2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)
Manipal University, Jaipur. Jan 19-20, 2023

Extracting potential Travel time information from raw GPS data and Evaluating the Performance of Public transit – a case study in Kandy, Sri Lanka

Shiveswarran Ratneswaran
Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka
shiveswarran.22@cse.mrt.ac.lk

Uthayasanker Thayasivam
Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka
rtuthaya@cse.mrt.ac.lk

2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT) | 978-1-6654-5357-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCT56969.2023.10075789

***Abstract*— The widespread use of location-enabled devices on public transportation vehicles produces a huge amount of geospatial data. The primary objective of this research study is to build a solution framework that can process a large amount of geospatial data obtained from GPS (Global Positioning System) receivers fixed on different buses on different routes, preprocess, clean, and transform that data for analysis. There are various challenges associated with the processing of GPS data, like discontinuities, non-uniformities, poor network coverage, and human errors. This study proposes two novel, simple algorithms to extract bus trip and bus stop sequences, from the crude raw data, incorporating those challenges. Moreover, the dwell times at the bus stops are estimated solely using this GPS data in three different possible scenarios in the data filtering process. When considering the previous related studies in this area, the proposed approaches are applied to GPS data obtained at a medium sample rate (for example, 15 seconds) for heterogeneous traffic conditions, and also with a unique dwell time estimation process. In addition, statistical methods are implemented to analyse a variety of novel public transit-system performance metrics, such as (i) excess journey time (EJT); (ii) excess dwelling time (EDT); (ii) excess running time (ERT); and (iv) segment idle time ratio (SITR), at different time horizons, where these metrics are developed in the absence of schedule data. These metrics facilitate the transport authorities in real-time bus monitoring, evaluating their performance, and identifying inappropriate driving behaviours. A detailed explanation is provided through a case study of two main routes in the Kandy district of Sri Lanka.**

Keywords—Travel time, GPS data, public transit, performance metrics

I. Introduction

Automatic Vehicle Location (AVL) systems identify and provide the location of moving vehicles with timestamps, which is a core source of Intelligent Transportation Systems (ITS). With the increased use of location-enabled devices (e.g., GPS (Global Positioning System)) on public transportation vehicles, a huge amount of bus trajectory data is being produced. However, the raw GPS data is crude. One major objective of this data is to estimate and analyze the travel time of public transits. There are several challenges in using GPS data for estimation and analysis. The key challenges are, noise due to the precision of the sensor, missing observations within a GPS trace, segmentation by terminals, and localization to point of interest. For this purpose, an efficient data pre-processing framework comprising solutions to the above challenges is required.

When considering Kandy city in Sri Lanka, public transport is the primary means of access to the city with 65% of the whole commuters entering the city for various purposes [1]. Kandy's traffic volume and congestion level have already significantly increased. A slight increase in the percentage of private cars on the road will create huge traffic congestion in the city. Therefore, there is a requirement for implementing management strategies with the help of performance metrics to improve public transportation reliability from both commuters' and operators' perspectives. In addition to that, Advanced Public Transport Systems (APTS) applications such as bus arrival time information in real-time, attract more passengers to bus transit. For the above two requirements, historical data on the travel time of buses should be available. With regard to the Strategic Cities Development Program (SCDP) by the Ministry of Urban Development, Sri Lanka, for solving the traffic problem within the Kandy city area, a pilot project was implemented by installing GPS sensors on 1100 buses in Kandy city. These buses would take 124 intra-provincial routes starting from the three Kandy central bus terminals located in the city core. Among them, two strategically important unique bus routes 654 and 690 were selected for this research based on the factors like heavy demand, high frequency, and running on the two main highways passing through Kandy city. These two routes carry a vast amount of traffic entering Kandy city.

GPS sensors are fixed in buses, which have unique device identification codes and emit signals once the engine is switched on and continue at a preset frequency (for example, every 15 seconds) until it goes off. Every GPS observation has spatio-temporal characteristics like the latitude and longitude values of the location with timestamps. In order to analyze this data usefully, the following key tasks are to be performed.

- Identify the beginning and end of the trips (terminals)
- Identify the dwellings at bus stops along the journey
- Handle outliers in the extracted trips
- Develop data quality metrics for the GPS data

There are various challenges in the above task when applying them to the available GPS dataset. The GPS data are a stream of vehicle location information. They can exhibit discontinuities or nonuniformities due to poor network coverage, especially in mountainous terrains. They can be coupled with actual locations on the digital map, but with system accuracy. The higher frequency of GPS records severely affects the matching with bus stops (localization to points of interest). Furthermore, the discontinuities in the data streaming especially at the end terminals cause to discard of some trips even before capturing them. These GPS traces may not be exactly distributed along the predefined bus route due to some personal trips other than the operational schedule of buses. Extracted trips may show some abnormal travel time values as outliers because of bus breakdowns, road construction works, accidents, extreme traffic congestion, public parades, road diversions or disruptions due to landslides, heavy rain, etc. In most of the previous studies, bus dwell times are not separately estimated, rather they are implicitly included in the total travel time of buses. An estimation of dwell times at each stop is required to analyse the bus stop pattern. Above all, extracting dwell time information solely from the GPS traces is a mammoth task.

Generally, public transit system performance metrics comprise a vast area including travel time reliability or service effectiveness, safety and security, vehicle utilization, cost efficiency etc. However, in this study, a novel methodology to evaluate public transportation-system performance using metrics such as 1) Excess journey time (EJT); 2) Excess dwelling time (EDT); 3) Excess Running time (ERT) 4) Segment Idle Time Ratio (SITR), which are focused on travel time are developed for the local conditions in Kandy, where there is no proper time scheduling at the bus stop level and no strict schedule adherence. The ERT and EDT indicate the additional time taken by buses on the road and at the stops more than a defined threshold value on a similar day and time. The SITR reflects the driver's behaviour and other possible reasons that cause the bus to stop or idle in the middle of its journey. These metrics help to observe the current performance when there is a stream of GPS data is fetched and immediate measures against the drivers who deliberately waste the valuable time of passengers, for the sake of their benefit, can be taken based on a reference value. The poor reliability of public transit towards passengers, creates dissatisfaction and hesitation, causing mode shift to them.

A variety of data sources can be used to derive public transit metrics such as automatic vehicle location (AVL) tracking, automated passenger counts (APC), automatic fare collection (AFC) devices, at bus stops and on-board etc. However, developing performance metrics solely based on GPS data is a hugely challenging task tackled in this research study. This paper provides three main contributions:

- Novel simple algorithms for extracting bus trips, bus stop sequences, and dwell times from the raw GPS data
- A novel solution framework and public transportation system performance metric system
- An open-source code that can be applied to any bus GPS datasets and a GPS trajectory dataset of buses in Kandy

The remainder of this paper is organized as follows: Section 2 discusses related works in bus GPS data processing and transportation system performance metrics, Section 3 describes the proposed novel approach, and Section 4 shows the results and a discussion. Finally, Section 5 presents conclusions and areas for future work.

II. Literature review

Earlier, techniques like loop detectors [2],[3], observation vehicles, or automatic vehicle recognition using video cameras[4] have been used to estimate journey times. Recent studies suggest that due to their limited coverage area, low data quality, and high implementation cost of the devices and instruments, those techniques may not be adequate to accurately estimate journey time [5]. With the evolution of AVL systems, a great deal of research is happening to estimate the travel time of probe vehicles (ex: bus transits, taxis, trucks, etc.), fixed with GPS sensors. Vehicle trajectories can be extracted from the GPS data by matching them on the digital map. The vehicle map-matching algorithms are categorized into four types in a survey[6]; (i) Geometric principle (concerned with the shape and geometric information of road segments integrating with Spatio-temporal proximity[7]); (ii) Topology (focus on network connectivity and continuity using local search algorithms[8]); (iii) Probability statistics (finding a matching path based on uncertainty/maximum probability in searching candidate segment or point of interest[9]); (iv) Advanced models (relatively complex using fuzzy logic methods, reinforcement learning[10]).

Furthermore, GPS data processing techniques are determined by the sampling frequency and type of probe vehicles. As per [6], the sampling data rate of greater than 30 seconds is called 'High-frequency sampling' while less than 10 seconds is 'Low-frequency sampling'. Low-frequency sampling data points are sparsely distributed, hence the probabilistic approach is suitable for map matching[11]. On the other hand, most current map-matching methods are suitable for high-frequency sampling data, despite of costing for big data transmission and storage[12]. But the available Kandy GPS data is at the sampling rate of 15 seconds, which is not aligned to whether a high or low rate. Therefore, a different approach needs to be utilized in trip extraction. Thereupon, bus stops should be identified in the extracted trajectory data. Bus stops along with their waiting time distribution were extracted using a density-based Spatio-temporal clustering algorithm and novel classification algorithm by [13] & [14]respectively.

In Sri Lanka, there is very little research done in the Bus GPS data processing. The authors of [15], propose algorithms based on geometric map matching to extract bus trips and estimate total travel time. Another study[16] expressed a probabilistic map-matching technique based on Hidden Markov Model (HMM). Yet, there are not any methodologies proposed in the Sri Lankan context to estimate dwell time at bus stops using GPS location and speed data, where there are chaotic road conditions and irregular schedules of buses.

Earlier studies estimate bus travel time In many developed cities throughout the world, the smart card-based AFC system has taken over as the primary way to collect fares for urban buses and rail transit[17]. It generates plenty of data related to bus arrival time at stops and journey travel times with higher accuracy. Much research has been done recently to estimate bus running and dwell times with the combination of AFC and AVL data[18], [19]. But, in the case of developing countries like Sri Lanka, such systems have not been introduced.

Hereafter, with the estimated travel times, various studies have been done to evaluate the performance of public transportation systems and their reliability. The most widely used transit performance metric is on-time performance, which is simply the per cent of schedule deviations within a threshold value. [20]. As well as additional time passengers spend waiting for buses and travelling in buses more than the schedule, is termed additional bus stop time (ABST) and additional travel time (ATT) respectively. Authors of [21] combined the above two aspects and used a metric for travel time performance, termed customer journey time performance (CJTP). They state that the data used for the metric are based on GPS data and an electronic fare collection system. In order to capture the variability in wait and travel times that tends to raise the inutility of transit travel, a new measure called the journey time buffer index (JTBI) was developed [22]. Authors of [23] further added that weekends/workdays, weather and time period of the day all have an impact on travel time variability and they incorporated them and added a threshold value for the travel time variability. The above work only targets customer-oriented mechanisms for performance evaluation. In contrast, our proposal is for the utilization by the public transportation authorities to control and regulate. Yet, it is necessary to develop public transit performance metrics solely based on GPS data, especially for the Sri Lankan local conditions.

III. Methodology

This section explains the data preparation steps proposed algorithms to extract the potential travel time-related information and finally the metrics that can be used to assess the performance of public transit.

A. Data gathering

The data collection was performed using AVL system. In this system, GPS receivers are fixed in the buses, and they are interfaced with a Global System for Mobile communications (GSM) modems and emit signals at a predefined sample rate to the servers located in an office remotely.

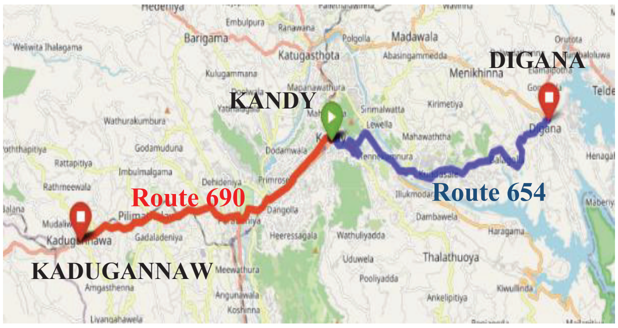

Fig. 1. The topographical view of the two bus routes (654&690) selected for the study

The GPS dataset used in this study was obtained from the project SCDP. The data were available from 1st October 2021 to 28th February 2022 (5 months) for bus lines 654 (Kandy-Digana) and 690 (Kandy-Kadugannawa) were selected for this study. The 16.6 km long route 654 and 15km long route 690 are located over the two main highways that connect Kandy city with East and Colombo, the capital city of Sri Lanka respectively. Fig. 1. shows that these two highways are among the three available options to enter the business city, Kandy.

B. GPS data processing

The overview of the approach of data processing to estimate the bus travel time, dwell time at bus stops and running time in between stops is shown in the Fig.2

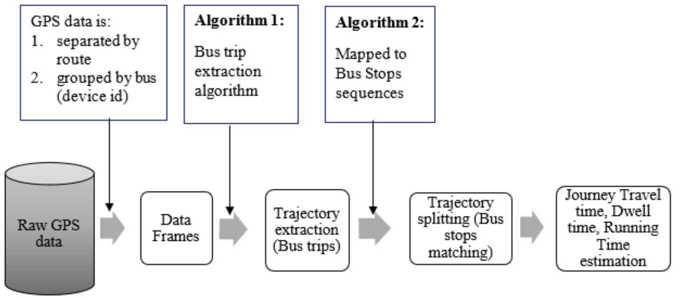

Fig. 2. The procedure to process the raw GPS data in order to estimate bus travel time related information

The crude raw GPS data was filtered out from the server for the required time window. Some data preprocessing techniques need to be carried down before deep diving into trip extraction. First, the removal of records with errors, such as (i) Sensor errors or precision errors caused some records with zero values for longitude and latitude columns, (ii) Erroneous dates (e.g., 01/01/2004) which are out of time window, might be caused by a device reset. Next, GPS records were sorted by date and time in chronological order, to identify trips sequentially. Later on, data were grouped by buses and *algorithm 1* (Bus trip extraction algorithm) was applied to filter out every trip and its trajectory coordinates. When slicing down further by localizing with the bus stops by applying *Algorithm 2* (Bus stop sequence matching), dwell time at the stops can be estimated. In scrutiny of the arrival and departure times at each bus stop, the running times in each route segment are estimated.

a) Bus trip extraction

At the very beginning, all route-wise data frames are converted to 'Geo-Data Frame' using 'Geo-Pandas' library based on python. This process converts latitude and longitude values as a single POINT coordinate and projects them in local Coordinate reference system EPSG:5234 of Sri Lanka. Using a suitable GPS data visualization package like Folium, we can project the coordinates on Open Street Map (OSM) to explore how the records are spread and to gain some insights and overview. There can be records outside the study area due to poor coverage in GPS positioning, human errors like drivers' mishandling the device, bus might be running out of route, for example private trips, parking, leaving, or returning to garage. The GPS receivers starts when the engine switched on, continuously emit signals and stop when the engine switched off. However, in real situations, the drivers without switching off the engine, wait in the terminal to begin their next return journey. Consequently, it is unable to break off one trip from the sequence of GPS records. Thus, a method to filter out possible trips using algorithm 1 is shown in fig 3. Bus GPS records are split into different trajectories where each trajectory represents a run of a bus from start to end stations.

Every valid trip should be bounded by coordinates indicating both end terminals. However, positioning errors are inherent in a GPS data. Therefore, pointing out exact records identical to locations of origin or destination terminals is impossible. Bus terminals generally represent an area, instead of a point. Therefore, a spatial threshold depicting buffer area of circle around a point is used to find matched coordinates in terms of space, in which $R_{terminal}$ is the radius of the buffer area. In this study, it is taken as 150m after a trial and error to filter maximum trips, also by considering facts like, length of the bus, variation of bus halt position within a terminal.

Algorithm 1: Bus trip extraction

```
Data: Data1: bus GPS records (p_i)
      Data2: bus terminals bt_1 , bt_2
Result: Trip(t) (Extracted Travel times of bus trips);
        Trajectory (Sequence of filtered GPS records)

Drop p_i.lat = 0 and p_i.long = 0
Sort Data1 by deviceid and time
for p_i in Data1 do
    for bt_i in Data2 do
        if p_i.geometry ⊂bt_i.geometry.buffer(R_terminal) then
            p_i.terminal = bt_i
        end
    end
end
/* Filtered records for 'end terminals' are grouped as
/* grouped_ends and O/D records are chosen
for g in grouped_ends do
    t.arrival_time = (min(g.time) & g.speed = 0)
    t.departure_time = max(g.time)
end
```

Fig. 3. The pseudocode of the proposed Bus trip extraction algorithm

Among the filtered coordinates, the start time t_{start} of a trip is defined as the last observed time point within the buffer area of the origin bus terminal, while the end of a trip, t_{end} is defined as the first observed time point of zero speed (s=0). The next step was to detect the directions of the trip by taking concern over both start and end terminals of each extracted trip. Trips starting from Kandy are called outbound trips while trips towards Kandy are inbound trips.

b) Bus stop matching

Similar to the approach of extracting trips, bus stop sequences are filtered out from each trajectory by mapping GPS points to the closest bus stop location. Here, the buffer radius around a bus stop is taken as R_{stop} =50m by considering the sampling rate of GPS data and length of a bus. When determining the bus arrival and departure times to a stop, there are three circumstances that must be considered as shown in the Fig 4. In the first case, bus has stopped at the stop to pick up and/or drop off the passengers. Here the arrival time at the bus stop is defined as the first record filtered within the buffer area where the speed equals zero. At the same time, departure point from the stop is selected where the speed of the bus becomes greater than zero (the successive record to the end of series of records with speed = 0) while the point still in the buffer zone. To interpret more the time point t_2 and t_3 (Fig 4. (i)) are chosen as arrival and departure respectively.

In the second case, there is never a speed equal to zero within the buffer area. The reasons put forwarded for this instance are, bus did not stop to pick up and/or drop off passengers, the sampling rate of our receiver (15 seconds) did not capture the shorter brief stop, or missing records due to poor coverage and any technical failure. As the solution, both the arrival and departure times are set to be the first record captured within this buffer, hence it makes dwell time equals zero. As shown in fig 4. (ii), time point t_2 is that chosen record.

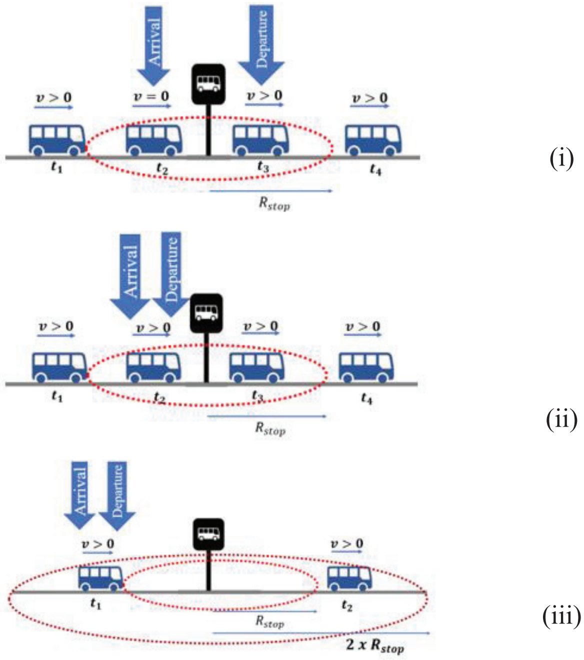

Fig. 4. Three different possible scenarios when capturing GPS records within the buffer area centered by bus stops

There is another rare case where the current buffer radius fails to capture any record. Here upon, the buffer radius is doubled and allowed some space to conquer a record from the trajectory. Here also, both the arrival and departure times are set to be the foremost point within this extended buffer. Despite the fact that the dwell time at this stop equals zero, the arrival time at each stop is required further to estimate running time along each route segments. Therefore, identifying a record closer to the bus stop location is very much important in this step.

C. Outlier analysis

After estimating the trip travel times from the trip extraction algorithm, they were plotted to visualize and analyze their distribution pattern. Outlier analysis is the technique to detect the anomalous value in the data set. First, erroneous values that can be observed in the histogram plot are removed completely from the dataset. These abnormal values can be obtained at the situation, when bus reached the end terminal, however, in the latter part of the journey device might be stopped working, then again when device starts to receive signals, bus might be still parked in the end terminal, or it might complete an around trip again and reached the end terminal. It was captured as the end point and hence, abnormal travel times were obtained.

When it has been assumed that the estimated travel times follows normal distribution, outlier detection is carried in accordance with 'three-sigma rule', in other words, any trips which were greater than three standard deviations from the mean value of their respective route, as shown in the equation (3) below.

$$\{|\, X - \alpha \,| > 3\sigma\} \tag{3}$$

D. GPS data quality metrics

Researchers have proposed several data quality metrics under different sets of dimensions like accuracy, consistency, completeness, integrity, currency, legibility, accessibility etc. In general, the purposes of GPS data are multi-faceted. The raw GPS data are transformed for various applications. Considering our intended purpose of travel time extraction, we propose two metrics that are measuring the quality of the GPS data in the dimension of completeness and consistency. Firstly, effective usable data (EUD) depicts the fraction of GPS records that are accommodated by the extracted trips from the whole raw GPS records (*equation 4*). In essence, EUD shows that remaining fraction of records might not produce a complete trip or corresponds to other private trips.

$$EUD = \frac{GPS\ records\ filtered\ with\ trips}{Total\ collected\ GPS\ records} \tag{4}$$

The next metric, uniformly sampled records (USR) indicate the fraction of records that are received from GPS device at the predefined sample rate (for example, 15 seconds) or lesser (*equation 5*). Besides, the remaining fraction is the records delayed receiving due to network issues, human errors etc. The two metrics can be applied individually to every buses and can check the quality of data being produced.

$$USR = \frac{GPS\ records\ at\ the\ uniform\ rate}{Total\ collected\ GPS\ records} \tag{5}$$

E. Public transit system performance metrics

The performance metrics are needed to be developed solely based on the available extracted historical travel time data, in the absence of schedule data. Bus travel times suffers significant temporal variation. The metrics that consider only the average travel time value obtained from entire data set and defining a threshold is practically meaningless. Thus, analyzing the temporal variation, cluster them based on time horizons and implementing the metrics can be apparent for authorities as well as passengers. Five time-horizons based on peak hours chosen for this study are morning (AM peak), evening (PM peak), afternoon (Noon peak), inter-peak and Off-peak. From the travel time analysis, the time window captured for these peak period horizons significantly varies with the routes and directions of travel. To put it differently, every route and its directions have varying peak hour slots.

Three transit performance metrics EJT, EDT, ERT are developed, considering the total journey time (JT) of a bus comprised of running time (RT) in between stops, dwelling time (DT) at the bus stops. The general formulae for the three metric is shown in the equation (6) below, where they are the excess of actual time value from the threshold. The threshold value is chosen as 90th percentile value, after observing the distribution pattern, which is right skewed, barely showing higher travel times.

$$\text{EJT}\ or\ EDT\ or\ ERT\ =$$
$$Actual\ time\ (JT/DT/RT) - 90^{th}\ percentile\ time \tag{6}$$

Running time analysis is a kind of stratifying the travel time along the segments of the route for further investigations to find out root cause of travel delays. Buses need to be stopped strictly at the bus stops to pick-up and drop-off passengers. When they stop on the way along the route segment except the bus stops, it will extend the travel time of on-board passengers and also causes obstructions to other general traffic using the road. These practices are commonly adhered in Sri Lankan conditions. Developing a metric using the speed values in the GPS data can be a suitable performance measuring strategy for this kind of issue. Another proposed metric, Segment Idle Time Ratio (SITR) measures the percentage of idle time (in other words, speed equals 0) of a trip except the dwell time at bus stops, over the total travel time of that particular trip. Similar to the dwell time estimation, for this metric measurement, the data points of each trip are regularized with 15 seconds interval, other incorporating records were removed. Consequently, every record indicates the state of bus for the 15 seconds time period. Hence, the calculation of segment idle time of a trip (i) is shown in equation and the ratio SITR in equation (7) and (8).

$$sit_i = n_i\ x\ f; (n_i = no.of\ records\ with\ speed\ 0\ \&\ stop = 0,\ f = sampling\ rate\ [15\ seconds] \tag{7}$$

$$SITR = \frac{sit}{Total\ Travel\ time} \tag{8}$$

IV. Results and Discussion

The summary of the processed GPS data along with the measured data quality metrics is shown in the table below

TABLE I. SUMMARY OF THE PROCESSED GPS DATA AND THE DATA QUALITY METRICS

Summary	**Routes**	
	Kandy- Digana (654)	***Kandy-Kadugannawa (690)***
Time period of data collection	October 2021 - February 2022	October 2021 - Februar2022
Total collected GPS records	6.656,394	5,472,668
Trips extracted	**10,224**	**11,150**
GPS records filtered with trips	4,054,163	4,265,762
Effective usable data (EUD)	0.61	0.78
Uniformly sampled records (USR)	0.77	0.89

When moving into in-depth analysis, the travel times and dwell times frequently fluctuate depending on the time of day and day of week and hence, different travel patterns might arise. Therefore, it is crucial to split the bus departure times by time period to analyze the processed data. They were pooled together in 15-minute intervals. Then, the average values of travel times in each interval were calculated and plotted for the route 654 in both in-bound and outbound directions as shown in the Fig. 5. It can be found that determining the time window for the peak and off-peak periods are very complex and it depends on the routes and directions of the travel. For example, in the Kandy-Digana as shown in the figure, the distinguished periods are taken as, AM peak as 6:00 to 7:00, Noon peak as 13:00 to 14:00, PM peak as 15:00 to 17:00, Off-peak as 7:00 to 9:00 and 17:00 to 19:00, while the remaining period is taken as Inter-peak (intermediate peak periods).

Moreover, day of the week also plays a crucial role in the travel time variation, which is shown in the Fig. 6 plotted with attributes of weekday and weekend. It takes about approximately 5 minutes lesser travel time in weekends in the Kandy-Kadugannawa bus route. Therefore, a separate approach can be used for the performance metric in weekends.

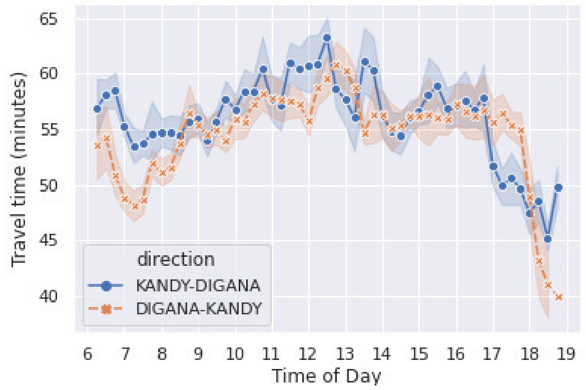

Fig. 5. Variation of travel time with the time of a day, for the two directions in the route 654, where out-bound trips have slightly higher travel time and in-bound direction have an extended peak hour after 17:00 hours

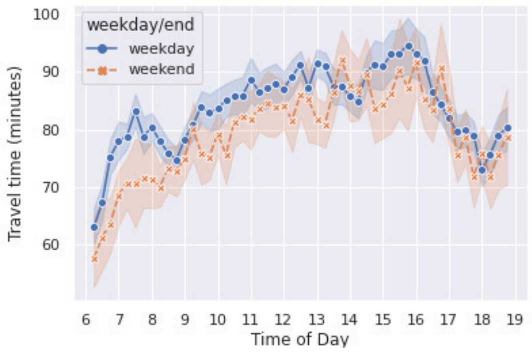

Fig. 6. Variation of travel time during weekdays and weekends for the route 690, where weekends having lower travel time in most time periods

The variation of dwell times at each bus stops in the route 654 is illustrated in the Fig. 7. The bus stops are clustered by demand, for example, (i) prominent stops (202,206,209), (ii) moderate stops (201,203,204,205,208,210) and (iii) brief stops (207,211,212,213). In contrast, brief stops also experience a significant higher dwell time at some instances. Therefore, further break down into peak and off-peak time periods for performance evaluation.

On the other hand, for the effective transit performance, authorities require metrics to evaluate and find out the worst-case scenarios of bus drivers taking too long travel time. The corresponding unruly drivers can be identified from the data set, since each buses have unique device identifier within it.

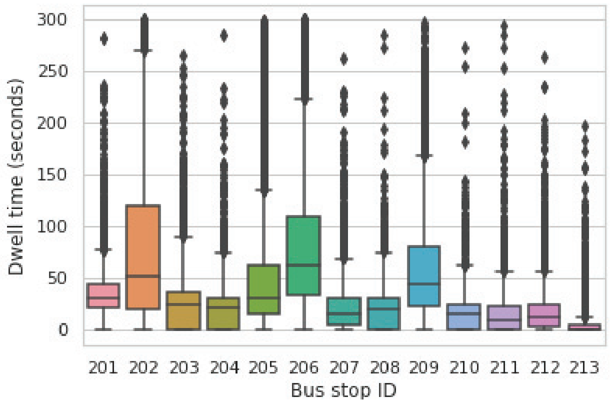

Fig. 7. Box plot showing the distribution of dwell times at the bus stops along Digana to Kandy direction, where some bus stops show higher dwell times

The excess travel time beyond the threshold values is accounted in the metrics EJT along with the unique trip identifier. This table can be used as reference for monitoring by transport authorities for Kandy-Kadugannawa route and similar charts can be produced for other routes.

TABLE II. MEAN TRAVEL TIME AND THE CALCULATED THRESHOLD VALUES FOR VARIOUS TIME PERIODS FOR THE METRIC **EJT**

Metric	*Time period*	*Sample size*	*Mean travel time (minutes)*	*Threshold value (minutes)*
EJT	AM peak	505	58	63
	PM peak	488	57	62
	Noon peak	974	61	69
	Inter-peak	2015	56	62
	Off-peak	1641	51	58

TABLE III. MEAN DWELL TIME AND THE CALCULATED THRESHOLD VALUES FOR VARIOUS TIME PERIODS FOR THE METRIC **EDT**

Metric	*Stop type*	*Sample size*	*Mean dwell time (seconds)*	*Threshold value (seconds)*
EDT	Prominent	14,625	80	190
	Moderate	29,039	31	57
	Brief	19,363	16	40

The time period clustering for EJT and ERT for this Kandy-Digana direction can be further narrowed as Noon peak, Off-peak and inter-peak from the calculated threshold values. As per the analysis, the bus journey time get extreme in the noon period, which lies in the schools closing time.

The metric SITR indicates that the percent of idle time except dwell time. When analysing Kandy-Digana route, the average of 11% value obtained for SITR, in other words, which is about 6 minutes delay in a 53 minutes long journey. However, this SITR value will be very high in city with higher traffic congestion. There are a few occurrences where this SITR value passes 20% of its journey, which indirectly implies that drivers cover the running time in shorter, might be with higher speed. Further research studies can be done with this direction in identifying rude behaviors of drivers.

The use cases and practical application of this research study is contributing to the General Transit Feed Specification (GTFS) by providing both the values of expected journey time and maximum journey time (threshold value) at various time period which will support effective planning and decision making for passengers.

V. Conclusions and Future Works

This paper has demonstrated the technique to extract the bus travel times and dwell times at the bus stops, solely utilizing the GPS data obtained from the devices fixed with buses at the sampling rate of 15 seconds. The proposed two algorithms for bus trip extraction and bus stop matching indicates the mitigation strategies to the possible challenges in the GPS data collection process such as poor network coverage, discontinuities and non-uniformities and the outlier analysis has tackled the extreme running trips. After critically observing the variation of travel times, a single metric enclosing the whole period of a particular route is practically meaningless. Therefore, the bus transit performance metrics EJT, ERT, EDT, SITR proposed in this study provide in-depth analysis of bus journey travel time segregated into dwell time, running time and segment idle time, moreover, into five time-horizons in a day. The output travel times, and their variation reflect the operational pattern of the buses in the 654 and 690 routes of Kandy ensuring the validity of the proposed two algorithms.

Future research direction includes a dashboard for public transportation management authority integrating a system for continuously and comprehensively monitoring punctuality and regularity of the transit performance meeting the reliability level. In addition, developing real-time prediction models for bus travel time and dwell time, hence, arrival time to the downstream bus stops, incorporating weather and other spatial temporal factors, can be done with this processed data.

Acknowledgment

This research work was supported by the Accelerating Higher Education funded by the World Bank. [grant number - Credit/Grant #:6026-LK/8743-LK].

References

[1] A. S. Kumarage, "Kandy_Transport Improvement Program (KTIP) Final Report," Ministry of Defense and Urban Development, Sri Lanka, Kandy, 2014. [Online]. Available: https://www.scdp.lk/pdf/Kandy_Transport%20_Improvement_Program_Final_Report_20140616.pdf

[2] S. Cohen and Z. Christoforou, "Travel Time Estimation Between Loop Detectors and Fcd: A Compatibility Study on the Lille Network, France," Transp. Res. Procedia, vol. 10, pp. 245–255, 2015, doi: 10.1016/j.trpro.2015.09.074.

[3] J.-S. Oh and Y. Chung, "Calculation of Travel Time Variability from Loop Detector Data," Transp. Res. Rec., p. 12, 1945.

[4] M. L. Tam and W. H. K. Lam, "Real-Time Travel Time Estimation Using Automatic Vehicle Identification Data in Hong Kong," in Advances in Hybrid Information Technology, vol. 4413, M. S. Szczuka, D. Howard, D. Ślęzak, H. Kim, T. Kim, I. Ko, G. Lee, and P. M. A. Sloot, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 352–361. doi: 10.1007/978-3-540-77368-9_35.

[5] L. M. Kieu, A. Bhaskar, and E. Chung, "Benefits and issues of bus travel time estimation and prediction," p. 17, 2012.

[6] Z. Huang, S. Qiao, N. Han, C. Yuan, X. Song, and Y. Xiao, "Survey on vehicle map matching techniques," CAAI Trans. Intell. Technol., vol. 6, no. 1, pp. 55–71, Mar. 2021, doi: 10.1049/cit2.12030.

[7] Teng, W and Wang, Y, "Real-Time Map Matching: A New Algorithm Integrating Spatio-Temporal Proximity and Improved Weighted Circle," Open Geosci., vol. 11, no. 1, 2019, doi: 10.1515/geo-2019-0023.

[8] "Simple Map-Matching Algorithm Applied to Intelligent Winter Maintenance Vehicle Data - Carola A. Blazquez, Alan P. Vonderohe, 2005." https://journals.sagepub.com/doi/10.1177/0361198105193500108 (accessed Sep. 21, 2022).

[9] L. Xue, W. YaNan, W. Min, and Y. Mei, "Map matching algorithm design based on GPS information," in 2014 IEEE Workshop on Electronics, Computer and Applications, May 2014, pp. 898–901. doi: 10.1109/IWECA.2014.6845766.

[10] W. Sun and T. Xiong, "A low-sampling-rate trajectory matching algorithm in combination of history trajectory and reinforcement learning," vol. 45, pp. 1328–1334, Nov. 2016, doi: 10.11947/j.AGCS.2016.20160046.

[11] Y. Liu and Z. Li, "A novel algorithm of low sampling rate GPS trajectories on map-matching," EURASIP J. Wirel. Commun. Netw., vol. 2017, Feb. 2017, doi: 10.1186/s13638-017-0814-6.

[12] S. Panichpapiboon and W. Pattara-atikom, An analysis of GPS sampling rates required in travel time estimation. 2009, p. 6. doi: 10.1109/VNC.2009.5416367.

[13] R. Mandal et al., "A system for stoppage pattern extraction from public bus GPS traces in developing regions," in Proceedings of the Third ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems - MobiGIS '14, Dallas, Texas, 2014, pp. 72–75. doi: 10.1145/2675316.2675323.

[14] N. Garg, G. Ramadurai, and S. Ranu, "Mining bus stops from raw GPS data of bus trajectories," in 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, Jan. 2018, pp. 583–588. doi: 10.1109/COMSNETS.2018.8328278.

[15] Munasinghe, Dhananjaya D D, Fernando W W P M, Kumarage A S, Perera A S, and Sivakumar T, "Bus Travel Time Estimation using GPS Data: A Case-Based Study for the Kandy City.," presented at the National Engineering Research Symposium (NERS 2020), Colombo, 2020.

[16] S. Nadeeshan and A. S. Perera, "A Map Matching Algorithm for Noisy, Low Frequent Public Transportation GPS Data," in 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT), Prague, Czech Republic, Jun. 2020, pp. 1081–1086. doi: 10.1109/CoDIT49905.2020.9263797.

[17] Z. Chen and W. Fan, "Extracting bus transit boarding stop information using smart card transaction data," J. Mod. Transp., vol. 26, no. 3, pp. 209–219, Sep. 2018, doi: 10.1007/s40534-018-0165-y.

[18] M. Schil, "Measuring Journey Time Reliability in London Using Automated Data Collection Systems," p. 162.

[19] Z. Dai, X. Ma, and X. Chen, "Bus travel time modelling using GPS probe and smart card data: A probabilistic approach considering link travel time and station dwell time," J. Intell. Transp. Syst., vol. 23, no. 2, pp. 175–190, Mar. 2019, doi: 10.1080/15472450.2018.1470932.

[20] T. S. Sen, S. H. S. University, and T. T. Institute, "Performance Measures for Public Transit Mobility Management," p. 202.

[21] E. Graves, S. Zheng, L. Tarte, B. Levine, and A. Reddy, "Customer Journey Time Metrics for New York City Bus Service using Big Data," Transp. Res. Rec. J. Transp. Res. Board, vol. 2673, no. 9, pp. 1–10, Sep. 2019, doi: 10.1177/0361198118821632.

[22] A. Gittens and A. Shalaby, "Evaluation of Bus Reliability Measures and Development of a New Composite Indicator," Transp. Res. Rec. J. Transp. Res. Board, vol. 2533, no. 1, pp. 91–99, Jan. 2015, doi: 10.3141/2533-10.

[23] Z. Zhuang, Z. Cheng, J. Yao, J. Wang, and S. An, "Bus Travel Time Reliability Incorporating Stop Waiting Time and In-vehicle Travel Time with AVL Data," In Review, preprint, Jan. 2022. doi: 10.21203/rs.3.rs-1144875/