WCTR 2013 RIO

SELECTED PROCEEDINGS

A COMPREHENSIVE ANALYSIS FOR BUS DWELL TIME PREDICTION

NAM H. VU, ASSOCIATE PROFESSOR, NATIONAL UNIVERSITY OF CIVIL ENGINEERING, HANOI, VIETNAM
ATA M. KHAN, PROFESSOR, CARLETON UNIVERSITY, OTTAWA, CANADA

This is an abridged version of the paper presented at the conference. The full version is being submitted elsewhere. Details on the full paper can be obtained from the author.

ISBN: 978-85-285-0232-9

13th World Conference on Transport Research

www.wctr2013rio.com

15-18 JULY 2013
Rio de Janeiro, Brazil

A COMPREHENSIVE ANALYSIS FOR BUS DWELL TIME PREDICTION

Nam H. Vu, Associate Professor, National University of Civil Engineering, Hanoi, Vietnam

Ata M. Khan, Professor, Carleton University, Ottawa, Canada

ABSTRACT

High quality predictive bus arrival time is recognized to be important for attracting and keeping choice riders. Although dwell time accounts for a smaller portion of total delay compared to running time, this cannot be neglected, especially when the buses serve high density population areas. In terms of real-time control, the predictive dwell time at stops can help bus dispatchers to have suitable proactive solutions once a crowded bus stop is detected in advance or the knowledge that bunching is likely to occur. This paper presents new dwell time prediction methods. The relationship between dwell time and a number of explanatory variables representing passenger characteristics, bus type, location of bus stop, and seasonal time period was modelled. Data were retrieved from automatic passenger counter (APC) and automatic vehicle location (AVL) systems of OC Transpo (Ottawa, Canada). The results show that the proposed methods enhance the predictive capability and can be applied for different types of buses equipped with AVL and APC systems.

Keywords: Bus transit, dwell time, predictive model, automatic passenger counter, automatic vehicle location system, bus bunching, traveller information.

INTRODUCTION

From the transit passenger point of view, one would argue that the predictions of dwell times, departure times and the number of passengers are not what a passenger wants to know. This standpoint is true; however, these predictions contribute to the accuracy of announced arrivals that passengers are interested in. Although dwell time accounts for a smaller portion of total delay (12-26%) compared to that of running time (48-75%), this cannot be neglected, especially when the buses serve high density population areas (Levinson, 1983).

From bus transit service provider's point of view, data on passenger activities at each stop, and the number of on-board passengers are extremely important for analyzing ridership and the relationship between passenger loading, running times and on-time performance (Pile et al., 1998). For real-time control, the predicted number of passengers boarding and alighting at a stop can help bus dispatchers to develop suitable pro-active solutions once a crowded bus stop is detected in advance or the knowledge that bunching is likely to occur.

Past research has explored the subject of bus delay prediction (Abdelfattah and Khan 1997, Cathy and Daily 2003, Bertini and El-Geneidy 2004, Vu and Khan 2010). For the purpose of developing bus arrival time prediction methods and algorithms, detailed modelling work has been reported that was mainly devoted to bus running time (Daily et al. 2001, Shalaby and Farhan 2004, Son et al. 2004, Vu and Khan 2010). In relative terms, dwell time has not received sufficient research attention.

Dwell time is influenced by many factors (Dueker et al. 2004). These can be classified into two groups. The first group relates to passengers and their activities such as the number of passengers boarding and alighting, and types of passengers (e.g., age, physical health, gender). The second group relates to bus service activities such as type of fare collection, number of doors, seating-capacity, type of bus (e.g., rigid-body bus, articulated bus, low-floor bus or high-deck bus), and service frequency.

Before the invention of the APC system, dwell time analysis was constrained due to time consuming and highly labour-intensive manual counting method. At present, the APC system is creating great potential for studies on dwell time by providing much data not only in terms of quantity but also in quality. This also enables the applications of real-time prediction method for predicting bus passenger activities and analyzing delays (Furth 2003, Horbury 1999, Mishalani et al 2000, Rajbhandari et al 2003, Schiavon 1999). The intent is to use technology and methodology to improve traveller information (TCRP 1998a, 1998b, 2000).

This paper presents new dwell time prediction methods. It is one part of an overall system of models developed for predicting bus real-time arrivals by using data collected by the automatic passenger counter (APC) and automatic vehicle location (AVL) system. For methodological advances in bus running time prediction, interested readers can refer to Vu and Khan (2010).

Modelling Bus Dwell Time

In order to build the best possible tool for predicting dwell time, two promising methods were defined. These are shown in Figures 1 and 2. The Dwell Time Model works as a real-time predictor for boarding and alighting passengers at a stop. Two of its sub-models, namely Real-time Boarding Passenger Prediction (RBSM) and Real-time Alighting Passenger Prediction (RASM) were developed separately. Their predictions are used as inputs for two other sub-models titled Regression Sub-Models (RESM) and Busiest door Sub-Model (BDSM).

By combining the sub-models two methods (A&B) were defined for predicting dwell time. Method A, shown in Figure 1, is the combination of BPSM, RASMAPSM, and RESM.

Method B, illustrated in Figure 2, is based on RBSM, APSM, and BDSM. Since RBSM and RASM are not the focus of this paper, their details are not included here. The details of the Regression Sub-Model (RESM) and the Busiest Door Sub-Model (BDSM) are provided due to our interest in the number of boarding and alighting passengers for each stop.

Variable selection and preparation

The variables selected for use in the regression models are shown in Table 1. Because data on some variables were not collected by the APC-AVL systems but could be estimated from other variables, it became necessary to prepare data before carrying out regression analyses. The variables were transformed and calculated (Equations 1 to 6). Explanations of variables are presented in Table 1.

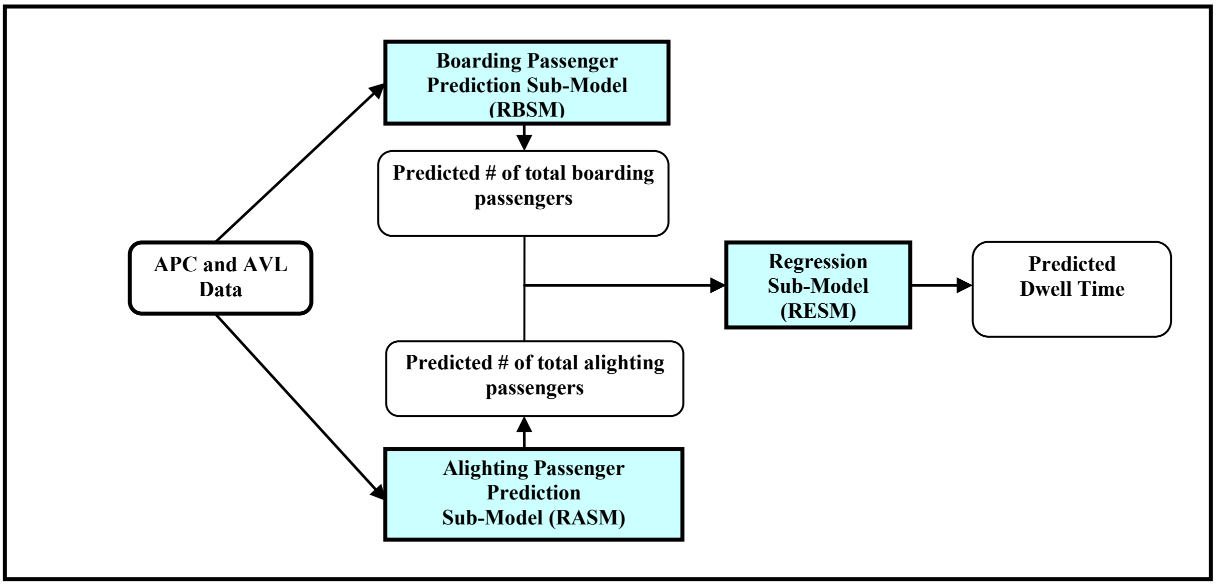

Figure 1- Dwell time Prediction Model: Method A

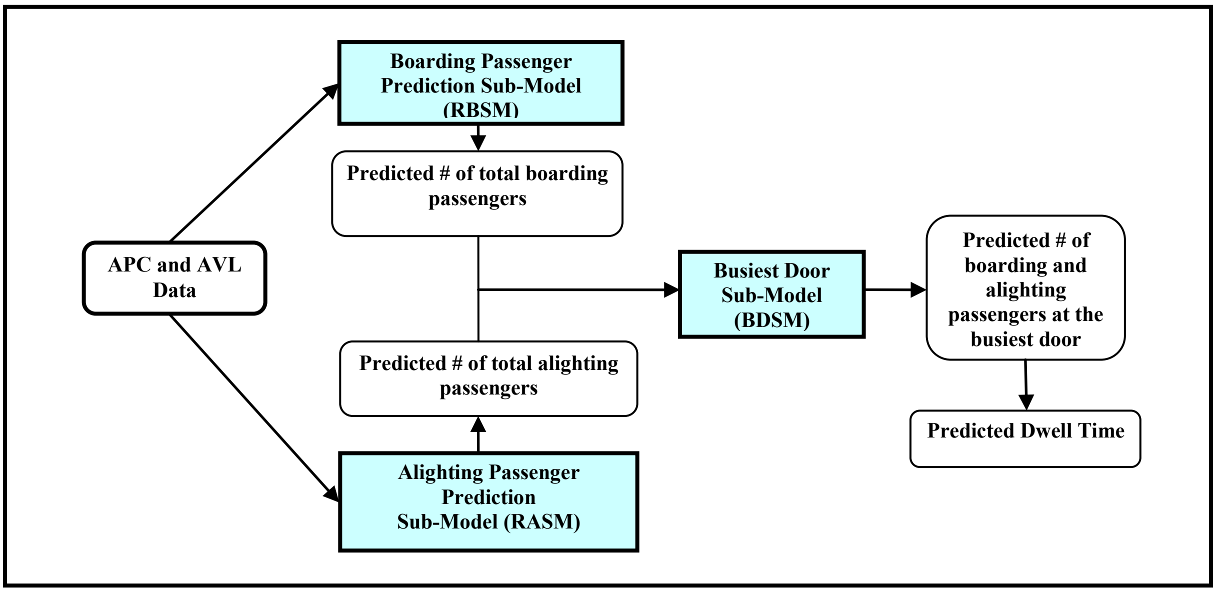

Figure 2 - Dwell time Prediction Model: Method B

Table 1- Variable selection for two models

Variable	Type of Variable	Code	Description of Variable	Reason for Selection
Boarding passengers	Numeric	*TOTAL_ONS*	The total number of passengers boarding at a stop	It has a major influence on dwell time.
Alighting passengers	Numeric	*TOTAL_OFFS*	The total number of passengers alighting at a stop	It has an influence on dwell time
Boarding and alighting	Numeric	*TOTAL_PASS*	Total passengers getting on and off at a bus stop	It has a major influence on dwell time.
On- board passengers prior stop	Numeric	*LOAD_ARR*	Number of passengers on the bus before it arrives or passes the stop.	This affects the circulation in the bus. Therefore it will influence dwell time
Loading factor	Numeric	*LF*	Ratio between the number of on-board passengers prior to stop and the capacity of the bus	A value of LF close to 1 means that the bus is likely full. Hence, this may cause a long dwell time.
Alighting passengers at front door	Numeric	*OFF_1*	Number of alighting passengers using front door	If passengers alight at front door, it will increase dwell time.
Punctuality	Numeric	*PUNT*	Measured by lateness and earliness of the bus.	A long lateness can cause a crowded downstream stop resulting in a long dwell time.
Bus Type	Dummy	*BUS_TYPE*	Articulated bus or rigid bus (all are low-floor buses). 1=Articulated Bus; 0 = Rigid bus	Dwell time depends on bus type
Time	Dummy	*TIME*	Time of day (Morning, noon, after noon). 1= Morning; 2= Noon; 3= afternoon	Dwell time depends on time of day
Season	Dummy	*SEASON*	0= Winter; 1 = Spring; 2=Summer;3= Fall	Dwell time may be different for summer and winter
Number of Doors	Numeric	*DOORS*	Self explanatory	Increasing number of doors reduces dwell time
Stop location	Dummy	*STOP_LOCA*	1= Stop located at CBD 0= Otherwise	The location of bus stop may affect dwell time

$$LOAD\_ARR: = (LOAD\_DEP) + (TOTAL\_OFFS) - (TOTAL\_ONS) \quad (1)$$

$$TOTAL\_ONS = (ON\_1) + (ON\_2) + (ON\_3) \quad (2)$$

$$TOTAL\_OFFS = (OFF\_1) + (OFF\_2) + (OFF\_3) \quad (3)$$

$$PUNT = (ACT\_TIME) - (EXPEC\_TIME) \tag{4}$$

$$TOTAL\_PASS = (TOTAL\_ONS) + (TOTAL\_OFFS) \tag{5}$$

$$LF = (LOAD\_ARR) / BUS\_CAPA \tag{6}$$

As noted earlier, variables are defined in Table 1, except the following:
BUS_ CAPA is the number of designated seats of transit bus; *BUS_CAPA* = 40 for rigid bus; = 65 for articulated bus.

Data were collected from APC units installed in buses in Ottawa (Canada). Two bus lines were selected, namely bus line 95 which features mainly articulated vehicles and bus-line 1 where most buses are rigid body type. Before going to analyses, dwell times recorded by APC units were processed to delete unreliable records. One of these records is the zero value, meaning that the bus does not stop at bus stops. Also, records of dwell time over 180 seconds were deleted without any concern because such time durations are usually the layovers.

After removing these records, the remaining data set was examined further to find additional records that should be deleted to upgrade prediction. The extreme cases and outliers were identified and dwell times lower than 57 seconds were kept for both routes. The sample size of the database for regression analysis reduced from 9983 cases to 8685 cases after outlier and zero value deletions.

The RESM Model

In this module, a number of regression types were studied in order to find the best predictor.

Type A-1: Simple Linear Regression

The general regression equation form is shown below.

$$\begin{aligned} dwell &= \beta_0 + \beta_1.(TOTAL\_ONS) + \beta_2.(TOTAL\_OFF) + \beta_3.(TOTAL\_PASS) \\ &+ \beta_4.(LOAD\_ARR) + \beta_5.(LF) + +\beta_6.(PUNT) + \beta_7.(BUS\_TYPE) + \beta_8.(TIME) \\ &+ \beta_9.(SEASON) + \beta_{10}.(DOORS) + \beta_{11}.(STOP\_LOCA) \end{aligned} \tag{7}$$

The backward stepwise regression method was selected in order to track the affect of each variable on dwell time. The suggested values of α = 0.15-0.2 (Menard, 2001, p. 64) was selected to keep balance in between taking unimportant variables and the risk of deletion of important variables. Eighty percent of the total cases were drawn randomly from the APC data.

After 3 runs, a stepwise regression method returned the best regression as shown in Equation 8. Two out of 9 variables were removed from the model (i.e. *TIME* and *SEASON*). All parameters are statistically significant. With *R-square* of 0.581, the model is satisfactory.

$$\underset{R^2=0.581}{dwell} = \underset{t=14.583}{16.175} + \underset{t=9.812}{2.126}(STOP\_LOCATE) + \underset{t=27.163}{0.438}(TOTAL\_OFFS)$$
$$+ \underset{t=75.856}{1.222}(TOTAL\_ONS) - \underset{t=-10.511}{4.397}(DOORS) + \underset{t=2.605}{0.074}(LOAD\_ARR) \quad (8)$$
$$- \underset{t=-3.085}{5.064}(LF) + \underset{t=5.116}{0.003}(PUNT)$$

Type A-2: Non- Linear Regression

Besides the multiple linear regression developed above, a series of non-linear regression types of functions were also examined (Table 2). As can be seen in the table, out of the equation types from A-2.1 to A-2.6, type A-2.3 has the highest *R-square*, meaning that it has the best predictive ability as compared with others. As a result, the use of type A-2.3 is used in this research to predict dwell time.

Table 2 - Proposed Non-linear Regression Models

Type	The best regression equations	R-square
A-2.1	$dwell = \underset{t=21.024}{13.177} + \underset{t=51.635}{1.928}(TOTAL\_ONS) - \underset{t=-20.636}{0.030}.(TOTAL\_ONS)^2$ $- \underset{t=28.834}{0.438}(TOTAL\_OFF) - \underset{t=-2.414}{0.830}(LF) - \underset{t=-17.342}{3.922}(DOORS)$ $+ \underset{t=9.684}{1.957}(STOP\_LOCA) + \underset{t=5.456}{0.003}(PUNT)$	0.610
A-2.2	$dwell = \underset{t=19.804}{12.843} + \underset{t=33.657}{2.351}(TOTAL\_ON) - \underset{t=-11.873}{0.071}(TOTAL\_ON)^2$ $+ \underset{t=7.050}{0.001}(TOTAL\_ONS)^3 + \underset{t=28.173}{0.429}(TOTAL\_OFF) + \underset{t=5.490}{0.03}(PUNT)$ $- \underset{-2.771}{0.954}(LF) - \underset{1.821}{0.153}(SEASON) - \underset{t=-17.953}{4.073}(DOORS) + \underset{t=8.987}{1.808}(STOP\_LOCA)$	0.618
A-2.3	$dwell = \underset{t=13.921}{14.434} + \underset{t=54.071}{2.014}(TOTAL\_ON) - \underset{t=-17.944}{0.026}.(TOTAL\_ON)^2$ $+ \underset{t=15.364}{0.516}(TOTAL\_OFFS) + \underset{t=3.387}{0.004}(TOTAL\_OFFS)^2$ $- \underset{t=-11.548}{0.022}(TOTAL\_ONS)(TOTAL\_OFFS)$ $+ \underset{2.982}{0.078}.(LOAD\_ARR) - \underset{-3.349}{5.047}(LF) + \underset{t=5.492}{0.003}(PUNT)$ $- \underset{t=-12.295}{4.748}(DOORS) + \underset{t=10.384}{2.087}(STOP\_LOCA) + \underset{t=2.419}{0.196}(SEASON)$	0.634
A-2.4	$dwell = 31.08(1 - 0.795\exp(-0.346.(10^{-1}.TOTAL\_PASS)^2))$	0.502
A-2.5	$dwell = 1184802 - 91.856\exp(-2.557(10^{-1}TOTAL\_PASS)^{-0.003})$	0.477
A-2.6	$dwell = (-146.259 + 470.907(TOTAL\_PASS)^{0.561})/(97.767$ $+ (10^{-1}TOTAL\_PASS)^{0.561})$	0.528

Although type A-2.6 has less predictive ability than type A-2.3, it is a compact model, requiring only the total of passenger activities recorded at bus doors or predicted. Therefore, it should be used when there is a lack of data on some variables. However, it should be noted that data on all variables that are used in this type of equation are all obtainable from the APC and AVL systems mounted on the buses.

The BDSM Model

If we assume that all passengers board at the front door and alight at the rear-door, the dwell time can be easily predicted as the maximum of the two service times. However, data collected from the APC system by the OC Transpo show that the on-board passengers alight at all doors while waiting passengers can get on the bus at the front door of a 2 door-bus and from all doors of the 3-door bus. This implies that the dwell time is actually the service time at the busiest door. The same suggestion can be found in the Highway Capacity Manual (Transportation Research Board 2000). However, how to find the busiest door was not discussed in literature so far. In our research, we propose a model to find busiest door for rigid body and articulated buses and dwell time as presented below.

Rigid-body bus

A rigid-body bus has usually two doors, one in front and the other in the rear. At a bus stop, passengers can only get on the bus from the front door while they can get off the bus at any door convenient for them. Before the bus completely opens the doors, alighting passengers tend to move to the doors for ease of exiting. Alighting passengers have priority to get off the bus first. If some passengers get off via the front door, on-ground passengers have to wait until the last alighter leaves the bus before getting on. This will lengthen dwell time compared to the case when all alighting passengers use the rear door and all boarding passengers use the front door. Obviously, if passengers get off at the front door, the dwell time is increased. If we can find the busier door, we will have a better prediction of dwell time of the bus.

On-board passengers are likely to choose the most convenient door for alighting. Of course, passengers near a door will usually use that door for alighting because of a short distance. Based on the distance and the ease to move to the doors, passengers in the middle of the bus will choose one of the doors for alighting.

Logically, passengers will use the door that offers them the highest convenience. Intuitively, the convenience of a door can be presented by several variables such as the number of boarding and alighting passengers, the number of standees on the bus, the stop location, time of day, season, and so on. The likelihood (or the probability) that a door is the most convenient and therefore the busier door depends on an unobserved variable (i.e. the convenience that a door offers to passengers at each bus stop) whose values can be quantified by a series of data retrieved from APC and AVL systems.

Let us call U as *"convenience"* function. Next, we assume that we have to relate the probability that the front door is the busier door (P_{front_door}) with convenience value that the front door offers to alighting passengers at bus stops.

The probability that the front door is the busier door can be calculated as shown below.

$$P_{front\_door} = P_1 = e^U \Big/ 1 + e^U \tag{9}$$

The probability that the rear door is the busier door (P_{rear_door}) is as shown below.

$$P_{rear\_door} = P_2 = 1 \Big/ 1 + e^U \tag{10}$$

$$P_1 + P_2 = 1 \tag{11}$$

The APC data of 7 bus stops for route 1 were used to develop the model. At each stop, the sums of passengers through the front door and the rear door were calculated and compared to each other for every bus trip. If the front door is the busier door, then it is coded as 1, otherwise as 0.

As noted previously, dwell times longer than 180 seconds or equal to zero are not included in the model. Also, the cases that recorded boarding passengers at the rear door are considered as "unusual events" and erased. Several types of *U* equations are proposed. The regression coefficients of each equation are estimated through an interactive maximum likelihood method and the best regression equation was found and tabulated in Table 3.

Table 3 - Best Logistic Regression equations

Type	The Best Logistic Regression Equation	-2LL	Nagelkerke R square	Overall Accuracy
B-L.1	$\underset{\chi^2=56267}{U} = -\underset{Sig=0.00}{1.023} + \underset{Sig=0.00}{1.205}(TOTAL\_ONS)$ $-\underset{Sig=0.00}{0.314}(TOTAL\_OFFS) + \underset{Sig=0.002}{0.017}(LOAD\_ARR)$ $-\underset{Sig=0.073}{0.001}.(PUNT) + \underset{Sig=0.000}{0.704}.(STOP\_LOCATE\ (1))$	1804.39	0.586	86.3%
B-L.2	$\underset{\chi^2=2520.600}{U} = \underset{Sig=0.063}{0.510} + \underset{Sig=0.000}{1.091}.(TOTAL\_ON)$ $-\underset{Sig=0.000}{0.018}.(TOTAL\_ON)^2 - \underset{Sig=0.000}{0.028}(LOAD\_ARR)$ $-\underset{Sig=0.001}{0.821}(STOP\_LOCA)$	948.378	0.469	89%
B-L.3	$U = \underset{Sig=0.00}{1.227} + \underset{0.00}{1.460}(TOTAL\_ON) - \underset{Sig=0.000}{0.665}(TOTAL\_OFF)$ $-\underset{Sig=0.000}{0.024}.(TOTAL\_ON)^2 + \underset{Sig=0.000}{0.018}(TOTAL\_OFFS)^2$ $-\underset{Sig=0.000}{0.022}(TOTAL\_ONS)(TOTAL\_OFFS)$ $-\underset{Sig=0.000}{0.042}.(LOAD\_ARR)$	861.847	0.529	89%

As can be seen in Table 3, all equation types are quite powerful in term of prediction. To obtain further information for selection, we use the R^2_L, which was suggested by Menard

(2001, p.27). According to this source, "R^2_L *is the most appropriate for logistic regression because it is conceptually closest to the Ordinary Least Square*".

$$R_L^2 = \chi^2 / \chi^2 - 2LL \qquad (12)$$

Where χ^2 = the *Chi-square* of the logistic regression model.

Type B-L.1 shows the highest R^2_L. As a result, it is used for this research study.

Table 4 presents prediction performance of type B-L.1 by using SPSS software. After 2-step running with the cut value of 0.5 the software returns the overall correct prediction of the model is up to 86.3 percent. Out of the 1740 cases recorded by the APC system that show the busier door to be the front door, this model predicted correctly 1552 cases. This results in an accuracy of up to 89.1 percent. Also, out of the observed 708 cases where the rear door is the busier door, the model predicted correctly 647 cases, resulting in a good prediction with up to 80.1% of accuracy.

Once the busier door has been defined, dwell time is determined by Equations 13 and 14.

$$dwell = P_1.(TOTAL\_OFFS).t_a + (TOTAL\_ONS).t_b + t_{op} \qquad if \quad P_1 \geq P_2 \qquad (13)$$

$$dwell = Max\{ P_1.(TOTAL\_OFFS).t_a + (TOTAL\_ONS).t_b \;;\; P2.(\ TOTAL\_OFFS).t_a \} + t_{op} \qquad if \quad P_1 < P_2 \qquad (14)$$

Where t_a, t_b are average service time in seconds (s) per passenger for alighting and boarding, respectively, and
t_{op} is average door opening and closing time (4 seconds).

Table 4 - Prediction performance of Type B-L.1

Classification Table[a]				
Observed		Predicted		
		BUSIER_DOOR		Percentage Correct
		Busier door is rear door	Busier Door is Front door	
BUSIER_DOOR	Busier door is rear door	647	161	80.1
	Busier Door is Front door	188	1552	89.2
Overall Percentage				86.3
a. The cut value is .500				

Articulated bus

Unlike a rigid bus, an articulated bus has three doors and passengers can board via rear doors. To predict the busiest door out of the three doors of the bus, multinomial *Logit* regression is applied. The multinomial logistic regression is an extension of the binary logistic regression applied to rigid body bus.

Similar to the explanations in the case of two-door bus, we use the *convenience functions* namely U_1 and U_2. Their variables are obtainable from APC and AVL systems data, one for each door (i.e., the front door and the first rear door) relative to the second rear door (i.e., the reference category). The probability that the front door is the busiest door is:

$$P_1 = e^{U_1} \Big/ e^{U_1} + e^{U_2} + 1 \qquad (15)$$

Similarly, the probability that the first rear door is the busiest door is:

$$P_2 = e^{U_2} \Big/ e^{U_1} + e^{U_2} + 1 \qquad (16)$$

and, the probability that the second rear door is the busiest door is:

$$P_2 = 1 \Big/ e^{U_1} + e^{U_2} + 1 \qquad (17)$$

Obviously that:

$$P_1+P_2+P_3=1 \qquad (18)$$

A number of U_1 and U_2 equations are examined in order to find the best *Logit* regression equations based on APC data belonging to 7 stops on bus route 95. Total number of boarding and alighting passengers at each door of the bus at each bus stop were recorded and compared with those of the other doors. If a door is the busiest door, it is coded as 1 and the remaining doors are coded as zero. Three types of equations named as B-ML.1, B-ML.2 and B-ML.3 were tested with stepwise method and the most powerful predictors are tabulated in Table 5.

As shown in Table 5, type B-ML.3 provides the largest *McFaddden* R^2_L of 0.224 and the overall accuracy of 65.9 percent. Therefore, type B-ML.3 is suggested for busiest door prediction of articulated bus. Also, it works fairly well, as indicated by χ^2 of 861.847 and *McFaddden* R^2_L of 0.224.

As for the results, the probabilities of front door and the first rear door are both reduced when more passengers get off the buses, but the front door has a larger reduction. In contrast, these probabilities increase when more passengers get on the bus and the influence on the front door has a larger increment. Interestingly, location of bus stop has a quite strong influence on passenger's door choice. This can be seen in *Logit* (*door_1_busiest*) and *Logit* (*door_2_busiest)* in the form of 1.824 and 1.033, if the stop is located in the CBD area.

Following the best multinomial logistic regression type, we can estimate the dwell time at the busiest door, as shown in Equations 18 and 19.

Table 5 - The Best Multinomial Logistic Regression equations

Type	The Best Logistic Regression Equation	χ^2- deviance	McFaden R square	Overall Accuracy
B-ML.1	$U_1 = 1.302 + 0.098 * TOTAL\_ONS - 0.308 * TOTAL\_OFFS$ $- 6.10^{-3} * LOAD\_ARR - 0.105 * TIME + 6.7.10^{-3} * SEASON$ $+ 0.549 * STOP\_LOCA$ $U_2 = -0.133 - 0.001 * TOTAL\_ONS - 0.023 * TOTAL\_OFFS$ $+ 0.004 * LOAD\_ARR - 0.114 * TIME - 0.022 * SEASON$ $+ 0.057 * STOP\_LOCA$	10360.49	0.193	62.3
B-ML.2	$U_1 = 1.602 + 1.50 * TOTAL\_ONS - 0.30 * TOTAL\_OFFS$ $- 0.003 * TOTAL\_ON^2 - 0.013 * LOAD\_ARR - 0.116 * TIME$ $+ 0.122 * SEASON + 0.659 * STOP\_LOCA$ $U_2 = -0.228 + 0.016(TOTAL\_ONS) - 0.023.(TOTAL\_OFFS)$ $+ 0.003 LOAD\_ARR) - 0.104.(TIME) - 0.011(SEASON)$ $+ 0.034.(STOP\_LOCA)$	8352.10	0.194	65.2
B-ML.3	$U_1 = 2.230 + 0.166 * TOTAL\_ONS - 0.513 * TOTAL\_OFFS$ $- 0.006 * TOTAL\_ON^2 + 0.006 * TOTAL\_OFF^2$ $+ 0.01 * TOTAL\_ONS * TOTAL\_OFFS - 0.016 * LOAD\_ARR$ $- 0.172 * TIME + 0.133 * SEASON + 0.601 * STOP\_LOCA$ $U_2 = -0.178 + 0.19 * TOTAL\_ONS - 0.032.* TOTAL\_OFFS$ $- 0.001 * TOTAL\_ON^2 + 0003 * LOAD\_ARR - 1.05 * TIME$ $- 0.01 * SEASON + 0.033 * STOP\_LOCA$	861.85	0.224	65.9

$$dwell = P_{Busiest\_door}(TOTAL\_ONS).t_a + P_{Busiest\_door}.(TOTAL\_OFFS).t_b + t_{op} \qquad (18)$$

$$P_{Busiest\_door} = Max\ (P_1, P_2, and\ P_3) \qquad (19)$$

Where

P_1, P_2, P_3 are the probabilities that the busiest door is the front door, the first rear door and the second rear door, and t_a, t_b, are in accordance with the TRB (2000) suggested values (Exhibit 27-9, p.27-10).

COMPARISON OF RESM AND BDSM MODULES

Two modules have been developed to predict dwell time in the previous sections. The RESM is based on the best dwell time regression function (i.e., type A-2.3) without clustering of data. BDSM is based on the revisions on busiest door selection method with two samples, one for rigid-body bus and the other for articulated bus.

In order to find advanced methodology for use in this research, two methods are compared, based on their prediction performance in terms of the overall accuracy by using the mean absolute percentage error (MAPE)

$$MAPE=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{|X_i - x_i|}{X_i}\right).100 \tag{20}$$

Where:

X_i = Counted value of dwell time by the APC system

x_i = predicted value

The prediction power of the two modules presented in Table 6 did prove that RESM is quite poor in predictions for rigid bus (i.e. MAPE up to 59.80%). But, it works well for predicting dwell time for articulated buses with MAPE as low as 35.79%. In contrast, BDSM outperforms RESM in case of rigid-body bus (MAPE= 32.42%) as compared to that of articulated bus (MAPE=45.72%).

As a result, RESM is suggested for predicting dwell time for an articulated bus while BDSM should be used for a rigid-body bus.

Table 6 - Mean Absolute Prediction Error of Method A vs. Method B

Bus Type	MAPE (%)		20% sample size
	RESM	BDSM	
Rigid-body-Bus	59.80	*32.42*	473 cases
Articulated Bus	*35.79*	45.72	1846 cases

DISCUSSION OF RESULTS

In previous research studies, dwell time was usually considered as delay and its values were defined by multiple regressions where the most influencing variables are total alighters and boarders.

Two models were developed to examine the fact that dwell time has a complicated relationship with many characteristics of passenger activities and service activities. A series of regression equation type, varying from multiple linear regression (i.e. non-linear regression, binary *Logit* regression, and multinomial *Logistic* regressions) were applied to explore this relationship.

Data collected by the APC system mounted on the OC Transpo buses were used as the inputs to these methods. The comparisons among them show that RESM (i.e. non-linear regression model) outperformed BDSM (i.e., Multinomial *Logit* regression) in predicting dwell time for the articulated bus. On the other hand, BDSM (i.e. binary *Logit*) is suitable for the rigid-body bus.

On the basis of the mean absolute percentage error (MAPE) statistical tests, successful models were identified for predicting dwell time for articulated buses and for rigid-body buses. The comparisons of the developed methods show that the non-linear regression model outperformed the Multinomial *Logit* regression in predicting dwell time for the articulated bus. On the other hand, the binary Logit is suitable for the rigid-body bus.

CONCLUSION

Tests made with two proposed models by using actual data show that these perform well and are robust. Therefore, these models can be useful to bus service providers for increasing the punctuality of their bus service for busy bus stops and ultimately these tools have the potential to enhance the transit ridership.

ACKNOWLEDGEMENTS

Financial assistance from the Natural Sciences and Engineering Research Council (NSERC) of Canada is acknowledged. The cooperation of City of Ottawa and the OC Transpo (Ottawa) in providing data is much appreciated. The views expressed are those of the authors.

REFERENCES

Abdelfattah, A. and Khan, A.M. (1997), Models for Predicting Bus Delays. *Transportation Research Record* 1623, Transportation Research Board, pp. 8-15.

Bertini, R., and El-Geneidy, A.M., (2004), Modeling Transit Trip Time Using Archived Bus Dispatch System Data. *Journal of Transportation Engineering*, Vol. 130, No.1, January 1, pp. 56-67.

Cathey, F.W., Dailey, D.J., (2003), A Prescription for Transit Arrival/ Departure Prediction using Automatic Vehicle Location Data. Transportation *Research Part C*, Vol.11, pp.241-264.

Dailey, D.J., Maclean, S.D., Cathey, F.W., and Wall, Z.R., (2001), Transit Vehicle Arrival Prediction, Algorithm and Large-Scale Implementation. *Transportation Research Record* 1771, Transportation Research Board, pp. 46-51.

Dueker, K.J., Kompel, T.J., Strathman, J.G., (2004), Determinants of Bus Dwell Time. *Journal of Public Transportation*, Vol.7, No.1, pp. 21-38.

Furth G.P., (2003). *Use of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential*. Transit Cooperative Research Program (TCRP) Web Document 23 (Project H-28), Transportation Research Board, Washington, DC.

Horbury, A.X.,(1999), Using non-real-time Automatic Vehicle Location Data to Improve bus Services. *Transportaton Research Part B*. V33. pp.559-579.

Levinson, H.S., (1983), Analyzing Transit Travel Time Performance. *Transportation Research Record* 915, Transportation Research Board, pp. 1-7.

Menard, S.W. (2002). *Applied Logistic Regression Analysis*, Thousand Oaks, California.
Mishalani, R.G., Lee, S., and McCord, M.R. (2000), Evaluating Real-Time Bus Arrival Information Systems. Transportation Research Record, 1731, pp. 81-87.

Pile, R., Niemayer, J. and Chrishom, R. (1998),*Transit Scheduling: Basic and Advanced Manuals.* TCRP Report 30, Transportation Research Board, Washington, D.C.

Rajbhandari, R., Chien, S.I., and Daniel, J.R. (2003), Estimation of Bus Dwell Times with Automatic Passenger Counter Information. *Transportation Research Record* 1841, Transportation Research Board, pp. 120-127.

Schiavon, J.J. (1999), *Understanding and Applying Advanced on Board Bus Electronics*. TCRP Report 43. Transportation Research Board.

Shalaby, A., Farhan, A. (2004), Prediction Models of Bus Arrival and Departure Times Using AVL and APC Data. *Journal of Public Transportation*, Vol.7, No.1, pp. 41-60.
Website: http://www.nctr.usf.edu/jpt/journalarticles.html

Son, B., Kim, H.J., Shin, C-H, Lee, S-K. (2004), *Bus Arrival Time Prediction Method for ITS Application*. Retrieved from the Website, 2009.

Transit Cooperative Research Program (TCRP) (2000). *Strategies for Improved Traveller Information*, Report 92, Transportation Research Board, Washington, DC.

Transit Cooperative Research Program (TCRP) (1998), *Understanding and Applying Advanced On-Board Bus Electronics*, Report 43, Transportation Research Board, Washington, DC.

Transit Cooperative Research Program (TCRP) (1998). *Passenger Counting Technologies and Procedures,* TCRP Synthesis 29, Transportation Research Board, Washington, DC.

Transportation Research Board (TRB) (2000). *Highway Capacity Manual. National Research Council, Washington, D.C.*

Vu, N.H., and Khan, A.M. (2010). Bus running time prediction using statistical pattern recognition technique. *Transportation Planning and Technology*. Vol.33, No.7, October 2010, 625-642.