DOI: 10.6977/IJoSI.202109_6(5).0005
W. C. Khoo, C. C. T. Cheok,etc/ Int. J. Systematic Innovation, 6(5), 44-54 (2021)

Finite mixture of Burr type XII for bus travel time in Klang Valley

Wooi Chen Khoo1*, Cynthia C. T. Cheok1 and Hooi Ling Khoo2
1 Department of Applied Statistics, School of Mathematical Sciences, Sunway University, Sunway City, 47500 Petaling Jaya, Selangor, Malaysia
2 Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia
* Corresponding author E-mail: wooichenk@sunway.edu.my

(Received 12 July 2021; Final version received 22 July 2021; Accepted 7 September 2021.)

Abstract

Bus travel time analysis is significant to provide useful information to the users for proper journey planning. This study investigated the bus travel time in Klang Valley, Malaysia, which includes the centered city Kuala Lumpur and the state of Selangor. Two bus routes T786 and 851 which cover suburban and urban region have been studied. The data is collected day-to-day. For each link, it is filtered according to morning peak and evening peak, and the travel time is found dual modes, which is well accommodated by the mixture of Burr distribution. The analysis results show that the two-component finite mixture of Burr distribution is viable solution to explain the link travel time. The aim is two-fold. First, the bus travel time data is fitted by the mixture of Burr distribution, subsequently the computation of reliability metrics is carried out. A skew-width method with median based buffer index is considered to measure the reliability. The buffer index of 851 which covers urban region is found relatively high compared to T786. The links which connected by the signalized intersections and junctions tend to show low reliability. For suburban route of T786, the low reliability may be due to the last station. It is found that an additional of 5 minutes waiting time is necessary.

Keywords: Burr, bus operator, mixture distribution, reliability, travel time

1. Introduction

Bus travel time is always a major concern for road users to plan ahead their travel journey. The travel time estimation is usually done intuitively without a proper study or an investigation, most of the users understand the bus arrival time based on experience. Although the travel time is easily known with real-time operating system nowadays, however, the understanding of the travel time pattern from the historical data is significant, as reliable analysis can be a good indicator for road users to gain better understanding on the traffic condition spatially and temporally, especially for the first-time users.

This paper focused on the study for bus travel time in Klang Valley region, which aims to provide a comprehensive analysis on travel time of two bus routes covering Central Business District (CBD) and suburban region. The data is collected day-to-day, from June to December in 2014 and it contains a total of ten bus routes in Klang Valley. For this paper, we examine the travel time of two particular bus routes of T786 and 851 temporally and spatially. T786 mainly covers residential and institution, while 851 passes through mostly governmental offices and commercial district. For each day, the travel time is filtered according to the peak hour; where the morning peak is defined by 7am-10am and the evening peak is defined by 5pm-8pm. For both routes, the travel time pattern of all links is analyzed and it is observed that some links presents dual modes.

Travel time exists in continuous domain. The spread of the data which contribute to pattern construction can be measured by distributional function. The travel time patterns habitually exhibit heavily right-skewed. However, dual peak is either common observed in the pattern because of the daily peak periods. It is of the interest to study the travel time patterns by fitting to some distribution function. It is aimed that more insight is gained upon the fitting, i.e. the skewness can be measured to determine the reliability. An implication is given based upon the analysis results of the reliability. It is found that the reliability of the bus route which covers urban region is relatively high compared to the bus route which covers suburban region. For some links, the waiting time at the particular bus stop could be up to 5 minutes. This study considers fitting the travel time data with a 2-component finite mixture of Burr distribution (MBD), and the favorable results have been obtained. It is believed that none of the relevant study found in using MBD fitting. This study suggests a plausible solution in bus travel time modelling.

Fitting bus travel time data with probability function is feasible for pattern diagnosis. The variability of the travel time gives uncertainty to the road users that they may not know the exact bus arrival time. Therefore, one can understand the travel time better with the distributional function. Probabilistic models provide plausible solution to measure the reliability. This could at least give certain confidence level to the road user to expect the arrival time. Over the past decades, many of the existing study focused on continuous distributions such as lognormal, gamma, Weibull, exponential and loglogistic to describe the travel time distribution ((Polus, 1979), (Xue et al., 2011), (Mazloumi et al., 2010), (Kieu et al., 2015))). Some applied normal distribution based on large sample size, which is the most common distribution has been used in the modelling, simply because of the elegant properties and readily applicable results. Such application may not be appropriate in terms of accuracy, neither suitable for the interpretation, as the data may not be certainly normally distributed. Burr XII distribution is an appropriate model to explain highly right skewed data because of its flexibility. (Susilawati et al., 2013) studied travel time variability on urban roads with Burr XII distribution. The approach is nonetheless appropriate for description because dual peaks are commonly seen in most of the cases of daily travel time. For instance, it is expected that the peaks are usually shown in the morning and in the evening during weekdays. For such bimodality, mixture model often takes good care of such phenomena and its interpretation. Here, we aim to discuss the fitting of day-to-day travel time data with the MBD distribution. The analysis shows promising results to outperform all unimodal distributions, and for some links it is competency with Gaussian Mixture Models (GMM).

It has been common to apply standard deviation to measure the reliability. In transportation modelling and reliability, standard deviation is used to measure the variation from the estimated mean travel time. Larger standard deviation simply means that the variation is higher, and hence the travel time reliability is low. Such approach is simple and direct, however it is not a good option for the skewed data, which is very likely happened in travel time data. The reliability metrics which uses skew and width based on percentile is considered instead to provide better interpretation (Van Lint et al., 2005), as the measure is based on median which is not sensitive to outliers. The percentile with probabilistic models based is calculated to measure the reliability, and it works well also for the data presenting skewness.

The objectives of this paper are to fit the travel time with MBD, and the analysis of the reliability is subsequently carried out. The secondary day-to-day data is provided by Rapid KL (a public transportation own by Prasarana), and the data is filtered for the peak in the morning 7am to 10am and the peak in the evening 5pm to 8pm. The result indicated that dual modes is presented in the travel time for both routes. For reliability analysis, it is expected that the route which covers urban shows relatively low reliability. See the analysis results in later section. The organization of the paper is as follows. Section 2 briefs some existing study in travel time modelling. Most of them emphasizes the fitting of unimodal distributions and GMM, none of them considered mixture of Burr distribution. Data resources and methodology are discussed in Section 3. Two routes are chosen here; one covers downtown KL which passing by the Central Business District (CBD), and another route covers suburban regions dominant by the residential areas and educational institutions. The details of all links and google map are provided along for better visualization. An explanation of the data collection is given. Section 4 reported significant findings. It is observed that the MBD outperformed all unimodal distributions, and it is competitive with GMM. Higher reliability is observed in the suburban bus route and implication is provided based on the analysis. Section 5 concludes.

2. Literature review

In developing country, travel time modelling is particular of the interest in transportation study as it plays an important role for a proper city transportation planning. Some modelling study is found in private transportation travel time modelling in various environments such as arterial roads, expressway and urban road network (Yu et al., 2020; Chalumuri et al., 2014; Shi et al., 2017) were conducted. This paper suggests a viable solution to explain the travel time modelling in public mode bus operators with mixture distribution. For public transportation, the study of travel time modelling especially in bus operation is necessary. Reliable estimation on bus travel time is not only to secure the confidence of bus passengers, it also beneficial to the bus-service company as it serves as an indicator for route condition. With the analysis results, the arrival time can be clearly tabulated at every bus stop station. The passengers can roughly understand the average travel time from station to station. Therefore, travel time efficiency can be improved if a proper travel time estimation and reliability analysis are carried out.

Time series methods have been widely used in estimating the travel time, as the travel time exist in time context, one can find many studies focus on using seasonal and trend patterns for description. Autoregressive integrated moving average (ARIMA) and Seasonal ARIMA are considered to address such problems. Comi et al. (2017) and Comi et al. (2020) applied time series approach in analyzing the travel time in Ukraine and Rome respectively. Despite of the analysis results provide promising prediction which could be offered to the authority for service enhancement, the approach is however a matured approach in past few decades, and the normality assumption is also a hurdle in the modelling as for such assumption is not always fulfilled by the real situation. Other approaches emerged in conjunction with time series, such as Kalman filtering and Machine Learning models of Artificial Neural Network (ANN). Fan and Gurmu (2015) compared the performance among historical average, Kalman filter and ANN. The results show that the ANN outperformed the counterparts. A mathematical-based model is developed by Wong (2009) to estimate regional bus travel time with ANN. Yuan et al. (2020) designed a mechanism of Recurrent Neural Network (RNN) to capture the dynamic temporal behavior and a Deep Neural Network (DNN) is used for travel time prediction. Noor et al. (2020) applied Support Vector Regression (SVR) to analyze the impact of explanatory variables on travel time of Urban City Bus data in Petaling Jaya, a main business district in Selangor. Result shows the weather has the least influence for travel time. Yu et al. (2017) applied survival models and regression analysis to predict the travel time of campus bus service. Such method resulted in good prediction on the travel time associated with uncertainties. Other relevant studies in travel time modelling involved simulation study to examine and to improve the bus service reliability, see Moosavi et al. (2020). Liyanage et al. (2020) suggested on-demand bus service rather than scheduled bus services, and the analysis results show superior benefits of the on-demand bus service. Both piece of works contributed significantly in sustainability practice.

Travel time exists in an uncertainty context. It is much appropriate to capture the travel time pattern by stochastic models. There have been many investigations in travel time modelling and reliability with probability density function in past three decades. The earlier research was carried out by Taylor (1982) who discussed the section travel times with normal distribution. Since then, the application of unimodal distributions such as normal, log-normal, log-logistic and Weibull has been commonly applied for travel time reliability. See Mazloumi et al. (2010), Ma et al. (2016), Shariat et al. (2019) and Büchel et al. (2020). Taylor and Susilawati (2012) shows that the bus travel time reliability is appropriately fitted by Burr distribution, and the travel time on urban roads often presenting dual modes (Susilawati et al., 2013). A recent study by Low et al. (2021) analyzed the bus travel time in Klang Valley region with Burr distribution. The result favored to the Burr distribution makes this study possible, where we may consider to explain the dual modes scenario of the travel time in Klang Valley region with a mixture of Burr distribution. The usage of applying unimodal distribution to capture travel time has been quite established. Highly right-skewed distribution could be easily explained by the heavy right-skewed distribution such as Weibull and its limiting distribution of Burr distribution. However, one of the important characteristics of the travel time which often presenting dual modes is usually neglected in the literature. Although the mixture of Gaussian models has been considered to deal with dual modes, but for both peaks to appear normal is subsequently a further argument. This study aims to fill up the gap of the study. A mixture of Burr distributions is introduced to handle the dual mode in travel time, which we found it appears in the day-to-day data, for both routes 851 and T786.

Mixture distribution is well-known to cater multimodality data in statistical study. In transportation modelling, it emerged as an important analysis tool for travel time modelling. Sun et al. (2018) classified the traffic flow in real-time with Gaussian mixture models (GMM) for better traffic operation and management. (Guo et al., 2019) analyzed the travel time collected from radio frequency identification technique (RFID) in urban road networks with GMM. Comparison has been done with the unimodal distributions and the results show that GMM defeats the counterparts. Similar study has been carried out by Ma et al. (2016) but focus on modelling the travel time variability for bus operations. Yang and Wu (2016) considered mixture models for fitting freeway travel time data. Three mixture models which have been discussed in the paper, i.e. mixture Gaussian, mixture gamma and mixture lognormal. The results show all mixture models are competence. Similar study has been found by Guessous et al. (2014). The mixture of two gamma and two normal distributions were considered to estimate the travel time distribution under different traffic conditions. To the best of our knowledge, the application of mixture distributions in transportation field is limited, especially in bus operating system. We aim to propose a mixture model, with 2-component of finite mixture of Burr distribution to fit the travel time data. The period of the data is 6 months long, and it is collected via Global Positioning System (GPS) within Klang Valley region. The analysis results show that mixture of Burr distribution is competent to the existing GMM, and it could be considered as a viable solution in further analysis.

3. Data resources and methodology

3.1 Data resources

The travel time data within Klang Valley is investigated in this study. Klang Valley is located in the centre of Malaysia. It comprises a Federal Territory of Kuala Lumpur and six districts in Selangor, i.e. Petaling, Klang, Gombak, Hulu Langat, Sepang and Kuala Langat. The secondary data is provided by Rapid KL, a public transportation system built by Prasarana Malaysia. We focus on the Rapid KL bus data. Two bus routes are considered in this paper, namely 851 (old route number B115) and T786 (old route number T786). Bus route 851 focuses on the service in Central Business District (CBD), while T786 covers the suburban area. RapidKL buses services covers 6 key areas of the Klang Valley. Route 851 is part of Damansara area coverage while T786 is operated under Lebuhraya Persekutuan area. For both bus routes, the AVL system provides the details such as bus stop ID, street name, route ID, stop ID, stop name and the distance between the bus stops. Some information such as stop ID, stop name and distance are given in Table 1.

Figure 1 shows the Google map of the bus routes. The total length for bus route 851 is 18.263km and for T786 is 10.277km. Both routes are considered intermediate and short length respectively, and they covered different area in Klang Valley. For bus route 851, it focuses on the service in Central Business District (CBD), while T786 covers the suburban area. RapidKL buses services covers 6 key areas of the Klang Valley. Route 851 is part of Damansara area coverage while T786 is operated under Lebuhraya Persekutuan area. The characteristics of the bus routes are given in Table 2. Both routes cover different area of Klang Valley. For 851, the route focus on Kuala Lumpur (KL) downtown area, which includes the central business district (CBD), such as Masjid Jamek and Pudu. In contrast, T786 operates in Petaling Jaya region, one of the main districts in Selangor. Comparable to 851, it covers suburban, which is dominated by the residential area and the institutions.

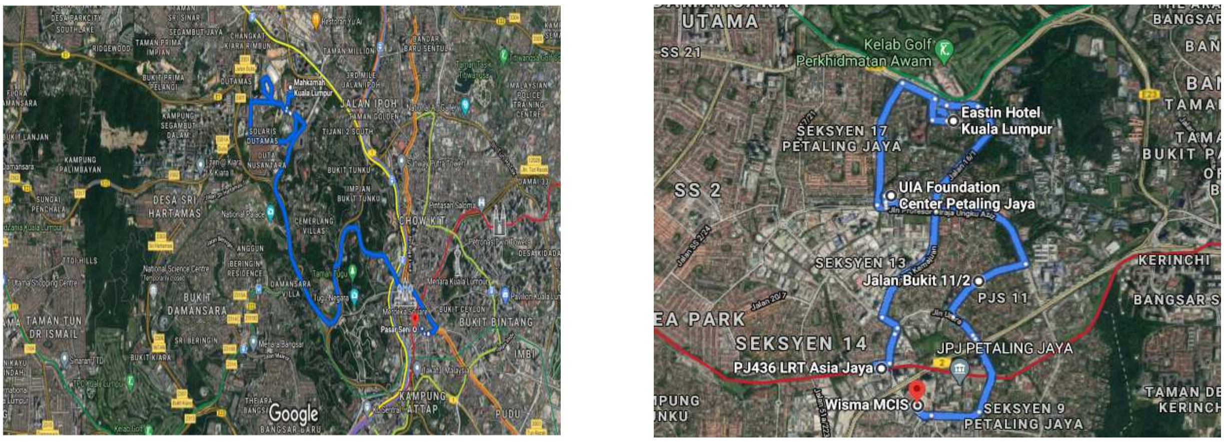
Table 1. Route and Stop ID.

Stop ID	Stop Name	Distance (km)
Route 851		
1004342	Pasar Seni 3 (Platform A1 - A2)	0.596
1002080	Mydin Sinar Kota	0.342
1001810	Lrt Masjid Jamek	0.406
1000230	Bank Negara Malaysia	0.616
1001070	Jkr	0.277
1001411	Kerja Raya	0.327
1001173	Jln Sultan Salahuddin	0.763
1001171	Jln Sultan Salahuddin	0.316
1000597	Lembaga Peperiksaan	4.383
1001099	Jln Duta	0.309
1001101	Jln Duta	0.237
1001623	Kompleks Mahkamah Kl	0.53
1001100	Jln Duta	2.378
1001375	Kedutaan India	1.406
1000598	Duta Vista	0.933
1001172	Jln Sultan Salahuddin	2.301
1001071	Jkr (Opp)	1.205
1000488	Cimb/Lrt Mjd	0.769
1001183	Jln Tun Perak	0.169
Route T786		
1001784	Lrt Asia Jaya	0.597
1001578	Renault	0.338
1000471	Century Battery	0.461
1000650	F&N	0.403
1000467	Castell	0.349
1000603	Eastin Hotel	1.658
1002183	Pangsapuri Tiara 2	0.355
1002920	Rumah No 26	2.178
1002675	Pusat Asasi Uia	0.31
1001955	Masjid Kolej Islam Malaya	0.293
1001007	Mahsa University / Intan	0.568
1004179	Universiti Tower	0.479
1003630	Srk Alam Shah	0.386
1002142	Pangsapuri Ehsan Ria	0.11
1004272	Yayasan Salam	0.563
1003857	Tasek Tmn Jaya	0.502
1001509	Kk Mart	0.727

Table 2. Characteristics of the bus routes.

Major Attraction	851	T786
Commercial		
Mydin	✓	
Publika/Solaris	✓	
Schools/Universities		
Mahsa University		✓
Pusat Asasi UIA		✓
SRK Alam Shah		✓
Transportation		
LRT Masjid Jamek	✓	
LRT Asia Jaya		✓
Governmental Offices		
Kompleks Mahkamah KL	✓	
Bank Negara	✓	
CIMB	✓	
Residential		
Pangsapuri Tiara		✓

The data is secondary, and it is filtered in the consideration of temporal and spatial aspects. It is a 6-month data set, from June to December 2014 and it was day-to-day collected by Global Positioning System (GPS) from 6:00AM to 12:00AM. For this study, the daily data is filtered for the peak in the morning (from 7:00AM to 10:00AM) and in the evening (from 5:00PM to 8:00PM). For both routes, we study the travel time in the temporal and spatial scale. The travel time for each link is in aggregated attributes for every 10 minutes. As an exemplification for aggregated time, the travel time captured by the GPS for busses travelling along the respected link from 7.00am to 7.10am will be accumulated. It is a subsequent accumulation for a duration of 10 minutes. This study combines the aggregated time for both the morning and evening peak hour periods for each link. For instance, given bus route 851, a link simply means that from the Stop ID 1004342 to the Stop ID 1002080, and the link length is 0.596km. We will investigate the travel time for every link of both routes and identify the links that show bimodality during this period. It is observed that dual modes appear in the travel time is observed.



(a) (b)

Fig. 1. Google map of bus routes (a) 851 and (b) T786.

3.2 Methodology

3.2.1 Mixture of Burr distribution

Mixture distribution has an established history in statistical field. It is also known as compound distribution. It has been defined to cater multimodality data. This study considers a mixture of Burr distribution to explain the travel time data. The proposal of using mixture of Burr distribution to fit the travel time data is the core contribution in this paper. See recent study for mixture of Burr distribution by (Aslam et al., 2018). To the best of our knowledge, fitting the travel time with mixture of Burr distribution is novel in transportation modelling. We aim to fit the dual peaks travel time with the mixture of Burr distribution.

θA m-finite mixture of a distribution is defined by the density with is the parameter

$$f(x) = \sum_{i=1}^{m} w_i p(x|\theta),$$

Where:

$w_i > 0\,; i = 1, \dots, m,\ \sum_{i=1}^{m} w_i = 1.$

This study considers a 2-component mixture of Burr distribution, where m = 2. The pdf is given as follows.

$$f(x) = w_1 \frac{c_1 k_1}{b_1}\left(\frac{x}{b_1}\right)^{c1-1}\left(1+\left(\frac{x}{b_1}\right)^{c_1}\right)^{-k_1-1}$$
$$+w_2 \frac{c_2 k_2}{b_2}\left(\frac{x}{b_2}\right)^{c_2-1}\left(1+\left(\frac{x}{b_2}\right)^{c_2}\right)^{-k_2-1}$$

Where $x > 0$, and c_1, c_2, k_1, k_2 are all positive. The corresponding cumulative distribution function is given by

$$F(x) = w_1\left(1-\left(1+\frac{x^{c_1}}{{b_1}^{c_1}}\right)\right)^{-k_1}$$
$$+w_2\left(1-\left(1+\frac{x^{c_2}}{{b_1}^{c_2}}\right)\right)^{-k_2}$$

The mixture of Burr distribution is carried out via Jupyter Notebook that supports Python programming language. A few modules like lmfit and numpy packages were used. It is a semi supervised learning approach. Maximum likelihood estimation is used to estimate the parameters. The Burr mixture model is then compared to other distributions such as GMM, including some unimodal distributions, which will be elaborated in next subsection. It is observed that the mixture of Burr caters well for the travel time of some links.

3.2.2 Evaluation approach

(A) Distribution fitting

Most of the existing reviews emphasizes the fitting of unimodal distributions. This is because the travel time usually performs highly skewed distribution. However, dual mode is possible presented in travel time, especially when the daily peak-hour framework is considered, as shown by Figure 2. For comparison purpose, we fitted the data with lognormal, Weibull, gamma, normal and GMM. Table 3 tabulates the probability distribution functions and the parameters for readers' reference.

Table 3. Probability density function and the parameters.

Distribution	Probability density function (pdf)	Parameters
GMM	$f(x) = \frac{a}{\sigma_1\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + \frac{1-a}{\sigma_2\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)$, $\mu_1, \mu_2 \in \mathbb{R}, {\sigma_1}^2, {\sigma_2}^2 > 0$	$\mu_1, \mu_2, \sigma_1, \sigma_2$
Gamma	$f(x\,; a\,, b) = \frac{x^{a-1}e^{-x/b}}{b^a \Gamma(a)}, x > 0, a, b > 0$	a, b
Burr	$f(x) = \frac{ck}{b}\left(\frac{x}{b}\right)^{c-1}\left(1+\left(\frac{x}{b}\right)^c\right)^{-k-1}, c > 0, k > 0$	b, c, k
Lognormal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left(-\frac{(ln\, x-\mu)^2}{2\sigma^2}\right), \mu \in \mathbb{R}, \sigma > 0$	σ, μ
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \mu \in \mathbb{R}, \sigma^2 > 0$	σ, μ
Weibull	$f(x) = \frac{B}{A}\left(\frac{t}{A}\right)^{B-1} e^{-\left(\frac{x}{A}\right)^B};\ x > 0, A, B > 0$	A, B

(B) Performance Evaluation

Model selection is always significant to evaluate the performance. Akaike Information Criteria (AIC) is applied here to justify the model viability. The formula of AIC is $2\boldsymbol{v}$ - 2 $ln\ L$ where $\boldsymbol{v}$ is the number of parameters of the distribution and $ln\ L$ is the estimated log-likelihood function.

3.3 Reliability

Standard deviation is usually carried out in transportation engineering field to measure the reliability. However, it is suitable for the data which is normally presentable. It is suggested to use other measurements such as range and interquartile for the data presents skewness. This paper applied skew-width methods by Van Lint et al. (2005) to measure the reliability. For such metrics it is considering to interpret the skewness by cumulative distribution function. The median-based Buffer Index (BI) represents an additional buffer time required in addition to the median of the travel time. For instance, considering a 95% confidence interval, with a median of 300 seconds and BI value of 0.6, an additional of 180 seconds are needed for travelers to arrive on-time. The BI is defined as:

$$BI_x = \frac{t_{95} - t_{50}}{t_{50}}$$

The skewness of the travel time is a ratio. The numerator subtracts the median from the 90th percentile, and the denominator subtract 10th percentile from the median. This ratio covers the range of 40% observations above the median, and thus it captures the tail and the skewness of the data Taylor et al. (2012). Similarly, the width of the travel time is also defined as the ratio, where the numerator takes the subtraction of the 10th percentile from the 90th percentile, and the denominator simple be the 50th percentile of the travel time. The equation of the skewness and the width travel time metrics are provided as follows.

$$\lambda_{skew} = \frac{t_{90}-t_{50}}{t_{50}-t_{10}} \quad (1)$$

$$\lambda_{var} = \frac{t_{90}-t_{10}}{t_{50}} \quad (2)$$

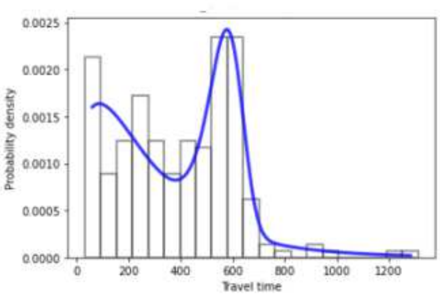
4. Results and discussion

4.1 Distributional fitting

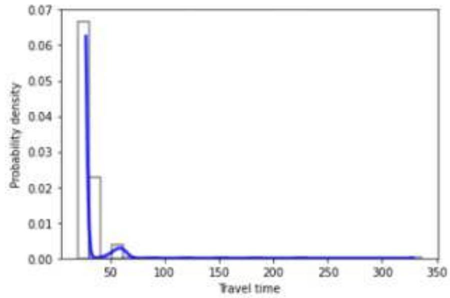
The link travel time is fitted by the (1) mixture of Burr distributions. Some existing distributions have been fitted with the data for comparison purpose. The distributions considered here are the well-know (2) Gaussian Mixture Model (GMM), the unimodal (3) Burr, (4) lognormal, (5) Weibull, (6) gamma and (7) normal distributions. We run the fittings for all links of both routes, but some results have been presented here due to the sake of brevity.

We reported the findings from the data fitting. First, it is found that the pattern of all links is outperformed by the mixture models, compared to the unimodal distributions. In addition to that, it is further justified by AIC that the mixture of Burr distribution obtained slightly lower than the GMM for certain links. The analysis shows that the hypothesis of performing mixture of Burr distributions in explaining the link travel time is plausible, as the lower AIC indicates a better fit for the model. We first discuss the route 851. The results show that for some links where the mixture of Burr distribution performed slightly better than GMM. For instances, the link 1000488 and 1001071 gives lower AIC values, which the AIC indicator suggests that the mixture of Burr is more presentable. The bimodality of the links presented is predicted due to the route characteristics which operates around several tourist attractions spots such as Merdeka Square, Tugu Negara and the Sultan Abdul Samad Building. These national monumental are located along the busy street of this bus link. For some links, the mixture of Burr distribution is observed to be competent with the GMM. For instance, for the link 1004342, the AIC values of MBD and GMM are 32.11 and 31.20 respectively. This is the starting point of the bus route situated at a busy area with LRT facilities and surrounded by the 4 signalized intersection street. The bus operating along this 0.6 kilometre link travels approaching Bus id 1002080. As for link 1002080, it has exhibited an evidence of bimodality because of its shorter length and the influence that the bus stop is located within a commercial area such as the mini market Mydin. Susilawati (2013) and Ma et al. (2016) states that bimodality usually present in shorter links of the routes. Result has proven that the travel time are not fitted by unimodal Burr distribution with finite parameters, and this concluded that the mixture models has captured a better result. It is enlightened to observe that for both links 1001183 and 1000597, GMM fails to carry out the distributional fitting due to the parameter constraint. Link 1000597 is the longest link in route T786, measuring 4.4 kilometres. This part of the route lies at the Board of Examination, Ministry of education and is interconnected with busy roads in Hartamas. Together with that is a mosque and an institute situated nearby. Burr Mixture model ranks as the best model to fit the data of link 1000597. Furthermore, it has also the advantage to cater the travel time for both short and long links. Results has also shown that Burr Mixture model fits nicely to the travel time data for Route T786.

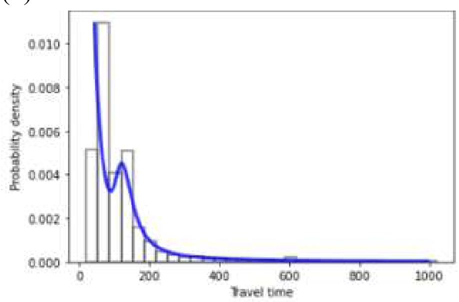
Route T786 is a shorter route compared to 851 which consists of links passing through educational institutes and residential areas, which the coverage is in suburban region. From the analysis it provides strong justification that the mixture of Burr distribution is a plausible model in the travel time analysis. MBD performs well to explain the pattern of both routes spatially (urban and suburban) and temporally (peak hours). See Table 4 for the performance evaluation and parameter estimation of the distributions. Also, the travel time of three links, i.e. 1002080, 1002142 and 1001509 has been presented together with MBD. Figure 2 shows that the MBD fits nicely to the travel time data. Our analysis provides strong justification that the mixture of Burr distribution is a plausible model in the travel time analysis.



(a)



(b)



(c)

Fig. 2. MBD fittings for links (a) 1002080, (b) 1002142, and (c) 1001509.

Table 4. Performance evaluation and parameter estimation.

Bus ID	Distribution	Parameter estimation	Log Likelihood	AIC
Route: 851				
1001071	MBD	$w_1 = 0.11, b_1 = 785.08, c_1 = 10, k_1 = 10$ $w_2 = 0.89, b_2 = 770.52, c_2 = 1.50, k_2 = 10$	-16.99	**33.97**
	GMM	$w_1 = 0.7, \mu_1 = 63.23, \sigma_1 = 14.80$ $w_2 = 0.3, \mu_2 = 129.52, \sigma_2 = 34.57$	-17.46	34.92
	Burr	$b = 74.48, c = 3.04, k = 0.72$	-2401.45	4808.9
	lognormal	$\mu = 4.51, \sigma = 0.69$	-2412.51	4829.02
	Weibull	$A = 130.25, B = 1.28$	-2484.9	4973.8
	gamma	$a = 1.96, b = 60.61$	-2460.59	4925.18
	normal	$\mu = 119.18, \sigma = 115.4$	-2676.23	5356.46
1002080	MBD	$w_1 = 0.65, b_1 = 999.99, c_1 = 1.32, k_1 = 5.35$ $w_2 = 0.35, b_2 = 738.38, c_2 = 9.99, k_2 = 10$	-16.49	32.97
	GMM	$w_1 = 0.005, \mu_1 = 63.58, \sigma_1 = 5.46$ $w_2 = 0.095, \mu_2 = 365, \sigma_2 = 266.18$	-15.18	**30.36**
	Burr	data are not fitted by Burr	-	-
	lognormal	$\mu = 5.61, \sigma = 0.89$	-16067.2	32138.4
	Weibull	$A = 406.39, B = 1.56$	-15769.2	31542.4
	gamma	$a = 1.87, b = 196.70$	-15830.2	31664.4
	normal	$\mu = 366.88, \sigma = 196.70$	15938.4	-31872.8
1000488	MBD	$w_1 = 0.32, b_1 = 35.62, c_1 = 9.99, k_1 = 0.1$ $w_2 = 0.68, b_2 = 60.23, c_2 = 9.99, k_2 = 0.2$	-15.83	**31.66**
	GMM	$w_1 = 0.37, \mu_1 = 68.99, \sigma_1 = 31.04$ $w_2 = 0.63, \mu_2 = 185.64, \sigma_2 = 118$	-17.07	34.14
	Burr	$b = 334, c = 1.62, k = 3.14$	-3697.43	7400.86
	lognormal	$\mu = 4.86, \sigma = 0.82$	-3692.71	7389.42
	Weibull	$A = 192.81, B = 1.35$	-3703.62	7411.24
	gamma	$a = 1.77, b = 99.5$	-3695	7394
	normal	$\mu = 175.81, \sigma = 99.5$	-3854.25	7712.5

1004342	MBD	$w_1 = 0.88, b_1 = 96.40, c_1 = 10, k_1 = 0.05$ $w_2 = 0.12, b_2 = 63.99, c_2 = 10, k_2 = 9.88$	-16.06	**32.11**
	GMM	$w_1 = 0.04, \mu_1 = 50.69, \sigma_1 = 2.37$ $w_2 = 0.96, \mu_2 = 121.37, \sigma_2 = 220.23$	-15.60	31.20
	Burr	$b = 1550.6, c = 1.15, k = 11.49$	-5216.94	10439.88
	lognormal	$\mu = 4.78, \sigma = 1$	-5190.22	10384.44
	Weibull	$A = 195.92, B = 1.09$	-5217.3	10438.6
	gamma	$a = 1.22, b = 154.70$	-5213.36	10430.72
	normal	$\mu = 188.72, \sigma = 175.85$	-5514.12	11032.24
1001183	MBD	$w_1 = 0.52, b_1 = 56.46, c_1 = 4.76, k_1 = 1.10$ $w_2 = 0.48, b_2 = 118.34, c_2 = 5.87, k_2 = 0.39$	-13.40	**26.80**
	GMM	Data are not fitted with GMM	-	-
	Burr	Data are not fitted with Burr	-	-
	lognormal	$\mu = 4.45, \sigma = 0.81$	-1070.33	2144.66
	Weibull	$A = 129.99, B = 1.24$	-1087.45	2178.9
	gamma	$a = 1.62, b = 73.93$	-1082.71	2169.42
	normal	$\mu = 120.41, \sigma = 110.48$	-1156.9	2317.8
1000597	MBD	$w_1 = 0.58, b_1 = 43.32, c_1 = 9.99, k_1 = 0.06$ $w_2 = 0.42, b_2 = 766.52, c_2 = 10, k_2 = 10$	-14.50	**28.99**
	GMM	data are not fitted with GMM	-	-
	Burr	data are not fitted with Burr	-	-
	lognormal	$\mu = 5.44, \sigma = 1.09$	-152.327	308.654
	Weibull	$A = 381.09, B = 1.21$	-150.887	305.774
	gamma	$a = 1.28, b = 279.75$	-151.023	306.046
	normal	$\mu = 358.73, \sigma = 282.857$	-154.905	313.81

Route: T786				
1001007	MBD	$w_1 = 6.45 \times 10^{-9}, b_1 = 388.68, c_1 = 0.19, k_1 = 9.93$ $w_2 = 0.99, b_2 = 71.41, c_2 = 5.99, k_2 = 0.33$	-18.08	**48.16**
	GMM	invalid	-	-
	Burr	$b = 66.23, c = 3.74, k = 0.45$	-16528.20	33062.40
	lognormal	$\mu = 4.64, \sigma = 0.73$	-16650.80	33305.60
	Weibull	$A = 151.33, B = 1.23$	-17140.60	34285.20
	gamma	$a = 1.79, b = 78.10$	-17003.70	34011.40
	normal	$\mu = 139.89\ , \sigma = 142.23$	-18516.60	37037.20
1002142	MBD	$w_1 = 0.95, b_1 = 28.14, c_1 = 10, k_1 = 3.85$ $w_2 = 0.05, b_2 = 74.84, c_2 = 9.99, k_2 = 10$	-18.38	48.75
	GMM	$w_1 = 0.95, \mu_1 = 23.88, \sigma_1 = 14.07$ $w_2 = 0.05, \mu_2 = 53.45, \sigma_2 = 4.42$	-18.58	**45.16**
	Burr	$b = 29.48, c = 116.52, k = 0.10$	-1365.68	2737.36
	lognormal	$\mu = 3.46, \sigma = 0.24$	-2118.24	4240.48
	Weibull	$A = 37.66, B = 1.86$	-2528.44	5060.88
	gamma	$a = 10.01, b = 3.35$	-2281.82	4567.64
	normal	$\mu = 33.58, \sigma = 19.54$	-2673.88	5351.76
1001509	MBD	$w_1 = 0.78, b_1 = 26.91, c_1 = 9.99, k_1 = 0.1$ $w_2 = 0.22, b_2 = 118.67, c_2 = 9.99, k_2 = 0.44$	-19.59	51.18
	GMM	$w_1 = 0.54, \mu_1 = 31.07, \sigma_1 = 10.64$ $w_2 = 0.46, \mu_2 = 120.16, \sigma_2 = 46.73$	-17.37	**42.73**
	Burr	$b = 53.29, c = 3.03, k = 0.60$	-11127.90	22261.80
	lognormal	$\mu = 4.29, \sigma = 0.72$	-11133.70	22271.40
	Weibull	$A = 106.84, B = 1.26$	-11437.90	22879.80
	gamma	$a = 1.87, b = 52.44$	-11339.90	22683.80
	normal	$\mu = 98.09, \sigma = 95.01$	-12339.70	24683.40

4.2 Interpretation and reliability metrics

Equation (1) and (2) are calculated based on percentile values for the travel time distribution. Equation (1) with more than one indicates that users has greater delay with respect to the travel time median. In general, as Equation (1) increases, it is more likely for a user to experience high travel time compared to the median. Equation (2) indicates the spread of travel time distribution. The larger the value, the higher the difference between two extreme values compared to the median. Large values of both parameters give an indication of low reliability, which means the uncertainty of the travel time is increased.

It is reported that the link 1001509 of route T786 and the link 1001071 of route 851 have poor reliability. The buffer time is 4.74 and 3.45 respectively, indicating that more additional travel time is required relatively to median. For instance, the travel time median for link 1001509 is about 1.23 minutes. The buffer time of 4.74 simply means that 4.74 times more than 1.23 minutes, which equivalently an additional of 5.83 minutes is needed for a traveler to travel the link. The analysis can be served as an indicator, which is very useful to the users to understand better the travel time and its reliability for each route link. It is believed that one of the contributing factors for link 1001509 could be the interconnectivity of this link to busy roads in Hartamas. Besides that, link 1001509 is the last station of route 851. The result could be caused by the congestions at the station. As for link 1001071, it is observed that the 1.205 km section is in between a stretch of two junctions. The bottleneck junction along this link which connects a street from where Asia School of Business-Residence could affect the flow of the traffic. Besides that, with the governmental offices (Public Works, Bahagian Perundingan Pengurusan Aset, JKR Malaysia) around might give rise to the low travel time reliability during peak hours. On the contrary, link 1002080 from route 851 and link 1002142 from route T786 had revealed the best travel time reliability among the links. This is in consideration of the fact that link 1002080 operates along a short stretch of a main road, travelling on a straight 0.3 km road with only 1 signalized intersection. To add on, there was no any disruptions flow caused by junctions or bottlenecks. The same condition had been observed for link 1002142. The link travels along a 0.1 km road with no signalized intersection in between. It is evident that these could be the factors that contribute to the high travel time reliability for these two links.

The reliability analysis has been analyzed for all links for both routes. Table 5 tabulates parts of results due to sake of brevity. The finding implies that for the links which connected by signalized intersections and junctions may get lower reliability. Therefore, it is reasonable to give tolerance of the travel time according to the buffer index for each respective links, i.e. an additional of 2 minutes bus arrival time is expected at the link 1004342.

Table 5. Performance evaluation and parameter estimation.

Link	λ_{var}	λ_{skew}	BI
Route 851			
1001071	4.26	2.14	3.45
1002080	0.71	1.40	0.74
1004342	2.56	2.89	2.08
1001183	3.29	2.25	2.69
1000597	1.00	1.77	1.05
Route T786			
1002142	1.00	0.36	0.34
1001509	3.64	2.76	4.74

5. Concluding remark

Since the quick and massive developments over the past few decades, road congestion has been a major issue in Klang Valley, especially the areas that located in the center, such as KL downtown and its connectivity of Petaling Jaya. The issue remains unsolved and the road traffic condition is getting deteriorated with improper traffic design as time goes by. Delay in bus travel time may cause inconvenience to the passengers, for the bus service company, losses is expected if passenger reduction happened due to the uncertainty of the travel time. Proper time management can be well done with a feasible solution on a reliable travel time estimation. This paper provided an insight for the travel time pattern of bus routes T786 and 851, which covers suburban and urban region respectively. The investigation started with the diagnostic of the travel time pattern, by using a mixture of Burr distribution. The mixture of Burr captured the travel time in temporal-spatial aspects. The MBD well performs to estimate the dual modes travel time for the links of both routes, with different route characteristics is taken into the consideration. Results showed that the travel time of some links is well fitted by the distribution, and it is competitive with the existing GMM. Certainly, the unimodal distributions (Burr, Lognormal, Weibull, Gamma and Normal) provide poor fitting when the data presents dual mode. Subsequently, we study the travel time variability with skew-width approach. Results indicated that low reliability is observed in the urban route. This may be affected by the attraction spots, the number of intersections and the length of the links. For bus route in suburban region, it is expected that a delay of 5 minutes is possible in the last station. The findings of this paper serve an important information for the users and the bus service company. Also, the information is a factor in order to carry out a relevant study of benefit-cost analysis in the near future.

6. Acknowledgement

The authors would like to acknowledge RapidKL for the assistance with data collection. This study is supported by Internal Research Grant (IRG) 2021 of Sunway University with the grant code GRTIN-IRG-122-2021, and the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS), Project No.: FRGS/1/2018/TK08/UTAR/02/1.

7. References

Aslam, Muhammad, Tahir, Minhas & Hussain, Zawar. (2018). Reliability analysis of 3-component mixture of distributions. Scientia Iranica. 25. 10.24200/sci.2017.4441.

Büchel, B. & Corman, F. (2020). Review on Statistical Modeling of Travel Time Variability for Road-Based Public Transport. Front. Built Environ. 6:70. doi: 10.3389/fbuil.2020.00070.

Comi, A. Nuzzolo, A. Brinchi, S. & Verghini, R. (2017). Bus travel time variability: some experimental evidences. Transportation Research Procedia 27, pp101-108.

Comi, A., Zhuk, M., Kovalyshyn, V., & Hilevych, V. (2020). Investigating bus travel time and predictive models: a time series-based approach. Transportation Research Procedia, 45, 692–699. doi:10.1016/j.trpro.2020.02.109.

Chalumuri, R.S. & Yasuo, A. (2014). Modelling travel time distribution under various uncertainties on Hanshin expressway of Japan. Eur. Transp. Res. Rev. 6, 85–92. https://doi.org/10.1007/s12544-013-0111-3

Fan, W. & Gurmu, Z (2015). Dynamic Travel Time Prediction Models for Buses Using Only GPS Data. International Journal of Transportation Science and Technology 4(4), pp 353-366, ISSN 2046-0430, https://doi.org/10.1016/S2046-0430(16)30168-x.

Guo, J., Li, C., Qin, X., Huang, W., Wei, Y., & Cao, J. (2019). Analyzing distributions for travel time data collected using radio frequency identification technique in urban road networks. Science China Technological Sciences, 62(1), 106-120. doi: 10.1007/s11431-018-9267-4.

Guessous, Y., Aron, M., Bhouri, N., & Cohen, S. (2014). Estimating Travel Time Distribution under Different Traffic Conditions. Transportation Research Procedia, 3, 339-348. doi:https://doi.org/10.1016/j.trpro.2014.10.014

Kieu, Le Minh, Bhaskar, Ashish, & Chung, Edward (2015) Public transport travel time variability definitions and monitoring. Journal of Transportation Engineering, 141(1), Article number: 04014068 1-9.

Liyanage, S. & Dia, H. (2020). An agent-based simulation approach for evaluating the performance of on-demand bus services. Sustainability 12, 4117. doi:10.3390/su12104117

Low, V. J. M., Khoo, H. L. & Khoo, W. C. (2021).

Statistical modelling of bus travel time with Burr distribution. ITM Web of Conference 36, 01011. ICMSA. https://doi.org/10.1051/itmconf/20213601011

Van Lint, J. W. C., & Van Zuylen, H. J. (2005). Monitoring and Predicting Freeway Travel Time Reliability: Using Width and Skew of Day-to-Day Travel Time Distribution. Transportation Research Record, 1917(1),54–62. https://doi.org/10.1177/0361198105191700107

Ma, Z., Ferreira, L., Mesbah, M., & Zhu, S. (2016). Modeling distributions of travel time variability for bus operations. Journal of Advanced Transportation, 50(1), 6-24. doi:10.1002/atr.1314

Mazloumi, E., Currie, G., & Rose, G. (2010). Using GPS data to gain insight into public transport travel time variability. Journal of Transportation Engineering, 136(7), 623-631.

Moosavi, S. M. H., Ismail, A. & Yuen, C. W. (2020). Using simulation model as a tool for analysing bus service reliability and implementing improvement strategies. Plos One. https://doi.org/10.1371/journal.pone.0232799

Md Noor, R., Seong Yik, N., Kolandaisamy, R., Ahmedy, I., Hossain, M.A., Alvin Yau, K., Md Shah, W. & Nandy, T. (2020). Predict Arrival Time by Using Machine Learning Algorithm to Promote Utilization of Urban Smart Bus. doi: 10.20944/preprints202002.0197.v1

Polus, A. (1979). A study of travel time and reliability on arterial routes. Transportation, 8(2), 141-151. doi:10.1007/BF00167196

Shariat Mohaymany, A., Ganjkhanloo, A., Bahaabadi, M. R., & Sayyad, A. (2019). Analysis of Travel Time Distribution for Varying Length of Time Interval. AUT Journal of Civil Engineering, 4(2), 10-10. doi:10.22060/ajce.2019.15596.5545

Shi, C. Y., Chen, B. Y. & Li, Q. (2017). Estimation of travel time distribution in urban road networks using low-frequency floating car data. ISPRS International Journal of Geo-Information, 6, 253, doi: 10.3390/ijgi6080253

Susilawati, S., Taylor, M. A. P., & Somenahalli, S. V. C. (2013). Distributions of travel time variability on urban roads. Journal of Advanced Transportation, 47(8), 720-736. doi:10.1002/atr.192

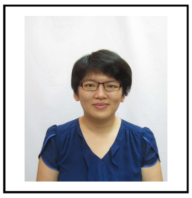
Taylor, M. (1980). Travel time variability – the case of two public modes. Transportation Science 16(4): 507-521. http://dx.doi.org/10.1287/trsc.16.4.507

Taylor, M., & Susilawati, S. (2012). Modelling Travel Time Reliability with the Burr Distribution. Procedia - Social and Behavioral Sciences, 54, 75–83. doi:10.1016/j.sbspro.2012.09.727

Wong, A. C. K. (2009). Travel time prediction model for regional bus transit. Master's thesis. Available from tspace.library.utoronto.ca

Sun, W., Hu, Z & Hong, L. J. (2018). Gaussian Mixture Model-Based Random a Search for Continuous Optimization via Simulation, Winter Simulation Conference (WSC), Gothenburg, Sweden, 2018, pp. 2003-2014, doi: 10.1109/WSC.2018.8632380

AUTHOR BIOGRAPHIES



Wooi Chen Khoo is a senior lecturer at Sunway University in Malaysia since 2017. Khoo received her Ph.D degree in Statistics in the field of applied probability and application from University of Malaya. She also holds BSc and MSc degrees in Mathematics from University of Science Malaysia. She is currently the University representative of data management. She deals with the Department of Statistics of Malaysia (DOSM). Her areas of interests are in applied probability and the application, parameter estimation, time series, statistical Inference and data analytics.