

Simulation-Based Regression Models to Estimate Bus Routes and Network Travel Times

Yaser E. Hawas, United Arab Emirates University

Abstract

This paper presents an approach to estimate bus route and network travel times using micro-simulation. This can be used in predicting the effectiveness of bus route designs using some network traffic measures or indicators. The used indicators are average network traffic intensity, posted speeds, route length, frequency of bus operation, and average passenger loadings (boarding and alighting). Regression models are calibrated to predict both route and overall network travel times. The prediction errors of these models were investigated and analyzed, and regression models were validated. Results indicated the validity of the calibrated regression models. Conclusions are made on how the devised models can be validated in reality and used for route planning purposes to determine best operating conditions such as the frequency.

Introduction

The performance of a transit network depends on the effective planning and design of transit routes. To ensure effective planning of transit networks, it is important to develop tools or methods to characterize network effectiveness as a function of frequency, route design, and other factors such as traffic network intensity and passenger loadings. Such methods or tools are eventually needed to assist transport agencies in transit planning applications, alteration of service

schedules, devising of enhancement policies, macro-management of operation, and, ultimately, better service for transit users.

Transit effectiveness measures are needed to quantify how efficiently transit system inputs are used in producing a given output (Nash 2006). Among the common effectiveness indicators of transit network design are the overall network and route travel times. The lesser the travel times of the designs needed to provide service to specific transit demand, the better is the design and the more attractive is the service to transit users. The effectiveness indicators are influenced by many factors, such as number of bus stops on routes, number of passengers boarding and alighting, speed restrictions, route length and alignments, etc.

In general, the factors that affect travel times include human, vehicular, and facility aspects. Different drivers and road conditions could cause large differences in journey times. For the same time interval and on the same link, different vehicles can have quite different travel times (Li and McDonald 2002). Free-flow travel speed is another factor that affects network travel time. Journey speed along an arterial road depends not only on the arterial road geometry but also on the traffic flow characteristics and traffic signal coordination (Lum et al. 1998). Other main factors cited in previous studies include incidents (Karl et al. 1999), signal delay (Wu 2001), weather conditions (Chien and Kuchipudi 2003), and traffic congestion levels (Lin 2005). Speed (Chien 2003), frequency, and number of boarding and alighting passengers of bus service (Tetreault 2010) have been used for route and network average travel time prediction.

The use of travel time information is essential for long-term design of transit service as well as scheduling. In relatively stable light traffic conditions, with light transit demand, fairly simple estimation procedures may be used to estimate travel times. On the other hand, in rapidly-changing traffic conditions, using sophisticated prediction models is essential (Van Grol et al. 1999). Different studies suggested different techniques for estimating or predicting travel times (Kwon et al. 2003; Chakraborty and Kikuchi 2004; Zhang and Rice 2003; El-Geneidy et al. 2010; Tetreault and El-Geneidy 2010).

Kwon et al. (2000) used linear regression and advanced statistical methods to develop models for predicting travel time. Chakraborty and Kikuchi (2004) developed a simple linear equation using regression analysis to predict automobile travel time based on bus travel time. Zhang and Rice (2003) developed a linear model with time varying coefficients for short-term travel time prediction. Multivariate

regression models were used to predict run time, schedule adherence, and reliability of the transit route (El-Geneidy et al. 2010; Tetreault and El-Geneidy 2010)

Predicted travel times using artificial neural networks have been found to be more accurate than the other methods (Waller et al. 2007). Huisken and Van Berkum (2003) developed a travel time prediction method based on artificial neural networks and compared them with the currently-used travel time prediction model for a corridor in the Netherlands. Van Lint et al. (2002) used recurrent neural networks to predict freeway travel time. Van Lint (2003) extended this work to develop an approach to quantify the uncertainty around the travel time predictions. Mark et al. (2004) conducted a comprehensive statistical analysis of the impact of various factors, such as temporal resolution of the data, speed, and flow on the experienced travel time predictions obtained using artificial neural networks in the presence of incidents. The data for this study were synthetically generated by simulation using a Cell Transmission Model as the traffic flow model.

Many researches used the Kalman Filtering (KF) algorithm for predicting travel time (Nanthawichit et al. 2003; Kuchipudi and Chien 2003; Chien and Kuchipudi 2003; Chen and Chien 2001). The KF algorithm was first applied by Okutani and Stephanedes (1984) to predict traffic volumes in an urban network. Nanthawichit et al. (2003) developed a method for short-term travel time prediction by combining a KF approach with a macroscopic traffic flow model. Kuchipudi and Chien (2003) developed a model in which both path-based data and link-based data are used to predict travel times using a KF framework. Chien and Kuchipudi (2003) used a KF algorithm for short-term prediction of travel time; the study used a combination of historical and real-time data. Chen and Chien (2001) used the KF technique for dynamic travel time prediction based on real-time probe vehicle data.

Simulation has become a popular and effective tool for analyzing a wide variety of dynamic problems that are associated with complex processes (Ni 2001). In the transportation field, the application of simulation is widely extending from small applications such as traffic signal optimization to wide-scale applications such as evaluating the national transport strategy. Simulation can capture statistics on the variability of the characteristics (Fishburn and Taaffe 1994). Several researchers have used simulation to generate traffic data (Anderson and Bell 1998). Chien and Kuchipudi (2003) used traffic simulation to develop a model for predicting travel times based on spot speed/volume data obtained from sensors on a freeway corridor. The spot speed/volume data were used to calibrate a traffic simulation model. The simulated travel times were fed into a KF framework to predict the future

travel times. Fernández (2010) developed a microscopic simulation model for the study of operations at public bus and light rail stops. Shalaby and Farhan (2003) developed microscopic simulation model to predict the travel time of public transport vehicles using automatic vehicle location data and automatic passenger counters; they used the KF algorithm to predict running time and dwell time of vehicles, and model verification was carried out using simulated data from VIS-SIM software. Zhang et al. (2008) developed some alighting and boarding micro-simulation model for passengers in Beijing metro stations. Toledo et al. (2010) used a mesoscopic simulation model for the evaluation of operations, planning, and control. Kachroo et al. (2001) developed travel time functions based on macroscopic models of highways. Burghout et al. (2005) developed a hybrid mesoscopic-microscopic model that applies microscopic simulation to areas of specific interest while simulating a large surrounding network in lesser detail with a mesoscopic model; the hybrid model integrates MITSIMLab, a microscopic traffic simulation model, and Mezzo, a mesoscopic traffic simulation model. Other used hybrid models include Hystra (Bourrel and Lesort 2003) and Micro-Macro link (Helbing et al. 2002), which combine dynamic macro with micro simulation.

Traditional macroscopic models are generally ineffective in evaluating strategies designed to influence travel choices and optimize system performance (Sbayti and Roden 2010). If the purpose of the study is to address congestion problems with operational and management strategies, microscopic models are much better positioned than macroscopic models to evaluate the effectiveness of the alternatives. While micro simulation models are definitely desirable, they require details about transportation facilities and flow entities at a granularity that is not typically available to transport agencies (Sbayti and Roden 2010).

In this paper, we propose an integrated approach that uses the microscopic model I-SIM-S (Hawas 2007a) to generate detailed levels of data on the facilities and flow entities that then can be used to characterize the effectiveness of bus route design in urban traffic networks. The adopted methodology entails using I-Sim-S to simulate different scenarios reflecting various network traffic intensities, posted speeds, route lengths, frequencies of bus operation, and average passenger loadings (boarding and alighting). The simulated scenarios are used to estimate the route effectiveness measures—namely, the bus route travel time and overall network travel time measures. This is then followed by developing regression models using the simulation-based effectiveness measures data. The developed models are then validated and assessed.

This paper consists of seven sections. The next section briefly reviews the features of the I-SIM-S micro simulator and presents the simulation-based experimental scenarios. The simulation results are analyzed in the following section, followed by presentation of the calibrated regression models. Validation of the regression models is highlighted, and conclusions and future research directions are highlighted.

Experimental Setup

I-SIM-S is an object-oriented program that allows for virtual detector installations at different locations and models different intersection layouts, traffic control types and timing, and link characteristics. It includes objects such as streets, lanes, detectors, vehicles, signals, and intersections, and each object is composed of data that represent the current values of object parameters and methods (or functions) that could be applied on the object (for example, add vehicle or remove vehicle for the lane object). The program is a hierarchical model in the sense that main or larger objects contain the sub (or smaller) objects; for example, each street object contains several lane objects, and each of these lane objects contains the vehicle objects (Hawas 2007a). I-SIM-S has extended capabilities to model and/or identify incidents accurately through multiple loop detectors on links of urban intersections. These capabilities were used to develop a fuzzy logic model for incident detection (Hawas 2007b). The I-Sim-S car-following model is coupled with a finite-state model that captures the lane switching behavior of the driver. The model also has the ability to capture lane switching when the lead vehicle is a bus stopping in a bus bay for boarding and alighting passengers. For more details on the I-Sim-S structure, mathematical formulation, and calibration, the reader is referred to Hawas (2007a). The simulator has been validated versus well-known microscopic simulators and in reality (Hawas and Abdul Hameed 2009).

One hypothetical network with 49 nodes, 14 origins, and 24 destinations was created for testing. The hypothetical network links were bidirectional with the same posted (free-flow) speed. All intersections were operated with pre-timed controllers. As shown in Table 1, the test scenarios were generated with four network configurations (different link lengths), four O-D flow patterns, two posted link speeds, three bus frequencies, and three levels of passenger loading.

Table 1. Simulation-based Experimental Scenarios

Link Length (m)	Source Node Volume (veh/hr)	Posted Link Speed (km/hr)	Frequency (#buses/hr/route)	# Passengers Boarding (or Alighting) per Hour at Each Bus Stop
H (300), V (300)	500 [18 scenarios]	60	2	50
				150
				300
			3	50
				150
				300
		4	50	
			150	
			300	
		80	2	50
				150
				300
	3		50	
			150	
			300	
	4	50		
		150		
		300		
1000	18 scenarios			
1500	18 scenarios			
2000	18 scenarios			
H (300), V (500)	72 scenarios			
H (500), V (300)	72 scenarios			
H (500), V (500)	72 scenarios			

H = length of horizontal link (m); V = length of vertical link (m)

The first column illustrates the values of the network link lengths used in the various scenarios. In these scenarios, horizontal and vertical link lengths are set as H (300), V (300); H (300), V (500); H (500), V (300), and H (500), V (500). The second column illustrates the source nodes demand volumes. Four different demand patterns are used generating 500, 1,000, 1,500, and 2,000 veh/hr at each of the 14 source (origin) nodes. In all tested scenarios, the source volumes were equally

distributed among all 24 possible destinations. These demand levels represent light to congested traffic conditions. The third column illustrates the link posted speed for all tested scenarios. Two speed limits of 60 and 80 km/hr were considered. The fourth column illustrates the bus frequencies used in the various simulation experiments. Three different bus frequencies (namely, 2, 3, and 4 buses/route) were tested. The last (fifth) column shows the number of passengers boarding (or alighting) per hour at each bus stop; three different levels were used (50, 150, and 300 passengers/hour). A total of 288 testing scenarios were generated (4 link lengths scenarios \times 4 source volume levels \times 2 speed scenarios \times 3 bus frequencies \times 3 boarding-alighting passenger levels).

At the beginning of the analysis period, details of the network structure, connectivity and characteristics, signal characteristics, and settings over the analysis period were provided as inputs to I-Sim-S simulator. The shortest path algorithm was “called” at the beginning of each cycle. Each vehicle (when generated) was assigned an O-D pair (in accordance with a pre-specified O-D matrix for the entire network). The generated vehicle is assigned the most recent shortest path of the O-D pair. Whenever a vehicle reaches the end of a link, its direction is determined according to its assigned path.

Table 2 shows a sample of the I-SIM-S output file for a test scenario with a total analysis period of 120 minutes, in a 49-node network. The link lengths for the tested grid network are 500 meters for the horizontal links and 300 for the vertical links. The posted speed in the network is set to 60 km/hr. The source node hourly volume is set to 1000 veh/hr for each source node. The frequency of bus is set to 3 buses per hour on each route. The passengers boarding and alighting at each bus stop are set to 50 and 50 passengers per hour, respectively.

Table 2 illustrates the output of route R1 (as denoted in the first column). The second column provides the bus number (BN) for each bus entering the network along R1. The number of entering buses depends on the frequency of buses per hour. For the tested scenario, with a frequency of three (3), and given an analysis period of 120 minutes, a total of six buses entered the network; five of which completed the trips (and as such their names were denoted by the status C letter) and one trip was incomplete by the end of the analysis period (denoted by the status I letter). The buses with incomplete trips were excluded from the calculation of average travel times. The third column presents the bus entry time into the network (ETN). This variable also depends on the frequency of bus per hour. According to the used frequency of three (3), on average a new bus will enter into the network every 20 minutes. Each operating bus on route R1 has six stops (S1 through S6) as

shown in the fourth column. The fifth column shows the stopping time at each stop (ST). The total stopping time (TST) of one bus along the route is presented in the sixth column, and it is estimated by adding the stopping times on all the bus stops. The seventh column presents the estimated total running time (TRT). The eighth column shows the total travel time (TT).

Table 2. Simulation-Based Bus Performance Output

Route	BN	ETN	BS	ST (min)	TST (min)	TRT (min)	TT (min)
R1	BUS No. 1_C	0.02	BS1 BS2 BS3 BS4 BS5 BS6	0.85 0.88 0.92 0.85 0.77 0.77	5.04	32.94	37.98
	BUS No. 2_C	19.98	BS1 BS2 BS3 BS4 BS5 BS6	0.92 0.8 0.83 0.92 0.75 0.88	5.1	33.02	38.12
	BUS No. 3_C	39.98	BS1 BS2 BS3 BS4 BS5 BS6	0.82 0.9 0.8 0.93 0.8 0.85	5.1	31.6	36.7
	BUS No. 4_C	59.98	BS1 BS2 BS3 BS4 BS5 BS6	0.83 0.92 0.82 0.82 0.83 0.93	5.15	32.28	37.43
	BUS No. 5_C	79.98	BS1 BS2 BS3 BS4 BS5 BS6	0.85 0.85 0.92 0.85 0.97 0.8	5.24	30.58	35.82
	BUS No. 6I*	99.98	BS1 BS2 BS3	0.88 0.85 0.92	2.65	15.13	17.78
Average Bus Performance					5.13	32.08	37.21

Analysis of Data

Figure 1 shows the variation of the average stopping time for different source volume level, various levels of passenger loading, and bus frequencies. It shows that increasing the passenger loading level increases the bus stopping time, whereas increasing the bus frequency decreases the average bus stopping time. The figure also illustrates that the source volume has no effect on bus stopping time.

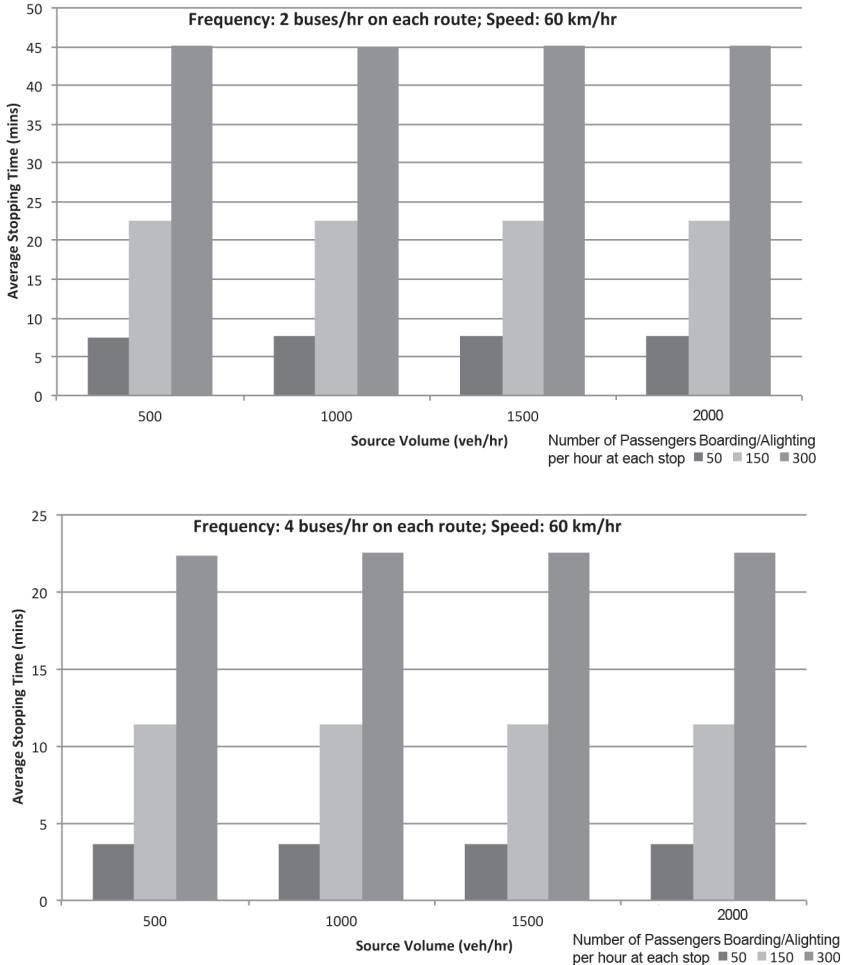


Figure 1. Average bus stopping time (mins) for network of horizontal and vertical link lengths of 300 and 300 m, link speed of 60 km/hr, under various source volumes and (a) frequency: 2 buses/hr, and (b) bus frequency: 4 buses/hr

Figure 2 shows the average bus stopping time (minutes) for various bus frequencies and passenger loading levels [for the network of horizontal and vertical link lengths of 300 and 300 meters, source volume of 500 veh/hr, link speeds of 60 km/hr]. It clearly illustrates that the average bus stopping time is affected by the bus frequencies and the passenger loading levels. The higher the bus's frequency, the lesser the bus's stopping time. On the other hand, the higher the passenger loading level, the higher the bus's stopping time.

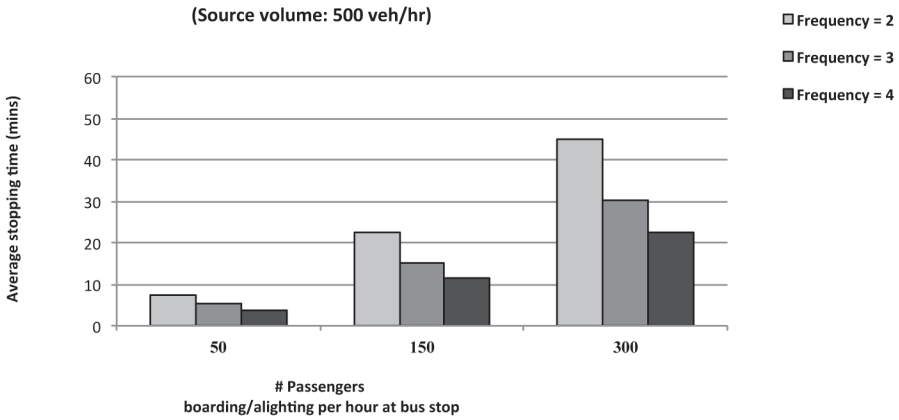
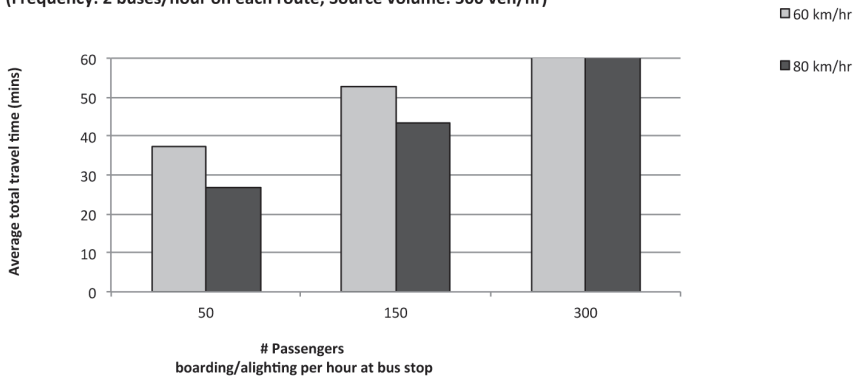


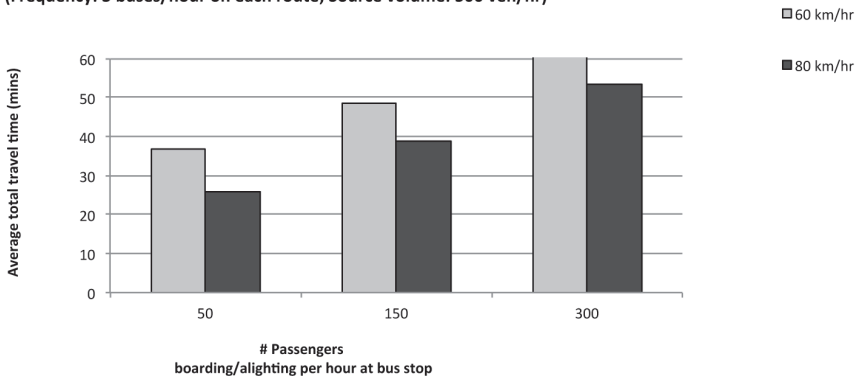
Figure 2. Average bus stopping time (mins) for various passenger loading levels and bus frequencies (scenario of network of horizontal and vertical link lengths of 300 and 300 m, source volume of 500 veh/hr, and link speed of 60 km/hr)

Figure 3 shows the average bus total travel time (minutes) for various link speeds, for the network of horizontal and vertical link lengths of 300 and 300 meters, source volume of 500 veh/hr, link speeds of 60 and 80 km/hr, and (a) frequency: 2 buses/hr, (b) bus frequency: 3 buses/hr and (c) bus frequency: 4 buses/hr.

(Frequency: 2 buses/hour on each route; Source volume: 500 veh/hr)



(Frequency: 3 buses/hour on each route; Source volume: 500 veh/hr)



(Frequency: 4 buses/hour on each route; Source volume: 500 veh/hr)

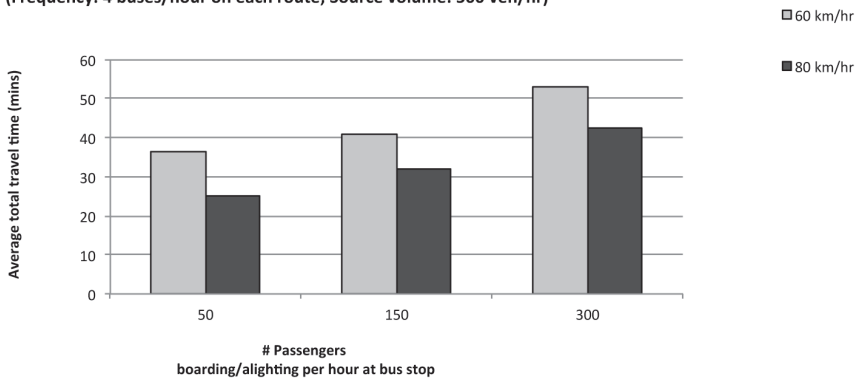


Figure 3. Average bus total travel time (mins) for network of horizontal and vertical link lengths of 300 and 300 m, source volume of 500 veh/hr, link speeds of 60 and 80 km/hr, and (a) frequency: 2 buses/hr, (b) bus frequency: 3 buses/hr, and (c) bus frequency: 4 buses/hr

Figure 4 shows the variation of the network average travel time for different source volume level, and various link speeds. Increasing source volume increases the average network travel time, whereas the increasing speed decreases the average network travel time.

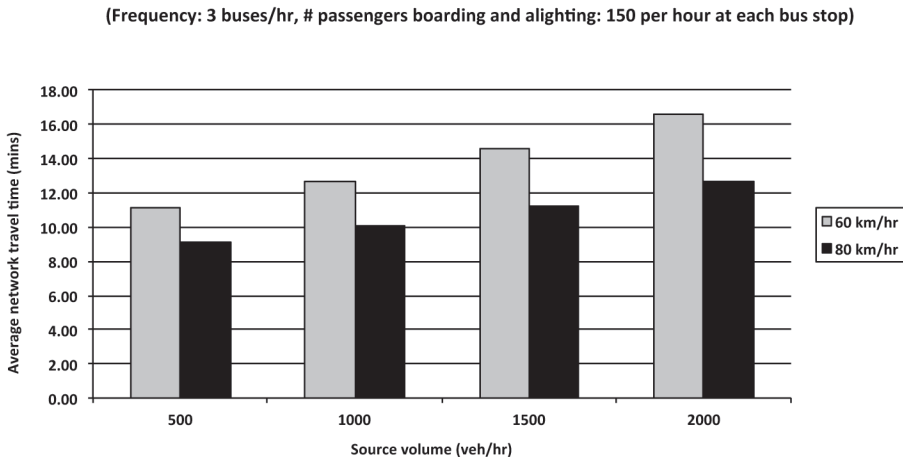


Figure 4. Average network travel time (in mins) for various source volumes and link speeds (bus frequency of 3 buses/hr and 150 passengers/hr boarding/alighting at each bus stop)

Calibration of Regression Models

The 288 cases/scenarios were simulated by I_Sim_S and used in developing linear regression models. Two linear regression models were developed to estimate the average bus route travel time and the overall network travel time.

The independent variables considered for regression analysis included total route length (m), speed (km/hr), frequency (buses/hr/route) (Chien 2003), average network intensity (veh/km), and number of passenger boarding and alighting/hr on an average route (Tetreault 2010).

The route length is the number of links per route times the length of each link. The average network intensity was calculated by dividing total network volume by total network length multiplied by number of lanes. Total network volume is number of source nodes times source volume per O-D pair. The study considered 14 source nodes, 4 different source volume levels, and 3 lanes per link. The total number of passengers boarding/alighting per hour on an average route was calculated by

multiplying number of passengers per stop by number of bus stops on the route, summed over all routes and then divided by number of routes. Four different levels of passenger loadings, four routes and six bus stops per route were considered for the test network. Table 3 shows the independent and dependent variables for the two regression models.

Table 3. Independent and Dependent Variables for Regression Analysis

Model	Independent Variables	Dependent Variables
I (ABTT)	Route length (m) (X1)	Average bus travel time (Y)
	Average network intensity (veh/lane.km) (X2)	
	Speed (km/hr) (X3)	
	Frequency (buses/hr/route) (X4)	
	# passengers boarding/alighting per hr on average route (X5)	
II (NATT)	Total network length (m) (X6)	Network average travel time (Z)
	Average network intensity (veh/ lane.km) (X7)	
	Speed (km/hr) (X8)	

Several iterations were done for calibrating Network Average Travel Time (NATT) model by excluding the insignificant variables. First, the linear regression model was run using five variables. It was concluded to include only the three variables shown in Table 3, based on the goodness-of-fit test (t-stat greater than 2). Two variables (percentage of buses in the network and passengers loading level) were excluded. The coefficients and the t-stat values for the final two models are summarized in Table 4.

The R-square values for both Average Bus Travel Time (ABTT model) and Network Average Travel Time (NATT model) are quite reasonable. The standard error for ABTT model is higher than that of the NATT model. The goodness-of-fit tests indicate that all the independent variables in both models are statistically significant (with t-stat greater than 2 and the p-values lesser than 0.05). The intercept values (a0 and b0) for the ABTT and NATT models are relatively high with good statistical significance values. This indicates the need to include more independent variables in both models.

Table 4. Coefficient Values and Goodness-of-Fit Measures (ATT and NATT)

ABTT Model				NATT Model			
Coeff		t-stat	p	Coeff		t-stat	p
a0	46.727	15.08	4.21E-38	b0	8.77	15.304	5.719E-39
a1	0.002	7.827	1.01E-13				
a2	0.007	2.392	0.017392	b6	0.003	21.583	7.996E-62
a3	-0.149	-5.06	7.49E-07	b7	0.012	19.742	3.295E-55
a4	-6.39	-17.8	4.7E-48	b8	-0.07	-12.53	5.959E-29
a5	0.020	41.26	1.7E-121				
Standard Error							
			4.97857				0.96365
R-square							
			0.88193				0.73362

The regression analysis predicted the signs of all coefficients logically. For example, if the length of the route (X1) is increased, the average bus travel time (Y) will increase (the a1 coefficient is positive). Similarly, if the total network length (X6) is increased, the network average travel time (Z) will increase (the b6 coefficient is positive). On the other hand, increasing the route’s average speed (X3) will result in decreasing the average bus travel time (Y) (the a3 coefficient is negative). The calibrated regression ABTT and NATT models are given below:

$$Y = 46.727 + 0.002X_1 - 0.007X_2 - 0.149X_3 - 6.398X_4 + 0.020X_5 \tag{1}$$

$$Z = 8.771 + 0.0002X_6 + 0.012X_7 - 0.071X_8 \tag{2}$$

All coefficient values are rounded up to three decimal places.

Residuals Analyses for Calibrated Models

The residuals were calculated as the difference between observed (extracted from the I-SIM-S simulator) and estimated values (from the regression models). Figure 5 shows the percentages of deviations between the simulation-based observed and the estimated values using the ABTT model. The negative and positive deviations represent 54 percent and 46 percent of the 288 scenarios, respectively. The average deviation is 0.66 percent, with the maximum and minimum deviations to be around 27 percent and -43 percent, respectively. A total of 252 scenarios (about 88% of the scenarios) exhibited values of deviation equal to or less than 15 percent. Only 36 scenarios (12%) have deviation values more than 15 percent. This means

that the ABTT regression model can be used for prediction, with an expected prediction error of 15 percent or less in 88 percent of the cases that the model is applied for. This is quite acceptable accuracy level for the purpose of planning/design of routes. The remaining 12 percent of the cases resulted in error values ranging between 15–30 percent. This suggested the probable need to include more independent variables.

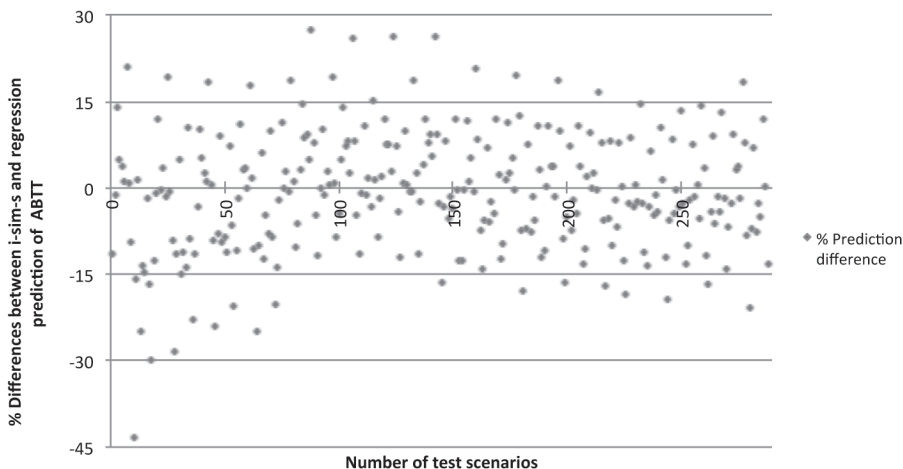


Figure 5. Percentages of deviations between predicted (using ABTT model, Eq. 1) and observed (I-SIM-S simulated) values of average bus travel time (for 288 calibration scenarios)

Figure 6 shows the percentages of deviations between observed and estimated values using the NATT model. The negative and positive deviations (errors) represent 55 percent and 45 percent of the 288 scenarios, respectively. The average deviation is -0.49 percent, with the maximum and minimum deviations to be around 13 percent and -13.9 percent, respectively. Only 55 scenarios (about 19% of the 288 scenarios) resulted in deviations more than 10 percent. This means that the NATT regression model can be used for prediction, with an expected prediction error of 10 percent, or less in 81 percent of the cases to which the model is applied. This is quite acceptable accuracy level for the purpose of planning/design of routes. It should be noted that the literature of the travel time prediction models indicates relatively larger error values. For instance, Shalaby and Farhan (2003) reported mean of error values ranging from 3–8 percent average error values (with a max error value of 23%) on individual links using the KF approach and a mean of error values ranging from 8–22 percent average error values (with a maximum error

value of 46%) on individual links on individual links using the regression modeling approach. The average deviation of the presented NATT model (-0.49%) with a maximum error of 13.9 percent is far below the reported models in Shalaby and Farhan (2003).

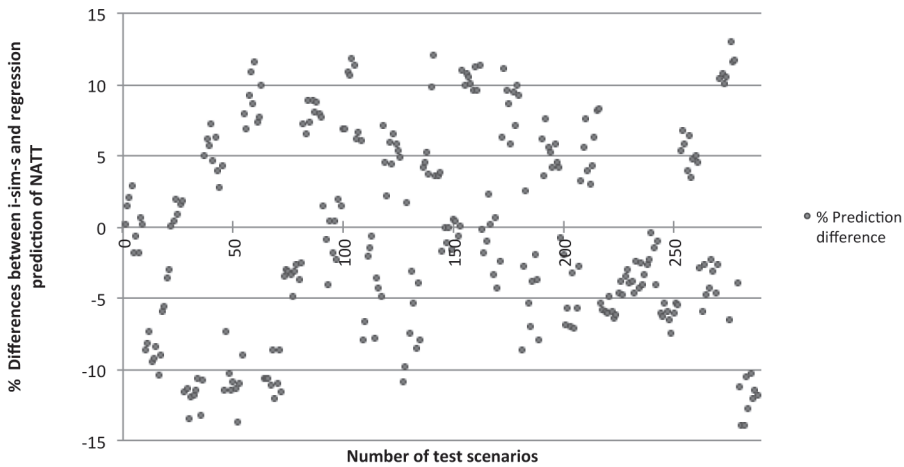


Figure 6. Percentage of deviation between predicted (using NATT model, Eq. 2) and observed (I-SIM-S simulated) values of network average travel time (for 288 calibration scenarios)

Validation of ABTT and NATT Models

The simulation model I-SIM-S was further used to generate 64 validation scenarios. These validation scenarios were generated with independent variable values different from those used in developing the 288 calibration scenarios (previously described in Table 1). Table 5 shows the values of the independent variables used in generating the validation scenarios. The rationale is to check the accuracy of the developed (ABTT and NATT) regression models in predicting the average bus travel time and the network average travel time in cases beyond those used in the calibration of these models.

Table 5. Values of Variables Considered for Validation Scenarios

Link Length (m)	Source Node Volume (veh/hr)	Posted Link Speed (km/hr)	Frequency (#buses/hr/route)	# Passengers Boarding (or Alighting) per Hour at Each Bus Stop
H (400 m) V (400 m)	750 veh/hr (16 scenarios)	70	1	200
			2	200
				400
			3	400
			4	200
		400		
		5	200	
			400	
		90	1	200
				200
	2		400	
			400	
	3		200	
			400	
	4	200		
		400		
5	200			
	400			
	1750 veh/hr	16 scenarios		
H (600m) V (600m)	32 scenarios			

Similar to the calibration scenarios, the validation experimental setup accounts for the variations in link lengths (column 1), O-D demand pattern represented by the source node volume (column 2), posted link speeds (column 3), bus frequency (column 4), and number of boarding and alighting passengers (column 5).

Two link lengths, two patterns of O-D flows, two speed limits, five bus frequencies, and two passenger loading levels were used in developing the validation scenarios, as shown in Table 5. It is important to note that these values are different than those used in calibration (previously shown in Table 1).

The above 64 validation scenarios were run by the simulator (I-SIM-S) to obtain the observed values of the average bus travel time and the network average travel time.

Furthermore, the developed regression models (ABTT and the NATT) were used to estimate the corresponding Y and Z values. The observed and estimated values were then used to carry on detailed residual analyses.

Residuals Analyses for Validation Scenarios

Figure 7 shows the percentages of deviations between the simulation-based observed and the estimated values using the ABTT model for all validation scenarios. The negative and positive deviations represent 52 percent and 48 percent of the 64 validation scenarios, respectively. The average deviation is 0.12 percent, with the maximum and minimum deviations to be around 33 percent and -24 percent, respectively. A total of 18 scenarios out of the 64 (about 28%) exhibited deviation values of more than 15 percent, as compared to 12 percent of the calibration scenarios with similar deviation values (as shown in Figure 5).

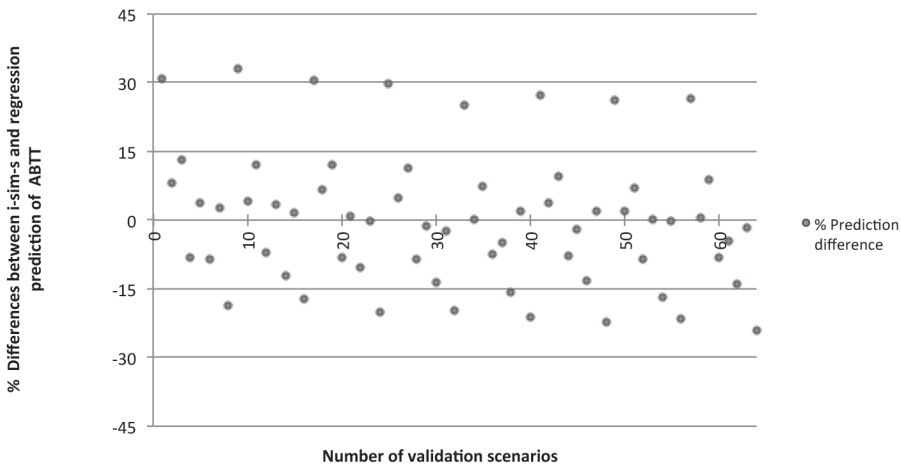


Figure 7. Percentages of deviations between predicted (using ABTT model, Eq. 1) and observed (I-SIM-S simulated) values of average bus travel time (for 64 calibration scenarios)

Figure 8 shows the percentages of deviations between the simulation-based observed and the estimated values using the NATT model for all validation scenarios. The negative and positive deviations represent 92.2 percent and 7.8 percent of the 64 validation scenarios, respectively. The average deviation is 12 percent, with the minimum deviations to be around -27 percent. A total of 44 scenarios out of the 64 (about 68%) exhibited deviation values of 15 percent or less.

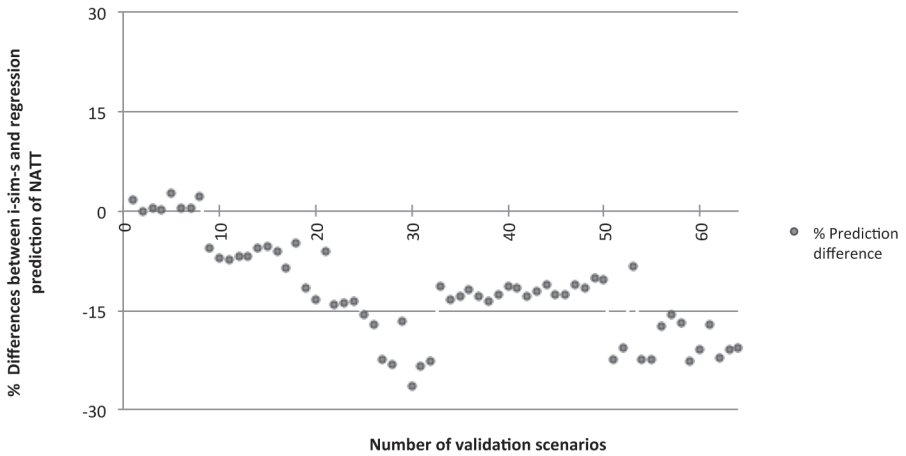


Figure 8. Percentage of deviation between predicted (using the NATT model, Eq. 2) and observed (I-SIM-S simulated) values of network average travel time (for 64 calibration scenarios)

Conclusions and Future Work

This paper presents an approach to characterize bus route design using micro-simulation and to predict the effectiveness of bus route designs using some general network measures and traffic indicators. The used indicators are average network traffic intensity, posted speeds, route length, frequency of bus operation, and average passenger loadings (boarding and alighting).

The developed models differ from those reported in the literature to predict travel times in several aspects. It explicitly captures bus operation rather than a general travel time prediction model for all vehicles in the network. Additionally, the developed regression models are simple, with few variables that can be developed/validated using field data if such data become readily available. The use of the detailed microscopic simulator for data generation and study of various scenarios enables developing a generalized model that can be applied to wide range of bus operation characteristics, instead of using limited field data. It also enables more accurate modeling and traffic measures and, therefore, better model validity. This is evident in the resulting acceptable validation error.

Regression models were calibrated and validated to predict both route and overall network travel times. The presented validation process is simulation-based. More work is still to be conducted for real-life validation. To validate the devised cali-

bration models in reality, data representing the independent variables of the two regression models need to be gathered. For instance, for the ABTT model, two particular independent variables (X2 and X5) will have to be measured through field surveys at bus stops and link traffic counts. Surveys should be conducted on a sample of operating bus routes during peak hours to count the number of passengers boarding/alighting at each bus stop. This will be used to estimate the average route ridership per hour (X2). For a network of N links, a sample of n links need to be randomly selected. The n value should be estimated using a sample size formula for a population of N links, according to a specific significance level and percent of error. On the randomly selected n links traffic counts should be done during the same peak hours during which the passenger surveys are instrumented. These link traffic counts can be used then to estimate the value of the average network intensity (X5). The other independent variables (X1, X3 and X4) can be easily measured also in reality or obtained directly from the operating bus agency itself. The observed independent variables can then be used for validating the ABTT model. Comparing actual bus trip times on surveyed routes with those estimated from the ABTT model using the field estimated independent variables (X1 through X5) can be used for assessing model accuracy. The NATT model can also be validated in reality using a similar approach. The X7 variable of the NATT model is similar (equivalent) to the variable X5 of the ABTT model. Currently, case study data are being collected for the operating bus routes of Al Ain and Abu Dhabi in the UAE. This will help determine whether the operating bus frequencies and route design are efficient.

The devised models can be used for route planning purposes to determine the best operational route characteristics such as bus frequency. Bus agencies can find the application of the two devised model quite beneficial to assess the efficiency of their operating routes. Also, the Y and the Z values of the ABTT and NATT models can be used further to develop some simplified mode split models to capture the effect of route design and bus frequencies on bus ridership. This could be an interesting application for these models that will allow agencies to obtain estimates of how changes in route design or operation will likely affect the market segment of bus users. This will be investigated in further research.

Acknowledgments

This work was funded by the Roadway, Transportation and Traffic Safety Research Center (RTTSRC) at UAE University. The author would like to acknowledge Eng. Nandita Basu, a former research assistant at RTTSRC, for her valuable assistance.

References

- Anderson, J., and Bell, M. 1998. Travel time estimation in urban road networks. *Proceedings, IEEE Conference*: 924-929.
- Bourrel, E., and Lesort, J. B. 2003. Mixing micro and macro representations of traffic flow: A hybrid model based on the LWR theory. *Transportation Research Record* 1852: 193-200.
- Burghout, W., H. N. Koutsopoulos, and I. Andreasson. 2005. Hybrid mesoscopic-microscopic traffic simulation. *Transportation Research Record* 1934: 218-225.
- Chakroborty, P., and S. Kikuchi. 2004. Using bus travel time data to estimate travel times on urban corridors. *Transportation Research Record* 1870: 18-25.
- Chen, M., and Chien, S. I. J. 2001. Dynamic freeway travel time prediction using probe vehicle data: link based versus path based. *Transportation Research Record* 1768: 157-161.
- Chien, S. I. J., and C. M. Kuchipudi. 2003. Dynamic travel time prediction with real-time and historical Data. *Journal of Transportation Engineering* 129(6): 608-616.
- El-Geneidy, A. M., J. Horning, and K. J. Krizek. 2010. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation*, www.interscience.wiley.com, DOI: 10.1002/atr.134.
- Fernández, R. 2010. Modeling public transport stops by microscopic simulation. *Transportation Research Part C* 18: 856-868.
- Fishburn, P. T., and K. M. Taaffe. 1994. Transportation applications. *Proceedings, 1994 Winter Simulation Conference*, J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, eds.: 22-25.
- Hawas, Y. E. 2007a. A microscopic simulation model for incident modeling in urban networks. *Transportation Planning and Technology* 30(2): 289-309.

- Hawas, Y. E. 2007b. A fuzzy-based system for incident detection in urban street networks. *Transportation Research Part C* 15: 69-95.
- Hawas, Y. E., and M. A. Hameed. 2009. A multi-stage procedure for validating microscopic traffic simulation models: A multistage procedure for validating microscopic traffic simulation models. *Transportation Planning and Technology* 32(1): 71-91.
- Helbing, D., A. Hennecke, V. Shvetsov, and M. Treiber. 2002. Micro- and macro-simulation of freeway traffic. *Mathematical and Computer Modelling* 35(5-6): 517-547.
- Huisken, G., and E. C. Van Berkum. 2003. A comparative analysis of short-range travel time prediction methods. *Proceedings, 82nd Annual TRB Meeting, Washington, DC.*
- Kachroo, P., K. Ozbay, and A. G. Hobeika. 2001. Real-time travel time estimation using macroscopic traffic flow models. *IEEE Intelligent Transportation Systems Conference Proceedings*: 132-137.
- Karl, C. A., S. Charles, and R. Trayford. 1999. Delivery of Real-Time and Predictive Travel Time Information: Experiences from a Melbourne Trial. *Proceedings, 6th World Congress on Intelligent Transport Systems, Toronto.*
- Kuchipudi, C. M., and S. I. J. Chien. 2003. Development of a hybrid model for dynamic travel time prediction. *Transportation Research Record* 1855: 22-31.
- Kwon, J., B. Coifman, and P. Bickel. 2000. Day-to-day travel time trends and travel time prediction from loop detector data. *Transportation Research Record* 1717: 120-129.
- Li, Y., and M. McDonald. 2002. Link travel time estimation using single GPS-equipped probe vehicle. *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems, Singapore*: 932-937.
- Lin, H. E., R. Zito, and M. A. P. Taylor. 2005. A review of travel-time prediction in transport and Logistics. *Proceedings of the Eastern Asia Society for Transportation Studies* 5: 1433-1448.
- Lum, K. M., Fan, H. S. L., Lam, S. H. and Olszewski, P. 1998. Speed-flow modeling of arterial roads in Singapore. *Transportation Engineering* 124(6): 213-222.

- Mark, C. D., A. W. Sadek, and D. Rizzo, D. 2004. Predicting experienced travel times with neural networks: A PARAMICS simulation study. *Proceedings of the 7th International IEEE conference on Intelligent Transportation Systems*: 906-911.
- Nanthawichit, C., T. Nakatsuji, and H. Suzuki. 2003. Application of probe vehicle data for real-time traffic state estimation and short term travel time prediction on a freeway. *Transportation Research Record* 1855: 49-59.
- Nash, A. 2006. Design of effective public transportation systems. 6th Swiss Transport Research Conference, Monte Verita / Ascona, Switzerland, 2006: 1-27. http://www.strc.ch/conferences/2006/Nash_STRC_2006.pdf.
- Ni, D. 2001. Panorama of transportation simulation. Georgia Institute of Technology. <http://gte710q.tripod.com/Research/panorama.pdf>.
- Okutani, I., and Y. J. Stephanedes. 1984. Dynamic prediction of traffic volume through Kalman Filtering theory. *Transportation Research Part B* 18(1): 1-11.
- Sbayti, H., and D. Roden. (AECOM). 2010. Best practices in the use of micro simulation models. NCHRP 08-36/Task 90, Research for the AASHTO Standing Committee on Planning. http://statewideplanning.org/wp-content/uploads/259_NCHRP-08-36-90.pdf.
- Shalaby, A., and A. Farhan. 2003. Bus travel time prediction model for dynamic operations control and passenger information systems. Paper presented at the Transportation Research Board 82nd Annual Meeting, Washington, DC.
- Tetreault, P. R., and A. M. El-Geneidy. 2010. Estimating bus run times for new limited-stop service using archived AVL and APC data. *Transportation Research Part A* 44: 390-402.
- Toledo, T., O. Cats, W. Burghout, and H. N. Koutsopoulos. 2010. Mesoscopic simulation for transit operations. *Transportation Research Part C* 18(6): 896-908.
- Van Lint, J. W. C. 2003. Confidence intervals for real-time freeway travel time prediction. *Proceedings of the IEEE Conference on Intelligent Transportation Systems, Shanghai, China* 2: 1453-1458.
- Van Lint, J. W. C., S. P. Hoogendoorn, and H. J. Van Zuylen. 2002. Freeway travel time prediction with state-space neural networks: Modeling state-space dynamics with recurrent neural networks. *Transportation Research Record* 1811: 30-39.

- Van-Grol, H. J. M., M. Danech-pajouh, S. Manfredi, and J. Whittaker. 1999. DAC-CORD: On-line travel time prediction. *World Conference on Transport Research Society (WCTRS) 2*: 455-467.
- Waller, S. T., Y. C. Chiu, N. Ruiz-Juri, A. Unnikrishnan, and B. Bustillos. 2007. Short-term travel-time prediction on freeways in conjunction with detector coverage analysis. FHWA/TX-08/0-5141-1. FHWA, U.S. Department of Transportation.
- Wu, C. F. 2001. The study of vehicle travel time estimation using GPS. Master's thesis, Department of Transportation Technology & Management, National Chiao Tung University, Taipei.
- Zhang, Q., B. Han, and D. Li. 2008. Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations. *Transportation Research Part C* 16(5): 635-649.
- Zhang, X., and J. A. Rice. 2003. Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies* 11: 2-4.

About the Author

YASER. E. HAWAS (y.hawas@uaeu.ac.ae) obtained his Ph.D. from the University of Texas at Austin in 1996 and worked there as a postdoctoral fellow before joining the United Arab Emirates University in 1998. He currently serves as a Professor in the Department of Civil and Environmental Engineering at the UAE University, where he also serves as the Director of the Roadway, Transportation and Traffic Safety Research Center (RTTSRC). His primary areas of expertise includes the development of ITS tools for modeling and analysis, advanced control systems, traffic management and operation, development of AI algorithms for real-time control, transit operation modeling and policies, behavioral modeling, advanced computational tools for network optimization, planning and design, and traffic safety analysis.