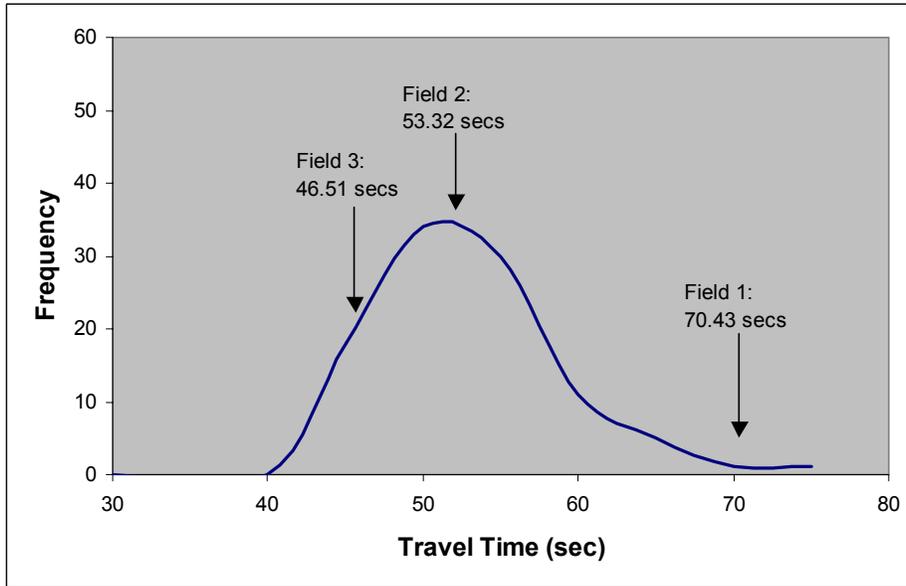**Figure 16. Convergence of Fitness Value with Generation (VISSIM, 1st Trial)**

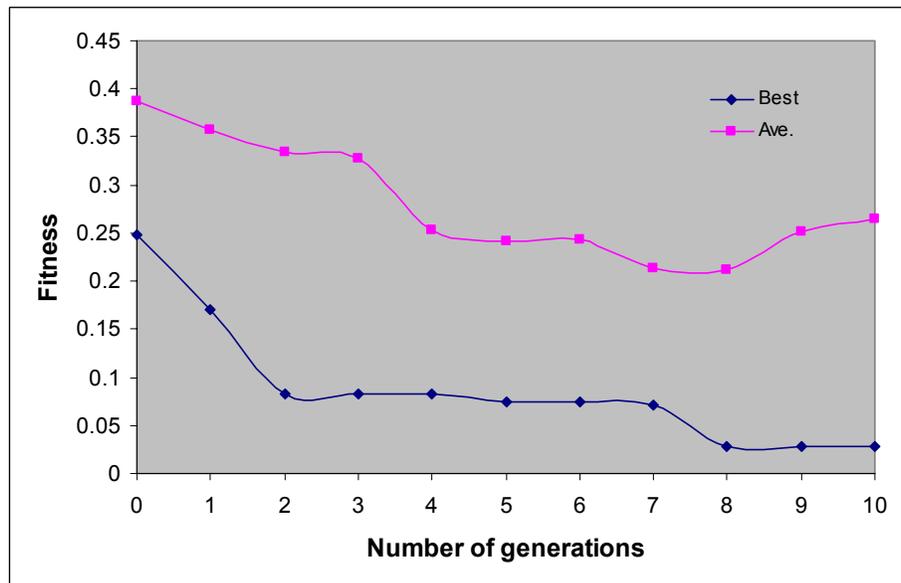**Figure 17. Travel Time Distribution of GA-based Parameter Set**

Figure 18. Convergence of Fitness Value with Generation (VISSIM, 2nd Trial)
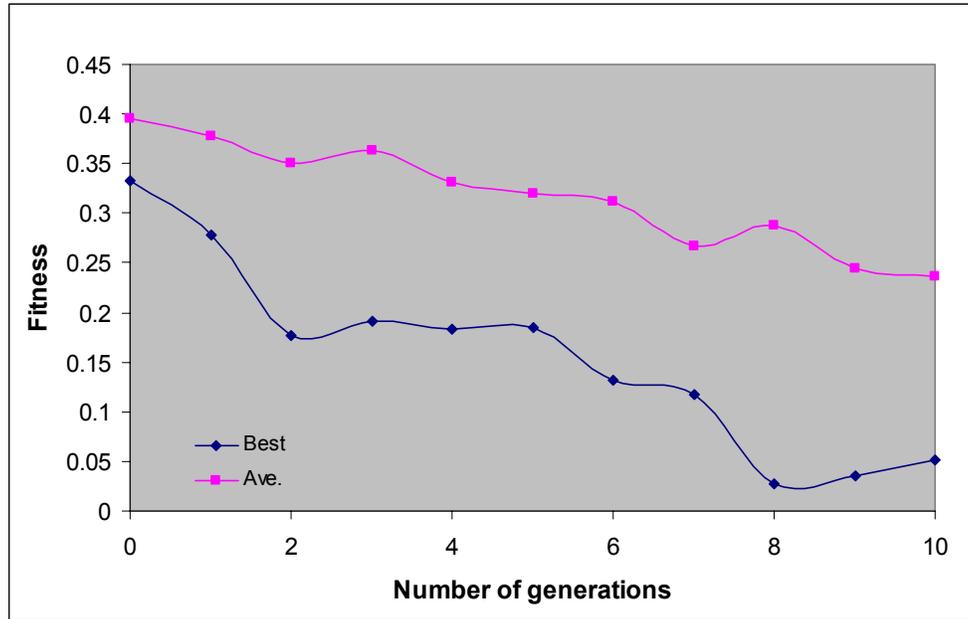
Figure 19. Convergence of Fitness Value with Generation (VISSIM, 3rd Trial)

Evaluation of the Parameter Sets

This section presents the comparison of 100 VISSIM simulation results based on default parameters, best-guessed parameters, LHD-based parameters, and GA-based parameters. The parameter values for each set were listed in Table 12.

Table 12. Four Parameter Sets for Site 15 with VISSIM

Site15 - VISSIM	Default	Best-Guessed	LHD based	GA based
Simulation resolution	5	10	2	6
Number of observed preceding vehicles	2	4	3	4
Maximum look ahead distance (meter)	250.00	300.00	297.80	215.15
Average standstill distance (meter)	2.00	2.00	4.80	3.85
Additive part of desired safety distance	3.00	5.00	4.47	5.0
Multiple part of desired safety distance	3.00	2.85	4.41	5.3
Priority rules – minimum headway (meter)	5.0	15.0	20.0	20.0
Priority rules – minimum gap time (second)	3.0	3.5	5.0	4.0
Desired speed distribution (mph)	Car: 40-50 HV: 30-40	Car: 40-50 HV: 30-40	Car: 27.5-42.5 HV: 15.5-18.6	Car: 27.5-42.5 HV: 15.5-18.6
Ave. travel time (second)	23.40	29.53	47.44	50.33

Evaluation of default parameters was performed by making multiple VISSIM runs without changing any parameter value. Users can also guess parameter values based on their experience with the study network and with the help of analytical tools such as HCM. For instance, in this case study, additional information such as saturation flow could be obtained from HCM procedure. According to the tables provided in the VISSIM manual, a linear regression model was derived with R square of 0.92 to estimate saturation flow rate:

$$Saturation\ flow\ rate = 2469.105 - 172.93*BX\_ADD - 65.51*BX\_MULT \qquad (3)$$

With the Solver function in MS Excel, an optimal combination of these two parameters was obtained to minimize the difference between the HCM and VISSIM saturation flow. The values for bx_add and bx_mult were 5 and 2.85 meters, respectively.

A comparison of the uncalibrated VISSIM models (default parameters and best-guessed parameters) and the calibrated VISSIM models (LHD-based and GA-based parameters) shows the importance of calibration for microscopic simulation models. Travel times along the southbound approach are compared in Figure 20. As seen in the figure, three average field travel times all fall within the distributions of simulation results using the calibrated models. The uncalibrated models generate much shorter travel times than observed in the field. A t test was conducted to compare the simulation results using GA-based parameter set with the other three parameter sets. The null hypothesis is that the means of the two samples are equal. The result shows that the calibrated parameter set using GA generated statistically significant results from all the other parameter sets. Animations of each parameter set were viewed in order to determine whether the animations were realistic or unrealistic. For the calibrated parameters, the animations at several travel time percentiles of the distribution were found acceptable. For default parameters, almost all vehicles passed the intersection without waiting, which was not realistic.

Validation

The "calibrated" models were then evaluated with a new set of field data under untried conditions including the input volumes and traffic composition. The MOE was again average travel time on the southbound approach, but on a different day. It was used for evaluation of the four parameter sets in Table 12. The comparison result is shown in Figure 21. Field average travel time, 51.53 sec, was approximately the mean of the simulated distributions of the calibrated models. Similarly, the uncalibrated models generated shorter travel times. The t test result confirms that the simulated distribution using GA-based parameter set is statistically significant from those using other three parameter sets.
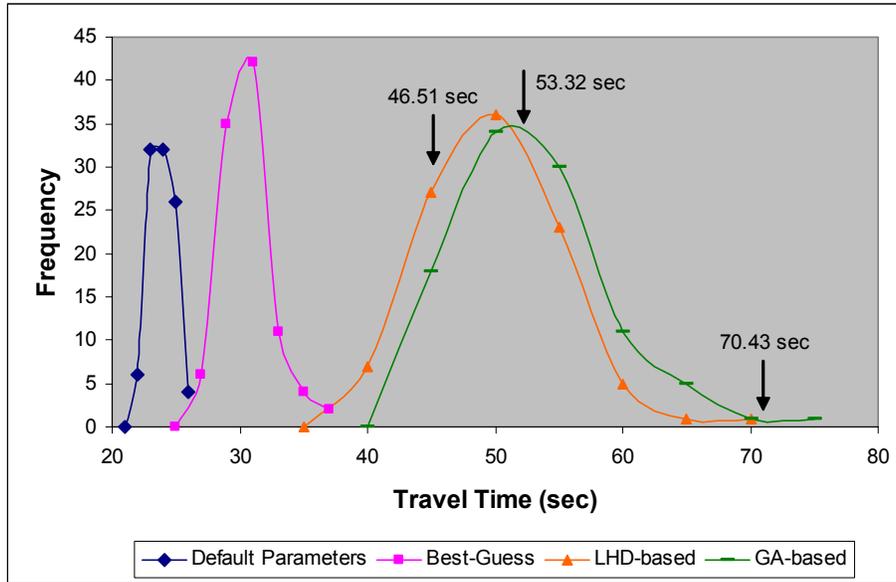
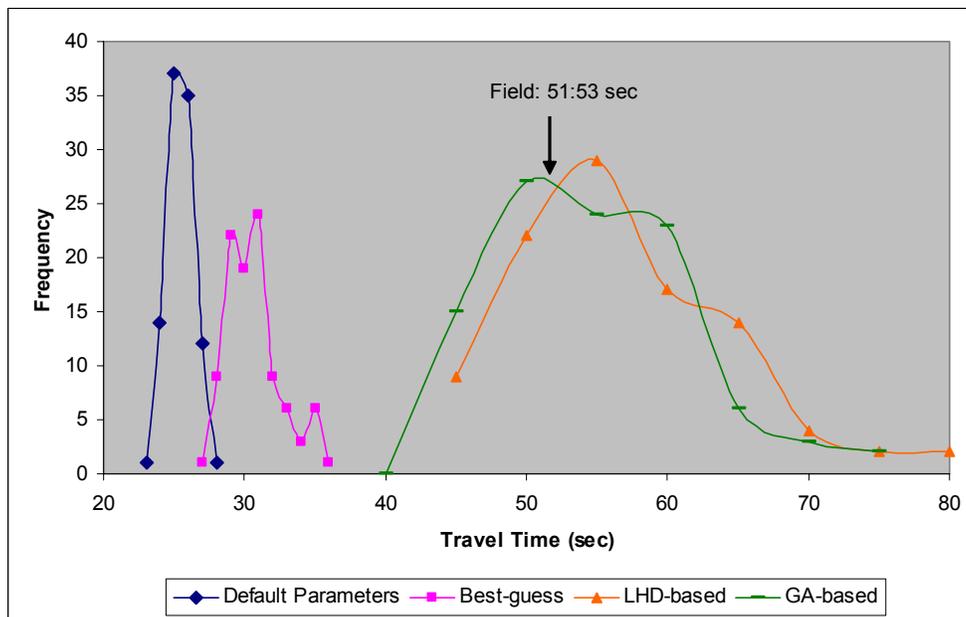**Figure 20. Comparison of Site 15 Travel Time with VISSIM**

Figure 21. Validation of VISSIM Using Site 15

Site 15 – PARAMICS

Identification of Calibration Parameters

The following is the first set of parameters and acceptable ranges used in the PARAMICS calibration process. Acceptable ranges were based on the researchers discretion and their

familiarity with PARAMICS. The detailed description of each parameter can be found in the PARAMICS manual (Quadstone, 2002).

1. Mean headway (sec): 0.6 – 2.2
2. Mean reaction time (sec): 0.3 – 1.9
3. Speed memory: 1-9
4. Curve speed factor: 1.0-5.0
5. Visibility (ft): 0 – 45
6. Headway factor: 0.6 – 1.4
7. Left turn speed (mph): 9 – 15
8. Link speed (mph): 30-40

Experimental Design for Calibration

The Latin hypercube design consisted of 200 cases using the first set of parameters and the values within the acceptable ranges.

Multiple Runs

Ten random seeded runs were conducted in PARAMICS for each of the 200 cases, for a total of 2000 runs. The increased number of runs is due to the fast PARAMICS simulation. The average travel time was recorded for each of the 2000 runs and the results from the 10 multiple runs were then averaged t represent of each of the 200 parameter sets.

Feasibility Test

The objective of this section is to check whether the simulated results based on the current parameters could capture the field data and identify the key parameters in the PARAMICS model. The travel time histogram of 200 cases is shown in Figure 22, which confirms that the results cover all three field data. It indicates that the current parameters and their ranges are able to generate the realistic results. Scatter plots of each parameter versus travel time from the simulation were then plotted to help identify key parameters.
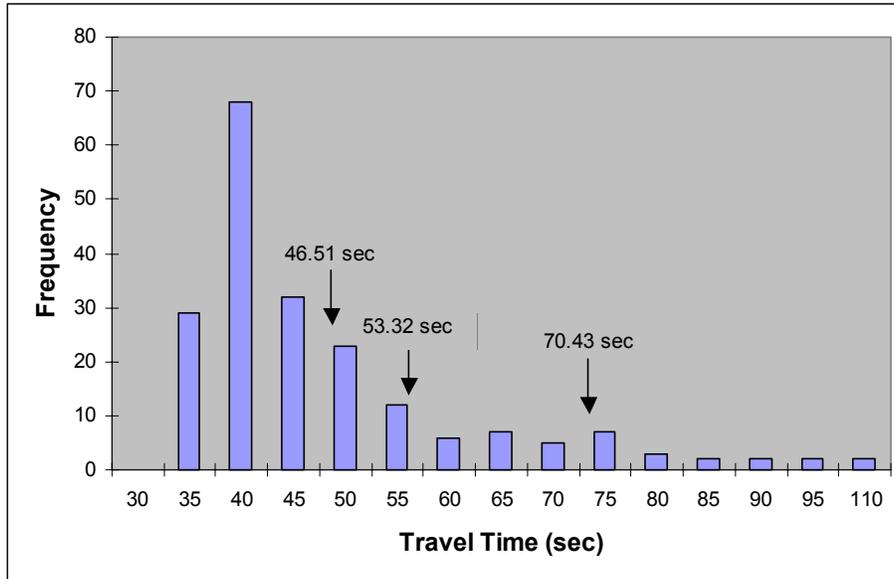
Figure 22. Feasibility Test Results for Site 15 with PARAMICS

Table 13. ANOVA Results for Site 15 with PARAMICS

Site15 – PARAMICS	Significance value (p value)	Significant mean differences (Sig. < 0.5)	SSR between groups
Mean Headway (sec)	0.000	1-2, 1-3*, 1-4*, 1-3*, 2-4*, 3-4	7341.2
Mean Reaction Time (sec)	0.000	1-3*, 1-4*, 2-3*, 2-4*, 3-4*	14827.3
Speed Memory	0.889	NA	231.5
Curve Speed Factor	0.400	3-4	598.5
Visibility (ft)	0.010	1-2*, 1-3*	1845.5
Headway Factor	0.100	1-2, 1-4	1259.5
Left Turn Speed (mph)	0.122	1-2, 2-3	850.2
Link Speed (mph)	0.000	7-8*, 7-9*, 8-9	4992.4

* Significant value is less than 0.05

Parameter Calibration Using Genetic Algorithm

A genetic algorithm was integrated with the PARAMICS model to find the optimal parameter values. The fitness value was calculated using Equation (1). The simulation travel time was the value averaged from ten PARAMICS runs per parameter set. To test the consistency of results from GA calibration, the GA process was repeated two more times with 20 generations and 20 populations each. The convergence results are shown in Figures 23 and 24. Huge fluctuations were observed in both best fitness value and average fitness value over generations due to the high variability of PARAMICS simulations. The parameter set with the best fitness value was selected to represent the local traffic conditions. Then PARAMICS was multiple run 100 times with this calibrated parameter set and average travel time was recorded for each run. Figure 25 shows the distribution of travel time. It covers all the field data.
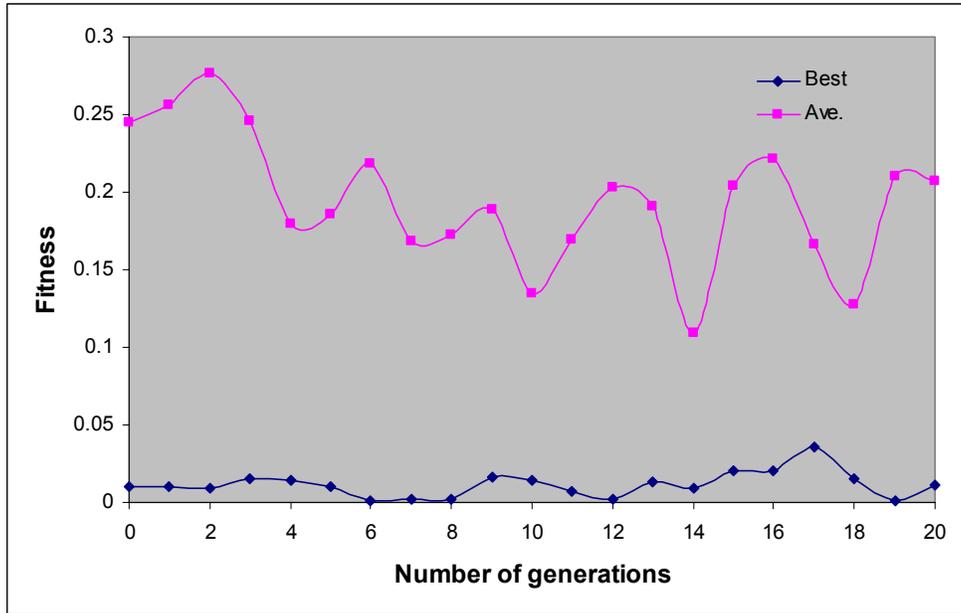
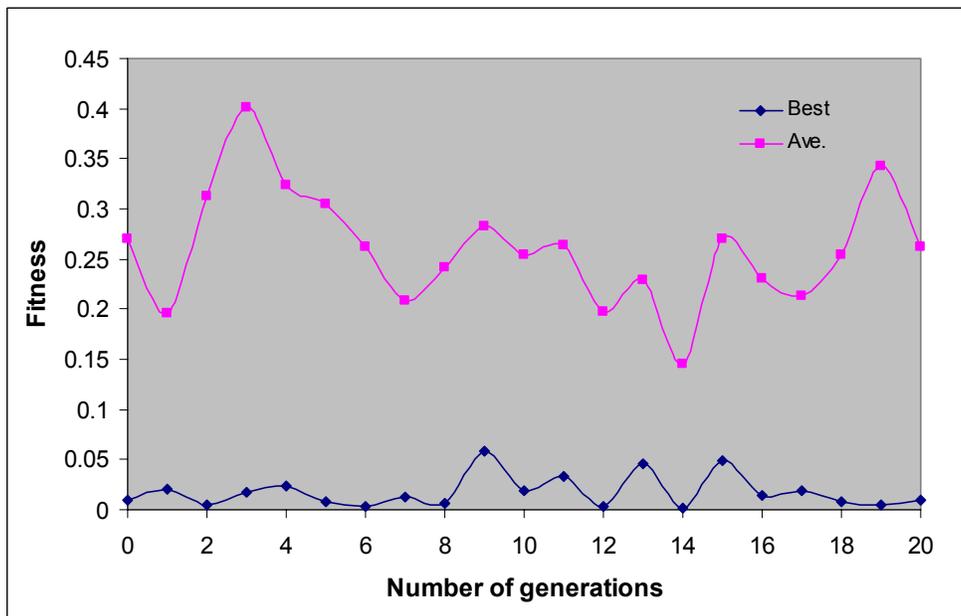**Figure 23. Convergence of Fitness Value with Generation (PARAMICS, 1st Trial)**

Figure 24. Convergence of Fitness Value with Generation (PARAMICS, 2nd Trial)
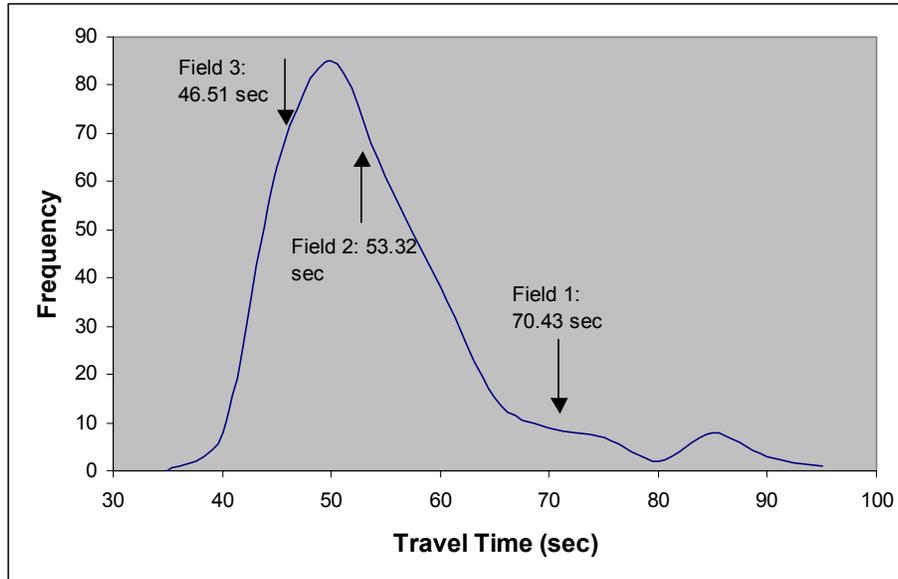
Figure 25. Travel Time Distribution of GA-based Parameter Set

Evaluation of the Parameter Sets

This section presents the evaluation results of simulation parameters, including default parameters, best-guessed parameters, LHD-based parameters and GA-based parameters. These parameter values are summarized in Table 14. Best-guessed parameter values in PARAMICS were given mainly based on the engineering judgment and observation of local traffic conditions by the researchers. LHD-based parameter set was the best parameter set that generated closest travel time to the field value.

Travel times from simulations based on these four parameters are compared in Figure 26. Different from VISSIM result, all the parameter sets except for the default parameters generated the comparable travel time distributions that capture the field data. The finding was consistent with the t test result. GA-based parameter set generated statistically significant results only from default parameter set while had similar performance as best-guessed parameter set, with a p value of 0.819, and LHD-based parameter, with a p value of 0.189. Animations of the model with calibrated parameters were viewed and regarded acceptable.

Table 14. Four Parameter Sets for Site 15 with PARAMICS

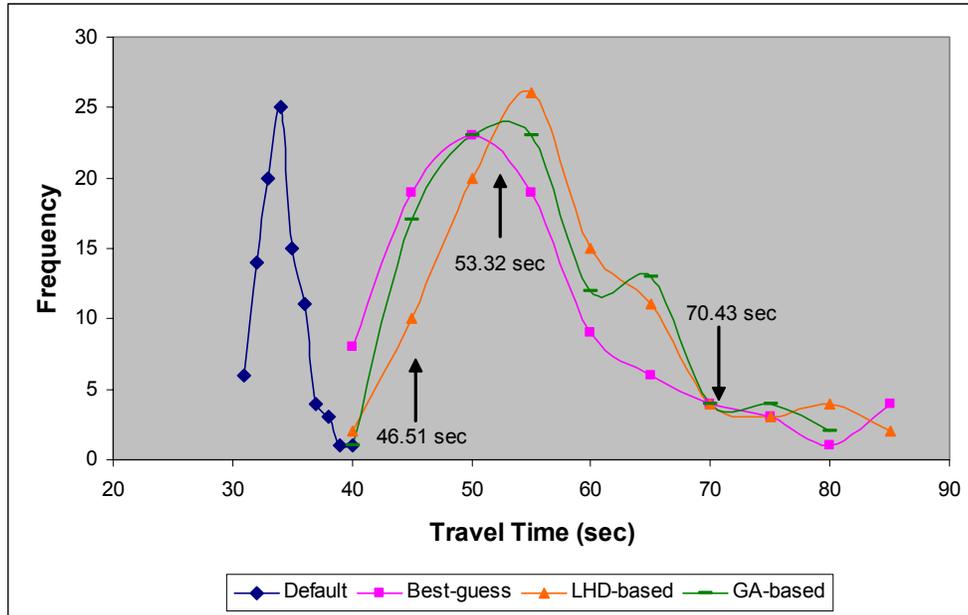| Site15 - PARAMICS	Default	Best-guessed	LHD	GA
Mean Headway (sec)	1.0	1.5	1.5	1.39
Mean Reaction Time (sec)	1.0	1.5	1.5	1.35
Speed Memory	3	3	3	8
Curve Speed Factor	1.0	1.5	2.9	1.44
Visibility (ft)	0.0	30	55	26.4
Headway Factor	1.0	1.0	0.7	1.3
Left Turn Speed (mph)	14	12	11	11
Link Speed (mph)	45	42	37	34
Ave. Travel Time (sec)	33.58	53.71	56.29	54.13

Figure 26. Comparison of Site 15 Travel Time with PARAMICS

Validation

A new set of field data collected on June 5, 2003, was used for evaluation of the calibrated PARAMICS model. Figure 27 shows the comparison result of simulation performance based on the four parameter sets in Table 14. As seen in Figure 27, the calibrated models were able to produce realistic traffic conditions. The result of t test shows that GA-based parameters generate similar performance with LHD-based parameters, with a p value of 0.593, but statistically different from the other two parameter sets.
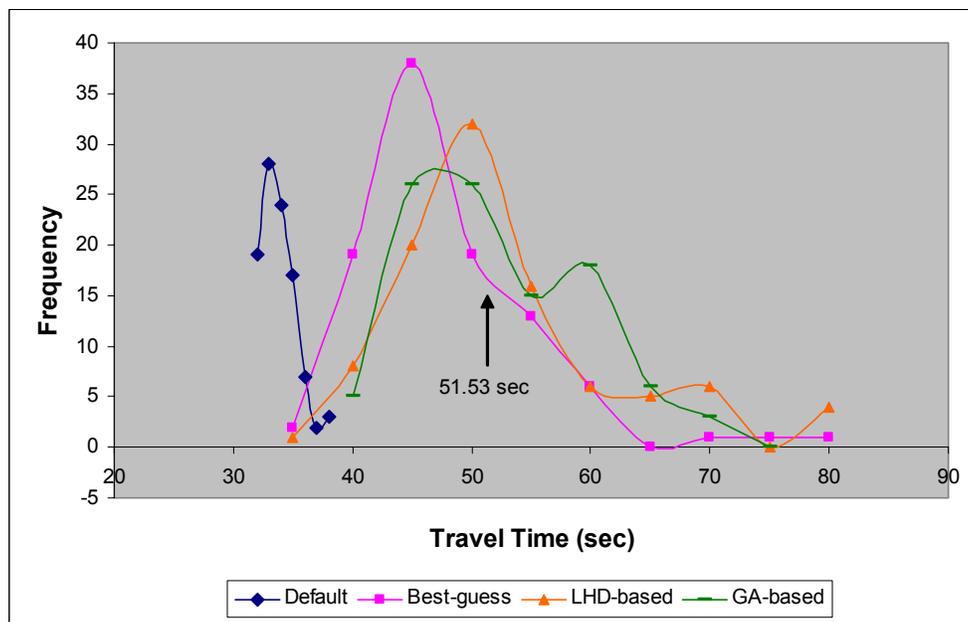

Figure 27. Validation of PARAMICS Using Site 15

Site 15 – CORSIM

Identification of Calibration Parameters

CORSIM includes numerous calibration parameters that allow users to fine-tune the model to replicate observed conditions. In addition to the parameters with acceptable ranges that are similar to those in the other two models, some parameters in CORSIM are represented by a discrete distribution (e.g. amber interval response) or by 10 percentile values (e.g. distribution of free flow speed by driver type) indexed by 10 driver types from conservative to aggressive. Furthermore, the sum of the 10 percentile values must equal 1,000. In order to increase aggressiveness or conservativeness for a population of drivers, the distribution alternatives of amber interval response, gap distributions for left turns, gap distribution for right turns, distribution of free flow speed by driver type, start-up lost time distribution, and discharge headway distribution are designed as shown below, along with the other parameters and their acceptable ranges.

1) Link mean free flow speed (mph): 35 – 45
2) Mean queue discharge headway (1/10 sec): 15 – 30
3) Mean start-up lost time (1/10 sec): 15 – 30
4) Left turn jumper probability (%): 10 – 40
5) Left turn speed (mph): 9 – 21
6) Right turn speed (mph): 9 – 21
7) Left-turn lagging within 2 seconds (%): 20 – 50
8) Left-turn lagging for 2-4 seconds (%): 5 – 15
9) Amber interval response (fpss)
   Default: 21 18 15 12 9 7 6 5 4 4
   Shift to left: 19 16 13 10 7 5 4 3 2 2
   Shift to right1: 23 20 17 14 11 9 8 7 6 6
   Shift to right2: 25 22 19 16 13 11 10 9 8 8
10) Gap distribution for left turns (sec)
    Default: 7.8 6.6 6.0 5.4 4.8 4.5 4.2 3.9 3.6 2.7
    Shift to left: 6.8 5.6 5.0 4.4 3.8 3.5 3.2 2.9 2.6 1.7
    Shift to right: 8.8 7.6 7.0 6.4 5.8 5.5 5.2 4.9 4.6 3.7
11) Gap distribution for right turns (sec)
    Default: 10 8.8 8.0 7.2 6.4 6.0 5.6 5.2 4.8 3.6
    Shift to left1: 9 7.8 7.0 6.2 5.4 5.0 4.6 4.2 3.8 2.6
    Shift to left2: 8 6.8 6.0 5.2 4.4 4.0 3.6 3.2 2.8 1.6
12) Distribution of free flow speed by driver type (%)
    Default: 75 81 91 94 97 100 107 111 117 127
    Narrow (0.8): 82 86 94 96 98 100 105 108 112 119
    Wider (1.2): 73 80 91 94 97 100 107 112 118 128
13) Start-up lost time distribution (%)
    Default: 218 140 125 118 102 86 78 63 47 23
    Narrow (0.8): 195 132 120 115 102 89 82 70 57 38
    Wider (1.2): 240 147 130 121 102 83 74 56 37 10
14) Discharge headway distribution (%)

47

Default: 170 120 120 110 100 100 90 70 70 50
Narrow (0.8): 156 116 116 108 100 100 92 76 76 60
Wider (1.2): 184 124 124 112 100 100 88 64 64 40

Experimental Design for Calibration

The Latin hypercube design consisted of 200 cases using the initial set of parameters and the values within the acceptable ranges.

Multiple Runs

Ten random seeded runs were conducted in CORSIM for each of the 200 cases, for a total of 2000 runs. The average travel time was recorded for each of the 2000 runs, and the results from the 10 multiple runs were then averaged to represent each of the 200 parameter sets.

Feasibility Test

A feasibility test was conducted to check whether the simulated results based on the current parameters could capture the field data and to identify the key parameters in CORSIM. The travel time histogram of 200 cases is shown in Figure 28. The field data fall within the top 5 percent of the distribution. It indicates that the selected parameters and their ranges were at the boundary. In order to shift the simulated distribution to the right to capture the field condition, key parameters that most affected the results were identified and expanded in ranges. According to the scatter plots of each parameter versus CORSIM travel time, mean queue discharge headway, gap distributions for left turns, and link mean free flow speed were primary factors.
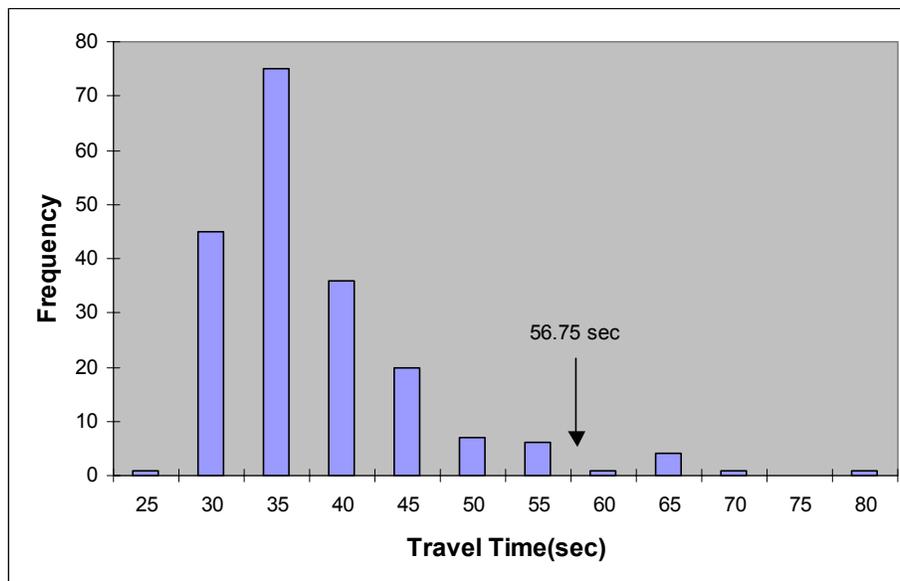

Figure 28. Feasibility Test Results for Site 15 with CORSIM

ANOVA was conducted to identify the key parameters. The results are summarized in Table 15. According to ANOVA, the key parameters were gap distribution for left turns, mean queue discharge headway, mean start-up lost time, and link mean free flow speed. Mean start-up lost time was missing in the previous result probably because it had a significant effect only when combined with mean start-up lost time distribution (see joint effect analysis in Table 16).

Table 15. ANOVA Results for Site 15 with CORSIM

Site 15 – CORSIM	Significance value (p value)	Significant mean differences (Sig. < 0.5)	SSR between groups
Link mean free flow speed	0.045	1-2, 1-3*	837.4
Mean queue discharge headway	0.000	1-2, 1-3*, 2-3*	6573.4
Mean start-up lost time	0.000	1-3*, 2-3*	3073.4
Left turn jumper probability	0.440	2-3	225.1
Left turn speed	0.806	NA	59.2
Right turn speed	0.292	2-3	336.3
Left-turn lagging within 2 seconds	0.747	NA	80.2
Left-turn lagging within 2-4 seconds	0.159	1-2	500.9
Amber interval response	0.913	NA	72.5
Gap distribution for left turns	0.000	1-2, 1-3*, 2-3*	6689.3
Gap distribution for right turns	0.303	2-3	326.6
Distribution of free flow speed by driver type	0.375	1-2	268.7
Start-up lost time distribution	0.347	1-2	290.0
Discharge headway distribution	0.935	NA	18.5

* Significant value is less than 0.05.

Table 16. Joint Effect between Mean Start-up Lost Time and Its Distribution

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4489.717[a]	8	561.215	4.738	.000
Intercept	236625.023	1	236625.023	1997.740	.000
V3	3815.707	2	1907.853	16.107	.000
V13	374.780	2	187.390	1.582	.208
V3 * V13	1132.578	4	283.145	2.390	.052
Error	22623.255	191	118.446		
Total	296415.375	200			
Corrected Total	27112.972	199			

a. R Squared = .166 (Adjusted R Squared = .131)

According to the field observations, left turn vehicles on the southbound approach sometimes waited at length to find an acceptable gap. Since it is a one-lane approach, these left turn vehicles blocked the whole approach and caused a long queue and delay for the vehicles behind them. In some occasions vehicles waited for two cycles before discharged. However, such phenomenon was never observed in the simulation. Therefore, one more option was

created for "gap distribution for left turn vehicles" considering longer gaps. Another 200 scenarios were generated, and multiple CORSIM simulations were conducted based on the new parameters and their ranges. Figure 29 shows the new results, which can be considered to cover the field data. The parameter set with closest travel time was selected in the final evaluation.
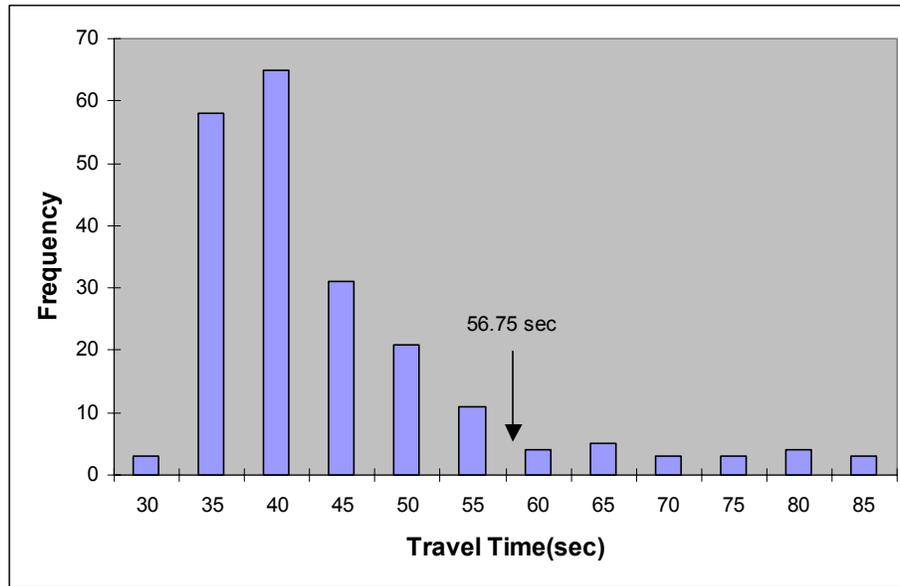

Figure 29. New Feasibility Test Results for Site 15 with CORSIM

Parameter Calibration Using Genetic Algorithm

A GA was integrated with CORSIM to find the optimal parameter values. The fitness function in Equation (2) was used. The simulation travel time was the value averaged from the ten CORSIM runs of each parameter set. Ten generations and ten populations were adopted in the GA process. The convergence of fitness value with generations is shown in Figure 30. Similar to PARAMICS, the parameters generated at the beginning generations showed good performance. The parameter set with the best fitness value was selected to represent the local traffic conditions and selected in the final evaluation.
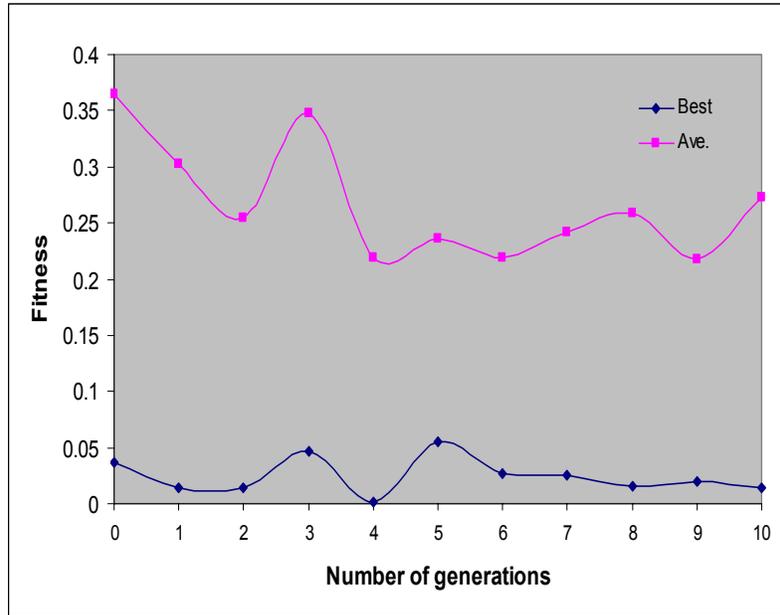
Figure 30. Convergence of Fitness Value with Generation (CORSIM)

Evaluation of the Parameter Sets

This section presents the comparison of simulation performance based on default parameters, best-guessed parameters, LHD-based parameters, and GA-based parameters. These parameter values are summarized in Table 17. Best-guessed parameter values in CORSIM were adjusted from the default values based on the observation of local traffic conditions and engineering judgment. Figure 31 shows the distribution comparison of simulation travel time based on the four parameter sets. The result shows that the calibrated parameters outperform the default and best-guessed parameters, and all three field data fall into the their distributions. The t test result shows that the calibrated parameter set using GA generates statistically significant results from the remaining three parameter sets. Animations of the calibrated model were viewed and found realistic.

Validation

The four calibrated CORSIM parameter sets were evaluated using field data that were not used in the calibration. Distributions of four parameter sets were compared in Figure 32. Similar to the calibration result, LHD-based and GA-based parameters could achieve simulated distributions with mean value close to the field data while the default parameters and best-guessed parameters produced shorter travel time from simulations. This finding was further confirmed via the t test, which shows that GA-based parameters generated similar results only with LHD-based parameters, with a p value of 0.12.

51

Table 17. Four Parameter Sets for Site 15 with CORSIM

| Site15-CORSIM	Default	Best-guessed	LHD	GA-Final
Link mean free flow speed (mph)	45	42	34	30
Mean queue discharge headway (sec)	1.8	2.0	2.7	2.8
Mean start-up lost time (sec)	2.0	2.3	2.4	2.8
Left turn jumper probability (%)	38	20	20	21
Left turn speed (ft/sec)	22	22	19	19
Right turn speed (ft/sec)	13	13	17	23
Left turn lagging within 2 seconds (%)	50	50	37	47
Left turn lagging within 2-4 seconds (%)	15	15	12	12
Amber interval response (fpss) Index	1	1	3	2
Gap distribution for left turns (sec) Index	1	1	3	1
Gap distribution for right turns (sec) Index	1	1	3	2
Distribution of free flow speed by driver type (%) Index	1	1	3	2
Start-up lost time distribution (%) Index	1	1	2	2
Discharge headway distribution (%) Index	1	1	1	1
Ave. Travel Time (sec)	28.17	31.7	63.09	58.41
Standard Deviation (sec)	1.82	2.16	16.06	16.03
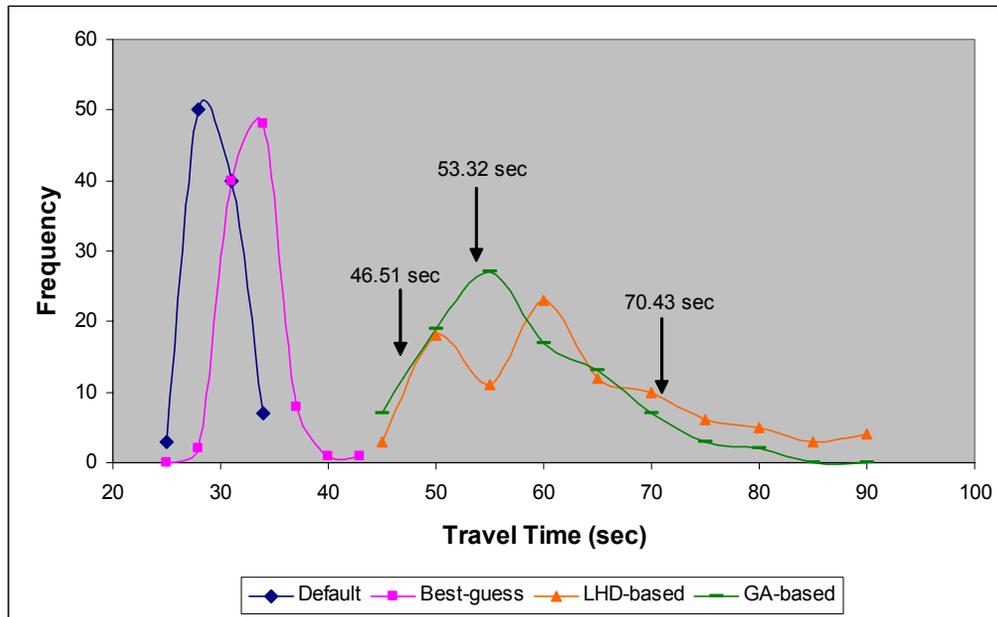
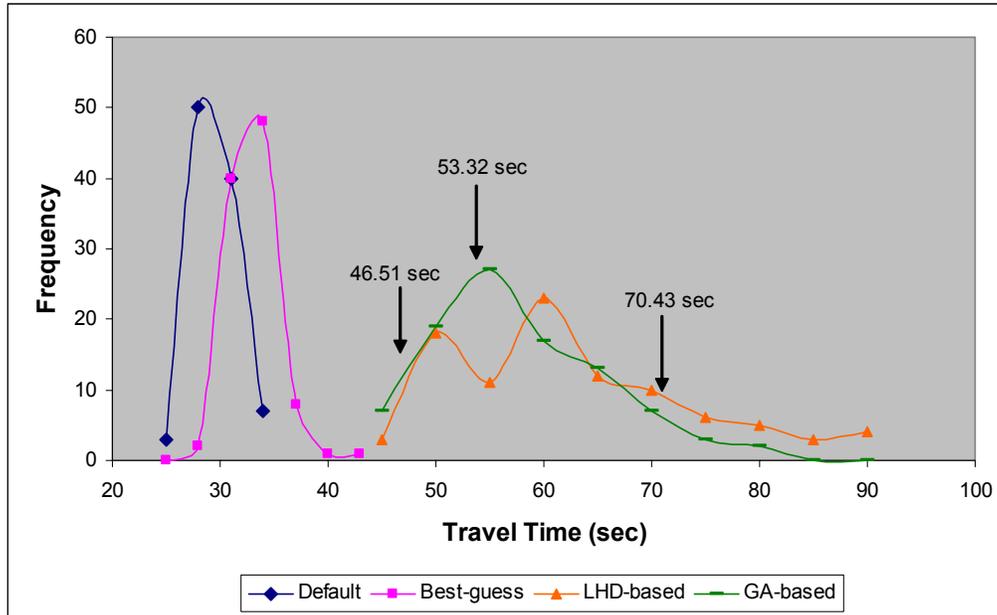**Figure 31. Comparison of Site 15 Travel Time with CORSIM**

52

Figure 32. Validation of CORSIM Using Site 15

Covington Network – VISSIM

Identification of Calibration Parameters

The following is the initial set of parameters and acceptable ranges used in the calibration process:

1)   Speed Index 1: 1-6 (65-70, 62.5-72.5, 60-75, 67.5-72.5, 65-75, 62.5-77.5 mph)
2)   Speed Index 2: 1-6 (55-60, 52.5-62.5, 50-65, 57.5-62.5, 55-65, 52.5-67.5 mph)
3)   Speed Index 3: 1-6 (45-50, 42.5-52.5, 40-55, 47.5-52.5, 45-55, 42.5-57.5 mph)
4)   Simulation Resolution: 1-9
5)   Waiting time before diffusion (second): 30-90
6)   Min. Headway (front/rear, meter): 0.1-0.9
7)   Max. Deceleration (m/s^2): -5.00 ~ -1.00
8)   Reduction Rate (meter per 1m/s^2): 20-80
9)   Accepted Deceleration (m/s^2): -3.0 ~ -0.2
10)   Number of observed preceding vehicles: 1-5
11)   Maximum look ahead distance (meter): 200 – 300
12)   CC0 – average standstill distance (meter): 1.0-2.0
13)   CC1 – headway at a certain speed (second): 0.5 ~ 3.0
14)   CC2 – longitudinal oscillation (meter): 0 ~ 15.0
15)   CC3 – start of the deceleration process (second): -30.0 ~ 0
16)   CC4 – minimal closing Δv (m/s): -1.0 ~ 0
17)   CC5 – minimal opening Δv (m/s): 0.0 ~ 1.0
18)   CC6 – $\pm dv/dx$ (10-4 rad/s): 0.0 ~ 20.0

53

19) CC7 – car following activities $\pm b$ (m/s2): 0.0 ~ 1.0
20) CC8 – acceleration behavior when starting (m/s2): 1.0 ~ 8.0
21) CC9 – acceleration behavior at v ~ 80 km/h (m/s2): 0.5 ~ 3.0

Parameters 1 to 3 set desired speed distributions along the Covington network at three locations where different posted speed limits were observed. They were indexed for the convenience of later experimental design. For instance, speed index 1 has six options to define the travel speed on the freeway where the posted speed limit is 65 mph. The values of CC0 to CC9, which are related to freeway traffic flow model, were obtained from the values presented at the Second Annual VISSIM Users Group Meeting.

Experimental Design for Calibration

The Latin Hypercube Sampling toolbox in MatLab was used to generate 200 scenarios using the initial set of parameters and their ranges.

Multiple Runs

Five random seeded runs were conducted in VISSIM for each of the 200 cases, for a total of 1000 runs. The average travel time was recorded for each of the 1000 runs. The results from the five multiple runs were then averaged to represent each of the 200 parameter sets.

Feasibility Test

In order to check whether the selected parameter set was able to produce field conditions, a feasibility test was conducted. Figure 33 shows the simulated travel time distribution of 200 cases. Since the field value fell within the acceptable range of the distribution, the ranges for current parameters were considered to be sufficient.
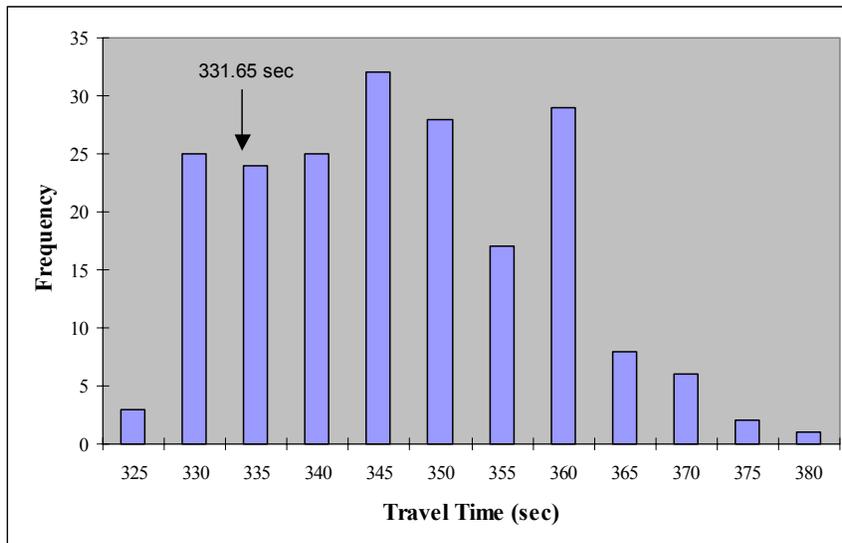

Figure 33. Feasibility Test Results for Covington Network with VISSIM

54

To identify the critical parameters, each parameter versus travel time from simulations was plotted. An apparent trend was observed for speed index 3, average standstill distance CC0, and headway at a certain speed CC1. In addition, ANOVA was used to identify the key parameters. Table 18 shows the ANOVA results. Based on the significance value of the *F* test and sum of squares (SSR) between groups, speed index 3, headway at a certain speed CC1, and speed index 2 were found to be the most important parameters to the result. Speed index 3 determines travel speeds on the longest portion of the network, with the posted speed limit of 45 mph, and speed index 2 determines the speeds on the second longest portion of the network, with the posted speed limit of 55 mph. CC1 is an important parameter to determine the safety distance between two continuous drivers on the freeway, which has strong influence on capacity especially in case of high volumes.

Table 18. ANOVA Results for Covington Network with VISSIM

Covington –VISSIM	Significance value (p value)	Significant mean differences (Sig. < 0.5)	SSR between groups
Speed Index 1	0.510	2-4	599.503
Speed Index 2	0.035	2-3, 2-4*, 2-5*, 2-6	1643.100
Speed Index 3	0.000	All*	24326.00
Simulation Resolution	0.809	NA	636.323
Waiting Time Before Diffusion (sec)	0.757	NA	78.023
Min. Headway (front/rear)	0.672	NA	217.022
Max. Deceleration	0.621	NA	248.612
-1m/s^2 per Distance	0.565	1-3, 2-3	478.399
Accepted Deceleration	0.142	1-2, 1-3	544.567
Observed Vehicles	0.757	NA	78.023
Look Ahead Distance (max)	0.011	1-3*, 2-3	1230.549
CC0	0.815	NA	57.364
CC1	0.016	1-5*, 2-5*, 2-5, 4-5	1663.209
CC2	0.011	1-2*, 1-3*	1242.380
CC3	0.037	1-2, 1-3*, 2-3	909.199
CC4	0.335	1-2	305.877
CC5	0.406	1-3	252.093
CC6	0.321	3-4	489.014
CC7	0.022	1-3, 2-3*	1056.758
CC8	0.538	NA	304.003
CC9	0.582	NA	401.303

* Significant value is less than 0.05.

Parameter Calibration Using Genetic Algorithm

With the parameters and acceptable ranges identified in the previous step, GA was integrated with VISSIM to find the optimal parameter set for the Covington network. GA procedure of 10 generations and 20 populations was repeated twice with different starting random seeds. The convergence of fitness values with generations is shown in Figures 34 and 35. It seems that GA can quickly converge to the optimal solution at the beginning generations. The parameter set with the best fitness value was selected in the final evaluation.
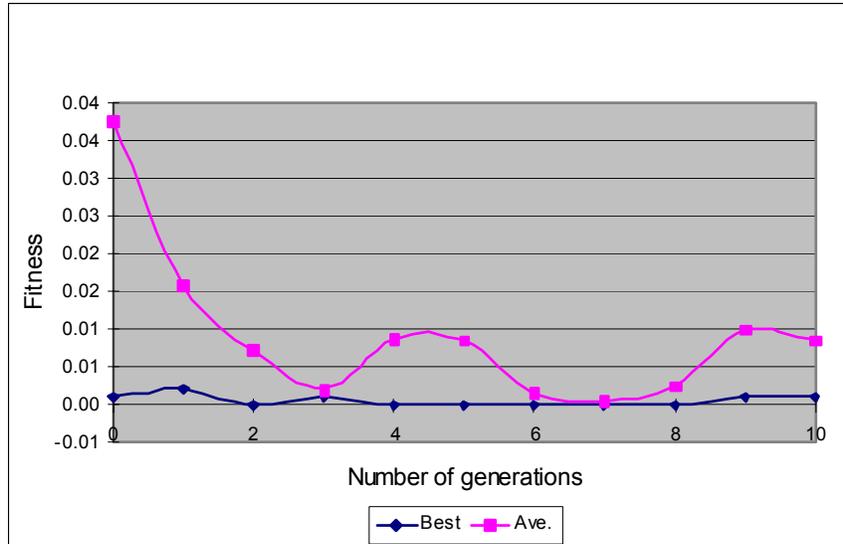
Figure 34. Convergence of Fitness Value with Generation (VISSIM, 1st Trial)
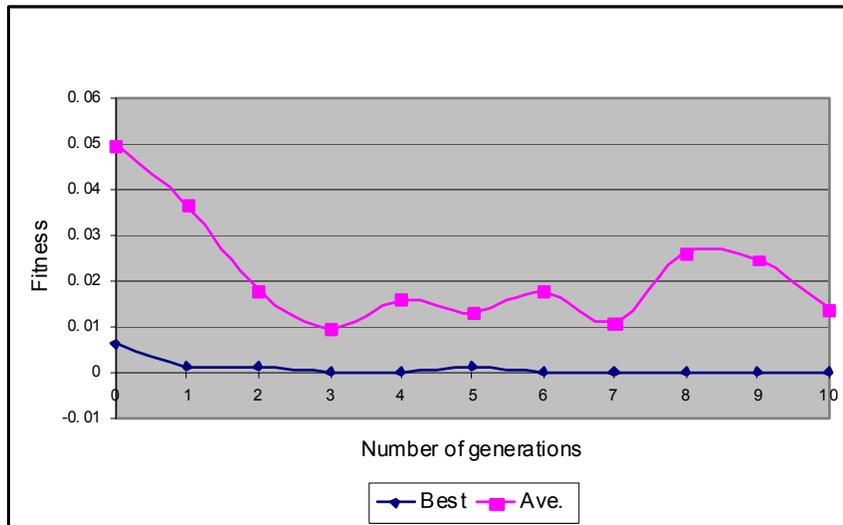

Figure 35. Convergence of Fitness Value with Generation (VISSIM, 2nd Trial)

Evaluation of the Parameter Sets

This section presents the comparison of 100 VISSIM simulation results based on default parameters, best-guessed parameters, LHD-based parameters, and GA-based parameters. The values of best-guessed parameters were determined based on engineering judgment and knowledge about local traffic conditions. The parameter values for each set, as well as the average travel time from simulation, are listed in Table 19. It was found that LHD-based parameters and GA-based parameters achieved almost identical simulation results in terms of mean average travel time, although the two parameter sets were quite different. The reason is

56

Table 19. Four Parameter Sets for Covington Network with VISSIM

Covington-VISSIM	Default	Best-guessed	LHD-based	GA-based
Speed Index 1	62.5-67.5 mph	65-70 mph	4 (67.5-72.5 mph)	5 (65.0-75.0 mph)
Speed Index 2	52.5-57.5 mph	55-60 mph	1 (55.0-60.0 mph)	2 (52.5-62.5 mph)
Speed Index 3	42.5-47.5 mph	45-50 mph	4 (47.5-52.5 mph)	4 (47.5-52.5 mph)
Simulation Resolution	5	10	3	9
Waiting Time Before Diffusion (sec)	60.00	60.00	72.89	68.2
Min. Headway (front/rear)	0.50	0.50	0.42	0.72
Max. Deceleration	-3.00	-3.00	-4.22	-1.04
-1m/s^2 per Distance	50.00	50.00	69.7	60
Accepted Deceleration	-1.00	-1.00	-0.5	-0.7
Observed Vehicles	2	4	5	3
Look Ahead Distance (max)	250.00	220.00	244.28	257.58
CC0	1.50	1.80	1.92	1.74
CC1	0.90	2.23	2.77	2.77
CC2	4.00	4.00	10.85	4.09
CC3	-8.00	-8.00	-21.59	-0.91
CC4	-0.35	-0.35	-0.48	-0.97
CC5	0.35	0.35	0.98	0.86
CC6	11.44	11.44	0.77	10.7
CC7	0.25	0.25	0.94	0.67
CC8	3.50	3.50	3.48	2.06
CC9	1.50	1.50	2.04	2.77
Ave. Travel Time (sec)	358.27	343.42	332.767	332.767

that the most important parameters, such as speed index 3, headway at a certain speed CC1, and speed index 2, have very close values in these two sets. In addition, the simulation is a stochastic process. Therefore, it is possible to obtain the same simulation results using different parameter sets as long as the key parameters are similar. However, the procedure may be repeated to confirm the performance of each parameter set.

As seen in Figure 36, the calibrated VISSIM models (LHD-based and GA-based parameters) provided simulation results similar to the field data, while the uncalibrated VISSIM models (default parameters and best-guessed parameters) generated higher travel time. The t test was conducted between GA-based parameter sets and the other three parameter sets. The result shows that GA-based parameters generated statistically significant results from all the other parameter sets. Animations of the calibrated VISSIM were viewed and deemed acceptable.
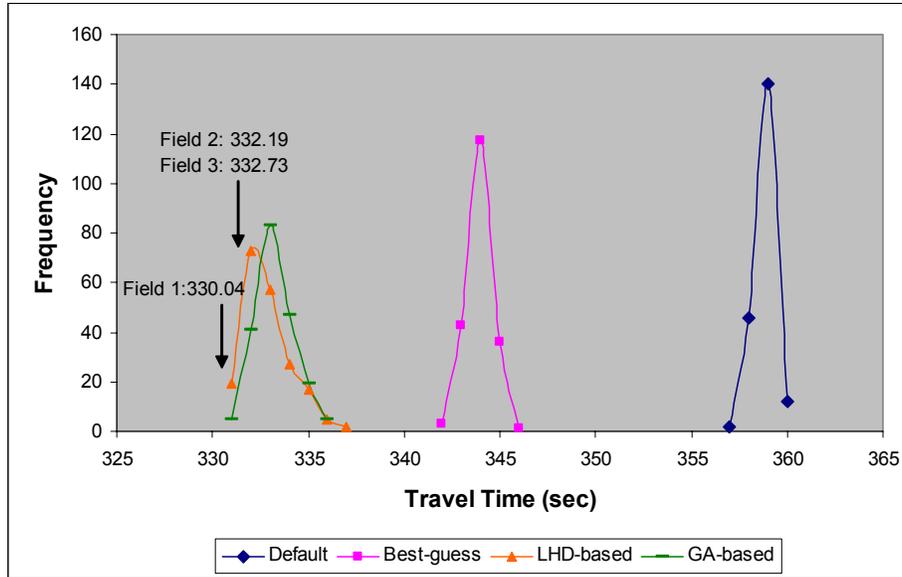
Figure 36. Comparison of Covington Network Travel Time with VISSIM

Validation

Traffic data collected on a different day were used for validation of parameter sets obtained from the calibration process. The field travel time was compared to the distributions of 100 runs using four parameter sets. Field data were a bit outside of the distributions of the calibrated parameters, but much closer to the distributions of LHD-based and GA-based parameters than that of best-guessed and uncalibrated parameters, as shown in Figure 37. The t test result indicates that the simulated distribution using GA-based parameters is statistically different from using the other three parameter sets.
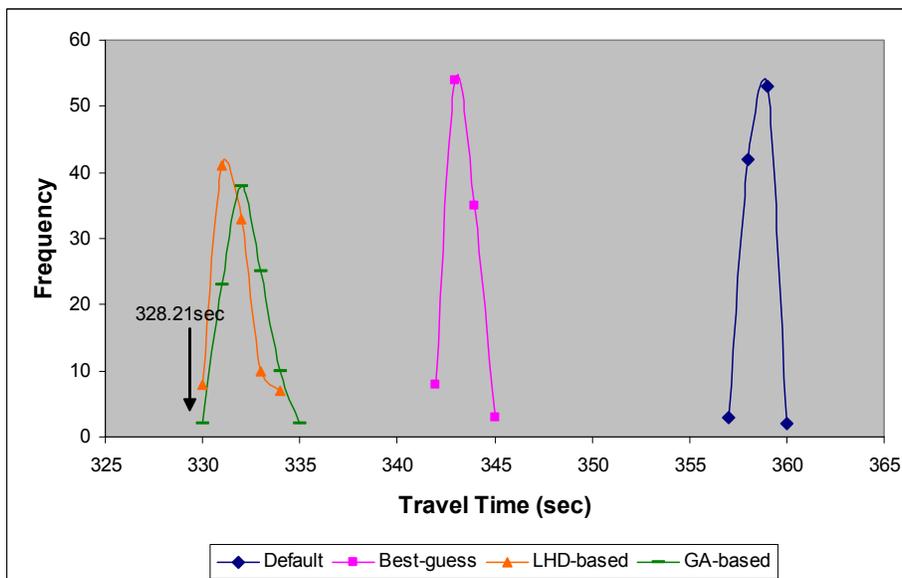

Figure 37. Validation of VISSIM Using Covington Network

Covington Network – PARAMICS

Identification of Calibration Parameters

       The following is the initial set of parameters and acceptable ranges identified in the calibration process. Acceptable ranges were based on the literature review and researchers' discretion.

1) Mean headway (sec): 0.6 – 2.2
2) Mean reaction time (sec): 0.3 – 1.9
3) Speed memory: 1 – 9
4) Curve speed factor: 1.0 – 5.0
5) Headway factor: 0.6 – 1.4
6) Link speed1 (mph): 60 – 70
7) Link speed2 (mph): 50 – 60
8) Link speed3 (mph): 40 – 50

where link speeds 1, 2, and 3 represent the speed limit of the links with different posted speed limits.

Experimental Design for Calibration

       The Latin hypercube design consisted of 200 cases of parameters and the values within the acceptable ranges.

Multiple Runs

       Ten random seeded runs were conducted in PARAMICS for each of the 200 cases, for a total of 2000 runs. The average travel time was recorded for each of the 2000 runs, and the results from the ten multiple runs were then averaged to represent each of the 200 parameter sets.

Feasibility Test

       In order to check the feasibility of selected parameters, the field data were compared to the simulated travel time distribution of 200 cases. As shown in Figure 38, the field data fall within the acceptable range of the distribution, which indicates that the current ranges for the selected parameters were sufficient. Based on the scatter plots of each parameter versus the travel times from simulations, mean headway and link speed 3 were considered as important parameters to the results. This finding was confirmed by ANOVA, as shown in Table 20. In addition, headway factor that determines the target headway for all vehicles on the selected links also was found to have certain impact on the result.
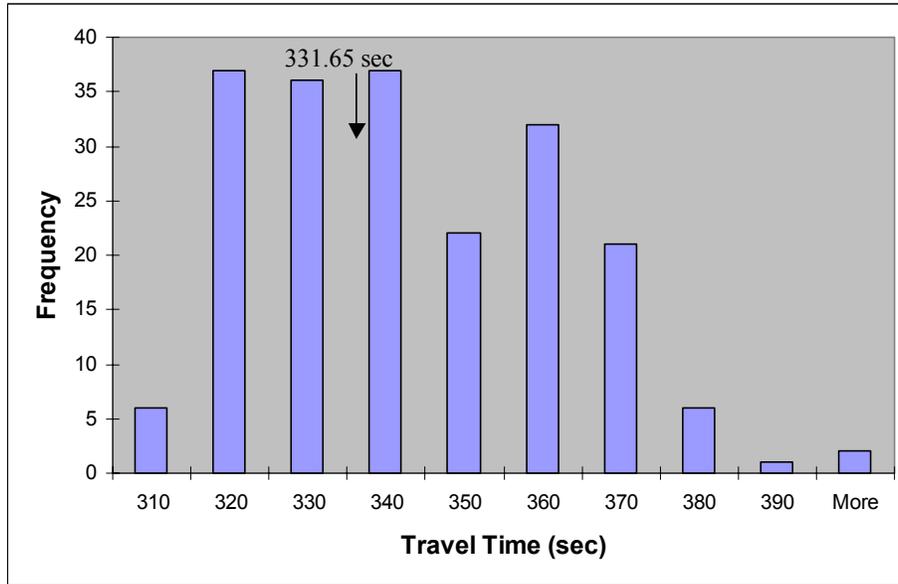
Figure 38. Feasibility Test Results for Covington Network with PARAMICS

Table 20. ANOVA Results for Covington Network with PARAMICS

Covington – PARAMICS	Significance value (ANOVA test)	Significant mean differences (Sig. < 0.5)	SSR between groups
Mean Headway	0.000	1-3*, 1-4*, 2-3, 2-4*, 3-4	6585.832
Mean Reaction Time	0.349	1-3	1195.959
Speed Memory	0.476	2-4	908.045
Curve Speed Factor	0.589	NA	700.924
Headway Factor	0.001	1-4*, 2-4*, 3-4*	5529.056
Link Speed1	0.354	1-2	755.875
Link Speed2	0.364	1-3	734.469
Link Speed3	0.000	ALL	50516.389

* Significant value is less than 0.05

Parameter Calibration Using Genetic Algorithm

GA procedure with ten generations and 20 populations were repeated twice to find the optimal parameter set. As shown in Figures 39 and 40, after about three generations, the fitness value is not improved much, just oscillates a bit. The parameter set with best performance was selected from the final generations.

Evaluation of the Parameter Sets

The comparison of 100 PARAMICS simulation results based on default parameters, best-guessed parameters, LHD-based parameters, and GA-based parameters was performed to evaluate the parameter sets. Again, best-guessed parameters were obtained from the knowledge of local traffic conditions. Table 21 summarizes these parameter values. As seen in Figure 41,
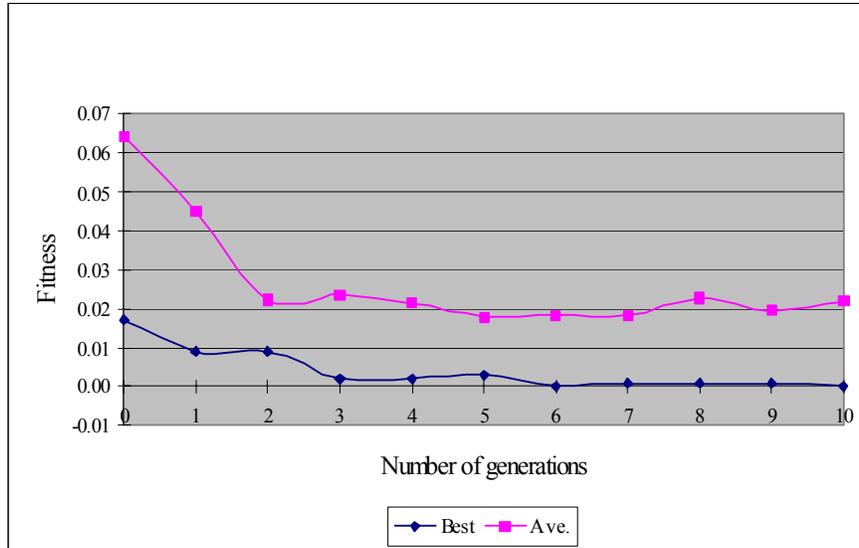
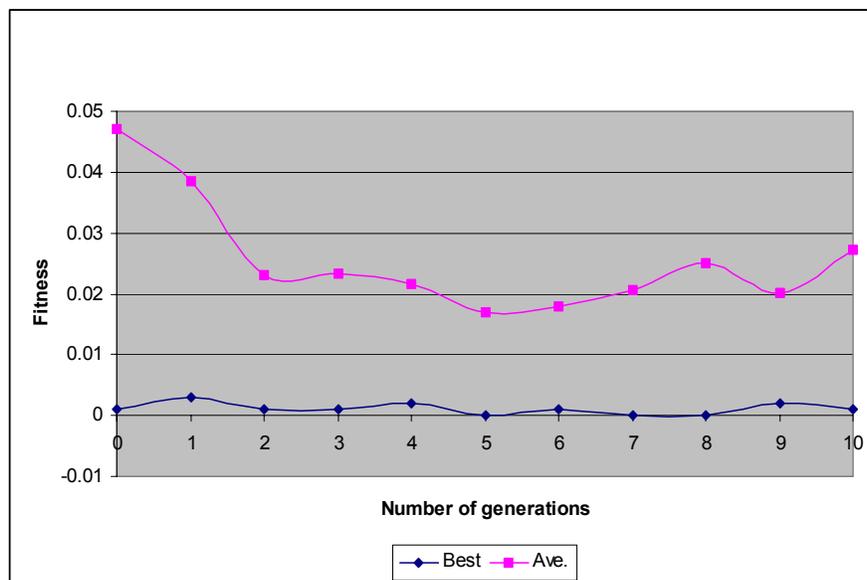**Figure 39. Convergence of Fitness Value with Generation (PARAMICS, 1st Trial)**

Figure 40. Convergence of Fitness Value with Generation (PARAMICS, 2nd Trial)

all parameter sets except for the best-guessed parameters produced a travel time distribution containing all field data. For the first time, the default parameters achieved the similar performance as the calibrated parameters, and the best-guessed parameters performed worse than the default parameters. The t test result shows that GA-based parameters generated similar results as the LHD-based parameters, with a p value of 0.019, but statistically different from the other two parameters. Animations of the calibrated PARAMICS were viewed and considered acceptable.

61

Table 21. Four Parameter Sets for Covington Network with PARAMICS

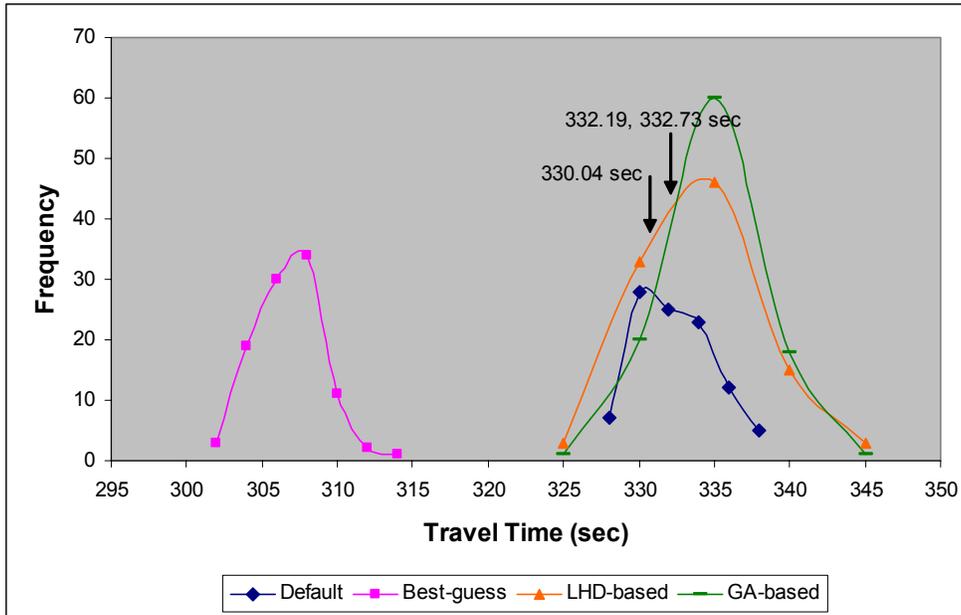| Covington - PARAMICS	Default	Best-guessed	LHD	GA
Mean Headway (sec)	1.0	1.2	1.9	1.33
Mean Reaction Time (sec)	1.0	1.0	1.9	0.99
Speed Memory	3	3	7	1
Curve Speed Factor	1.0	2.0	4.2	4.27
Headway Factor	1.0	1.2	1.13	1.12
Link Speed1 (mph)	65	70	61	62
Link Speed2 (mph)	55	60	59	55
Link Speed3 (mph)	45	50	48	46
Ave. Travel Time (sec)	331.4	305.88	331.3	332.5

Figure 41. Comparison of Covington Network Travel Time with PARAMICS

Validation

Another new data set that was not used in the calibration, including volume and travel time, was used for validation of parameter sets obtained from the calibration process. A similar pattern was found: Except for the best-guessed parameter set, the other three parameter sets generated good distributions containing the field data, as shown in Figure 42. However, according to the t test result, only the results of the default parameters were statistically equal to those of GA-based parameters, with a p value of 0.273.
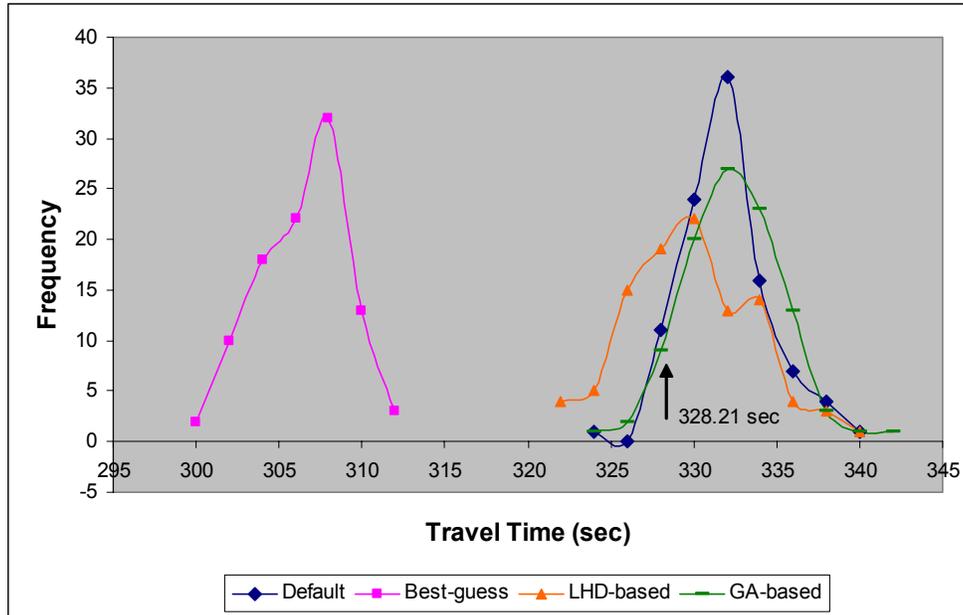
Figure 42. Validation of PARAMICS Using Covington Network

Covington Network – CORSIM

Identification of Calibration Parameters

The initial parameters with acceptable ranges and distribution alternatives are listed as follows:

1) Link mean free flow speed 1 (mph): 65-70
2) Link mean free flow speed 2 (mph): 55-60
3) Link mean free flow speed 3 (mph): 45-50
4) Car Following Sensitivity:
   Default(1): 1.25 1.15 1.05 0.95 0.85 0.75 0.65 0.55 0.45 0.35
   Small1(2): 1.15 1.05 0.95 0.85 0.75 0.65 0.55 0.45 0.35 0.25
   Large1(3): 1.35 1.25 1.15 1.05 0.95 0.85 0.75 0.65 0.55 0.45
   Narrow (4):1.16 1.08 1.00 0.92 0.84 0.76 0.68 0.60 0.52 0.44
   Wider (5): 1.34 1.22 1.10 0.98 0.86 0.74 0.62 0.50 0.38 0.26
5) Pitt car following constant (feet): 3-10
6) Lag acceleration (second): 0.3 – 2.0
7) Lag deceleration (second): 0.3 – 1.5
8) Time to complete a lane-change maneuver (second): 1 – 4
9) Gap acceptance parameter: 1 – 6
10) Percent of drivers desiring to yield to merging vehicles (%): 5 – 30
11) Multiplier for desire to make a discretionary lane change: 0.1 – 0.9
12) Advantage threshold for discretionary lane change: 0.1 – 0.9
13) Minimum separation for generation of vehicles (sec): 0.2 – 2.0
14) Distribution of free flow speed by driver type

Default: 88 91 94 97 99 101 103 106 109 112
Narrow (0.8): 90 93 95 98 99 101 102 105 107 110
Wide (1.2):   86 89 93 96 99 101 104 107 111 114

Parameters 1 to 3 set the link mean free flow speeds for links with different field posted speed limits. Index 1 represents 65 mph while index 3 represents 45 mph. For those parameters having a distribution such as car following sensitivity, several options were created to represent different driver behaviors.

Experimental Design for Calibration

The Latin hypercube design consisted of 200 cases using the selected parameters and the values within the acceptable ranges.

Multiple Runs

Ten random seeded runs were conducted in CORSIM for each of the 200 cases, for a total of 2000 runs. The average travel time was recorded for each of the 2000 runs, and the results from the ten multiple runs were then averaged to represent each of the 200 parameter sets.

Feasibility Test

The field data were compared with the simulated distribution of 200 cases in Figure 43. The figure shows that all the travel times from simulations are higher than the field data, which indicates that the initial ranges for the selected parameters were not sufficient to achieve the field conditions and needed adjustment.
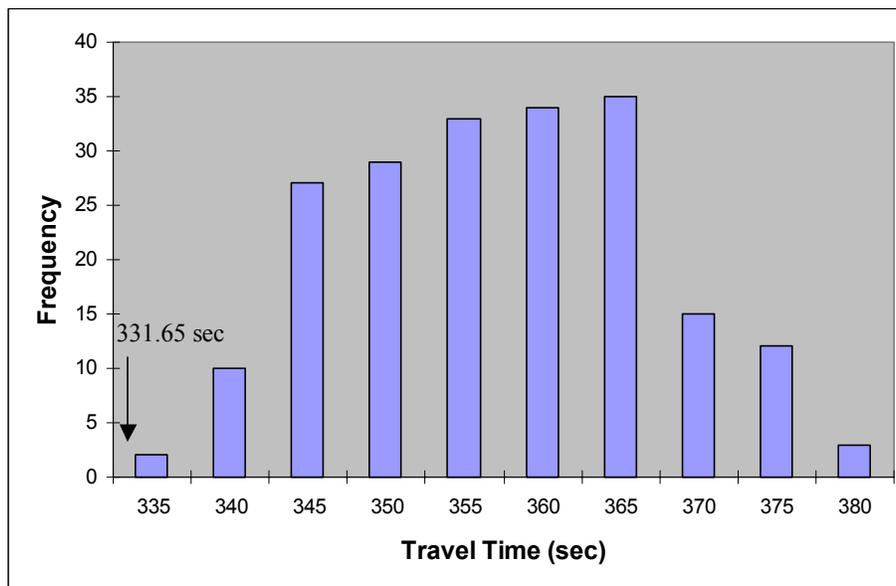

Figure 43. Feasibility Test Results for Covington Network with CORSIM

According to the scatter plots showing each parameter versus travel times from the simulations, the key parameters showing the apparent trend were link mean free flow speed 3, link mean free flow speed 2, and distribution of free flow speed. The finding was consistent with the ANOVA results, as seen in Table 22. These three parameters were found to have biggest SSR and smallest significance value. Joint effect analysis (MANOVA) also was conducted for the parameters that might have interaction effects, and the result did not reveal any significant joint effect among them.

Table 22. ANOVA Results for Covington Network with CORSIM

Covington – CORSIM	Significance value (p value)	Significant mean differences (Sig. < 0.5)	SSR between groups
Link mean free flow speed 1	0.858	NA	30.1
Link mean free flow speed 2	0.085	1-2, 1-3	480.1
Link mean free flow speed 3	0.000	ALL*	12752.8
Car following sensitivity	0.958	NA	63.3
Pitt car following constant	0.222	1-2, 2-3	294.0
Lag acceleration	0.683	NA	74.9
Lag deceleration	0.102	1-3, 2-3	444.6
Time to complete a lane-change maneuver	0.431	1-2	165.0
Gap acceptance parameter	0.963	NA	7.4
Percent of drivers desiring to yield to merging vehicles	0.730	NA	62.0
Multiplier for desire to make a discretionary lane change	0.665	NA	80.2
Advantage threshold for discretionary lane change	0.366	2-3	197.0
Minimum separation for generation of vehicles	0.117	1-3, 2-3	418.6
Distribution of free flow speed by driver type	0.000	ALL*	5212.9

* Significant value is less than 0.05

In order to shift the simulated distribution to the left to capture the field travel time, key parameters such as mean free flow speeds need to be adjusted towards higher values. Therefore, the range of link mean free flow speeds 2 and 3 were modified from 55-60 mph to 57.5-62.5 mph, from 45-50 mph to 47.5-52.5 mph, separately. The new LHD result is shown in Figure 44, in which field data fall within the acceptable range of the distribution. Thus, the new parameters were considered to be able to produce the realistic traffic condition.
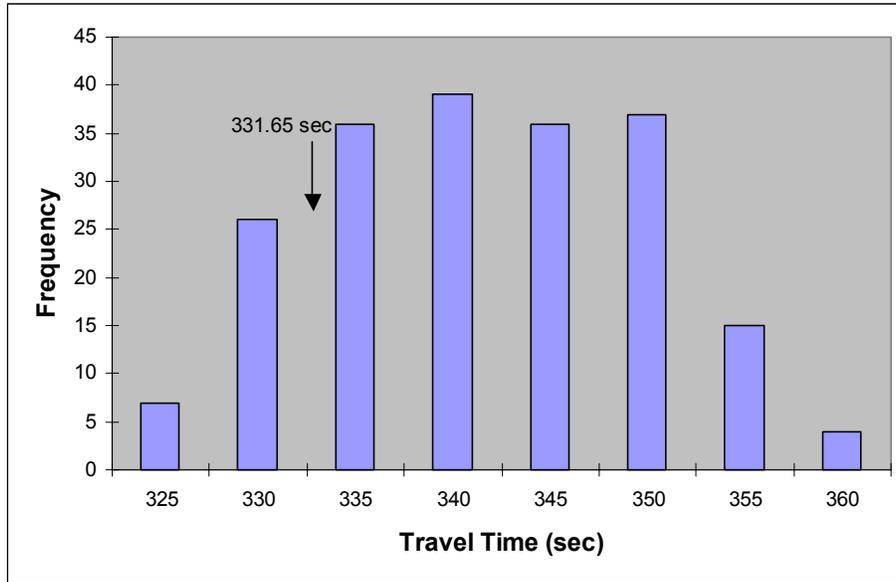
Figure 44. New Feasibility Test Result for Covington Network with CORSIM

Parameter Calibration Using Genetic Algorithm

A genetic algorithm was integrated with the CORSIM model to find the optimal parameter values. The simulation travel time was the value averaged from ten CORSIM runs per parameter set. Ten generations and 20 populations were adopted in the GA process. The convergence of fitness value with the generations is shown in Figure 45. Similar to PARAMICS, the best fitness value converged at the beginning generations and did not improve much later. The parameter set with best performance was selected from the final generation.
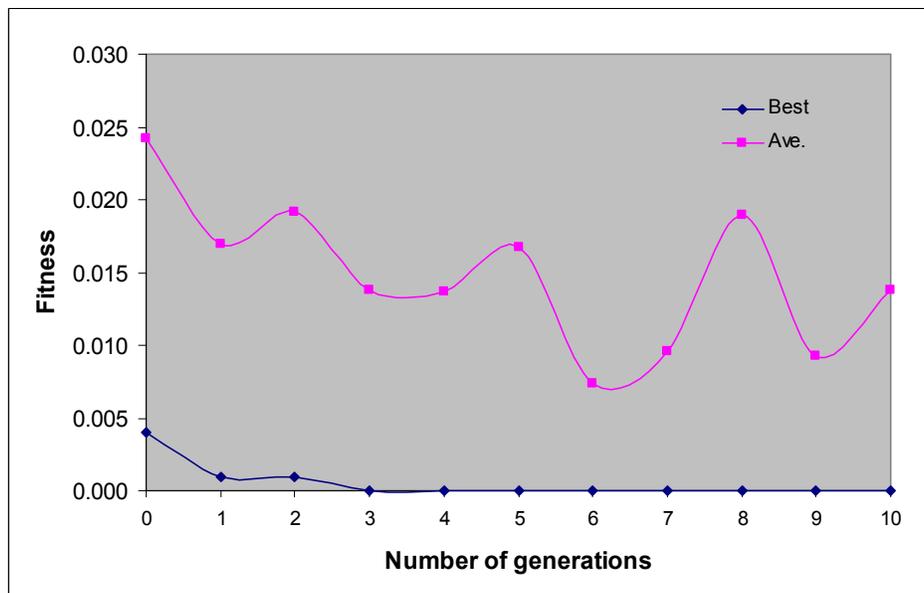

Figure 45. Convergence of Fitness Value with Generation (CORSIM, 1st Trial)

Evaluation of the Parameter Sets

This section presents the comparison of 100 CORSIM simulation results based on default parameters, best-guessed parameters, LHD-based parameters, and GA-based parameters. Table 23 lists these parameter values. As seen in Figure 46, the calibrated parameters outperformed the default and best-guessed parameters, and all the field data fall within the distributions of calibrated models. The t test result shows that the calibrated parameter set using GA generated statistically significant results from the other three parameter sets. Animations of the calibrated CORSIM were viewed and exhibited realistic driver behavior.

Table 23. Four Parameter Sets for Covington Network with CORSIM

| Covington -CORSIM	Default	Best-guessed	LHD	GA-Final
Link mean free flow speed 1 (mph)	65	70	70	66
Link mean free flow speed 2 (mph)	55	60	61	57
Link mean free flow speed 3 (mph)	45	50	51	51
Car following sensitivity Index	1	1	4	1
Pitt car following constant (ft)	10	10	8	3
Lag acceleration (sec)	0.3	0.3	0.8	1.2
Lag deceleration (sec)	0.3	0.3	0.6	0.5
Time to complete a lane-change maneuver (sec)	2.0	1.5	1.1	1.0
Gap acceptance parameter	3	3	3	4
Percent of drivers desiring to yield to merging vehicles (%)	20	30	11	20
Multiplier for desire to make a discretionary lane change	0.5	0.5	0.2	0.4
Advantage threshold for discretionary lane change	0.4	0.4	0.8	0.8
Minimum separation for generation of vehicles (sec)	1.6	1.6	1.9	1.3
Distribution of free flow speed by driver type Index	1	1	1	2
Ave. Travel Time (sec)	372.23	336.69	332.29	331.67
Standard Deviation (sec)	1.13	0.90	1.10	0.93
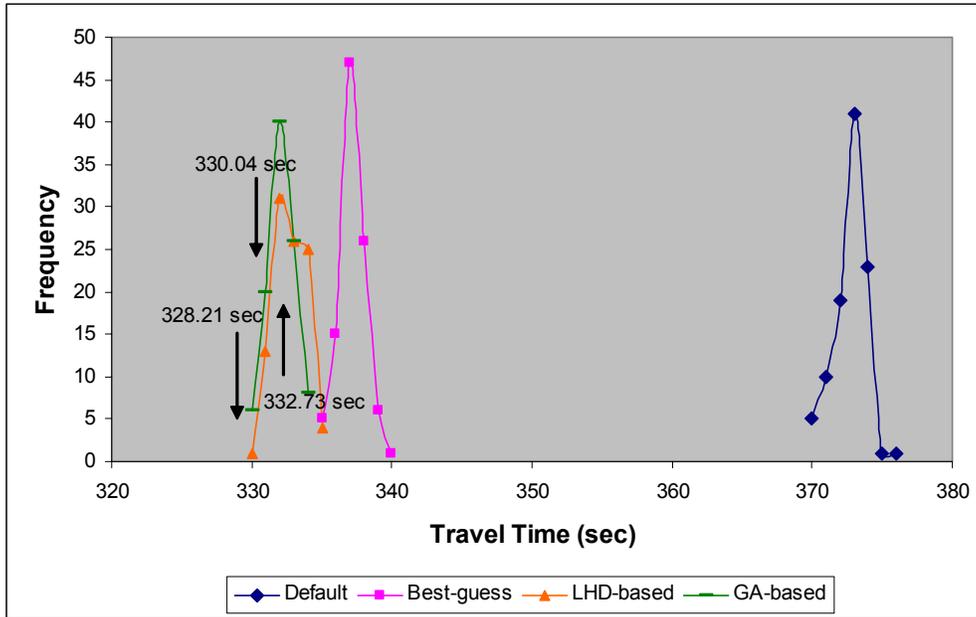
Figure 46. Comparison of Covington Network Travel Time with CORSIM

Validation

A similar pattern was found in the validation of the four parameter sets. The calibrated models generated satisfactory distributions containing the field data, as seen in Figure 47. The t test result indicated that only LHD-based parameters could achieve similar result as GA-based parameters, with a p value of 0.41.
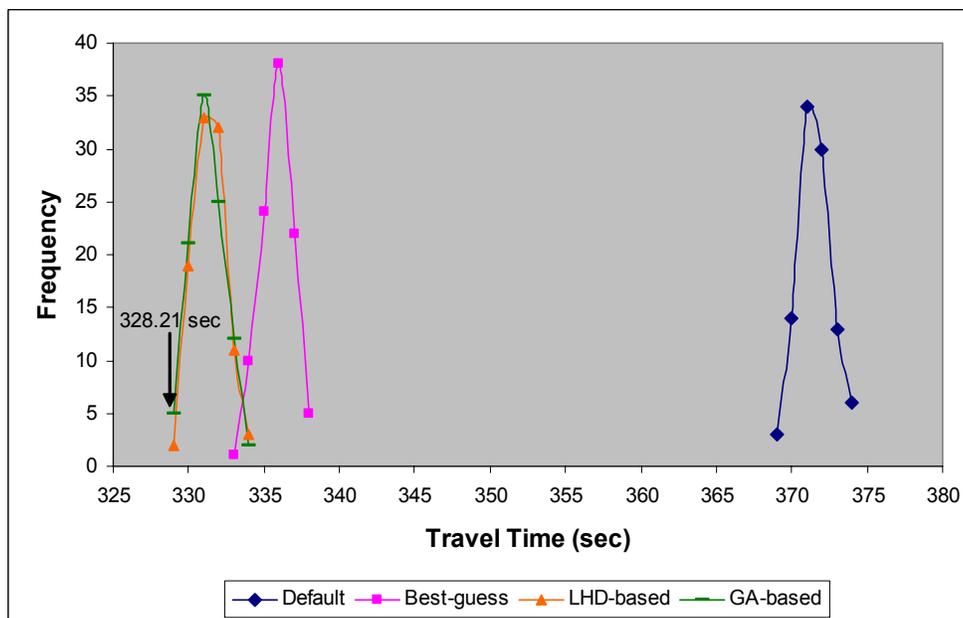

Figure 47. Validation of CORSIM Using Covington Network

CONCLUSIONS

This project developed and evaluated a procedure for calibration and validation of microscopic simulation models. The proposed procedure consisted of eight steps: (1) simulation model setup; (2) initial evaluation of default simulation output; (3) initial calibration comprising identification of calibration parameters, experimental design for calibration and multiple runs; (4) feasibility test and comparison to criteria; (5) parameter calibration using GA; (6) evaluation of the parameter sets, including simulated distributions, statistical tests, and visualizations; (7) model validation; and (8) feedback. The procedure was applied to three widely used microscopic simulation models – VISSIM, PARAMICS, and CORSIM – through two case studies under an actuated signalized intersection and a freeway segment with a lane-closure work zone in Virginia.

The performance of calibrated models was evaluated by comparing a distribution of simulation output to the multiple days of field data. When the variability of simulation output over multiple runs was studied in this research, PARAMICS generated wider distribution than VISSIM and CORSIM in both Site 15 and the Covington network. Thus, PARAMICS needs to make more replications to obtain consistent performance measures. This study shows that the proposed procedure produced acceptable results for all applications and thus confirms that the procedure is effective for different networks and simulation models used in this study. However, the performance of default parameters resulted in significant discrepancies between simulated and field data, while best-guessed parameters and LHD-based parameters usually generated better results than those of default parameters.

Different simulation models provide different sets of adjustable parameters. For example, to assign travel speeds to the vehicles, VISSIM uses desired speed distributions, PARAMICS uses link speed, and CORSIM combines link mean free flow speed and distribution of free flow speed by driver types. Generally, PARAMICS has the least calibration parameters, while CORSIM and VISSIM have more calibration parameters. In order to obtain realistic output from the simulations, it is important that users understand the definition and impact of each parameter. In addition, depending on the characteristics of the study site, key parameters that significantly affect the performance measures could vary. For example, gap distribution for left turns is critical when a large proportion of left-turn vehicles exist. Generally, car-following and lane-changing parameters are important to mimic real driver behavior for most scenarios. Appendix A provides a list of important parameters in each selected simulation model obtained from the case studies. The impact of key parameters cannot be overstated, and the appropriate values need to be assigned. An accurate definition of study scope and purpose and a thorough understanding of the study networks could help users select parameters for calibration.

Feasibility testing was found to be very useful in identifying reasonable and appropriate ranges of calibration parameters. It was performed as an initial check whether the selected parameters and their ranges were appropriate to achieve the field condition or not. If so, it was very likely that optimal parameter values could be obtained within these ranges. Otherwise, it would be difficult to obtain the realistic results even with an extensive search. The statistical analysis, such as ANOVA, was useful to identify key parameters. Sometimes, the best parameter

set provided by the feasibility test alone achieved similar performance as the optimized parameters using GA – and it requires less time and effort.

RECOMMENDATIONS

1. As demonstrated in this project, calibrated parameters obtained from the proposed procedure were able to replicate field conditions, while default parameters showed significant discrepancies. The importance of calibration and validation for simulation models is manifested. Therefore, simulation models need to be calibrated and validated before any further simulation-based traffic analyses are conducted.

2. The proposed model calibration and validation procedure was successfully implemented in two different networks with three selected microscopic simulation models. The study results indicated that such a systematic approach should be used to calibrate and validate microscopic simulation models. Specifically, the methods used to identify significant calibration parameters are useful for traffic engineers whenever a microscopic simulation model-based evaluation is conducted. Based on the lessons learned from saturation flow parameters in the calibration of VISSIM using Site 15, extra thought should be given to determining the initial parameters and the acceptable ranges. Both field data and analytical tools should be used to help determine initial parameter ranges.

3. Because this study used advanced statistical technique such as Latin Hypercube Design (LHD) and software programs including MATLAB and SPSS that are unfamiliar to VDOT engineers, the proposed procedure may not be practically adopted by VDOT traffic engineers. Immediate follow-up research should be conducted to help VDOT engineers exercise the proposed simulation procedure for their day-to-day simulation-based studies. The study should develop a well-documented manual for the calibration and validation procedure and conduct a hands-on short course and a workshop.

4. To account for day-to-day variability, multiple days of field data should be used. A single day of data could be higher or lower than the true mean of field value and may not represent the general field conditions. For example, the mean travel times of Site 15 from three days for calibration were 70.43 sec, 53.32 sec, and 46.51 sec, which showed a large variance. Using any single value would have resulted in very different parameter values. However, 4 days of data provide a paucity of traffic data in the calibration and validation procedure. A comprehensive range of data collection using advanced equipment, as well as ITS technologies, should be implemented to fully calibrate and validate the simulation models.

5. Further research is recommended in the long term that would to include more MOEs in the calibration and validation process. This study used travel time measure for calibration and validation. However, the performance of other MOEs is uncertain. The calibration and validation methodologies should be applied to a variety of performance measures besides travel time. In addition, field data under untried conditions should be considered for ultimate validation. As networks tested in this project were fairly simple, more complex networks

should be considered to verify the performance of the proposed procedure. Doing so will provide greater insight on the feasibility of the proposed procedure. Finally, with an increased trend of simulation applications for large and complex networks-based problems, it is necessary to calibrate not only driver behavior parameters but also dynamic origin and destination demand and route choice parameters. The proposed procedure for simulation model calibration and validation could be enhanced to include more functions and be evaluated under these scenarios.

ACKNOWLEDGMENTS

REFERENCES

Beasley, D., Bull, D.R., and Martin, R.R. An Overview of Genetic Algorithms: Part I, Fundamentals. *University Computing*, Vol. 15, No. 2, 1993, pp. 58-69.

Chen, A., and Yang, C. *A Stochastic Transportation Network Design Problem with Spatial Equity Constraint.* Transportation Research Board, Washington, DC, 2004.

Cheu, R.L., Jin, X., Ng, K.C., Ng, Y.L., and Srinivasan, D. Calibration of FRESIM for a Singapore expressway using genetic algorithm. *Journal of Transportation Engineering*, Vol. 124, No. 6, 1998, pp. 526-535.

Chu, L., Liu, Henry X., Oh, J-S, and Recker, W. *A Calibration Procedure for Microscopic Traffic Simulation.* Transportation Research Board, Washington, DC, 2004.

Cohen, S. L. *An Approach to Calibration and Validation of Traffic Simulation Models.* Transportation Research Board, Washington, DC, 2004.

Dowling, R., Skabardonis, A., Halkias, J., Hale, G. M., and Zammit, G. *Guidelines for Calibration of Microsimulation Models: Framework and Applications.* Transportation Research Board, Washington, DC, 2004.

Federal Highway Administration. *CORSIM User's Manual*. U.S. Department of Transportation, Office of Safety and Traffic Operations, McLean, VA, 1997.

Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Publishing Co., Inc., Reading, MA, 1989.

Hegazy T., Elbeltagi, E., and Hatem, E-B. *Bridge Deck Management System with Integrated Life Cycle Cost Optimization.* Transportation Research Board, Washington, DC, 2004.

Hellinga, B.R. Requirement for the Calibration of Traffic Simulation Models. Department of Civil Engineering, University of Waterloo. http://www.civil.uwaterloo.ca/b.PDF. Accessed March 3, 2003.

Kim, K-O, and Rilett, L.R. *A Genetic Algorithm Based Approach to Traffic Micro-simulation Calibration Using ITS Data.* Transportation Research Board, Washington, DC, 2004.

KLD Associates, Inc. *WATSim Wide-Area Traffic Simulator User's Guide.* TR-320A, Huntington Station, NY, February 2002.

Lee, D. H., Xu, Y., and Chandrasekar, P. *Parameter Calibration for PARAMICS Using Genetic Algorithm.* Transportation Research Board, Washington, DC, 2001.

Liu, Y. H. and Mahmassani, H. S. Global Maximum Likelihood Estimation Procedure for Multinomial Probit (MNP) Model Parameters. *Transportation Research Part B,* Vol.. 34, 2000, pp. 419-449.

M. Van Aerde &Assoc., Ltd., *INTEGRATION Release 2.30 User's Guide.* Kingston, Ontario, 2002.

The Mathworks, Inc. *MATLAB Users' Manual.* Massachusetts, 1999.

McKay, M.D.. and Beckman, R. J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, Vol. 21, No. 2, 1979, pp. 239-245.

Milam, R. T. *Recommended Guidelines for the Calibration and Validation of Traffic Simulation Models.* Fehr & Peers Associates, Inc. http://www.fehrandpeers.com/publications/traff_simulation_m.html. Accessed March 10, 2003.

Park, B., Yun, I., and Choi, K. Evaluation of Microscopic Simulation Programs for Coordinated Signal System. Paper presented at the 13th ITS America's Annual Meeting, Minneapolis, MN, May 19-22, 2003.

Park, B. and Schneeberger, J.D. *Microscopic Simulation Model Calibration and Validation: A Case Study of VISSIM for a Coordinated Actuated Signal System.* Transportation Research Board, Washington, DC, 2003.

Park, B., Santra, P., Yun, I., and Lee, D-H. *Optimization of Time-of-Day Breakpoints for Better Traffic Signal Control.* Transportation Research Board, Washington, DC, 2004.

PTV Planug Transport Verkehr AG. *VISSIM Version 3.6 Manual.* Innovative Transportation Concepts LLC, Karlsruhe, Germany, December 2001.

Quadstone Limited. *PARAMICS Modeler Version 3.0 User Guide and Reference Manual.* Edinburgh, U.K., 2002.

Sacks, J., Rouphail, N., Park, B., and Thakuriah, P. Statistically Based Validation of Computer Simulation Models in Traffic Operations and Management. *Journal of Transportation and Statistics*, 2002, Vol. 5, No. 1, 2002, pp. 1-24.

Smith, L. *Overview of TRANSIMS: the Transportation Analysis and Simulation System.* Los Alamos National Laboratory, May 1995.

SPSS Inc. *SPSS for Windows Release 10.1.0.* Chicago, 2001.

Sun, D., Benekohal, R.F., and Waller, S.T. *Intersection Signal Timing Optimization Using Multi-Objective Evolutionary Algorithm – NSGA II.* Transportation Research Board, Washington, DC, 2004.

Trafficware. *Synchro 6.0 User Guide.* Albany, CA, 2003.

Trafficware. *SimTraffic 6.0 User Guide.* Albany, CA, 2003.

TSS-Transportation Simulation System. *AIMSUN User Manual.* Barcelona, Spain, 2002.

Tsai, B-W, Kannekanti, V.N., and Harvey, J.T. *The Application of Genetic Algorithm in Asphalt Pavement Design.* Transportation Research Board, Washington, DC, 2004.

White, T. General Overview of Simulation Models. Paper presented at the 49th Annual Meeting of the Southern District Institute of Transportation Engineers, Williamsburg, VA, January 2002.

Wiedemann, R. Modeling of RTI-Elements on Multi-lane Roads. In *Advanced Telematics in Road Transport,* edited by the Commission of the European Community, DG XIII, Brussels, 1991.

Yang, Q., and Koutsopoulos, H.N. A Microscopic Traffic Simulator for Evaluation of Dynamic Traffic Management System. *Technical Report*, 1996.

Yang, S., Hamedi, M., and Haghani, A. *An Integrated Approach for the EMS Location and Assignment Problem.* Transportation Research Board, Washington, DC, 2004.

Zhang, Y. and Owen, L.E. *Systematic Validation of a Microscopic Traffic Simulation Program.* Transportation Research Board, Washington, DC, 2004.

APPENDIX A: IMPORTANT PARAMETER LIST

VISSIM

1. Waiting Time Before Diffusion
2. Minimum Headway (front/rear)
3. Maximum Deceleration
4. Reduction Rate
5. Accepted Deceleration
6. Number of Observed Vehicles
7. Maximum Look Ahead Distance
8. Average Standstill Distance (Urban)
9. Additive Part of Desired Safety Distance (Urban)
10. Multiple Part of Desired Safety Distance (Urban)
11. CC0 – average standstill distance (Freeway)
12. CC1 – headway at a certain speed (Freeway)
13. CC2 – longitudinal oscillation (Freeway)
14. CC3 – start of the deceleration process (Freeway)
15. CC7 – car following activities (Freeway)

PARAMICS

1. Mean Headway (Global)
2. Mean Reaction Time (Global)
3. Curve Speed Factor (Global)
4. Visibility Distance (Local)
5. Headway Factor (Local)
6. Link Endspeed (Local)

CORSIM

Urban Network
1. Queue Discharge Headway
2. Distribution of Multiplier for Discharge Headway Percentage
3. Start-up Lost Time
4. Distribution of Multiplier for Start-up Lost Time Percentage
5. Distribution of Freeflow Speed by Driver Type
6. Maximum Left and Right Turning Speeds
7. Gap Acceptance for Left and Right Turns
8. Lance Change Parameters

Freeway Network
1. Mean Start-up Lost Delay at Ramp Meters
2. Distribution of Freeflow Speed by Driver Type
3. Car-following Sensitivity Factor
4. Lane Change Gap Acceptance Parameters