Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 16 (2011) 282–292

Procedia
Social and Behavioral Sciences

6th International Symposium on Highway Capacity and Quality of Service
Stockholm, Sweden June 28 – July 1, 2011

Derivation of ATS and PTSF Functions for Two-lane, Rural Highways in Brazil

José Elievam Bessa Jr.[a], José Reynaldo Setti[a]*

[a]*University of São Paulo, São Carlos School of Engineering, Department of Transportation Engineering, Av. Trabalhador São-carlense 400, São Carlos, SP, Brazil – 13566-590*

Abstract

The HCM2000 uses average travel speed (ATS) and percent time spent following (PTSF) as measures of effectiveness to assess the level of service on two-lane highways. While ATS is difficult to observe directly, it is virtually impossible to obtain PTSF from traffic observations; thus, the HCM2000's procedure uses ATS-flow and PTSF-flow functions derived from simulation to estimate the level of service. The highway environment in Brazil is sufficiently different from the North-American one to invalidate the use of the original HCM2000 ATS and PTSF functions. The objective of this study was to adapt HCM2000's ATS and PTSF functions for two-lane rural highways in Brazil. Traffic data were collected in 11 locations, capturing a wide range of road and traffic conditions. These data were used to calibrate and validate TWOPAS, using a genetic algorithm (GA). Synthetic traffic data, similar to those collected by in-road traffic detectors, were generated for a wide range of conditions. The generation of synthetic data uses solutions created by the GA during the calibration of TWOPAS that were just marginally worse than the optimal calibration parameter set. The synthetic data set obtained consists of distributions of flow-ATS-PTSF observations for a hypothetical road segment representing a two-lane rural highway with ideal conditions in Brazil. The HCM2000 functions were then recalibrated using these synthetic data. New mathematical relationships were also investigated: a concave ATS-flow function and an exponential function for PTSF-flow. Besides being capable of representing several curve shapes (including those in the HCM2000), these new models are more compatible with the experimental data and are thus able to better represent the behavior traffic streams on Brazilian roads. These models could potentially be used to develop an adaptation of the HCM for Brazil.

Keywords: two-lane highways; speed-flow relationships; average travel speed; percent time spent following; simulation of traffic flows

1. Introduction

In Brazil, as in many other countries, the HCM2000 is used to estimate the level of service (LOS) and to assess the operational conditions of rural two-lane highways. The quality of service criteria for this class of highways is based on the average travel speed (ATS) and the percent time spent following (PTSF). Both these two measures are

* Corresponding author. Tel.: +55-16-3373-9596; fax: +55-16-3373-9602
E-mail address: jrasetti@usp.br.

1877-0428 © 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.sbspro.2011.04.450

difficult to be obtained directly from field observations, since they require observations of vehicles along road segments and not at specific points. To overcome this problem, the HCM2000 provides PTSF vs. flow and ATS vs. flow functions obtained from simulations to estimate LOS.

The differences between the road environment in Brazil and in the United States are enough to invalidate the use of the HCM2000 PTSF and ATS functions, which were developed based on data collected on North-American highways. The purpose of this paper is to present the method used to adapt HCM2000's ATS and PTSF functions for two-lane rural highways in Brazil. The proposed method uses a genetic algorithm (GA) to calibrate the traffic simulation model used (TWOPAS) and to generate synthetic traffic data, which is used to derive the ATS and PTSF functions. The method uses GA-generated data to reproduce the variability found in data collected by detectors, thus improving the reliability of the resulting functions. The paper is structured as follows: initially, it presents a brief review of the speed-flow models used by the HCM over the time. Afterwards, the overall structure of the proposed method is described, followed by detailed descriptions of each step in the process. The ATS and PTSF obtained for Brazilian two-lane, rural highways are discussed and compared to the ones provided by the HCM2000, followed by some concluding remarks.

2. Literature review

Chapter 20 of the HCM2000, which establishes the procedures for capacity and LOS estimation for two-lane highways, have been adapted to local conditions in Germany (FGSV, 2001) and South Africa (Van As, 2003). In Brazil, two studies have proposed adaptations of Chapter 20 (Egami, 2006; Mon-Ma, 2008). These studies used traffic simulation software, which was recalibrated and validated, to reproduce local traffic behavior and characteristics. Using the recalibrated versions of TRARR (Egami, 2006) and TWOPAS (Mon-Ma, 2008), new values for the parameters and equation coefficients were found, reproducing the simulation experiments used to develop the HCM2000 chapter on two-lane highways (Harwood et al., 1999).

One of the key aspects of the HCM approach to estimate the level of service for uninterrupted facilities is the speed-flow relationship. Since 1950, every edition of the HCM uses a speed-flow relationship to characterize operational conditions on two-lane highways. This approach is not only adopted in the HCM; the German highway capacity manual, the HBS2001, uses a similar approach in its procedure to estimate LOS on two-lane highways (FGSV, 2001).

Figure 1 illustrates the evolution of the speed-flow relationship used in the HCM. In the 1950 and 1965 editions, ideal conditions included average travel speed of 113 km/h (70 mph). The 1985 edition adopted 96 km/h (60 mph) as the average travel speed under ideal conditions. HCM2000 used a free-flow speed (FFS) of 110 km/h, under ideal conditions, as well as allowing for lower FFS (as low as 70 km/h), if conditions were less than ideal (Figure 1 shows only the speed-flow curve for ideal conditions). The speed-flow relationship adopted in HBS2001 is also shown in Figure 1. The general format of all functions shown in Figure 1 is basically the same: the maximum speed occurs under very low traffic and the speed decreases steadily as the traffic flow increases. Some models are linear (HCM2000 and 1965 HCM); the others are slightly concave (HBS2001, 1985 HCM and 1950 HCM). The shape of the function reflects the data used to describe the operational conditions. The fact that the HBS2001 function has a different shape than that used in HCM2000 shows the importance of adapting the speed-flow function to local conditions.

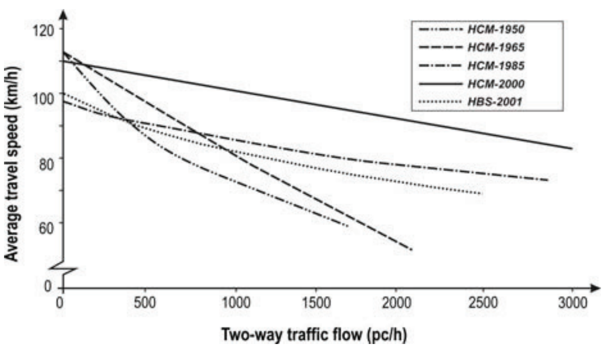

Figure 1: Speed-flow relationships used in the estimation of level of service on two-lane highways

McLean (1989) provides an interesting review of the empirical research on speed-flow relations. Summarizing conclusions of several studies, the conclusion is that most of them fitted linear functions to speed-flow data and that the general form of the model is:

$$V_1 = c - a \cdot Q_1 - b \cdot Q_o , \tag{1}$$

where V_1 is the mean speed of traffic in the primary direction (km/h); Q_1 is the primary direction flow (veh/h); Q_o is the opposing direction flow (veh/h) and a, b and c are constants. Most studies reviewed have suggested that the primary flow has a stronger influence on the speed than opposing flow, meaning that $a > b$. McLean, however, points out that several mathematical models of traffic flow on two-lane roads infer that the opposing flow should influence speed more than the primary flow, for equilibrium conditions on sections where overtaking is not impeded by sight distance restrictions. Empirical data collected in Canada supports the assertion that $a < b$ under these conditions (McLean, 1989).

The concave model adopted in the 1985 HCM has also been used to describe the relationship between speed and flow (Luttinen, 2000; Brilon and Weiser, 2006). Its general form is

$$V_1 = c - a \cdot \sqrt{Q_1} - b \cdot \sqrt{Q_o} \ . \tag{2}$$

Models fitted with data from highways in Finland (Luttinen, 2000) indicate that the effect of the opposing flow on speed is quite small (i.e., $a > b$); in Germany, where the practice is to use bidirectional analyses, the effect of opposing flow is disregarded – i.e., $b = 0$ (Brilon and Weiser, 2006). In both studies, the concave model fitted better than the HCM2000 linear model. Empirical data in both countries and a theoretical model applied to the German data confirmed the non-linear decrease of speed as the flow increased.

Research to develop the chapter on two-lane highways in the HCM2000 (Harwood et al., 1999) and in the HBS2001 (Brilon and Weiser, 1998) was based on the use of computer simulation. The main reason for this approach is the impracticality of obtaining empirical data to cover the full range of parameters (e.g., grade magnitudes and lengths, traffic volumes and so on), which can be easily overcome by using simulations. The next sections of this paper present a method to use a traffic simulation model to generate synthetic traffic data that reproduces the natural variability of actual data collected by road sensors. The use of this data can lead to models that better represent the real traffic.

3. The proposed method for synthetic traffic data generation

Traditionally, calibration of a simulation model is the search for a set of model parameter values which minimizes the differences between the observed and the simulated traffic streams (FHWA, 2004; Egami et al., 2006). Variability among runs is then achieved by using different random number sequence seeds; each replication of the simulation will produce slightly different results. This approach, however, is not capable of creating the same range of values found in real traffic streams. This section presents a method to generate a synthetic traffic data set whose variability is similar to that of data sets obtained from traffic detectors placed on highways.

The flowchart diagram shown in Figure 2 resumes the steps in the proposed method. Initially, two datasets are needed: one, from the roadway detector, usually covering a sufficiently long time period, and another one, comprising data that will be used for the simulation model calibration and validation. The detector data consist of speed, flow and traffic mix observations for 15-minute intervals collected over several months. As no weather information was available, traffic data could not be disaggregated for wet/dry pavement conditions. The segment where the detector was installed was level and its road cross section consisted of two, 3.5 m wide traffic lanes (one in each direction) and 2.5 m wide paved shoulders.

The calibration/validation data sets are collected over 4- to 6-hour intervals and are more detailed, including travel time (and average travel speed), headway and vehicle class for each vehicle that travels along the section (Egami et al., 2006).

For the research herein reported, the detector data are filtered to select the observations containing less than 3% heavy vehicles, as the PTSF and ATS functions are developed for streams containing only passenger cars. The filtered data are used to obtain the cumulative distribution of observed traffic flows which is used, in connection with the cumulative distributions of calibration parameter values, to generate scenarios.

A genetic algorithm (GA) is used to calibrate the simulation model, which, in this research, was TWOPAS. In the process to calibrate the model, the GA generates thousands of sets of calibration parameter values, whose quality is assessed by a *fitness function*. The best solution produces the highest fitness; a very large number of solutions, however, produce fitness values that are marginally lower than the maximum. In the proposed approach, these solutions are called *feasible solutions* and are used to generate cumulative distributions of feasible values for the calibration parameters.

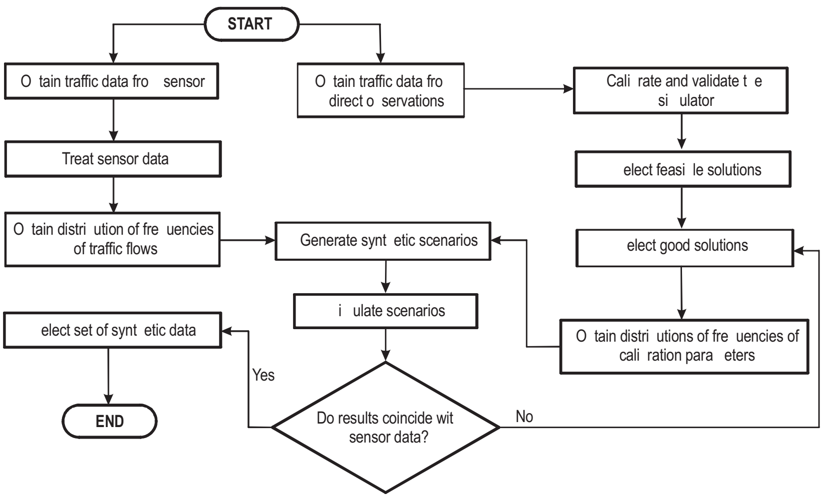

Figure 2: Flowchart diagram for the proposed approach to generate synthetic traffic data

In the next step, simulation scenarios are generated using Monte Carlo sampling. A scenario consists of a traffic flow rate and a set of values for the calibration parameters, randomly generated using the corresponding cumulative probability functions obtained in the previous steps.

Next, the simulation model is used to generate the synthetic data. Using the set of calibration parameter values and the flow rate, the simulation is replicated a number of times, using different seeds for the random number sequence. Once all scenarios are simulated, a visual comparison is carried out: if the synthetic data matches the observed data, the process is finished; otherwise, another set of feasible solutions should be chosen to generate the synthetic data. The next sections explain the proposed approach in greater detail.

3.1. Traffic Sensor Data Filtering

The treatment of the traffic data obtained from inductive loops installed on the road has two purposes: (1) to select the observation intervals for which the traffic was composed only by passenger cars; and (2) to select the observation intervals presenting uncongested traffic.

As the objective of the research was to obtain PSTF and ATS models for ideal conditions, only intervals presenting a negligible fraction of heavy vehicles were selected. Thus, only intervals with truck percentages equal to or less than 3% were selected. The next step in the data filtering was to separate observations of congested traffic from observations of free flow. For this, the data set that passed the first stage filter was segmented into 50 veh/h flow rate bands – i.e., 0 to 50 veh/h, 50 to 100 veh/h and so on. Cluster analysis was used to separate the congested flow observations from the free flow observations. Speed was adopted as the grouping criterion and the Euclidean distance was used as the distance metric (Bessa Jr. and Setti, 2010).

At the end of this stage, the distribution of observed flow rates for uncongested flow composed only by passenger cars was obtained. This distribution is used to create simulation scenarios using Monte Carlo sampling.

3.2. TWOPAS Calibration and Validation

Traffic simulation software is an important tool that has been used to develop the HCM since the 1985 edition because it allows complete control of traffic conditions. The simulation software used must, however, be able to represent the traffic stream behavior with a high degree of confidence, which can only be achieved by recalibrating the simulator to local conditions. Calibration is the process of adjusting the value of model parameters using observed data so that the model can realistically represent the system being simulated (Kim and Rilett, 2001). Traffic simulation software has a set of user-adjustable parameters for the purpose of calibrating the model to local conditions (FHWA, 2004). Calibration can also be considered as an optimization problem, where one searches for the set of parameter values that minimize the differences between the simulated and the observed traffic streams.

An optimization technique that has been successfully used in the automatic calibration of traffic simulation models is based on genetic algorithms (GA) (Kim and Rilett, 2001; Egami et al., 2006; Hollander and Liu, 2008) and searches for the best solutions not from a single point, but from multiple points, which increases the likelihood of finding a global optimum. A GA is an iterative procedure that starts with a randomly generated population. This population is a set of solutions and each solution is a vector with values for the calibration parameters, called chromosome. During each iteration (also called generation), each solution is evaluated and selected according to a fitness function which measures the quality of the results. Three genetic operators are used to create a new generation that will replace the previous population: crossover, selection and mutation. The first one uses two solutions (parents) to create new solutions (offspring) by exchanging segments of the parental chromosomes. The most common selection procedure is called elitism, in which one selects the best solution in a generation to produce the next generation by exchanging genetic material with the other individuals, thus assuring the transmission of the best traits from one generation to the next. Mutation is an operation that introduces unexpected variation in the solution to increase the chances of finding a global optimum (Goldberg, 1989).

The program used for the automatic calibration of TWOPAS is based on a previous version (Egami et al., 2006) and contains three modules: control, simulation and GA. The control module handles input/output data used or generated by the other modules; the simulation module controls TWOPAS runs; and the GA module applies genetic operators to chromosomes. More details about the GA can be found elsewhere (Egami et al., 2006).

The GA parameters include 250 generations as the stop criterion, crossover ratio of 0.5, mutation probability of 0.3 and an initial population of 20 chromosomes, created randomly. Each chromosome is a binary string sequentially combining the calibration parameters that have the greatest effect on the adopted performance measures and which were chosen using sensitivity analysis. Table 1 shows the definition, search range and number of bits for the calibration parameters.

The fitness function $F(I)$ used to assess how well the calibration parameter set (or chromosome) I can reproduce the observed traffic flow is calculated using:

$$F(I) = \alpha \cdot \exp\left[-\beta \cdot AMAER(I)\right], \tag{3}$$

in which α and β are constants whose values were adopted to be 100 and 5, respectively; and $AMAER(I)$ is the average mean absolute relative error ratio found for the k sites when using the calibration parameter set I (Egami et al., 2006):

$$AMAER(I) = k^{-1} \sum\nolimits_{i=1}^{k} MAER(I,i), \tag{4}$$

in which k is the number of sites used for the calibration data collection and $MAER(I,i)$ is the mean absolute error ratio for site i using chromosome I. $MAER(I,i)$ is calculated by:

$$MAER(I,i) = N^{-1} \sum_{j=1}^{N} \left| OBS(j) - SIM(j) \right| / OBS(j), \tag{5}$$

where N is the number of measures of performance used to compare the simulated and the observed streams; $OBS(j)$ is the observed value of performance measure j and $SIM(j)$ is the value of measure of performance j obtained from the simulation.

The performance measures used are percent traffic following, the average, the 15th-percentile and the 85th-percentile of the travel speed distribution. Percent traffic following is defined as the fraction of vehicles travelling with headways less than or equal to 3 seconds. For sections without climbing lanes, it was measured at both ends of the section (4 observation points); for segments containing climbing lanes, it was also measured at two other points. The average travel speed was obtained from the travel time in each direction. In addition to the average travel time, V_{15} and V_{85} were included in the calculation of $MAER$ to increase the fidelity of the simulation model. The use of these two percentiles of the travel speed tries to incorporate more information concerning the shape of the travel speed distribution in the model calibration. Taking this into account, the GA gives more importance to solutions that, besides providing a good estimate for average travel speed, also ensure that the travel speed distribution created by the simulation is similar to that observed in the field.

Table 1: Definitions, search range and length for TWOPAS parameters used by the automatic calibration system

Parameter	*Parameter definition*	*Search range*	*Bits*
PREC	probability that driver will reconsider starting a pass during one review period	$0.1 \leq PREC \leq 0.5$	3
VEAN	specified mean desired speed	$82 \leq VEAN \leq 104$	5
VSIG(1,1)	standard deviation for trucks in direction 1	$5 \leq VSIG(1,1) \leq 20$	4
VSIG(1,3)	standard deviation for cars in direction 1	$5 \leq VSIG(1,3) \leq 20$	4
VSIG(2,1)	standard deviation for trucks in direction 2	$5 \leq VSIG(2,1) \leq 20$	4
VSIG(2,3)	standard deviation for cars in direction 2	$5 \leq VSIG(2,3) \leq 20$	4
VBI(1,1)	bias to be added to mean desired speed (VEAN) for trucks in direction 1	$-22 \leq VBI(1,1) \leq 5$	5
VBI(1,3)	bias to be added to mean desired speed (VEAN) for cars in direction 1	$-5 \leq VBI(1,3) \leq 22$	5
VBI(2,1)	bias to be added to mean desired speed (VEAN) for trucks in direction 2	$-22 \leq VBI(2,1) \leq 5$	5
VBI(2,3)	bias to be added to mean desired speed (VEAN) for cars in direction 2	$-5 \leq VBI(2,3) \leq 22$	5
ZKCOR	car-following sensitivity factor	$0.6 \leq ZKCOR \leq 1.0$	3
Δcomp	used to calculate stochastic driver type factor BKPM(k), for k = 1 to 10	$-0.20 \leq \Delta comp \leq 0.20$	6
η	used to calculate effective weight/net horsepower ratio for trucks	$0.60 \leq \eta \leq 1.20$	6

At the end of the calibration procedure, the best solution produced $AMAER = 0.068$ and $F = 71.17$, while the worse solution showed fitness F close to 44. To validate the calibration, a second data set (collected on a different occasion) was used to compare the results of the recalibrated TWOPAS to the real traffic. This comparison produced $AMAER = 0.061$ and $F = 73.71$, which indicates that the recalibrated version of TWOPAS is capable of realistically simulating the behavior of a typical two-lane highway in Brazil.

The solutions generated by the AG were stored and classified according to their fitness value. The 15th percentile of the solutions corresponded to $F \geq 69$ (up to 3% worse than the best solution) and the 50th percentile corresponded to $F \geq 66$, or 7% lower than the best solution fitness. The next section explains their use in the generation of the synthetic traffic data.

3.3. Synthetic Traffic Data Generation

The pseudocode used for the generation of traffic synthetic data is:

```
Read the desired number of simulation replications M;
Read the vector of random number seeds R(M);
Read the desired number of observations N;
For i = 1 to N:
   Generate the flow rate V(i) by sampling the cumulative distribution of flow rate
      frequency;
   Generate values for the vector of calibration parameters C(i) by sampling the cumulative
      distributions of frequency for each calibration parameter;
   For j = 1 to M:
        Run TWOPAS using the traffic flow rate V(i), calibration parameter vector C(i) and
           random number seed R(j);
        Obtain traffic flow rate Vsim(j) and average travel speed ATS(j) from TWOPAS output;
   End;
End;
```

The generation of the flow rate and the calibration parameter values uses Monte Carlo sampling. A random number r between 0 and 1 is generated; the parameter takes the value that corresponds to the r-th percentile of the cumulative distribution of the parameter being sampled. The distributions of calibration parameter values used in the

Monte Carlo sampling are obtained from the best solutions created by the GA during the calibration of TWOPAS. Taking a generic calibration parameter as an example, Figure 3(a) shows the cumulative distributions of its values obtained using the 15% best solutions created by the GA – which, in this case, corresponds to those with fitness $F \geq 69$. Figures 3(b) and 3(c) respectively show the cumulative distributions of parameter values associated with the 50% best solutions ($F \geq 66$) and all the solutions created by the GA ($F \geq 44$). The cutoff level of fitness used to characterize the best solutions should be determined by a sensitivity analysis, which compares the variability of the synthetic data to the variability of the sensor data.

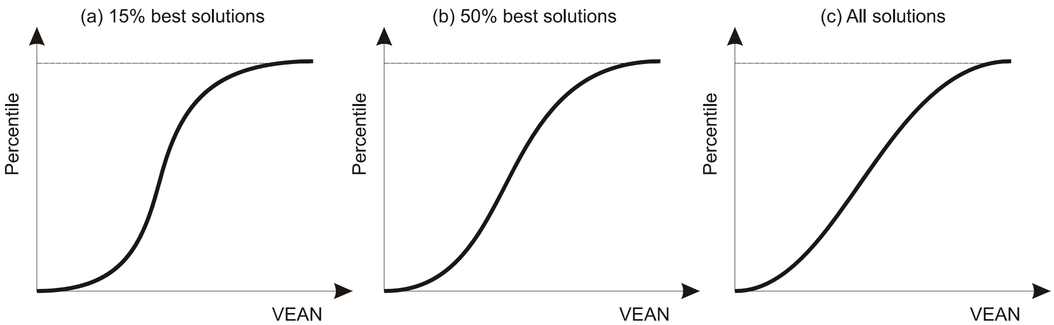

Figure 3: Cumulative distributions of VEAN values obtained from the solutions created by the GA for the calibration of TWOPAS

Plots of the synthetic data used for the sensitivity analysis for the case study are shown in Figure 4(a) to (d). A total of 6000 traffic flow rates were simulated, which corresponds roughly to the number of observations of 15-min intervals with less than 3% trucks. Each flow rate was replicated five times and the directional split was randomly chosen. It can be noted from the data plots that the variability of the synthetic data increases with the number of solutions used to create the distribution of calibration parameter values. The synthetic data generate with solutions with $F \geq 69$ (15% best GA solutions) present less variation than the observed data; whereas the variability of the synthetic data produced using all solutions is significantly greater than the variability of the sensor data. A visual inspection shows that the use of calibration parameter sets with $F \geq 59$ (85% best solutions) produces approximately the same variability associated with the sensor data. Thus, this set of solutions was used to create the distributions of calibration parameter values used in the proposed method.

4. ATS and PTSF model estimation

Once TWOPAS was recalibrated, the synthetic data were used to calibrate ATS and PTSF functions for Brazilian two-lane highways. Synthetic data were used instead of sensor data because of two reasons: the small number of road sections for which data under ideal conditions were available and the scarcity of observations close to capacity. Road sensor data typically includes a very large number of observations at low flow rates and very few observations of flows close to capacity. To avoid this problem, a set of synthetic data points covering the range of flows was generated as explained now.

A total of 500 flow rates randomly chosen flow rates between 0 to 1700 passenger cars/hour (for directional analyses) or between 0 and 3200 pc/h, for bidirectional analyses. A 50/50 directional split was adopted for the data used to estimate bidirectional models. For data used to estimate directional models, the flow rate in the direction of interest and on the opposite direction were chosen randomly, subject to the restriction that the sum of the two flow rates should be less than or equal to 3200 veh/h.

The synthetic data set was created simulating a traffic stream containing only passenger cars over a 10-km section with ideal geometric conditions for each of those 500 flow rates. Each flow rate simulation was replicated five times using different random number series seeds. Each synthetic data ser contained 2500 data points, each consisting of flow rate, ATS and PTSF.

Two formulations for the ATS model were analyzed: the linear model used in the HCM2000 and the model adopted by the 2001 edition of the German Highway Capacity Manual (HBS2001). In the HCM2000 linear model, the relationship between the average travel speed *ATS* and the bidirectional traffic flow rate q (in pc/h) is:

$$ATS = FFS - a \cdot q \,, \tag{6}$$

where FFS is the free flow speed (in km/h) and a is a constant. The linear model for the directional average travel speed ATS_d is:

$$ATS_d = FFS_d - a_1 \cdot q_d - a_2 \cdot q_o \,, \tag{7}$$

with FFS_d being the free flow speed in the direction of interest; q_d, the flow rate in this direction; q_o, the flow in the opposite direction; and a_1 and a_2 being constants.

The HBS2001 model for the speed-flow relationship is (Brilon and Weiser, 1998):

$$ATS = FFS - b \cdot \sqrt{q} \,, \tag{8}$$

with b being a constant. The model for the average travel speed ATS_d in one direction is:

$$ATS_d = FFS_d - b_1 \cdot \sqrt{q_d} - b_2 \cdot \sqrt{q_o} \,, \tag{9}$$

in which b_1 and b_2 are calibration constants.

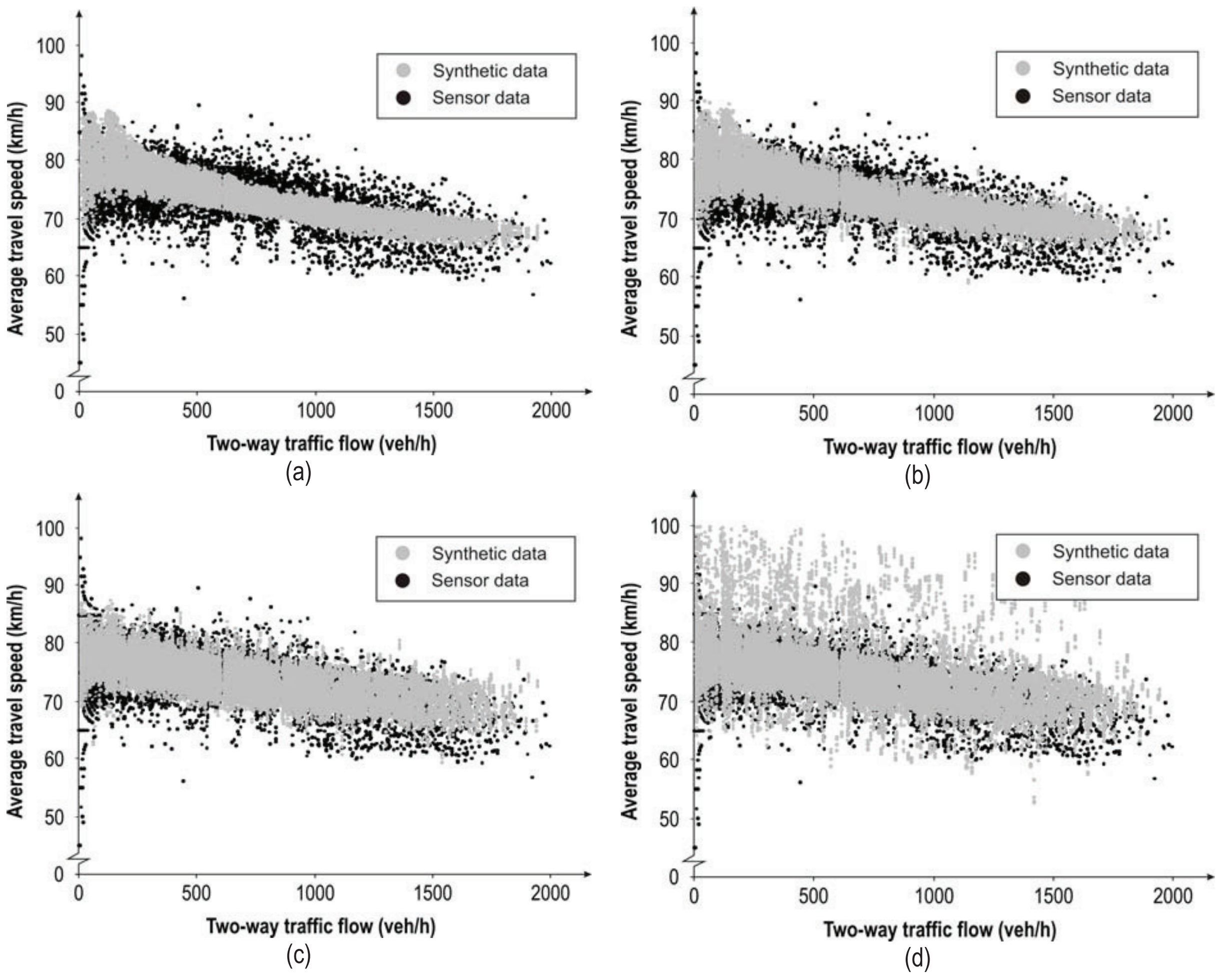

Figure 4: Results of the sensitivity analysis conducted to find the solutions used to create the distributions of calibration parameter values. The analysis focused on the variability of synthetic data generated using (a) only solutions with $F \geq 69$ (15th percentile); (b) only solutions with $F \geq 66$ (50th percentile); (c) only solutions with $F \geq 59$ (85th percentile); and (d) 100% of the solutions created by the GA ($F \geq 44$)

Using synthetic traffic data to estimate the regression coefficients allows for the incorporation of greater behavioral variability into the analysis than by using the traditional approach. Figure 5(a) shows passenger car speed as a function of two-way traffic flow rate, using the traditional approach, in which the best values of the calibration parameters are used in every scenario, which is then replicated a number of times using different random number sequences to account for the natural stochasticity of traffic flows. Figure 5(b) shows that the synthetic data generated using the proposed approach can better reproduce the natural variability of the traffic stream.

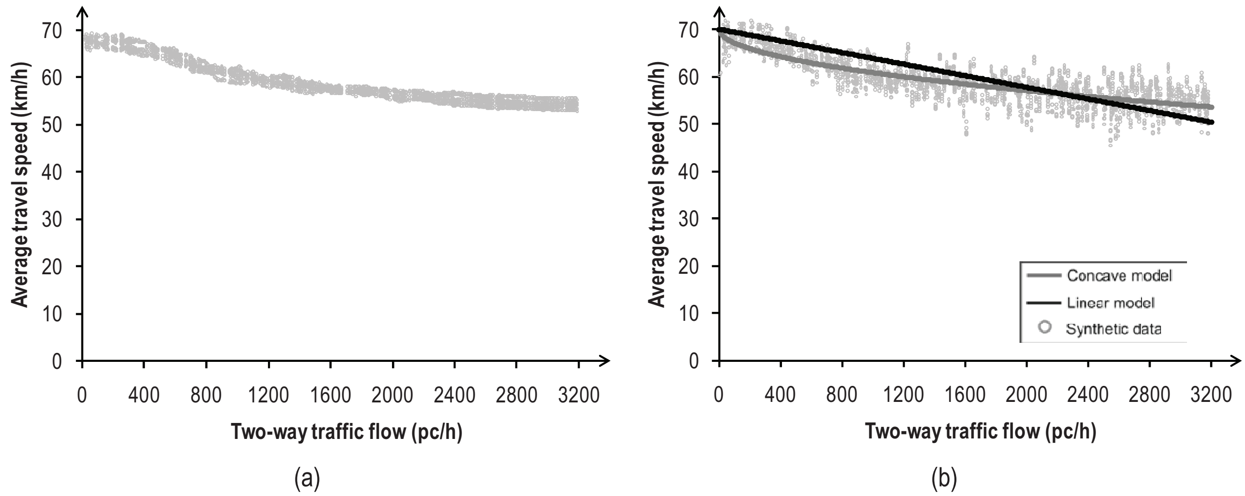

Figure 5: Synthetic data for the calibration of ATS models: (a) generated using the traditional approach; and (b) generated using the proposed approach

Table 2 summarizes the models, listing values for the constants and R^2, which were obtained by linear regression. It can be noted that the concave models (Eq. 8) provided a marginally better fit to the data than the linear models (Eq. 6).

Table 2: Linear regression coefficients and goodness-of-fit for ATS models for bidirectional and directional analysis

(a) Bidirectional flow

FFS	Linear		Concave	
(km/h)	*a*	R^2	*b*	R^2
70	0.0062	0.92	0.2903	0.95
80	0.0058	0.92	0.2749	0.95
90	0.0056	0.92	0.2665	0.96
100	0.0054	0.91	0.2598	0.95
110	0.0053	0.90	0.2536	0.95

(b) Unidirectional flow

FFS	Linear			Concave		
(km/h)	a_1	a_2	R^2	b_1	b_2	R^2
70	0.0047	0.0063	0.92	0.1361	0.2193	0.93
80	0.0046	0.0055	0.93	0.1393	0.1897	0.94
90	0.0046	0.0052	0.91	0.1398	0.1813	0.92
100	0.0044	0.0052	0.92	0.1310	0.1810	0.94
110	0.0042	0.0053	0.93	0.1228	0.1853	0.95

The structure of the percent-time-spent-following the model used in the HCM2000 for bidirectional analysis is:

$$PTSF = 100\left[1 - \exp(-a \cdot q)\right], \qquad (10)$$

and, for directional analysis, is

$$PTSF_d = 100\left[1 - \exp(-b \cdot q_d^c)\right], \qquad (11)$$

where *PTSF* and $PTSF_d$ are the percent time travelling in platoons; q is the bidirectional traffic flow rate; q_d is the traffic flow rate in the direction of interest; and a, b and c are constants.

The synthetic data for the estimation of PTSF functions were generated in a similar way to the data used for the ATS models. Flow rates in the direction of interest were randomly chosen for the data used to calibrate the directional models; flows on the opposite direction varied from 200 to 1600 veh/h in 200 veh/h increments.

Figure 6(a) shows the PTSF model obtained by fitting Eq. (10) to the synthetic data. Although the coefficient of determination was found to be quite good (R^2 = 0.96), it can be noted that the model overestimates PTSF for bidirectional flows less than 400 cpe/h and underestimates it for flows between 500 and 1500 pce/h.

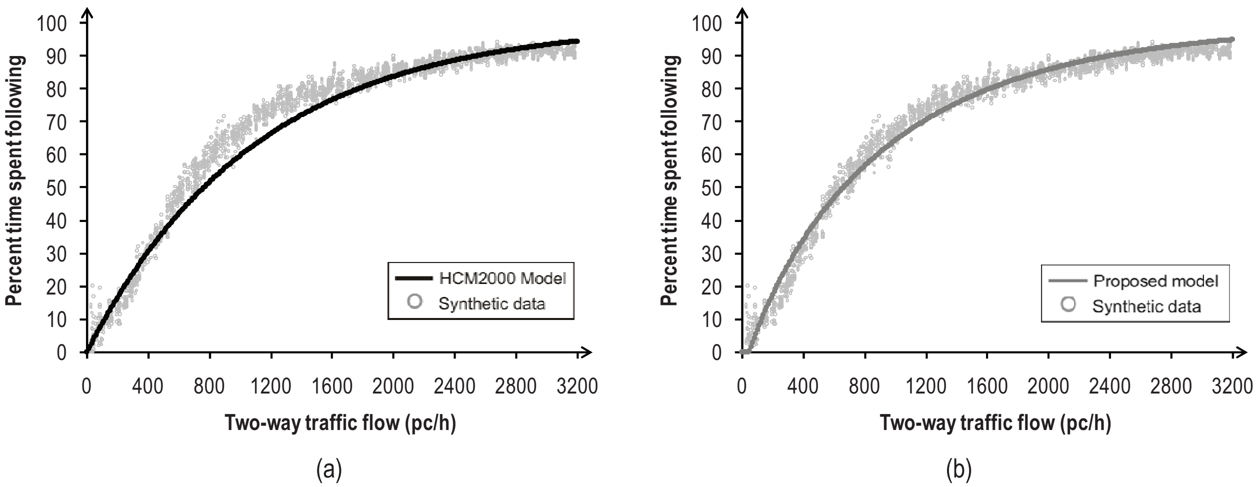

Figure 6: Percent time spent following models for bidirectional analysis: (a) model fitted using the same structure as in the HCM2000; and (b) proposed model

An alternative structure for the PTSF model was used, in which the bidirectional model is given by

$$PTSF = 100\left[1 - a \cdot \exp(-b \cdot q^c)\right], \tag{12}$$

and the directional model is

$$PTSF_d = 100\left[1 - a \cdot \exp(-b \cdot q_d^c)\right]. \tag{13}$$

It should be pointed out that Eq. (10) is the same as Eq. (12), when $a = 1$ and $c = 1$, and that Eq. (11) and Eq. (13) are the same if $a = 1$. Moreover, because Eqs. (12) and (13) can produce $PTSF < 0$ when $a \neq 1$, a restriction should be imposed such that

$$PTSF = \begin{cases} 0, \text{if } 100\left[1 - a \cdot \exp(-b \cdot q^c)\right] < 0 \\ 100\left[1 - a \cdot \exp(-b \cdot q^c)\right], \text{else.} \end{cases} \tag{14}$$

Figure 6(b) illustrates the proposed model for bidirectional analysis, in which $a = 1.079$; $b = 0.0026$ and $c = 0.878$ ($R^2 = 0.98$). The fitted PTSF models for directional analysis are shown in Table 3 as a function of the opposing flow. Table 3(a) shows values for model parameters for the HCM2000 approach and Table 3(b) shows parameter values for the proposed model structure. For all cases, the proposed model is better fitted to the data than the HCM2000 model, as measured by R^2. It should be noted in Table 3(b) that the values found for the model coefficient a are significantly different from 1, which is its value in the HCM2000 PTSF model.

Table 3: Percent-time-spent-following model parameters for directional analysis

(a) HCM2000-type model

Opposing flow (pc/h)	***b***	***c***	R^2
200	-0.004	0.871	0.98
400	-0.008	0.777	0.95
600	-0.011	0.745	0.95
800	-0.014	0.721	0.95
1000	-0.014	0.724	0.94
1200	-0.014	0.731	0.95

(b) Proposed approach

Opposing flow (pc/h)	***a***	***b***	***c***	R^2
200	1.084	-0.007	0.783	0.98
400	1.119	-0.017	0.680	0.96
600	1.162	-0.026	0.631	0.96
800	1.193	-0.034	0.598	0.95
1000	1.212	-0.037	0.594	0.95
1200	1.196	-0.034	0.608	0.95

1400	-0.014	0.729	0.94	1400	1.211	-0.037	0.599	0.95
1600	-0.013	0.739	0.94	1600	1.197	-0.033	0.613	0.95

The observation of Figures 6(a) and (b) shows that the new PTSF model is slightly better than the HCM2000-type model; likewise, the results summarized in Table 3 also show the same tendency. A question that still remains, however, is whether the improvements are enough to justify the adoption of the new model.

5. Concluding remarks

This paper has shown the derivation of ATS and PTSF models for Brazilian two-lane highways using synthetic data. While the ATS and PTSF functions obtained are applicable only to two-lane highways in Brazil, the method used for the generation of synthetic data can be applied anywhere. This method is based on parameter values used by a genetic algorithm while calibrating the model to the local conditions. The synthetic data set has the same degree of variability found in actual traffic data collected using road sensors, thus increasing the quality of the models. If weather information be available, the method can be easily modified to generate synthetic data for dry conditions.

The ATS and PTSF functions resulting from this research should replace the ones given in the HCM2000, for directional and bidirectional analysis. The best model for ATS was a concave model similar to the one used in the HBS2001. An alternative form for the PTSF functions proved slightly better than the HCM2000-type models.

Acknowledgements

Financial support for this research was provided by CNPq (National Council for Scientific and Technological Development) through grants 303341/2008-4 and 131007/2007-6. Traffic data and support for data collection at weigh stations was provided by DER-SP.

References

Bessa Jr., J. E. & Setti, J. R. (2010) Synthetic traffic data generation using a genetic algorithm and microscopic simulation (in Portuguese). *Transportes*, paper accepted for publication in a forthcoming issue.

Brilon, W. & Weiser, F. (1998) Capacity and speed-flow relationships on rural two-lane highways in Germany. *Transportation Research Circular: Third International Symposium on Highway Capacity*, p. 199-218.

Brilon, W. & Weiser, F. (2006) Two-lane rural highways: the German experience. *Transportation Research Record: Journal of the Transportation Research Board, v. 1988*, p. 38-47.

Egami, C. Y. (2006) *Adaptation of HCM2000 for Level of Service Estimation of Two-Lane Highways in Brazil* (in Portuguese). PhD thesis, University of São Paulo, São Carlos School of Engineering, São Carlos, SP, Brazil.

Egami, C. Y., Mon-Ma, M. L., Setti, J. R. & Rilett, L. R. (2006) Automatic Calibration of Two-Lane Highway Traffic Simulation Models Using a Genetic Algorithm. In: *9th International Conference on Applications of Advanced Technology in Transportation*, ASCE Press, p. 510–515.

FGSV (2001). *Handbuch zur Bemessung von Strassenverkehrsanlagen* (HBS, German Highway Capacity Manual). Forschungsgesellschaft für Straßen- und Verkehrswesen, Cologne, 2001.

FHWA (2004). *Traffic Analysis Toolbox Volume III: guidelines for applying traffic microsimulation modeling software*. Federal Highway Administration.

Goldberg, D. E (1989). *Genetic algorithms in search, optimization and machine learning,* Addison-Wesley, Inc., Massachussets.

Harwood, D. W., May, A. D., Anderson, I. B., Leiman, L. & Archilla, A. R. (1999) *Capacity and quality of service of two-lane highways.* Publication NCHRP Project 3-55 (3). Midwest Research Institute.

Hollander, Y. & Liu, R. (2008) The principles of calibrating traffic microsimulation models. *Transportation, v. 35*, p. 347-362.

Kim, K & Rilett, R. L. (2001) Genetic algorithm based approach for calibration microscopic simulation models. *IEEE Intelligent Transportation Systems Conference Proceedings*, p. 698-704.

Luttinen, R. T. (2000) Level of service on Finnish two-lane highways. *Transportation Research Circular E-C018: Fourth International Symposium on Highway Capacity*, p. 175-187.

McLean, J. R. (1989) *Two-lane highway traffic operations – Theory and practice*. Gordon and Preach Science Publishers, Inc., Amsterdam.

Mon-Ma, M. L. (2008) *Adaptation of HCM-2000 to typical two-lane highways with auxiliary lanes in the state of São Paulo* (in Portuguese). PhD thesis, University of São Paulo, São Carlos School of Engineering, São Carlos, SP, Brazil.

Van As, C. (2003) *The development of an analysis method for the determination of level of service of two-lane undivided highways in South Africa - Project Summary*. South African National Roads Agency Limited.