|  |  |
| --- | --- |
| Building Tomorrow’s Society  *Bâtir la Société de Demain* | csce_logo |
| Fredericton, Canada  June 13 – June 16, 2018/ *Juin 13 – Juin 16, 2018* |

**Calibration and Validation of Micro-Simulation Models Using Measurable Variables**

Tawfeek, Mostafa H.1,4, El Esawey, Mohamed2, El-Araby, Khaled3 and Abdel-Latif, Hatem3

1 Ph.D. Candidate, University of Alberta, Canada

2 Associate Professor, Ain Shams University, Egypt

3 Professor, Ain Shams University, Egypt

4 [mostafa.h.tawfeek@ualberta.ca](mailto:mostafa.h.tawfeek@ualberta.ca) (corresponding author)

**Abstract:** Traffic micro-simulation models have been extensively used to evaluate the impacts of traffic design alternatives. These models help stakeholders make informed decisions about any new change in the transport network. Calibration and validation of micro-simulation models are essential steps to ensure the reliability of the models. Most of the existing calibration efforts focus on experimental designs of driver behavior and lane changing parameters that have no physical meaning, or at least, cannot be easily measured. This paper suggests an approach for calibrating micro-simulation models using measurable variables. We advocate using an easy-to-understand, yet appropriate measures of the traffic stream such as spot speed as a calibration parameter. The approach was applied to a suburban corridor of about 12 km located in Greater Cairo Region. Video data were collected for about 9 hours on the corridor for both directions and were used to obtain traffic volumes and composition. Spot speeds were collected using a speed gun for a sample that comprised more than 2,000 vehicles belonging to different vehicle classes. VISSIM was used to model the corridor for only one direction that encompassed uninterrupted flow conditions. Spot speeds distributions were edited extensively in VISSIM to replicate the empirical distributions of the collected data. Furthermore, the distributions were developed for each vehicle type to achieve a higher level of accuracy. Statistical tests were carried out to ensure that the modeled speed distributions match the observed distributions. Corridor travel time was used as a validation parameter. Error measurements were computed for different analysis intervals (e.g., average travel time during a particular range), and it was shown that the validation errors decrease as the analysis period increases.

**Key Words:** Calibration and Validation, VISSIM, micro-simulation models

# **Introduction**

Micro-simulation models have been widely used by transportation experts and professionals to evaluate the effects of new geometric and traffic modifications. A comparison can hence be carried out between the current and the future situation in terms of the operational performance and potential gains. Simulation models should be reliable in a fashion that replicates real-life conditions so that the simulation results would enable making informed decisions. Model calibration and validation are two necessary steps to ensure the reliability of the developed model. The accuracy of the model outputs mainly depends on the quality of these two steps. This research explores the feasibility of using measurable variables for the calibration and validation of simulation models. In this approach, one or more variables are selected and measured with a high level of detail to enable a reliable calibration. One of the issues that are also explored in this research is the effect of the aggregation/measurement interval on the value of the error measurements used in evaluating the validity of the developed model. To elaborate, in the validation of micro-simulation models, data on one or more measures of effectiveness (MOEs) are collected and used for model validation. Examples of these measures include travel time, queue length, etc. In addition, this paper assessed the validity of the simulation model based on measures during different time periods. We hypothesize that the length of the aggregation interval could affect the estimation accuracy of the model and hence lead to a wrong assumption of a validated model.

# **Literature Review**

Micro-simulation has been a powerful tool for testing transportation-related solutions and designs before implementation. Applications of these models include studying route choice behavior (Talaat et al. 2007), analysis of unconventional intersection designs (El Esawey and Sayed 2007), testing and evaluating new signal optimization strategies (Kesur 2009), and, recently, assessing the impacts of intelligent vehicles applications (Tawfeek and El-Basyouny 2018). Recognizing the importance of calibrating and validating micro-simulation models, a considerable body of literature exists to address several issues related to the development, calibration, and validation of different micro-simulation models. In the calibration efforts of small network model (e.g., an isolated intersection, a corridor of a series of intersections, or a section of a corridor), the focus is usually on driver behavior and lane change parameters. The FHWA Guide (Dowling et al. 2004) defines calibration as the adjustment of a set of parameters in the simulation model to predict real-life conditions. The guide further recommends keeping this set of parameters as small as possible to minimize the calibration effort. Accordingly, the observed data will be non-adjustable variables in the calibration process to reduce the number of adjustable parameters. The FHWA Guide recommends the following three steps for calibration: capacity calibration, route choice calibration and finally system performance calibration.

Most of the current procedures for calibrating small network models start with identifying a set of calibration parameters along with their ranges. Some MOEs (e.g., travel time) are defined and used to compare the simulation-measured values of each scenario to field observations. An optimization algorithm is then utilized to determine the best set of parameters to make the model output matches field observations. Calibration procedures usually differ in terms of the number of calibration parameters and their ranges, the method for conducting the experimental design, and the optimization algorithm used to obtain the best set of parameters. For example, while calibrating a micro-simulation model for a congested freeway using VISSIM, Gomes et al. (2004) used Wiedemann 99 module and focused on three necessary lane change parameters (i.e., look back distance, emergency stop distance and waiting time before diffusion) and four car-following parameters. The selection of parameters was based on iterations with visual evaluation of the results and manual adjustment of the parameters. In another study by Kim et al. (2005), a simulation model of a signalized arterial was calibrated by adjusting Wiedemann 74 car-following and lane changing parameters. The calibrated car following parameters included the number of observed preceding vehicles, look ahead distance, average standstill distance, and desired safety distance. Adjusted lane changing parameters, on the other hand, included only lane change distance. The selected ranges for the calibration parameters were determined based on engineering judgment. A genetic algorithm was applied to choose the best set of calibration parameters.

Lownes and Machemehl (2006) aimed to examine the variations of the capacity of a congested freeway when changing driver behavior parameters. The authors used the same car following parameters presented in (Gomes et al. 2004) to replicate real-life conditions. Capacity, demand, and queue characteristics were used as measures to evaluate the calibration quality. A visual inspection was carried out to compare the queue characteristics of the model and compare it to real-life conditions. A sensitivity analysis was undertaken to study the impact of eleven driver behavior parameters on the estimated capacity. Chitturi and Benekohal (2008) developed a procedure to calibrate simulation models based on the relationship between capacity and two driver behavior parameters in VISSIM; standstill distance and time gap between vehicles. Although speed distribution was used, the focus was on the change in the capacity of freeways or work zones when the values of the parameters were modified. In a recent effort by Manjunatha et al. (2013), a methodology was presented to calibrate simulation models of heterogeneous traffic networks. Although the case study was for two intersections in an urban area, Wiedemann-99 car following model was used for its flexibility. The optimum values of the calibration parameters were chosen by an optimization formulation solved by a genetic algorithm. A sensitivity analysis showed that a combination of some car following parameters had a significant impact on the capacity. These parameters were defined for each vehicle type. The model was calibrated and validated using time delays as a measure of effectiveness.

Although many calibration procedures are currently available in the existing literature, it is noteworthy that these approaches depend mainly on adjusting model parameters that have no physical meaning or are difficult to obtain from field observations. In this paper, an approach is presented to utilize measurable variables such as spot speeds as the calibration parameters. High level of disaggregation is proposed where the measured variable is estimated for each particular vehicle class in the simulation model. The motivation behind this research is to apply a calibration methodology that is based on realistic data that have natural meaning and can be easily captured from the field.

# **Calibration and Validation Methodology**

Successful utilization of micro-simulation models is mainly dependent on their accuracy and reliability. Accordingly, it is essential to accurately calibrate and validate micro-simulation models before their use to derive critical design decisions. Calibration of micro-simulation models is the process by which model parameters are adjusted to minimize the differences between the model outputs and the observed values of some traffic parameters (Dowling et al. 2004). A similar definition is presented by Hollander and Liu (Hollander and Liu 2008) where the authors suggested that the purpose of calibration is to fine-tune a subset of driver behavior parameters so that the model output matches field observed data.

In the absence of a robust calibration and validation procedure, the accuracy of the model will always be questionable. Micro-simulation software packages often include several modules that interact with one another to generate the simulation visualization and output. Each of these modules has enormous parameters that can be manipulated to change the simulation results. A modeler seeking to calibrate a micro-simulation model has to be careful not to get jammed in an infinite loop trying to fix a problem by tweaking some parameters just to find another issue popping up elsewhere. As a result, it is necessary to develop a strategy that is based on a series of logical and sequential steps to calibrate the model promptly without compromising the required accuracy level. As well, it is also preferable to base the calibration on measurable variables rather than ad-hoc procedures using non-measurable calibration parameters. In this study, we advocate focusing on traffic variables to act as “calibration parameters” instead of manipulating user default parameters. The calibration variables should satisfy the following conditions 1) measurable in the field, 2) have a physical meaning, 3) representative of the traffic state of the modeled network, and 4) editable in the simulation model. Examples of these variables include spot speed distributions, acceleration and deceleration profiles, etc. We further hypothesize that detailed representation of the measured variable could lead to a well-calibrated model without the need to tweak any default parameter. The following steps describe the overall calibration and validation procedure:

1. Specify the dominant traffic conditions on the modeled network during the simulation period (e.g., free flow, congested, near capacity, etc.),
2. Select one or more calibration variables that are measurable, have physical meaning, editable in the simulation model, and most representative of the traffic conditions (e.g., spot speeds for free flow conditions, queue length for congested conditions),
3. Specify the evaluation criteria, based on which, the success of the calibration process will be judged (e.g., the similarity between the statistical distributions of simulated outcomes and field measures of the variable),
4. Recognize different sets of calibration variables (i.e., controllable inputs) that have an impact on the evaluation criteria (e.g., speed distribution of each vehicle class)
5. Set acceptable error margins and calibration targets,
6. Edit the calibration variables in the simulation model so that they become as close as possible to the field-measured values,
7. Run the model initially with the default parameters of the simulation model and evaluate the results,
8. If the results of the run satisfy the calibration targets, the model is considered calibrated. Otherwise, the model needs further calibration,
9. Edit the calibration parameter(s) in the model and repeat steps 6 and 7 to satisfy the calibration targets,
10. Specify the measures of effectiveness (MOEs) which will be used for model validation,
11. Compare the outputs of the calibrated model with the real-life values to validate the model, and
12. If the results were not satisfactory, repeat the calibration using more detailed data or another calibration variable until an appropriate accuracy level is achieved.

## **Evaluation Measures**

Evaluation criteria used to assess the quality of micro-simulation models’ calibration can be categorized into two major groups: traffic demand measures and system performance measures. Traffic demand measures are often represented by link volumes. Intuitively, an analyst cannot accept a micro-simulation model that generates traffic volumes which are not comparable to those measured from the field. In micro-simulation models of medium/large networks, differences in traffic volumes become an issue because these models are based on dynamic traffic assignment where demand is modeled using an Origin-Destination (OD) matrix. System performance measures, on the other hand, incorporate a wide range of traffic-related variables that can be both measured in the field and generated from the simulation model. These measures/variables can be used for both model calibration and validation. The extent (i.e., period and sample size) of data collection for model calibration is constrained by budget and time. The collected dataset should be split into two subsets: one for model calibration and one for model validation. This split can be either temporal or spatial. For example, if the dataset is collected for two days, data from the first day can be used for the calibration, while data from the second day can be used for the validation. Alternatively, if the data were collected from many entities of the network (e.g., a number of locations), data of some locations can be used for the calibration, while data of the other locations can be used for the validation. For a comprehensive literature review of different criteria used in the calibration of micro-simulation models, readers are directed to consult (Hollander and Liu 2008). The current study is concerned only with micro-simulation models with static routing (i.e., demand is modeled using traffic volumes rather than OD matrices). Hence, the calibration is aimed to assess system performance of the simulation model compared to real-life data.

## **Calibration Targets**

Calibration targets are determined according to the minimum performance requirements for the micro-simulation model keeping in mind the aim of developing the model and the available resources (*18*). The FHWA Guide (Dowling et al. 2004) includes examples of calibration targets. Although these targets were developed and used specifically for freeways, some of them have been widely applied in simulation models of urban areas (Oketch and Carrick 2005, Choi et al. 2008). The FHWA Guide (Dowling et al. 2004) suggests that calibration targets can vary according to the potential use of the model, the available resources of the analyst, and time constraints.

## **Model Validation**

Validation of a simulation model is the next stage after ensuring that the model is well-calibrated. Validation is defined as the process of matching the output of the calibrated simulation model with some real-life measurements that were not used in the calibration. The evaluation measure(s) used in this process should be different from the measure(s) used in the calibration process. Alternatively, the measurements of the same variable can be utilized but for various locations, time periods, and/or traffic conditions (Park and Schneeberger 2003). The model is deemed “valid” if the chosen MOE from the unused real-life dataset is close enough to the simulation values. Otherwise, the calibration process has to be re-executed until comparable results are achieved.

# **CASE STUDY: VISSIM MODEL OF 26th OF JULY OF CORRIDOR**

## **Test Site**

The corridor chosen for this study is a 12 km-suburban expressway section of the 26th of July corridor which stretches between the Ring Road East and Cairo–Alexandria Desert road (CAD road) west. Two locations were selected for the data collection where each location is about 1.5 km away from the starting/ending point of the corridor. This ensured that the flow is uninterrupted and unaffected by the bottlenecks and changes of the road cross section at the starting/ending points of the road.

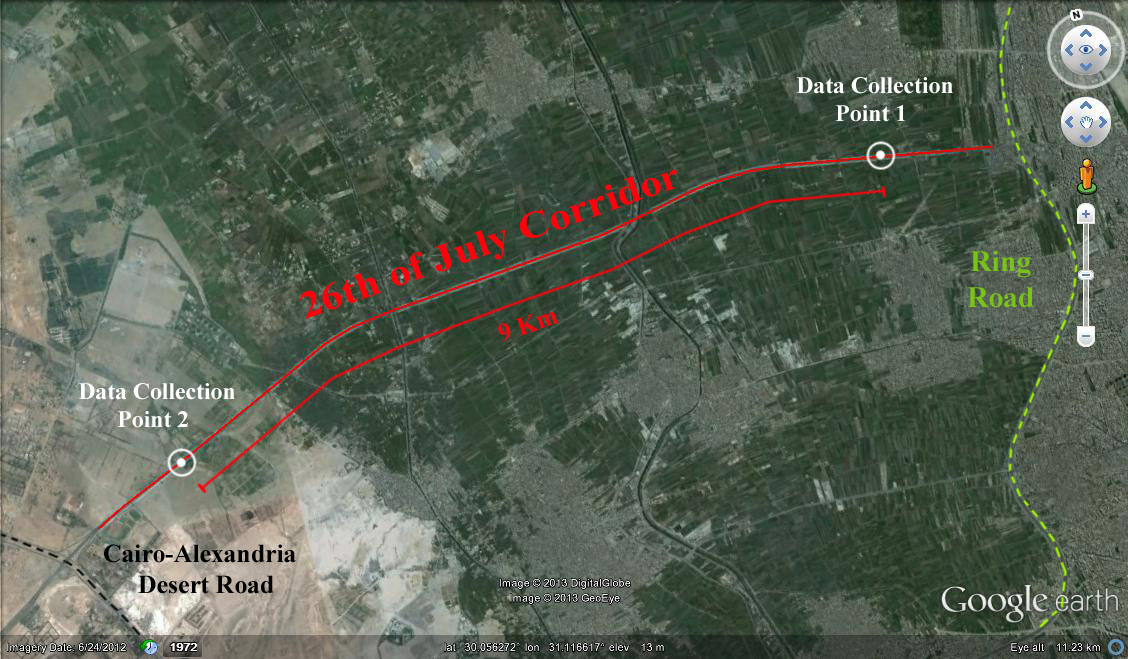


Figure : 26th of July corridor alignment and location of data collection points

The road cross section is divided and separated by New Jersey concrete barriers, where each direction consists of four travel lanes. The corridor is a freeway section as it has no U-turns crossovers, no access points, and no at-grade intersections. The average lane width is 3.5 m with no shoulders, and the posted speed limit is 90 km/hr. Figure 1 shows the alignment of the section and the data collection points locations.

## **Data Description**

Developing a reliable micro-simulation model requires the availability of a vast amount of traffic data. At the basic level, data of traffic volumes/composition and speeds have to be available. For the calibration and validation of the model, more detailed data have to be collected such as travel times and spot speed distributions. Video cameras were installed at the data collection points where the video recording took place between 9:00 a.m. and 6:00 p.m. to cover the peak hour at each direction. The data collection took place on a typical working day in the summer of 2012. Spot speed data were collected via a handheld radar speed gun of an accuracy of ±1.6 km/hr for a sample that included more than 2,000 vehicles.

### **Traffic Counts and Composition**

Volume data and traffic compositions were extracted at 15-minute intervals by manually observing the captured videos. No queues were observed at the data collection point located in the WB direction of the corridor. Nonetheless, long queues were observed at the other location between 3 p.m. and 6 p.m. As the corridor represents a divided roadway, each direction was modeled separately according to the prevailing traffic conditions. As mentioned earlier, the selection of the calibration variables mainly depends on the type of traffic conditions. Accordingly, measures such as spot speeds can be used for the calibration of the WB direction where no queue took place. On the other hand, the model of the EB direction can be calibrated using a measure such as the queue length. In this paper, we focus our analysis only on the WB direction which was operating at uncongested flow conditions. The recorded peak hour flow was approximately 5,400 vehicles for the hour between 9:00 and 10:00 a.m. with a peak hour factor of about 0.963 showing a uniform traffic distribution along the hour. Assuming an average lane capacity of 1800 veh/hr, the v/c ratio during the peak period would be 0.757 showing a reasonable utilization of the corridor. More than 33,000 vehicles were counted during the data collection period. Uniform distribution of the volumes was noticed during the afternoon period; 12:00 p.m. to 5:00 p.m. Corridor traffic was classified into six distinct vehicle classes: passenger cars (P.C.), which included private cars and taxis, microbuses (M.B.), heavy vehicles (H.V), buses, Motorcycles, and other vehicles (e.g., emergency vehicles). The traffic composition of the total traffic volume is 74.8%, 13.5%, 6.5%, 2.2%, 2.5%, and 0.4% for P.C., M.B., H.V., buses, Motorcycles, and other vehicles respectively.

### **Spot Speeds**

Spot speeds were measured for a sample that comprised 2,012 vehicles during the 9 hours of data collection. This sample represented about 6% of the total number of counted vehicles (i.e., 33,291). The sample size distribution per vehicle class is matching the actual traffic composition.

Figure 2: Average spot speeds (left) and average travel times (right) per vehicle class per hour

Figure 2 shows, on the left, the average spot speed per hour for each vehicle class type. Average spot speeds for P.C. varied between 91 km/hr and 97 km/hr while it ranged between 90 km/hr and 95 km/hr for M.B. It is also evident from Figure 2 that no significant differences existed between the recorded average spot speeds of P.C. and M.B. throughout the measurement period (i.e., speed attributes almost remained constant). Noteworthy is that the average spot speeds of buses and H.V. are less than the average spot speeds of P.C. and M.B. just due to the differences in vehicle characteristics.

### **Travel Times**

Travel times were obtained by matching vehicles passages in front of the two cameras and calculating the time elapsed by their passages. This method is similar to the concept of time stamps used to calculate travel times from Automatic Vehicle Identification (AVI) tags. However, the process was carried out manually in this study by human observers. Travel times were determined for each vehicle class and each hour for a total of 783 vehicles. The measured travel times were used to calculate the average values for each vehicle class and each hour of data collection. Figure 2 shows, on the right, the average travel time as estimated from the measured data. The free flow travel time on the corridor is about 360 seconds (i.e., 9 km length and a speed limit of 90 km/hr). As shown in Figure 2, the highest travel times for all vehicle classes were recorded between 10:00 and 11:00 a.m. which was not the peak hour of the traffic volume. This indicated abnormal conditions occurring within the section between the two cameras. Unfortunately, point measurements do not enable clearly identifying the reason for the high travel times during this hour. Nonetheless, it is expected that a number of slow vehicles or maybe H.V. and buses were running together on the corridor at the same time and resulted in temporary delays for all other vehicles.

## **Development of the Micro-simulation Model**

VISSIM is a stochastic time step behavior-based microscopic simulation software package that uses a Wiedemann psycho-physical car following logic to model traffic on urban streets and freeway environments. VISSIM is versatile and provides the flexibility to model any geometric configuration, a wide range of traffic operations in both the interrupted and uninterrupted traffic environment, and unique driver behaviors encountered within the transportation system (PTV 2011). VISSIM was chosen to simulate the selected section of the corridor where the behavior type was set to be freeway (i.e., free lane selection) which uses Wiedemann 99 car following model. The following sub-sections provide more details on the model development.

### **Network Coding**

Aerial photos were obtained from Google Earth TM, patched together and used as a background for network coding. Photos were scaled using the background scale feature in VISSIM to adjust the dimensions of the section. The number of lanes and lane widths were identified from field visits as mentioned earlier. Hourly traffic volumes were modeled for each of the 9 hours (i.e., from 9:00 to 18:00) of the simulation period. Traffic composition was edited separately for each hour according to the collected data. Transit lines schedules were created according to the observed headways during the data collection period.

### **Model Calibration**

In general, the target of calibration is to ensure that the simulation outputs of the evaluation variable match the field-observed data. Instead of calculating average errors to assess the quality of the calibration, a more robust approach is applied where the “matching” is evaluated using a suitable hypothesis statistical test. The null hypothesis is that both the simulation results and the field values of the calibration variables follow the same distribution. If the simulation output fits the observed data statistically, then the model is calibrated. Otherwise, the calibration variable would need further fine-tuning.

Spot speed per vehicle class was chosen as the calibration variable in the current analysis where the desired speed distribution was selected as the editable calibration parameter. Spot speed distribution is maybe the most influential calibration parameter for uninterrupted flow conditions of a simulation model. Spot speed distribution is usually specified in a simulation model according to the speed limit on the modeled network. The speed limit on the modeled corridor is 90 km/hr. Nevertheless, commuters usually drive at a speed higher than 90 km/hr due to the lack of enforcement as shown in Figure 2. Hence, it became essential to model the speed distributions according to the collected data regardless of the speed limit. The cumulative distribution of field-observed spot speeds was created for each vehicle class according to the collected data. As well, the cumulative distributions of simulated spot speeds were developed. The model will be a calibrated model if a statistical agreement is found between the cumulative distributions of the simulated and field-observed spot speeds.

Several Goodness-of-Fit (GoF) tests can be used to test the null hypothesis that the observed values of a variable are a random sample from a specified distribution. The most widely used test is the Chi-squared test that can be used for testing against discrete and continuous distributions. This test requires the sample of observations to be large enough for the approximations to be valid. Other types of GoF tests are based on the Empirical Distribution Function (EDF). The EDF tests offer advantages over the traditional Chi-squared GoF test, including improved power and invariance with respect to the histogram midpoints (D’Agostino and Stephens 1986). Kolmogorov-Smirnov (K-S) is an example of EDF tests which assumes that the observations come from the same statistical distribution as the null hypothesis. This hypothesis is rejected if the test statistic is higher than the tabulated critical value (D’Agostino and Stephens 1986). The K-S test was used in this study to assess the statistical agreement between the two cumulative distributions. The K-S test is only appropriate for testing data against a continuous distribution. As well, the K-S test is distribution free as the critical values do not depend on the specific distribution being tested. The K-S statistic is computed as the maximum of D+ and D-, where D+ is the largest vertical distance between the EDF and the distribution function when the EDF is higher than the distribution function, and D- is the largest vertical distance when the EDF is less than the distribution function. This is expressed mathematically as (Kim et al. 2005):

Where *F1*(*x*) and *F2*(*x*) are the cumulative probability density functions of *x*1 and *x2* (i.e., observed and simulated spot speeds). If the computed statistic *D* is greater than the critical tabulated value, then the null hypothesis at the chosen level of confidence is rejected.

### **Model Validation**

In this study, corridor travel times were selected as an MOE for model validation. Average travel time was determined for each vehicle class for both the observed and the simulated travel times. Instead of calculating a single value for the average travel time during the entire simulation period, average travel times were computed for different measurement (i.e., aggregation) intervals including 0.5-hour, 1 hour, and 3 hours. The purpose was to examine the effect of the length of aggregation interval on the validation error. It is expected that increasing the aggregation interval will decrease the error and it was desirable to quantify this effect. Different error measurements were used to evaluate the quality of the calibrated model. These included:

The Mean Absolute Percent Error (MAPE), calculated as:

The Root Relative Squared Error (RRSE), calculated as follows:

The Normalized Root Mean Square Error (RMSN), calculated as Balakrishna et al. (*17*):

Where:

*Travel Time Sim*= Simulated Travel Time,

*Travel Time RL*= Observed real-life Travel Time, and

*N* = Number of validation observation

Each of these error measures was applied to travel times of each vehicle class and to each time interval to make the validation more trustworthy. The FHWA Guide (Dowling et al. 2004) presented calibration criteria which generally suggest that a 15% error margin can be acceptable in similar exercises.

# **Results**

## **Model Calibration**

Initially, ten simulation runs were executed with the default parameters of VISSIM (i.e., with no calibration). The cumulative frequency distributions of the field-observed (i.e., actual) and the simulated spot speeds for different vehicle classes were compared using the K-S test which rejected the null hypothesis of distributions similarity. The results indicated that running the model with the default parameters would not be appropriate as shown in Table 1. Therefore, to calibrate the model, the desired speed distributions were manually edited for all vehicle classes to match the observed spot speeds cumulative frequency distributions closely. The model was re-run, and the simulated speed distributions of the calibrated model were estimated. From K-S test statistics, a close matching was found between the observed and simulated speed distributions as presented in Table 1. As shown in Table 1, the statistics of the calibrated model indicate a significant improvement over the results of the default model. The hypotheses of similarity were accepted, and hence, the model was successfully calibrated. Validation of the simulation model will be the last step to ensure that the model replicates field conditions.

Table 1: Results of K-S Test for Model Calibration

|  |  |  |  |
| --- | --- | --- | --- |
| Vehicle class | *Dmax* | | Improvement (%) |
| No calibration | After calibration |
| Passenger Cars | 0.283 | 0.12 | 58% |
| Microbuses | 0.28 | 0.099 | 65% |
| Heavy vehicles | 0.352 | 0.067 | 81% |
| Buses | 0.434 | 0.038 | 91% |

## **Model Validation**

The average travel times of the corridor were used for model validation. The simulation average travel times of the calibrated model were compared to the field-observed travel times. Different error measurements were computed for each vehicle class, and aggregation intervals and the results are presented in Table 2.

Table 2: Error measurements for model validation by vehicle class

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Vehicle class | MAPE | | | RRSE | | | RMSN | | |
| 0.5 hour | 1 hour | 3 hours | 0.5 hour | 1 hour | 3 hours | 0.5 hour | 1 hour | 3 hours |
| Passenger Cars | 3% | 3% | 3% | 5% | 5% | **3%** | 6% | 5% | **3%** |
| Microbuses | 5% | 4% | 3% | 7% | 5% | 4% | 7% | 6% | 4% |
| Heavy vehicles | 11% | 10% | 9% | 12% | 10% | 9% | 12% | 10% | 9% |
| Buses | 9% | 8% | 7% | 10% | 8% | 7% | 9% | 8% | 7% |

As shown in Table 2, all the error measurements were below 12% indicating a reasonable matching between the simulated and the actual travel times. Comparing the error values of different vehicle classes, it was found that the lowest validation errors were observed for the travel times of P.C. and M.B. Conversely, the travel times of H.V. exhibited the highest differences between the simulation and the field values. The reason is maybe that most of the H.V. observed on the corridor were unloaded, and hence they run faster than the simulated H.V. Another reason could be the small sample size of the measured travel times for the H.V. which might magnify the magnitude of validation errors. As also shown in Table 2, the validation errors decrease with the increase of the aggregation period (e.g., the error calculated from the average travel times during 1 hour is less than the error calculated during 0.5 hours). Nevertheless, the decrease in the amount of error is not significant (i.e., around only 2%). Overall, all validation results were satisfactory with minimal errors. Therefore, it can be concluded that the model is successfully calibrated and validated.

# **SummAry and conclusions**

A simple-yet-powerful procedure for the calibration and validation of micro-simulation models based on measurable variables was presented. The proposed method starts with running the model with the default parameters and comparing the simulated values and the field-observed ones for the calibration variable. If close matching is found between the two datasets, the model is deemed suitable and of no need for further calibration. Otherwise, the default parameters of the model would require fine-tuning. Unlike other calibration procedures that heavily depend on experimental designs and optimization algorithms for the selection of the best set of parameters, the method presented in this study relies on traffic variables that are measurable and have natural meaning. The procedure starts with identifying one or more measurable variables that are most representative of the traffic conditions on the network. The variable(s) also has to be editable in the simulation model and easy to capture in real-life. Examples of these variables include spot speeds for uninterrupted free flow environments and queue length for uninterrupted congested flow conditions. Detailed data have to be collected about that variable to enable a more reliable calibration. Such detailed data could include a larger sample, disaggregate data for different vehicle classes, or data for different times of the day. A case study was presented using a VISSIM model of a suburban freeway in Greater Cairo. Extensive traffic volume and spot speed data were collected for 9 hours on the corridor. Desired speed distribution per vehicle class was chosen as a measurable calibration parameter instead of using driver behavior parameters which are difficult to measure in real-life. All cumulative speed distributions were edited in VISSIM to closely match the observed spot speed distribution of each vehicle class. Statistical tests were performed to compare the simulated and the field values. The results showed the superiority of the calibrated model compared to the model with default parameters in terms of improved test statistics. The model was finally validated using average corridor travel time per vehicle class. Average travel times were calculated for different aggregation periods for both the simulated and the real-life values. Different error measurements were computed to assess the validity of the model. In general, all the errors were below 12% showing a reasonable matching between the observed and the simulated travel times. It was also revealed that the error decreases as the measurement interval increases, but the drops in error were insignificant. This conclusion should be treated with caution as the studied corridor exhibits uninterrupted flow conditions and hence the differences in error for different measurement intervals might not be large. In summary, the model could be calibrated and validated with reasonable trade-offs between accuracy and modeling efforts. The approach is logical as it depends only on measurable variables for calibrating simulation models. However, the method requires an extensive amount of data to ensure reliable calibration. The paper discussed some practical and theoretical issues that are essential for the development and application of micro-simulation models. These concepts are beneficial for both practitioners and researchers. As an extension of this paper, the calibration process using measurable and tangible variables from the field with interrupted/congested conditions will be explored.

References

Chitturi, M. V., and Benekohal, R.F. 2008. Calibration of Vissim for Freeways. Transportation Research Board 87th Annual Meeting,.

Choi, K., Jayakrishnan, R., Kim, H., Yang, I., and Lee, J. 2008. Dynamic OD Estimation using Dynamic Traffic Simulation Model in an Urban Arterial Corridor. *In* Transportation Research Board 88th Annual Meeting.

D’Agostino, R.B., and Stephens, M.A. 1986. Goodness-of-Fit Techniques.

Dowling, R., Skabardonis, A., and Alexiadis, V. 2004. Traffic Analysis Toolbox Volume III : Guidelines for Applying Traffic Microsimulation Modeling Software. Rep. No. FHWA-HRT-04-040, U.S. DOT, Federal Highway Administration, Washington, D.C, **III**(July): 146.

El Esawey, M., and Sayed, T.A. 2007. Comparison of Two Unconventional Intersection Schemes: Crossover Displaced Left-Turn and Upstream Signalized Crossover Intersections. Transportation Research Record: Journal of the Transportation Research Board, (2023): 10–19.

Gomes, G., May, A., and Horowitz, R. 2004. Congested Freeway Microsimulation Model Using VISSIM. Transportation Research Record, **1876**(1): 71–81.

Hollander, Y., and Liu, R. 2008. The principles of calibrating traffic microsimulation models. Transportation, **35**(3): 347–362.

Kesur, K.B. 2009. Advances in Genetic Algorithm Optimization of Traffic Signals. Journal of Transportation Engineering, **135**(4): 160.

Kim, S.-J., Kim, W., and Rilett., L.R. 2005. Calibration of microsimulation models using nonparametric statistical techniques. Journal of the Transportation Research Board 1935, (1): 111–119.

Lownes, N., and Machemehl, R. 2006. Sensitivity of Simulated Capacity to Modification of VISSIM Driver Behavior Parameters. Transportation Research Record: Journal of the Transportation Research Board, **1988**: 102–110.

Manjunatha, P., Vortisch, P., and Mathew, T. 2013. Methodology for the Calibration of VISSIM in Mixed Traffic. Transportation Research Board 92nd Annual Meeting,: 11. Available from http://docs.trb.org/prp/13-3677.pdf.

Oketch, T., and Carrick, M. (2005). 2005. Calibration and validation of a micro-simulation model in network analysis. In: TRB annual meeting, Washington, DC,.

Park, B., and Schneeberger, J. 2003. Microscopic Simulation Model Calibration and Validation: Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal System. Transportation Research Record: Journal of the Transportation Research Board, **1856**: 185–192.

PTV, A.G. 2011. VISSIM 5.40 user manual. *In* Karlsruhe, Germany.

Talaat, H., Masoud, M., and Abdulhai, B. 2007. A Simple Mixed Reality Infrastructure for Experimental Analysis of Route Choice Behaviour under ITS Applications. *In* Transportation Research Board 86th Annual Meeting.

Tawfeek, M.H., and El-Basyouny, K. 2018. Network-level Comparison of Various Forward Collision Warning Algorithms. Simulation: Transactions of the Society for Modeling and Simulation International,.

DOI: 10.1177/0037549718777613