REQUIREMENTS FOR THE CALIBRATION OF TRAFFIC SIMULATION MODELS

Bruce R. Hellinga
Department of Civil Engineering
University of Waterloo
Waterloo, ON N2L 3G1

ABSTRACT

The recent emphasis on utilising advanced technologies to make more efficient use of existing transportation infrastructure, coupled with the continuing advances in desktop computing technologies, has created an environment in which traffic simulation models have the potential to provide a cost-effective, objective, and flexible approach for assessing design and management alternatives. However, the models must be demonstrated to be valid, and they must be adequately calibrated for local conditions. While the processes of verification, validation and calibration are certainly not new, based on the lack of literature on this topic, it appears that the application of these processes to traffic simulation models is not well defined.

This paper examines the issues related to primarily the calibration of traffic simulation models. It attempts to answer, or at least clarify, several key questions, including the following: What is model calibration? What measures of performance should be used in calibration? and When is a model adequately calibrated? The discussion of these questions stems from the goal of obtaining model results that are seen as credible, reliable, and useful, particularly by people who are not modellers. The distinct roles and responsibilities of model developers and model users are discussed. Examples are provided to illustrate the model calibration process.

KEYWORDS: traffic, simulation, modelling, methodology, confidence limits

INTRODUCTION

Increasing complexities of roadway design and traffic management requirements have created the need for analysis method that are more flexible and robust than traditional approaches. This need, combined with continuing improvements desktop computer capabilities, has created an environment in which traffic simulation models have the potential to provide a cost-effective, objective, and flexible approach to assessing design and management alternatives. The use of simulation models is particularly appealing for the estimation of quantities not easily estimated using more traditional approaches, such as air quality impacts, fuel consumption rates, accident risk factors, and toll revenues, and for the evaluation of emerging technologies, such as intelligent transportation systems.

However, traffic simulation modelling, and transportation modelling in general, is often viewed by non-modellers as an inexact science at best, and as unreliable "black-box" technology, at worst. In many cases, this scepticism of the value of simulation modelling results from (a) unrealistic expectations of the capabilities of the simulation model, (b) use of a poorly validated or verified model, and/or (c) use of a poorly calibrated model. Unfortunately, there are few standards by which the level of calibration, validation, or verification of a model can be measured.

The processes of model verification, validation and calibration are critical to the credibility and reliability of the model results. However, in current traffic engineering practice, there appears to be little uniformity in the definition and conduct of these process elements. Unlike many other engineering tasks, such as materials testing and structural design, no generally accepted standards exist to determine when a model can be considered to be suitably validated or calibrated.

This paper examines the issues related primarily to the calibration of traffic simulation models. Discussions pertaining to aspects of model verification and validation are available elsewhere in the literature (Rakha *et al.*, 1996). To provide context, this paper provides definitions of model verification, validation and calibration. However, the key question of what constitutes adequate model calibration is addressed.

These issues and questions are addressed from the perspective of a model developer and model user. The goal is to provide an understanding of the calibration process, and where possible, a methodology by which model users can obtain results that are seen as credible, reliable, and useful, particularly by people who are not modellers. Examples are provided to illustrate the model validation process.

Definition of terminology

The terms validation, verification, and calibration are in common use by traffic modellers and non-modellers alike. However, the meanings associated with these terms have not always been clear and consistent. Consequently, this section provides a definition of each term.

Model validation is defined to be the process of determining if the model logic proposed by the model developer, is correctly represented by the computer code. Validation simply ascertains that the outputs from the computer code are consistent with the model logic. Model validation does not make any assessment of the validity of the proposed model logic or of the theory on which the logic is based.

Model verification is defined to be the process of determining to what extent the model's underlying theory and logic reflect reality. For example, can the aggregate link volume-delay functions utilised by the model, produce correct estimates of trip travel time, queue sizes, and route choices? The model verification process presumes that the model has already been validated, such that the computer program correctly reflects the model developer's intended modelling logic.

Model calibration is defined as the process by which the model user establishes input parameter values in order to reflect the local traffic conditions being modelled.

Structure of paper

This paper is organised as follows. The next section describes the proposed calibration process and identifies the component steps that constitute the process. Subsequent sections describe each of these components in turn, beginning with the definition of study scope, goals, and objectives, the considerations in choosing measures of performance and evaluation criteria, the calibration of input data, the evaluation of simulation output, and the re-calibration of input data on the basis of this evaluation. Finally, conclusions are drawn, and recommendations are made.

DESCRIPTION OF CALIBRATION PROCESS

The proposed calibration process, consisting of three main phases and eight component steps, is illustrated in Figure 1.

Phase 1 comprises those tasks and activities that are conducted prior to the commencement of any modelling. These tasks consist of the definition of the study goals and objectives; the identification of required field data; the identification of measures of performance that are consistent with the study goals, available field data, and simulation model capabilities; and the specification of criteria for the evaluation of the calibration process.

Phase 2 consists of the initial calibration of model parameter values on the basis of available field data. Typically, parameters requiring calibration include network coding, including the specification of the location of zones and nodes; link characteristics such as macroscopic speed-flow-density relationships; driver behaviour characteristics such as routing strategies and gap acceptance requirements; and origin-destination traffic demands.

In Phase 3, the results from the model are compared to field conditions and tested against the previously established criteria. If these criteria are met, then the model is considered to be adequately calibrated and the model can be used for evaluating non-base case scenarios. However, if the established criteria are not met, the model calibration is considered to be unacceptable, and refinements and modifications must be made to the model parameter values.

Each of these three phases is described and illustrated in more detail in the following sections.

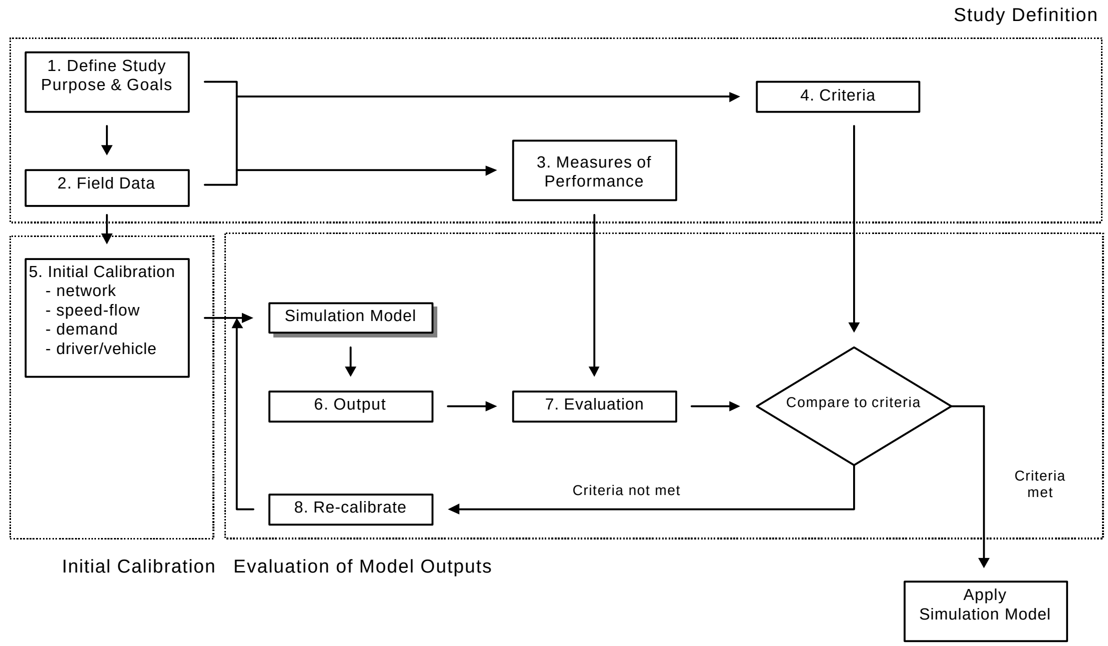

Figure - 1: Proposed calibration process

PHASE – 1: STUDY DEFINITION

Defining Study Goals and Objectives

The decision to use a traffic simulation model within a traffic operations or design study should be made after careful consideration of the relative advantages and disadvantages of simulation modelling as compared to other non-simulation techniques. Typically, the decision to use simulation modelling results from one or both of the following:

(a) The need to estimate measures of performance that can not be adequately estimated by other techniques, and/or

(b) The need to evaluate a complex network configuration, highly dynamic traffic demand characteristics, and/or a unique control strategy that cannot be adequately analysed using non-simulation techniques.

The reasons for choosing a simulation technique over a non-simulation technique should be consistent with the stated study purpose and objectives. Furthermore, these objectives should influence the choice of field data collected (Step 2), the measures of performance that are chosen (Step 3), and the criteria that are to be used to assess the sufficiency of the model calibration (Step 4).

Determining Required Field Data

In most studies in which simulation models are used, very limited resources exist for the collection of field data for calibrating the model to base conditions. And even when adequate resources exist, some data (e.g. emission rates, route choice behaviour, and origin-destination demands) are difficult to measure directly in the field. Thus, it is important that a priority ranked list of required field data be established. This list must identify the type of data (e.g. link flows), the temporal resolution (e.g. aggregated to 15-minute average flows), the spatial resolution (e.g. counts between each major intersection), the temporal coverage (e.g. 5 AM to 11 AM), and the spatial coverage (e.g. both directions of Yonge Street from Finch Ave. to Steeles Ave.). Each item in the list should be given a relative priority to indicate its relative importance in the calibration process. The development of such a list enables decisions to be made regarding the resources that should be allocated to the collection of field data. Furthermore, if decisions are made to not collect certain data, the consequences of these decisions on the calibration process can be realistically defined.

To illustrate, consider Table 1, in which an example list of data required to calibrate a model is provided. A similar list of data required for the development of evaluation criteria is provided in Table 2. These tables were created for a study of Interstate 696 in Detroit, Michigan, in which the goal of the study was to calibrate a simulation model for the purposes of evaluating the impacts of various advanced traffic management strategies and advanced traveller information systems (Hellinga, 1996).

Table 1: Data required for calibration of model (Source: Hellinga, 1996)

Category	**Parameter**	**Data Source**[1]	**Comments**	**Status**
Network Coding	Link length	plan drawings	Particularly, length of turning bays and ramps.	Required
	Number of lanes	Plan drawings/local collection	Existing coding of I696 needs to be reviewed. May require local verification if plans not available.	Required
	Lane connectivity	base map/road map	Connectivity in existing coding is considered correct.	Required
	Lane use restrictions, lane striping, and turn prohibitions	OC/local collection		Required
	Signal locations and timings	OC	If timings not available can model using signal optimisation	Locations required timings desired
Emergent Behaviour	Speed-Flow-Density Relationships	Loop detector data	Require 5 - 15 minute aggregated speed and volume data for at least 4 freeway locations (which experience congestion), covering from 6 AM to 8 PM.	Desired (if not available, will need to use typical freeway values)
O-D Demands	Link traffic flows	Loop detector data	Desire 15 minute aggregated flows for all detector locations within study area, including service roads. 1-hour manual counts will suffice if loop data not available.	Required
	Turning movements	Manual counts	Any existing turning movement data for intersections within study network	Desired
	Existing O-D data	MDOT/OC	Obtain any existing data.	Desired
	Average trip length	MDOT/OC	Obtain any existing data.	Desired
Driver Routing	Traveller information	Local knowledge	Allegorical data describing quality of traveller information available (e.g. radio traffic reports during rush hour every 5 minutes)	Desired

[1] OC = Oakland County; MDOT = Michigan Department of Transportation

Table 2: Data required to develop criteria for evaluating calibration
(Source: Hellinga, 1996)

Data	**Source**[1]	**Comments**	**Status**
Detector speed, volume and occupancy	MDOT/OC	Maximum 15 minute level of aggregation. Require data for entire PM peak period for all available stations. Should have data from several days.	required
Travel time	MDOT/local collection	Obtain any existing travel time data.	desired
Average intersection delay	OC/local collection		desired
Congestion patterns during PM peak	MDOT/OC	Data expected to be allegorical, not quantitative.	required

[1] OC = Oakland County; MDOT = Michigan Department of Transportation

Choosing Measures of Performance

Simulation models can generally provide numerous measures of performance (MOPs). Potential MOPs include: link volume, link speed, queue size and location, link travel time, trip travel time by origin and destination and by time of departure, total travel time, average trip length, average number of stops, average fuel consumption, tailpipe emissions of hydrocarbons (HC), carbon monoxide (CO), and nitrous oxide (NO_x), and average accident risk.

The choice of the appropriate set of MOPs is influenced by the study objectives, the capabilities of the model, and the available field data. Furthermore, the set of MOPs used to evaluate the model calibration process may be different from the set used to evaluate the traffic management or geometric configuration alternatives within the study.

The development of the list of required field data is also influenced by the choice of MOPs to be used within the evaluation process. For example, if the delay experienced by drivers on the approaches to intersections is to be used as one of the measures of performance in the evaluation process, then data that quantify this delay must be available.

After establishing the list of required data, the resources required to collect the listed data must be compared to the resources available. If the available resources are insufficient, then lowest priority data needs are eliminated from the list.

Establishing Evaluation Criteria

Perhaps the most difficult element of traffic modelling is the establishment of criteria by which the adequacy of model results can be determined. Unlike many other engineering tasks, such as material testing and structural design, no generally accepted standards or criteria exist to determine when a model can be considered to be suitably calibrated. Furthermore, the choice of the appropriate measure of performance to use, and the establishment of an acceptable range of values for this measure, is often made more complex by a lack of knowledge of the impact that exceeding the established criteria would have on the ultimate decision making process.

To illustrate, consider a traffic impact study of a proposed large urban shopping centre. The goal of the study is to estimate the expected traffic impacts of this development and to evaluate strategies for mitigating these impacts. The traffic generated by this development is expected to directly impact the adjacent arterial, as well as a nearby highway. The measure of performance that will be used to evaluate the impact of the traffic generated by the development on the arterial will be intersection level of service, as determined from average stopped delay. The impact on the highway will be evaluated by level of service as determined by volume/capacity ratio. Since the measures of performance of average stopped delay and highway volume are to be used to evaluate the impact of the development on the network, and assuming appropriate field data are available, it may also be prudent to use these MOPs in the evaluation of the model calibration. The difficulty in doing so, is in establishing acceptance criteria. If the pre-development field data indicate that the average stopped delay is 18 seconds/vehicle, what results must be obtained from the simulation model in order for the model to be considered adequately calibrated? If the model results indicate an average stopped delay of 26 seconds/vehicle can the model be considered to be adequately calibrated?

As illustrated in Figure 2, when sufficient field data are available, it may be possible to quantify the mean and variance of the measure of performance of interest, and from these data establish statistical confidence limits that can be used as calibration criteria. However, in practice, field data in sufficient quantity and of sufficient quality are rarely available to permit this type of rigorous statistical approach to more than a few MOPs. Furthermore, study budgetary constraints rarely permit model users to conduct rigorous statistical analysis of field data for the purposes of establishing calibration criteria.

Consequently, in practice, calibration acceptance criteria tend to be subjective, rather than purely objective. Model results may be evaluated qualitatively by members of the project team that have personal experience and knowledge of the network conditions. Conversely, quantitative MOPs may exist for both the field data and the simulation results, but the assessment of the degree to which the model results reflect the field data may be qualitative. In these situations, major discrepancies between model results and field observation are identified, and subjectively assessed as to whether the calibration is acceptable or not. This approach tends to result in the use of qualitative terms such as *reasonable*, *adequate*, and *representative*, to describe the calibration results.

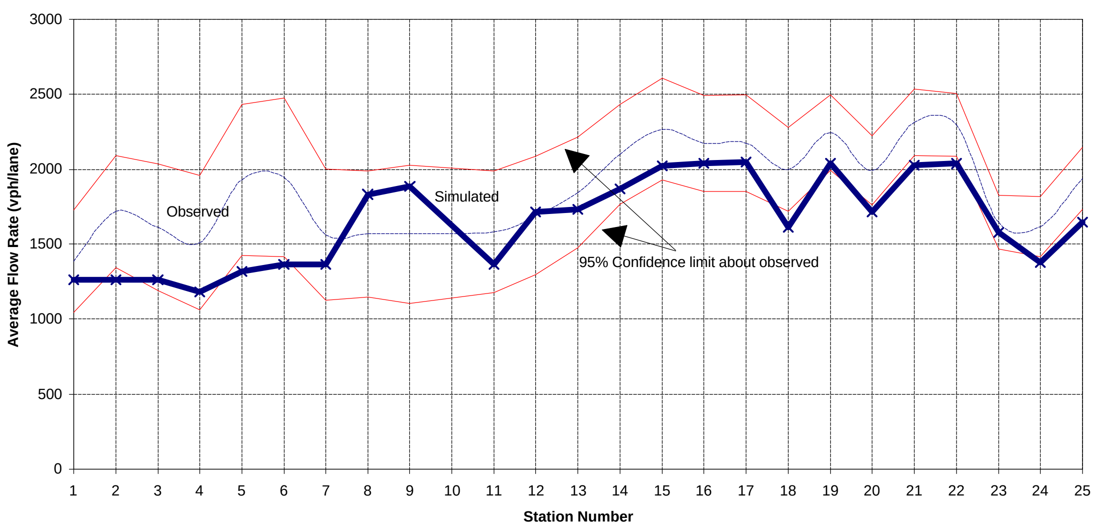

Figure 2: Development of calibration acceptance criteria based on statistical confidence limits (Source: Van Aerde and Rakha, 1995)

PHASE – 2: INITIAL CALIBRATION

Initial calibration of the model consists of four major elements – namely the representation of the network, the selection of appropriate macroscopic speed-flow-density relationships, the specification of driver routing behaviour, and the development of O-D traffic demands. Each of these elements is described in the following sections.

Network Representation

Traffic models require that the road network be abstracted into a form that can be represented within the model. Typically this representation is in the form of links, representing roadway segments; nodes, which represent the intersection of road segments; and zones which are sources and/or sinks for vehicles. Choices must be made regarding the network spatial extents and the level of detail represented within the network. In making these choices, it must be recognised that bottlenecks not part of the modelled network will not have any impact on the model results, but may be observed to influence field data if the bottleneck causes congestion to spill-back into the area of the network that is modelled. Location of zones and the specification of links connecting zones to the traffic network can also have a significant influence on the modelled routing behaviour, turning movements, and level of congestion in the area immediately surrounding the zone.

Macroscopic flow characteristics

Microscopic traffic simulation models typically require the specification of macroscopic speed-flow-density relationships. Many models permit these relationships to be specified for each link, however, in practice, links are often classified into several categories, and a unique speed-flow-density relationship is specified for each category. Modifications to these relationships might be made for specific links if conditions warrant (e.g. steep vertical grade, restrictive geometry, etc.).

Figure 3 illustrates a calibrated speed-flow relationship for a section of Interstate 4 in Orlando, Florida (Van Aerde and Rakha, 1995). The observed data were obtained from loop detectors and aggregated to a 5-minute average.

The choice of speed-flow-density relationship has implications for travel speed, but also for queue size and speed of queue growth. Thus, specifying a realistic speed-flow-density relationship is critical to obtaining an adequately calibrated model. Free speed, or the uncongested speed associated with very low flows, is likely the least critical parameter, since it only impacts travel times, and can usually be estimated from the posted speed limit with acceptable accuracy. However, the impacts of the speed at capacity flow, capacity, and the jam density, are significant, as they influence the formation, extent, and dissipation of congestion.

Driver Routing Behaviour

The issue of driver routing choice behaviour is an important consideration for studies in which the modelled network presents drivers with more than one viable route choice. In these cases, accurate representation of drivers' route choices is essential for adequate calibration. Even when modelling a linear corridor, such as Highway 401 in Toronto, routing is an issue, since this freeway facility consists of separated express and collector lanes, enabling drivers to make route choices.

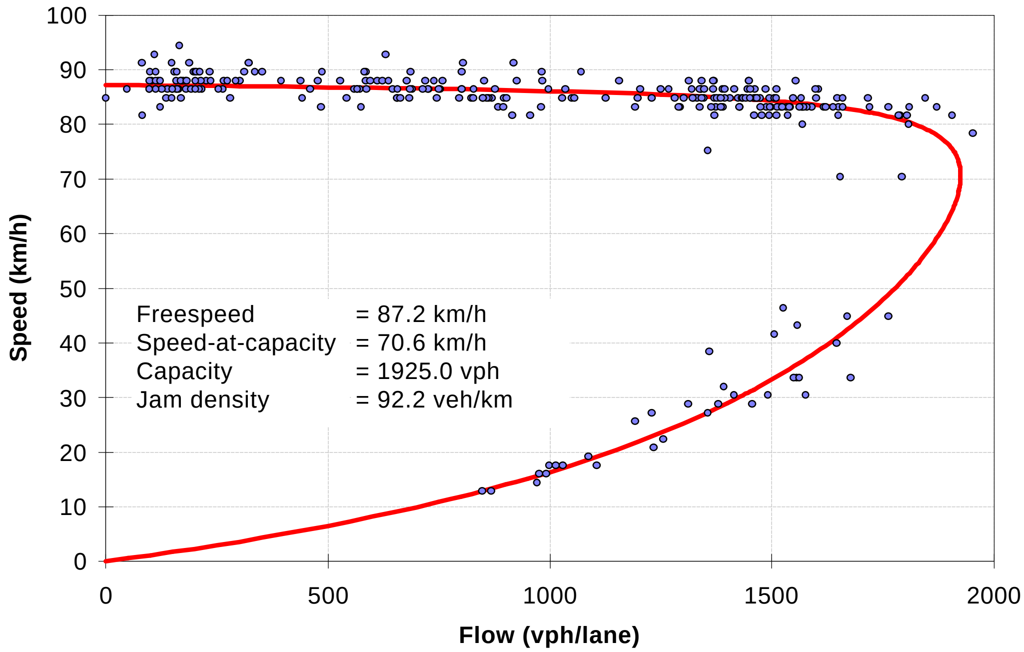

Figure 3: Calibrated speed-flow relationship for Interstate 4 in Orlando, Florida
(Source: Van Aerde and Rakha, 1995)

The calibration of route choice parameters is largely a function of the capabilities of the model being used. Some models, such as INTEGRATION, provide the user with several routing strategies, which each have two or more parameters that must be specified. An inappropriate choice of either routing strategy, or parameter value, can result in model outputs that are clearly unrealistic, even when all of the other model parameters are adequately calibrated. Unfortunately, field data can rarely be collected that directly measure driver route choosing behaviour. If link flows are collected, then the outcome of all of drivers' routing decisions may be determined (e.g. the fraction of the total flow that chose to turn left, go straight, and turn right at a given intersection) however, this does not generally provide direct evidence of the routes chosen by individual vehicles (rather it provides average turning movements), nor does it provide valuable insights into the process by which drivers make these decisions (e.g. the perceived travel time advantage of the chosen route over the rejected route; the criteria used to choose the route – travel time or something else). Therefore, the appropriate selection of routing strategy and the associated parameter values is based largely on experience rather than a direct calibration to observed data.

Demand characteristics

O-D demands define the amount of trip making demand between each origin zone and each destination zone during a particular departure time period. Typically, traffic simulation models represent only vehicle traffic (as opposed to person trip demands), so these demands are generally represented in terms of vehicle trips per hour. Usually, O-D demands cannot be directly observed (an obvious exception to this is when a roadway or network area is equipped with dedicated short range communication infrastructure, such as electronic toll tags, by which O-D data can be directly collected). Unfortunately, most traffic simulation models require O-D data as input, so the O-D demands must be derived via other means. Many methods have been developed for estimating O-D demands, but the most relevant for traffic simulation is the estimation of demands on the basis of link traffic counts. A review of the state-of-the-art of O-D demand estimation is available elsewhere (Hellinga, 1994).

Despite the significant research efforts directed at developing O-D estimation methodologies, several unresolved issues remain. First, there is generally no measure available by which the accuracy of the estimated demand can be quantified with respect to the true demand. Surrogate measures, such as those that compare the estimated and actual link flows, are useful, but they do not directly measure the accuracy of the estimated demand. Thus, there is often no means of determining how reflective the estimated demand is of the field conditions.

Second, the evaluation of advanced traffic management strategies generally requires the explicit modelling of the dynamic nature of the traffic flow characteristics and control strategies. Thus, dynamic, rather than static, O-D demands are typically required. The obvious question then is, How dynamic should the demands be?

The answer to this question is not obvious, but the two extremes provide boundaries. The most dynamic demand is one which is most disaggregate in terms of the departure time period duration. Since fractional vehicles have little physical meaning, the most disaggregate time period is the time period during which a single vehicle departs. Thus, the most dynamic demand is a list of each vehicle’s individual departure time.

The least dynamic demand has the longest departure time period duration. For practical purposes, this may be considered for traffic performance modelling (as opposed to planning), to be the duration of the peak period (say 3-4 hours).

The optimal departure period duration generally lies between these two extremes, and is usually in the range of 15 minutes to 1 hour, depending on the characteristics of the network and the study objectives.

PHASE – 3: EVALUATION OF MODEL OUTPUTS

This last phase attempts to reconcile model outputs with observed field data and determine if the model is adequately calibrated. If the model is not adequately calibrated, the question remains as to what must be done to reduce the discrepancies between the simulation model output and the field data?

If care has been taken during Phases 1 and 2 to choose appropriate MOPs, to establish quantitative criteria, and to complete the initial calibration, then the evaluation of the model results and the adequacy of the model calibration may be a well defined process. However, if the model results do not satisfy the stipulated criteria, some form of action is required. Normally, this action takes the form of re-calibrating one or more of the model input parameters. This is the most difficult part of the process, as it requires a well developed understanding of traffic flow theory, the local network conditions, and the model being used. No *turnkey* approaches exist by which the model user can immediately identify which of the input parameters require re-calibration.

The model user must examine the discrepancies between the model results and the field data and decide which simulation model input parameters to re-calibrate. Discrepancies may result from a single parameter (e.g. incorrect signal timing), but more typically result from the interaction of several inputs (e.g. incorrect representation of routing behaviour and O-D demands). Since many of these individual inputs cannot be evaluated directly (e.g. the accuracy of the estimated O-D demands cannot be directly assessed), the impact of any inaccuracies within the estimated input parameters cannot be separated from impacts from other input parameters.

Within Model Variability

An additional consideration is the impact of variability within the simulation model itself. Most simulation models, including INTRAS, FRESIM, NETSIM, CORSIM, and INTEGRATION, are stochastic models that use random number seeds within the program to generate a sequence of stochastic values. When the random number seed changes, the sequence changes, and the results provided by the model also change. Most practitioners tend to run the model once for each scenario being tested, without considering that each simulation result is a single point within a distribution of values. Studies of NETSIM have shown that misleading results can be obtained if the variability of simulation results is not considered (Benekohal and Abu-Lebdeh, 1994).

While studies have also been conducted to determine methods for reducing the number of simulation runs required to estimate the mean and confidence limits of a specific MOP (Joshi and Rathi, 1995), it is not clear that these methods can be easily incorporated within the objectives and constraints typically associated with non-research oriented studies.

CONCLUSIONS AND RECOMMENDATIONS

Policy makers and non-modellers often view simulation model results with scepticism. This scepticism often results from a lack of understanding of, and confidence in the model validation, verification, and calibration process. This confusion results from a general lack of clarity and consistency in the use and meaning of the terms *validation*, *verification*, and *calibration*, by modellers and non-modellers alike.

The processes of model validation and verification are primarily the responsibility of the model developers. Model calibration is the responsibility of the model user.

Successful model calibration requires that a systematic approach be followed, and that criteria that are consistent with the objectives of the study be developed.

In practice, the calibration activity is typically limited by constraints on resources for field data collection. Therefore, it is essential that data requirements be clearly and realistically defined, and that these requirements be prioritised on the basis of their relative importance to the calibration process. The creation of such a list enables the model user to make efficient use of available data collection resources, and also enables the model user and the client to identify the impact that increases or decreases in data collection effort has on the calibration process.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the financial support provided by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Benekohal, R.F and Abu-Lebdeh, G. (1994) *Variability Analysis of Traffic Simulation Outputs: Practical Approach for TRAF-NETSIM*. Transportation Research Record 1457. pp. 198 – 207.

Hellinga, B. (1996) *Test plan for the calibration of the INTEGRATION model to the I-696 Corridor*. Interim report prepared for MitreTek Systems, Inc., Washington DC.

Hellinga, B. (1994) *Estimating Dynamic Origin-Destination Demands from Link and Probe Counts.* Ph.D. Dissertation, Queen's University, Kingston, Ontario, Canada. (275 pages)

Hellinga, B., and Van Aerde, M. (1994) *An Overview of a Simulation Study of the Highway 401 Freeway Traffic Management System*. Canadian Journal of Civil Engineering, Volume 21. pp. 439 - 454.

Joshi, S.S. and Rathi, A. (1995) *Statistical Analysis and Validation of Multipopulation Traffic Simulation Experiments*. Transportation Research Record 1510. pp. 35 – 44.

Rakha, H., Hellinga, B., Van Aerde, M., and Perez, W. (1996) *Systematic Verification, Validation and Calibration of Traffic Simulation Models.* Presented at the 1996 Transportation Research Board Annual Meeting.

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