European Transport \ Trasporti Europei (2023) Issue 94, Paper n° 2, ISSN 1825-3997

https://doi.org/10.48295/ET.2023.94.2

Application of Public Transit AVL Data for Evaluation of Delay Variability

Harsha M. M.[2], Raviraj H. Mulangi[1*], Panditharadhya B. J.[2]

[1*]*Assistant Professor, National Institute of Technology Karnataka, Surathkal, Karnataka – 575025, India*
[2]*Assistant Professor, Siddaganga Institute of Technology, Tumakuru, Karnataka – 572103, India*

Abstract
The travel time is the significant factor in evaluating efficiency and performance of public transit system. A greater percentage of travel time is accounted by bus stop delays which depends on passenger count, bus stop characteristics, traffic condition, bus performance, etc. Many of the Indian transit agencies store the passenger details stage wise not stop wise, which makes it difficult to evaluate delay variability at bus stop level. In this connection, Automatic Vehicle Location (AVL) data from Intelligent Transport System (ITS) implemented at Mysore, India is considered for evaluating bus stop delay variability. The collected data is used for estimating delay at five stops by adopting trajectory-based formulation. The probability distributions have been utilized to model the variability in delay. The performance has been analysed using Kolmogorov-Smirnov (KS) test. The daily variability of delay at bus stops has been evaluated using Coefficient of Variation (COV). The results of the performance evaluation of delay distributions show that the Generalized Extreme Value (GEV) distribution is the best descriptor of the delay variability in terms of accuracy, robustness, and survival capacity. In the absence of passenger data collection systems, method of evaluation of delay using AVL data presented in this study is helpful.

Keywords: Automatic Vehicle Location (AVL), Bus Stop Delay, Delay Variability, Intelligent Transport System (ITS), Public Transit.

1. Introduction

Urbanization refers to the introduction of employment opportunities, modern amenities, well-built infrastructure facilities in a region. Even though there are many advantages of urbanization, the increasing density of population and vehicle ownership in metropolitan cities have given rise to problems such as traffic congestion, environmental pollution, etc. When usage of transportation infrastructure facilities increase beyond the limit to which it is designed, it results in reduced efficiency of traffic movement, causes accidents and environmental impacts. Increase in the number of vehicles is one of the cause behind many of the traffic problems. Even though, the metropolitan cities are facilitated with well-connected public transport systems, people tend to use the private transport to avoid delays. The most sustainable solution for this problem is to make the public transit system more efficient and reliable to attract the commuters. Intelligent Transport System (ITS) is helpful in achieving the goal of influencing the mode choice of commuters towards public transportation.

*Corresponding author: Raviraj H Mulangi (ravirajmh@nitk.edu.in)

The travel time is the significant factor in evaluating efficiency and performance of public transit system. Major proportion of travel time in public transit and its level of service is accounted by bus stop delays which has a huge impact on the mode choice of passengers (Chen et al., 2013; Padmanaban et al., 2009; Rashidi et al., 2014). The most accurate method to compute the delay at bus stops is based on number of passengers boarding and alighting the bus. The automated methods to count the boarding and alighting passengers, such as Automatic Passenger Counting (APC), have not been implemented completely in most of the Indian cities. The adaptation of advanced technologies such as Bus Rapid Transit System (BRTS) and ITS for public transit services in India has created a rich source of data in the form of Automatic Vehicle Location (AVL). In the absence of APC techniques, the AVL data can be used as an alternative for the computation of bus stop delays. AVL data consists of the vehicle coordinates recorded at equal time intervals by the GPS units mounted in the buses.

The time consumed by the passengers during boarding/alighting depends on several factors such as, age, mode of payment, amount of luggage they are carrying, bus stop characteristics and the traffic environment (Currie et al., 2012). These factors make the delay calculation a critical one. Also, the variability in bus stop delay reduces the confidence of commuters in choosing the public transport mode. Thus, it is necessary to limit the delay at bus stops to promote the use of public transport system.

Due to the lack of automated data collection technologies studies on bus stop delays are limited in India. In Mysore city bus transit system, the public bus transit agency stores passenger data, which is not stop wise, but stage wise (In a stage, more than one bus stops are combined on the basis of distance, fare, and other characteristics). Hence, the system is not capable of recording the passenger data at individual bus stops. Also, the previous studies on bus stop delays considered only the dwell time on the basis of passenger count, without taking into account the delays during stopping and restarting (deceleration and acceleration) of the buses. The present study aims to address the research gaps mentioned above using the AVL data from ITS wing of Mysore city public transit system. The total bus stop delay including the delays due to acceleration and deceleration of buses, has been computed using AVL data. The delay variability analysis is carried out with probability distributions using Kolmogorov-Smirnov (KS) test to evaluate their goodness of fit.

2. Literature Review

The technologies such as BRTS and ITS have been adapted to draw the commuter's modal choice towards public transit systems which could be a possible solution to congestion problems. The understanding of the reliability of public transit travel time, delay and the variability associated with them is essential for the implementation of such a system (Harsha et al., 2020). The advanced techniques that provide real time information of bus arrival timings, such as Advanced Public Transport Systems (APTS), contribute to the understanding of public transit travel time reliability. The travel time prediction, inclusive of total running time and delays, works as the basis for such systems. Algorithms have been designed to predict travel times with Kalman filters considering explicitly the dwell time at bus stops, under Indian traffic conditions (Padmanaban et al., 2009). The punctuality of the buses is generally affected by the unstable travel times and bus stop delays that cumulatively form a vicious cycle (Qu et al., 2014). The total travel time of a bus moving along a route is the combination of driving time and dwell time (Dorbritz et al., 2009; Meng and Qu, 2013). The driving time is nothing but the running time of the bus from one stop to other and dwell time is the boarding/alighting time of the passengers which includes the time taken for opening and closing of doors at bus stops. The efficiency of bus reduces as the factors affecting the delays increase (Chen et al., 2013).

The manual collection of dwell time data is tedious and laborious (Dueker et al., 2004). Instead, the data recorded by systems such as Automatic Passenger Count (APC) at stop level or AVL data of buses can be used for delay analysis. The stop wise APC data consists of number of passengers alighting/boarding at every bus stop as well as the arrival and departure times at the level of time-points. Time-points are those stops at which the departures are matching with the scheduled timings (Wong and Khani, 2018). The AVL data recorded by the GPS units in the buses are useful in the analysis of travel time (Li et al., 2018), bus stop delays (Wang et al., 2016), and control delay at junctions (Ko et al., 2008).

Many of the previous studies discussed the usefulness of AVL/APC data in bus dwell time analysis. APC data has been used by Arhin et al. (2016) in which stop time at bus stops has been modeled using multivariate regression. The dwell time model developed by Dueker et al. (2004) using APC data, has been applied to simulate dwell times for different conditions (time periods of day, route characteristics, and number of passengers). Both AVL and APC data have been used to develop multilinear models to estimate the boarding/alighting times of passengers at bus stops in a study by AlHadidi and Rakha (2019). The aforementioned studies were conducted in other countries where traffic conditions are different than that in India. The APC systems have not been implemented yet in everywhere in India, instead evaluation of bus stop delay can be done using AVL data. For linking delay with travel time, AVL data with a resolution of 30 seconds is considered (Wang et al., 2016). Free flow, congestion, stoppage at intersections and bus stops have been considered as the components of travel time.

The bus stop delay variability has been analysed using statistical models (Khoo, 2013). An analytical model was presented for estimating delay variability of intersections controlled by fixed time traffic signal (Chen et al., 2017). Many researchers have worked on variability analysis of delay using probability distributions such as Generalized Extreme Value (GEV) (Esfeh et al., 2020), Normal (Bunker, 2022; Olszewski, 1994), Log normal (Buchel and Corman, 2022), Logistic (Wong and Khani, 2018), Gamma (Khoo, 2013), Uniform (Wong and Khani, 2018), Exponential (Bergstrom and Kruger, 2013; Khoo, 2013), Log-logistic (Li et al., 2020), Weibull (Khoo, 2013), Burr (Susilawati and Yatmar, 2020) and Erlang distributions (Khoo, 2013) have been used to model bus stop and intersection delays.

In the current Indian transport infrastructure, the APC technologies are not implemented in most of the cities and data regarding boarding and alighting of passengers at bus stops is not available. Hence, the only alternative is the AVL data of public transit buses. In this connection, Automatic Vehicle Location (AVL) data from Intelligent Transport System (ITS) implemented at Mysore, India is considered for evaluating bus stop delay variability. The variability in delay has been analysed using probability distributions. Also, the performance has been analysed using KS test with respect to robustness and accuracy of each distribution.

3. Data Interpretation

The Mysore city in India has been chosen as the study area for the present study. The ITS infrastructure was implemented in Mysore during 2012, to improve the efficiency of public transit system. The Mysore City Transport Division (MCTD) operates around 500 buses equipped with GPS devices serving different routes of the city network. The GPS devices record the information at a frequency of 10 seconds. The movement of buses is monitored by MCTD. GPS device records a lot of information among which, trip number, bus identification number, date of the schedule, coordinates of the bus and time of recording are the parameters of interest in this study.

For the present study, the data from the bus route connecting Mysore City Bus Stand (CBS) and Yelwala (route length – 18.46 km) has been used. The data pertaining to May and June 2018 (two months) is extracted for the selected route with the help of QGIS software. The average journey time and frequency of buses are 45 minutes and 8 minutes respectively. This route consists of 23 bus stops, out of which five bus stops have been considered for the analysis (i.e., Belawadi, BM Hospital, Yelwala Government Hospital, Daba and SRS Hootagalli) in Yelwala to CBS direction, as shown in Figure 1. The bus stops are selected in such a way that, they are away from the major junctions and signalized intersections which can affect the speed of the buses.

Figure 1: Bus stop locations (Source: Open Street maps).

4. Methodology

The step wise methodology adopted for the bus stop delay analysis in this study is presented in Figure 2. The AVL data from ITS - Mysore has been processed to remove the outliers and unreasonable values beforehand. Using the trajectories of buses, the components of bus stop delay (deceleration delay, stopped delay, acceleration delay) have been computed. Further, Coefficient of Variation (COV) has been calculated to evaluate the variation in delay. The probability distributions have been utilized to model variability in delay and their performance is measured in terms of robustness and accuracy using KS test.

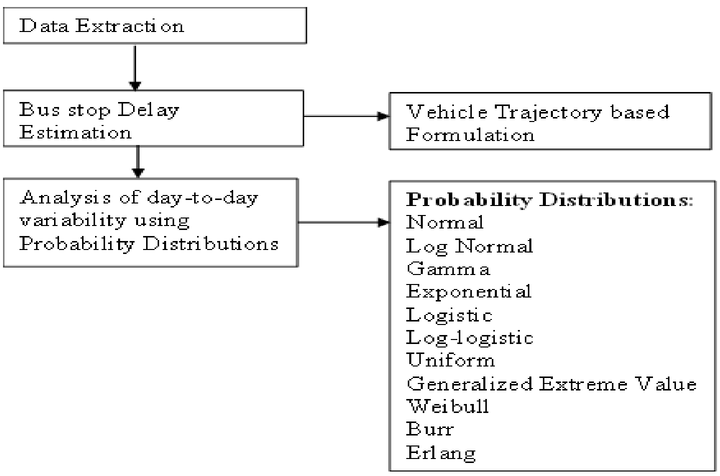

Figure 2: Framework of methodology.

4.1 Data Extraction

The data in Sequential Query Language (SQL) format have been collected from Mysore ITS for the estimation of delay and variability analysis. The collected data is imported to a local server using SQL. This data consists of errors, incomplete information and needs pre-processing before using for the application. During pre-processing, the outliers have been removed and only the data with complete information (unique time stamps, trip IDs) have been retained for further analysis.

The pre-processed data is visualized with QGIS using map matching technique. The data pertaining to the chosen bus stops up to a length of 700 m on both the sides are extracted. There are approximately 2000 trips obtained for each bus stop. Every trip was analysed with trajectory, speed, and acceleration of the bus. The speed profiles have been obtained using the instantaneous speed values recorded by GPS. The acceleration values at every data point are computed using Equation 1 representing backward difference formula,

$$a_i = \frac{V_i - V_{i-1}}{t_1 - t_{i-1}} \tag{1}$$

Where, a_i = acceleration at ith data point; v_i= speed at ith data point; v_{i-1} = speed at i-1th data point; and t_i= time stamps at ith data point; and t_{i-1} = time stamps at i-1th data point.

As per Equation 1, the reduction of speed creates the dip in the acceleration profile and the magnitude of fall in the acceleration profile depends on the amount of reduction in speed of ith point with respect to the i-1th point (As presented in Figure 4 and 5). The reduction in speed of bus is an indication of the bus experiencing delay at bus stop. The trips in which there is no speed reduction (speed values not lesser than 5 m/s) near the bus stop are not used for the delay analysis.

4.2 Bus stop delay components and their computation

The bus stop delay is composed of deceleration, stopped and acceleration delays (Figure 3). The methodology to estimate bus stop delay is based on bus trajectories. The components of bus stop delay have been computed using Equation 2, 3 and 4. Further, they are combined to calculate the total delay at the bus stop (Equation 5). The formulation for direct computation of bus stop delay is given by Equation 6.

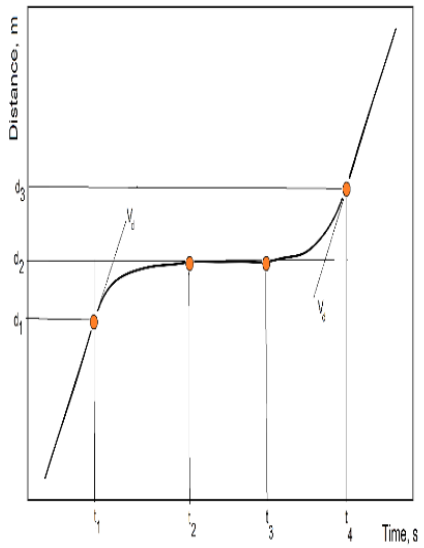

Figure 3: Sample distance profile of a bus in bus stop.

$$Deceleration\ delay\ = (t_2 - t_1) - \frac{d_2 - d_1}{v_d} \tag{2}$$

$$Stopped\ delay = t_3 - t_2 \tag{3}$$

For trajectories without stoppage (Figure 3), t_3 will be equal to t_2.

$$Acceleration\ delay\ = (t_4 - t_3) - \frac{d_3 - d_2}{v_d} \tag{4}$$

Bus stop delay = deceleration delay + stopped delay + acceleration delay (5)

$$Bus\ stop\ delay = (t_4 - t_1) - \frac{d_3 - d_1}{v_d} \tag{6}$$

Where, t_1, t_2, t_3 and t_4 are known as the critical points: t_1 = time stamp when deceleration of the bus begins; t_2 = time stamp at which bus stops completely; t_3 = time stamp when the acceleration of bus begins; t_4 = time stamp when acceleration of the bus stops and it maintains the speed further ; (d_2 - d_1) = distance covered by the bus during deceleration phase before it comes to rest; and (d_3 – d_2) = distance covered by the bus during acceleration phase till it maintains the speed; v_d = desired speed.

The inventory survey conducted in Mysore city during the selection of study area reported that the maximum speed acquired by the buses is 60 kmph. Thus, 60 kmph is considered as the desired speed in the study. Firstly, the estimation of delay using bus trajectory needs the critical points (t_1, t_2, t_3 & t_4) to be identified. The possible locations of critical points have been identified by distance profile as shown in Figure 3. The critical points indicating stopped delay (t_2 and t_3) is described by the significant variation in speed values of buses. Hence, the speed profiles have been plotted and analysed to identify the critical points at the beginning and end of the curve segment where the speed values fall below 5 kmph (Mousa, 2002).

The buses with such segments in their speed profile with speed values lesser than 5 kmph near bus stop are considered as stopped. To identify the critical points t_1 and t_4 (the beginning of deceleration and end of acceleration respectively), acceleration profiles of the buses have been used.

Figure 4 represents speed and acceleration profiles of a bus with stopped delay. The speed and acceleration profiles of a bus without experiencing the stopped delay are shown in Figure 5.

The reason behind the absence of stopped delay can be due to the tendency of drivers not to stop the bus completely in case of no passengers at the bus stop. But the deceleration phase is commonly observed before making the decision to stop at the bus stop. Thus, both the cases with and without stopped delay have been considered for the analysis.

The acceleration profile shows that during the deceleration phase, the positive acceleration values reduce to zero and then reach negative values. The acceleration values then start increasing from negative to positive and reach zero or slightly negative value once the bus reaches the desired speed, which indicates the acceleration phase. The point where acceleration profile reaches zero at the beginning of the deceleration phase is identified as t_1 and that point indicating zero acceleration at the end of the acceleration phase is identified as t_4. The critical point t_1 is the last non-negative point during deceleration phase and t_4 is the last non-positive value during acceleration phase respectively. The points where the acceleration values change from positive to negative are identified as the critical points.

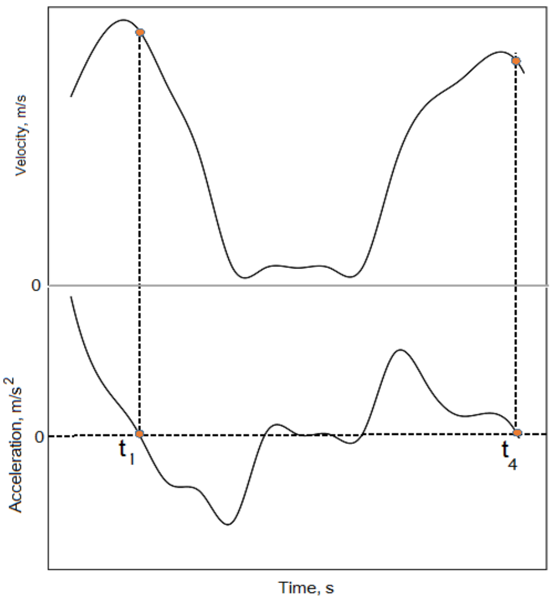

Figure 4: Speed and Acceleration Profile with Stopped Delay.

It is also observed in few of the cases that the vehicle may slowly accelerate without reaching the desired speed during acceleration phase. The changeover of acceleration curve positive to negative is not possible in such cases. Therefore, a threshold value for the distance is used to identify the critical point at the end of acceleration phase. For this purpose, the distance from bus stop after which the buses mostly maintain uniform speed has been analysed from speed profiles. Even though, the bus has a slightly positive acceleration values, if it has reached the threshold distance, the nearest time stamp is considered as the critical point t_4, indicating the end of acceleration phase.

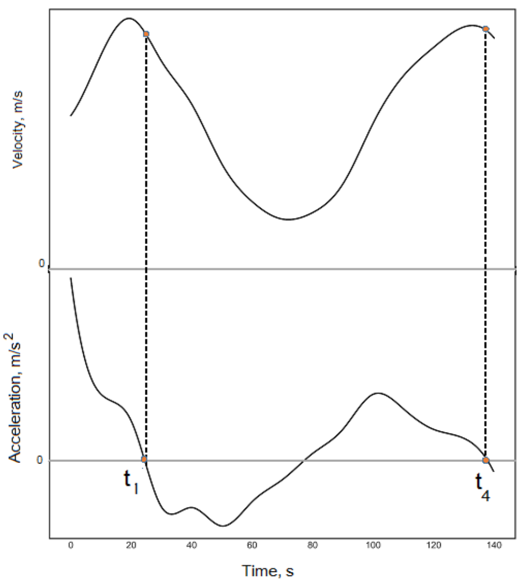

Figure 5: Sample of Speed and Acceleration Profile without Stopped Delay.

4.3 Delay Variability

The delay values of each bus stop are aggregated based on the hourly delay data and the aggregated data has been utilized for the delay variability analysis using descriptive statistics and probability distributions. The selected distributions: Normal, Uniform, Logistic, Burr, Exponential, GEV, Log-logistic, Weibull, Gamma, Log-normal and Erlang, are fitted to the aggregated delay data. The estimation of probability density function for every distribution is done on the basis of Maximum likelihood method. The distribution fit was evaluated using KS test as a measure of goodness of fit on the data. The EasyFit software has been used to fit the distributions and to determine the KS p-value. A distribution fit is considered as significant if p-value is higher than the significance level (i.e., 0.05). For a distribution fit considered, higher the value of KS p-value, better is the fit. The distributions are assigned ranks based on their fitting ability on the delay data. The performance evaluation of distributions is done by considering robustness and accuracy (Ma et al., 2016). The ability of a distribution to fit the data with minimum error is indicated by the accuracy value computed using the assigned ranks and descriptive statistics of p-values obtained from KS test. The ability of the distribution to fit the data from different conditions with low error is implied by robustness. Robustness is determined based on the number-of-cases wherein the distribution has passed the hypothesis test.

The survivor function has been used in the evaluation of alternative distributions to fit the public transit delay data. The distributions are analysed with the survival probability of p-value, computed using Equation 7.

$$S(KS\ p-value\ ) = 1 - F(KS\ p-value\ ) \tag{7}$$

Where, S (KS p-value) = Survival function
F (KS p-value) = Cumulative density function.

5. Results and Discussion

The methodology described in the previous section has been applied to the data for bus stop delay analysis. The variability in bus stop delay is estimated by computing the COV. The delays at bus stops have been aggregated based on the days of a week for the detailed analysis. The COV of day wise aggregated delays at bus stops have been visualized using heat maps as shown in Figure 6 and 7. The visualization using heat maps provides a two-dimensional notion of the magnitude of the variable plotted. The colour codes have been used to represent the variation of COV values along the different time periods of different days in week. The variation in the intensity of the chosen colour code (red colour in the present study) describes the changes in the COV values of delay.

The heat maps of delay COV for Yelwala and Belawadi stops are shown in Figure 6. Among the bus stops considered in this study, highest COV of 79.73% is observed for Belawadi stop, during 8.00 AM to 9.00 PM of Tuesday and highest average COV value of 35.70 % was also observed at Belawadi bus stop. At Yelwala Government Hospital stop, high COV values are noticed during morning hours, especially from 8.00 AM to 1.00 PM. Since the stop is in residential area, the variability will be higher during morning hours as people commute to work towards CBS. On Sunday and Monday, high COV values can be observed throughout the day. The trend is similar for weekdays and weekends. High COV values were observed during few of the evening time periods. In Belawadi stop, the variability is more during morning hours, especially on Tuesday morning, variability is observed to be higher during 8 AM to 10 AM. On Saturday, high values are observed for the whole day, among which the highest values are during evening hours, after 3.00 PM. On Sunday, COV values are lower for evening hours.

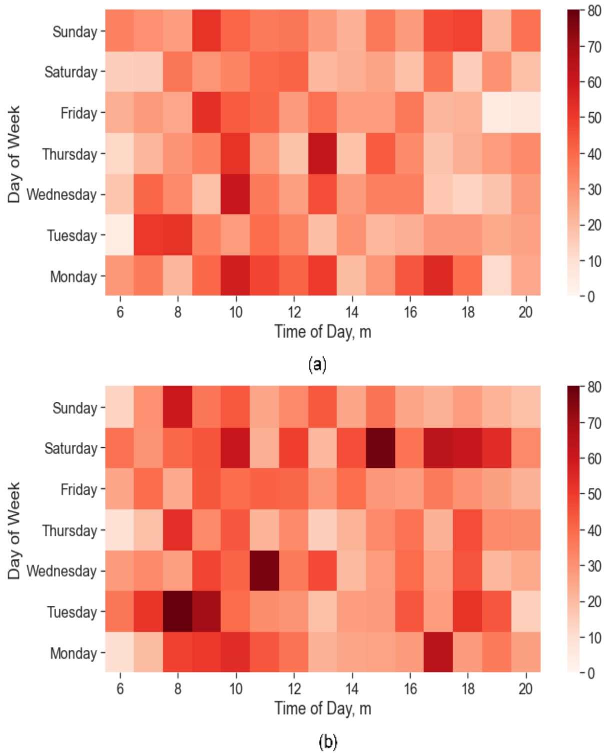

Figure 6: Variation of Delay COV a) Yelwala Government Hospital bus stop b) Belawadi bus stop.

The heat maps of delay COV for Daba, SRS Hootgalli and BM Hospital stops are shown in Figure 7. Low and uniform COV values are observed during all the time periods at BM Hospital bus stop. In Daba stop, COV values are low during most of the time periods. Only a few higher COV values are observed, especially during 9.00 AM to 5.00 PM time. On Saturday, COV values are smaller during the whole day and higher values of COV can be observed from 11.00 AM to 1.00 PM, on Sunday. Reasonable and uniform COV values are observed throughout in SRS Hootagalli stop. Not many variations are found in the variability data of this stop. During weekends, variability has been observed to be higher during afternoon and evening hours, especially after 2.00 PM due to recreational trips.

Among all the stops considered, Belawadi bus stop has been observed to be having a higher delay variability, Daba stop has less delay variability and the delay variability of BM Hospital stop is consistent during all the days. In few of the bus stops, lower delay variability (Dhaba and BM Hospital stop) is found during weekends and SRS Hootgalli and Belawadi stops show a higher variability during weekends. Therefore, weekends and weekdays were considered separately for studying delay distributions.

Among selected probability distributions, the best fit probability distribution is identified. The hourly aggregated delay data from weekdays and weekends has been analysed separately. Figure 8 shows the distribution fit for the sample aggregated delay data, using the chosen probability distributions. Table 1 and 2 present the outcomes of distribution fit for weekday and weekend delay data respectively.

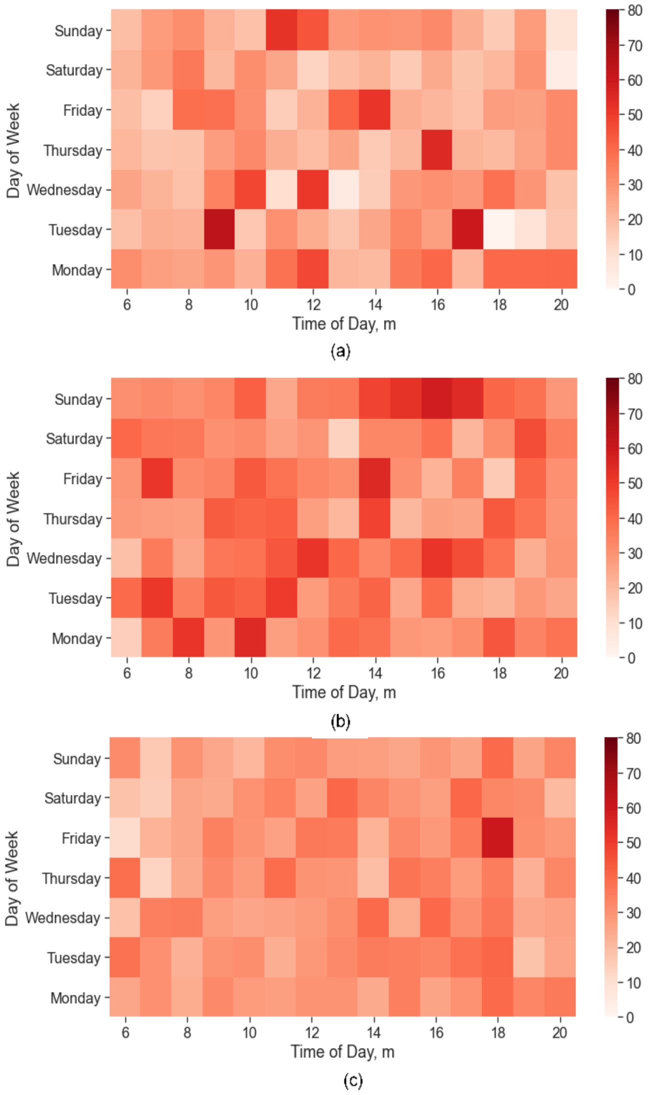

Figure 7: Variation of Delay COV a) Daba bus stop b) SRS Hootgalli bus stop c) BM Hospital bus stop.

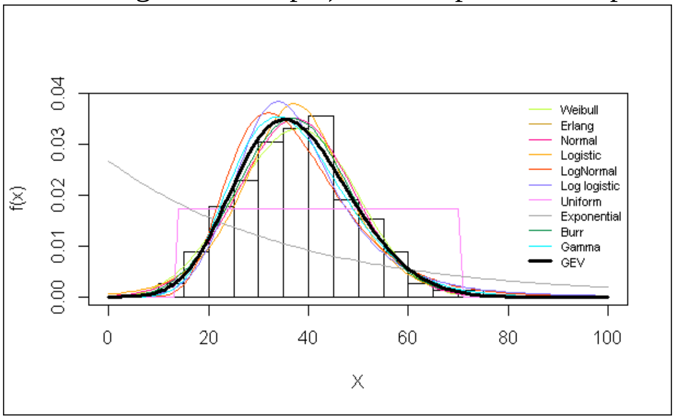

Figure 8: Probability density function of 11 statistical distributions.

Table 1: Summary of K-S test for weekdays.

Distribution	Mean p value	SD of p value	Cases pass ratio (%)	Cases top 3 ratio (%)	Cases top 1 ratio (%)
Burr	0.788	0.241	99%	77%	25%
Erlang	0.339	0.310	71%	3%	1%
Exponential	0.015	0.094	1%	0%	0%
GEV	0.821	0.215	100%	79%	47%
Gamma	0.560	0.319	90%	14%	1%
Log-logistic	0.593	0.292	96%	35%	3%
Log-normal	0.619	0.296	99%	26%	6%
Logistic	0.476	0.299	86%	16%	5%
Normal	0.463	0.309	84%	13%	4%
Uniform	0.339	0.305	78%	9%	5%
Weibull	0.628	0.273	96%	27%	3%

Table 2: Summary of K-S test for weekends.

Distribution	Mean p-value	SD of p-value	Cases pass ratio (%)	Cases top 3 ratio (%)	Cases top 1 ratio (%)
Burr	0.788	0.241	99%	77%	25%
Erlang	0.339	0.310	71%	3%	1%
Exponential	0.015	0.094	1%	0%	0%
GEV	0.821	0.215	100%	79%	47%
Gamma	0.560	0.319	90%	14%	1%
Log-logistic	0.593	0.292	96%	35%	3%
Log-normal	0.619	0.296	99%	26%	6%
Logistic	0.476	0.299	86%	16%	5%
Normal	0.463	0.309	84%	13%	4%
Uniform	0.339	0.305	78%	9%	5%
Weibull	0.628	0.273	96%	27%	3%

GEV distribution has passed 99% of the KS tests and in most of the cases it stands within the top 3 distributions on weekends. GEV has the highest number-of-cases being in top position i.e., 52%, Burr distribution is ranked one in 19% of the cases. Also, Burr distribution has passed the tests 96% of the times and is within the top 3 positions in 78% of the cases. Then, the passing ratios of Log-logistic, Log-normal and Weibull distributions are found to be 90%, 88% and 81% respectively. The passing ratio is found to be the least for Exponential distribution and so are the mean and standard deviation of its p-values. Exponential distribution has no instance of being ranked 1 or within the top 3 ranks. This indicates the Exponential distribution is not useful in explaining the delay variability. GEV distribution is found to have the largest mean p-value among all the distributions fits, i.e., 0.78 and is followed by Burr distribution with mean p-value of 0.722. The standard deviation of p-value is found to be highest in case of Gamma i.e., 0.349 and further it is in the decreasing order for Weibull, Normal, Logistic, Log-normal, and Log-logistic distributions. The distribution with least standard deviation of p-value is GEV distributions, implying its consistency.

In the case of weekends, all the cases are passed by GEV distribution i.e., 100% pass ratio and in 79% of the cases GEV is among the top 3 rankings. Burr distribution too is observed to pass most of the cases i.e., 99% pass ratio. Log-normal distribution and Burr distribution have an equal number of cases passing the KS test. Burr distribution is ranked number 1 in 25% of the cases and in 77% of the cases it is ranked within the top 3 distributions. The distributions in decreasing order of their pass ratios can be listed as: Log-logistic, Weibull, Gamma, Logistic and Normal. In 47% of the cases of tests, GEV distribution is ranked 1. The Exponential distribution is not proved to be useful for delay distribution analysis during weekends also, by having the smallest mean p-value and number-of-cases passing the test.

The largest mean p-value 0.821 is found in GEV distribution and the second highest mean p-value (0.788) in Burr distribution. GEV has highest mean and least standard deviation of p-values, like weekdays. The highest standard deviation of p-values is from Erlang distribution. Normal, Logistic, Log-logistic and Gamma distributions are having standard deviations of p-values in the descending order. The performance of GEV distribution is better compared to other distributions chosen in analyzing bus stop delays.

The distributions have been then subjected to performance evaluation using survivor function (Ma et al., 2016). The p-values have been plotted along x-axis and the corresponding values of survivor function are plotted along y-axis as shown in Figure 9. The probability of a distribution showing better performance on the data analysed for the corresponding p-value is given by the survivor plot. The larger value of survival function indicates higher accuracy of the probability distribution.

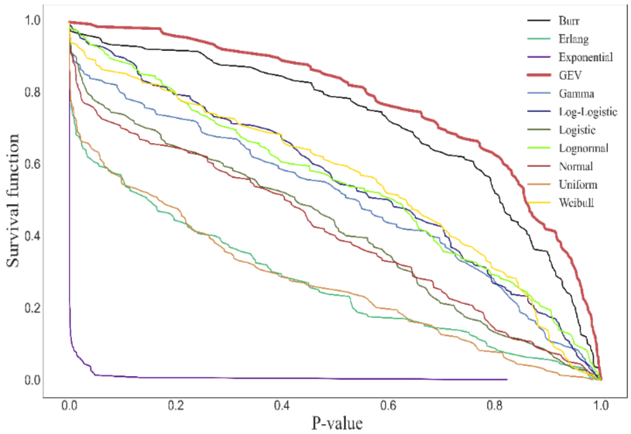

Figure 9: Survival plot of delay distributions.

In Figure 9, GEV distribution is the most accurate distribution having the highest values of survivor function. In this analysis, p-values are more than 0.6 in 80% of cases fitted with GEV distribution. The poor performance of Exponential distribution can be observed from the survival plot of delay distributions. Erlang and Uniform distributions stand next to Exponential distribution in their performance. The other distributions (Logistic, gamma, Log-normal, Weibull and Log-logistic) perform similarly and Weibull distribution's performance is somewhat better. After p-value of 0.3, the performance of Burr and GEV distributions is mostly similar. Burr survivor curve drops by 5 % initially at zero p-value. This drop in the initial stage of the survivor plot indicates the underperformance of Burr distribution being unable to estimate the delay variation accurately, which would affect the application of Burr distribution to delay variability problems under practical circumstances.

6. Conclusions

An attempt has been made in this study to extract different components of delay experienced at bus stops using AVL data of public transit buses from ITS, Mysore. Also, the delay variability has been analysed with the help of probability distributions. The delay data has been aggregated for every hour on different days of the week. The COV of the aggregated delay data has been computed to analyse the variability. The distributions of delay data represent the nature of variation and patterns observed in the variations. Eleven probability distributions have been selected for delay variability analysis. Also, the aggregated delays are analysed separately for weekdays and weekends. The survivor function has been applied to evaluate the distributions with respect to their performance. The key conclusions from the study are listed as follows:

- AVL data of public transit buses is useful in estimating the components of delay at bus stops, deceleration delay, stopped delay and acceleration delay based on the distance; speed and acceleration profile of the buses.
- The variability of bus stop delay in terms of COV with respect to different time periods of different days of week, is visualized using heat maps.
- The distribution of bus stop delay variability is estimated using probability distributions and survivor function is used in evaluating their performance.
- The performance of GEV and Burr distributions in estimating the delay variability distribution has been found to be the best during weekdays as well as weekends.
- Exponential distribution showed a very poor performance in delay variability distribution fit for both weekday and weekend data.
- The survivor plots indicate that the accuracy of GEV and Burr are higher than other distributions. But the dips in the survivor curves of Burr distributions at zero p-values represent the cases of failure in better fitting of delay data.

The bus stop delay analysis carried out in the present study is an initial step towards exploring the delay estimation techniques with the available sources of data. The public transit reliability is dependent on the travel time and its alignment with the scheduled arrival/departure timings. Hence, the estimation of delay and its variability has a potential scope in elevating the public transit reliability. Further, the techniques proposed in present study can be used to predict the travel time accurately.

Conflicts of Interest

The authors declare that they have no conflicts of interest associated with the article or the work in it.

Acknowledgements

The authors are grateful to the Mysore Intelligent Transport System (MITRA) and Karnataka State Road Transport Corporation (KSRTC) for providing the data.

References

AlHadidi, T., Rakha, H.A. (2013) "Modeling bus passenger boarding/alighting times: A stochastic approach", Transportation Research Interdisciplinary Perspectives, 2, 100027. https://doi.org/10.1016/j.trip.2019.100027

Arhin, S., Noel, E., Anderson, M.F., Williams, L., Ribisso, A., Stinson, R. (2016) "Optimization of transit total bus stop time models", Journal of Traffic and Transportation Engineering (English Edition), 3 (2), pp. 146-153. https://doi.org/10.1016/j.jtte.2015.07.001

Bergström, A., Krüger, N. (2013) "Modeling passenger train delay distributions: evidence and implications", Working papers in Transport Economics, CTS - Centre for Transport Studies Stockholm (KTH and VTI), 2013 (3), pp. 1-27.

Büchel, B., Corman, F. (2022). "Modeling conditional dependencies for bus travel time estimation", Physica A: Statistical Mechanics and its Applications, 592, 126764.

Bunker, J. M. (2022) "Probabilistic Modeling of Bus Rapid Transit Station Loading Area Selection for Bus Capacity Estimation", Transportation Research Record, 03611981221082553.

Chen, P., Sun, J., Qi, H. (2017) "Estimation of delay variability at signalized intersections for urban arterial performance evaluation", Journal of Intelligent Transportation Systems, 21 (2), pp. 94-110. https://doi.org/10.1080/15472450.2016.1216319

Chen, S., Zhou, R., Zhou, Y., Mao, B. (2013) "Computation on bus delay at stops in Beijing through statistical analysis", Mathematical Problems in Engineering, 745370, 2013. https://doi.org/10.1155/2013/745370

Currie, G., Delbosc, A., Reynolds, J. (2012) "Modeling dwell time for streetcars in Melbourne, Australia, and Toronto, Canada", Transportation Research Record, 2275 (1), pp. 22-29. https://doi.org/10.3141%2F2275-03

Dorbritz, R., Lüthi, M., Weidmann, U., Nash, A. (2009) "Effects of onboard ticket sales on public transport reliability", Transportation Research Record, 2110 (1), pp. 112-119. https://doi.org/10.3141%2F2110-14

Dueker, K.J., Kimpel, T.J., Strathman, J.G., Callas, S. (2004) "Determinants of bus dwell time", Journal of public transportation, 7 (1), pp. 21-40. http://doi.org/10.5038/2375-0901.7.1.2

Esfeh, M.A., Kattan, L., Lam, W.H., Esfe, R.A., Salari, M. (2020) "Compound generalized extreme value distribution for modeling the effects of monthly and seasonal variation on the extreme travel delays for vulnerability analysis of road network", Transportation research Part C: Emerging technologies, 120, 102808. https://doi.org/10.1016/j.trc.2020.102808

Harsha, M.M., Mulangi, R.H., Kumar, H.D. (2020) "Analysis of Bus Travel Time Variability using Automatic Vehicle Location Data", Transportation Research Procedia, 48, pp. 3283-3298. https://doi.org/10.1016/j.trpro.2020.08.123

Khoo, H.L. (2013) "Statistical modeling of bus dwell time at stops", Journal of the Eastern Asia Society for Transportation Studies, 10, pp. 1489-1500. https://doi.org/10.11175/easts.10.1489

Ko, J., Hunter, M., Guensler, R. (2008) "Measuring control delay components using second-by-second GPS speed data", Journal of Transportation Engineering, 134 (8), pp. 338-346. https://doi.org/10.1061/(ASCE)0733-947X(2008)134:8(338)

Li, J., Zhang, H., Zhang, Y., Zhang, X. (2020) “Modeling Drivers’ Stopping Behaviors during Yellow Intervals at Intersections considering Group Heterogeneity”, Journal of advanced transportation, 2020, 01766140. https://doi.org/10.1155/2020/8818496

Li, Z., Kluger, R., Hu, X., Wu, Y.J., Zhu, X. (2018) “Reconstructing vehicle trajectories to support travel time estimation”, Transportation research record, 2672 (42), pp. 148-158. https://doi.org/10.1177%2F0361198118772956

Ma, Z., Ferreira, L., Mesbah, M., Zhu, S. “Modeling distributions of travel time variability for bus operations”, Journal of Advanced Transportation, 50 (1), pp. 6-24.

Meng, Q., Qu, X. (2013) “Bus dwell time estimation at bus bays: A probabilistic approach”, Transportation Research Part C: Emerging Technologies, 36, pp. 61-71. https://doi.org/10.1016/j.trc.2013.08.007

Mousa, R.M., “Analysis and modeling of measured delays at isolated signalized intersections”, Journal of transportation engineering, 128(4), 2002, pp.347-354. https://doi.org/10.1061/(ASCE)0733-947X(2002)128:4(347)

Olszewski, P.S. (1994) “Modeling probability distribution of delay at signalized intersections”, Journal of advanced transportation, 28 (3), pp. 253-274. https://doi.org/10.1002/atr.5670280306

Padmanaban, R.P.S., Vanajakshi, L., Subramanian, S.C. (2009) “Automated delay identification for bus travel time prediction towards APTS applications”, Second International Conference on Emerging Trends in Engineering and Technology, IEEE, pp. 564-569.

Qu, X., Oh, E., Weng, J., Jin, S. (2014) “Bus travel time reliability analysis: a case study”, Proceedings of the Institution of Civil Engineers-Transport, 167 (3), pp. 178-184. Thomas Telford Ltd. https://doi.org/10.1680/tran.13.00009

Rashidi, S., Ranjitkar, P., Hadas, Y. (2014) “Modeling bus dwell time with decision tree-based methods”, Transportation Research Record, 2418 (1), pp. 74-83. https://doi.org/10.3141%2F2418-09

Susilawati, Ramli, M. I., Yatmar, H. (2020) “Delay distribution estimation at a signalized intersection”, IOP Conference Series: Earth and Environmental Science, 419 (1), 012090.

Wang, H., Zhang, G., Zhang, Z., Wang, Y. (2016) “Estimating control delays at signalised intersections using low-resolution transit bus-based global positioning system data”, IET Intelligent Transport Systems, 10 (2), pp. 73-78.

Wong, E., Khani, A. (2018) “Transit delay estimation using stop-level automated passenger count data”, Journal of Transportation Engineering, Part A: Systems, 144 (3), 04018005. https://doi.org/10.1061/JTEPBS.0000118