Research Article

Predicting Dwell Time by Bus Stop Type and Time of the Day

Stephen A Arhin*, Errol C Noel, Melissa Anderson, Lakeasha Williams, Asteway Ribbiso and Regis Stinson

Department of Civil and Environmental Engineering, Howard University, Washington, DC, 20059, USA

Abstract

Bus dwell time (DT), which is defined as the time interval between the opening and closing its doors to serve passengers at the bus stop, is an important element in improving the travel time between end terminals of bus routes. DT on bus routes in dense urban areas varies by time of day. Also, DT could be measured or estimated using mathematical models. This study aimed at developing innovative DT models for bus stops located in dense urban areas taking in consideration the bus stop type (located near intersections and at mid-blocks), and by time of the day (morning, mid-day and evening). The models were developed using simple ordinary least squares methods with all statistical inferences at 95% confidence interval. The results of the data analysis showed that DT, on average, was higher at bus stops near intersections than those at mid-blocks. The models obtained for DT were determined to be statistically significant at 95% confidence level, based on the R^2, F-Statistic and model validation tests. The Kolmogorov-Smirnoff, normal probability and residual plots were used to confirm the adequacy of the models. The analysis also revealed that the models were significantly different by time of day and by bus stop type. It should also be noted that the models were based on bus transit operation in a dense urban area and may not be appropriate for predicting DT in non-similar settings.

Keywords: Bus reliability; Dwell time; Urban areas

Introduction

Several variables such as traffic congestion, weather condition, dwell time and number of passengers boarding or alighting directly affect the efficiency of a transit bus system along its route. The Highway Capacity Manual [1] defines dwell time (DT) as the time that a transit vehicle stops for the purpose of serving passengers. It includes the total passenger service time between the opening and closing of doors. The DT at bus stops represents a significant portion of route operating time, and its variability is linked to the reliability of the service being provided. Regional bus transit systems, such as the Washington Metropolitan Area Transit Authority (WMATA), measure DT since it affects overall transit reliability. WMATA provides bus transit system users with timely information regarding bus arrivals and travel times online. Bus passengers in some metropolitan areas in the U.S. could use their smartphones, standard computers and a variation of portable information devices to access real-time information on the arrival status of buses at any bus stop. WMATA recently incorporated Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) systems to improve its bus information system. The reliability of a bus route service is generally gauged by determining whether the transit system is compliant with its advertised schedules. Since DT is a critical factor in reliability assessments, it is essential to predict its value along bus transit corridors. Some studies [2] developed DT models which are thought to be monolithic and as a result, only one model is typically used in transit reliability decision-making processes. Such studies did not take in consideration factors such as time of the day and bus stop type into consideration. This research aimed at developing DT models using field data for WMATA's bus stops along heavily travelled corridors in the District of Columbia by time of the day and bus stop type.

Objectives

This research was aimed at developing DT models that take in consideration time of the day and bus stop types; mid-block and near intersections. Dwell time estimation for peak and non-peak travel could be used to improve estimated route travel time. The outcome could contribute toward more reliable bus transit schedules, improved published route schedules and increased accuracy in accounting for travel time variation along bus routes.

Literature Review

Generally, due to the cost and time involved in manual data collection, studies on bus dwell time are sparse. As a result, some of the dwell time studies were based on small sample sizes and mostly focused on evaluating various issues that could result in bus delays, including time of the day, method of payment and passenger load. Some studies developed dwell time models based on small bus stop sample sizes (less than 30). In addition, several studies used ordinary least squares (OLS) regression methods to determine the relationship between dwell time and various bus stop operation variables such as bus lift operations, passenger boarding and alighting, and payment methods deemed to likely influence dwell time [2]. Milkovits [3] developed dwell time regression model for transit buses in Chicago, IL using data from the automatic fare counting, automatic passenger counting, and automatic vehicle location systems. In addition to developing the model, the study analyzed the impact of the secondary factors on dwell time. The outcome of the study indicated that smart media fare cards recorded, on average, 1.5-s faster transaction time than magnetic strip tickets when the buses are uncrowded. When the bus is crowded, (i.e., number of onboard passengers exceeds the seating capacity), there is no statistically significant difference between the fare media types [3]. In a study conducted by Rajbhandari et al. [4], passenger demand was determined to be the main factor influencing dwell time. The authors stated that more time could be saved by reducing dwell time than by installing bus priority systems [4]. They also identified the total number of mobility impaired passengers, the time it took for passengers to alight when the bus was packed, and the length of time the door was opened for passengers to board, as variables that influenced bus dwell

time. Based on ordinary least squares (OLS) method, four regression models were developed using different variables. These models did not specifically cover time of the day and by bus stop type. Hooi Ling Khoo [5], presented bus dwell time models developed based on data collected for 20 bus stops in Malaysia. The models were aimed at determining the best statistical distribution to explain and describe dwell time variability [5]. The study locations were selected based on an estimation of passenger demand, type of location and number of bus routes served by the bus stop. Data was collected during peak (8:00 A.M. to 9:00 A.M. and 5:00 P.M. to 6:00 P.M.) and off peak hours (9:00 A.M. to 10:00 A.M. and 4:00 P.M. to 5:00 P.M.) in 2010 and 2011 using video recording. Statistical analysis was conducted to determine which variables influenced dwell time the most. The independent variables considered were: time of day (peak hour/off-peak hour), platform crowding level, payment method, and number of passengers boarding and alighting. The results indicated that the more crowded the platform was, the longer the dwell time. In addition, dwell time was less disperse during off peak hours than peak hours due to traffic congestion. Finally, the study also confirmed that time of the day, the number of passengers boarding and alighting directly impacted dwell-time variability. Kraft and Bergen found that the time required to service passengers for morning and evening peak hours are similar, while midday peak hours' required higher service time. From the study, it was also concluded that boarding times are higher than alighting times [6]. The author led another study where seven major variables that have a direct impact on bus dwell time were identified. These are: climate, operating practices, operating policies, mobility, human, modal and other system elements [7]. From the study, it was determined that the time a bus spends at a bus stop location could be influenced by right lane volume, curb lane usage, right lane configuration, gaps in traffic, vehicle classifications, presence of parking, and the length of the space the bus had to maneuver in and out of the bus stop. A study aimed at establishing bus way capacity and determining the variables that affect bus way performance was conducted by Gardner and Cornwell [8]. The authors determined that dwell time is one of the most relevant variables when evaluating bus transit service and bus way performance. The research revealed that there is a strong relationship between the number of onboard passengers and dwell time. Dwell times were found to be longer when the bus arrived at the bus stop with a large number of passengers on-board. A transit travel time performance study [9] conducted determined that the dwell time is, on average, 5 seconds when no passengers alight or board. However, an additional 2.75 seconds per passenger is added to the dwell time with boarding and alighting passengers. Similarly, Guenthner [10] reported that there is a 10-20 second penalty for each stop plus a 3-5 second penalty for each passenger boarding or alighting. However, due to the small sample sizes (<30), the results of both studies had low explanatory power. Chen [11] studied the impact of delay on service associated with bus stops types such as bay-side and curb-side bus stops in Beijing, China. The study involved field data collection regarding the operation of buses at stops and the magnitude of delay. The study also examined variables associated arrival and departure times at the bus stops. The arrival and departure times are related to leaving and returning of a bus to the main traffic stream from the bus stop. Data was collected on twelve buses for A.M., mid-day, and P.M. peak periods in 2011 for the following variables: number of passengers on board, boarding, and alighting. It was determined from the results that dwell time is linearly related to passenger activities at the doors, regardless of the magnitude of load factors. The literature review shows a variety of dwell time models. However, most of them were not based on time of day and by bus stop type. Variables such as the number of passengers boarding and alighting were noted to greatly influence dwell times, in addition to crowding level, method of payment, time of the day, and bus stop location. As a result, the same variables were considered in this study for the models by time of day and type of bus stop.

Research Methodology

Selected bus stop locations

Sixty bus stops in the District of Columbia were selected for this study based on WMATA's Bus Stop Usage Report, published in 2014 [12]. The report ranked the bus stops according to the number of passengers boarding and alighting. Only highly ranked bus stops were selected to ensure the occurrence of bus stopping events during the data collection process. Two bus stop types were considered: near intersections and at mid-block. Thirty (30) bus stops for each type were selected.

Data collection

Since Automatic Passenger Count (APC) and Automatic Vehicle Location (AVL) data were not readily available, the research team relied on manual field data collection at the bus stops. The filed data collection technicians were trained over a 2-week duration before embarking on the data collection. Data was collected at the 60 bus stops on weekdays in 2014, at the same bus stops for the three periods in a day: morning (7:00 A.M. to 10:00 A.M.), mid-day (12:00 P.M. to 2:30 P.M.) and evening (4:00 P.M. to 6:00 P.M.).

Bus stop characteristics data obtained for each selected bus stop included:

- Number of approach lanes

- Bus pad's length

- Presence of on-street parking near bus stop

The DT for each bus stop event was computed by calculating the difference between the time the doors opened and closed. The presence of on-street parking near the bus stop was considered as either permitted or not permitted, using the codes 1 or 2 respectively. A minimum of 10 bus stop events were recorded per period resulting in a total of 1,783 bus stop events being observed at the 60 bus stops. Field data collection sheets were returned and reviewed, after which the data was imported to Microsoft Excel and SPSS for analysis. The data error was determined to be approximately 3%.

Statistical and regression analyses

Descriptive statistics of the pertinent variables were conducted which included the mean, 95% Confidence Interval, and standard deviation. These were computed for both bus stop types and for all three periods (morning, mid-day and evening).

The generalized regression model for Dwell Time was determined to assume the following form:

$$D_t = P_b.k_1 + P_a.k_2 + N.k_3 + B_p.k_4 + P_k.k_4 + \varepsilon$$

Where,

P_b = number of passengers boarding

P_a = number of passengers alighting

P_k = presence of parking near the bus stop

N = number of lanes

B_p = bus pad length

And ε= associated model error: $\varepsilon \sim N(0, \sigma^2)$

The DT models for the bus stops for the morning, mid-day and evening peak hours were developed by employing basic multivariate regression analysis. The statistical significance of the regression coefficients of the resulting model were tested at 5% level of significance. In addition, the overall statistical significance of each regression model for each bus stop type was tested using the F-test (ANOVA) at 5% level of significance. The R^2 value was used to determine the goodness of fit of the model, while the F-test was used to evaluate the null hypothesis that all regression coefficients are equal to zero versus the alternative that at least one does not. A significant F-test would indicate that the observed R^2 is reliable, and is not a spurious result of oddities in the data set. The regression models were also checked for homoscedasticity (constant variance). In addition to reviewing the normal probability plots, the two-sample Kolmogorov-Smirnov test (KS test) was used to evaluate the hypothesis that there is no difference between the cumulative distribution functions (CDFS) of the two-sample data vectors (predicted and actual dependent variables). The two-sided test uses the maximum absolute difference between the CDFS of the distributions of the two sample sets. The test statistic is

$$D^* = \max_X \left(\left| F_1(x) - F_2(x) \right| \right),$$

Where $F_1(x)$ is the proportion of x1 values less than or equal to x and $F_2(x)$ is the proportion of x2 values less than or equal to x.

Finally the Root Mean Square Error (RMSE) was computed to measure the difference between the predicted and observed values of the models. The RMSE of a prediction model with respect to the estimated variable can be defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}}$$

where X_{obs} is the observed values and X_{model} is the predicted values [13].

Results

Descriptive statistics

The summaries of the descriptive statistical analyses are presented by bus stop type and by time period in Tables 1 and 2 which shows the key descriptive statistics: the mean, standard deviation and 95% confidence interval.

The results for the bus stops located at intersections show that, on average, the DT ranged from 22.7 seconds to 32.5 seconds. The highest mean DT was also observed during mid-day. For bus stops located at mid-blocks, the average DT during the day ranged from 19.7 to 22.1

	AM	MID	PM
Mean	22.7	32.4	30.7
Stand. Dev.	9.9	19.8	20.14
95% C. Int	19 – 26.4	25 – 39.8	23.2 – 38.2

Table 1: Dwell time descriptive statistics for bus stops located near intersections.

	AM	MID	PM
Mean	20.3	22.1	19.6
Stand. Dev.	11.4	13.43	12.6
95% C. Int	16 – 24.6	17.1 – 27.1	14.9 – 24.3

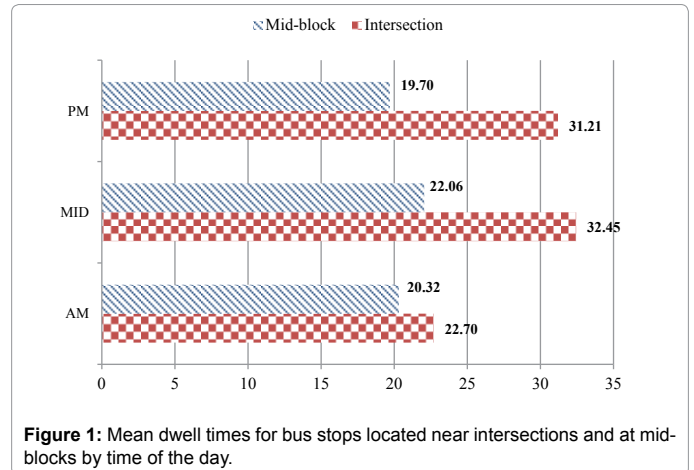**Table 2:** Dwell time descriptive statistics for bus stops located at mid-blocks.

Figure 1: Mean dwell times for bus stops located near intersections and at mid-blocks by time of the day.

Variables	DT$_{AM}$	DT$_{MID-DAY}$	DT$_{PM}$
P$_b$	0.00	0.00	0.00
P$_a$	0.02	0.00	0.00
P$_k$	0.40	0.09	0.55
L$_n$	0.45	0.12	0.89
B$_p$	0.81	0.54	0.19
Intercept	0.06	0.12	0.63

Table 3: Significance of model coefficients (p-values) for bus stops near intersections.

Variables	DT$_{AM}$	DT$_{MID-DAY}$	DT$_{PM}$
P$_b$	0.00	0.00	0.00
P$_a$	0.00	0.10	0.00
P$_k$	0.02	0.20	0.40
L$_n$	0.21	0.29	0.88
B$_p$	0.32	0.93	0.68
Intercept	0.79	0.38	0.86

Table 4: Significance of model coefficients (p-values) for bus stops located at mid-blocks.

seconds with the highest mean DT (22.1 seconds) being observed during mid-day. Figure 1 shows the mean DT by time of the day. From the figure, it was observed that the mean DT values at mid-blocks were generally lower than those for the bus stops located at intersections for the three periods observed. Furthermore, the highest DT value was observed during the mid-day period.

Regression analysis

The regression models were developed by bus stop location and by time of the day using the data obtained for the 60 bus stops. The DT models were determined based on the following:

$$DT = f\left(P_b, P_a, P_k, L_n, B_p \right) + \varepsilon$$

Where $\varepsilon \sim N(0, \sigma^2)$ with the independent variables previously defined. The following Tables 3 and 4 show the statistical significance of the model coefficients based on the variables listed.

The results presented in Tables 3 and 4 shows that only P_b and P_a contribute significantly to the prediction of dwell time. Hence, the assumption of an ideal DT model will be of the following form:

$$DT = f\left(P_A, P_B \right) + \varepsilon$$

Based on this, the resulting regression models and associated statistics are as follows.

DT Models	R²	F-value	p-value	RMSE
DT_{AM}= 1.13(P_a) + 3.34(P_b)	0.72	34.12	0.00	5.17
DT_{MID}= 1.89(P_a) + 5.48(P_b)	0.79	50.04	0.00	8.98
DT_{PM}= 2.61(P_a) + 4.29(P_b)	0.88	107.42	0.00	6.73

Table 5: Dwell time models by time of the day for bus stops located near intersections.

DT Models	R²	F-value	p-value	RMSE
DT_{AM}= 1.13(P_a) + 3.34(P_b)	0.86	79.53	0.00	4.27
DT_{MID}= 1.89(P_a) + 5.48(P_b)	0.68	28.45	0.00	7.49
DT_{PM}= 2.61(P_a) + 4.29(P_b)	0.92	144.36	0.00	3.63

Table 6: Dwell time models by time of the day for bus stops located at mid-blocks.

Variables	DT_{AM}	$DT_{MID-DAY}$	DT_{PM}
P_b	0.01	0.01	0.00
P_a	0.00	0.00	0.00

Table 7: Significance of model coefficients for bus stops near intersections.

Variables	DT_{AM}	$DT_{MID-DAY}$	DT_{PM}
P_b	0.00	0.10	0.00
P_a	0.00	0.00	0.00

Table 8: Significance of model coefficients for bus stops at mid-block.

Tables 7 and 8 present the statistical significance of the model coefficients based on the variables listed in the resulting model. The adequacy and significance of the regression models were tested at 5% level of significance.

The results showed that the models could explain relatively high percentages of the variations in the data, based on the R² values (68-92%). The highest R² values for the DT models were observed during the P.M. periods for both bus stop types. Furthermore, the p-values for the regression models' F-statistics were determined to be less than 0.05, indicating that the coefficients are not equal to zero at 5% level of significance. For all the models, the residual plots showed evenly distributed random plots about the zero line, confirming that the models fit the data sets well. Also, the normal probability plots show a line along the points, thus an assumption of normality would be reasonable for the data sets. Finally, the D-statistics and corresponding p-values of the KS tests for each model showed that there is no statistical difference between the CDFS of the observed and predicted values for each model. The different RMSEs also suggest that the models are different.

Discussion of Results

Dwell time has been identified as a measure that should be monitored and managed to improve service performance of bus transit. The DT models developed for WMATA's transit buses in the District of Columbia were based on the following independent variables: number of passengers alighting and boarding, bus pad length, presence of on-street parking and number of approach lanes. For the DT models developed for bus stop types (mid-block and near intersection) and for time of the day (morning, mid-day and evening peak hours), the prominent independent variables were number of passenger alighting and boarding. From the results of the analysis for bus stops located at mid-blocks, the maximum average DT was 22.1 seconds, which was observed during the mid-day peak period. For bus stops located at intersections, the average DT was 22.7 seconds in the morning, 32.5 seconds during mid-day and 31.2 seconds in the evening. The mean DT at intersections was generally higher at the bus stops at intersections than those located at mid-blocks. This could be attributed to the potential influence of intersection interactions including traffic, signal

operations, pedestrian crossing, congestion, parking maneuvers, etc. The overall mean DT was determined to be 29 seconds at the bus stops near intersections. For bus stops located at mid-blocks, the overall mean DT was 21 seconds. The regression models by bus stop type yielded statistically significant regression models within the margin of error (5% level of significance), with high R² values for DT (73%-95%). The results of the ANOVA tests also showed statistically significant F - statistics ($p<0.05$).

For all the DT models, the number of passengers alighting and boarding contributed significantly to the model based on the statistical significance of their coefficients. The study also revealed that the number of approach lanes, presence of parking and bus pad length do not significantly influence dwell time. Residual plots for all the models also showed randomness about the zero line indicating their viability, in addition to the normal probability plots showing points near a straight line. Moreover, the KS Test results indicated that the models adequately predicted the observed values. The RMSE differ for each of the models, thus suggesting that the models are significantly different from each other. As a result, a morning peak dwell time model should not be used to predict that for evening peak. The two sets of models (Tables 3, 5 and 6) show that DT can be adequately predicted based on the independent variables listed. However, in the first set of models, there are several independent variables that were not contributing significantly to the prediction of DT.

Conclusions

From the analysis results, it can be concluded that DT's differ based on the bus stop location and time of the day. Based on the field data, the proposed regression models have a high explanatory power, at 95% confidence interval. Being able to predict bus DT will enable transit decision-makers to improve bus schedule planning and overall bus reliability. These models should be updated every 2 years due to potential changes in traffic and new infrastructure. Future studies will include the comparison and validation of the two sets of models. In addition, these models can be implemented in order to shorten DT at both types of bus stops.

References

1. National Research Council (U.S.) HCM (2010) Highway Capacity Manual. Washington, D.C: Transportation Research Board.

2. Dueker JK, Kimpel TJ, Strathman JG, Callas S (2004) Determinants of Bus Dwell Time. J Public Transportation 7: 21-40

3. Milkovits MN (2008) Modeling the Factors Affecting Bus Stop Dwell Time: Use of Automatic Passenger Counting, Automatic Fare Counting, and Automatic Vehicle Location Data. Transportation Res Record 2072: 125-130.

4. Rajbhandari R, Chien S, Daniel J (2003) Estimation Dwell Times with Automatic Passenger Counter Information. Transportation Res Record: J Transportation Research Board 1841.

5. Khoo HL (2013) Statistical Modeling of Bus Dwell Time at Stops. Eastern Asia Society Transportation Stud Volume 9.

6. Kraft W, Bergen T (1974) Evaluation of Passenger Service Times for Street Transit Systems. Transportation Res Record 505: 13-20.

7. Kraft W (1975) An Analysis of the passenger Vehicle Interface of Street Transit Systems with Applications to Design Optimization. Ph. D dissertation. New Jersey Institute of Technology.

8. Gardner G, Cornwell PR, Crackwell JA (1991) The Performance of Busway Transit in Developing Cities. Transport and Road Research Laboratory, Department of Transport, Report 329.

9. Levinson HS (1983) Analyzing Transit Travel Time Performance. Transportation Res Record 915: 1-6.

10. Guenthner RP, Sinha KC (1983) Modeling Bus Delays Due to Passenger

Boarding and Alighting. Transportation Res Record 915: 7-13.

11. Chen S, Zhou R, Zhou Y, Mao B (2013) Computation on Bus Delays at Stops in Beijing through Statistical Analysis. Mathematical Problems Eng3013, Article ID745370.

12. Washington Metropolitan Area Transit Authority (WMATA) (2014) Weekday Stop Usage Report .

13. Petch MG (2008) Prognostics and Health Management of Electronics, John Wiley & Sons.