RESEARCH REPORT 2

Technical Memorandum on Calibration/Validation of Traffic Microsimulation Models

Prepared for

Orange County Transportation Authority

December 2002


Report prepared by the Institute of Transportation Studies, University of California, Irvine. Authors of this report were MG McNally and Jun-Seok Oh.

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600

Contact:
MG McNally <mmcnally@uci.edu>
949 824-8462
ABSTRACT

Traffic simulation models have been increasingly utilized for the evaluation of traffic control systems and the prediction of future traffic conditions. Recent enhancements in computational capability have led to the development of high fidelity microscopic simulation models and this has accelerated their application in traffic engineering and, increasingly, in transportation planning. Microscopic simulation models, however, have very complex data requirements and numerous model parameters. Although most microscopic simulation programs provide data input guidelines and default model parameters, these models nevertheless need to be calibrated for the specific study area and the intended applications. The purpose of this memorandum is to provide some guidelines for appropriate use of microscopic simulation models. The specific approach is to discuss general issues in model calibration and validation from the perspective of global model parameters as well as from the perspective of location-specific direct adjustments of input data. These two perspectives can each critically influence the performance of the simulation model and the resulting validation exercise. Key issues addressed in this memorandum include overall model calibration/validation procedures. Discussion of specific model parameters to be calibrated, calibration methods, and model validation guidelines.

A subsequent report will provide an application of selected calibration/validation techniques as part of a Paramics case study for the I-5/SR-55 interchange.
Table of Contents

1. Overview ........................................................................................................................ 1
   1.1 Definition of Model Calibration/Validation .......................................................... 1
   1.2 Model Components ............................................................................................... 2
   1.3 Overall Procedure ............................................................................................... 4
   1.4 Model Calibration/Validation Techniques .......................................................... 6

2. Basic Data Input and Error Checking ......................................................................... 8

3. Driving Behavior Model Calibration/Validation ....................................................... 10
   3.1 Model Parameters to be Calibrated .................................................................. 10
   3.2 Data for Model Calibration/Validation ............................................................... 12
   3.3 Measures of Effectiveness (MOEs) .................................................................. 13
   3.4 Calibration Examples .......................................................................................... 14

4. Route Choice Behavior Model Calibration/Validation ........................................... 16
   4.1 Route Choice Behavior Model Parameters ....................................................... 16
   4.2 Data and Measures of Effectiveness for Route Choice Behavior Models. .. 17

5. Stand-alone Model Validation .................................................................................. 19
   5.1 Tuning local parameters ..................................................................................... 19
   5.2 OD demand adjustment ...................................................................................... 20
   5.3 Statistical Analysis Methods .............................................................................. 22
   5.4 Validation Standard ............................................................................................. 25

6. Conclusions ................................................................................................................. 29

References ...................................................................................................................... 30
List of Tables

Table 1 Classification of model components ................................................................. 3
Table 2 Car-following model parameters ................................................................. 11
Table 3 Lane-changing model parameters .............................................................. 11
Table 4 Summary of car-following model calibration results (Paramics) ................. 14
Table 5 Route choice model parameters .................................................................. 17
Table 6 An Example of Validation Guidelines (Milam & Choa, 2001) .................. 26
Table 7 Freeway Model Validation Criteria (Wisconsin DOT, 2002) ..................... 27

List of Figures

Figure 1. Role of Model Validation .............................................................................. 2
Figure 2 Overall Procedure for Model Calibration/Validation ............................. 6
1. Overview

Traffic simulation models have been increasingly utilized for the evaluation of traffic control systems and the prediction of future traffic conditions. Recent enhancements in computational capability have led to the development of high fidelity microscopic simulation models and this has accelerated their application in traffic engineering and, increasingly, in transportation planning. Microscopic simulation models, however, have very complex data requirements and numerous model parameters. Although most microscopic simulation programs provide data input guidelines and default model parameters, these models nevertheless need to be calibrated for the specific study area and the intended applications.

The purpose of this memorandum is to provide some guidelines for appropriate use of microscopic simulation models. The specific approach is to discuss general issues in model calibration and validation from the perspective of global model parameters as well as from the perspective of location-specific direct adjustments of input data. These two perspectives can each critically influence the performance of the simulation model and the resulting validation exercise.

Key issues in this memorandum include:

- Overall model calibration/validation procedure
- Model parameters to be calibrated
- Calibration methods
- Model validation guidelines

1.1 Definition of Model Calibration/Validation

Traffic models are used for predicting traffic conditions and they need to be able to replicate observed traffic conditions before being used for future forecasts. The credibility of the models depends on how well the models are validated. The validation procedure can be differently defined and each the procedures may have the meanings. As described by FHWA (Barton-Aschman Associations, Inc. and Cambridge Systematics, Inc., 1997), the model validation can be explained with four steps: estimation, calibration, validation, and application. Even though these steps are used for validation of conventional four step models, the concept illustrated in Figure 1 can also be applied to the validation of micro-simulation models.

The four steps -- estimation, calibration, validation, and application -- are often used with different meanings and overlap in their objectives. While being consistent with the definitions used in static planning model validation (Barton-Aschman Associations, Inc. and Cambridge Systematics, Inc., 1997), these steps can be defined as follows:
**Model estimation**: Statistical estimation procedures are used to find the values of the model parameters that maximize the likelihood of fitting observed travel data. The focus is on correctly specifying the form of the model and determining the statistical significance of the variables.

**Model calibration**: After the model parameters have been estimated, calibration is used to adjust parameter values until predicted travel matches observed travel pattern in the region.

**Model validation**: In order to test the ability of the model to predict future behavior, validation requires comparing the model predictions with information other than that used in estimating the model. This step is typically an iterative process linked to model calibration.

**Model application**: Although the model may replicate base year conditions, the application of the model to future year conditions and policy options requires checking the reasonableness of projections, so there is a link between application and validation as well. The sensitivity of the models in response to system or policy changes is often the main issue in model application.

![Figure 1. Role of Model Validation](image)

### 1.2 Model Components

Microscopic traffic simulation models include many model components that cover from vehicle and roadway modeling to various ATMIS (advanced traffic management and information systems) application modeling. This memorandum classifies such modeling components into six categories: travel demand models, vehicle modeling, roadway modeling, traffic flow modeling, traffic control systems modeling, and route choice modeling. These components do not comprise an
exhaustive list, but cover most major components being used in microscopic simulation models.

In this classification (Table 1), the first three steps in conventional four-step modeling process are included in the travel demand modeling components. Of course, these steps can be separately included in the microsimulation, but most microscopic simulation models pay more attention to the traffic modeling and traffic assignment. In fact, the early traffic simulation models were mostly oriented towards traffic flow modeling or traffic control systems. Only in recent times, microscopic models have begun to include traffic assignment functionality.

<table>
<thead>
<tr>
<th>Model Components</th>
<th>Models</th>
<th>Corresponding Step in the Static 4-Step Model</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel demand modeling</td>
<td>Departure time choice¹</td>
<td>Trip generation Trip distribution Mode choice</td>
<td>Usually input data</td>
</tr>
<tr>
<td>Vehicle modeling</td>
<td>Vehicle characteristics, emission and energy modeling, AVCSS</td>
<td>-</td>
<td>Usually input data</td>
</tr>
<tr>
<td>Roadway modeling</td>
<td>Network geometry, toll lane, bike lane</td>
<td>Highway network coding</td>
<td>Network</td>
</tr>
<tr>
<td>Traffic flow modeling</td>
<td>Car-following, lane-changing, gap-acceptance models</td>
<td>Link performance function</td>
<td>Core model</td>
</tr>
<tr>
<td>Traffic control systems modeling</td>
<td>Traffic surveillance, signal control, ramp metering, stop and yield signs, VMS</td>
<td>Link performance function</td>
<td>Usually input data or externally programmed</td>
</tr>
<tr>
<td>Route choice modeling</td>
<td>Route choice behavior model</td>
<td>Traffic assignment</td>
<td>Core model</td>
</tr>
</tbody>
</table>

¹) Usually input data are time-dependent OD demand.

These model components can also grouped into three categories according to their use associated with overall model construction and calibration/validation. The first group is components treated as basic input data. This group includes travel demand, network geometry, vehicle data, and traffic control systems. These are mostly basic input data for microscopic simulation. The second category is core models governing vehicular traffic movement, which characterizes the microscopic simulation model. The last group is the route choice modeling that differentiates recent large-scale microsimulation models from the early traffic simulation models. The inclusion of traffic assignment step
in microsimulation promoted microscopic traffic simulation from a traffic engineering tool to transportation applications.

- Group I: includes travel demand, network geometry, vehicle data, and traffic control systems
- Group II: Driving behavior modeling (traffic flow model)
- Group III: route choice behavior modeling (traffic assignment)

Usually the first group is treated as basic input for micro-simulation that does not require any model calibration or validation. Instead, accuracy of data and/or availability of data are more important. However, these inputs directly affect the performance of the microsimulation models, and the accuracy of such inputs needs to be taken into consideration during the model calibration and validation procedures. It should be noted that model calibration/validation is meaningless without such accurate inputs, and these can be adjusted during model validation as well.

The second group, driving behavior modeling (traffic flow modeling), is the core component of the microsimulation model and needs to be calibrated for the specific region. Traffic flow modeling in microsimulation is based on drivers’ driving behavior that could be different depending on the region and/or transportation environment. Most calibration efforts in microsimulation have focused on these driving behavior models.

Presence of the third group makes the calibration/validation procedure complicated because of following three reasons. First, networks requiring such traffic assignment step are usually large-scale. Second, the route choice behavior model is hard to calibrate due to the lack of data. Third, the route choice model directly affects the overall traffic pattern and its effects are mixed with other model components. This is especially true when O/D demand also needs to be estimated within the simulation because O/D demand pattern and the route choice model interact with each other.

1.3 Overall Procedure

This overall procedure of model calibration/validation eventually becomes the overall scheme of building a microsimulation model. This memorandum begins with basic data and includes two calibration and validation procedures: one for traffic flow modeling, the other for route choice and/or OD estimation, as depicted in Figure 2.

The first step is to input basic data. This step affects the overall procedure of the model development process. Inaccurate and inadequate data input will make overall calibration/validation process tedious and time-consuming. This kind of adjustment may ask for repetitions of the overall process.
This memorandum suggests a two-stage calibration/validation procedure. The first stage is to calibrate and validate traffic flow (driver behavior) models, and the second stage is to calibrate and validate route choice behavior model and/or OD estimation if needed. Since the first stage deals with only traffic flow aspect without consideration of network-wide level effect, spatial scope of this calibration stage is small enough to remove the route choice and OD demand effects. During the first stage calibration, global parameters associated with traffic movement, such as car-following and lane-changing, are calibrated within the sub-network level. Once the global parameters are calibrated, the model is validated and adjusted to reflect the local characteristics. This validation needs to be conducted locally (intersection-by-intersection or junction-by-junction) if high level of accuracy is required.

It is recommended that the second stage be conducted only after finishing the first stage calibration/validation. The second stage is to calibrate the route choice behavior model, which is conducted on a network-wide level. The route choice behavior model can be calibrated using either disaggregated data or aggregated data. The OD estimation and/or overall model validation is an iterative process to match simulation results with the aggregated traffic observations, being carried out after finishing model calibration. If the OD estimation process is needed, then all model parameters need to be fixed first since this process is based on the traffic assignment matrix that is affected by any change in simulation input or parameters. Even though most local characteristics are adjusted during the flow model validation process, a final fine-tuning may still be required. This fine-tuning is to reflect network-level local effects.

Model validation is typically an iterative process linked to model calibration. It involves checking the model results against observed data and adjusting parameters until model results fall within an acceptable range of error. Individual validations are used as part of calibration to show that each component reasonably reproduces observed travel characteristics. Validation of the overall models tests the effects of compounding errors. Overall measure of model performance should be viewed with the possibility of error propagation.
1.4 Model Calibration/Validation Techniques

In the traditional process of tuning calibration parameters, selected parameters are adjusted until reasonable (qualitative and quantitative) correspondence between the model and field-measured microscopic and macroscopic behavior is achieved. Such a process of multiple calibrations is a time-consuming process. The trial-and-error approach has been used for model validation, which makes the overall validation process inefficient and tedious. This highlights the need for a high level of automation, which will reduce the workload and the possibility for human error in the parameter tuning process. Therefore, it is important to develop an easy-to-use method for tuning algorithms' parameters. Recently, the genetic algorithm (GA) optimization method has been applied to model calibration.
calibrations (Cheu ET al., 1998; Yang, 1999; Lee ET al, 2001; Ma ET al., 2002). This method approximates a “survival of the fittest” selection regime as follows:

- An initial population of candidate solutions (parameter sets) starts the process.
- Each solution is given a set of relevant attributes (calibration parameter values) within the reasonable range of the search space.
- Each solution’s fitness is then tested (by how well the resulting simulation with those parameter values compares with real-world data).
- Proportions of the “least-fit” solutions are removed from the population.
- Other solution sets “reproduce” with similar properties plus a random “mutation” or “crossover.”

The new population is tested for fitness and the cycle continues until we obtain an acceptable solution (parameter set) with a fitness level above a certain threshold. We are aware of the limitations of GA techniques as well. First, there is no guarantee that this method will result in the global optimum. That is, for a certain simulation model and set of real-world data, the process may converge to a reasonable calibration parameter set; however a very different set with somewhat better correspondence to the real-world behavior might be overlooked. Secondly, the computational complexity of this method may be intractable in many cases, even with today’s high-speed computers. The more parameters being tuned with the GA, the greater the variability in simulation output (which can result in a need for multiple simulation runs). These are well-known deficiencies of GA techniques; however, GA is one of the candidates for calibrations primarily due to their generality and applicability to a wide variety of calibration scenarios involving a wide array of core algorithm categories.
2. Basic Data Input and Error Checking

The first step is to prepare basic input data for microsimulation. In this stage, data accuracy is of the most concern because accuracy of data directly affects the result of model calibration and validation. Data can be classified into four categories: travel demand, vehicle data, network and geometry, and traffic control systems. It is also necessary to check if these data are correct before proceeding to model calibration. That is, checking errors in basic data is the prerequisite for the model calibration. In this memorandum, basic input data are briefly described to help understand overall model calibration and validation.

The travel demand in microsimulation model is time-dependent traffic data or OD demand. In a simple network, the time-dependent traffic data can be directly used from observed traffic volume data. However, for the multiple origin and destination case, the demand needs to be inferred/estimated from observed traffic data or borrowed from the planning OD demand data. At this stage, the OD demand is assumed to be available.

Vehicle data are also basic input in microsimulation modeling. Individual vehicle types are modeled with detailed vehicle characteristics. Some simulation model packages also require distribution of vehicle characteristics. These are considered a basic input for the vehicles in the region.

Network and geometry are basic component of roadway modeling. These are basic input of the transportation network. Even though each microsimulation model differently represents its network and geometry with different level of detail, basic network representations are similar in most simulation models. While many programs use node and link representations, some programs use section and connector representations. Such link or section characteristics directly affect vehicle movement in the simulation, so such local characteristics need to be adjusted during model validation process.

Traffic control systems are major components affecting vehicular movements in microsimulation. Traffic control systems are core elements of microsimulation modeling. Thanks to the modeling of individual vehicle movements, traffic control systems can be operated as real-world traffic system by mimicking traffic control algorithms and corresponding traffic movements. Special attention needs to be paid to the junction traffic control systems, such as intersection traffic control system and ramp metering system, which affect overall traffic movements in the network. Some parameters involved in the intersection traffic control need to be locally adjusted to reflect local intersection characteristics.

These four components are generally considered basic inputs for micro-simulation. Even though some of data need to be adjusted during the model validation stage to reflect local conditions, most of them can be treated as basic inputs that can be observed without any calibration effort.
Inadequate data inputs that affect the overall performance of simulation model can be identified and adjusted during the validation stage, but this kind of adjustment makes calibration/validation tedious. Therefore, it is essential to input accurate and adequate basic data at the beginning for efficient model development. Dowling et al. (2002) provides a simple checklist to verify that input data has been coded accurately. The checklist is classified into three categories as follows:

**Software**
- Check the software and user group websites for known errors.

**Network**
- Check basic network connectivity. Are all connections present?
- Check link geometry (lengths, number of lanes, free-flow speed, facility type, etc.)
- Check intersection controls (control type, control data)
- Check for prohibited turns, lane closure and lane restrictions at intersections and on links.

**Demand**
- Check the identified sources and sinks (zones) for traffic
- Verify zone volumes against traffic counts.
- Check vehicle occupancy distribution (if modeling high occupancy vehicles).
- Check vehicle mix proportions and vehicle characteristic descriptions.
3. Driving Behavior Model Calibration/Validation

As described in the previous section, this memorandum takes a two-stage approach in model calibration and validation. While the second stage deals with network level problems, such as route choice behavior model and OD estimation process, the first stage calibrates driving behavior model parameters that can be done within the sub-network level. It should be also noted that model parameters vary by simulation programs since each simulation program takes different models.

3.1 Model Parameters to be Calibrated

In microscopic simulation model, car-following and lane-changing models govern individual vehicle movements. Table 2 and Table 3 show model parameters associated with car-following and lane-changing behavior models in three representative microsimulation programs: Paramics, VISSIM, and AIMSUN. These are major model parameters to be calibrated for car-following and lane-changing models used in each program. Each program provides a default value with a range of values for each parameter; however, these default values may not be appropriate for the area. As shown in Table 2 and Table 3, each program requires a different set of parameters due to the difference of model applied.
### Table 2. Car-following Model Parameters

<table>
<thead>
<tr>
<th>Program</th>
<th>Parameters</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paramics</td>
<td>. Mean target headway</td>
<td>1.0 second</td>
</tr>
<tr>
<td></td>
<td>. Mean reaction time</td>
<td>1.0 second</td>
</tr>
<tr>
<td></td>
<td>. Queue speed</td>
<td>2 m/s</td>
</tr>
<tr>
<td></td>
<td>. Queue distance</td>
<td>10 m</td>
</tr>
<tr>
<td>VISSIM</td>
<td>Urban traffic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>. Average standstill</td>
<td>2.0 m</td>
</tr>
<tr>
<td></td>
<td>. Maximum deceleration</td>
<td>-9.99 m/s²</td>
</tr>
<tr>
<td></td>
<td>. Additive part of desired safety distance</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>. Multiplicative part of desired safety distance</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>Freeway traffic (10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>. Distance bumper-to-bumper at standstill</td>
<td>1.50 m</td>
</tr>
<tr>
<td></td>
<td>. Headway time</td>
<td>1.30 s</td>
</tr>
<tr>
<td></td>
<td>. Added distance more than the desired safety distance</td>
<td>4.00 m</td>
</tr>
<tr>
<td></td>
<td>. Time to start deceleration process</td>
<td>-12.00</td>
</tr>
<tr>
<td></td>
<td>. Sensitivity parameter for speed difference during following</td>
<td>-0.25, 0.35</td>
</tr>
<tr>
<td></td>
<td>. Parameter for the distance on speed oscilliation</td>
<td>6.00</td>
</tr>
<tr>
<td></td>
<td>. Actual acceleration during the oscillation process</td>
<td>2.00 m/s²</td>
</tr>
<tr>
<td></td>
<td>. Desired acceleration from standstill</td>
<td>1.50 m/s²</td>
</tr>
<tr>
<td>AIMSUN</td>
<td>General</td>
<td></td>
</tr>
<tr>
<td></td>
<td>. Reaction time</td>
<td>0.5 sec</td>
</tr>
<tr>
<td></td>
<td>. Reaction time at stop</td>
<td>1.0 sec</td>
</tr>
<tr>
<td></td>
<td>. Queuing up speed</td>
<td>1.0 m/s</td>
</tr>
<tr>
<td></td>
<td>. Queue leaving speed</td>
<td>4.0 m/s</td>
</tr>
<tr>
<td></td>
<td>2 lane car-following</td>
<td></td>
</tr>
<tr>
<td></td>
<td>. Number of vehicles to be considered</td>
<td>4 vehicles</td>
</tr>
<tr>
<td></td>
<td>. Maximum distance</td>
<td>100 m</td>
</tr>
<tr>
<td></td>
<td>. Maximum speed difference</td>
<td>50 km/h</td>
</tr>
<tr>
<td></td>
<td>. Maximum speed difference on-ramp</td>
<td>70 km/h</td>
</tr>
</tbody>
</table>

### Table 3. Lane-changing Model Parameters

<table>
<thead>
<tr>
<th>Program</th>
<th>Parameters</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paramics</td>
<td>Sign-posting, headway factor, ramp headway factor, minimum ramp time, ramp aware distance</td>
<td>(local</td>
</tr>
<tr>
<td></td>
<td>(local parameters)</td>
<td>parameters)</td>
</tr>
<tr>
<td>VISSIM</td>
<td>Waiting time before diffusion</td>
<td>60.00 s</td>
</tr>
<tr>
<td></td>
<td>Minimum headway (minimum distance to the front vehicle that must be available for a lane change)</td>
<td>0.50 m</td>
</tr>
<tr>
<td></td>
<td>To slower lane if collision time above (minimum time headway towards the next vehicle on the slow lane so that a vehicle on the fast lane changes to the slower lane)</td>
<td>11.00 s</td>
</tr>
<tr>
<td>AIMSUN</td>
<td>Percent Overtake, Percent Recover, On-Ramp model</td>
<td>0.90, 0.95. 3.3</td>
</tr>
</tbody>
</table>
3.2 Data for Model Calibration/Validation

In general, driving behavior models in microsimulation models are formulated as mathematical relationships between microscopic traffic characteristics based on understanding of microscopic behavior. Model calibration is an effort to find model parameters to reflect traffic flow characteristics that can be described by either disaggregated (microscopic) data or aggregated (macroscopic) data. Hence, the model calibration can be carried out either microscopically or macroscopically.

Driving behavioral models are developed from a close investigation of microscopic traffic behavior, so microscopic analysis might be an important source for detailed model calibration. However, such data acquisition is rather difficult and expensive since microscopic data are collected from individual vehicles. These characteristics can be represented by:

- vehicle speed;
- vehicle length;
- vehicle acceleration;
- relative speed: speed difference between two consecutive vehicles;
- time headway: time between the passage of time between two consecutive vehicles;
- time gap: the passage time between the back of leading vehicle and the front of following vehicle;
- distance headway;
- distance gap.

Microscopic data are rarely available due to difficulty of data collection. In driving behavior model studies, the lack of reliable microscopic data has been a significant problem. A few methods have been used for collection of such microscopic data:

- Aerial photogrammetric technique
- Video data collection from roadside or helicopter
- Video image processing
- Instrumented vehicles
- Driving simulator
- GPS based vehicle

One of the early approaches is to use aerial photos taken in short time intervals at a high elevation. This aerial photogrammetric technique (Treiterer, 1975) determining location and speed of a vehicle is very labor intensive technique. Another approach is using video data from the roadside or a helicopter in order to collect information on dynamic continuous relationships. To process these data, video image processing techniques have been widely developed and applied. An alternative approach is to use laboratory-based simulators where driver’s reactions are measured within a fully controlled virtual environment. Even though
Macroscopic data are aggregated over time and measured at cross sections. Most surveillance systems, such as inductance loop detectors, provide macroscopic data. Largely due to ease of data acquisition, macroscopic data have been widely used in model calibration and validation. The macroscopic quantities are:

- flow rate: the number of vehicles passing a cross section per unit of time;
- space mean speed: the harmonic average speed of traffic on cross sections;
- occupancy: the fraction of time occupied by vehicles;
- density: the number of vehicles per unit distance.

A major concern in model calibration/validation is error inherent in the collection of input data. Problems with input data or validation data can lead to erroneous corrections to models that will damage model performance, credibility and results. The reliability of a model validation effort is always constrained by the quality and quantity of validation data available. Source of significant uncertainty or potential error should be identified early in an effective calibration/validation process. It is important to recognize that uncertainty is inevitable, and to avoid confusing precision with accuracy.

### 3.3 Measures of Effectiveness (MOEs)

Although microscopic data provides detailed driving behavior characteristics and useful resources for model development, validating a driving behavior model with microscopic data is a challenging task. In many cases, graphical comparison of the speed change pattern and trajectory was considered to be sufficient for microscopic validation and objective validation was not used because of a strong correlation between successive points on the speed profile or trajectory plots of a vehicle (Benekohal, 1991; Aycin and Benekohal, 1999). Most microscopic comparisons (Ludmann et al., 1997; Chakroborty and Kikuchi, 1999; Fellendorf and Vortisch, 2001) rely on visual comparison of following vehicle’s behavior between observed and predicted. Representative characteristics compared for microscopic model validation are:

- Distance vs. speed difference during starting and stopping process
- Speed difference vs. headway between the vehicles
- Speed profile of following vehicle under given leading vehicle’s speed profile
- Acceleration/deceleration rate of following vehicle

In this comparison, however, it should be noted that the observed headway distribution could be distorted by lane changing actions as Cohen (2001) pointed out.

The macroscopic validation is rather straightforward in its comparison. It compares macroscopic traffic characteristics between observed and predicted. Commonly used traffic characteristics are:

- Speed and flow relationship
- Distribution of traffic volume by lane
- Speed distribution
- Average speed
- Travel time
- Density
- Throughput

### 3.4 Calibration Examples

This report summarizes some of these calibration efforts on Paramics car-following model. In Paramics car-following model calibration, the main parameters are mean target headway and mean reaction time. These examples used macroscopic traffic data from loop detectors and found parameters that minimize the deviation between real-world observation and simulation output.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean target headway (seconds)</th>
<th>Mean reaction time (seconds)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default values</td>
<td>1.000</td>
<td>1.000</td>
<td>UK</td>
</tr>
<tr>
<td>Trial and error (Abdulhai et al., 1999)</td>
<td>1.650</td>
<td>0.420</td>
<td>I-405, California</td>
</tr>
<tr>
<td>Genetic Algorithm (Lee et al., 2001)</td>
<td>0.625</td>
<td>0.415</td>
<td>I-5, California</td>
</tr>
<tr>
<td>Empirical trials (Gardes et al., 2002)</td>
<td>1.00</td>
<td>0.60</td>
<td>I-680, California</td>
</tr>
<tr>
<td>GENOSIM (Ma et al., 2002)</td>
<td>0.83</td>
<td>0.66</td>
<td>Downtown, Toronto, Canada</td>
</tr>
</tbody>
</table>
The car following and lane changing models on Paramics were developed over 5 years to show close correlation to British data as well as visual observation using video and the ‘mind’s eye’ judgment. The default values for both the mean headway and the mean reaction time are set to 1 second. Abdulhai et al. (1999) simultaneously calibrated the mean headway and reaction time by applying an iterative two dimensional search process that minimizes the discrepancy between the simulation output and field observations. The calibrated values were 1.65 for mean headway and 0.42 for mean reaction time. Later, Lee et al. (2001) applied a genetic algorithm for model calibration using 30-second loop-detector volume and occupancy on one-mile section of freeway I-5, California. The fitness function in GA was a combination of average relative error of flow and occupancy between simulation output and field data. The population size for each generation was 20 and the number of generation was 12. The final calibration results were 0.415 for mean reaction time and 0.615 for mean target headway, which were much lower than the default value 1.0. As another empirical effort, Gardes et al. (2002) calibrated Paramics car-following parameters based on data from I-680, California. They found a range of combinations of target headway and reaction time values that give acceptable results for their maximum volume and average speed criteria. They chose one set of parameters: a mean headway of one second and a mean reaction time of 0.6 second. They validated the Paramics model by comparing speed contour maps between simulation and field data via visual comparison and the chi-square test. They also observed that the simulation produced acceptable speed-flow relationships. As for an arterial case, Ma and Abdulhai (2002) calibrated the model using GENOSIM, a tool that they developed by incorporating Genetic Algorithm. They used the global relative error as the misfit function for the GA, and calibrated mean headway, mean reaction time, dynamic feedback interval, the level of familiarity, and the level of perturbation.

\[
GRE = \frac{\sum_{i=1}^{n} |Q_{real,i} - Q_{sim,i}|}{\sum_{i=1}^{n} Q_{real,i}}
\]

where:
- \(Q_{real,i}\) = actual value of turning counts at each observation location
- \(Q_{sim,i}\) = simulation value of turning counts at each observation location
- \(n\) = number of time-space points
4. Route Choice Behavior Model Calibration/Validation

Modeling the route choice behavior is one of the most challenging tasks in transportation demand modeling. In static traffic assignment, user equilibrium or stochastic user equilibrium concept has been widely accepted and applied. However, in the dynamic traffic assignment process that involves non-stationary property of travel time, it is questionable whether this kind of equilibrium approach is applicable under the situation when real-time travel time information is not available. It is also crucial to understand how drivers perceive their travel times that change over time.

In the network level traffic analysis, traffic pattern is, in fact, governed by the traffic assignment process that is modeled as a form of drivers’ route choice behavior model. However, there are many difficulties in calibrating modeling drivers’ route choice behavior model under congested traffic condition.

This report does not provide details of route choice behavior model calibration. Rather, this report focuses more on characteristics of the route choice model in microsimulation and how to set model parameters.

4.1 Route Choice Behavior Model Parameters

Route behavior models govern traffic assignment in microsimulation. As shown in Table 5, each simulation model provides several traffic assignment methods, and parameters to be calibrated are dependent on the assignment method applied. Unlike Paramics and AIMSUN that calculate their paths internally, VISSIM requires external path input or turning fraction data, which does not require route choice model calibration but path flow pattern. The path flow pattern for VISSIM can be obtained from static traffic assignment or dynamic traffic assignment that requires iterative solution seeking procedure.

Both Paramics and AIMSUN employ basically similar assignment methods in the sense that they allow user to use fixed, stochastic, and dynamic routing. Compared to Paramics that uses direct perturbation error, AIMSUN’s strength is in the employment of econometric models in its route choice behavioral modeling.

Prior to route choice model calibration, users need to decide what type of traffic assignment methods is appropriate for the area. While the stochastic and dynamic feedback type is most appealing in that experience drivers’ capability and drivers’ misconception can be incorporated, finding the appropriate method and parameters is one of the most challenging tasks in microsimulation modeling.
Table 5. Route Choice Model Parameters

<table>
<thead>
<tr>
<th>Program</th>
<th>Assignment methods</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paramics</td>
<td>All-or-nothing</td>
<td>Percentage of familiar drivers</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>Generalized cost coefficients (time, distance, toll)</td>
</tr>
<tr>
<td></td>
<td>Dynamic feedback</td>
<td>Smoothing factor for feedback costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Perturbation error in stochastic assignment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feedback interval for dynamic feedback</td>
</tr>
<tr>
<td>VISSIM</td>
<td>Turning fraction-based routing</td>
<td>(For iterative dynamic assignment)</td>
</tr>
<tr>
<td></td>
<td>Predefined path-based routing</td>
<td>Evaluation interval</td>
</tr>
<tr>
<td></td>
<td>Iterative dynamic assignment</td>
<td>Smoothing factor for iteration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>General cost parameter (time, cost, other)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitivity parameter of Kirchhoff distribution parameter for route choice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logit scaling factor and lower limit for parking lot choice</td>
</tr>
<tr>
<td>AIMSUN</td>
<td>Fixed routing</td>
<td>Percentage of vehicles guided</td>
</tr>
<tr>
<td></td>
<td>Variable routing</td>
<td>Update cycle</td>
</tr>
<tr>
<td></td>
<td>Dynamic routing (en-route decision)</td>
<td>Number of intervals</td>
</tr>
<tr>
<td></td>
<td>Route choice models</td>
<td>Capacity weight</td>
</tr>
<tr>
<td></td>
<td>. Binomial</td>
<td>Maximum number of routes</td>
</tr>
<tr>
<td></td>
<td>. Multinomial</td>
<td>Parameters of model (Binomial, Multinomial, C-logit)</td>
</tr>
<tr>
<td></td>
<td>. C-logit</td>
<td></td>
</tr>
</tbody>
</table>

4.2 Data and Measures of Effectiveness for Route Choice Behavior Models

In general, route choice behavior model can also be calibrated either microscopically or macroscopically. The microscopic route choice behavior has been analyzed via the survey on drivers' route choice. Acquisition of the microscopic (disaggregate) route choice data, however, is a very expensive and difficult task. Survey data for disaggregate route choice behavior include:

- A set of available alternative routes
- Drivers' perceived travel time and cost on the alternative routes
- Driver characteristics

However, except the parameter for route choice model (Binomial, Multinomial, C-logit) in AIMSUN, most parameters in route choice model need to be defined by users, rather than being calibrated. Even though conventional disaggregate approaches have been widely used in analyzing route choice behavior, their results may not be directly applied to microsimulation. Therefore, the route choice model calibration relies more on macroscopic analysis.
The macroscopic analysis is mostly based on traffic counts or splits observed between alternative routes. Data for model calibration and validation could be either sample fraction between alternatives or point measurements. The sample fraction between alternatives can be obtained via license plate survey or advanced vehicle tracing technologies; however, such efforts have seen only limited application for model validation purposes. Instead, most validation efforts were utilizing point traffic counts that are readily available from loop detectors.
5. Stand-alone Model Validation

The process of validation consists of comparison between model outputs and measured real-world behavior. Such comparisons can be done qualitatively and quantitatively. The qualitative approach consists of static and animated Turing tests. A static Turing test involves side-by-side graphical comparison, and the animated Turing test is an effective and intuitive way of finding anomalous model behavior (Rao et al., 1998). While the qualitative approach provides relies on users’ intuitive decision via visual comparison, the quantitative approach is based on measurements from simulation results and real-world observations. This section deals with a model validation process based on quantitative approach, including tuning local parameters, OD adjustment, statistical analysis methods, and validation standard.

In microsimulation modeling, a stand-alone model validation represents an overall validation of the individually calibrated/validated models, so the error at this stage may be compounded. Traffic assignment, human model integration, and O/D estimation are systematically related to each other. However, the main focus of the model validation at this stage is to reflect local traffic conditions and adjust OD demands assuming that the individual models are already correctly calibrated and validated. That is, the OD adjustment can be regarded as the last parameter adjustment.

5.1 Tuning local parameters

There are many local parameters affecting vehicles’ movements. After fixing global parameters, local parameters need to be adjusted in order to reflect local characteristics. These local parameters are complementary inputs for the basic network data, being used to change specific behavior at links, on lanes or at junctions.

In Paramics, usually the major adjustments to a model include moving kerb and stop-line control points, and coding forced lane changes to override default Paramics lane changing. The main link, lane, and junction parameters include:

- Changing the hazard warning distance (sign-posting)
- Including link gradients
- Coding junction visibility
- Changing link headway factor
- Coding link end speeds
- Coding lane end stop time
- Forced lane changes
- Forced merges
- Stay in lane
- Lane and turn restrictions
In AIMSUN, section parameters influence all vehicles, regardless of type, when driving in a particular section of the network. Examples of parameter influences include the following:

- **Speed Limit influences:**
  - Average Speed (higher speed limit => higher average section speed)
  - Travel Times: (higher speed limit => lower average section travel time)

- **Turning Speed influences:**
  - Turning Capacity (higher Turning speed => higher capacity)
  - Travel Times (higher Turning speed => lower travel times)
  - Average Speed (higher Turning speed => higher average speed)

- **Visibility Distance influences:**
  - Yield (Give Way) Sign Behavior

- **Distance Lane Changing Zones influences:**
  - Turning proportions
  - Blocking situations

- **Distance On-Ramp influences:**
  - On-ramp Capacity
  - Use of lane as slow lane

### 5.2 OD demand adjustment

Traditionally, the O-D trip tables are estimated through large-scale surveys such as home interview surveys, roadside interviews, license plate surveys, etc. These survey techniques are expensive, time-consuming, and labor-intensive. In conventional travel modeling, OD demand is considered an initial input data obtained from travel survey. In microsimulation, OD demand can also be treated as initial input. However, due to the difficulty in estimating time-dependent (dynamic) OD demand, many researchers have tried to develop methods estimating time-dependent OD demand. In fact, OD demand is the most influential input data in network-wide traffic simulation. Despite numerous studies on this time-dependent OD estimation, this problem remains still one of the most challenging research problems. This section does not provide complete methodology for OD estimation, but reviews the problem and some of methodologies in literature.

The OD estimation problem is basically identical to the validation process in that comparing the predicted traffic measures to the observed ones. The simplest form of the problem can be defined as an error minimization problem. The true (observed) traffic measures ($X$) can be expressed as:

$$ X = A \cdot Q + \varepsilon $$

where:

- $X$ = observed traffic measure
- $A$ = traffic assignment pattern matrix
The problem is defined as a minimization problem that minimizes the sum of squared differences between estimates and observations.

\[ \min \varepsilon = (X - AQ)'(X - AQ) \]

Then, the least square estimate of OD demand (Q) can be expressed is given by

\[ \hat{Q} = (A'A)^{-1}A'X. \]

In the OD estimation problem, the traffic assignment matrix (X) is the main variable, and the problem can be differently categorized depending on the traffic assignment matrix (X). In microsimulation, the OD estimation problem can be sub-divided into categories associated with the traffic assignment logic.

- Static OD estimation with predefined traffic assignment pattern
- Time-dependent OD estimation with predefined traffic assignment pattern
- Time-dependent OD estimation with time-dependent traffic assignment pattern

If the dynamic feedback is not applied (network costs are not updated during simulation) and the observed measurements are aggregated measures (e.g. one hour volume in a one-hour simulation), the problem becomes a simple steady-state linear problem. This is the simplest form of OD estimation problem. The second case is an extension of the first problem by taking the time-dependent effect into consideration. To account for the time-dependent effect, a number of researchers have examined the time-dependent O-D trip estimation problem with the assumption of proportional assignment. These methods basically extend the entropy maximization (Willumsen, 1984) and generalized least squares (Cascetta, 1993) from steady state to time-dependent case or the Kalman filtering (Okutani, 1987; Ashok and Ben-Akiva, 1993, 2000) from a simple network to a general network with route choice. In a general network, the last case, the proportional assignment assumption does not hold because route choice proportions and the O-D trip table are interdependent. There exists an inconsistency between the assignment assumptions made in the O-D estimation process and the O-D trip table assigned to the network. Thus, it is necessary to introduce a route choice model into the O-D trip table estimation process to resolve this inconsistency. To solve this problem, there have been three approaches: extension of the bi-level programming approach (Janson and Southworth, 1992), extension of the linear path flow estimator (Sherali and Park, 2001), and extension of the nonlinear path flow estimator (Bell et al., 1996; Bell and Grosso, 1998).
5.3 Statistical Analysis Methods

The quantitative analysis requires statistical analysis because all measured traffic data as well as the outputs of many simulators are subject to random variability. Simple statistical measures are not sufficient for more insightful validation efforts. The primary characteristic of a traffic simulation model is that the modeled outputs are general time series, which implies that the outputs may exhibit correlation structures. The objective is always to have simulation outputs close to measured data for the same context. A general formula to compute deviations is:

\[ E = \sum_{i=1}^{j=1} \sum_{k=1}^{m} (\alpha E_i^2(i, j, k) + \beta E_2^2(i, j, k) + \gamma E_3^2(i, j, k)) \]

Here \( E_i \) is the error between simulated and field outputs; \( i, j, \) and \( k \) are indices for variables such as, perhaps, lane location, link and time of day and \( E \) is the accumulated Least Square Error. If the validation is in a large network context, fewer variables than for say all the lanes and links can be selected than what is in the complete network context. Also when validation is done at a macro level in our proposed scheme for integrated-testing, we may be interested in link volumes and not interested in specific lane volumes, for instance. In another case the time dimension may be less important. The \( \alpha, \beta, \) and \( \gamma \) parameters are weights that can be normalized to the maximum possible value for each \( E_i \).

This memorandum introduces five statistical measures that recently introduced in previous studies (Transport Simulation Systems, 2000; Rao and Owen, 2000):

- Root mean square error
- Theil’s U-Statistic
- Two-sample t-test
- Two-dimensional two-sample K-S test
- Error analysis using ARIMA (autoregressive integrated moving average)

Validation MOE is based on the observation that the measured and the simulated series, \( v_{ij} \) and \( w_{ij} \) respectively, are time series. In that case the quality of the simulation model can be established in terms of the quality of the prediction, and that would mean resorting to time series forecasting techniques. A way of achieving that could be through regression analyses. In this case, the error of the \( j \)-th “prediction” for the observation point \( i \) is \( d_i = w_{ij} - v_{ij}, j=1,\ldots,m \), then a typical measure of effectiveness of the predictions for the observation point \( i \) is the Root Mean Square Error (RMSE) defined by:

\[ rmse_i = \sqrt{\frac{1}{m} \sum_{j=1}^{m} (w_{ij} - v_{ij})^2} \]

This MOE is perhaps the most widely used in traffic simulation: the smaller \( rmse_i \) is, the better the model is. However, it has a quite important drawback, because
it squares the error and so emphasizes large errors. It would be helpful to have a measure that considers both the disproportionate weight of large errors and also provides a basis for comparison with other methods.

Theil’s U-Statistic is a measure of effectiveness that achieves these objectives. In general, if \( X_j \) is the observed and \( Y_j \) the forecasted series, \( j = 1, \ldots, m \), then Theil’s U-Statistic is defined as:

\[
U = \frac{\sum_{j=1}^{m-1} (\text{FRC}_{j+1} - \text{ARC}_{j+1})^2}{(m - 1)}
\]

where:
\[
\text{FRC}_{j+1} = \frac{Y_{j+1} - Y_j}{Y_j} \quad \text{is the forecasted relative change, and}
\]
\[
\text{ARC}_{j+1} = \frac{X_{j+1} - X_j}{X_j} \quad \text{the actual relative change.}
\]

Theil’s U-Statistic has an immediate interpretation, as follows:

- \( U = 0 \) \iff \( \text{FRC}_{j+1} = \text{ARC}_{j+1} \), then the forecast is perfect
- \( U = 1 \) \iff \( \text{FRC}_{j+1} = 0 \), the forecast is as bad as possible

Therefore, the closer to zero Theil’s U-Statistic is, the better the forecasted series is or, in other words, the better the simulation algorithm. When Theil’s U-statistic is close to 1 the forecasted series, and therefore the simulation algorithm, should be rejected.

When the forecast efficiency is based on the regression model \( E(v|w) = \beta_0 + \beta_1 w + \varepsilon \) (with \( \varepsilon \) a random error term) the most efficient forecast would correspond to \( \beta_0 = 0 \) and \( \beta_1 = 1 \); this can be tested by the application of variance analysis to the regression model as indicated earlier. Taking into account that the average squared forecast error:

\[
D_m^2 = \frac{1}{m} \sum_{j=1}^{m} (Y_j - X_j)^2
\]

can be decomposed (cf. Theil) in the following way:

\[
D_m^2 = \frac{1}{m} \sum_{j=1}^{m} (Y_j - X_j)^2 = (\bar{Y} - \bar{X})^2 + (S_Y - S_X)^2 + 2(1 - \rho)S_Y S_X
\]

where \( \bar{Y} \) and \( \bar{X} \) are the sample means of the forecasted and the observed series respectively, \( S_Y \) and \( S_X \) are the sample standard deviations and \( \rho \) is the sample correlation coefficient between the two series, the following indices can be defined:
Here $U_M$ is the “Bias proportion” index, which can be interpreted in terms of a measure of systematic error. $U_S$ is the “variance proportion” index, which provides an indication of the forecasted series ability to replicate the degree of variability of the original series or, in other words, the simulation algorithms’ ability to replicate the variable of interest of the actual system. Finally $U_C$ or “Covariance Proportion” index is a measure of the unsystematic error. The best forecasts, and hence the best simulation model, are those for which $U_M$ and $U_S$ do not differ significantly from zero and $U_C$ is close to unity. It can be shown that this happens when $\beta_0$ and $\beta_1$ in the regression do not differ significantly from zero and unity respectively.

Recent studies (Rao et al., 1998; Rao and Owen, 2000) propose three levels of statistical tests for quantitative approach. In the first level, they suggested two-sample $t$-tests between the simulation and the real-world data. They suggested two-dimensional two-sample Kolmogorov-Smirnov (K-S) test as the second-level test. Particularly a univariate non-seasonal autoregressive-integrated-moving-average (ARIMA) modeling approach is proposed in the third level statistical test. Three levels of statistical tests are briefly described here. Details of computational algorithms can be found from Rao et al. (1998) and Rao and Owen (2000).

The first level is to compare averages of standard traffic flow characteristics (e.g. speed, density, and volume) between the simulation and the real-world observation. A two-sample $t$-test is conducted based on the average measures of multiple simulation results. The second level is a nonparametric K-S test that does not require explicit distributional assumptions about the underlying processes. The significance level of an observed value $d$ of the test statistics $D$ is given approximately by the formula:

$$ \text{Probability}(D > d) = Q_{KS} \left( \frac{\sqrt{Nd}}{1 + \sqrt{1 - r^2} (0.25 - 0.75 / \sqrt{N})} \right) $$

where

$Q_{KS} = a$ complex monotonic function

$D = K-S$ statistic, defined as the maximum value of the absolute difference between the two two-dimensional cumulative distribution functions;
\[ N = \frac{N_1 N_2}{N_1 + N_2}, \text{ where } N_1 \text{ and } N_2 \text{ are sample sizes;} \]

and where \( r \) is the sample correlation coefficient between the two distributions. This probability expression can be used to test the agreement between the real world and the simulation data.

Most simulation output data over time are correlated with each other and the underlying output processes of almost all simulation models are non-stationary in nature. These two properties make it difficult to apply classical statistical methods. To avoid such issues, a univariate non-seasonal time-series model, ARIMA, is introduced as the third level test. An error analysis involves an analysis of the differences between two time series resulting from a simulation model and the simulated system:

\[
\varepsilon(t) = \left[ X_1(t) - X_2(t) \right] / X_2(t) \times 100
\]

where

- \( t \) = integer that refers to time when an observation occurs,
- \( X_1(t) \) = observation of user-specified MOE at time \( t \) from model,
- \( X_2(t) \) = observation of MOE at time \( t \) from real world, and
- \( \varepsilon(t) \) = error (in percentage) at time \( t \).

For convenience, it is assumed that the original time \( t = 0 \). Let \((0, T)\) be the time period for historical data; \( \varepsilon(0, T) \) represent a historical error time series; \((T, T+H)\) be the time horizon of analysis (THA), which represents the prediction part of the process; and \( \varepsilon(T, T+H) \) be the errors for the THA. In the proposed error analysis, \( \varepsilon(0, T) \) are used to generate an ARIMA model that characterizes a historical error time series, and \( \varepsilon(T, T+H) \) are then computed on the basis of the model. The range of \( \varepsilon(T, T+H) \) represents the interval characterization of the historical error for the THA. Thus a methodology can be constructed in which, if the range is within a user-specified acceptable range of error (ARE), then confidence is gained in the model.

### 5.4 Validation Standard

Under ideal conditions, the calibration of individual components of a simulation model will improve the simulation model’s ability to replicate traffic flow results that match field conditions within an acceptable range of error. Traffic characteristics and statistical analysis methods that can be used in model validation have been discussed in previous sections. These MOEs and statistical methods might be a direction to set a validation guideline that defines the acceptable range of error for these characteristics. It will be essential to develop
a validation guideline for microsimulation models as applications of microsimulation become popular in transportation planning and traffic analyses.

An example of validation guideline can be found from Milan and Choa (2001). They have set a validation guideline based on some critical measures as in Table 6. This guideline may be a good starting point for discussing guidelines of the microsimulation even though the guideline lacks statistical justification to determine if they provide an acceptable range of error.

**Table 6. An Example of Validation Guidelines (Milam & Choa, 2001)**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Validation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume Served</td>
<td>Percent difference between input volume and model output or assigned volume</td>
<td>95 to 105 percent of observed value</td>
</tr>
<tr>
<td>Average Travel Time</td>
<td>Standard Deviation between floating car average travel times and model simulated average travel time for a series of links</td>
<td>1 Standard Deviation</td>
</tr>
<tr>
<td>Average Travel Speed</td>
<td>Standard Deviation between floating car average travel speed and simulated average travel speed for individual links</td>
<td>1 Standard Deviation</td>
</tr>
<tr>
<td>Freeway Density</td>
<td>Percent difference between observed freeway density (from volume counts and floating car travel speed) and simulated density</td>
<td>90 to 110 percent of observed value</td>
</tr>
<tr>
<td>Average and Maximum Vehicle Queue Length</td>
<td>Percent difference between observed queue lengths and simulated queue lengths</td>
<td>80 to 120 percent of observed value</td>
</tr>
</tbody>
</table>

Recently Wisconsin Department of Transportation (2002) has suggested criteria for freeway model validation. The guideline is based on the GEH statistic computed as follows:

\[
GEH = \frac{(V - C)^2}{(V + C)/2}
\]

where

- \(GEH\) = The statistic
- \(V\) = model estimated directional hourly volume at a location.
- \(C\) = directional hourly count at a location.
### Table 7. Freeway Model Validation Criteria (Wisconsin DOT, 2002)

<table>
<thead>
<tr>
<th>Criteria &amp; Measures</th>
<th>Acceptability Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hourly Flows, Model vs. Observed</strong></td>
<td></td>
</tr>
<tr>
<td>Individual Link Flows</td>
<td></td>
</tr>
<tr>
<td>Within 15%, for 700 vph &lt; Flow &lt; 2700 vph</td>
<td>&gt; 85% of cases</td>
</tr>
<tr>
<td>Within 100 vph, for Flow &lt; 700 vph</td>
<td>&gt; 85% of cases</td>
</tr>
<tr>
<td>Within 400 vph, for Flow &gt; 2700 vph</td>
<td>&gt; 85% of cases</td>
</tr>
<tr>
<td>Total Link Flows</td>
<td></td>
</tr>
<tr>
<td>Within 5%</td>
<td>All Accepting Links</td>
</tr>
<tr>
<td>GEH Statistic – Individual Link Flows</td>
<td></td>
</tr>
<tr>
<td>GEH &lt; 5</td>
<td>&gt; 85% of cases</td>
</tr>
<tr>
<td>GEH Statistic – Total Link Flows</td>
<td></td>
</tr>
<tr>
<td>GEH &lt; 4</td>
<td>All Accepting Links</td>
</tr>
<tr>
<td><strong>Travel Times, Model vs. Observed</strong></td>
<td></td>
</tr>
<tr>
<td>Journey Times Network</td>
<td></td>
</tr>
<tr>
<td>Within 15% (or one minute, if higher)</td>
<td>&gt; 85% of cases</td>
</tr>
<tr>
<td><strong>Visual Audits</strong></td>
<td></td>
</tr>
<tr>
<td>Individual Link Speeds</td>
<td></td>
</tr>
<tr>
<td>Visually acceptable Speed-Flow relation</td>
<td>To analyst’s satisfaction</td>
</tr>
<tr>
<td>Bottlenecks</td>
<td></td>
</tr>
<tr>
<td>Visually acceptable Queuing</td>
<td>To analyst’s satisfaction</td>
</tr>
</tbody>
</table>

These validation standards are based on hourly measurements similar to static models rather than time-dependent measurements that are available in microsimulation models. As discussed in previous section, unlike static demand modeling, autocorrelation and non-stationary issues in microsimulation outputs make it difficult to apply classical statistical methods. Accordingly, the validation criteria for microsimulation have to be different from those for static demand models and need to be further studied for the time-dependent systems.

The validation standard can be described by summarizing aforementioned MOEs and statistical analysis methods. In a validation process, comparison is made mostly based on macroscopic data are macroscopic data due to easiness of data acquisition. MOEs to be examined for model validation are:

- Speed and flow relationship
- Distribution of traffic volume by lane
- Speed distribution
- Average speed
- Travel time
- Density
- Throughput
Validation criteria for those MOEs could be defined via statistical analysis methods as discussed in the previous section. Suggested statistics are:

- Root mean square error
- Theil’s U-Statistic
- Two-sample t-test
- Two-dimensional two-sample K-S test
- Error analysis using ARIMA

Basically, establishing a validation guideline is to set up the acceptable range of those statistics. However, there are no absolute measures or thresholds that can be achieved to declare a model or its components validated. The level of accuracy expected of a model depends on the size of network and the intended application of the model. Especially in microsimulation, the data aggregation interval is an important factor affecting overall statistic measures. Therefore, it is necessary to investigate many factors affecting simulation statistics and to carry out sensitivity analyses in developing the validation guideline for microsimulation.
6. Conclusions

Calibration/validation is difficult, time consuming, and a challenging problem in traffic simulation modeling. This report has discussed an overall approach and general procedures for microsimulation calibration/validation. Due to the general nature of this memorandum, it does not provide detailed methods and variables to be adjusted, but does provide representative model characteristics often used in microsimulation models. A four-step approach has been suggested, including basic data input, driving model calibration/validation, route choice model calibration, and stand-alone model validation.

In model validation, a major concern is error inherent in basic input data and data used for model validation. Problems in input data or validation data can lead to erroneous corrections to models that will damage model performance, credibility and results. Possible source of error resulting from development and calibration of travel models (Barton-Aschman and Cambridge Systematics, 1997) includes:

- **Measurement error**: inherent in the process of measuring data resulting from poor data quality control
- **Sampling error**: bias introduced in the process of selecting the set of observations from the population
- **Computational errors**
- **Specification error**: improper structure of the model, such as omission of relevant variables
- **Transfer error**: when a model or parameter developed for one context or region is applied in a different one
- **Aggregation error**: arising from the need to forecast for groups of individuals while modeling needs to be done at the level of the individual

It is important to recognize the source of error in model calibration and validation processes. Erroneous corrections to models can be avoided by identifying uncertainty and such potential error in basic input or validation data.

Microsimulation models mimic real-world traffic condition and produce numerous outputs as in the real world. The fundamental objective of model validation is to narrow the gap between real-world measures and predicted values. A major difficulty in establishing validation standard involves statistical issues in microsimulation outputs that are auto-correlated and non-stationary. Such properties make it difficult to apply conventional statistical analysis. This memorandum has provided major MOEs and statistical analysis methods to help establish a guideline for microsimulation model calibration/validation. However, it is necessary to carry out extensive investigation on statistically acceptable range of error and further studies need to be performed on MOEs in order to establish a reasonable guideline for a microsimulation model.
References


Brackstone, M., M. McDonald, and B. Sultan Dynamic Behavioral Data Collection Using an Instrumented Vehicle, Transportation Research Record 1689, 1999, pp 9-17


Wisconsin Department Of Transportation, District 2. Freeway System Operational Assessment, Technical Report I-33, Paramics Calibration & Validation Guidelines, DRAFT, June 2002