VTRC
Virginia Transportation
Research Council

We Bring Innovation to Transportation

Methodology for Calibration and Validation of Mesoscopic Traffic Simulation Models

http://www.virginiadot.org/vtrc/main/online_reports/pdf/22-r7.pdf

MO ZHAO, Ph.D.
Research Scientist

JUSTICE APPIAH, Ph.D., P.E.
Senior Research Scientist

Final Report VTRC 22-R7

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
vtrc.virginiadot.org

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 22-R7	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: Methodology for Calibration and Validation of Mesoscopic Traffic Simulation Models		5. Report Date: August 2021
		6. Performing Organization Code:
7. Author(s): Mo Zhao, Ph.D., and Justice Appiah, Ph.D., P.E.		8. Performing Organization Report No.: VTRC 22-R7
9. Performing Organization and Address: Virginia Transportation Research Council 530 Edgemont Road Charlottesville, VA 22903		10. Work Unit No. (TRAIS):
		11. Contract or Grant No.: 114314
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation 1401 E. Broad Street Richmond, VA 23219	Federal Highway Administration 400 North 8th Street, Room 750 Richmond, VA 23219-4825	13. Type of Report and Period Covered: Final
		14. Sponsoring Agency Code:
15. Supplementary Notes: This is an SPR-B report		

16. Abstract:

Traffic simulation models are increasingly being used in the transportation engineering profession—often, to solve complex problems that may not lend themselves to traditional analysis techniques. The application of traffic simulation models has traditionally been at the individual vehicle (microscopic) level or aggregate traffic stream (macroscopic) level. Recently, the Virginia Department of Transportation and other agencies have shown interest in mesoscopic traffic simulation models, which allow for a level of detail higher than macroscopic models and model execution times better than those of microscopic models.

This study proposed a procedure for mesoscopic simulation model calibration and validation. The proposed procedure was demonstrated on a test bed along I-95 in the City of Richmond and Chesterfield County, Virginia, using Aimsun Next. Results of the case study indicated that the proposed procedure appears to be properly calibrating and validating the mesoscopic simulation model of the test bed.

17 Key Words: Mesoscopic simulation, calibration, validation		18. Distribution Statement: No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.	
19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages: 43	22. Price:

FINAL REPORT

METHODOLOGY FOR CALIBRATION AND VALIDATION OF MESOSCOPIC TRAFFIC SIMULATION MODELS

Mo Zhao, Ph.D.
Research Scientist

Justice Appiah, Ph.D., P.E.
Senior Research Scientist

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

August 2021
VTRC 22-R7

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

ABSTRACT

Traffic simulation models are increasingly being used in the transportation engineering profession—often, to solve complex problems that may not lend themselves to traditional analysis techniques. The application of traffic simulation models has traditionally been at the individual vehicle (microscopic) level or aggregate traffic stream (macroscopic) level. Recently, the Virginia Department of Transportation and other agencies have shown interest in mesoscopic traffic simulation models, which allow for a level of detail higher than macroscopic models and model execution times better than those of microscopic models.

This study proposed a procedure for mesoscopic simulation model calibration and validation. The proposed procedure was demonstrated on a test bed along I-95 in the City of Richmond and Chesterfield County, Virginia, using Aimsun Next. Results of the case study indicated that the proposed procedure appears to be properly calibrating and validating the mesoscopic simulation model of the test bed.

TABLE OF CONTENTS

FINAL REPORT

METHODOLOGY FOR CALIBRATION AND VALIDATION OF MESOSCOPIC TRAFFIC SIMULATION MODELS

Mo Zhao, Ph.D.
Research Scientist

Justice Appiah, Ph.D., P.E.
Senior Research Scientist

INTRODUCTION

When used properly, traffic simulation models provide a convenient environment for evaluating alternative traffic operations and transportation planning solutions in a relatively inexpensive and risk-free manner. Transportation system analyses using traffic simulation models have traditionally focused on the macroscopic and microscopic levels of detail. Macroscopic models provide a high-level view of the transportation system and are typically used to evaluate system-wide issues such as the amount of travel between areas (or zones)—sometimes segregated by trip purpose, time of day, or travel mode—and route choice from one area to another. Microscopic models, on the other hand, focus on evaluating the effects of localized travel demand, geometry, and traffic control on the operational performance (traffic flow rates, queuing, speeds, delay, etc.) of transportation facilities by modeling the movements of individual vehicles on a network. Even though microsimulation models provide good estimates of traffic operations, they are often not practical for analyzing large areas because the time and effort needed for model development and computational needs to execute large area models can be prohibitive.

Mesoscopic modeling is performed at an intermediate level of detail between macroscopic and microscopic simulation models. Compared with macroscopic models, mesoscopic models can simulate more details of the movements of individual vehicles and produce more accurate simulation results. Compared with microscopic models, mesoscopic models can provide significant savings in modeling time and effort, especially for large networks, without unduly compromising the accuracy of results. For example, Dixit et al. (2008) evaluated contraflow hurricane evacuation plans for an approximately 90-mile section of I-4 in Florida between Tampa and the Orange County line. Five scenarios were evaluated using both the microscopic version of Vissim and a mesoscopic cell transmission model proposed by Daganzo (1994). The results of mesoscopic and microscopic simulation analyses were similar (within 5% of each other) for all the metrics reported, including total travel time, evacuation clearance time, and excess number of vehicles that were not able to be loaded onto the network. However, the computing time for all five scenarios was 100 hours for the microscopic simulation analysis and 7.5 minutes for the mesoscopic simulation (Dixit et al., 2008).

The basic architecture of traffic simulation models includes inputs on both the supply (i.e., links, nodes, and traffic control system) and demand (point-to-point trip movements, link volumes, and vehicle routes) components of the transportation network. Because of the wide variety of end users, traffic simulation models are often built and calibrated for specific applications, and it is important that the resulting model produces output that is as close to reality as possible. This requires that the default driver behavior and other model parameters are adjusted to match those of the specific application and driver population for which the model is being developed. Although it is relatively straightforward to build the supply component and run a traffic simulation model, it is considerably more difficult to calibrate the model such that the model output matches field-observed data as closely as possible. This becomes even more difficult when large networks are simulated, as in the case of many mesoscopic modeling applications.

A study by Appiah et al. (2018) at the Virginia Transportation Research Council (VTRC) evaluated the feasibility of using commercially available mesoscopic traffic simulation software packages to analyze projects that may have a regional impact such as high occupancy vehicle / high occupancy toll (HOV/HOT) lanes and other intelligent transportation system / integrated corridor management strategies. That study compared two mesoscopic simulation software packages, Aimsun Next (hereinafter "Aimsun") and Vissim, in detail and concluded that "both seem to provide reasonable simulations of traffic operations and to be suitable for use as analysis tools when evaluating potential strategies for large scale networks." One major challenge for the Virginia Department of Transportation (VDOT) to mainstreaming the deployment of mesoscopic models is a lack of well-established procedures for model calibration and validation. As a logical extension of the earlier feasibility study, there is a need for a systemic procedure to calibrate and validate mesoscopic simulation models.

PURPOSE AND SCOPE

The purpose of this study was to propose a procedure to calibrate and validate mesoscopic traffic simulation models that can be implemented by VDOT. The objectives were as follows:

1. Identify a calibration and validation procedure that is generic and applicable to commercially available mesoscopic simulation models.

2. Summarize lessons learned that VDOT engineers can use when developing or evaluating mesoscopic simulation projects.

The proposed procedure was demonstrated using an Aimsun model of a test bed in the City of Richmond and Chesterfield County, Virginia. This study was focused on methods for calibration and validation of mesoscopic models. It did not define specific applications where mesoscopic models should be used over microscopic or macroscopic models; those decisions should be made by the project manager.

METHODS

Four tasks were performed to achieve the study objectives:

1. Conduct a literature review.
2. Identify the test bed, and develop a base simulation model.
3. Propose a calibration and validation procedure.
4. Demonstrate the proposed procedure.

Task 1: Conduct Literature Review

The literature on the latest developments with regard to the calibration and validation of mesoscopic traffic simulation models was identified. Results of the literature review helped the research team formulate the calibration procedure proposed in this study. The Transportation Research International Documentation (TRID) database was used to search the literature.

Task 2: Identify Test Bed and Develop a Base Model

This task involved selecting a test network and developing a base model to demonstrate the proposed calibration and validation procedure. Preference was given to selecting a site that included many of the different geometric and traffic control configurations—at-grade and grade-separated intersections, ramps, and merge and diverge areas—that were likely to be encountered in any type of transportation network and therefore likely to cover many of the parameters involved in the behavioral models of mesoscopic simulation models. Preference was also given to identifying a site for which relevant input data including network geometry, posted speeds, traffic signal timings, and origin-destination (OD) demand data were either readily available or could be collected with reasonable effort. The test network was determined in consultation with staff of VDOT's Central Office.

Task 3: Propose a Calibration and Validation Procedure

This task involved proposing a method to calibrate and validate mesoscopic simulation models. The proposed approach was a combination of (1) the three-stage process proposed by Kundé (2002), and (2) the implementation by Park and Schneeberger (2003) of a generic framework for the systematic calibration of traffic simulation models originally proposed by Hellinga (1998) (see Figure 1).

Similar to Kundé's (2002), the approach proposed in this study involved sequentially calibrating the network at increasing levels of complexity, starting with individual segments (disaggregate level) and progressing through subnetworks of interconnected links to the network as a whole (system level).

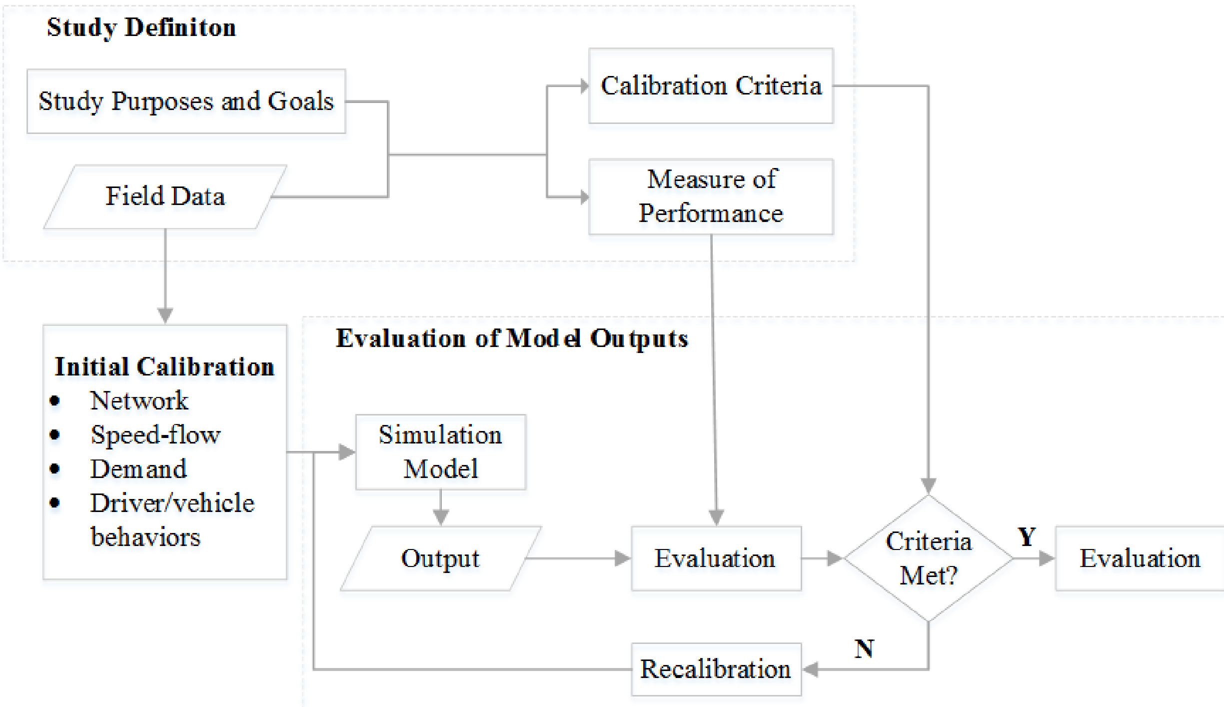


Figure 1. Generic Calibration Framework for Traffic Simulation Models. Adapted from Hellinga (1998).

At the system level, calibration is viewed as an optimization problem for which the problem solver seeks to find the set of parameter values that minimizes the discrepancy (as measured by the mean absolute percentage error) between the simulation output (travel times, link volumes, etc.) and field-observed data. Kundé (2002) proposed the simultaneous perturbation stochastic approximation (SPSA) method for system level calibration. However, the SPSA method was considered ill-suited for the purpose of this study, which was to propose a methodology that can be used by practitioners. As a consequence, an approach based on the optimization of a surface function (or metamodel) was proposed and is intended to be a simpler alternative to the SPSA and other more complex optimization methods. Examples of metamodels include linear and nonlinear regression models (Ciuffo et al., 2013; Park and Schneeberger, 2003).

A regression function similar to that of Park and Schneeberger (2003) was used to approximate the input-output response surface relationship of the simulation model and help identify candidate "optimal" parameter sets. This metamodeling approach was selected because it has been shown to be effective for the calibration of a Vissim microscopic traffic simulation model (Park and Schneeberger, 2003) and has been successfully applied to the sensitivity analysis of an Aimsun mesoscopic model (Ciuffo et al., 2013). Further, it was considered a reasonable balance between the commonly used trial and error method, which can be tedious and inefficient, and complex optimization techniques such as genetic algorithms and SPSA, which may not have great appeal among practitioners. A detailed description of the proposed procedure is provided later in this report.

Task 4: Demonstrate Proposed Procedure

The proposed procedure was demonstrated on a test bed in Richmond, Virginia, using the Aimsun mesoscopic traffic simulation software. The model was calibrated to field-observed conditions and validated against independent data sets from the test bed. Aimsun was selected as the software for the case study in consultation with staff of VDOT's Central Office.

RESULTS AND DISCUSSION

Literature Review

This section reviews the literature pertaining to characteristics, applications, and calibration of mesoscopic simulation models.

Overview of Mesoscopic Simulation Models

Network Representation

In mesoscopic transportation simulation models, the transportation network is usually represented as links (roadway segments) and nodes (junctions). Roadway segment characteristics that may be coded include road type, maximum speed, and number of lanes. A node serves as the transfer module between two adjacent links. Intersection elements such as traffic signal control settings and right-of-way assignment (priority rules) can also be coded. Traffic control at a signalized intersection is generally simplified such that the green time and the yellow time (or a portion of it) represent a "go" condition and the remainder of the cycle indicates a "stop" condition. Because of this simplification, actuated signal control operation may not be fully implemented in mesoscopic simulation. Conceptually, network links have two sections: the running section and the queuing section (Barceló, 2010). Queues may form at the nodes and spillback onto the link. The length of the queuing section can be determined from the vehicle length and the standstill distance. The running section of the link is not affected by the traffic spillback.

Traffic Demand

To build a mesoscopic simulation model, users generally need to define the traffic demand and the network configuration. Traffic demand is typically specified as a set of OD matrices that provides the number of vehicles of a given class going from every origin to every destination in the network during a specified time interval. Traffic belonging to the various OD pairs enters and exits the network at predefined locations.

For a given demand going from a given origin to a given destination, there are usually several possible routes to travel from the origin to the destination, and the vehicles must be distributed among these alternative routes. Assigning traffic to different routes is generally complicated by the fact that, in reality, the travel demand patterns change throughout the day

(Cascetta, 2009). Moreover, the road network may also have time-varying characteristics such as different traffic signal control strategies during different times of the day. All of these time dependencies must be carefully considered in developing a useful simulation model. A number of techniques are available for assigning traffic to the network. For example, in the Vissim mesoscopic simulation, an iterative simulation approach is used to assign traffic to the network dynamically (PTV AG, 2016). That is, the modeled network is simulated repeatedly and drivers choose their routes based on the travel costs (travel time, delay, toll, etc.) experienced from the preceding simulations. The iterations are typically continued for a pre-specified maximum number of iterations or until the traffic volumes and travel times on all edges do not change significantly from one iteration to the next.

Driving Behavior

Vehicles are simulated as traveling through the network in one of two ways in mesoscopic models (Barceló, 2010). The first approach considers vehicles as a package or platoon moving along links. Mesoscopic software such as CONTRAM adopts this approach (Burghout, 2005). The second approach uses a simplified car following model that enables some basic information about individual vehicles to be tracked in the simulation. Mesoscopic simulation tools such as Aimsun and Vissim use this approach. The volume of traffic on a link is heavily influenced by a critical road section parameter: the jam density. This user-defined parameter is the number of vehicles per mile per lane when the link is congested. It is used to define the traffic flow-density relationship.

Two driving behavior parameters are critical to modeling car following behavior in mesoscopic models. The first is the reaction time factor (also called the response time factor), which may be the same for all modeled network links and nodes or may be two separate values, one for nodes and another for links (Barceló, 2010). The reaction time parameter determines how fast the simulated vehicles react to traffic conditions, be it moving in a platoon or responding to changes in traffic signal indication. The second parameter is the look ahead distance. It is used to define how far a vehicle is going to look for the next possible turn, thus making necessary lane changes or selecting the proper lane at an upstream intersection.

The lane changing behavior of vehicles in mesoscopic simulation also varies by software (Burghout, 2005). In general, mesoscopic simulation software packages do not consider lane changing behavior within the link. Instead, the route for a vehicle is determined at the beginning of the simulation; when entering a link that is on its path, the vehicle uses the lane that best serves the need to stay on its predetermined path. Some mesoscopic software, for example, Aimsun, allows the vehicle to make lane selection when it leaves one link and enters another link (Aimsun, 2019).

Applications of Mesoscopic Models

Dixit et al. (2008) evaluated contraflow hurricane evacuation plans for an approximately 90-mile section of I-4 in Florida between Tampa and the Orange County line. Five scenarios were evaluated using both the microscopic version of Vissim and a mesoscopic cell transmission model proposed by Daganzo (1994). The results of the mesoscopic and microscopic simulation analyses were similar (within 5% of each other) for all the metrics reported, including total travel time, evacuation clearance time, and excess number of vehicles that were not able to be loaded onto the network. However, the microscopic simulation analysis took about 800 times longer than the mesoscopic simulation (Dixit et al., 2008).

Hou et al. (2013) used mesoscopic simulation to incorporate the impacts of adverse weather in traffic condition estimation and prediction. Mesoscopic simulation models were established for Irvine, California; Chicago, Illinois; Salt Lake City, Utah; and Baltimore, Maryland. Chiu et al. used mesoscopic simulation to evaluate contra-flow and phased evacuation strategies in a large area over a long time period and applied the proposed mesoscopic simulation model to a Houston-Galveston hurricane evacuation scenario (Chiu et al., 2008). De Palma and Marchal demonstrated the capabilities of mesoscopic simulation in modeling both within-a-day and day-to-day dynamics of large-scale transportation systems (de Palma and Marchal, 2002). Kristoffersson also used mesoscopic simulation to evaluate the impacts of cordon pricing on travelers' departure time, mode choice, and route choice behaviors in Stockholm (Kristoffersson, 2013).

Burghout (2005) used a hybrid simulation model to simulate traffic operations on a mixed freeway–urban network in Stockholm. The simulation took approximately 20% less time when compared to that of microscopic simulation (Burghout, 2005). Chen used the hybrid simulation to assess the impacts of a large development in New South Wales. The mesoscopic simulation was used to evaluate changes in route choice behavior in a large area surrounding the development (Chen, 2014). Liu et al. used the Aimsun mesoscopic simulation model to evaluate traffic dynamics of the Minneapolis–Saint Paul road network in Minnesota after the unexpected collapse of the I-35 Bridge (Liu et al., 2011). Casas et al. demonstrated the advantages of using the hybrid microscopic and mesoscopic simulation to evaluate the impacts of an advanced traveler information system and tested the approach on parts of the Madrid network that included 327 centroids, 1,375 intersections, and 3,591 sections (Casas et al., 2011).

Sun et al. (2020) used Aimsun and Vissim to demonstrate the practical application of mesoscopic simulation models as tools for evaluating traffic management strategies. A network consisting of a system of freeways in Richmond, Virginia, was prepared and used to evaluate the potential impacts of various levels of HOV lane usage on system operations. Both Aimsun and Vissim mesoscopic simulation produced results that were "deemed to be reasonable."

Calibration of Simulation Models

Hellinga (1998) proposed a framework for systematic calibration of traffic simulation models consisting of nine steps: (1) defining study goals and objectives, (2) determining required field data, (3) choosing measures of performance, (4) establishing evaluation criteria, (5) establishing network representation, (6) determining macroscopic speed-flow-density relationships, (7) determining driver routing behavior, (8) determining demand characteristics, and (9) evaluating model outputs. Hellinga (1998) provided useful guidance but did not include a direct application of the process to an actual network.

Park and Schneeberger (2003) proposed a nine-step procedure for calibration and validation of microscopic traffic simulation models, which was generally consistent with Hellinga's (1998) framework. The nine steps were (1) select measure of effectiveness; (2) collect data; (3) identify calibration parameters; (4) develop experimental design; (5) run simulation; (6) develop surface function; (7) generate candidate parameter set; (8) evaluate; and (9) validate through new data collection. An example case study using data from a network in Fairfax, Virginia, and the Vissim microsimulation model was successfully used to demonstrate application of the procedure.

Kundé (2002) proposed a three-stage approach to calibrating mesoscopic traffic simulation models, with a focus on calibrating the supply side of such models. At the disaggregate level (first stage), segment speed-density relationships were estimated using field data. These relationships were then refined in carefully selected subnetworks (collections of links and nodes) so that interactions among segments could be accounted for (second stage). The ideal subnetwork had either a limited number of route choice options or no route choice options (Kundé, 2002). The use of subnetworks with minimal or no route choice options facilitated the estimation of accurate OD demand (at the subnetwork level) and allowed for better simulation of traffic dynamics, without the errors inherent in demand estimation in the presence of route choice (Kundé, 2002). The third stage used an optimization approach at the entire network level to incorporate demand-supply interactions into the calibration process. The approach was demonstrated on a network in Irvine, California, that consisted of 298 nodes, 618 links, and 1,373 segments using the DynaMIT mesoscopic simulation model. A similar approach was adopted for calibration of a DynaSmart mesoscopic model of Hampton Roads, Virginia, by Park et al. (2010).

Balakrishna et al. (2007) extended the work of Kundé (2002) by incorporating demand estimation into the model calibration process. They used the SPSA algorithm to estimate simultaneously "all demand and supply inputs and parameters" for the DynaMIT mesoscopic model. The methodology was demonstrated on a network from Los Angeles, California, with 243 nodes, 606 links, and 740 segments.

Shafiei et al. (2017) discussed efforts to calibrate a large-scale Aimsun mesoscopic model of Melbourne, Australia, using data from multiple sources. Traffic flow fundamental diagrams were calibrated based on empirical data using statistical clustering and classification techniques. The OD demand was obtained through a bi-level optimization process that generated OD split proportions at the lower level and minimized the gap between observed and estimated traffic data at the upper level.

Sensitivity Analysis

Ciuffo et al. (2013) performed a sensitivity analysis of the model parameters of the Aimsun mesoscopic model. Five different test networks, i.e., roundabout, unsignalized intersection, signalized intersection, on-ramp, and merge-diverge area, were modeled in a case study. Collectively, the five networks were considered representative of the main configurations available in a typical transportation network and deemed to "cover all parameters involved in the behavioral models" that underlie the Aimsun mesoscopic model. A metamodel-based sensitivity analysis technique was used to identify the relative importance of model parameters. Seven parameters were considered in the study: reaction time, reaction time at stop, vehicle length, jam density, maximum give-way time, maximum acceleration, and random seed. The main findings were as follows:

- The metamodel-based approach produced results that were considered reliable for all but the roundabout scenario.
- In all scenarios, changes in reaction time had the largest effect on the output. This was especially true for the on-ramp, merge-diverge, and signalized intersection networks where the "reaction time alone accounts for approximately 90% of the variance of the mean travel time."
- Vehicle length accounted for "a certain share in the output variance for almost all network configurations."
- Maximum give-way time accounted for "a significant amount of the output variance" in the unsignalized intersection scenario.
- Reaction time at stop accounted for "just a small amount of output variance" in the signalized intersection scenario.
- Random seed accounted for "some 10% of the output variance."

Test Bed and Model Development

Test Bed

The network selected for use as the test bed is shown in Figure 2. It extends more than 13 miles along I-95 between State Route 10 (Exit 61) in Chesterfield County and State Route 195 (Exit 74) in the City of Richmond. The network consists of I-95, part of the Route 1 corridor that is parallel to I-95, and urban arterials that intersect with I-95. It includes seven interchanges on I-95, including three interchanges with intersecting limited-access highways (Route 195, Route 150, and Route 288). The test bed also includes 25 intersections on arterials and/or freeway ramps. This test bed was used in a previous study for VDOT: *I-95 Existing Conditions Vissim Model Development* (Kimley-Horn and Associates, 2017). In the remainder of this report, this study is referred to as "I-95 existing conditions study." More details about the study area can be found in the study report (Kimley-Horn and Associates, 2017).

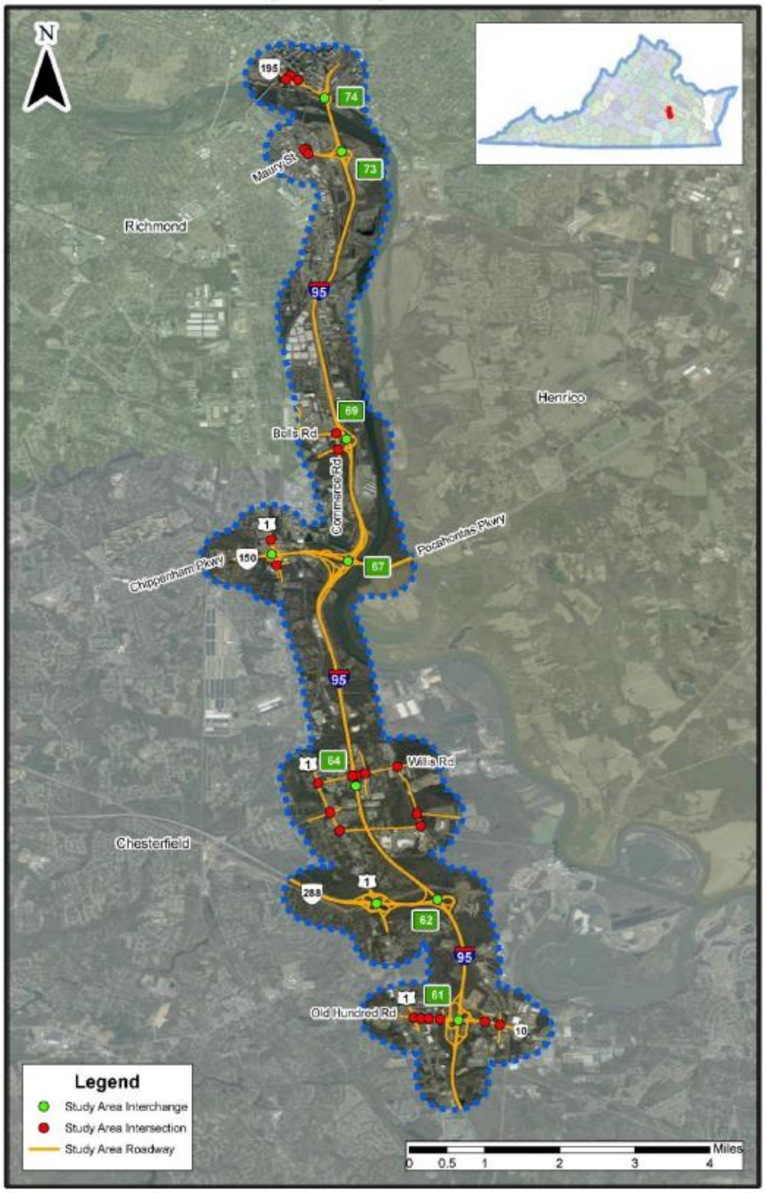


Figure 2. Test Bed Along I-95 From Downtown Richmond to Route 10 (Kimley-Horn and Associates, 2017)

Overview of the Aimsun Mesoscopic Model

The software used to demonstrate the proposed calibration and validation procedure was Aimsun Next, Version 8.4.1 (Aimsun, 2019). The Aimsun traffic simulator integrates macroscopic, mesoscopic, and microscopic simulations and a hybrid simulation (a combination of the mesoscopic and microscopic simulations). An overview of the Aimsun mesoscopic model including network representation and behavioral models is provided here.

Network Representation

Aimsun models the transportation network as a directed graph consisting of four geometric elements: centroids, sections, nodes, and turns (Aimsun, 2019). These are defined as follows:

- Centroids are the source and sink of vehicles within the simulation.

- Sections represent the roadway links and segments, and nodes represent the junctions and intersections (Nevada Department of Transportation [DOT], 2018). Section characteristics that may be coded include number of lanes, speed limit, and jam density.

- Within a node, the turns serve as connectors transferring vehicles between two adjacent sections. Vehicles are assumed to travel at free-flow speed within turns. The turn speed and turn length are used to calculate the turn travel time.

Behavioral Models

Vehicle movements in the Aimsun mesoscopic model are governed by three behavioral models: car following, gap acceptance, and lane changing / lane selection models (Aimsun, 2019). Aimsun calculates the time at which each vehicle enters and exits a section (discrete event simulation), rather than the position of each vehicle over time.

A simplified car following model is used to calculate the earliest time at which a vehicle exits from a section and thus the section travel time (Mahut, 2001). The number of vehicles in a section is bounded by the section capacity, which is the product of the jam density, number of lanes, and section length. Similarly, the turn capacity is calculated as the product of the jam density of the origin section, turn length, and number of destination lanes. The simplified car following model produces a dual-regime speed-density relationship such that the speed $V(k)$ remains constant during the free-flow regime but decreases once the density k exceeds a critical value and enters the congested regime (Shafiei et al., 2017). Mathematically this is expressed as shown in Equation 1:

$$V(k) = min\left[V_f, \frac{1}{R}\left(\frac{1}{k} - \frac{1}{k_{jam}}\right)\right] \qquad \text{[Eq. 1]}$$

where

V_f = the free-flow speed (mph)
k_{jam} = the jam density (vehicle/mile/lane)
R = the reaction time (hour).

Equation 1 indicates that for a given value of jam density, the shape of the model (and thus calibration of the speed-density relationship) is jointly determined by the free-flow speed V_f and the reaction time R.

The lane changing model is used to select entrance and exit lanes when the vehicle exits from the upstream section. However, no lane change occurs inside the section. At the beginning of the simulation, the mesoscopic simulator generates a default lane sequence. During simulation, the lane selection model is used to determine whether a vehicle changes to another lane by evaluating the downstream lanes in terms of connector distances, density, and lane changing cost (Aimsun, 2019). The default downstream entrance lane choice is the non-full lane that provides the shortest connection. However, when the vehicle is within the look ahead distance, preference is given to the lane closest to the target downstream exit, thus increasing the likelihood that the vehicle is routed to its intended destination.

The gap acceptance model is used to model give-way behavior. When there are two vehicles that could be in conflict within a node, the gap acceptance model determines which one has priority. The gap acceptance model has several parameters, including maximum give-way time, initial safety margin, final safety margin, and give-way time factor. The vehicle waits in the minor section until it has a gap greater than the initial safety margin. Once it has waited the maximum give-way time multiplied by the give-way time factor, the minimum gap required to execute the movement is linearly reduced to the final safety margin.

Model Development

Input data required for the Aimsun model were input for the supply (i.e., links, nodes, and traffic control system) and demand (i.e., point-to-point OD flows, link volumes, and vehicle routes) components of the transportation network. The sources of these data are discussed here.

Supply Data

Supply data are usually collected from databases maintained by transportation agencies. In this study, the I-95 existing conditions study report (Kimley-Horn and Associates, 2017) and a Vissim microscopic model of the test bed from that study were provided to the research team and served as the major sources of data. These data reflect the morning peak hour conditions at the test bed in 2016. The major attributes of the supply elements coded included the following:

- link (or section) attributes such as number of lanes, length, width, grade, and any other distinguishing features of the link such as the presence and location of speed reduction zones and how links connect to each other and to nodes in the network

- node attributes such as number of approaches, lane widths, and degree and length of curvature

- relevant information on the traffic system such as the type of intersection control (i.e., signal controlled, stop controlled, or uncontrolled) and the corresponding attributes of each control type (e.g., signal and detector locations, signal timing plans, and priority rules for conflicting traffic).

Demand Data

The primary demand data required for mesoscopic simulation are centroid-to-centroid trip movements as represented by an OD trip matrix. The input OD demand data may be available from a number of sources including the U.S. Census Bureau's American Community Survey data; data from private sector companies such as StreetLight; and travel demand models. The selection of the OD data source may be based on factors such as ease of acquisition, cost, and data quality. Where OD data are unavailable or outdated, incomplete, or inaccurate, reliable procedures exist to calculate or update the OD matrices from traffic counts obtained from a subset of the network links. Cascetta (2009) and de Dios Ortúzar and Willumsen (2011) described some of the commonly used methods. Some simulation software including Aimsun have some of these synthetic OD matrix estimation procedures built into the software package.

For this study, demand data in the form of centroid-to-centroid OD flows were not available. Instead, OD demand data were estimated using Vissim vehicle input information including volumes entering the network every 15 minutes on the entrance links, vehicle compositions, and static routing decisions for predefined paths. The centroids were defined to be generally consistent with the vehicle entrance and exit locations in the Vissim model. Each centroid was connected to entrance and exit links that represent the traffic analysis zone entrance and exit, respectively. A total of 63 centroids were defined; each centroid was connected to one entrance link and one or more exit links. A base OD matrix for every 15 minutes from 6:30 to 8:30 AM was estimated for cars and trucks separately.

Network Coding, Debugging, and Demand Adjustment

The network structure was built by importing a base map from OpenStreetMap. The roadway characteristics such as road classification, lane configuration, connections and turning movements at nodes, and speed limit were also imported from OpenStreetMap and then edited based on the test bed data obtained from the I-95 existing conditions study report (Kimley-Horn and Associates, 2017).

Turning speeds at nodes were automatically calculated by Aimsun and manually checked and adjusted based on the turning speeds retrieved from the Vissim model. Traffic control plans were coded based on the data from the Vissim model.

The network was checked to verify that the configurations were consistent with the field conditions. The data coded into the network, including roadway geometry, traffic control, speed setting, vehicle composition, and traffic demand, were reviewed to identify and rectify coding errors. The "Check and Fix Network Tool" in Aimsun was used to identify network coding errors such as unconnected lanes, conflicting turns, missing stop or yield signs, and improper turning speed. The Aimsun mesoscopic simulator is based on dynamic traffic assignment (Barceló, 2010), and it provides two approaches to dynamic traffic assignment: stochastic route choice, and dynamic user equilibrium (Aimsun, 2019). The stochastic route choice (SRC) mode in Aimsun uses a discrete choice model (binomial, proportional, logit, C-logit, or user-defined models) to simulate driver behaviors. It is more suitable for evaluating a process where drivers adjust their behaviors according to day-to-day learning. The SRC mode is not equilibrium based and thus does not have a convergence criterion. Multiple runs (replications in the Aimsun SRC mode experiment) must be conducted for each scenario to reduce stochastic variability. With the dynamic user equilibrium (DUE) mode in Aimsun, the behavioral assumption on how drivers choose the routes is consistent with the DUE principle (Aimsun, 2019). The simulator calculates the shortest path (taking into account the estimated link costs on available paths) in each iteration/interval and updates the path flows using the method of successive averages (MSA) or gradient-based algorithms. No route choice models are involved in the DUE mode. The simulation time of one simulation run may be longer in the DUE mode than in the SRC mode depending on the convergence criterion used, but the DUE mode may require fewer replications than the SRC mode because a DUE mode experiment in Aimsun runs a simulation multiple times, adjusting the path assignments on each iteration. The iterative nature of the DUE mode reduces the reliance on results from a single seed simulation run (Aimsun, 2019). In this study, the DUE mode was used. Three user equilibrium algorithms (gradient-based method, MSA, and weighted MSA) are available in Aimsun. The MSA implemented in Aimsun was proposed by Florian et al. (2002). The MSA algorithm redistributes the flows among the available paths in an iterative procedure; a new shortest path from an origin to a destination at a time interval is computed at iteration n and then the path flows are updated (Aimsun, 2019). The weighted MSA implemented in Aimsun was proposed by Liu et al. (2007). The difference between MSA and weighted MSA is that in iteration n of the algorithm, $1/n$ of the demand is moved in MSA and $2/(n+1)$ of the demand is moved in weighted MSA. The MSA was selected in this case study because this algorithm is computationally efficient. The time interval between recalculations of the shortest path was set to 15 minutes to be consistent with the interval of OD matrices. The stopping criterion was set to a maximum of 50 iterations or a 0.5% of relative gap.

Field-observed peak hour volumes at critical road sections were obtained from the I-95 existing conditions study report (Kimley-Horn and Associates, 2017). With these peak hour volumes as ground truth, the base OD matrices estimated from Vissim vehicle input information were adjusted using the "Static OD Adjustment Tool" in Aimsun such that traffic volumes resulting from the Aimsun matrix adjustment process matched the field-observed data as closely as possible. The comparison between peak hour volumes assigned for the critical sections by the Aimsun adjusted demand and the field-observed data is provided in Figure 3.

In general, the assigned volumes from the adjusted demand matched field data well (R^2 = 0.97). However, there were a few locations that showed a significant mismatch between the field (or reference) volumes and the Aimsun adjusted volumes. These outliers were all on low volume arterial roads. One possible reason for these outliers comprises the errors already in the Vissim vehicle input data. Also, actuated signal timing plans retrieved from the Vissim model could not be duplicated in Aimsun because of the simplification inherent in the Aimsun mesoscopic signal state generator.

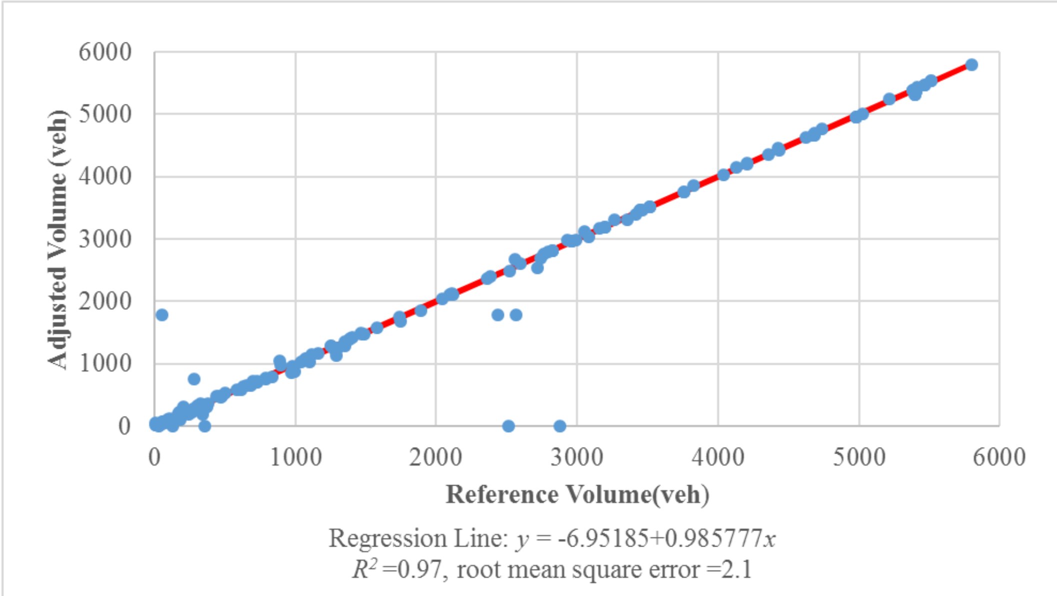


Figure 3. Assigned Peak Hour Volumes vs. Field-Observed Peak Hour Volumes. The red line is the 45 degree line.

Description of Proposed Calibration and Validation Procedure

The proposed procedure consists of seven steps: (1) determine performance measures and evaluation criteria; (2) collect model calibration and validation data; (3) identify calibration parameters; (4) conduct disaggregate level calibration; (5) conduct subnetwork level calibration; (6) conduct system level calibration; and (7) validate the model with external data. The procedure is consistent with the generic framework for the systematic calibration of traffic simulation models proposed by Hellinga (1998).

Step 1: Determine Performance Measures and Evaluation Criteria

Determine Performance Measures

In this substep, performance measures for model calibration and validation are identified that are consistent with the application for which the model is being developed. The performance measure could be an average travel time between two locations in the network or the distribution of speeds, traffic volumes, or travel times on critical segments. Other performance measures may be used depending on the application that exists on the study network. It is important that the specific performance measure(s) selected are consistent with the objectives of the model and can be collected with relative ease both in the field and from the simulation software that will be used.

Determine Evaluation Criteria

In this substep criteria are established by which the adequacy of the model results can be determined. Although there are no universally accepted objective criteria to determine when a model can be considered to be suitably calibrated and fit for the purpose (or validated). there are many non-statistical (e.g., mean absolute percent error, root mean squared percent error, and GEH statistic) and statistical (e.g., t-tests, F-tests, and non-parametric tests) methods that can be used to measure how close the simulation output "matches" field-observed data (Hellinga, 1998; Rilett, 2020).

Hellinga (1998) noted that budgetary constraints and a general lack of field data in sufficient quantity and quality often make it difficult to use the more rigorous statistical approaches. Irrespective of the method used to assess the "goodness-of-fit," personal experience and knowledge of the network by the project team play a crucial role in determining whether the model is fit for the purpose. Several agencies have, often through stakeholder deliberations, established thresholds that serve as guidance for determining the adequacy of model calibration and validation. These thresholds are often geared toward evaluating microsimulation models, and a literature search of state DOT traffic analysis manuals / guidance documents did not identify any published calibration and validation criteria specific to mesoscopic simulation. The absolute percent error and the GEH statistic are two of the common metrics that are used in these guidance documents (Department of Planning, Transport and Infrastructure, 2019; VDOT, 2020). VDOT has established calibration thresholds and criteria for microscopic simulation in VDOT's *Traffic Operations and Safety Analysis Manual* (TOSAM), but no considerations for mesoscopic simulation are currently included (VDOT, 2020).

Step 2: Collect Data for Model Calibration and Validation

In this step, the data needed for calibration and validation are collected. Depending on the performance measures selected in Step 1, there are a number of data elements that can be collected, including link travel times, volumes, speeds, and detector occupancy.

Step 3: Identify Calibration Parameters

Mesoscopic traffic simulation software contains many models (e.g., car following, lane change, gap acceptance, lane selection, and route choice) that the user can modify to mimic local traffic conditions more accurately (Barceló, 2010; PTV AG, 2016). In this step, all relevant parameters are identified and acceptable ranges for each parameter are determined. Reasonable parameter ranges may be determined through a review of the literature and/or sensitivity analyses.

Step 4: Conduct Disaggregate Level Calibration

Flow propagation in commercially available mesoscopic simulation models such as Aimsun and Vissim is often based on a macroscopic speed-density function that is derived from a simplification of the Gipps car following model (Mahut, 2001; Shafiei et al., 2017). In this step, parameters of the speed-density function are calculated (see Eq. 1) for all network links.

Besides using the default parameter values provided by the chosen mesoscopic simulation model, there are a number of ways that the speed-density function parameters may be determined, including the following:

- Fit a regression model to probe/sensor data (speed, occupancy/density) collected for every network segment and aggregated over 5- to 15-minute intervals. This requires large amounts of data that are unlikely to be available for all (or even a majority) of the network links. Besides, since each link can have a separate speed-density function, the total number of parameters to be calibrated is potentially very large.

- For simplicity, and to make the best use of potentially limited data, classify network links into groups based on "similarity" and derive a single speed-density function for each group using data from all members of the particular group.

- Use "typical" or published values from similar studies.

Step 5: Conduct Subnetwork Level Calibration

An assumption inherent in the individual segment level calibration is that there is no interaction between contiguous segments. To account for these interactions, subnetworks consisting of collections of links and nodes with several interactions are selected and evaluated. The purpose of this step is to assess the accuracy of the speed-density function parameters from the disaggregate level calibration and to adjust the parameter values where necessary. If the simulation output "matches" field-observed data for the selected subnetwork(s) links, then the speed-density function parameters are considered appropriately calibrated. Otherwise, the parameters may be adjusted and evaluated again or they may be re-calibrated with all other model parameters at the system level (see Step 6). When successful, the disaggregate and subnetwork level calibration steps can significantly reduce the number of parameters and the effort/resources needed to conduct calibration of all relevant parameters at the entire network level.

In selecting subnetwork(s) for calibration, it is important that the selected subnetwork has limited or no route choice possibilities so that the subnetwork OD demand can be accurately estimated with minimal concern about potential propagation of errors attributable to the OD estimation process (in the presence of route choice). It is also important that there are field data (e.g., link volumes) available on the selected subnetwork to which the simulation output can be compared.

Step 6: Conduct System Level Calibration

In this step, calibration of all relevant parameters (with the possible exception of the speed-density function parameters calibrated in the previous steps) is done at the system or entire network level. System level calibration follows the three-step process proposed by Park and Schneeberger (2003) described here.

1. *Conduct an experimental design.* An experimental design method such as the Latin hypercube (Viana, 2016) is used to identify a collection of possible combinations of the parameter values spanning the parameter space. The use of Latin hypercube sampling helps limit the number of combinations to a practical level while still providing good coverage of the parameter space. For example, a full factorial design using five parameters involves 3,125 possible combinations (5^5) even if only five levels per parameter were considered. In contrast, a Latin hypercube design using five parameters would require only approximately 50 combinations without compromising coverage of the parameter space (Santner et al., 2003). The Latin hypercube sample can be easily generated in statistical software such as JMP, SAS, R (package "lhs"), and Python (packages "pyDOE" and "surrogate modeling toolbox") by specifying lower and upper bounds for the various parameters. The simulation is then run for each sampled combination of parameter values and the output performance measure(s) are noted.

2. *Develop a surface function or metamodel.* The results of the simulation runs are used to develop a response surface function or metamodel that relates the simulation output to the inputs. The dependent variable of the surface function is the performance measure selected in Step 1 and the values of the performance measure are obtained from the simulation runs. The independent variables are the parameters to be calibrated and the values of the parameters are obtained from the experiment design in Step 6.

3. *Evaluate candidate solutions.* The estimated surface function is used to determine several combinations of parameter values that provide output close to the field-observed performance measure. These constitute candidate solutions to the optimization problem. The simulation is re-run for each candidate solution. The combination that minimizes the discrepancy (as measured by the root mean square error, mean absolute percentage error, mean absolute deviation, GEH statistic, etc.) between the simulation output performance measure (travel times, speeds, queue lengths, etc.) and field-observed performance measure is selected.

Step 7: Validate the Model With External Data

In this step, the calibrated model is checked for validity by comparing simulation output with independent field-observed data that were not used as part of the calibration process.

Case Study: Application of Proposed Procedure to Test Bed

A mesoscopic simulation model for the test bed shown in Figure 2 was constructed in Aimsun. The study period was 6:30 to 8:30 AM, with the peak hour set from 7:15 to 8:15 AM. This section demonstrates an application of the proposed procedure to the test bed with a focus on demonstrating the calibration and validation for the peak hour.

Performance Measure and Evaluation Criteria

Two performance measures were selected for the calibration and validation process in this case study. These were selected based on the availability of data. As noted previously, other measures may be selected based on the specific details of the study for which a model is being calibrated. Travel times along I-95, Route 150, and Route 288 on the test bed were obtained from INRIX probe vehicle data. Peak hour travel time on one of these paths, I-95 Northbound between Route 10 (Exit 61) and Route 195 (Exit 74A), was used as the performance measure for model calibration. Peak hour volumes on critical roadway sections and those for critical intersection turning movements were used as performance measures for the validation process. Peak hour volumes on relevant sections were also used for the subnetwork level calibration.

The evaluation criteria used in this case study were the "closeness" of the simulation output to field-observed data as measured by the absolute percent error (VDOT, 2020) and the GEH statistic (Aimsun, 2019; HDR, Inc., 2018). The evaluation criteria generally depend on the purpose of the model. VDOT's TOSAM (VDOT, 2020) established the criteria for microscopic simulation (Table 1). Aimsun adapts the GEH statistic for model validation using traffic volumes (Aimsun, 2019). The GEH statistic is defined as indicated in Equation 2:

$$GEH = \sqrt{\frac{2(m-o)^2}{(m+o)}} \quad \text{[Eq. 2]}$$

where

m = the modelled hourly flow
o = the observed hourly flow. This statistic can be used only for hourly volumes.

In Aimsun, a GEH less than 5 is considered a good fit of the field-observed data. A GEH greater than 10 is considered unacceptable, and a GEH between 5 and 10 indicates that investigation is needed. Similarly, GEH was used for model validation in a microsimulation study for the Nevada DOT (HDR, Inc., 2018); the validation criteria are shown in Table 2.

Table 1. VDOT's TOSAM Microsimulation Model Calibration Criteria

Simulated Measure	Acceptance Threshold
Traffic Volume (vehicles per hour) 85% of the network links and/or turning movement, and a select number of critical links and/or turning movements, as determined by the District Traffic Engineer or his/her designee, shall meet the calibration thresholds	Within ±20% for <100 vph Within ±15% for ≥100 vph to <1,000 vph Within ±10% for ≥1,000 vph to <5,000 vph Within ±500 vph for ≥5000 vph

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020).

Table 2. Nevada DOT's GEH Criteria for Traffic Volume Validation

Simulated Measure	Acceptance Threshold
Traffic Volume (vehicles per hour)	GEH < 5 for >85% of individual links
	GEH < 10 for 100% of individual links
>95% of links/turns shall meet GEH acceptance thresholds	GEH < 5 for >75% of all turns
	GEH < 10 for >95% of all turns

Calibration and Validation Data

Field-observed peak hour volumes at critical road sections were obtained from the I-95 existing conditions study report (Kimley-Horn and Associates, 2017). In addition, peak hour turning movement volumes at 11 critical intersections and entering volumes at 12 other intersections were also obtained from the report. These data were used for calibration and validation.

One year of detector data, including average speed and occupancy collected in 15-minute intervals, was collected from 15 detectors on the I-95 mainline in the study area. These data were used to estimate the speed-density relationship for the disaggregate level calibration. Ramp detector data were collected but not used because speed readings were missing.

Probe speed and travel time from INRIX were collected for sections on I-95, Route 150, and Route 288 in the study area. These data were collected in 15-minute intervals for 2016.

Calibration Parameters

Calibration parameters and their thresholds should be identified at the beginning of a project (Nevada DOT, 2018). Based on the review of literature, seven parameters and their thresholds were identified for calibration in this case study (Appiah et al., 2018; Ciuffo et al., 2013). These parameters were (1) free-flow speed, (2) reaction time, (3) reaction time at traffic light, (4) jam density, (5) maximum give-way time, (6) look ahead distance for freeway sections, and (7) look ahead distance for other roadway types. These parameters are defined as follows:

- Free-flow speed is the speed that a vehicle travels when unimpeded in low volume conditions. Free-flow speeds ranging between 30 mph and 75 mph were considered in this study.

- Reaction time is the time it takes a driver to react to speed changes in the preceding vehicle. It affects the maximum throughput and the queue propagation speed (Oriol, 2018). The default value is 1.2 s. The range of reaction time values used in this study was 0.5 s to 2.0 s.

- Reaction time at traffic light is the time it takes for the first vehicle stopped at a traffic light to react to a change from red to green (Aimsun, 2019). The default value in Aimsun is 1.6 s. The range of values used in this study was 0.5 s to 3.0 s.

- Jam density is a local section parameter that sets the maximum traffic density in the section and it is used to determine whether a lane is full and when a vehicle can join the back of the queue after the first vehicle leaves a full queue (Oriol, 2018). Jam density was assumed to vary between 140 veh/mi/ln and 250 veh/mi/ln.

- Maximum give-way time is a vehicle parameter that sets the maximum time it takes to resolve node events with conflicting vehicle movements when a vehicle is in a give-way situation. This parameter affects the gap-acceptance model and lane-change model. Values in the range of 1.0 s to 50.0 s were considered in this study.

- Look ahead distance is the distance at which vehicles begin positioning for downstream lane changes. It is a turn parameter and the default values vary by road type. The look ahead distance was allowed to vary between 300 ft and 700 ft for freeway sections and 100 ft and 500 ft for all other road types including arterials, ramps, and connectors in this study.

Disaggregate Level Calibration

The speed and occupancy data collected from 15 detectors on I-95 were used to estimate parameters of the dual-regime model (see Eq. 1). Density was estimated using Equation 3:

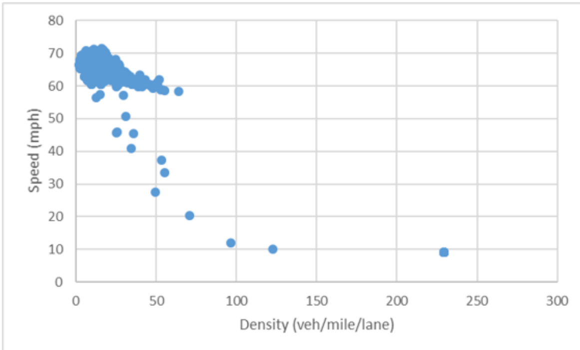
$$k = \frac{52.8}{L_v + L_s} \times o \qquad \text{[Eq. 3]}$$

where

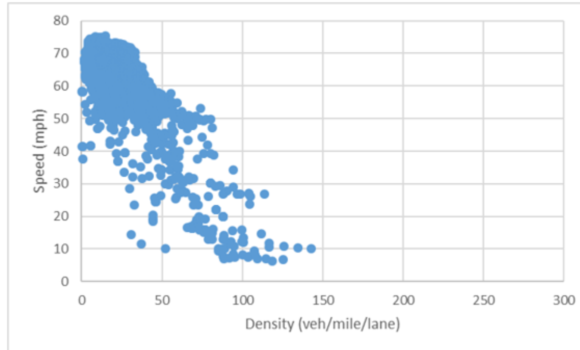
k = density
Lv and Ls = average lengths for vehicles and detectors
o = occupancy (%).

Assuming an Lv of 16.4 ft and an Ls of 6.5ft (Shafiei et al., 2017), the density and speed relationships were plotted for different lane configuration and speed limit combinations as shown in Figure 4.

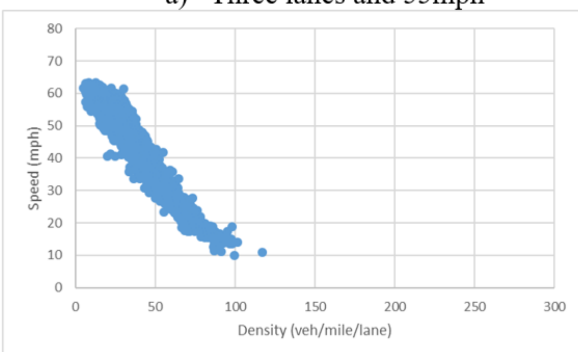
Detector data were available only on the I-95 mainline. The available data for the congested regime were not sufficient, as shown in Figure 4, which could lead to inaccurate calibration results. Therefore, estimated parameters of the dual-regime model for various classes of roadways from Shafiei et al. (2017) were considered for this study. These parameters, shown in Table 3, were estimated using 1 year of data from freeway loop detectors assuming a jam density of 230 veh/mi (143 veh/km). For the lane and speed combinations for which data were available on I-95, the calibrated values from the Shafiei et al. study fit reasonably well. Thus, the Shafiei et al. (2017) values were considered acceptable for this case study. Detector data were not available to develop speed-density relationships on arterial roads; therefore, a reaction time of 1.2 s was used based on a 2018 mesoscopic simulation study for the area near the test bed (Appiah et al., 2018) conducted at VTRC. A summary of the reaction times used in this study is given in Table 4.



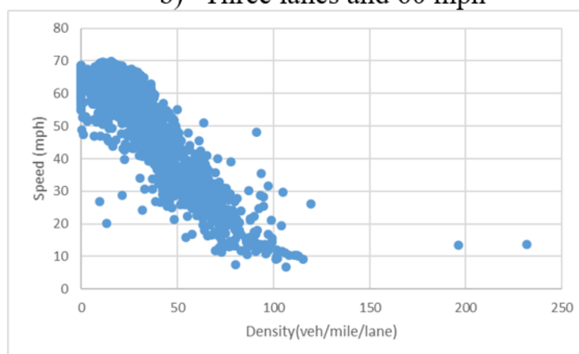
a) Three lanes and 55mph



b) Three lanes and 60 mph



c) Four lanes and 55 mph



d) Four lanes and 60 mph

Figure 4. Speed-Density Relationships

Table 3. Calibrated Parameters in Shafiei et al. (2017)

Lane and Speed Settings	Free-flow Speed (kmph)	Reaction Time (s)
2 lane and 100 kmph	96	1.9
3 lane and 100 kmph	94	1.8
4 lane and 100 kmph	93	1.8
5 lane and 100 kmph	87	2.3
3 lane and 80 kmph	72	1.4
4 lane and 80 kmph	76	1.6

Table 4. Reaction Times Used in This Study

Lane and Speed Settings	Reaction Time (s)
2 lane and 55 mph	1.4
3 lane and 55 mph	1.4
4 lane and 55 mph	1.6
2 lane and 60 mph	1.9
3 lane and 60 mph	1.8
4 lane and 60 mph	1.8
5 lane and 60 mph	2.3
2 lane and 65 mph	1.9
3 lane and 65 mph	1.9
Others	1.2

Subnetwork Level Calibration

A subnetwork consisting of the interchange at I-95 and Route 10 (Figure 5) was selected for demonstration of subnetwork level calibration. This subnetwork was selected because of data availability and because there were limited route choice options. To estimate the local traffic demand of this subnetwork, the centroids were first defined by the sections of the base network with either a starting node or an ending node inside the subnetwork and also by centroids of the base network with at least a connection attached to a node with coordinates lying in the subnetwork or to a section inside the subnetwork (Aimsun, 2019). Then the OD matrices for this subnetwork were generated using the "Static Traversal OD Matrix Generation Tool" in Aimsun.

The simulation of the subnetwork was run with the jam density and reaction time parameters equal to the values from the disaggregate level calibration. The calibration results are shown in Table 5. The simulated link flows were all within 5% of equivalent field values. Therefore, the speed-density parameter values (reaction times and free-flow speeds in this case study) were considered reasonable, so that further adjustment was not needed.

Figure 5. Example of Subnetwork for Calibration

Table 5. Subnetwork Calibration Results

Object	Observed Volume (vph)	Simulated Volume (vph)	Relative Difference (%)
On-ramp to I-95	324	337	-3
Off-ramp to Route 10	169	164	4
Mainline diverge section	4205	4104	2
Mainline between ramps	4036	3952	2
Mainline merge section	4360	4304	1

System Level Calibration

Parameters calibrated at the system level were jam density, reaction time at traffic light, maximum give-way time, look ahead distance for freeways, and look ahead distance for arterials. Peak hour travel time was used as a performance measure for system level calibration in this case study.

Experimental Design

The five parameters to be calibrated were difficult to measure in the field. The number of combinations of reasonable values of these parameters is very large and it is not practical to run simulations for all of these scenarios. The experimental design is used to reduce the number of parameter combinations to be tested while achieving reasonable calibration accuracy. The Latin hypercube space-filling design, a widely used experimental design method, was used in this case study. There are no specific criteria for sample size. A common rule of thumb is to use about 10 points per dimension (Santner et al., 2003). For the five calibration parameters, a minimum of 50 combinations is desired for Latin hypercube design. Fifty cases of the parameter combinations were generated in a Latin hypercube design using JMP statistical software. These parameter combinations are shown in the Appendix.

Development of Surface Function

Peak hour travel time on I-95 Northbound between Route 10 (Exit 61) and Route 195 (Exit 74A) was used as a performance measure for calibration. The average peak hour travel time on this path was 739 s based on INRIX probe vehicles data. Mesoscopic simulation was run for each of the 50 cases and the performance measure from the simulation was noted. Using the five parameters as independent variables and the performance measure as the dependent variable, a surface function was created. Both linear and nonlinear regression models were tested, and the linear regression model expressed in Equation 4 was selected because it was simple and provided a reasonably good fit of the data. The R^2 was modest at 0.4; however; the model was statistically significant overall based on the F-statistic (F = 3.18, p-value = 0.037) (Minitab, 2015).

$$Y = 479.74 + 2.13X_1 - 1.43X_2 + 0.34X_3 + 0.17X_4 + 0.63X_5 \qquad \text{[Eq. 4]}$$

where

Y = the performance measure (simulated travel time on I-95 Northbound)
X_1 = reaction time at traffic light (s)
X_2 = maximum give-way time (s)
X_3 = look ahead distance on freeways (ft)
X_4 = look ahead distance on other roads (ft)
X_5 = jam density (vehicle/mile/lane).

All variables were statistically significant at the 5% significance level except X_1 and X_2, which were significant only at the 10% (p-value = 0.09) and 20% (p-value = 0.18) levels of significance, respectively. The finding that the reaction at traffic light and maximum give-way time variables were not highly statistically significant was not surprising given that the response variable was measured along a freeway facility. Overall, the goodness of fit of this surface function was considered reasonable for the purpose of identifying candidate parameter sets for further evaluation. It is noted that the regression model (Eq. 4) is specific to the test bed and may not transfer to other networks.

Evaluation of Candidate Solutions

With the surface function, it is possible to identify an optimal combination of the five parameters that could produce a good match to the field-observed performance measure. Microsoft Excel's Solver function was used to find candidate solutions (combinations of parameter values) for which the surface function output (Y) equaled the field-observed performance measure (739 s). Five such solutions were identified (see Table 5). Case 6 was one of the design points from the Latin hypercube experimental design and it was selected because of good travel time results (708 s).

The simulation model was run for all six cases; the values of the performance measure are shown in the last row of Table 6. The result of Case 6 was closest to the field-observed travel time, but the differences among all cases were small and the relative difference between simulated and field-observed travel time was within 10% for all cases. Therefore, further evaluation using an independent data set (i.e., model validation) was needed to identify the best case.

Table 6. Model Parameters and Simulated Travel Time for the Candidate Cases

Model Parameter	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Reaction time at stop (s)	1.6	1.6	1	1.8	2	1.2
Maximum give-way (s)	20	35	17	20	10	16
Look ahead distance on freeways (ft)	300	300	300	300	300	692
Look ahead distance on other roads (ft)	100	100	100	100	100	182
Jam density (veh/mi/ln)	230	230	190	230	238	149
Simulated travel time on I-95 N (s)	687	686	661	684	682	708

Model Validation

Peak hour volume was used as a performance measure for model validation in this example. The field-observed peak volumes included the following:

- peak hour volumes on critical links
- peak hour volumes for turning movements at critical intersections
- peak hour entering volumes at other intersections.

Both VDOT's TOSAM criteria (Table 1) and the Nevada DOT study validation criteria (Table 2) were checked for this case study. The validation results are shown in Tables 7 through 9.

Table 7. Critical Links Validation Results

Case No.	Volume Level	No. of Sections	TOSAM Threshold	% Meet TOSAM	% GEH <5	% GEH <10
Case 1	Volume ≥5000	15	Within ± 500 vph	100	100	100
	5000>volume≥1000	72	Within ± 10%	89	89	100
	1000>volume≥100	36	Within ± 15%	83	100	100
	volume<100	1	Within ± 20%	100	100	100
Case 2	Volume ≥5000	15	Within ± 500 vph	100	100	100
	5000>volume≥1000	72	Within ± 10%	87	87	100
	1000>volume≥100	36	Within ± 15%	81	97	100
	volume<100	1	Within ± 20%	100	100	100
Case 3	Volume ≥5000	15	Within ± 500 vph	46	47	47
	5000>volume≥1000	72	Within ± 10%	50	44	69
	1000>volume≥100	36	Within ± 15%	61	76	96
	volume<100	1	Within ± 20%	100	100	100
Case 4	Volume ≥5000	15	Within ± 500 vph	100	100	100
	5000>volume≥1000	72	Within ± 10%	86	89	94
	1000>volume≥100	36	Within ± 15%	81	97	100
	volume<100	1	Within ± 20%	100	100	100
Case 5	Volume ≥5000	15	Within ± 500 vph	100	100	100
	5000>volume≥1000	72	Within ± 10%	86	80	94
	1000>volume≥100	36	Within ± 15%	83	97	100
	volume<100	1	Within ± 20%	100	100	100
Case 6	Volume ≥5000	15	Within ± 500 vph	87	87	100
	5000>volume≥1000	72	Within ± 10%	49	44	68
	1000>volume≥100	36	Within ± 15%	59	70	89
	volume<100	1	Within ± 20%	100	100	100

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020).

Based on the closeness to field-observed data, the parameters from Case 1 in Table 5 were identified as the best option. The results of Case 1 met the GEH criteria used in the Nevada DOT study (HDR, Inc., 2018). The peak hour volumes for critical links and minor intersection approaches also met VDOT's TOSAM microsimulation criteria (VDOT, 2020), but the peak hour turning movement volumes did not. It is worth mentioning that several attempts at recalibration and validation including adjusting the input OD matrices multiple times (using the "Static OD Adjustment Tool" in Aimsun) did not improve the results. Mesoscopic simulation in Aimsun is an event-based simulation where an event occurs when a vehicle is entering or leaving a section or node and the intervening movement is not simulated (Aimsun, 2019). Mesoscopic simulation is much faster for large networks but may provide less accurate results than the microscopic simulation, especially for intersections where vehicle movements can be complicated. For low volume links and turning movements, the calibration was more likely to meet the GEH criteria than the TOSAM criteria. The results for volumes greater than 1,000 vph were mostly consistent for the two criteria. When the number of links / turning movements under a given volume level was very low (less than 10), the calibration was also more likely to meet the GEH criteria.

Table 8. Critical Turning Movement Volumes Validation Results

Case No.	Volume Level	No. of Movements	TOSAM Threshold	% Meet TOSAM	% GEH <5	% GEH <10
Case 1	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	41	Within ± 15%	80	98	100
	volume<100	18	Within ± 20%	78	100	100
Case 2	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	41	Within ± 15%	78	92	100
	volume<100	18	Within ± 20%	73	94	100
Case 3	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	41	Within ± 15%	42	53	60
	volume<100	18	Within ± 20%	72	94	94
Case 4	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	41	Within ± 15%	66	78	89
	volume<100	18	Within ± 20%	61	94	100
Case 5	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	41	Within ± 15%	66	78	90
	volume<100	18	Within ± 20%	67	94	100
Case 6	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	41	Within ± 15%	25	45	62
	volume<100	18	Within ± 20%	72	94	100

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020).

Table 9. Minor Intersection Approach Entering Volumes Validation Results

Case No.	Volume Level	No. of Approaches	TOSAM Threshold	% Meet TOSAM	% GEH <5	% GEH <10
Case 1	5000>volume≥1000	9	Within ± 10%	89	100	100
	1000>volume≥100	16	Within ± 15%	95	100	100
	volume<100	12	Within ± 20%	83	100	100
Case 2	5000>volume≥1000	9	Within ± 10%	89	89	89
	1000>volume≥100	16	Within ± 15%	80	100	100
	volume<100	12	Within ± 20%	83	100	100
Case 3	5000>volume≥1000	9	Within ± 10%	78	89	100
	1000>volume≥100	16	Within ± 15%	81	95	100
	volume<100	12	Within ± 20%	67	100	100
Case 4	5000>volume≥1000	9	Within ± 10%	89	89	89
	1000>volume≥100	16	Within ± 15%	81	100	100
	volume<100	12	Within ± 20%	83	100	100
Case 5	5000>volume≥1000	9	Within ± 10%	44	44	78
	1000>volume≥100	16	Within ± 15%	91	95	100
	volume<100	12	Within ± 20%	83	100	100
Case 6	5000>volume≥1000	9	Within ± 10%	89	100	100
	1000>volume≥100	16	Within ± 15%	66	95	95
	volume<100	12	Within ± 20%	75	100	100

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020).

In addition to system level validation, the simulation results of Case 1 were assessed for several subareas (hereinafter "local networks") of the test bed. This was to ensure that the network was appropriately calibrated and that generally good validation results at a system or global level were not masking poor performance at the local level. The polygons in Figure 6 represent four local networks on the test bed.

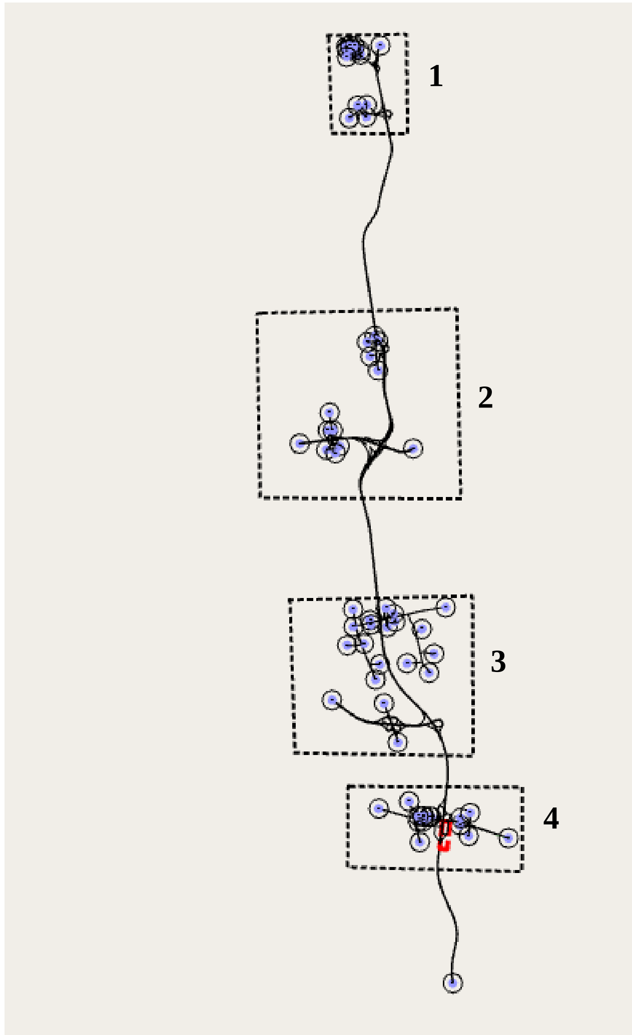


Figure 6. Four Local Networks for Model Validation

The validation results are given in Tables 10 through 12. Similar to the system results, the local network validation results met the GEH criteria but not VDOT's TOSAM criteria, and the turning movement volume results were worse than those for link volumes and intersection approach entering volumes.

Summary

This case study demonstrated a procedure for calibration and validation of mesoscopic simulation models. The proposed procedure includes seven steps: (1) determine performance measures and evaluation criteria; (2) collect model calibration and validation data; (3) identify calibration parameters; (4) conduct disaggregate level calibration; (5) conduct subnetwork level calibration; (6) conduct system level calibration; and (7) validate the model with external data. Using Aimsun, the procedure was applied to a mesoscopic simulation model for the test bed along I-95 in the City of Richmond and State Route 10 in Chesterfield County that included roadways of various functional classes and many different geometric configurations.

Table 10. Local Network Critical Links Validation Results

Local Network	Volume Level	No. of Sections	TOSAM Threshold	% Meet TOSAM	% GEH <5	% GEH <10
1	Volume ≥5000	4	Within ± 500 vph	100	100	100
	5000>volume≥1000	15	Within ± 10%	80	100	100
	1000>volume≥100	7	Within ± 15%	71	100	100
	volume<100	0	Within ± 20%	-	-	-
2	Volume ≥5000	5	Within ± 500 vph	100	100	100
	5000>volume≥1000	20	Within ± 10%	100	100	100
	1000>volume≥100	16	Within ± 15%	87	100	100
	volume<100	0	Within ± 20%	-	-	-
3	Volume ≥5000	4	Within ± 500 vph	100	100	100
	5000>volume≥1000	15	Within ± 10%	100	100	100
	1000>volume≥100	7	Within ± 15%	57	100	100
	volume<100	1	Within ± 20%	100	100	100
4	Volume ≥5000	1	Within ± 500 vph	100	100	100
	5000>volume≥1000	10	Within ± 10%	100	100	100
	1000>volume≥100	6	Within ± 15%	100	100	100
	volume<100	0	Within ± 20%	-	-	-

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020).

Table 11. Local Network Critical Turning Movement Volumes Validation Results

Local Network	Volume Level	No. of Movements	TOSAM Threshold	% Meet TOSAM	% GEH <5	% GEH <10
1	5000>volume≥1000	0	Within ± 10%	-	-	-
	1000>volume≥100	9	Within ± 15%	78	100	100
	volume<100	5	Within ± 20%	80	100	100
2	5000>volume≥1000	0	Within ± 10%	-	-	-
	1000>volume≥100	1	Within ± 15%	100	100	100
	volume<100	4	Within ± 20%	75	100	100
3	5000>volume≥1000	0	Within ± 10%	-	-	-
	1000>volume≥100	15	Within ± 15%	73	100	100
	volume<100	3	Within ± 20%	67	100	100
4	5000>volume≥1000	2	Within ± 10%	50	100	100
	1000>volume≥100	16	Within ± 15%	88	94	100
	volume<100	6	Within ± 20%	83	100	100

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020).

Table 12. Local Network Minor Intersection Approach Entering Volumes Validation Results

Local Network	Volume Level	No. of Approaches	TOSAM Threshold	% Meet TOSAM	% GEH <5	%GEH <10
1	5000>volume≥1000	0	Within ± 10%	-	-	-
	1000>volume≥100	1	Within ± 15%	100	100	100
	volume<100	1	Within ± 20%	0	100	100
2	5000>volume≥1000	1	Within ± 10%	0	10	100
	1000>volume≥100	6	Within ± 15%	83	100	100
	volume<100	4	Within ± 20%	100	100	100
3	5000>volume≥1000	0	Within ± 10%	-	-	-
	1000>volume≥100	8	Within ± 15%	100	100	100
	volume<100	2	Within ± 20%	50	100	100
4	5000>volume≥1000	8	Within ± 10%	100	100	100
	1000>volume≥100	1	Within ± 15%	100	100	100
	volume<100	5	Within ± 20%	80	100	100

TOSAM = VDOT's *Traffic Operations and Safety Analysis Manual* (VDOT, 2020); - = not applicable.

The adequacy of the calibration and validation exercise was evaluated by comparing the simulation output to field-observed data using VDOT's TOSAM criteria and GEH criteria from a Nevada DOT study. The results showed that the proposed procedure produced acceptable results for the case study network. More than 85% of the peak hour volumes on links and at turns where ground truth data were available had a GEH less than 5, and 100% had a GEH less than 10. The simulated travel times were within 20% for average observed travel times on freeways as required in VDOT's TOSAM (VDOT, 2020). The peak hour volumes on critical links and minor intersection approaches met the TOSAM calibration criteria; however, the percentage of peak hour turning movement volumes that met the TOSAM thresholds (79%) was less than the requirement in TOSAM (85%). Multiple attempts were made to recalibrate/validate the model by adjusting the input OD matrices; however, these did not improve the results. Based on the results of this case study, it appears that mesoscopic simulation models may not achieve the same accuracy as the microsimulation models for all network elements (e.g., individual link vs. turning volumes). A separate set of calibration thresholds and criteria, other than those currently available in TOSAM for microsimulation analysis, may need to be developed for mesoscopic simulation model applications.

CONCLUSIONS

- *The proposed calibration and validation procedure appears to be effective based on the results of the case study*. The calibrated model produced acceptable results based on GEH criteria. The model parameters selected through the proposed method produced results that were closer to field-observed data than the default parameters in Aimsun.

- *VDOT's TOSAM (VDOT, 2020) microsimulation calibration thresholds and criteria may not be directly applicable for mesoscopic model application*. The peak hour volumes on critical links and minor intersection approaches met the TOSAM calibration criteria; however, the percentage of peak hour turning movement volumes that met the TOSAM threshold (79%) was less than the requirement in TOSAM (85%) even after several attempts at recalibration.

RECOMMENDATIONS

1. *VDOT's Traffic Engineering Division (TED) should consider establishing separate criteria for mesoscopic model evaluation in future versions of TOSAM*. Currently, VDOT has not established calibration and validation criteria or methods for mesoscopic simulation. The volume threshold criteria for microsimulation in the current version of TOSAM seem to work well for network links and intersection approaches but not as well for intersection turning movements and may need to be adjusted. Metrics other than the percentage difference, such as GEH, could be considered.

IMPLEMENTATION AND BENEFITS

Implementation

To implement Recommendation 1, VDOT's TED will convene a technical advisory committee (TAC), similar to the TOSAM TAC, to discuss fully all issues pertaining to mesoscopic simulation model applications in Virginia, including the establishment of calibration and validation methods and criteria for such models. This group will convene within 1 year of the publication of this report. The outcome of the TAC deliberations will be clear and concise language regarding mesoscopic model applications in TOSAM. VTRC will work with the TED and the TAC to establish criteria for evaluating mesoscopic simulation model applications. The findings in this study are based on a single case study, so additional sensitivity analyses and case studies may be required to establish final calibration and validation thresholds.

The TAC will also discuss calibration procedures for mesoscopic models for potential inclusion in TOSAM. The methods included in this report represent robust techniques to achieve the highest possible level of calibration. They are a balance between the commonly used trial-and-error methods and more complex optimization methods, but they may still be deemed complex by some practitioners. Based on the discussions related to applications and calibration/validation thresholds, it is possible that the TAC will want to pursue less complex methods to achieve mesoscopic calibration/validation in light of adopted thresholds. TED, the TAC, and VTRC will discuss final methods to be included in TOSAM.

Benefits

The benefits of implementing Recommendation 1 will be the consistent and accurate application of traffic analyses using mesoscopic simulation models. Updated mesoscopic simulation criteria in VDOT's TOSAM will help VDOT ensure the quality of work by staff and contractors and also ensure that thresholds are reasonable for mesoscopic models. Better analysis will better meet VDOT's business needs. Implementation of the proposed procedure will help VDOT analysts and consultants calibrate simulation models effectively and thus better utilize mesoscopic models for traffic analysis, especially for large networks where microscopic models may not be practical.

ACKNOWLEDGMENTS

The authors thank the following VDOT personnel who served on the technical review panel for this study: Hyun Cho (VTRC), Jungwook Jun (Transportation Mobility and Planning Division), Sanhita Lahiri (Traffic Engineering Division), Jungtaek Lee (Northern Region Operations), Paul Szatkowski (Operations Division), and Sharad Uprety (Transportation Mobility and Planning Division).

REFERENCES

Aimsun. *Aimsun Next 8.4 User's Manual*. Aimsun SL, 2019.

Appiah, J., Sun, B., and Park, B. *Mesoscopic Modeling Guidance and Pilot Test*. Unpublished Report. Virginia Transportation Research Council, Charlottesville, 2018.

Balakrishna, R., Ben-Akiva, M., and Koutsopoulos, H.N. Offline Calibration of Dynamic Traffic Assignment: Simultaneous Demand-and-Supply Estimation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2003(1), 2007, pp. 50-58.

Barceló, J. Models, Traffic Models, Simulation, and Traffic Simulation. In *Fundamentals of Traffic Simulation, International Series in Operations Research & Management Science*. Springer, New York, 2010, pp. 1-62.

Burghout, W. *Mesoscopic Simulation Models for Short-Term Prediction*. PREDIKT Project Report. 2005.

Casas, J., Perarnau, J., and Torday, A. The Need to Combine Different Traffic Modelling Levels for Effectively Tackling Large-Scale Projects Adding a Hybrid Meso/Micro Approach. *Procedia-Social and Behavioral Sciences,* Vol. 20, 2011, pp. 251-262.

Cascetta, E. *Transportation Systems Analysis: Models and Applications*. Second Edition. Springer, 2009.

Chen, B. Application of Transport Hybrid Modelling for New Town Planning. In *Australian Institute of Traffic Planning and Management (AITPM) National Conference, 2014.* Adelaide, South Australia, Australia, 2014.

Chiu, Y.-C., Zheng, H., Villalobos, J.A., Peacock, W., and Henk, R. Evaluating Regional Contra-Flow and Phased Evacuation Strategies for Texas Using a Large-Scale Dynamic Traffic Simulation and Assignment Approach. *Journal of Homeland Security and Emergency Management*, Vol. 5, No. 1, 2008, pp. 1-29.

Ciuffo, B., Casas, J., Montanino, M., Perarnau, J., and Punzo, V. Gaussian Process Metamodels for Sensitivity Analysis of Traffic Simulation Models: Case Study of AIMSUN Mesoscopic Model. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2390(1), 2013, pp. 87-98.

Daganzo, C. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent With Hydrodynamic Theory. *Transportation Research, Part B*, Vol. 28, No. 4, 1994, pp. 269-287.

de Dios Ortúzar, J., and Willumsen, L.G. *Modelling Transport*. 4th Edition. John Wiley & Sons, New York, 2011.

de Palma, A., and Marchal, F. Real Cases Applications of the Fully Dynamic METROPOLIS Tool-Box: An Advocacy for Large-Scale Mesoscopic Transportation Systems. *Networks and Spatial Economics*, Vol. 2, No. 4, 2002, pp. 347-369.

Department of Planning, Transport and Infrastructure. *Metropolitan Adelaide Traffic Simulation and Assessment Model (MATSAM) Traffic Simulation Model Development Guidelines: Aimsun Next*. South Australia, 2019.

Dixit, V., Ramasamy, S., and Radwan, E. Assessment of I-4 Contraflow Plans: Microscopic Versus Mesoscopic Simulation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2041, 2008, pp. 89-97.

Florian, M., Mahut, M., and Tremblay, N. Application of a Simulation-Based Dynamic Traffic Assignment Model. In *Simulation Approaches in Transportation Analysis-Recent Advances and Challenges.* Springer, 2005, pp. 1-22.

HDR, Inc. *2017 Aimsun Next Model Development and Calibration Report*. Nevada Department of Transportation, Carson City, 2018.

Hellinga, B.R. Requirements for the Calibration of Traffic Simulation Models. *Proceedings of the Canadian Society for Civil Engineering*, Vol. 4, 1998, pp. 211-222.

Hou, T., Mahmassani, H., Alfelor, R., Kim, J., and Saberi, M. Calibration of Traffic Flow Models Under Adverse Weather and Application in Mesoscopic Network Simulation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2391, 2013, pp. 92-104.

Kimley-Horn and Associates, Inc. *I-95 Existing Conditions Vissim Model Development*. Virginia Department of Transportation, Richmond, 2017.

Kristoffersson, I. Impacts of Time-Varying Cordon Pricing: Validation and Application of Mesoscopic Model for Stockholm. *Transport Policy,* Vol. 28, 2013, pp. 51-60.

Kundé, K.K. *Calibration of Mesoscopic Traffic Simulation Models for Dynamic Traffic Assignment*. Ph.D. Dissertation, Massachusetts Institute of Technology, 2002.

Liu, H.X., He, H., and He, B. Method of Successive Weighted Averages (MSWA) and Self-Regulated Averaging Schemes for Solving Stochastic User Equilibrium Problem. *Networks and Spatial Economics,* Vol. 9, No. 4, 2007.

Liu, H.X., Danczyk, A., and Xiaozheng, H. *Development of the Next Generation Metro-Wide Simulation Models for the Twin Cities' Metropolitan Area: Mesoscopic Modeling*. CTS 11-03. University of Minnesota, 2011.

Mahut, M. *A Discrete Flow Model for Dynamic Network Loading*. Ph.D. Thesis, University of Montreal, Canada, 2001.

Minitab. What Is the F-TEST of Overall Significance in Regression Analysis? Minitab Blog, June 11, 2015. https://blog.minitab.com/en/adventures-in-statistics-2/what-is-the-f-test-of-overall-significance-in-regression-analysis. Accessed July 26, 2021.

Nevada Department of Transportation. *Aimsun Next Modeling Guidelines.* Carson City, 2018.

Oriol, L. Meso Over Micro: When Good Meso Calibration Can Save the Need for Microsimulation: Part Two. Technical Note #33. May 2018. https://www.aimsun.com/tech_notes/hybrid-micro-meso-2/. Accessed April 17, 2021.

Park, B., and Schneeberger, J.D. Microscopic Simulation Model Calibration and Validation: Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal System. *Transportation Research Record*: *Journal of the Transportation Board*, No. 1856(1), 2003, pp. 185-192.

Park, B., Pampati, D.M., and Balakrishna, R. Architecture for On-Line Deployment of Dynamit in Hampton Roads, VA. In *Applications of Advanced Technology in Transportation.* American Society of Civil Engineers, 2006, pp. 605-610.

Park, B., Jones, T.K., and Griffin, S.O. *Traffic Analysis Toolbox Volume XI: Weather and Traffic Analysis, Modeling and Simulation*. FHWA-JPO-11-019. Federal Highway Administration, Washington, DC, 2010.

PTV AG. *VISSIM 9 User Manual*. Karlsruhe, Germany, 2016.

Rilett, L.R. Using Simulation to Estimate and Forecast Transportation Metrics: Lessons Learned. In *CIGOS 2019, Innovation for Sustainable Infrastructure*, pp. 23-33. Springer, Singapore, 2020.

Santner, T.J., Williams, B.J., and Notz, W. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.

Shafiei, S., Gu, Z., Sarvi, M., Saberi, M., and Board, T.R. Deployment and Calibration of Large-Scale Mesoscopic Dynamic Traffic Assignment Model of Melbourne, Australia. *TRB 96th Annual Meeting Compendium of Papers*, Washington, DC, 2017.

Sun, B., Appiah, J., and Park, B. Practical Guidance for Using Mesoscopic Simulation Tools. *Transportation Research Procedia*, Vol. 48, 2020, pp. 764-776.

Viana, F.A.C. A Tutorial on Latin Hypercube Design of Experiments. *Quality and Reliability Engineering International*, Vol. 32, No. 5, 2016, pp. 1975-1985.

Virginia Department of Transportation. *Traffic Operations and Safety Analysis Manual (TOSAM)*, Version 2.0. Richmond, 2020.

APPENDIX

CALIBRATION PARAMETERS GENERATION FROM LATIN HYPERCUBE DESIGN

	Reaction Time at Traffic Light (s)	Max. Give-Way Time (s)	Look Ahead Distance on Freeways (ft)	Look Ahead Distance on Other Roads (ft)	Jam Density (veh/mi/ln)
1	0.76	45.00	544.90	328.57	200.61
2	2.39	39.00	308.16	459.18	225.31
3	1.32	9.00	577.55	475.51	178.16
4	0.91	40.00	504.08	116.33	211.84
5	2.03	7.00	553.06	271.43	162.45
6	1.11	1.00	389.80	304.08	214.08
7	1.37	4.00	585.71	148.98	209.59
8	1.62	11.00	340.82	108.16	232.04
9	0.60	33.00	536.73	500.00	171.43
10	1.72	37.00	381.63	491.84	169.18
11	1.83	24.00	430.61	279.59	198.37
12	1.67	29.00	626.53	197.96	196.12
13	2.64	30.00	659.18	361.22	146.73
14	2.69	46.00	651.02	206.12	173.67
15	1.01	21.00	487.76	230.61	243.27
16	2.29	50.00	520.41	369.39	160.20
17	1.98	12.00	397.96	426.53	142.24
18	2.59	14.00	675.51	165.31	236.53
19	2.08	22.00	667.35	385.71	205.10
20	2.80	27.00	495.92	467.35	182.65
21	2.44	6.00	348.98	173.47	184.90
22	0.55	25.00	332.65	214.29	207.35
23	2.18	32.00	463.27	132.65	238.78
24	2.23	8.00	471.43	246.94	241.02
25	1.27	34.00	634.69	353.06	151.22
26	1.06	48.00	610.20	157.14	155.71
27	0.50	20.00	593.88	287.76	187.14
28	2.13	41.00	422.45	124.49	180.41
29	0.65	44.00	316.33	402.04	189.39
30	1.93	13.00	365.31	451.02	227.55
31	1.57	17.00	446.94	100.00	164.69
32	1.42	47.00	373.47	263.27	220.82
33	2.90	23.00	528.57	189.80	193.88
34	1.78	38.00	512.24	442.86	218.57
35	3.00	19.00	324.49	312.24	223.06
36	0.96	28.00	700.00	344.90	229.80
37	1.88	49.00	683.67	336.73	202.86
38	1.47	42.00	618.37	222.45	247.76
39	2.54	36.00	300.00	320.41	166.94
40	0.70	26.00	406.12	434.69	234.29
41	0.86	18.00	357.14	393.88	175.92
42	2.95	15.00	602.04	410.20	250.00
43	2.85	43.00	479.59	295.92	216.33
44	1.52	10.00	569.39	418.37	245.51
45	1.16	35.00	414.29	255.10	144.49
46	2.74	2.00	438.78	377.55	191.63
47	2.34	31.00	561.22	140.82	140.00
48	1.21	16.00	691.84	181.63	148.98
49	0.81	3.00	455.10	238.78	157.96
50	2.49	5.00	642.86	483.67	153.47