Date : 14 August 2005
Paper Version : 1.2

Paper Title :

Systematic Verification, Validation and Calibration of Traffic Simulation Models

Authors :

H. Rakha, B. Hellinga, M. Van Aerde, and W. Perez

Destination of Paper :

Transportation Research Board
75th Annual Meeting
January 1996
Washington, D.C.

Systematic Verification, Validation and Calibration of Traffic Simulation Models

H. Rakha[1], B. Hellinga[1], M. Van Aerde[1], and W. Perez[2]

[1] Department of Civil Engineering
Queen's University
Kingston, Ontario, Canada K7L 3N6
Fax: (613) 545-2128

[2] SAIC
1710 Goodridge Drive
McLean, Virginia 22102
Fax: (703) 734-5934

ABSTRACT

Traffic simulation models have the potential to provide a cost-effective, objective, and flexible approach to assessing design and management alternatives, particularly when these alternatives involve the emerging technologies of Intelligent Transportation Systems (ITS). However, in order for this potential to be realized, models must be valid for the application and must provide results that are credible and reliable. The process of ensuring validity, credibility, and reliability typically consists of three elements - verification, validation and calibration.

In current traffic engineering practice, there appears to be little uniformity in the definition of each of these three process elements. There also appears to be a lack of consensus among both model developers and model users regarding the actions required to carry out each process element and the division of responsibilities between the two groups.

This paper attempts to provide traffic model developers and users with a framework for the verification, validation and calibration of traffic models. Examples are provided to illustrate the model verification and validation processes. Furthermore, each process element is clearly defined as is the role of model developers and model users.

INTRODUCTION

Background

With the rapid advancement of computer technology, numerical modeling has become a valuable tool in the design and evaluation of scenarios across different engineering disciplines. In some instances, numerical modeling is the only means available for evaluating and testing designs. For example, field experimentation with hazardous immiscible liquids is prohibited by law in many regulatory jurisdictions and thus numerical modeling is the only available means by which remedial actions on field-scale processes can be evaluated.

In the field of traffic engineering, financial and legal constraints frequently prevent the use of field experiments to evaluate alternative traffic design and management schemes. A significant body of empirical data have been accumulated during the past five decades from the field tests that have been conducted. These data have been critical for the development of fundamental theory and general relationships, but are often less useful for evaluating specific alternatives, particularly when these alternatives involve the emerging technologies of Intelligent Transportation Systems (ITS). Under these conditions, traffic simulation models have the potential to provide a cost-effective, objective, and flexible approach to assessing design and management alternatives.

Traffic simulation models may be used to evaluate the impacts of changes to both network infrastructure (e.g. adding a lane to an existing roadway or adding a road to a network), traffic control devices (e.g. re-timing of traffic signal settings or installing a ramp metering scheme), and advanced forms of ITS (e.g. evaluating the impact of equipping a large fraction of the driver population with a dynamic route guidance system). Traffic simulation models are able to provide traffic, environmental and safety Measures of Performance (MOP) including average travel time, average travel distance, average number of vehicle stops, total fuel consumption, vehicle emissions of HC, CO, and NO_x, and vehicle accident risk.

Typically, developers create models and perform initial model evaluations to provide model users with a level of assurance that the model is reliable and realistic. Model users typically need to select values for input parameters that reflect the specific local conditions to be modeled. In addition, model users often desire, or are required, to demonstrate that the model results are realistic and credible.

In view of the increasingly important role that traffic simulation models play in the evaluation of advanced forms of ITS, and in view of the significant levels of effort required to conduct these evaluations, it is important that a systematic procedure for the verification, validation and calibration of traffic models be clearly defined and understood by both model developers and model users.

Objectives of Paper

Different engineering disciplines have defined the terminology in developing, verifying, validating and calibrating their respective numerical models (ASCE, 1982). However, there appears to be a lack of consensus among traffic model developers as to the terminology to be used in the traffic model development and testing exercise. While Benekohal (1991) has provided a first step in creating a framework for the verification and validation procedure, the objective of this paper is to provide traffic model developers and users with a more general standard for traffic modeling.

This paper has five specific objectives:

1. Define a standard nomenclature in traffic model verification, validation and calibration.
2. Establish a framework for the systematic verification, validation and calibration of traffic models.
3. Distinguish between the role of the traffic model developer and the traffic model user in the verification, validation and calibration process.
4. Demonstrate the need to develop a standard testbed for traffic model validation.
5. Demonstrate the need to develop a guide on traffic model calibration techniques.

Paper Layout

The next section of this paper defines the proposed standard verification, validation and calibration nomenclature in the traffic modeling context. This section also demonstrates how the verification, validation and calibration processes interact. The objective of this section is to provide the reader with a background prior to discussing the individual verification, validation and calibration elements in detail.

The third section examines the process of traffic model verification in addition to proposing a traffic model verification procedure. Traffic model validation is examined in the fourth section. This section proposes a two-step validation procedure, in which the traffic model is first compared to validated analytical theory, and then to field data. Example illustrations are provided to further describe the analytical and field validation processes. While these validation examples are specific to the INTEGRATION model, the examples are equally valid for other traffic models. The fifth section examines the process of traffic model calibration. Finally, in the sixth section, the conclusions and recommendations of the paper are presented.

OVERVIEW OF THE VERIFICATION, VALIDATION & CALIBRATION PROCEDURE

Terminology

The first hurdles that model developers and model users must overcome is the acceptance of a standard set of terminology to describe verification, validation, and calibration. Currently, the terms *verification*, *validation*, and *calibration* are often poorly understood and misused or used out of context by traffic modelers. The objective of this section is to propose a standard clear definition to these processes in order to address the misuse of this terminology.

Model verification is defined to be the process of determining if the logic that describes the underlying mechanics of the model, as specified by the model designer, is faithfully captured by the computer code. Model verification therefore determines if, independent of the validity of the logic or the theory from which the logic is derived, the corresponding computer program produces the desired outputs (in terms of accuracy, magnitude, and direction). For example, if the model designer specifies that $A = B + C$, then model verification determines if the computer code computes A as the sum of B and C. Model verification does not attempt to determine whether this relationship adequately captures reality or if A should be equal to something other than the sum of B and C.

Model validation is considered to be the process of determining to what extent the model's underlying fundamental rules and relationships are able to adequately capture the targeted emergent behavior, as specified within the relevant theory and as demonstrated by field data. In other words, can the car-following, lane-changing and gap acceptance rules utilized by the model produce the corresponding capacities, queue sizes, speed distributions and weaving effects.

Model calibration is considered to be the process of determining to what extent the model user is able to, or is required to, modify the default input parameter values, that describe the underlying mechanics, in order to reflect the observed local traffic conditions being modeled.

Procedure Framework

The individual tasks of model verification, validation, and calibration interact with each other, requiring that they be carried out in a particular sequence, as illustrated in Figure 1. Model verification, which is the first element of the process, is the responsibility of the model developer. Model verification entails compiling the program successfully, running the model error-free and without excessive mathematical approximation. Model validation, which is also the responsibility of the model developer, is initiated once model verification that been successfully completed. Model validation entails comparing the model output to generated analytical solutions and to collected field data. A failure in validation requires that some modifications be made to the model, and thus the verification process must be repeated. The iterative cycle is conducted by the model developer until the validation process is successful. At this point the model developer has fulfilled his/her obligations in providing a model that has been successfully verified and validated. It must be emphasized that it is impossible to conclusively demonstrate that a model is valid, since it is impossible to execute the model for every possible combination of input data. Instead the model developer demonstrates that the model is not invalid for the scenarios studied. Care must be taken to ensure that the scenarios that are studied are representative of typical scenarios for which model users are expected to encounter.

The model developer should provide model users with a document that describes how to run the model, together with a description of the different input parameters, their impact, acceptable ranges, and default values. In addition the model developer should provide documentation describing validation tests that have been conducted.

Model users, prior to applying the model to their local study, must engage in a calibration exercise. Traffic model calibration consists of selected input parameter values that reflect the local study area's network, climatic, and driver characteristics. Three causes can be identified that can lead to an unsuccessful model calibration. First, input parameter values may not have been computed correctly from the existing field data. Second, insufficient field data, or data of insufficient quality may have been used to estimate the parameter values. Third, the underlying model logic may be inadequate to capture some traffic behavior phenomena. Unfortunately, it is not generally clear which one, or more, of these factors causes this discrepancy with the local traffic conditions.

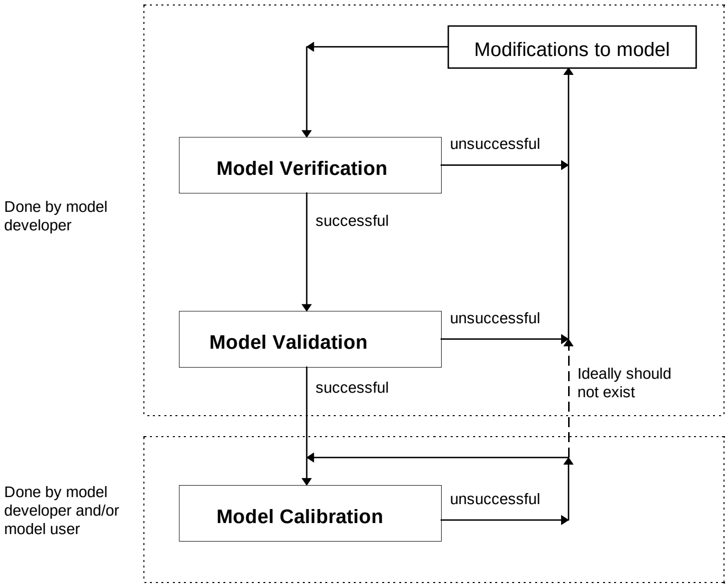

Figure 1: General verification, validation and calibration framework

TRAFFIC MODEL VERIFICATION

Verification is the process of determining if the computer code, that implements the modeling logic, produces the desired output for a given set of input data. A model is considered to be successfully verified if the computer model results are consistent, in terms of both magnitude and direction, with results from the direct application of the logic on which the computer code is based. The level of accuracy is defined by the model developer and is generally a function of the model component that is being verified.

Verification does not require that the actual logic on which the model is based, be accurate in capturing the relevant theory, nor does it require that actual field data be used as input to the model. Consequently, verification can be performed independent of field data and without a comprehensive understanding of traffic engineering theory and practice. However, field data should be considered in order to ensure that the verification is performed for a range of input parameter values, and input parameter combinations, that are consistent with typical field conditions.

Traffic model verification has two objectives:

1. Ensure that, for a given input, the program code provides output that are consistent with the logic on which the program code is based.
2. Conduct limited sensitivity testing to verify that outputs are consistent over the range of typical input values.

Figure 2 illustrates the traffic model verification process in detail. The verification process can be divided into four sequential steps.

The first step involves the selection of model input parameter values. These values should be selected so that they encompass the expected domain of application of the model. In this step, each input parameter value is checked independently for consistency with typical field data. For example, consider the verification of a freeway model component which requires two input parameters: the free-speed and speed-at-capacity. Initial parameter values of 100 km/h and 50 km/h are chosen for the free-speed and speed-at-capacity. Field data indicate that the free-speed ranges from 80 km/h to 120 km/h, while the speed-at-capacity ranges from 40 km/h to 90 km/h. The initial independent check determines if the values chosen for the free-speed and speed-at-capacity are each within the typical range exhibited by field data.

Following the initial independent check for consistency, an additional check is made in step 2 to test if the combination of the selected model input parameter values is consistent with field data. For example, a free-speed of 80 km/h and a speed-at-capacity of 90 km/h, which are both separately consistent with the field measurements, together would not be consistent with field data because it is impossible to have a speed-at-capacity that exceeds the free-speed. Thus, the joint assessment ensures that the combination of selected model input parameter values is consistent with typical field data values.

In step 3, results are generated by the simulation model for the selected input parameters. Results are also generated through the direct application of the model logic, independent of the computer code.

The forth step consists of comparing the computer model output to the output generated by direct application of the model logic. If the results are within the level of accuracy specified by the developer, the verification process is considered successful. Otherwise the model requires modifications and the verification process must be repeated with the modified model.

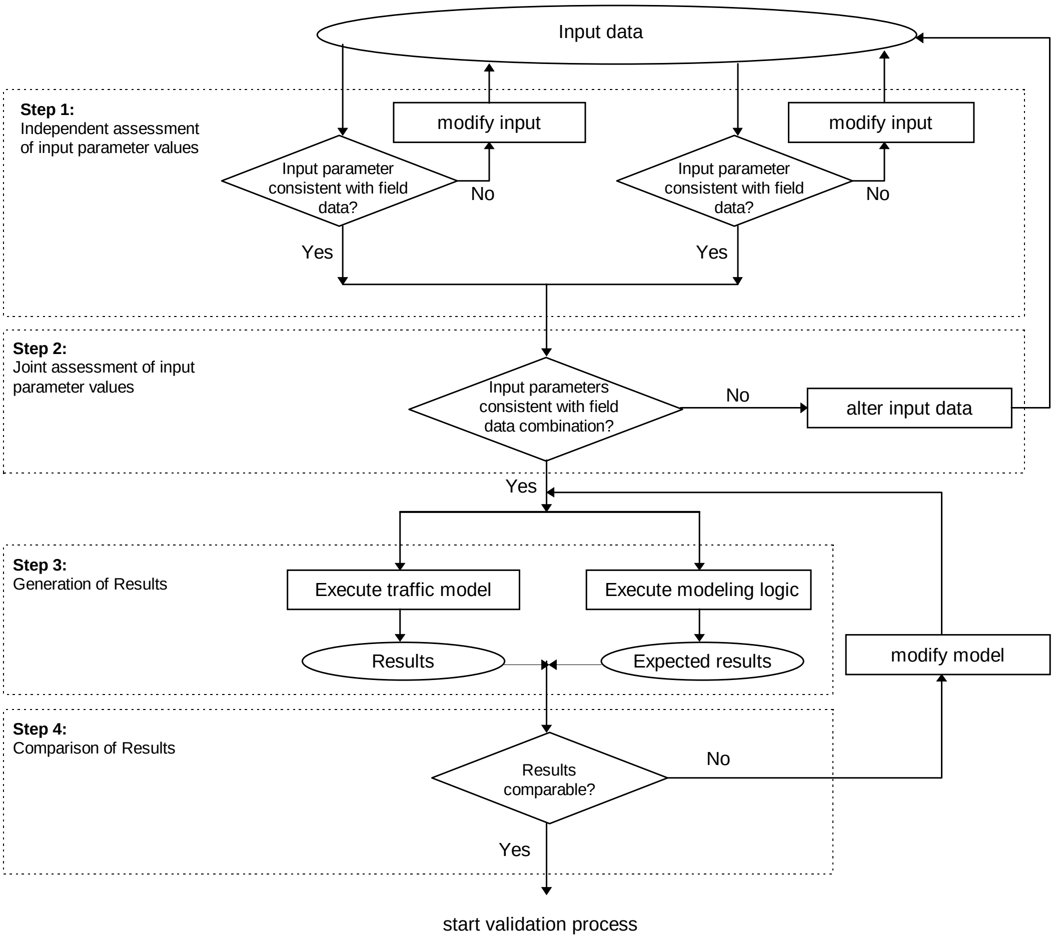

Figure 2: Schematic of the verification process

TRAFFIC MODEL VALIDATION

Validation is the second element in the verification, validation and calibration process. Model validation attempts to determine if the hypothesized relationship between the underlying behavioral rules, that are captured by the model, and the consequent emergent behavior can be demonstrated to be consistent with the prevailing theory and field data. Model validation should be done by the model developer and should ideally not be repeated by the model user. Model users should only need to calibrate the model to their particular set of local conditions and need not repeat the validation exercise in order to estimate the expected level of modeling error.

There are three objectives to the validation process:

1. Provide measures that reflect the model's ability to match the selected benchmark (analytical solution or field data) for a particular application domain.
2. Provide a sample of default parameters, together with the range of inputs, for which the validation is applicable.

3. Provide the results of a sensitivity analysis of the model about the default parameters in order to indicate the potential rate at which the error increases for a given calibration error level.

As indicated by Figure 3, the six steps of the validation process are divided into two groups, defined by analytical validation and field validation. Analytical validation, which is conducted first, examines simple network problems in which interaction effects can be limited and for which analytical solutions can be generated. Subsequently, field validation is conducted, in which an actual real-world network is modeled, and model results are compared to field data.

In step 1, a comprehensive suite of strategic hypothetical scenarios are defined, each of which is designed to test a specific model feature. The suite of scenarios must be sequenced so that the most fundamental elements of the model are tested first. The design of this suite of scenarios is critical to the success of the analytical validation process, and as such, a comprehensive examination of this testbed suite of scenarios will appear in a future paper.

Step 2 consists of the generation of model results and analytical solutions for each of the testbed scenarios in turn. Care must be taken to ensure that assumptions made by the analytical solution are identified, and if possible, captured in the model by the selection of appropriate input parameter values. In the event that several analytical techniques are available, each should be used to generate solutions for comparison to the model results. Furthermore, the results of the model can be compared to the results of other validated models.

In step 3, suitable MOPs are identified and the analytical solutions and simulation model results are compared. The degree of similarity is quantified and the cause of any discrepancies identified. This process is critical, since it clearly identifies the capabilities and limitations of both the simulation model and the analytical techniques. As such, this process requires a comprehensive understanding of the traffic simulation model, the analytical techniques employed, and traffic engineering theory and practice. In the event that the comparison is unsuccessful the developer has to make modifications to the model and return to the verification process.

Steps 4, 5, and 6 comprise field validation, which parallels analytical validations with four significant differences. First, the scenarios examined are actual real-world networks, with the result that model components cannot be examined in isolations and higher order iteration effects cannot be controlled for. Second, the state of the network is not determined from analytical solutions, but from field data, which must be collected. Third, the generation of model results first requires that input parameter values be calibrated using the collected field data. Fourth, the comparison of model results and field data often provide little insight into the accuracy and applicability of the model. Discrepancies can result from incorrect input parameter value selection (i.e. calibration), or from a fault with the underlying model logic, or both. The developer must discern the cause of the discrepancy and then make the relevant modifications to the model and/or to the calibration of the input parameters. Even if the discrepancies between the model results and the field data are very small, it is possible that compensating errors in the selection of the input parameter values, may mask fundamental flaws in the model logic.

Given the complexity of traffic models and the stochastic variability that is evident in real world traffic data, the level of accuracy required in the analysis is typically less than in the case of analytical validation.

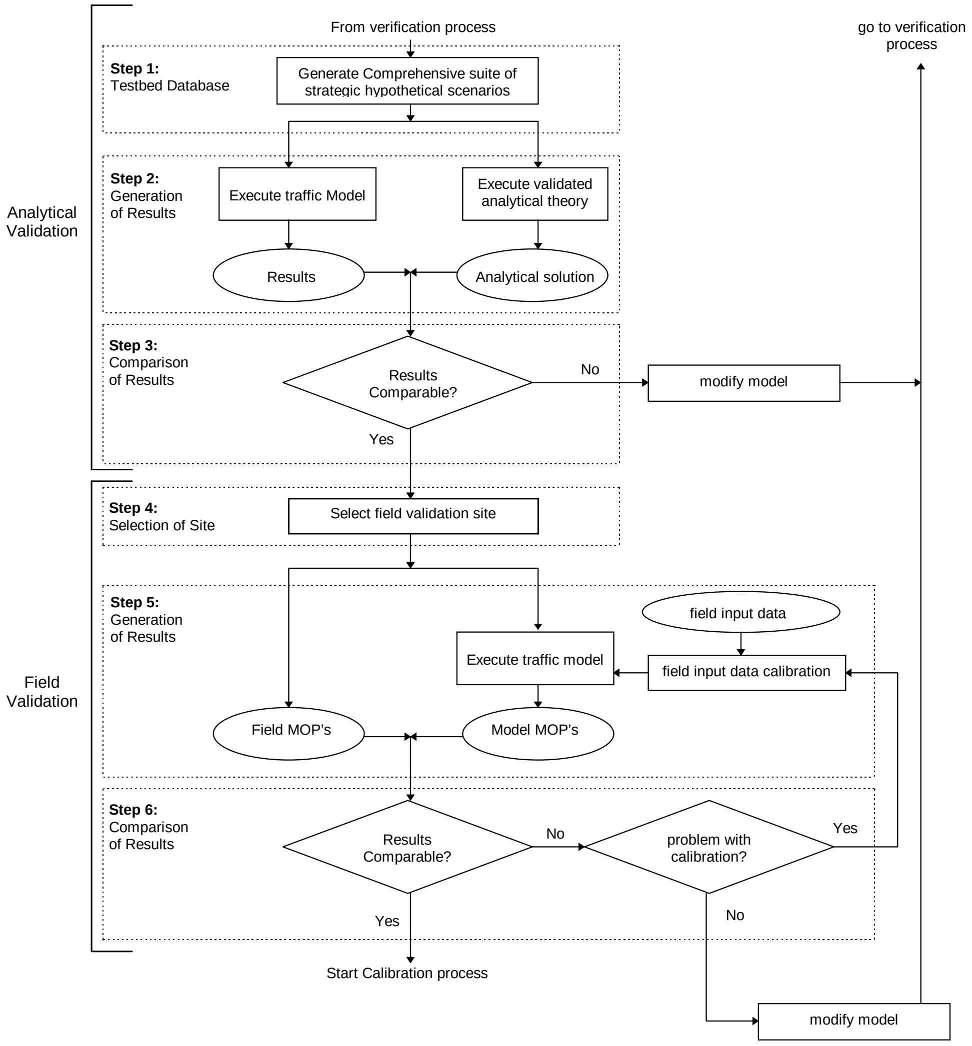

Figure 3: Schematic of the validation process

Example Illustration of Analytical Validation

This section describes an analytical validation process that was conducted to validate the INTEGRATION traffic simulation model (Van Aerde, 1985; Van Aerde and Yagar, 1990) for a simple hypothetical signalized network. Results from the INTEGRATION simulation model were compared to a validated and accepted signalized model, TRANSYT (Robertson and Bretherton, 1991), and to two validated analytical solutions (Rakha and Van Aerde, 1995).

The hypothetical network utilized for the analytical validation consists of a single arterial connecting two zones, as illustrated in Figure 4. Traffic signal 1, which was located at node 10, was considered to be the master signal to which the offset of traffic signal 2, at node 11, was referenced.

All links were assigned a constant speed of 60 km/h in order to replicate TRANSYT's implicit assumption that the link travel time is independent of the link flow. All the links were composed of two lanes in each direction, each with a 1500 vph/lane saturation flow rate. The Origin-Destination (O-D) demands used in the analysis were as follows: 1200 vph from 1 to 2, 1200 vph from 2 to 1, 1200 vph from 3 to 4, and 800 vph from 6 to 5.

The macroscopic analytical solution represented the flow of vehicles as a continuous flow of fluid as opposed to modeling the flow as discrete entities. This macroscopic solution corresponds most closely to the macroscopic approach of TRANSYT. The macroscopic analytical solution was derived using the optimum cycle length and phase split signal settings that were generated by the TRANSYT model, but for offsets of signal 2 relative to signal 1 ranging from 0 to 50 seconds at 10 second increments. The total delay for each approach was calculated by integrating the area under the queue/time relationship.

The microscopic analytical solution, which was developed on a spreadsheet, tracked each vehicle from its point of entrance to its final exit from the network. The intent was to execute an analytical procedure that was most similar to the microscopic representation of the INTEGRATION model.

It is evident from Figure 5 that the total travel time estimates were consistent in terms of magnitude and trends for the two models and the two analytical solutions. However, it must be emphasized that this validation exercise does not conclusively demonstrate that the INTEGRATION simulation logic is valid for all signalized network scenarios, instead it demonstrates that the simulation logic is not invalid for the conditions that were analyzed.

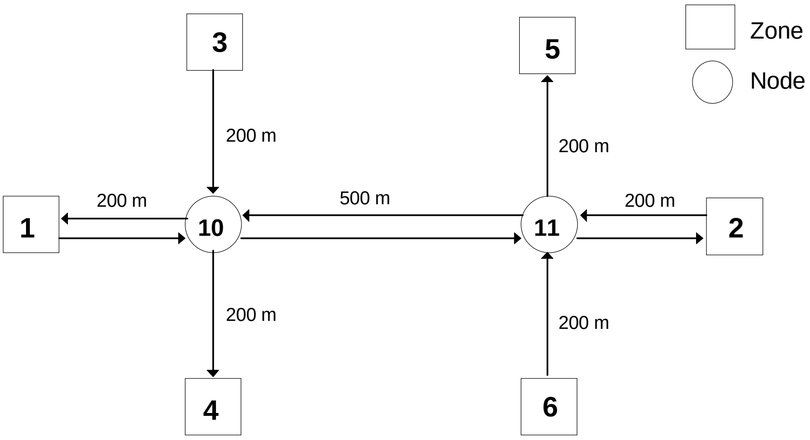

Figure 4: Network layout for analytical validation

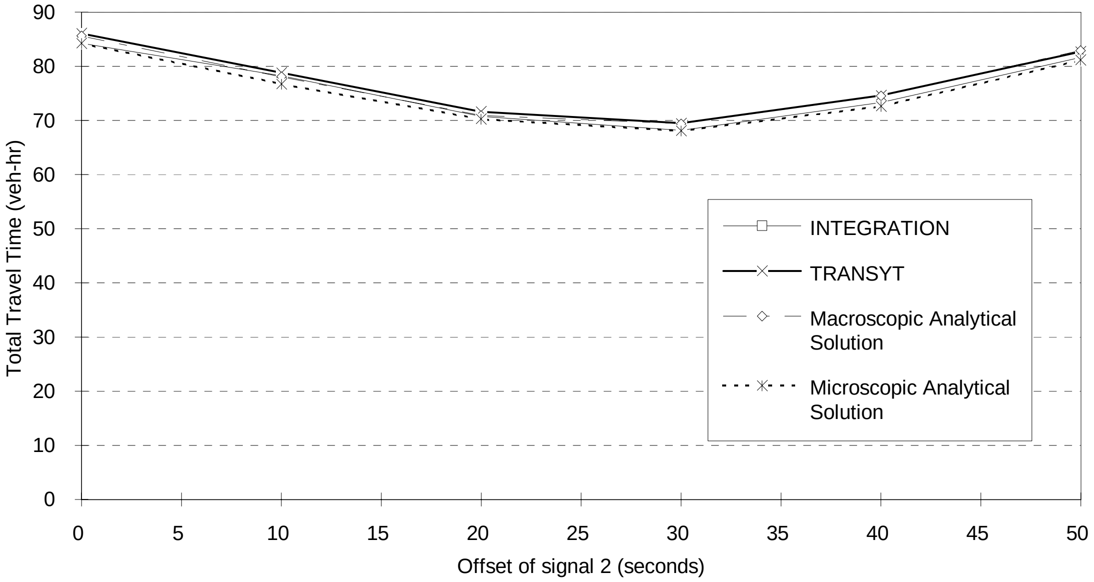

Figure 5: Total travel time estimated by analytical solutions, INTEGRATION and TRANSYT

Example Illustration of Field Validation

This section describes a field validation exercise that was conducted using the INTEGRATION simulation model. The validation exercise used actual real-time information from loop detectors along the I-4 freeway in Orlando, Florida. Freeway Management Center (FMC) data were available for 22 non-incident weekdays. The FMC recorded 30-second flow, speed and occupancy measurements for 24 stations in both directions along the I-4 (Van Aerde and Rakha, 1995).

This validation exercise was conducted by generating 15-minute flow and speed estimates from the INTEGRATION model that were compared to the 15-minute flow and speed measurements obtained from the loop detectors.

Figure 6 demonstrates the spatial variation in flow along the detectorized section of I-4 at the conclusion of a 2-hour simulation of the PM peak (at 5:00 PM). The dashed line represents the average typical flow variation for 22 non-incident weekdays. Based on a Chi squared type goodness of fit test, it was found that the measured 15-minute flows were not statistically different from the expected outcome of a normal distribution at the 95 percent confidence level. Based on the assumption of normality, the 95 percent confidence limits for the average typical measured flows were estimated and plotted in Figure 6. It appears from this figure that most of the simulated results were within the 95 percent confidence limits, and that simulated flows that were outside the bounds were only marginally so.

In addition, Figure 7 demonstrates the spatial variation in simulated and measured speed estimates for the I-4 freeway. It appears that the simulated speed estimates were within the estimated confidence limits at stations 8 through 25, however, the simulated speed estimates were outside the bounds at stations 1 through 7, indicating larger queue spill back conditions from station 7 within the simulation model.

Figures 6 and 7 demonstrate two issues. First, the selection of MOPs to be utilized in the validation process is of utmost importance. In this example illustration, although the spatial representation of the flow along the I-4 freeway indicated a high degree of correlation between the simulated and field data, the spatial representation of speed indicated a much lower degree of correlation.

Secondly, the example illustration poses further questions that need to be addressed, namely: what is the cause of this discrepancy? Did the discrepancy result from an erroneous calibration of the density of the queue (jam density)? or an error in the Origin-Destination demand? or an error in the routing trees? or an error in the roadway capacities? or some combination of these errors? Furthermore, what should be done to correct this unsuccessful field validation. At present, there does not exist a systematic diagnostic technique that can be applied to answer these questions, implying that further research remains to be done in the area.

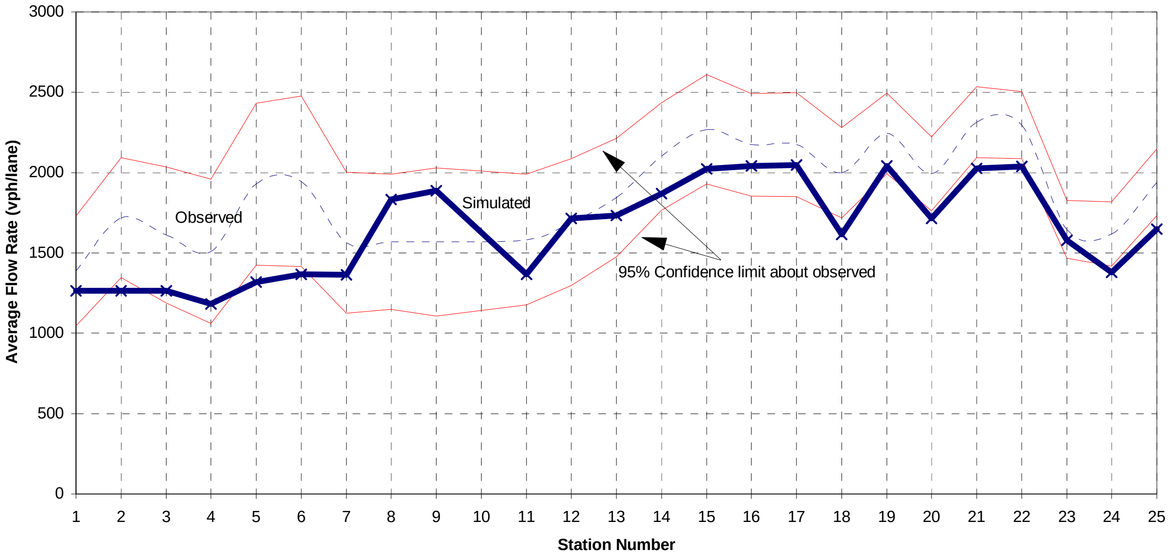

Figure 6: Spatial variation in simulated and average flow rate estimates at 5:00 PM along the eastbound direction of I-4

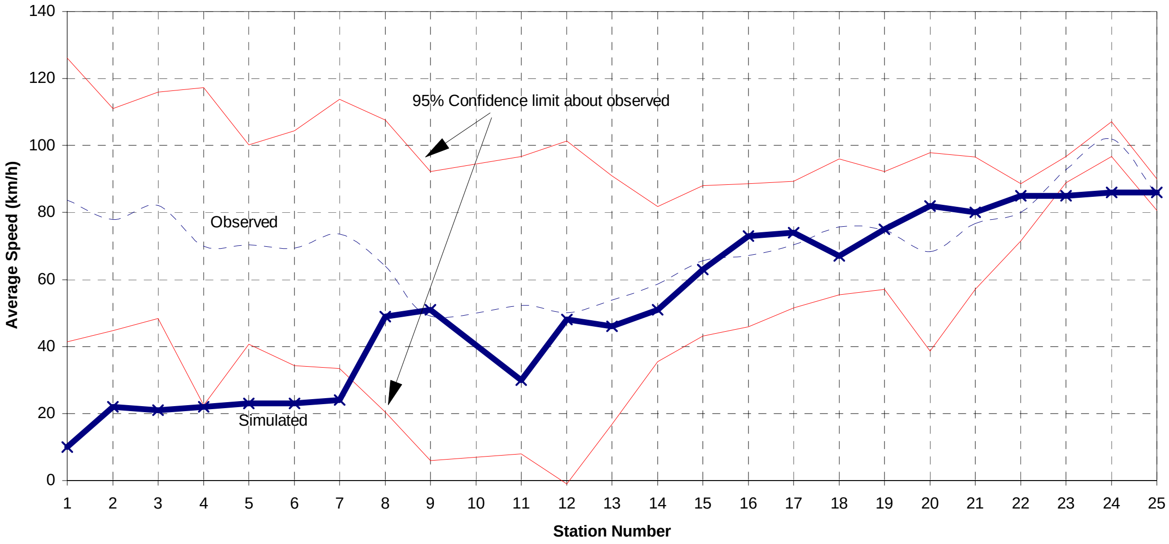

Figure 7: Spatial variation in average simulated and average observed speed at 5:00 PM along the eastbound direction of I-4

TRAFFIC MODEL CALIBRATION

In this paper, calibration is defined as the process of selecting the best set of model input parameters to address the most important differences between the model's default assumptions/conditions, and those actually observed locally.

Ideally, calibration consists solely of determining values of input parameters on the basis of available field data. Unfortunately, it is difficult to determine the success or failure of the calibration process by strictly examining the field data and the selected input parameter values. The accuracy of the input parameter value often can be quantified, but the impact of this degree of accuracy on the model's MOPs is much more difficult to estimate. For example, the mean speed of vehicles traveling in uncongested, low volume conditions, may be considered to be the best estimate of the link free speed. The accuracy of this estimate might be described by the standard deviation of individual vehicle speeds about this mean speed. However, of more interest, what is the expected impact of having a standard deviation of 10% of the mean versus 25% of the mean? More importantly, does a standard deviation of 10% of the mean represent an acceptable level of error?

Currently, these questions can rarely be answered *a priori*, with the result that model users typically estimate input parameter values and then utilize these to provide model MOPs which are subsequently compared to field conditions. This process now represents field validation, and encounters the same difficulties described previously.

In general, model users are faced with two dilemmas when calibrating traffic simulation models. First, the data available are frequently limited in quality and quantity. Detector data, providing volume, occupancy, and speed, are the most common data available, but even they are usually available for only limited portions of the network. Furthermore, the detector data do not provide direct knowledge of important input parameters such as driver routing behavior and origin - destination demands.

Second, model users are often not able to determine the impact that input parameter values have on the selected model MOPs. This may arise from several causes, including a lack of understanding of traffic engineering principles, a lack of understanding of the model, poor model documentation, or a combination of these.

These dilemmas often result in a calibration exercise that evolves into a form of the field validation process and thus involves a large amount of comparison of model outputs to actual field data.

Using the limited data and quality of data available, typically, traffic model users select the default input parameters to run the model and compare the output of the model to observed field data. Prior to comparing the model output to field data, a decision must be made on a number of issues (as was the case in the field validation process). First, what measures of performance (MOPs) should be used in the comparison? Second, what is the level of accuracy required in the calibration process? Third, what should be done if the outputs differ? There appear to be no satisfactory answers, in the literature, to these questions. Thus, further research is required to develop a standardized procedure by which a modeler can calibrate traffic simulation models to local conditions.

CONCLUSIONS AND RECOMMENDATIONS

This paper has attempted to define a standard nomenclature for the verification, validation and calibration of traffic simulation models. This proposed standard nomenclature could serve to provide some consensus among traffic model developers and users as to the terminology to be used in traffic model development, testing, calibration and application.

In addition, the paper has established a systematic framework for the verification, validation and calibration procedure of traffic models. Furthermore, the paper has also defined the role of the model developer and user. The model developer should be responsible for conducting the model verification and validation exercises and providing the model user with the following documentation:

- A manual describing how to use the model, together with a description of the different input parameter fields and default values.
- Documentation of the model validation exercises that were conducted. This document should provide three specific elements:
    1. A measure of the error between the model results and the selected benchmark (analytical or field data).
    2. Sample default input data values.
    3. Results of a sensitivity analysis of the model about the default parameter values.

The model user should be responsible for calibrating the model input parameters to the field data.

It is recommended that a standard testbed of traffic networks and traffic scenarios (O-D demands, incidents, etc.) be established in order to be utilized in validating different model components. This testbed would serve as a standard for all models, permitting users to interpret the validation results and compare these results for different models.

It is further recommended that a standard calibration framework be developed to assist users in the calibration of traffic models. This framework should provide the user with some strategies to be utilized in order to address different calibration issues. Furthermore, this framework must be general such that it can be used to calibrate any traffic model.

REFERENCES

ASCE (1982), *Modeling Hydraulic Phenomena: A Glossary of Terms*, American Society of Civil Engineers (ASCE), Vol. 108, No. HY7, July, pp. 845-852.

Benekohal R.F. (1991), *Procedure for Validation of Microscopic Traffic Flow Simulation Models*, Transportation Research Record 1320, pp. 190-202.

Rakha H.A. and M.W. Van Aerde (1995), “A Comparison of the Simulation Modules of the TRANSYT and INTEGRATION Models,” Submitted to 1996 Transportation Research Record (TRB) 75th Annual Meeting, January.

Robertson D.I. and Bretherton R.D. (1991), *Optimizing Networks of Traffic Signals in Real-Time -- The SCOOT Method*, IEEE Transactions on Vehicular Technology, Vol. 40, No. 1, February, pp. 11-15.

Van Aerde M. (1985), *Modelling of Traffic Flows, Assignment and Queueing in Integrated Freeway/Traffic Signal Networks*, Ph.D. thesis, Department of Civil Engineering, University of Waterloo, Waterloo, Canada.

Van Aerde M. and Yagar S. (1990), “Combining Traffic Management and Driver Information in Traffic Integrated Networks,” IEE Third International Conference on 'Road and Traffic Control,' London, pp. 11-16.

Van Aerde M. and H. Rakha (1995), *TravTek Evaluation Modeling Study*, FHWA-RD-95-090.