ANALYSIS OF TRENDS IN TRANSIT BUS DWELL TIME DATA

Isaac K. Isukapati
The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania
isaack@cs.cmu.edu

Hana Rudová
Faculty of Informatics
Masaryk University
Czech Republic
hanka@fi.muni.cz

Gregory J. Barlow
The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania
gjb@cmu.edu

Stephen F. Smith
The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania
sfs@cs.cmu.edu

ABSTRACT

Transit vehicles create special challenges for urban traffic signal control. Signal timing plans are typically designed for the flow of passenger vehicles, but transit vehicles, with frequent stops and uncertain dwell times, may have very different flow patterns that fail to match signal coordination plans. The presence of transit vehicles stopping on urban streets can also restrict or block other traffic on the road, resulting in further disruption to coordination. These factors can result in increased overall wait times and delays throughout the system for transit vehicles and other traffic. Transit signal priority (TSP) systems are often used to mitigate some of these issues, primarily addressing delay to the transit vehicles. However, predominant existing TSP strategies give unconditional priority to transit vehicles, thereby exacerbating quality of service for other modes.

In areas where transit vehicles have significant effects on traffic congestion, particularly in urban areas, using more realistic models of bus behavior in traffic signal control strategies could reduce delay for all travel modes, particularly in a connected vehicle context using adaptive control. However, estimating the arrival time of a transit vehicle at an intersection requires an accurate model of transit stop dwell times. As a first step toward developing a dwell time model for purposes of predicting bus arrival times, this paper analyzes trends in automatic vehicle location (AVL) data provided by the Port Authority of Allegheny County (PAAC) collected over the two year period from September 2012 – August 2014 for two major bus routes. Our analysis enables several inferences to be drawn. First, the statistical properties of dwell times are similar (for most stops) across years for a given season and hence it is fine to join the data for the same season (or month) across years. Second, the probability of a non-zero dwell time varies from stop to stop in a given route suggesting that buses need not be given same priority at all signalized intersections. Third, cumulative density functions (CDFs) of dwell time distributions do provide insights into reliability of dwell times for a given stop; this information is especially useful in real-time control decisions; Fourth, fifteen minute interval dwell time CDFs of peak hour demonstrate the highly stochastic nature of dwell times. Based on this trend analysis, we argue that an effective predictive dwell time distribution model must treat independent variables as random or stochastic regressors.

INTRODUCTION

Control of urban transportation networks is complicated by the multi-modal nature of traffic dynamics, involving passenger cars, pedestrians, transit vehicles, bicyclists, and other modes of travel. Transit vehicles in particular create special challenges for urban traffic signal control.

Most traffic signal timing plans are designed for the flow of passenger vehicles, but transit vehicles may have very different driving behaviors, particularly in urban areas. Unlike passenger vehicles, transit vehicles may have frequent stops with uncertain dwell times, producing flow patterns that fail to match signal coordination plans. The presence of transit vehicles stopping on urban streets can also restrict or block other traffic on the road depending on stop locations, resulting in further disruption to coordinated traffic flow. These factors can result in increased overall wait times and delays throughout the system. This decrease in mobility, in turn, deeply impacts the reliability of transit vehicle schedules, degrading the experience of the large number of travelers who depend on transit for mobility (which ultimately affects ridership levels).

To help mitigate these delays to transit vehicles, cities often deploy transit signal priority (TSP) systems. In these systems, transit vehicles are equipped with a device that has the ability to communicate priority requests to the roadside signal control infrastructure; TSP control logic in turn responds to this request by implementing one of two actions: 1) holding the green for transit vehicles (if priority is requested on the existing phase in service), or 2) abruptly ending the current phase in order to serve the phase that the priority request came from. Even though these systems can significantly improve the mobility of transit vehicles in urban signalized networks, they have several shortcomings. First, the use of unconditional priority in signal control systems is most appropriate for safety reasons (e.g., rail preemption or emergency vehicles). Giving strict priority to transit vehicles ignores the rest of the traffic on the road, and can have deleterious effects on overall system wait times and throughput. Second, in circumstances with competing priority requests, the adverse effects to overall traffic flow are compounded by a basic first come, first serve policy. Third, special priority needs only be given to buses when they are running behind schedule.

One way to address these shortcomings is to consider a better model of transit vehicle behaviors in the broader context of real-time traffic signal control, and attempt to factor in other traffic flows when optimizing bus movements. A step in this direction has been taken recently in (*1*) where bus priority is considered together with vehicle platoon "coordination" priority to make real-time phase change decisions. However, this scheme assumes a specification of the relative importance of different priority requests, and hence still gives strict preference to higher priority requests. We believe that a weighted priority scheme, such as that employed in the adaptive traffic signal control strategy of the Surtrac system (*2, 3*) can provide an alternative basis for optimizing the movements of buses that does not require strict stratification of priority requests. Priority is most often based on vehicle delay, but the fullness of a transit vehicle could allow its weight (priority) to be based on person delay instead. This approach is capable of better balancing competing traffic flows (e.g., should a bus always have priority over a vehicle platoon, regardless of platoon size?), particularly in dense transit areas where the needs of competing transit vehicles must be balanced.

To model the movement of transit vehicles, reliable prediction of dwell times is essential. While transit vehicles differ in other ways from passenger vehicles, such as slower acceleration and lower maneuverability, the frequency of stops and the length of those stops are what most set transit vehicle trajectories apart from smaller passenger vehicles. Dwell times — the length of time a transit vehicle is stopped to unload and load passengers — are determined by many factors, such as: the number of passengers boarding or alighting, the types of passengers (handicapped passengers often require more time), or the way a passenger pays for their trip (cash payments require longer dwell times than smart cards or off-board payment). In trying to predict dwell times, we must first understand the statistical properties of dwell times in order to determine the feasibility of taking this approach to multi-modal traffic signal control.

In this paper, we take the first step toward developing a dwell time model for the purpose of predicting bus arrival times, by analyzing trends in automatic vehicle location (AVL) data. Using data provided by the Port Authority of Allegheny County (PAAC) collected over the two year period from September 2012 – August 2014, we consider two major bus routes, focusing on a segment of the route through a corridor of interest where both a vehicle-to-infrastructure (V2I) communications testbed using dedicated short-range commuication (DSRC) and the Surtrac adaptive traffic signal control system are currently deployed. Ultimately, a system to incorporate bus movement prediction into Surtrac is envisioned, using DSRC as a mechanism to detect bus locations.

This paper is organized as follows. First, we present a review of related literature. We then describe the transit AVL dataset and present analysis of the data. A more thorough look at how this analysis could be incorporated into a real-time, model-based optimization approach to traffic signal control is presented, along with other potential applications of the analysis in this paper. Finally, we present some conclusions.

LITERATURE REVIEW

Transit bus dwell time is defined as the duration a transit vehicle is stopped for serving passengers, including the time needed to open and close passenger doors (*4*). It is widely accepted that bus dwell times play an instrumental role in transit operations (*5*). Widespread implementation of automatic passenger counting (APC), automatic fare counting (AFC), and automatic vehicle location (AVL) systems data provide a basis for rich statistical insights into dwell times. Even though dwell times are highly correlated with the number of passengers boarding and alighting, secondary factors such as crowding or fare type (e.g., card vs cash) may also have a large effect on dwell times. Transit operators are interested in understanding bus dwell times with an intention of coming up with better strategies to improve service. Hence, previous research efforts on dwell time prediction models have focused on the factors influencing average dwell times (*6–12*). Various approaches such as regression models (*13*), probabilistic approaches (*14*), decision trees (*11*), or time series models (*15*) have been studied to compute the bus dwell time, though not always using information available during real-time operations. A model for the computation of the bus dwell times based on information available in real-time was published in (*16*) as a part of a study devoted to the computation of bus arrival and departure times. A prediction model was developed using Kalman filters for passenger arrival rate and the headway, where headway corresponds to the actual arrival time of the last bus minus the predicted arrival time of the next bus. Algorithms employ historical data over several days as well as the data from the previous bus. More recent work inspired by this model in (*17*) developed a basis for predicting the number of boarding and alighting passengers, based on historical data and the information of the previous bus on the current day, and these predictions were then used together with crowding effects to compute the bus dwell time. Evaluation in both studies (*16, 17*) was completed using AVL and APC data from just a few days, where one set of days (four days in (*16*), two days in (*17*)) were used for the model calibration and the performance evaluation was completed using the data of the last day.

Transit signal priority (TSP) has been studied in the United States since the 1970s (*18*). A recent review of TSP is available at (*19*). Various studies (*20–22*) have investigated the problem of priority from the standpoint of giving spatial priority for buses (e.g., design of bus lanes), while other studies (*23, 24*) have focused on optimal detector placement for TSP. There are numerous studies (*25–31*) that explore various control strategies for TSP. Recent studies on TSP proposed strategies using connected vehicle technology (*32, 33*) and demonstrated the advantages of real-time information. Most recently, Ding et al. (*34*) strives to integrate bus dwell times into transit signal priority. While this is an interesting study, the authors' dwell time prediction model is based on only three days of video detection and station survey data.

APPLICATION CONTEXT

As mentioned earlier, the analysis of bus dwell times described in this paper is part of a larger effort aimed at incorporating knowledge of buses into the real-time adaptive signal control strategy of the Surtrac system and using this information to optimize bus movements in an integrated way. Surtrac takes a decentralized, online planning approach to intersection control. In brief, each intersection senses its approaching traffic and (in real-time) constructs a phase schedule (i.e., timing plan) that moves approaching traffic through the intersection with minimal cumulative wait time. The head of this phase schedule is executed in rolling horizon fashion and its tail is recomputed and extended every few seconds. Each time a new phase schedule is generated, expected outflows are communicated to downstream neighbors, providing an expectation of what traffic is coming behind current locally sensed traffic, thus enabling coordinated activity at the network level. In its initial 9-intersection deployment in the East Liberty area of Pittsburgh PA, Surtrac achieved substantial reductions in travel times (25%), wait times (40%), and emissions (projected at 20%) (*2*). This deployment has subsequently been expanded several times and Surtrac currently controls a network of 50 interconnected intersections. Dedicated short-range communication (DSRC) road side equipment (RSE) units were installed at 24 intersections during the most recent expansion of the system to provide a test bed for integrating Surtrac adaptive signal control with connected vehicle technology. One initial focus is to equip some number of buses that move through this test bed corridor with DSRC on board units (OBUs) and exploit real-time mode information to move buses more effectively.

Surtrac's intersection scheduling approach provides a natural basis for exploiting real-time mode information to give active attention to bus movements. The key to its online planning effectiveness is its formulation of the intersection control problem as a special type of "single-machine" scheduling problem (*3*). Detected approaching traffic is aggregated into sequences of clusters (queues and platoons), which preserves the non-linear nature of traffic flows while allowing efficient (sub-second) computation of long horizon schedules. This scheduling model can be augmented in two ways to account for and optimize bus movements. First, mode information can be incorporated into the aggregate cluster representation and used together with knowledge of bus stop locations and dwell time information to more accurately project when a bus (and those vehicles traveling behind it) will actually arrive at the intersection; this information alone can lead to signal control decisions that improve bus movements. Second, clusters can also be weighted by mode, to give preference to clusters that contain buses without unconditionally favoring them.

TRANSIT DATA

As a first step toward developing and utilizing dwell time information to more accurately predict when traffic clusters will arrive at the intersection, this paper analyzes trends in AVL data provided

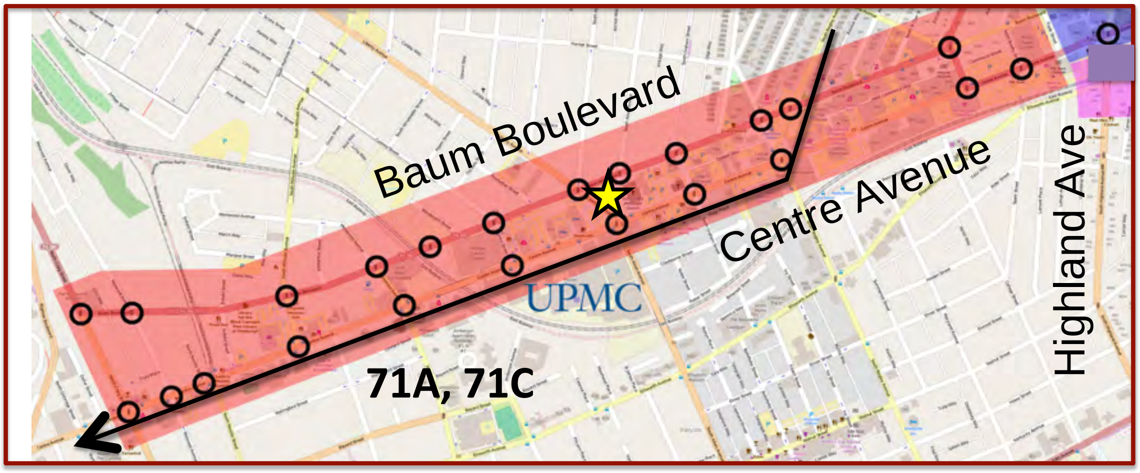


FIGURE 1 Surtrac Connected Vehicle Test Bed

by the Port Authority of Allegheny County (PAAC) for the period from September 2012 – August 2014 for two major bus routes - 71A and 71C - that travel through the Surtrac connected vehicle test bed corridor. The test bed corridor and bus routes are depicted in Figure 1. These data are recorded for weekdays, and each record corresponds to one bus stopping. The information recorded in each record includes the route (e.g., 71A), the direction of the trip (inbound directs to downtown, outbound from downtown), trip identification by start time and date of the trip, and the bus stop (identified by the unique PAAC identification number and the sequential stop number is assigned to each bus stop in the trip). There is also a bus dwell time, specified as a difference between door open and door close times.

Of course there are various artifacts in real-life data. We were not able to obtain August 2013 and November 2013 data; hence our seasonal trend analysis relies on three other months (February, July, and October). There was a small fraction of the data marked as invalid bus stops for the route, which was deleted from the data set used. The name of one bus stop used in our study was changed over the lifetime of the data, so we standardized on the older name which was present in most files (i.e., Centre Ave at Craig St NS is now called Craig St at Centre Ave FS). Data were missing for some sequential stop numbers (see for example the sequential stop number 23 in Figure 2a) implying that a small portion of bus stops is changing over the years (and the sequential stop number may change for the given PAAC number).

We restrict our analysis here to inbound trips only, i.e., the buses toward Downtown Pittsburgh. One set of graphs shows the data for all day (see Figure 6); the remaining graphs show the data for the morning rush hours from 7 am to 10 am. In this latter case we included those records where the bus arrival time was within this specified interval. Finally, we only analyze bus dwell times ≤ 100 seconds and discard any outliers for better presentation, since these outliers are likely not regular stops (likely layovers, breakdowns, etc.). Information about the number of records for each route 71A and 71C trip in the inbound direction is given in Table 1. Included is the total number of records and percentages of used and removed records for both all day and 7 – 10 am.

TABLE 1 The total number of records for routes 71A and 71C inbound and percentages of used and removed records for all day and morning peak.

	All day			7 – 10 am	
	# total	% used	% removed	% used	% removed
71A	995,709	98.71	1.29	21.80	0.29
71C	1,024,518	98.75	1.25	20.73	0.27

DATA ANALYSIS

In the subsections below we analyze several aspects of the dwell time information provided by PAAC. The first set of analyses focuses on seasonal variations in dwell times: the purpose of this analysis is to understand whether statistical properties of dwell times are similar or significantly different from year to year for a given season. Next, we present cumulative density functions (CDFs) of dwell times for stops within the connected vehicle test-bed; we draw some conclusions with respect to reliability of dwell times based on CDFs stability. Lastly, we present trends in dwell times over the course of the day. This information is useful especially to understand time of day trends in transit ridership.

Longitudinal analysis of dwell times

Bus dwell times are often affected by seasonal variations. In cold weather climates, ridership patterns look different during the summer and winter. In cities with large student populations, ridership changes drastically during the school year. The question is, whether or not the statistical properties of dwell times remain the same for a given month year to year. For the purpose of this analysis we considered February 2013 and 2014 peak-hour route 71C inbound data. Even though our main focus is on understanding the dwell time distributions in our connected vehicle test-bed, it is useful to look at these distributions for all the stops in the route and Figure 2 summarizes the descriptive statistics of dwell times for each bus stop. It contains two subplots; each subplot presents standard box plot for the dwell times for each bus stop; small red squares represent average dwell times, whereas red circles represent median values for dwell times. The values within the box represent values within inter-quartile range, and blue diamond markers represent outliers (data that are not within the range of 5 – 95% are considered outliers). Several observations can be made from these plots. First, for bus stops in the connected vehicle corridor (from Negley Ave at #370 to Centre Ave at Craig St), the distribution of dwell times for the years 2013 and 2014 look similar. Kolmogorov-Smirnov (KS) tests were performed to check for any statistical differences in these distributions and Table 3 summarizes these results; test results suggest that at 95% confidence interval, these distributions are in fact similar. Second, the dwell times are very low for stops in Uptown. Since Uptown has a smaller population and few employers, it makes sense that dwell times would be lower as not many people board the bus there in the morning. Similar analysis is conducted for two other seasons (Summer and Fall) and the results suggest that statistical properties of dwell times stay similar for most stops year to year and for a given season. Given the similarities in statistical trends, it is reasonable to join the data for the same season (or month) across years.

TABLE 2 Results of Kolmogorov-Smirnov tests for route 71C in inbound direction, 7:00 – 10:00

Stop	February 2013 vs. 2014		July 2012 vs. 2013		October 2012 vs. 2013	
	max	crit	max	crit	max	crit
Negley Ave at #370	0.080	0.481	0.094	0.555	0.032	0.480
Negley Ave at Centre Ave	0.092	0.267	0.053	0.272	0.009	0.257
Centre Ave at Graham St	0.071	0.351	0.054	0.430	0.037	0.377
Centre Ave at Aiken Ave	0.046	0.284	0.033	0.284	0.061	0.262
Centre Ave opp Shadyside hos.	0.123	0.481	0.028	0.453	0.004	0.453
Centre Ave at Cypress St	0.016	0.430	0.083	0.430	0.175	0.430
Centre Ave at Morewood Ave	0.091	0.297	0.023	0.340	0.073	0.297
Centre Ave at Millvale Ave	0.006	0.321	0.000	0.340	0.031	0.321
Centre Ave opp Neville St	0.096	0.377	0.043	0.453	0.069	0.351
Centre Ave at Melwood Ave	0.095	0.363	0.058	0.410	0.106	0.363
Centre Ave at Craig St Ns	0.072	0.321	0.051	0.340	0.023	0.312

The next sub-analysis highlights the likelihood that a bus will dwell at a given bus stop. Figure 3 summarizes these results. It contains two subplots; each subplot contains comparative bar graph of total observed records per stop (red) and a subset of those records with non-zero dwell times (light green). Probability of a bus stopping at a bus stop is the ratio of number of records with non-zero dwell times and the total number of observed records. Subplots a and b present the results for February 2013 and February 2014 respectively. As one might notice, there is a higher probability that the bus might stop in the connected vehicle corridor, Oakland, or Downtown and that probability a bus stops in Uptown is low. What this really means is that a bus need not stop at every stop, and the locations where it stops more frequently also have higher activity of passengers boarding and alighting the bus. Therefore one can argue the deleterious affects that a given bus has at a given bus stop needs to be taken into consideration before giving it an undue priority.

Cumulative density functions of dwell times

Next, we take a closer look at the trends in dwell time distributions for the intersections in the connected vehicle test-bed. The corridor starts at Negley Ave at #370, located just north of the intersection of Negley Ave with Baum Boulevard, and ends at Centre Ave at Craig Street, prior to turning left on Craig Street. A bus passes through 10 intersections on this portion of the corridor, and there are 11 possible bus stops (one stop, Centre Ave opposite Shadyside Hospital, is mid-block). AM peak hour data from three months (each representative of a season) is considered for

TABLE 3 Results of Kolmogorov-Smirnov tests for routes 71A and 71C in inbound direction, 7:00 – 10:00. Both years are joined for each route.

Stop	February 71A vs. 71C		July 71A vs. 71C		October 71A vs. 71C	
	max	crit	max	crit	max	crit
Negley Ave at #370	0.010	0.453	0.085	0.514	0.040	0.453
Negley Ave at Centre Ave	0.072	0.241	0.069	0.237	0.022	0.227
Centre Ave at Graham St	0.006	0.330	0.064	0.363	0.034	0.330
Centre Ave at Aiken Ave	0.061	0.248	0.031	0.248	0.008	0.233
Centre Ave opp Shadyside hos.	0.026	0.410	0.063	0.393	0.028	0.377
Centre Ave at Cypress St	0.053	0.377	0.079	0.363	0.053	0.340
Centre Ave at Morewood Ave	0.045	0.267	0.003	0.297	0.025	0.262
Centre Ave at Millvale Ave	0.016	0.278	0.056	0.304	0.004	0.278
Centre Ave opp Neville St	0.084	0.312	0.136	0.363	0.089	0.321
Centre Ave at Melwood Ave	0.069	0.321	0.110	0.330	0.043	0.297
Centre Ave at Craig St Ns	0.075	0.278	0.016	0.272	0.089	0.267

the analysis. February, July and October were chosen as representative months for winter, summer and fall respectively. The 71A and 71C buses share exactly the same route on this section of the corridor and the remainder of the route toward Downtown. Since these two buses are essentially interchangeable during this section of their routes, it makes sense to join these two data sets as long as the dwell time distributions of the 71A are statistically similar to those of the 71C for each bus stop in the corridor. Again, Kolmogorov-Smirnov (KS) tests were performed at 95% confidence interval to check for any statistical differences in these distributions; As summarized in Table 3, a total of 66 KS tests were performed (given that there are a total of eleven stops and three different months). Test results suggest that these distributions are in fact similar. Therefore, these two data sets were combined for the purposes of subsequent analysis.

Figure 4 presents cumulative density functions (CDFs) of dwell time distributions for three different months (each representative of a season). It contains three subplots, one for each month. The distributions with curves furthest to the left have smaller variance in dwell time distributions and hence are more reliable. The following inferences can be drawn from these plots. First, for all three seasons dwell time distributions have the largest variance at Negley Ave at Centre Ave (CDF in red), followed by Centre Ave at Aiken Ave (CDF in blue) and Centre Ave at Morewood Ave (CDF in cyan). The combination of college student apartments, a large grocery store, and one of the main cancer hospitals in the region results in many transit passengers at these three intersections. Second, dwell time distributions have slightly larger variation during the school year (February and October) as opposed to summer (July). Third, as one might expect, the bus stops with larger variations in dwell times also have a high probability of stopping for a given bus (please refer to Figure 3). This information is useful for transit system planners for assess reliability of dwell times at various stops.

Figure 5 presents cumulative density functions (CDFs) of dwell time distributions for three different bus stops for the AM peak in fifteen minute intervals. It contains three plots: subplot (a) presents distribution of dwell times for Negley Ave at Centre Ave; subplots (b) and (c) present similar results but for Centre Ave at Aiken Ave and Centre Ave opposite Shadyside Hospital stops respectively. Each plot in turn contains three subplots, one for each month. Again as mentioned earlier, the lowest variance in dwell time is reflected by the curves furthest to the left. The purpose of presenting these plots is not to make an inference about which fifteen minute window one needs to pay more attention to but to illustrate the variance in dwell times. Given these distributions of dwell times, an ordinary linear regression-based predictive model might not be the best choice.

Trends in dwell times over a day

Finally, we take a closer look at the trends in dwell time distributions during the course of the day. The same three bus stops were considered in this analysis, using February (2013 & 2014) dwell time distribution data for bus routes 71A and 71C. Typically, bus service starts at 5:30 AM and ends around midnight, so data were binned into 15 minute intervals for this duration of time. Figure 6 summarizes the descriptive statistics of dwell times for each fifteen minute time window. It contains three subplots; each subplot presents standard box plot for the dwell times for each fifteen minute window; small red squares represent average dwell times, whereas red circles represent median values for dwell times. The values within the box represent inter quartile range, and blue diamond markers represent outliers (data that is not within the range of $5-95\%$ is considered an outlier). Subplots a, b and c present the results for the bus stops at Negley Ave at Centre Ave, Centre Ave at Aiken Ave, and Centre Ave opposite Shadyside Hospital respectively. The following inferences can be drawn from these plots. First, the bus stop at Negley Ave at Center Ave has, in general, high average and median dwell times throughout the day. Trends during the AM peak can be attributed to college students commuting to classes. Higher dwell times observed in other time periods can be attributed to people boarding the bus after finishing grocery shopping at the large grocery store present at the corner of this intersection. Second, the bus stop at Centre Ave at Aiken Ave has high average and median dwell times during the AM peak (again this can be attributed to students commuting to Oakland). Third, the bus stop at Centre Ave opposite Shadyside Hospital has lower average and median dwell times than the other two bus stops. Slightly higher average dwell times between 15:00–16:00 may be explained by a shift change at the hospital across the street.

Note on choice of statistical models

As mentioned earlier, reliable prediction of bus dwell times plays an instrumental role in multimodal traffic signal control. Historically, bus dwell times are modeled as a function of load, number of boarding, and alighting of the passengers. More specifically, previous research efforts have focused on understanding factors that affect dwell times in order to analyze existing transit operational strategies. Therefore, most efforts used fixed effects regression models to understand these relationships. While such type of modeling is helpful to understand the average trends, it might not be very effective for use in real-time operations for the following reasons: First, dwell times are typically positive numbers; ordinary linear regression may not be the best choice unless these times are first transformed in a way that removes this restriction. Second, ordinary linear regression methods solely focus on understanding how the expectations of an outcome (also known as dependent) variable Y depends on one or more predictors (also known as independent variables, regressors or covariates) X. In other words, these models mainly concern themselves with exploring the nature of relationship between Y and X (e.g., linear or quadratic) and how well regressors can explain the variance in Y. Furthermore, here Y is modeled as a random variable whereas X are treated as fixed variables i.e., these variables are assumed to be measured without measurement error. However, this assumption does not typically hold in the context of real-time signal control. For example, the number of boarding and alighting of passenger information will be erroneous under the scenarios of high level crowding on-board or when the sensors are misaligned. So any regression model that treats X as fixed variables will likely predict highly erroneous dwell times. In that sense, any predictive dwell time distribution model should treat independent variables as random or stochastic regressors. Therefore, special attention needs to be paid while choosing an appropriate dwell time model for real-time operations.

APPLICATIONS

As mentioned earlier, the analysis of bus dwell times presented in this paper is part of a larger effort aimed at incorporating knowledge of buses into real-time adaptive signal control strategy such as Surtrac. However, reliable bus dwell time information has much broader applicability.

Bus dwell time data is immensely useful for transit system planners and schedulers. Transit planners have the responsibility for determining where major improvements are needed, where the buses should go, and how much service should be provided. In that regard, a service planner wants to determine which routes in the system have the poorest reliability (or in other words, the most variability in on-time performance) at the stops. This information is a useful supplement to passenger demand analysis when designing express bus routes.

On the other hand, a Transit scheduler is interested in generating reliable bus schedules. In transit scheduling it is not so much the travel time that matters, but the reliability of the travel time. The objective is to build schedules that can actually be followed. Otherwise, the bus drivers cannot make the stops at the scheduled times. Riders too cannot get to their destinations at the times listed in the schedule and there is a possibility that they might miss their transfers. Hence, the published schedule has no value. More often than not, dense urban networks have a deleterious affect on transit travel times. In that sense, reliable dwell time prediction can play an instrumental role not only in better managing traffic but also giving priority to buses (when required) to improve reliability of transit schedules.

Finally, a better understanding of dwell time information can enable the development of more sophisticated Transit Signal Priority (TSP) systems. An understanding of probability of stopping information at various bus stops will enable more targeted and cost-effective TSP system deployments. Even further, more accurate characterization of bus dwell times, together with real-time communication of mode and schedule information, will promote the development of TSP systems that are capable of conditioning bus priority on whether buses are ahead of or behind schedule.

CONCLUSIONS

In this paper, we present some of our findings on statistical properties of bus dwell times based on Pittsburgh Port Authority's AVL transit data spanning from September 2012 - August 2014. For the purpose of this analysis, we considered inbound AM peak hour (7:00 – 10:00 AM) data for a bus route that goes through our connected vehicle testbed. The following inferences are drawn based on our analysis: first, the statistical properties of dwell times are similar (for most stops) across years for a given season and hence it is fine to join the data for the same season (or month) across years. Second, the probability of a non-zero dwell time varies from stop to stop in a given route suggesting that buses need not be given the same priority at different signalized intersections. Third, cumulative density functions (CDFs) of dwell time distributions do provide insights into reliability of dwell times for a given stop; this information is especially useful in real-time control decisions; Fourth, fifteen minute interval dwell time CDFs of peak hour demonstrate the highly stochastic nature of dwell times. Fifth, we presented trends in dwell times over a day and how they are influenced by factors like student housing, market district etc. Finally, based on these trends, we argued that a useful predictive dwell time distribution model must treat independent variables as random or stochastic regressors. In future work we intend to explore the efficacy of Bayesian random effect modeling techniques for bus dwell time prediction and field test the efficacy of those models in our connected vehicle testbed.

REFERENCES

[1] Feng, Y., M. Zamanipour, K. L. Head, and S. Khoshmagham, Connected vehicle based adaptive signal control and applications. In *Transportation Research Board 95th Annual Meeting*, 2016, 16-3070.

[2] Smith, S., G. Barlow, X.-F. Xie, and Z. Rubinstein, Smart urban signal networks: Initial application of the SURTRAC adaptive traffic signal control system. In *International Conference on Automated Planning and Scheduling*, 2013.

[3] Xie, X.-F., S. F. Smith, L. Lu, and G. J. Barlow, Schedule-driven intersection control. *Transportation Research Part C: Emerging Technologies*, Vol. 24, 2012, pp. 168–189.

[4] Transportation Research Board, *HCM 2010 highway capacity manual*. Transportation Research Board, Washington, DC, 2010.

[5] Levinson, H. S., Analyzing transit travel time performance. *Transportation Research Record*, , No. 915, 1983, pp. 1–6.

[6] Zografos, K. C. and H. S. Levinson, Passenger service times for a no-fare bus system. *Transportation Research Record*, , No. 1051, 1986, pp. 42–48.

[7] Marshall, L. F., H. S. Levinson, L. C. Lennon, and J. Cheng, Bus service times and capacities in Manhattan. *Transportation Research Record*, , No. 1266, 1990.

[8] Lin, T.-m. and N. H. Wilson, Dwell time relationships for light rail systems. *Transportation Research Record*, , No. 1361, 1992.

[9] Rajbhandari, R., S. I. Chien, and J. R. Daniel, Estimation of bus dwell times with automatic passenger counter information. *Transportation Research Record*, Vol. 1841, 2003, pp. 120–127.

[10] Dueker, K. J., T. J. Kimpel, J. G. Strathman, and S. Callas, Determinants of bus dwell time. *Journal of Public Transportation*, Vol. 7, No. 1, 2004, pp. 21–40.

[11] Rashidi, S., P. Ranjitkar, and Y. Hadas, Modeling bus dwell time with decision tree-based methods. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2418, 2014, pp. 74–83.

[12] Wang, C., Z. Ye, Y. Wang, and Y. Xu, Modeling bus dwell time and time lost serving stop in China. *Journal of Public Transportation*, Vol. 19, No. 3, 2016, pp. 55–77.

[13] Tirachini, A., Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers. *Transportmetrica A: Transport Science*, Vol. 9, No. 1, 2013, pp. 28–49.

[14] Meng, Q. and X. Qu, Bus dwell time estimation at bus bays: A probabilistic approach. *Transportation Research Part C: Emerging Technologies*, Vol. 36, 2013, pp. 61–71.

[15] Rashidi, S. and P. Ranjitkar, Estimation of bus dwell time using univariate time series models. *Journal of Advanced Transportation*, Vol. 49, No. 1, 2015, pp. 139–152.

[16] Shalaby, A. and A. Farhan, Prediction model of bus arrival and departure times using AVL and APC data. *Journal of Public Transportation*, Vol. 7, No. 1, 2004, pp. 41–60.

[17] Zhang, C. and J. Teng, Bus dwell time estimation and prediction: A study case in Shanghai-China. *Procedia-Social and Behavioral Sciences*, Vol. 96, 2013, pp. 1329–1340.

[18] Evans, H. and G. Skiles, Improving public transit through bus preemption of traffic signals. *Traffic Quarterly*, Vol. 24, No. 4, 1970.

[19] Lin, Y., X. Yang, N. Zou, and M. Franz, Transit signal priority control at signalized intersections: A comprehensive review. *Transportation Letters: The International Journal of Transportation Research*, Vol. 7, No. 3, 2015, pp. 168–180.

[20] Eichler, M. and C. F. Daganzo, Bus lanes with intermittent priority: Strategy formulae and an evaluation. *Transportation Research Part B: Methodological*, Vol. 40, No. 9, 2006, pp. 731–744.
[21] Zhou, G. and A. Gan, Performance of transit signal priority with queue jumper lanes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1925, 2005, pp. 265–271.
[22] Xu, H. and M. Zheng, Impact of bus-only lane location on the development and performance of the logic rule-based bus rapid transit signal priority. *Journal of Transportation Engineering*, Vol. 138, No. 3, 2011, pp. 293–314.
[23] Zhou, G., A. Gan, and X. Zhu, Determination of optimal detector location for transit signal priority with queue jumper lanes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1978, 2006, pp. 123–129.
[24] Liu, H., A. Skabardonis, W.-b. Zhang, and M. Li, Optimal detector location for bus signal priority. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1867, 2004, pp. 144–150.
[25] Skabardonis, A., Control strategies for transit priority. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1727, 2000, pp. 20–26.
[26] Smith, H. R., B. Hemily, and M. Ivanovic, *Transit signal priority (TSP): A planning and implementation handbook*. ITS America, 2005.
[27] Dion, F., H. Rakha, and Y. Zhang, Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial. *Journal of transportation engineering*, Vol. 130, No. 3, 2004, pp. 294–303.
[28] Chang, G.-L., M. Vasudevan, and C.-C. Su, Bus-preemption under adaptive signal control environments. *Transportation Research Record*, , No. 1494, 1995, pp. 146–154.
[29] America, I., A. T. M. S. Committee, et al., *An overview of transit signal priority, final draft*. ITS America, 2002.
[30] Sunkari, S. R., P. S. Beasley, T. Urbanik, and D. B. Fambro, Model to evaluate the impacts of bus priority on signalized intersections. *Transportation Research Record*, , No. 1494, 1995, pp. 117–123.
[31] Furth, P. G. and T. H. J. Muller, Conditional bus priority at signalized intersections: Better service with less traffic disruption. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1731, 2000, pp. 23–30.
[32] Hu, J., B. B. Park, and Y.-J. Lee, Coordinated transit signal priority supporting transit progression under connected vehicle technology. *Transportation Research Part C: Emerging Technologies*, Vol. 55, 2015, pp. 393–408.
[33] Hu, J., B. B. Park, and Y.-J. Lee, Transit signal priority accommodating conflicting requests under connected vehicles technology. *Transportation Research Part C: Emerging Technologies*, Vol. 69, 2016, pp. 173–192.
[34] Ding, J., M. Yang, W. Wang, C. Xu, and Y. Bao, Strategy for multiobjective transit signal priority with prediction of bus dwell time at stops. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2488, 2015, pp. 10–19.

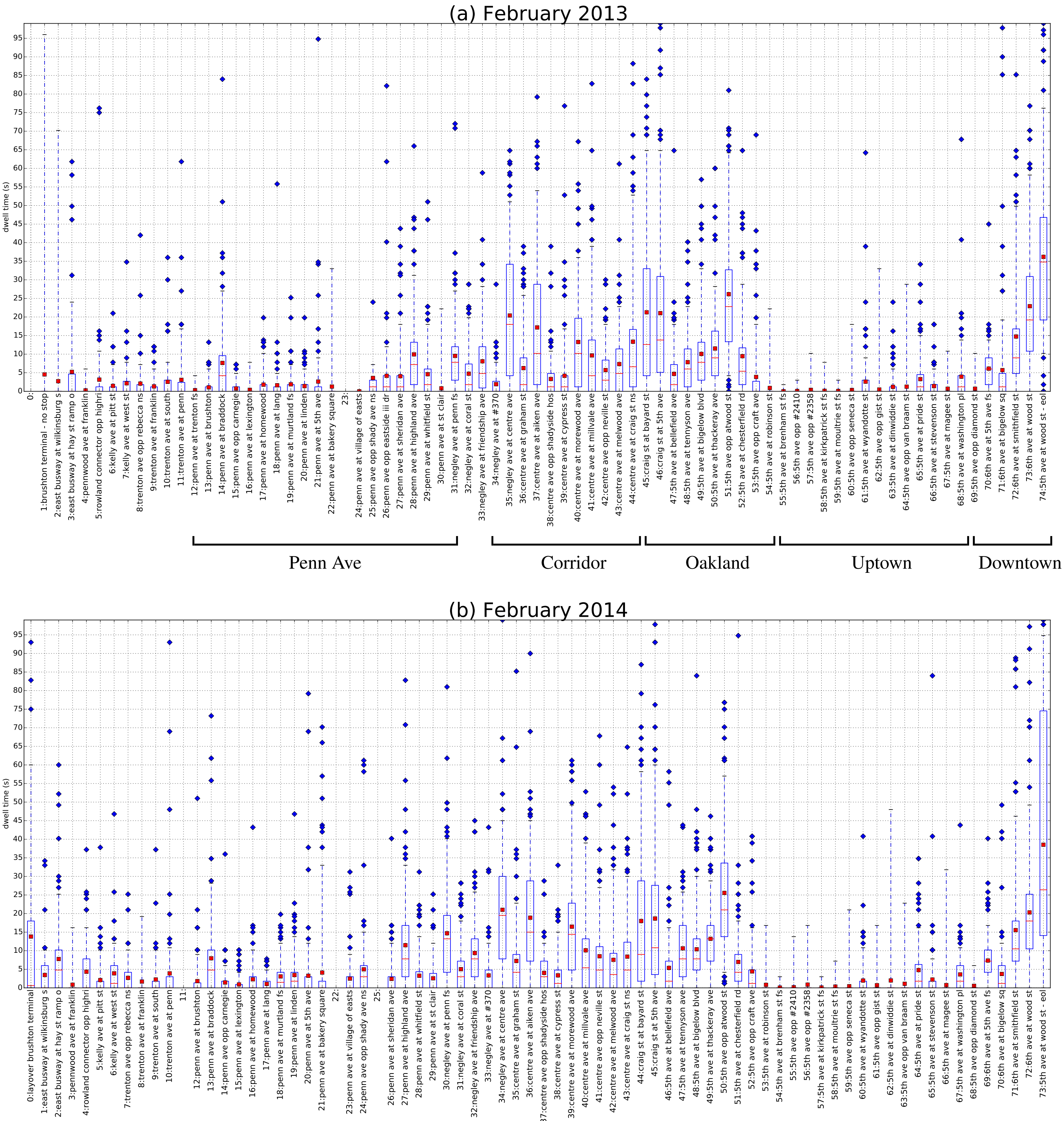


FIGURE 2 Dwell times (less than 100 s), February 2013 (a) vs. February 2014 (b), route 71C inbound, 7:00 – 10:00 AM.

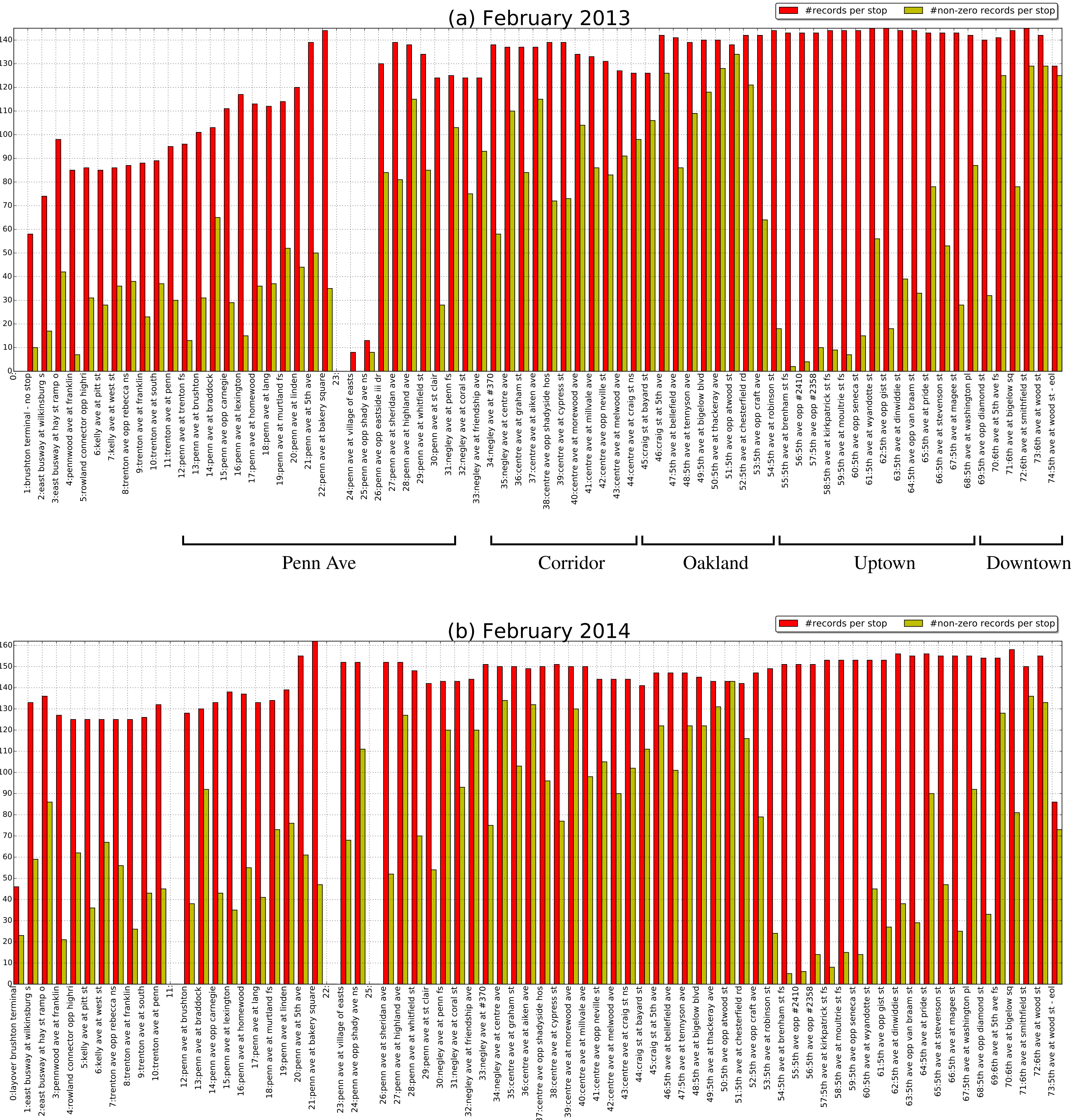


FIGURE 3 Number of records and number of non-zero records (for dwell times less than 100 s), February 2013 (a) vs. February 2014 (b), route 71C inbound, 7:00 – 10:00 AM.

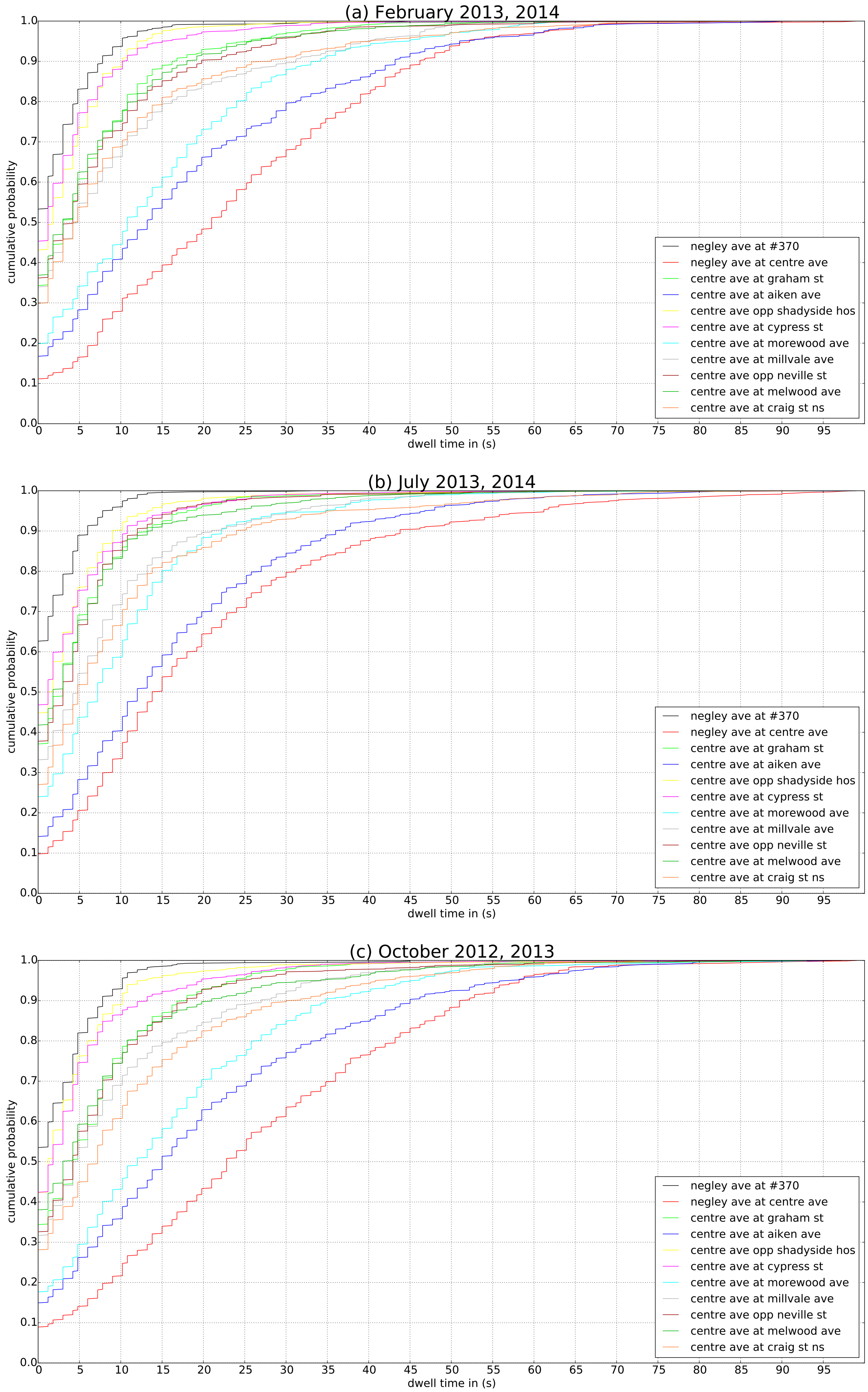
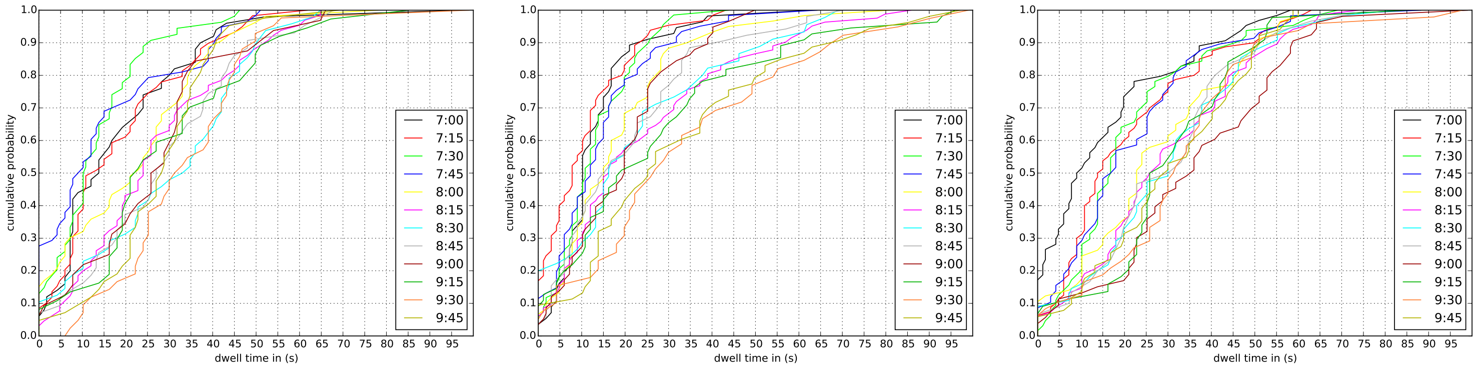
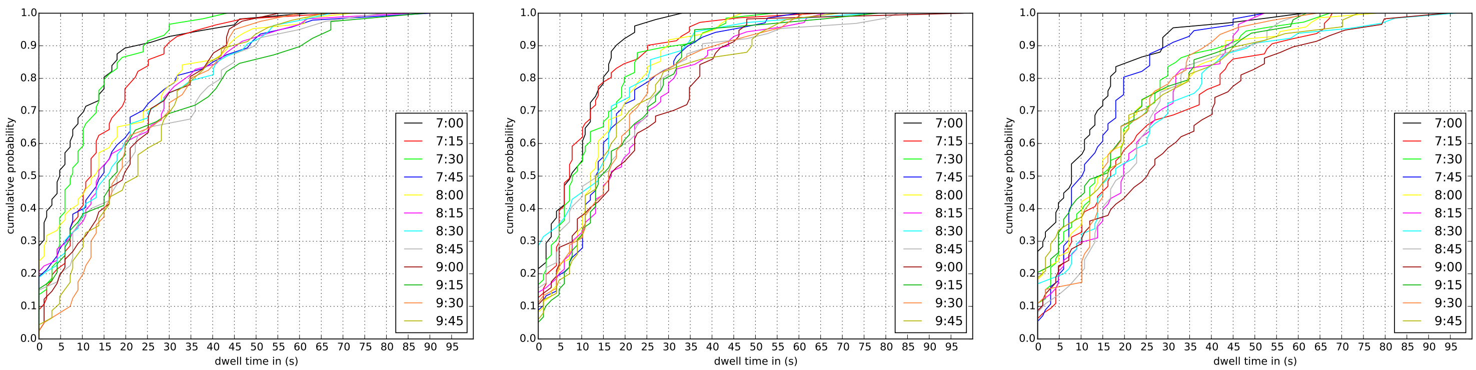


FIGURE 4 Cumulative distribution function for dwell times (less than 100 s) for the corridor, routes 71A+71C inbound, 7:00 – 10:00 am, February 2013+2014 (a), July 2013+2014 (b), October 2012+2013 (c).

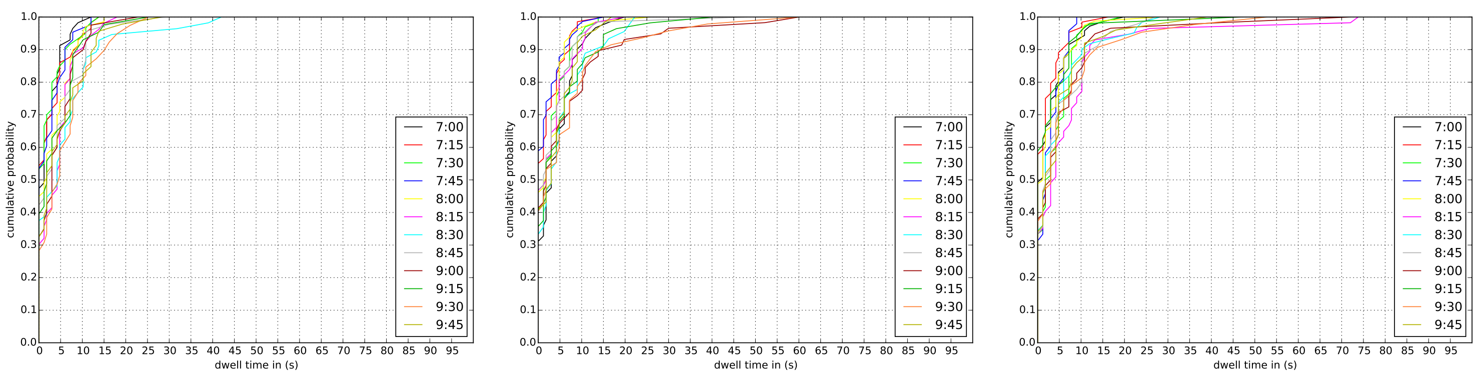
February 2013, 2014 July 2013, 2014 October 2012, 2013



(a) Negley Ave at Centre Ave



(b) Centre Ave at Aiken Ave



(c) Centre Ave opp Shadyside hospital

FIGURE 5 Cumulative density function for dwell times (less than 100 s), routes 71A+71C inbound, 7:00 – 10:00 am, left column: February 2013+2014, middle column: July 2013+2014, right column: October 2012+2013, (a) top row: bus stop Negley Ave at Centre Ave, (b) middle row: bus stop Centre Ave at Aiken Ave, (c) bottom row: bus stop Centre Ave opp Shadyside hospital.

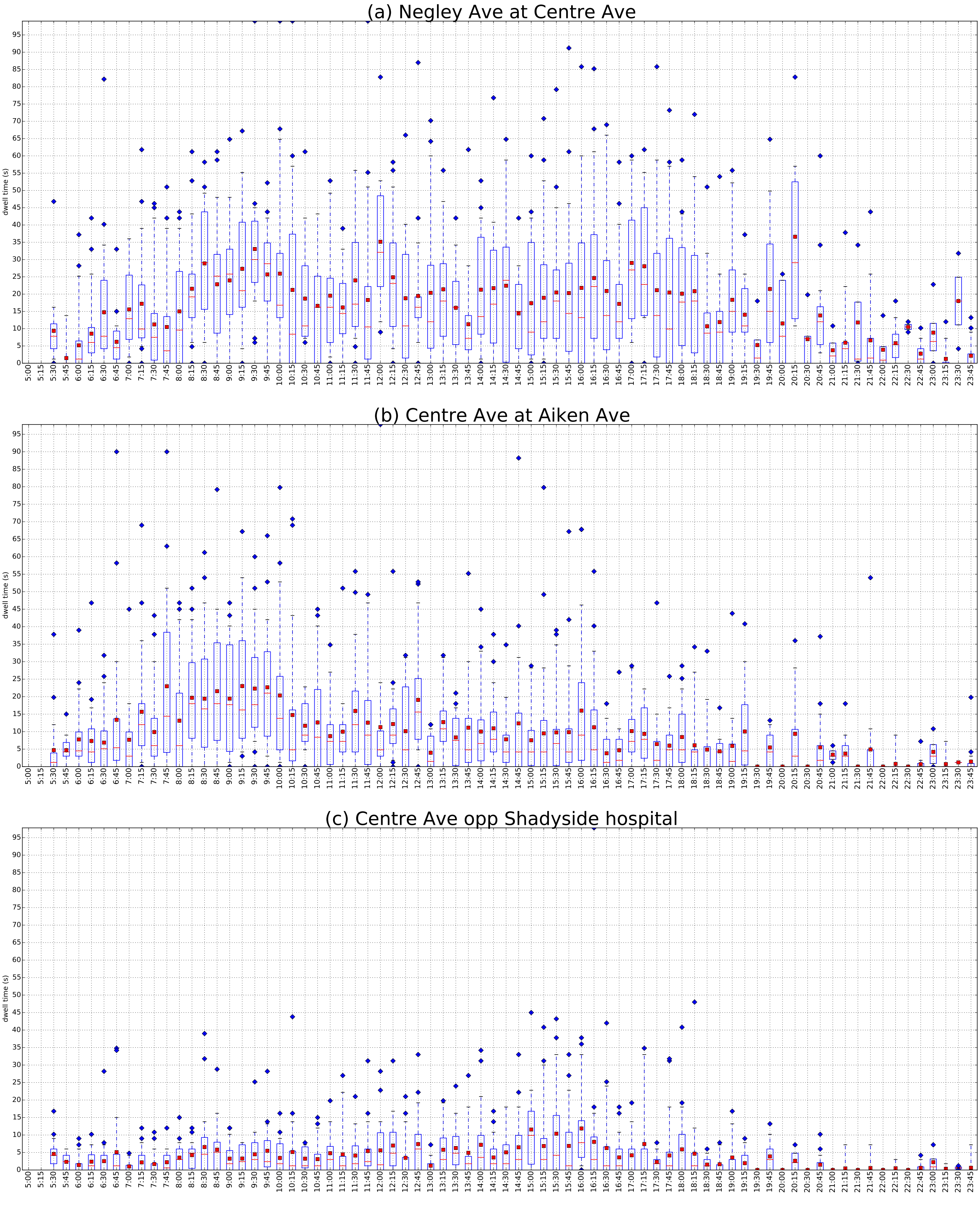


FIGURE 6 Dwell times (less than 100 s), February 2013+2014, routes 71A+71C inbound.