

Evaluation of Bus Dwelling Patterns Using Bus GPS Data

İlgin Gökaşar

Department of Civil Engineering
Boğaziçi University
İstanbul, Turkey
ilgin.gokasar@boun.edu.tr

Yiğit Çetinel

Department of Civil Engineering
Boğaziçi University
İstanbul, Turkey
yigit.cetinel@boun.edu.tr

Abstract— Most of major cities have extensive bus network systems in congested areas. Despite the important contribution of buses on reducing the number of vehicles in the traffic, interaction of buses with the other vehicles may cause some deterioration in the network. Long bus dwelling times can be listed as one of the most crucial ones, since bus dwellings on bus stops have negative impact on traffic congestion, when there is no separated lane for public transportation. Some of the reasons that cause long dwelling times on the bus stops can be listed as high public transit demand, long boarding and alighting time for congested buses and geometric features of bus stop area. In this study, the effects of bus stop characteristics and attributes on bus dwelling patterns will be investigated. Similar dwelling patterns are clustered by taking the properties of the area around bus stops into consideration. It is crucial to spot the problematic bus stops and to find the solutions to eliminate their negative impact on road network in congested cities like Istanbul where the commuting time doubles during peak hours. GPS trajectory data of the bus fleet will be analyzed in order to find dwelling patterns and to observe such similarities or differences, and correlations between dwelling times and station density obtained from public transportation smart card daily data on the selected network. Any possible improvement for the network will be reported and transferability to other cities will be discussed briefly.

Keywords— *Bus trajectory; GPS; bus dwell pattern recognition*

I. INTRODUCTION

Public busses as a mode of transportation has a high importance for most of major cities. To operate this system in its most efficient way, necessary infrastructure must be available in the transportation network of a city, such as appropriate bus stops, bus lanes, bus bays, and so on. These improvements are necessary in order to minimize the effects of the public transportation vehicles on the overall traffic.

Dwell time at bus stops is one of the critical components of bus travel time. It can have a negative impact on the overall traffic, effecting the other vehicles' travel times as well. Number of boarding and alighting passengers is highly correlated with the dwell time and because of this fact Min-Tang Li et al. [1] had developed a simulation model to estimate the bus dwell time by considering the number of disembarking and boarding passengers. Milkovits [2] has modelled the other factors that are affecting the bus stop dwell time, using the automatic passenger

and fare and vehicle location data. With this model the secondary effects, such as crowding, fare type and bus design, have been inspected in order to determine if these factors have a significant relationship with the dwell time.

With each additional bus stop on the roads, the nearby locations experience adverse effects. Koshy and Arasan [3] has conducted a research for the influence of bus stops on the mixed traffic by constructing a model. They have determined that the curbside bus stops that reduces 25% of the general traffic speed must be converted to a bus stop with bus bays. The result actually shows the potential harm of bus stops on traffic and provides a valid solution. These impacts on traffic not only effects the other non-public transport vehicles, but also other busses as well. The dwell times also have an impact on overall travel time, which reduces the quality level of service of the public busses [4]. Turnquist [5] in their research has proposed two alternative solutions to reduce the deviation from schedule due to stop dwell time; to increase the stop spacing by eliminating some stops along a route and zone scheduling.

In order to analyze the effects, there are many methods to collect data from the busses. The most popular method is to use the GPS data provided from the special equipment installed in the busses. This data is used for daily operation purposes such as looking up where the bus is currently at, as well as to measure the level of service of the public bus transportation system. Lin, and Zeng [6] used the GPS data to study the real-time bus arrival time prediction used for the level of services. Pinelli et al. [7] has also used GPS data to construct a robust bus-stop identification methodology. Stoll et al. [8] has used GPS data to visualize and identify congestion hot spots in urban arterials.

Although utilizing GPS data is very common; however, it can be challenging to work with. Csiszár and Sándor [9] has studied a method for analyzing and predicting dwell times at stops in a local bus network. It was noted in this study that creating a database based on vehicle tracking system and separating move and dwell phases was the research's most challenging task. It is also important to know how to maintain and manage the large quantities of GPS data. Stopher et al. [10] developed methods to manipulate data collected from GPS devices carried by the travelers and produced records.

In this research, GPS data of the public busses were analyzed in order to find dwelling patterns and to find correlations between the characteristics of bus stops and the dwelling patterns. With the results it is aimed to determine the problematic stations and find out the reason behind it. By determining these stations and reporting any possible improvement, the negative impact of these bus stops on the overall traffic is expected to be decreased.

II. METHODOLOGY

GPS trajectory data of buses in service during April 2016 is obtained from Istanbul Electric Tram and Tunnel Company (IETT) in comma separated vector (CSV) format. Data consist of time stamp, latitude, longitude values and bus door number of the selected bus route shown in Fig. 1.

The selected bus route (30A) consists of a variety of traffic density and the number of lanes. It is a ring route between two densely populated areas, namely Mecidiyeköy and Beşiktaş, passing through a network, consisting of 3-lane arterial, 1- and 2-lane local roads.

The selected bus route data is analyzed using R-Stats software. 421 data files of single daily bus logs are recorded in 30 days. On that specific route, mostly 14 different buses are operating in a day. Each data log has 5 thousand data points on the average with a total of 2.1 million data points. Data points are collected approximately 4 times in a minute, the median of the sample rate is 0.0625 Hz (16 second).

Distance between geographic (geodetic) coordinates of successive data points is calculated based on an ellipsoid (spheroid) model of the world. WGS84 ellipsoid model is used in the calculation, which is using inverse geodesic function in ‘Geosphere’ package of R. Then, using the time and position differences, average speeds are simply calculated for each data point (Fig 2).

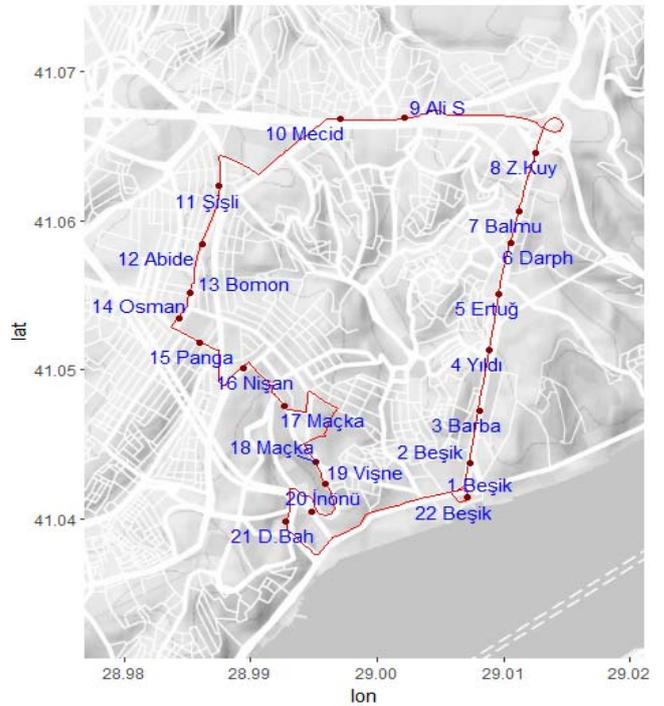


Fig. 1. The selected bus route (Beşiktaş-Mecidiyeköy) and its bus stops

Threshold speed for determining whether a bus is stopping is needed, since the bus can stop without having zero speed value because of the sampling rate. The threshold speed V_S is calculated with simple bus stopping simulation. The bus deceleration and acceleration rates are taken from the literature[11] as 0.19g and 0.15g. In the shortest stopping interval case, it is assumed that bus is stopping only 1 second. In 16 seconds period, which is median sampling period of data, the average bus speed is calculated as 5.9 km/h.

All data points, which have less speed value than threshold speed V_S , are labelled as stopped. With longer stopping cases, average bus speeds will be decreasing. Thus, it is claimed that data with less than V_S gives stopping of the buses. Consecutive stopping point durations are summed up to determine durations of moments when buses stop.

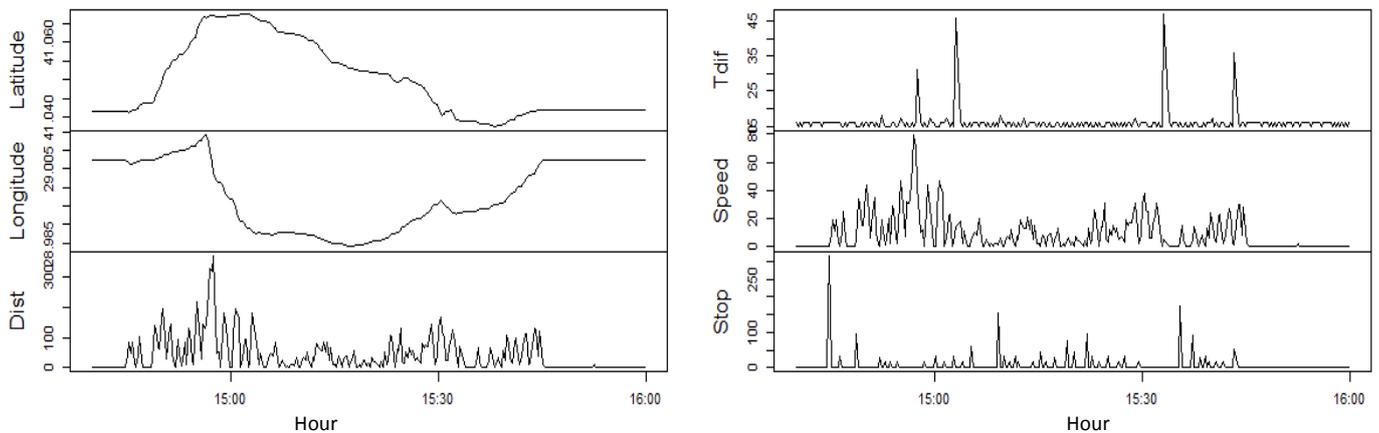


Fig. 2. Sample data information: Latitude, longitude, distance, time difference, speed and stop duration

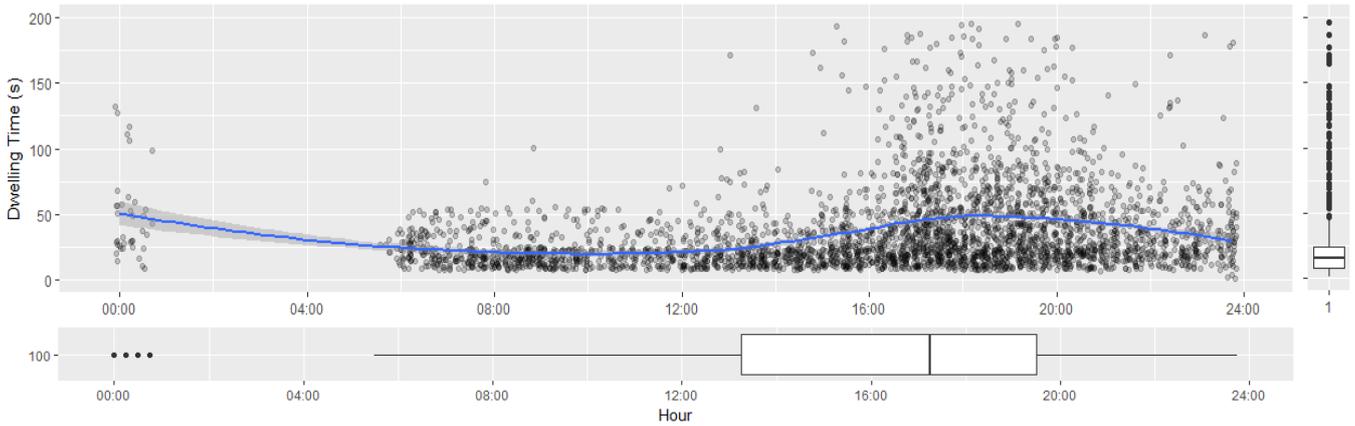


Fig. 3. Sample bus stop dwelling pattern and boxplots

Sample trip of a bus is extracted and plotted in Fig. 2. The trip started at 14:40, ended at 15:50 as seen in ‘Dist’ subplot that shows displacement of bus. ‘Tdif’ (time difference) shows that there are irregularities in the data recording frequency although the most recurrent logging period is 16 seconds. There is a peak at the beginning of the ‘Stop’ subplot because of waiting time before trip start. Excluding the first peak, other peaks show durations of separate stopping moments.

Box plot of the stopping times of the sample bus stops shows density of the data (Fig. 3). The box plot on the x axis presents that the most of the data are between 1:00AM and 7:30PM. It is spotted with the box plot on the y axis that there are outliers after 50 seconds dwell time. Mean of the data is 5PM on time axis and 15 seconds on dwell time axis.

The explained analysis are implied to all bus trips. Then, the stoppings are matched to bus stops to find dwelling patterns. For a demonstration purpose, the third bus stop (Barbaros Bulvarı) is focused on. All dwell times of the stop plotted with record time on x axis in Fig. 3. Bus operation hours starting from 6 AM and there is not much data after midnight. The data is smoothed with LOESS (Local Polynomial Regression Fitting) model. A polynomial surface is fit using local fitting. As a result, there can be seen a lift in dwell time in the evening peak.

All dwelling patterns of bus stops on selected route are plotted starting from 6AM to midnight in Fig. 4. The first plot (a) shows dwell patterns of all bus stops on the selected route. The bus stops are selected as outliers when their dwell time is exceeding one minute at any time (b). The third plot (c) shows regular bus stop patterns except outlier ones.

Most of the stops have increased dwell times in the evening peak starting from 4PM to 7PM, and they also have peak dwell times around 6PM. When some of them have saddle point of dwell times around 1PM, some others reaches their peak points. Although a significant increase of dwell pattern is observed on one bus stop at the end of the day, the rest of them have decreasing pattern around 11PM.

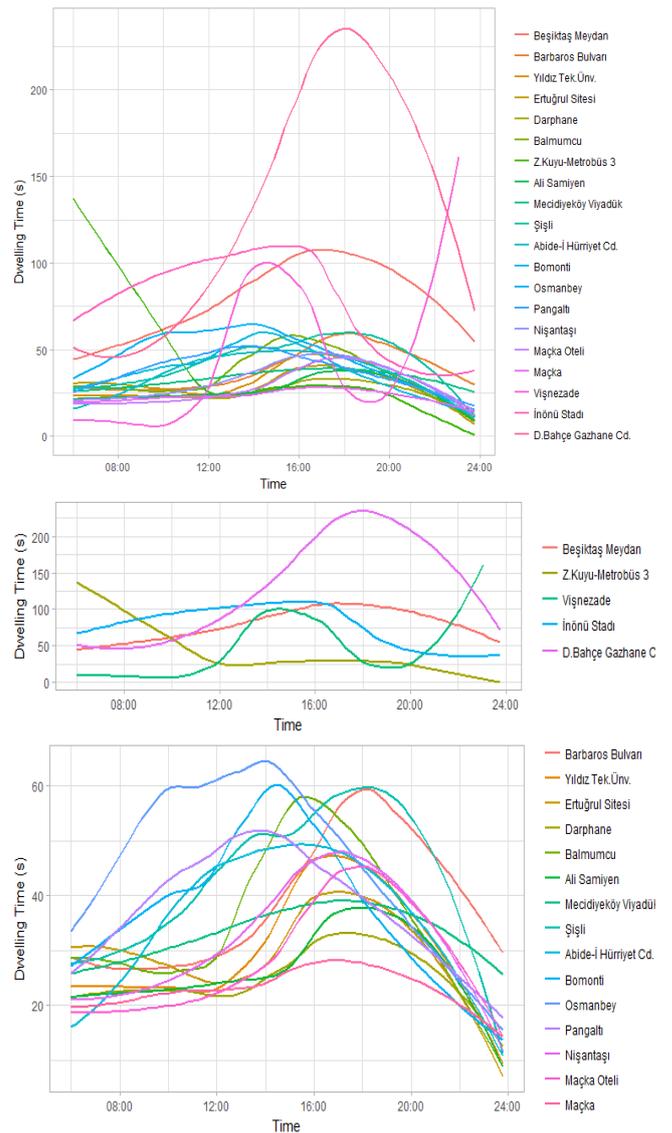


Fig. 4. (a) Dwelling patterns of bus stops on selected route (b) Dwell patterns of outlier bus stops (c) Close up to dwell patterns after extracting outlier bus stops

TABLE I. BUS STOPS INFORMATION

No	Bus Stop Attributes			
	Name	Lane No	Pocket	Ridership ^a
2	Beşiktaş Meydan	3	1	858
3	Barbaros Bulvarı	3	1	695
4	Yıldız Tek. Üniv.	3	1	801
5	Ertuğrul Sitesi	3	0	708
6	Darphane	3	0	690
7	Balmumcu	3	0	729
8	Z.Kuyu-Metrobüs 3	3	1	334
9	Ali Samiyen	1	0	757
10	Mecidiyeköy Viyadük	4	1	868
11	Şişli	2	0	853
12	Abide-İ Hürriyet Cd.	1	0	676
13	Bomonti	1	0	569
14	Osmanbey	2	1	635
15	Pangaltı	1	0	850
16	Nişantaşı	2	1	665
17	Maçka Otelı	1	0	567
18	Maçka	1	0	494
19	Vişnezade	1	0	209
20	İnönü Stadi	3	0	160
21	D.Bahçe Gazhane Cd.	3	1	331

^a. Transportation card use of the selected bus route in April 2016

Dwelling pattern similarities are determined using hierarchical clustering method. Euclidian distances (1) among bus stops' patterns are calculated for each moment. The patterns are clustered using maximum of the calculated distances (2).

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \quad (1)$$

$$\max \{d(a, b) : a \in A, b \in B\}. \quad (2)$$

The dendrogram of the clusters is shown in Fig. 5. When using 7 clusters, all outliers are separated from regular bus stops. While the cluster number 1, 4, 5, 6 and 7 are outlier clusters, each with 1 member, the cluster number 2 and 3 includes bus stops with regular patterns.

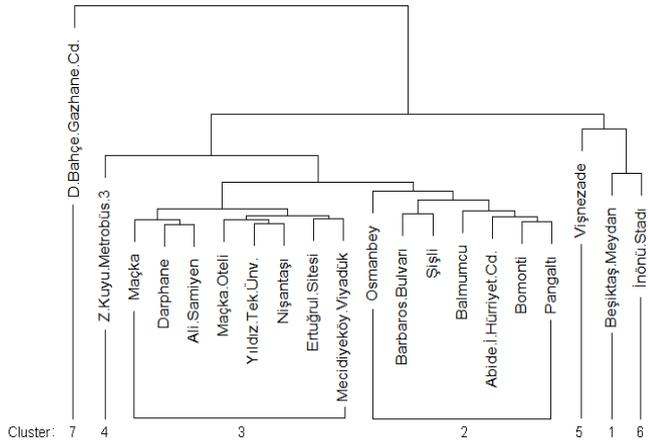


Fig. 5. Bus dwell patterns cluster dendrogram

Bus stops characteristics are shown in Table 1. The stops are sorted according to the route order. Number of lanes where the bus stops locate, whether pockets of the bus stops exist and passenger demands of the bus stops are obtain as bus stops characteristics.

Bus ridership is calculated using transportation smart card data (İstanbulkart), which consists of passengers boarding public transportation. A timestamp, user id and bus stop id are recorded when a passenger uses the card while boarding a bus. It is used to calculate ridership of each bus stop monthly, daily or hourly. Transportation smart card data of April 2016 for the selected route is aggregated to calculate total bus ridership for each bus stop.

All vehicles which are in the selected bus route are almost identical. They have one door for entrance which allows only one person to board and two exit doors which allow two or three people to alight. Thus, only the network information around the bus stops and the ridership are considered as attributes.

The first stop (number 1 and 22) is the terminal station of the selected ring route. The terminal station has long waiting times in order to fit on schedules of the route. Since these waiting times are different from dwelling times, the first stop is eliminated from the analysis.

After a bus finishes a trip, it waits in the terminal until its next schedule. 5-10 mins before the scheduled time, the bus door is opened to allow passengers to board. The time is usually much more than enough for all passengers to board. If the bus driver sees an extraordinary demand in the terminal, he/she opens the door even earlier so that there will be no delay in the schedule.

In the selected route, although majority of the bus stops locate in 3-lane road, 35% of them are in one lane road. 40% of bus stops have separated pocket. More than 13.3 thousand people, and 622 people per bus stop, are using this bus route monthly when including the terminal station (bus stops number 1) which is outside the scope of this analysis. Dwell pattern clusters and characteristic clusters are compared in Fig. 6.

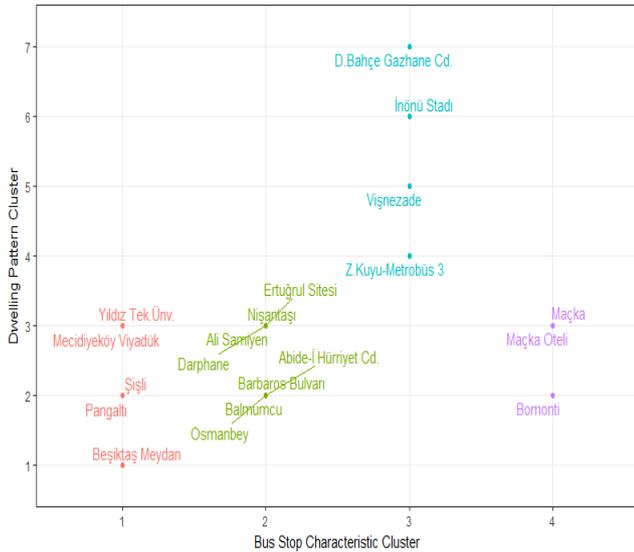


Fig. 6. Bus dwell patterns cluster and characteristic cluster comparison

Bus stops characteristic are clustered using hierarchical clustering method. When using 4 clusters, all outliers are separated from regular bus stops except the first stop and the outliers are clustered together in bus characteristic cluster number 3.

III. CONCLUSION

In the study, it is aimed to find dwelling patterns of the bus stops from the bus trajectory data and correlate the patterns with the bus stop features. Hierarchical clustering with complete linkage method is used to find similar clusters. Two main groups are detected apart from outliers. When the features are clustered and matched with the pattern clusters, most of the outliers are separated into one group.

Since bus stop characteristics are affecting bus dwelling times, the dwell pattern of a new bus stop can be predicted by comparing dwell pattern of the similar bus stops. This can help planners to locate a new bus stop in process of route planning. To avoid the long dwelling times of a bus stop, it can be relocated into an area with similar characteristics of bus stops with normal dwell patterns.

For further studies, analysis can be applied with more features. In addition to hierarchical clustering, new methods can be adapted to cluster dwell patterns and features.

ACKNOWLEDGMENT

This project was supported by the Boğaziçi University Research Fund (BAP) with the project number 11660 and the project code 16A04P2. The authors also would like to thank IETT for providing bus trajectory data greatly assisting the research.

- [1] M.-T. Li, F. Zhao, L.-F. Chow, H. Zhang, and S.-C. Li, "Simulation Model for Estimating Bus Dwell Time by Simultaneously Considering Numbers of Disembarking and Boarding Passengers," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1971, no. 1, pp. 59–65, 2006.
- [2] M. Milkovits, "Modeling the Factors Affecting Bus Stop Dwell Time: Use of Automatic Passenger Counting, Automatic Fare Counting, and Automatic Vehicle Location Data," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2072, no. 1, pp. 125–130, 2008.
- [3] R. Z. Koshy and V. T. Arasan, "Influence of Bus Stops on Flow Characteristics of Mixed Traffic," *J. Transp. Eng.*, vol. 131, no. 8, pp. 640–643, 2005.
- [4] E. Mazloumi, G. Currie, and G. Rose, "Using GPS data to gain insight into public transport travel time variability," *J. Transp. Eng.*, vol. 136, no. July, pp. 623–632, 2009.
- [5] M. A. Turnquist, "Strategies for improving bus transit service reliability," *Transp. Res. Rec.*, 1982.
- [6] W.-H. Lin and J. Zeng, "An Experimental Study on Real Time Bus Arrival Time Prediction With GPS Data," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1666, no. 1, pp. 101–109, 1999.
- [7] F. Pinelli, F. Calabrese, and E. P. Bouillet, "Robust bus-stop identification and denoising methodology," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, no. Itsc, pp. 2298–2303, 2013.
- [8] N. Stoll, T. Glick, and Figlio, "Utilizing High Resolution Bus GPS Data to Visualize and Identify Congestion Hot-spots in Urban Arterials," *Transp. Res. Rec.*, no. 2539, pp. 20–29, 2016.
- [9] C. Csiszár and Z. Sándor, "Method for analysis and prediction of dwell times at stops in local bus transportation," *Transport*, vol. 4142, no. April, pp. 1–12, 2016.
- [10] P. Stopher, Q. Jiang, and C. FitzGerald, "Processing GPS data from travel surveys," *28th Australas. Transp. Res. Forum*, pp. 1–21, 2005.
- [11] M. Kirchner, P. Schubert, and C. T. Haas, "Characterisation of Real-World Bus Acceleration and Deceleration Signals," *J. Signal Inf. Process.*, vol. 2014, no. February, pp. 8–13, 2014.