Research article

Public Transit Bus Travel Time Variability Analysis Using Limited Datasets: A Case Study

Ashwini Bukanakere Prakash*, Ranganathaiah Sumathi and Honnudike Satyanarayana Sudhira

Department of Computer Science and Engineering, Siddaganga Institute of Technology, Tumakuru, 572103, India

Curr. Appl. Sci. Technol. 2024, Vol. 24 (No. 5), e0257174; https://doi.org/10.55003/cast.2024.257174

Received: 16 December 2022, Revised: 4 March 2024, Accepted: 20 April 2024, Published: 13 June 2024

Abstract

Public transit service is a sustainable and eco-friendly alternative for commuting, and promoting its usage is the need of the day. An understanding of the variability of travel time can aid service operators to improve the reliability and ridership of public transport. Gaining insights into the variability of travel time is a data-intensive task, and most of the existing studies utilize multiple traffic-related datasets. However, most cities lack the infrastructure to collect multiple data sets, hence in the current study, the location data of public transit buses were used for the analysis. The study was conducted in Tumakuru city, India at two spatial levels, namely route and segment, and further at temporal levels such as the day-of-the-week and departure time window. Wilcoxon signed-rank test was applied to identify similar spatial-temporal aggregations, and a few aggregations demonstrated similarity. Consistent with the existing literature, six statistical distributions were selected to fit the data through the Kolmogorov-Smirnov test. The results emphasized that the Logistic distribution is the best fit at all spatial-temporal aggregation levels, and the lognormal and GEV distributions offered better fit for a few aggregation levels. Logistic distribution is recommended for operations planners and researchers to conduct reliability analysis and travel time forecasting in the future.

Keywords

travel time variability;

public transit reliability;

statistical distributions;

spatial-temporal aggregations

1. Introduction

Most tier-1 and metropolitan cities have more than one mode of mass transit such as local trains, metro rails, buses, etc. However, buses are the sole mode of mass transit in several smaller cities and urban areas. Hence, providing a better quality of service to commuters in such areas becomes even more

*Corresponding author: E-mail: ashvinibp@sit.ac.in

crucial. Changing the mode of travel of commuters from private to public is a gigantic task. Transit operation planners [1] play a vital role in bringing the modal shift as they are responsible for rolling out operations such as timetable generation, routing, providing dynamic information to passengers, evaluating the services.

Most of the existing research on travel time analysis was carried out on homogeneous traffic conditions with good infrastructure and of high technological maturity. India has heterogeneous traffic composition, which makes it challenging to provide new solutions to travel-related problems. Existing research on travel time analysis in India was mainly conducted in tier-1 cities [2-4]. Tier-1 cities have multi-mode public transit services and better infrastructure compared to tier-2 and tier-3 cities [5]. There are eight tier-1 cities in India [6] and more than hundreds of tier-2 and tier-3 cities that are experiencing rapid development. In the majority of tier-2 and tier-3 cities of India, the predominant public transit mode is bus.

The variability in travel times and speeds under various spatial-temporal scales [7] was characterized to explore the scale of prediction [8], suitable prediction models [9], and the technology that will be required for developing smart applications [10] in the future. In this context, the travel time and speed of public transit buses in Tumakuru, a tier-2 city in a southern state of India, was analyzed. A comprehensive study of a selected Tumakuru city service route was conducted: (i) to discover similar spatial-temporal aggregations that could be modelled together, (ii) to identify the best fit distribution for further analysis for researchers and service providers, (iii) to study the impact of spatial aggregation that is based on the land use pattern of the route, and (iv) to highlight the impact of the temporal aggregations at the day-of-the-week and DTW levels for future analysis.

Analysis of the travel time behavior of public transit services is a well-researched topic. Several researchers analyzed the reliability [11] and variability [12] of travel time [7] at various spatial-temporal scales using a variety of data sources [13]. Further, researchers analyzed public transit bus data for estimation and prediction of travel time using statistical [2,14, 15], mathematical [14, 16], data mining [17, 18], machine learning [19, 20-25], and hybrid [3, 26] approaches and used these predictions to forecast bus arrival time at bus-stops. From the above discussion, it is clear that characterizing travel time and speed is one of the essential tasks for any travel-related analysis. An insightful analysis of travel time can lead to various applications such as public transit information systems [27] to provide a cohesive schedule and dynamic information of the arrival of the buses at bus stops.

There were several studies on the variability of travel time of public transit buses conducted worldwide in the past. In a study conducted in Paris, France, Taylor [28] analyzed the variability of public transit buses and metro rails by conducting a chi-square test and it was concluded that buses had high variability compared to metro trains and that the travel times of buses followed log-normal distributions. Mazloumi *et al*. [29] analyzed the variability of bus travel time (BTT) at two temporal scales, namely short departure time window (DTW) and long DTW, and found that the BTT followed a normal distribution for both long and short DTWs except during off-peak hours of long DTWs.

In a study by Susilawati *et al*. [30], data logs of buses in Adelaide, Australia were analyzed and it was concluded that the Burr distribution represented the variations in the BTT better than lognormal, Weibull, Gamma, and Generalised Pareto. Findings from the statistical distribution analysis done by Chepuri *et al*. [31], it indicated that travel times during peak hours were better described using normal distributions. Rahman *et al*. [9] identified lognormal and normal distributions as the better models, and a combination of statistical distributions was also recommended as an alternative to a single distribution to provide better estimates of bus arrival time. In a study by Mazloumi *et al*. [7], GPS location data for a bus route in Melbourne, Australia, and flow data using loop detectors were used in a framework developed to predict variance and mean travel time. Birr *et al*. [32] assessed the impact of external factors on the travel time of public transit buses on the roads of a Tri-City Agglomeration in northern Poland and concluded that for developing models to estimate the travel time of public transit vehicles, the sections of the networks had to be analyzed for traffic behavior and the available infrastructure, taking into account the dwell time of vehicles. Comi *et al*. [8] conducted an analysis of the location and vehicle counter data in Rome in mixed traffic conditions and discovered a similarity between the time of travel and temporal patterns and concluded that it could help in short-term travel time forecasting.

Rahman *et al*. [9] proposed a methodology to analyze the statistical distributions of bus travel times using the GPS logs of buses and found that lognormal and normal distributions were best suited for 7-8 kms horizon. Furthermore, a combination of probability distributions according to different horizons for improving the bus arrival estimates was emphasized. Kathuria *et al*. [5], used location data from the bus transit system of Ahmedabad, India, and performed a variability analysis at three levels; first at route level, second at segment level, and third at the level of a reliability analysis. Ashwini and Sumathi [33] conducted variability analysis of public transit buses in Tumakuru, India, and examined the various distributions that fitted the travel time at route level.

Overall, from the literature, it was inferred that the existing research work was conducted at various locations globally, and most studies were location specific; there was no generic method that could be used to analyze travel time variability. Most researchers used statistical approaches for variability analysis and tried to fit several distributions to identify the best fit distribution for modelling the data, and the identified distribution was again different for each location. Most of the work was carried out using multiple data sets and in areas with mature technology related to traffic. The current work was conducted in a limited data scenario in a small-scale city to gain insights into the travel time variability based on a statistical approach.

2. Materials and Methods

When variability is studied, the knowledge gained can enhance the predicting ability, certainty, and steadiness associated with travel conditions. Working in this direction, travel time data were aggregated for various spatial-temporal aggregations, and preliminary analysis, sample comparison, and data modeling were performed to check the goodness-of-fit for each spatial-temporal aggregations in the study area.

2.1 Spatial-temporal aggregations of travel times

In general travel, data exhibits spatial-temporal characteristics. Aggregating data at various spatial-temporal scales [34] have a significant influence on both the choice of best-fitting probability distribution and the basic statistics. It is necessary to identify the spatial-temporal aggregations that suit the travel parameters. According to the existing literature, travel networks were analyzed at four spatial aggregations: at the network level, the route level [29], the segment level [9], and the bus-stop level [35]. Route level is from the origin of the route to the destination. Segment level is either based on the location of intersections or at the end of a set of bus stops and bus-stop level is between two consecutive bus stops. From the literature review, it was apparent that common temporal scales of analysis were at the day-of-the-week, weekday and weekend, time-of-the-day, and DTW levels [28, 36]. A few researchers considered the peak and off-peak hours [24] for analysis and neglected the rest of the day.

In the current study the travel time and speed were analyzed at two spatial aggregations:

1) Route level (from origin to destination of the route)

2) Segment level (based on the land use pattern of the city along the route)

For each spatial scale, the data is further aggregated at two temporal scales:

1) Day-of-the-week (all days of the week)
2) Departure time window (1-hour window from 7:00 to 20:00)

Note:10:00 AM -11:00 AM was considered as peak hour in a few analyses [37, 38].

2.2 Study area and data

The study was conducted in Tumakuru, which is a small-scale city in Karnataka, India. All city service buses in the Tumakuru city service have GPS equipped bus tracking systems. The location data of the buses during March 2022 were used for this analysis. The route number 201, Tumakuru bus stand (TBS) to Kyathasandra (KYA) was the sample route studied. Further, the route was split into four segments:, the Central Business District (CBD) segment 1(S1), which is the core city region, from the TBS to Bhadramma Choultry (BC), the Inner City (IC) region segment 2 (S2) from BC to Shivakumara Swamiji Circle (SSC), the Inner Suburban (ISU) region segment 3(S3) from SSC to Batawadi (BW), and finally segment 4(S4) the Outer Suburban (OSU) region from BW to KYA [39]. The details of route TBS-KYA are given in Table 1, and a map of the route highlighting the bus stops and the margins of the segments is shown in Figure 1. Eight hundred trips spread over various times of the day and on all days of the week were extracted for the study.

Table 1. TBS-KYA route details

Parameters	Segment 1	Segment 2	Segment 3	Segment 4	Route
Intersections	3	1	1	1	6
Length	1.76 km	1.0 km	2.09 km	2.05 km	6.9 km
Lanes	2	2	2	3	2
Bus stops	3	2	3	1	9

Figure 1. Route map of TBS-KYA along with bus-stops

2.3 Preliminary analysis

The basic descriptive statistics of the travel time and speeds of the route 'Tumakuru Bus Stand (TBS) to Kyathasandra (KYA)' are summarized in Table 2. The standard deviation and coefficient of variation of each segment illustrate that the segment behavior was distinctive. Skewness is a common measure used in statistics that reveals the extent to which the given data is oriented towards one side in a probability distribution. A non-zero skewness value generally indicates that the data is

Table 2. Descriptive statistics of travel time and speed data at route and segment level

	Travel Time in Seconds					Speed in Kilometer per Hour				
Statistic	Route	S1	S2	S3	S4	Route	S1	S2	S3	S4
Minimum	839.00	240.0	132.0	182.0	122.0	6.56	6.52	9.34	10.2	3.30
Maximum	1801.0	734.0	521.0	529.0	539.0	27.3	21.4	26.8	31.0	46.0
1^{st} Quartile	1186.7	392.7	254.7	253.0	216.0	16.4	11.2	12.9	17.4	26.5
3^{rd} Quartile	1383.0	496.0	344.0	312.0	274.2	19.6	14.2	16.9	20.7	33.2
Median	1273.0	445.0	292.0	284.0	244.5	18.0	12.4	14.8	19.2	29.8
Mean	1286.2	447.5	298.9	286.2	250.7	18.1	12.7	15.1	19.1	30.0
Standard deviation	158.3	77.90	60.54	51.78	53.41	2.51	2.28	2.96	2.87	5.80
Variation coefficient	0.12	0.17	0.20	0.18	0.21	0.14	0.18	0.20	0.15	0.19
Skewness	0.50	0.38	0.29	0.79	1.21	0.15	0.44	0.77	0.25	-0.4
Kurtosis	0.60	0.30	-0.11	2.11	3.12	1.97	0.34	0.85	1.25	2.35
Mean AD	121.4	62.28	49.34	38.52	39.22	1.87	1.83	2.35	2.16	4.28
Median AD	95.0	51.50	43.00	29.00	29.00	1.55	1.50	2.01	1.62	3.30

not perfectly symmetrical. The skewness values of the travel time and travel speed data of the buses were non-zero, indicating nonsymmetrical data distribution. Kurtosis is another statistical measure that indicates the shape of the probability distribution. A kurtosis value higher than 3 indicates fat tails and possible presence of outliers. In the current study the kurtosis value was high for the travel time of segment 4. Further in the study, Kolmogorov-Smirnov tests [40] were conducted to find the best fit distribution for the travel time data. Based on probability densities, a histogram of travel time data of Mondays is shown in Figure 2.

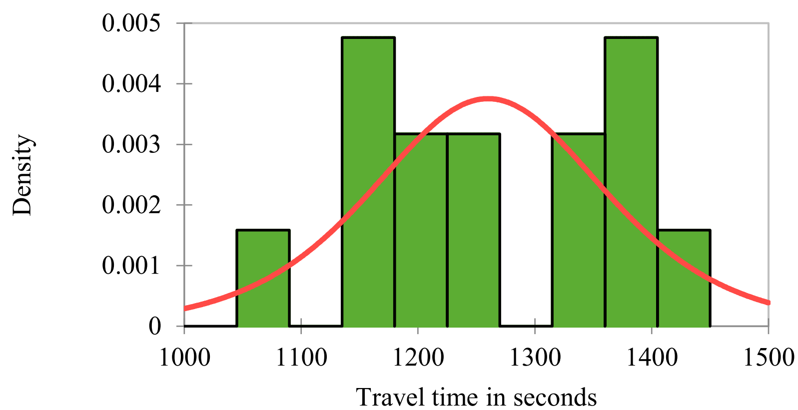

Figure 2. Histogram of travel time data of the route on Mondays

2.4 Comparative analysis

A non-parametric Wilcoxon signed-rank test is used to get deeper insights into data. The Wilcoxon signed-rank test verifies whether the two samples drawn have an equal median; in other words, whether the two-samples follow the same distribution. If they are unequal, it is evident that one sample is either left or right-skewed as compared to the other. A one-to-one comparison of travel time data samples at all the spatial-temporal scales was performed. Travel time data of each day-of-the-week were paired with every other day. For 7 days of the week, a total of 21 unique pairs were formed. Similarly, the travel times of each DTW (totally 13) were paired with every other DTW leading to a total of 78 pairs. All the pairs were compared using the Wilcoxon signed-rank test to identify possible similar sample pairs. The test interpretation and the results are presented subsequently.

Test Interpretation of Wilcoxon signed-rank test:
Null hypothesis: The two samples follow the same distribution
Alternative hypothesis: The distributions of the two samples are different
Level of significance =0.05 (p-value less than 0.05 will reject the null hypothesis and accept the alternative hypothesis and vice versa)
Statistic: $W = \sum_{i=1}^{N}[sgn(x_{2,i} - x_{1,i}).R_i]$

Where N is the sample size, *sgn* is the sign function, x_{1i} and x_{2i} is the corresponding instance in samples 1 and 2, and R_i is the rank assigned to the sorted difference of samples.

The p-values of the Wilcoxon signed-rank test [41] are tabulated in Table 3 (the significant p-values are in bold). The null hypothesis was accepted for 8 pairs out of 21 pairs compared. The p-values of comparing travel time among each DTW for all days of the week are tabulated in Table 4. The null hypothesis was accepted for 31 pairs of DTWs out of 78 pairs.

2.5 Modelling the data

The statistical distributions of any data portray the inbuilt variability in it. As discussed in the previous section, only travel time data were considered for further study. Consistent with the literature, the travel times at all aforementioned spatial-temporal scales were studied to understand the underlying patterns by fitting statistical distributions such as Gamma [42], GEV [36], Lognormal [29], Logistic [43], Normal [29], and Weibull [35] distributions.

Kolmogorov-Smirnov test [40] was used to check if a data sample is drawn from a population that fitted a given theoretical distribution.

Test Interpretation of Kolmogorov-Smirnov (KS) test:
Null hypothesis: The sample data follow a given distribution
Alternative hypothesis: The sample data does not follow the given distribution
Level of significance = 0.05 (p-value less than 0.05 will reject the null hypothesis and accept the alternative hypothesis and vice versa)
Statistics: $D = \max_{1 \le i \le N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right)$

Where the data has N ordered data points $Y_1, Y_2, ..., Y_N$ and F is the theoretical cumulative distribution.

2.5.1 Modeling based on the spatial aggregation levels

Initially, the KS test was conducted at spatial aggregations (route level and segment level). Travel time at 5 aggregation levels was tested to check whether they fitted the 6 distributions. The p-values are shown in Table 5.

Table 3. Day-of-the-week comparison

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
Monday	<0.0001					
Tuesday	<0.0001	**0.287**				
Wednesday	<0.0001	0.002	**0.101**			
Thursday	<0.0001	0.014	**0.103**	**0.864**		
Friday	<0.0001	0.036	**0.133**	**0.797**	**0.819**	
Saturday	0.001	<0.0001	0.000	0.023	0.034	**0.088**

Table 4. Departure time window comparison

	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19
8-9	<0.0001											
9-10	<0.0001	**0.138**										
10-11	<0.0001	**0.847**	**0.445**									
11-12	<0.0001	0.006	0.008	0.009								
12-13	<0.0001	<0.0001	0.010	0.000	0.045							
13-14	<0.0001	0.000	0.010	0.000	**0.067**	**0.825**						
14-15	<0.0001	**0.050**	**0.880**	0.032	**0.369**	0.003	0.016					
15-16	<0.0001	**0.053**	**0.431**	0.045	**0.797**	**0.052**	**0.121**	**0.591**				
16-17	<0.0001	<0.0001	0.001	<0.0001	0.003	**0.241**	**0.069**	0.002	0.011			
17-18	<0.0001	<0.0001	<0.0001	<0.0001	0.001	**0.342**	**0.207**	0.000	0.018	**0.735**		
18-19	<0.0001	0.000	0.013	0.002	**0.083**	**0.815**	**0.825**	0.013	**0.096**	**0.071**	**0.223**	
19-20	<0.0001	0.012	**0.214**	0.013	**0.889**	0.032	**0.053**	**0.552**	**0.770**	0.005	0.007	**0.074**

Table 5. p -values of the KS test at the spatial aggregation level

	Sample Size	Gamma	GEV	Log-normal	Logistic	Normal	Weibull
Route	800	0.348	<0.0001	0.158	**0.874**	0.278	0.006
S1	800	0.332	<0.0001	0.128	0.544	**0.821**	0.343
S2	800	0.522	<0.0001	0.202	0.348	**0.535**	0.382
S3	800	0.442	0.001	0.172	**0.853**	0.414	0.019
S4	800	0.042	<0.0001	0.118	**0.146**	0.001	<0.0001

*Note: The highest p-value for each aggregation is in bold.

2.5.2 Modeling based on spatial-temporal aggregation levels

Retaining the spatial aggregation levels, travel time data were further divided based on temporal aggregations as day-of-the-week and DTW. Modeling at each spatial-temporal aggregation through fitting distribution is presented as follows.

Day-of-the-week: The data segmented at the day-of-the-week temporal scale for each spatial scale were subjected to KS-test to check the adherence to the distributions selected in the study, and the p-values of the tests are given in Table 6.

Departure time window: The temporal scale of day-of-the-week that was further reduced to DTW of 1 h, was considered for segmenting data. The KS-test was repeated for all the segments of data to identify the best fit distribution, and to highlight whether the further reduced temporal scale had provided better insights into data. The *p-values* of each test are given in Table 7.

3. Results and Discussion

Advances in technology and its global acceptance and adoption by service providers have made public transit bus data readily available. This has opened an opportunity for researchers to gain insights into the travel time in road network cities with minimum ICT infrastructure related to traffic management and public transit service. The study was focused on understanding the underlying trend and variability of travel time and speed, and in this direction, a comprehensive study on a selected route in Tumakuru city was conducted. The raw location data were prepared and segmented at various spatial-temporal aggregation levels in line with the existing literature. Spatial data was segmented at route and segment levels, and preliminary analysis revealed that the aggregates of the segments demonstrated skewness compared to the route level. The segment level division was retained, and the data was further divided into two temporal scales: day-of-the-week and DTW.

The proportion of the segment lengths to route length and segment travel time to route travel time is depicted in Figure 3. The percentage of segment lengths to route length, when compared to individual travel time, implied that the travel times of segments 1 and 2 were high compared to segments 3 and 4. It was observed that the travel speed of segment 1 was the least. This may be due to congestion caused by land use pattern and delays at major intersections [44, 45] in segment 1, which influenced the speed of the segment to be the slowest of all. Segment 4 was the fastest having an OSU land use pattern, and the road used by the bus was a national highway with 3 lanes. Providing dedicated lanes [4] for buses and prioritizing them at the intersections [46, 47] can improve travel speed and time. The segments of the route belong to the national highway, state highway, municipality, and urban development authority. Any decision to improve the bus route needs common guidelines for the institutions owning the segments. Furthermore, they have to participate in integrated planning and execution. The Ministry of Urban Development, India has suggested the 'Urban Roads Code' [48] for this purpose and these institutions have to collaborate in this direction to handle the situation.

Further, to identify the pairs of spatial-temporal aggregations that were similar, the Wilcoxon signed-rank test was conducted. Knowing whether two samples are behaving similarly can aid in *optimizing* service operations, and reduce the overhead of *decision making* and *computation*. For example, the p-values of Wednesdays, Thursdays, and Fridays when mutually compared gave high values, indicating that they could be modeled together whereas data of Sundays, Mondays, Tuesdays, and Saturdays were different compared to all other days. Similarly, the consecutive DTWs of the morning (8-9, 9-10, and 10-11), noon DTWs (12-13 and13-14),(14-15 and 15-16), and evening DTWs (16-17, 17-18, and 18-19) could not be ruled out as being similar,

Table 6. p -values of the KS test at day-of-the-week temporal aggregation level

Day-of-the-week	Sample Size	Gamma	GEV	Log-normal	Logistic	Normal	Weibull
Route-level							
Sunday	92	0.797	0.042	0.730	0.829	**0.909**	0.653
Monday	126	0.454	0.628	0.595	**0.942**	0.211	0.048
Tuesday	126	0.683	0.145	**0.790**	0.598	0.455	0.130
Wednesday	124	0.551	0.481	0.642	**0.809**	0.375	0.092
Thursday	114	0.925	0.193	0.831	**0.991**	0.986	0.769
Friday	120	0.317	**0.777**	0.431	0.654	0.156	0.125
Saturday	98	0.025	0.037	0.095	**0.685**	0.001	0.001
Segment 1							
Sunday	92	**0.985**	0.308	0.976	0.951	0.953	0.767
Monday	126	0.297	0.376	0.509	**0.916**	0.051	0.021
Tuesday	126	0.936	0.089	0.870	0.886	**0.961**	0.731
Wednesday	124	0.978	0.185	0.904	0.991	**0.996**	0.943
Thursday	114	0.782	0.162	0.637	0.963	0.977	**0.999**
Friday	120	0.609	0.238	0.492	0.710	**0.756**	0.630
Saturday	98	0.981	0.287	0.920	0.978	**0.991**	0.811
Segment 2							
Sunday	92	0.326	0.149	0.204	**0.815**	0.662	0.434
Monday	126	0.432	0.175	0.289	**0.906**	0.802	0.864
Tuesday	126	**0.817**	0.454	0.774	0.696	0.745	0.589
Wednesday	124	0.816	0.354	**0.841**	0.775	0.738	0.827
Thursday	114	0.870	0.098	0.666	0.928	0.973	**0.978**
Friday	120	**0.889**	0.170	0.882	0.832	0.850	0.629
Saturday	98	0.595	0.033	0.436	0.642	0.679	**0.707**
Segment 3							
Sunday	92	0.267	**0.662**	0.382	0.316	0.116	0.040
Monday	126	0.800	0.594	**0.852**	0.704	0.713	0.549
Tuesday	126	0.363	0.190	0.345	**0.762**	0.168	0.054
Wednesday	124	0.516	0.094	0.385	**0.865**	0.793	0.607
Thursday	114	0.790	0.218	0.624	**0.870**	0.820	0.401
Friday	120	**0.980**	0.798	0.936	0.938	0.840	0.147
Saturday	98	0.951	0.474	0.976	**0.996**	0.824	0.317
Segment 4							
Sunday	92	0.007	<0.0001	0.004	0.003	0.000	<0.0001
Monday	126	0.655	0.039	0.426	0.698	**0.730**	0.165
Tuesday	126	0.023	**0.281**	0.056	0.197	0.003	0.001
Wednesday	124	0.090	0.044	0.048	**0.211**	0.107	0.010
Thursday	114	0.001	0.022	0.002	0.007	<0.0001	<0.0001
Friday	120	<0.0001	0.015	<0.0001	0.003	<0.0001	<0.0001
Saturday	98	0.014	0.017	0.027	**0.103**	0.003	0.000

*Note: The highest p-value for each aggregation is in bold.

Table 7. p -values of the KS test at departure time window temporal aggregation level

DTW	Sample Size	Gamma	GEV	Log-normal	Logistic	Normal	Weibull
Route level							
7-8	63	0.428	0.066	0.374	**0.788**	0.542	0.437
8-9	62	0.551	0.559	0.625	**0.750**	0.415	0.131
9-10	68	0.686	0.288	0.630	**0.900**	0.582	0.222
10-11	60	0.309	0.470	0.372	**0.617**	0.211	0.037
11-12	60	0.304	0.002	0.261	**0.881**	0.424	0.511
12-13	64	0.403	0.001	0.363	0.319	**0.412**	0.111
13-14	55	0.095	0.049	0.100	**0.111**	0.069	0.022
14-15	64	0.608	0.612	0.670	**0.815**	0.486	0.161
15-16	58	0.541	**0.925**	0.641	0.710	0.374	0.277
16-17	60	0.631	**0.973**	0.732	0.790	0.443	0.196
17-18	59	0.269	0.004	0.259	0.278	**0.296**	0.213
18-19	62	0.696	0.017	0.699	**0.749**	0.698	0.477
19-20	65	0.129	0.182	0.163	**0.346**	0.079	0.024
Segment1							
7-8	63	0.342	**0.955**	0.442	0.748	0.197	0.017
8-9	62	0.470	**0.935**	0.582	0.755	0.284	0.129
9-10	68	0.373	0.368	0.502	**0.686**	0.184	0.077
10-11	60	0.852	0.874	**0.944**	0.928	0.586	0.376
11-12	60	0.211	0.032	0.165	0.412	0.341	**0.776**
12-13	64	0.547	0.427	0.634	**0.757**	0.392	0.240
13-14	55	0.023	0.000	0.031	0.063	0.014	0.036
14-15	64	0.618	0.095	0.553	0.602	**0.636**	0.443
15-16	58	0.811	0.609	0.799	**0.849**	0.829	0.527
16-17	60	0.902	0.438	0.943	**0.977**	0.774	0.364
17-18	59	0.146	**0.231**	0.163	0.160	0.119	0.148
18-19	62	0.651	0.225	**0.701**	0.549	0.557	0.433
19-20	65	0.263	<0.0001	0.277	**0.330**	0.243	0.193
Segment 2							
7-8	59	0.772	0.090	0.673	**0.965**	0.932	0.939
8-9	64	0.989	0.951	0.986	**0.993**	0.989	0.802
9-10	69	0.745	0.005	**0.787**	0.712	0.656	0.600
10-11	64	0.888	0.845	**0.927**	0.866	0.709	0.563
11-12	62	0.797	0.216	0.657	**0.977**	0.862	0.600
12-13	69	0.154	0.000	0.111	**0.597**	0.332	0.384
13-14	52	0.091	**0.162**	0.150	0.146	0.029	0.016
14-15	58	0.775	**0.863**	0.846	0.726	0.503	0.268
15-16	54	0.719	0.047	**0.758**	0.591	0.639	0.351
16-17	64	0.517	**0.915**	0.634	0.550	0.320	0.237
17-18	62	0.164	0.001	0.116	0.502	0.309	**0.601**
18-19	60	0.548	0.159	**0.562**	0.491	0.521	0.473
19-20	63	0.032	**0.177**	0.043	0.107	0.015	0.015

Table 7. p -values of the KS test at departure time window temporal aggregation level (continued)

DTW	Sample Size	Gamma	GEV	Log-normal	Logistic	Normal	Weibull
Segment 3							
7-8	56	0.826	0.117	**0.849**	0.784	0.780	0.832
8-9	66	0.348	0.151	0.399	**0.661**	0.243	0.065
9-10	68	**0.670**	0.305	0.543	0.603	0.454	0.220
10-11	62	0.147	0.081	0.099	**0.296**	0.091	0.032
11-12	59	0.368	0.219	0.299	0.340	0.442	**0.479**
12-13	68	0.238	**0.364**	0.283	0.243	0.156	0.125
13-14	57	0.216	**0.396**	0.298	0.203	0.108	0.065
14-15	63	0.437	0.049	0.339	0.802	0.682	**0.933**
15-16	55	0.530	**0.664**	0.631	0.627	0.361	0.423
16-17	63	0.986	0.566	**0.993**	0.961	0.949	0.655
17-18	61	0.035	**0.179**	0.051	0.142	0.015	0.006
18-19	60	0.489	0.001	0.420	**0.819**	0.638	0.724
19-20	62	0.366	0.401	**0.439**	0.382	0.248	0.145
Segment 4							
7-8	54	0.014	0.046	0.010	0.032	0.030	0.042
8-9	65	0.945	0.269	0.893	0.953	**0.994**	0.976
9-10	69	**0.075**	0.038	0.050	0.070	0.154	0.126
10-11	63	0.091	0.064	**0.120**	0.092	0.051	0.020
11-12	61	0.010	0.003	0.007	0.005	0.010	0.004
12-13	66	0.283	0.281	**0.412**	0.274	0.136	0.071
13-14	56	0.360	**0.617**	0.473	0.435	0.198	0.140
14-15	62	0.035	0.123	0.054	**0.195**	0.018	0.005
15-16	57	0.255	0.210	0.303	**0.492**	0.040	0.026
16-17	62	0.059	**0.297**	0.159	0.288	0.005	0.008
17-18	60	0.039	0.006	0.050	**0.070**	0.022	0.026
18-19	63	0.196	**0.454**	0.254	0.285	0.115	0.061
19-20	62	0.002	0.000	0.004	0.011	0.000	0.003

*Note: The highest p-value for each aggregation is in bold number.

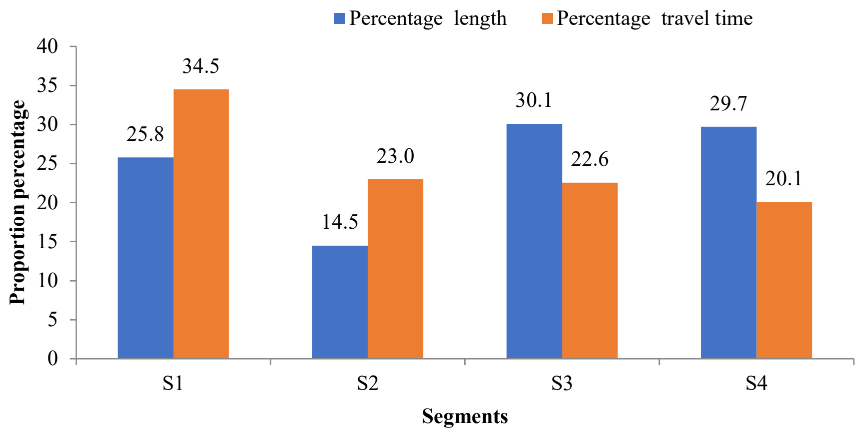

Figure 3. Proportion of segment length to its average travel time in percentage

but 7-8 DTW was not similar to any other DTW. For a further study, temporal aggregation can be redesigned with possible combination of several DTWs for modeling and forecasting.

The travel time data were modelled using the statistical distributions identified through the literature survey. The distributions of data at all spatial-temporal aggregations were tested using the KS-test and it was found that Logistic distribution was the best distribution that best suited travel time. The number of cases in which each distribution selected in the study best suited data is summarized in Figures 4 and 5. In Figure 4, the results of the day-of-the-week temporal aggregation are summarized and those of DTW are shown in Figure 5. For both temporal aggregations, segment 4 demonstrated diverse results, where a few aggregations did not fit any distribution, and included were Sundays, Thursdays, and Fridays and DTWs 7-8, 11-12, and 19-20. The land use pattern of segment 4 was OSU and the road in this segment was part of a national highway; thus, the segment was different from all other segments. Hence, this segment has to be treated individually for further analysis and forecasting. From the results, it can be also inferred that for day-of-the-week temporal aggregation, the data were best fitted by the Logistic or normal distribution, demonstrating that higher temporal aggregation data is more symmetrical, whereas at lower temporal aggregations (DTW), a few samples fitted the GEV as the second-best suitable distribution. The GEV distribution can handle extreme travel times during the afternoon and evening DTWs. The top 1 and top 2 distributions that fit overall at all spatial-temporal aggregations are summarized in Figure 6. It can be concluded that Logistic is suitable at all spatial-temporal levels, or individual distribution can be considered for each spatial-temporal aggregation.

The allotted travel time for the route was 1200 s but it was observed that on weekdays (except for DTW 7-8), the travel time exceeded the allotted time in most cases. Based on the observations, the allocation of varying trip times at various spatial-temporal aggregations can be done during timetable preparation. The results of modeling the data by fitting to various distributions can be applied for optimizing service operations in the future.

Overall, the major finding of the research and recommendations to the service providers are as follows:

- Most of the spatial-temporal aggregation data fit Logistic distributions, indicating symmetry but long and fat tails. Hence, while modeling in the future these cases need to be addressed, and also for the activation function for a few ML neural network models, the logistic function can be a possible activation function.

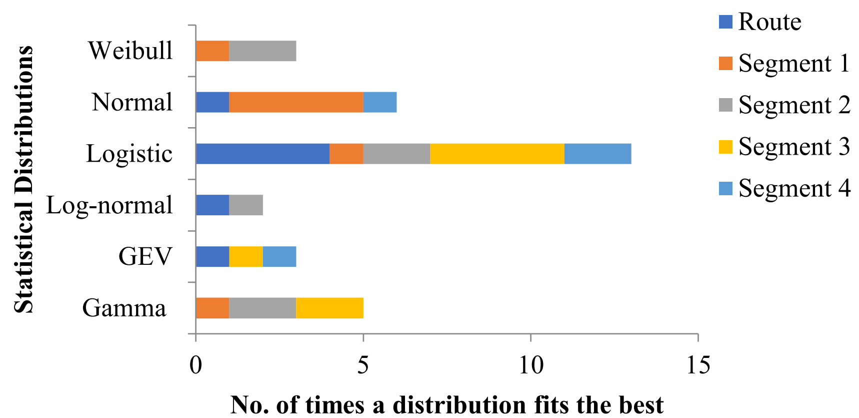

Figure 4. Count of best fit distribution at day-of-the-week level

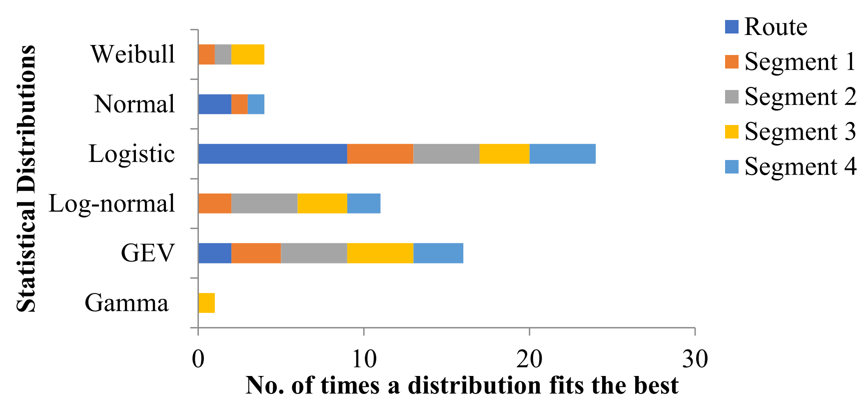

Figure 5. Count of best fit distribution at 1-h DTW level

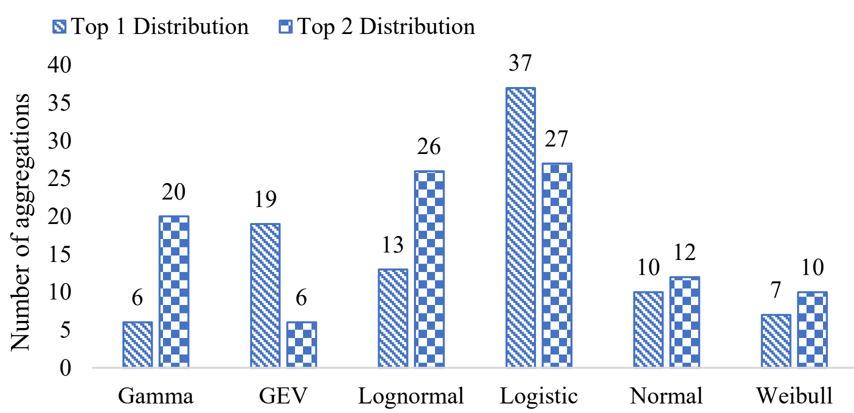

Figure 6. Top 1 and top 2 distribution that fit the data at all aggregation levels

- A few aggregations fit best to the GEV and Weibull distributions, indicating the presence of extreme values. Models that will be developed in the subsequent research work should address these extreme values.
- The trip aggregates were split based on the spatial-temporal characteristics and the results emphasized that better insights were gained at lower spatial aggregation levels. Future models must emphasize spatial-temporal characteristics for model development.
- The allotted time for a trip according to the schedule was 1200 s and from the preliminary analysis, it was found that in most cases the actual travel time exceeded the allotted time. Hence, allocating varying trip times [49] based on day-of-the-week and time of the day when generating the timetable is recommended.

Variability analysis was performed for a vital route passing through major intersections of the city. However, the study can be extended to all the bus routes and the overall variability of travel time in the city can be analyzed. The insights gained can be useful to service providers to understand travel time behavior and implement solutions based on the insights. The study was conducted in a tier-2 city with buses as a major public transit mode. There are several cities across India and Asia that are similar to this city, and the approach carried out in this work can be extended to those cities.

4. Conclusions

Promoting public transit services is the need of the day; these are sustainable and eco-friendly alternatives for private modes of transport. The reliability of the public transit system plays an important role in bringing a modal shift from private to public modes. In this regard, the variability of the public transit buses of Tumakuru city, a smart city in the southern region of India was studied. The travel time data were divided based on spatial-temporal aggregations. The results of *segment*-level aggregation proved to provide better insights into data. Retaining the spatial aggregation, travel time data was further aggregated at the day-of-the-week and DTW scales, and the DTW results represented the data well; overall, spatial-temporal aggregations gave better insights into data. A one-to-one comparison of spatial-temporal aggregations was performed using the Wilcoxon signed-rank test to discover similar/dissimilar datasets of the study. It was found that there were a few aggregations as similar pairs that can be analyzed collectively in the future. Based on the literature survey, six distributions, Gamma, GEV, Lognormal, Logistic, Normal, and Weibull were identified, and the travel time data were compared against these theoretical distributions using the KS test. It was found that the Logistic distribution fitted the best compared to all other distributions at all spatial-temporal aggregation levels, and the Lognormal and GEV distributions also fitted better for a few aggregation levels. For further travel time analysis, Logistic distribution should be considered by operations planners and researchers. Other spatial aggregations such as bus-stop level and intersection levels, and temporal aggregations such as peak off-peak hours can be explored to observe variability in the future. This work can also be extended to analysis of all city routes and to assessment of variability at the city level.

5. Acknowledgements

Special thanks to Tumakuru Smart City Limited for sharing the GPS logs of the city service buses in the study area for analysis.

References

[1] El-Geneidy, A., Hourdos, J. and Horning, J., 2009. Bus transit service planning and operations in a competitive environment. *Journal of Public Transportation*, 12(3), 39-59, https://doi.org/10.5038/2375-0901.12.3.3.

[2] Shaji, H.E., Tangirala, A.K. and Vanajakshi, L., 2020. Prediction of trends in bus travel time using spatial patterns. *Transportation Research Procedia*, 48, 998-1007, https://doi.org/10.1016/j.trpro.2020.08.128.

[3] Kumar, B.A., Vanajakshi, L. and Subramanian, S.C., 2018. A hybrid model based method for bus travel time estimation. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 22(5), 390-406, https://doi.org/10.1080/15472450.2017.1378102.

[4] Raj, G.G., Sekhar, C.R. and Velmurugan, S., 2013. Micro simulation based performance evaluation of Delhi bus rapid transit corridor. *Procedia - Social and Behavioral Sciences*, 104, 825-834, https://doi.org/10.1016/j.sbspro.2013.11.177.

[5] Kathuria, A., Parida, M. and Chalumuri, R.S., 2020. Travel-time variability analysis of bus rapid transit system using GPS data. *Journal of Transportation Engineering, Part A: Systems*, 146(6), https://doi.org/10.1061/jtepbs.0000357.

[6] Munhoz, P.A.M.S.A., Dias, F.D.C., Chinelli, C.K., Guedes, A.L.A., dos Santos, J.A.N., Silva, W.D.S., and Soares, C.A.P., 2020., Smart mobility: The main drivers for increasing the intelligence of urban mobility. *Sustainability*, 12(24), https://doi.org/10.3390/su122410675.

[7] Mazloumi, E., Rose, G., Currie, G. and Sarvi, M., 2011. An integrated framework to predict bus travel time and its variability using traffic flow data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 15(2), 75-90, https://doi.org/10.1080/15472450.2011.570109.

[8] Comi, A., Nuzzolo, A., Brinchi, S. and Verghini, R., 2017. Bus travel time variability: Some experimental evidences. *Transportation Research Procedia*, 27, 101-108, https://doi.org/10.1016/j.trpro.2017.12.072.

[9] Rahman, M.M., Wirasinghe, S.C. and Kattan, L., 2018. Analysis of bus travel time distributions for varying horizons and real-time applications. *Transportation Research Part C: Emerging Technologies*, 86, 453-466, https://doi.org/10.1016/j.trc.2017.11.023.

[10] Ashwini, B.P., Savithramma, R.M. and Sumathi, R., 2022. Artificial intelligence in smart city applications: An overview. *6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, May 25-27, 2022, pp. 986-993.

[11] Ashwini, B.P., Sumathi, R. and Sudhira, H.S., 2023. Analyzing travel time reliability of a bus route in a limited data set scenario: A case study. *International Journal of Intelligent Systems and Applications in Engineering*,11(2), 30-39.

[12] Liu, Y., Zhang, H., Jia, J., Shi, B. and Wang, W., 2023. Understanding urban bus travel time: Statistical analysis and a deep learning prediction. *International Journal of Modern Physics B*, 37(4), https://doi.org/10.1142/S0217979223500340.

[13] Ashwini, P.B. and Sumathi, R., 2020. Data sources for urban traffic prediction: A review on classification, comparison and technologies. *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, December 3-5, 2020, pp. 628-635. https://doi.org/10.1109/ICISS49785.2020.9316096.

[14] Billings, D. and Yang, J.S., 2006. Application of the ARIMA models to urban roadway travel time prediction- A case study. *Proceedings of IEEE International Conference on Systems, Man and Cybernetic*s, Taipei, Taiwan, October 8-11, 2006, pp. 2529-2534, https://doi.org/10.1109/ICSMC.2006.385244.

[15] Al-Suleiman, T.I., Al-Torkman, R. and Mujalli, R., 2023. Performance analysis of public bus transport services in rural areas. *Jordan Journal of Civil Engineering*, 17(2), 272-292, https://doi.org/10.14525/JJCE.v17i2.09.

[16] Hans, E., Chiabaut, N., Leclercq, L. and Bertini, R.L., 2015. Real-time bus route state forecasting using particle filter: An empirical data application. *Transportation Research Procedia*, 6, 434-447, https://doi.org/10.1016/j.trpro.2015.03.033.

[17] Cristóbal, T., Padrón, G., Quesada-Arencibia, A., Alayón, F., de Blasio, G.D. and Garc, C.R., 2019. Bus travel time prediction model based on profile similarity. *Sensors*, 19(3), https://doi.org/10.3390/s19132869.

[18] Liu, D., Sun, J. and Wang, S. 2020. Bus time: Which is the right prediction model for my bus arrival time? *5th IEEE International Conference on Big Data Analytics (ICBDA),* Xiamen, China, May 8-11, 2020, pp. 180-185.

[19] Ashwini, B.P., Sumathi, R. and Sudhira, H.S. 2022. Bus travel time Prediction: A comparative study of linear and non-linear machine learning models. *Journal of Physics: Conference Series*, 2161(1), https://doi.org/10.1088/1742-6596/2161/1/012053.

[20] Chien, S.I.J., Ding, Y. and Wei, C. 2002. Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering*, 128(5), 429-438, https://doi.org/10.1061/(ASCE)0733-947X(2002)128:5(429).

[21] Jeong, R. and Rilett, R., 2004. Bus arrival time prediction using artificial neural network model. *7th International IEEE Conference on Intelligent Transportation Systems,* Washington, USA, October 3-6, 2004, pp. 988-993.

[22] Han, Q., Liu, K., Zeng, L., He, G., Ye, L. and Li, F., 2020. A bus arrival time prediction method based on position calibration and LSTM. *IEEE Access*, 8, 42372-42383, https://doi.org/10.1109/ACCESS.2020.2976574.

[23] Chen, B., Bai, R., Li, J., Liu, Y., Xue, N. and Ren, J., 2023. A multiobjective single bus corridor scheduling using machine learning-based predictive models. *International Journal of Production Research*, 61(1), 131-145, https://doi.org/10.1080/00207543.2020.1766716.

[24] Yang, M., Chen, C., Wang, L., Yan, X. and Zhou, L., 2016. Bus arrival time prediction using support vector machine with genetic algorithm. *Neural Network World*, 26(3), 205-217, https://doi.org/10.14311/NNW.2016.26.011.

[25] Islek, I. and Oguducu, S.G., 2019. Use of LSTM for short-term and long-term travel time prediction. [online] Available at: https://ceur-ws.org/Vol-2482/paper24.pdf.

[26] Yu, B., Yang, Z.-Z., Chen, K. and Yu, B., 2010. Hybrid model for prediction of bus arrival times at next station. *Journal of Advanced Transportation*, 44(3), 193-204.

[27] Savithramma, R.M., Ashwini, B. and Sumathi, R., 2022. Smart mobility implementation in smart cities: A comprehensive review on state-of-art technologies. *4th International Conference on Smart Systems and Inventive Technology*, Tirunelveli, India, January 20-22, 2022, pp. 10-17.

[28] Taylor, M.A.P., 1982. Travel time variability - The case of two public modes. *Transportation Science*, 16(4), 507-521, https://doi.org/10.1287/trsc.16.4.507.

[29] Mazloumi, E., Currie, G. and Rose, G., 2010. Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering*, 136(7), 623-631, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000126.

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

[31] Chepuri, A., Ramakrishnan, J., Arkatkar, S., Joshi, G. and Pulugurtha, S.S., 2018. Examining travel time reliability-based performance indicators for bus routes using GPS-based bus trajectory data in India. *Journal of Transportation Engineering, Part A: Systems*, 144(5), https://doi.org/10.1061/jtepbs.0000109.

[32] Birr, K., Jamroz, K. and Kustra, W., 2014. Travel time of public transport vehicles estimation. *Transportation Research Procedia*, 3, 359-365, https://doi.org/10.1016/j.trpro.2014.10.016.

[33] Ashwini, B.P. and Sumathi, R., 2022. Variability analysis of public transit bus travel time using GPS logs: A case study. *Proceedings of the Third International Conference on Information Management and Machine Intelligence. Algorithms for Intelligent Systems*, Jaipur, India,

December 23-24, 2021, pp 567-577.

[34] Büchel, B. and Corman, F., 2020. Review on statistical modeling of travel time variability for road-based public transport. *Frontiers in Built Environment*, 6, https://doi.org/10.3389/fbuil.2020.00070.

[35] Wong, E. and Khani, A., 2018. Transit delay estimation using stop-level automated passenger count data. *Journal of Transportation Engineering, Part A: Systems*, 144(3), https://doi.org/10.1061/jtepbs.0000118.

[36] Harsha, M.M., Mulangi, R.H. and Kumar, H.D.D., 2020. Analysis of bus travel time variability using automatic vehicle location data. *Transportation Research Procedia*, 48, 3283-3298, https://doi.org/10.1016/j.trpro.2020.08.123.

[37] Directorate of Urban Land Transport, Government of Karnataka, India, 2013. *TUMKUR CITY BUS SERVICE- Evaluation Report.* [online] Available at: https://dult.karnataka.gov.in/uploads/media_to_upload1637327838.pdf.

[38] Directorate of Urban Land Transport, Government of Karnataka, India, 2014. *Comprehensive Mobility Plan for CMA Final Report*. [online] Available at: https://www.cmdachennai.gov.in/pdfs/ComprehensiveMobilityPlan-CMA.pdf.

[39] Fadlallah, B., Fadlallah, A., Razafsha, M., Karnib, N., Wang, K. and Kobeissy, F., 2019., Robust detection of epilepsy using weighted-permutation entropy: Methods and analysis In: F. Kobeissy, A. Alawieh, F.A. Zaraket and K. Wang, eds. *Leveraging Biomedical and Healthcare Data.* Cambridge: Academic Press, pp. 91-106.

[40] Ashwini, B.P., Sumathi, R. and Sudhira, H.S., 2022. A dynamic model for bus arrival time estimation based on spatial patterns using machine learning. *International Journal of Engineering Trends and Technology*, 70(9), 185-193, https://doi.org/10.14445/22315381/IJETT-V70I9P219.

[41] Suarez, R., Faghri, A. and Li, M., 2014. Evaluation of the accuracy and automation of travel time and delay data collection methods. *Journal of Transportation Technologies*, 4(1), 72-83, https://doi.org/10.4236/jtts.2014.41007.

[42] Dai, Z., Ma, X. and Chen, X., 2019. Bus travel time modelling using GPS probe and smart card data: A probabilistic approach considering link travel time and station dwell time. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 23(2), 175-190, https://doi.org/10.1080/15472450.2018.1470932.

[43] Durán-Hormazábal, E. and Tirachini, A., 2016. Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in Santiago, Chile. *Research in Transportation Economics*, 59, 26-39, https://doi.org/10.1016/j.retrec.2016.06.002.

[44] Savithramma, R.M. and Sumathi, R., 2022. Comparative analysis of waiting delay estimation models for signalized intersections: A case study Tumakuru City. *Proceedings of the Third International Conference on Information Management and Machine Intelligence. Algorithms for Intelligent Systems*, Jaipur, India, December 23-24, 2021, pp. 579-589.

[45] Savithramma, R.M., Sumathi, R. and Sudhira, H.S., 2022. A comparative analysis of machine learning algorithms in design process of adaptive traffic signal control system. *Journal of Physics: Conference Series*, 2161(1), https://doi.org/10.1088/1742-6596/2161/1/012054.

[46] Girijan, A., Vanajakshi, L.D. and Chilukuri, B.R., 2021. Dynamic thresholds identification for green extension and red truncation strategies for bus priority. *IEEE Access*, 9, 64291-64305, https://doi.org/10.1109/ACCESS.2021.3074361.

[47] Thodi, B.T., Chilukuri, B.R. and Vanajakshi, L., 2022. An analytical approach to real-time bus signal priority system for isolated intersections. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 26(2), 1-23, https://doi.org/10.1080/15472450.2020.1797504.

[48] Indian Road Congress, 2013. *Urban Road Code of Practice*. [online] Available at: https://mohua.gov.in/upload/uploadfiles/files/Urban%20Road%20Code%20of%20Practice.pdf.

[49] Wu, Y., Tang, J. and Zhang, Y., 2013. Optimizing reliable timetable for bus transit network: Model formulation and solution. *Proceedings of 25th Chinese Control and Decision Conference*, Guiu\yang, China, May 25-27, 2013, pp. 1835-1840.