Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Developing a data-processing method to estimate the travel time based on Bluetooth detections

Golnoosh Farajollahi[1], Nicholas Holyoak[1], Rocco Zito[1]

[1] College of Science and Engineering, Flinders University

Email for correspondence (presenting author): fara0050@flinders.edu.au

1. Introduction

Traffic data collection is a fundamental input for effective transport planning and management. Indicators such as travel time, vehicle speed and traffic flow volume are necessary for traffic management and investigating traffic congestion (Jedwanna & Boonsiripant, 2022). With the advent of intelligent transport systems (ITS), traditional data collection approaches such as field surveys are being superseded with technologies such as Bluetooth. Information extracted from traditional manual approaches are quickly outdated representation of actual traffic conditions. Bluetooth technology has been recognised as an alternative, cost-effective method to supply travel time (Bhaskar et al., 2015) and other valuable traffic information (Barcelö et al., 2010, Remias et al., 2017, Kulkarni et al., 2023) such as origin-destination demands (Behara et al., 2018), routing choice (Kottayil et al., 2020) and vehicle trajectories (Advani et al., 2021). There has been increasing interest in the use of Bluetooth sensors in traffic management due to the rapid growth in the number of portable electronic devices that incorporate the technology.

Bluetooth devices and scanners are being investigated by transportation agencies as potential additional data sources for enhancing transport analysis and forecasts (Bhaskar & Chung, 2013) and using this technology has been recognised as a viable technique for collecting real-time travel data (Sichani et al., 2022). A key Bluetooth sensor performance metric is coverage area, defined as the distance for a Bluetooth Media Access Control Scanner (BMS) to identify and connect with other Bluetooth devices. In transportation, Bluetooth coverage zones can be utilised to identify the presence of Bluetooth devices in passing cars. The coverage area is searched in order to read the Media Access Control addresses (MAC-ID) of discoverable Bluetooth devices at a particular point in time at a particular location. The MAC address uniquely identifies Bluetooth devices and enables the communication between them (Friesen & McLeod, 2015) and Bluetooth devices can be followed across multiple locations, allowing for the recognition of the same device and providing crucial travel time, speed and route information (Effinger et al., 2013). While there are many advantages to using Bluetooth technology in transport, it is important to acknowledge and address potential limitations such as outliers and filtering issues that can arise during data analysis and impact the accuracy of analysis results (Michau, Nantes et al. 2013). This research aims to address deficiencies with traditional survey-based data collection methods with the use of data generated by emerging technologies such as Bluetooth for travel time estimation. To achieve this goal and utilise these cost-effective and up-to-date data, we must employ a variety of filtering techniques to create a reliable database and apply it to travel time estimation.

2. Review of existing outlier filtering methods

An observation that considerably deviates from the expected travel time of a vehicle is referred to as an outlier. Measurement errors, alternative paths, stops between sensors and various modes of travel are among the numerous sources of outliers. These issues are commonplace, especially in metropolitan regions and can significantly affect traffic flow data (Van Boxel et al., 2011). Accurate Bluetooth data analysis therefore relies on the filtering of outliers, which may ultimately introduce bias into results. Under congested and/or unstable traffic conditions the outlier identification process can be challenging. A range of outlier-filtering approaches has been proposed in the literature, largely classified as statistical and smoothing techniques in this research. As an example of statistical methods, in research conducted by (Carrese et al., 2021) different filtering techniques including the comparison of extracted travel time with data from Floating Car Data (FCD) or Google API, the filtering around the median or mode, and the boxplot method, have been evaluated. The paper suggested that the mixed procedure using the maximum travel time, the maximum absolute difference with the mode, and the boxplot method resulted in the most accurate results. The performance of three filtering methods including median absolute deviation, modified z-score, and boxplot by considering the defined GPS dataset as ground truth data has been evaluated in (Mathew et al., 2017). According to this study, the modified z-score approach had the best performance with successful removal of 70% of the confirmed outliers and incorrect removal of only 5% of the confirmed nonoutliers. Respectively, the median absolute deviation and boxplot were recognized as the most and least aggressive methods with the highest and lowest removal of both outliers and non-outliers.

Using the standard residual of the robust Greenshield model to determine the confidence interval has been proposed in (Van Boxel et al., 2011). A proactive adaptive outlier detection algorithm employing both current and historical data from *k* nearest neighbours to predict the travel time validity window has been introduced by (Moghaddam & Hellinga, 2014). A non-parametric outlier filtering method called “outskewer” has been proposed by (Khedher et al., 2021) to address the limitation of low-rate sample size and enable real-time outlier detection. In the smoothing technique category, well known methods including the TRANSGUIDE approach based on an adaptive, non-linear regression model (Sw, 1998), TranStar (Vickich, 2001) and TransMIT (Mouskos et al., 1998) have been proposed over the time. The enhanced TRANSGUIDE outlier filtering method has been proposed by (Dion & Rakha, 2006). The paper suggested a low-pass adaptive filtering approach for determining the average travel time and removing outliers. The other improvement of the TRANSGUIDE method has been conducted in (Khedher & Yun, 2017) to overcome the original model's shortcomings in unstable traffic conditions. Additionally, the LOWESS-based outlier filtering technique has been proposed by (Wu et al., 2020). The efficiency of this method with removing 93% of the confirmed outliers in comparison to other methods such as moving average, moving standard deviation, modified z-score, MAD, and boxplot has been approved. a four-step filtering method based on smoothed histogram to identify unusual travel time has been suggested by (Haghani et al., 2010) and based on the comparison against drive test, the performance of the algorithm has been approved.

From this review a modified outlier detection algorithm was developed and used with the Bluetooth data sets described below.

3. Methodology and case study application

This research utilises the Adelaide road network and the Adelaide Bluetooth network. Bluetooth data records were collected from sensors installed along the road network in order to detect Bluetooth signals sent by passing vehicles. The collected data included Bluetooth Media Access Control (MAC) addresses, time stamps, and duration each vehicle spent in the sensor's coverage area. This data was collected over the course of a week, from June 8th to June 14th, 2017, in order to account for both weekdays and weekends. The traffic management system provided the signal phasing information to estimate the duration of traffic signal phases. Additionally, the sensors and traffic data were georeferenced using a geographic information system (GIS) for spatial analysis.

3.1 Single MAC-ID removal

In order to compute the origin-destination matrix and travel time, each MAC-ID must be collected by a minimum of two sensors. Single records are ineffective because they do not provide the complete information about vehicle's travel and movement.

$$\textit{Detect per MAC-ID} \geq N_{Threshold} \qquad (1)$$

Where: $N_{Threshold} = 2$

3.2 Multiple MAC-Id capture method

According to the Bluetooth sensors' coverage area (100-150m), there is a possibility for each MAC-ID to be detected more than one time at each sensor. Based on the chosen detection time among all detections, various scenario can be defined as first-first, last-last, average-average, stop-stop, first-last or last-first. According to the analysis conducted by (Bhaskar & Chung, 2013), the accuracy of average travel time estimation is lower for the First-First section compared to the last-last and stop-to-stop methods. For this study, the last-last method was used as it was found to be the most accurate among all the methods in a previous study conducted by (Saeedi et al., 2013).

3.3 Adjacent sensors issue

If there is an area of overlap between two sensors detection areas (two sensors closer than 100m) and a vehicle stops in this location, it is considered traveling and may result in incorrect travel time as both sensors record it at different times. This issue can be resolved by defining a threshold for two sensors by using the detection radius or to improve the accuracy of determining the threshold, other variables such as the type of Bluetooth sensor properties, the road network characteristic, and the sensors' location on the network. In this study 200m is chosen as a threshold by considering the sensors' detection radius. If two sensors are positioned closer than the defined threshold and are located in the same transport zone, they are treated as a single sensor. Subsequently, all corresponding records are consolidated and transferred to this combined sensor.

3.4 Travel time matrix estimation for each pair of sensors

The travel time matrix estimation stage involves calculating all travel times between each pair of sensors in order to detect atypical travel times and eliminate outliers.

$$\Delta T_{AB} = t_B - t_A \qquad (2)$$

Where ΔT_{AB} is travel time between two sensors A and B, t_A is detection time at sensor A and t_B is detection time at sensor B. Figure 1 illustrates a sample presentation of the calculated travel time for a single day between two Bluetooth sensors.

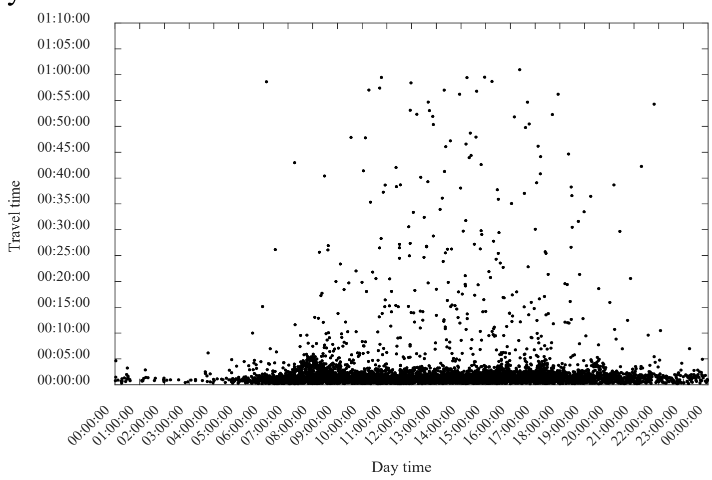


Figure 1: Travel time between pair of sensors

3.5 Rule-based filtering

Before importing the calculated travel time to the outlier elimination method, it is preferable to remove obviously incorrect trips from the database to reduce the effects on statistical indicators such as the mean and median, which influence the results of the procedures. To identify erroneous trips, the speed can be estimated based on the distance between two sensors and the time calculated in the previous step. If the speed in urban area is greater than 120 km/h or less than 1 km/h (Michau et al., 2013) it is unlikely to be the correct trip.

$$S_{AB-i} = \frac{Dist(A,B)}{\Delta T_{AB}} \tag{3}$$

$$DB_T = \{t_{\mathrm{B}i} - t_{\mathrm{A}i} \mid 1 \leqslant S_{AB-i} \leqslant 120\} \tag{4}$$

Where $S_{(AB-i)}$ is speed and ΔT_{AB} is travel time of vehicle *i* between sensors *A* and *B*; *Dist(A,B)* is the distance between two sensors (Euclidean or network distance) and DB_T is filtered travel time dataset.

3.6 Outlier elimination and filtering

The Medcouple boxplot approach is a reliable statistical technique for finding data set outliers based on the robust skewness measurement concept of the medcouple (Brys et al., 2003). The medcouple is determined by dividing the difference between the median and the lower quartile by the interquartile range (IQR). The position of the medcouple relative to the data distribution is then used to identify and identify outliers on the scatter plot. Outliers are data points that are greater than 1.5 or 3 times the IQR. The medcouple boxplot method has a variety of advantages over standard boxplots and other outlier detection methods and is more robust to outliers than the traditional boxplot. This is because it detects outliers using a robust measure of skewness (the medcouple) rather than the IQR. Additionally, the medcouple boxplot approach provides a more accurate representation of a dataset's central tendency and distribution, making it more useful for making conclusions about the dataset (Brys et al., 2004). Considering ΔT_{ijas} a travel time between two sensors *i* and *j*:

$$med = median(\Delta T_{ij}) \tag{5}$$

$$MAD = median\left(abs(\Delta T_{ij} - med)\right) \tag{6}$$

$$Q1 = median(\Delta T_{ij}[\Delta T_{ij} \leq med]) \tag{7}$$

$$Q3 = median(\Delta T_{ij}[\Delta T_{ij} \geq med]) \tag{8}$$

$$g1 = (Q3 - med) \,/\, MAD \tag{9}$$

$$g2 = (med - Q1) \,/\, MAD \tag{10}$$

if: $g1 > -g2$:

$$MC = \frac{(g1 + g2)}{2} \tag{11}$$

else:

$$MC = -\left(\frac{(abs(g1) + abs(g2))}{2}\right) \tag{12}$$

$$H1 = median(\Delta T_{ij}[\Delta T_{ij} \leq med - MC * MAD]) \tag{13}$$

$$H3 = median(\Delta T_{ij}[\Delta T_{ij} \geq med + MC * MAD]) \tag{14}$$

$$Wl = \max(\min(\Delta T_{ij}[\Delta T_{ij} \geq H1 - 1.5 * MC * MAD]), \min(\Delta T_{ij})) \tag{15}$$

$$Wu = \min(\max(\Delta T_{ij}[\Delta T_{ij} \leq H3 + 1.5 * MC * MAD]), \max(\Delta T_{ij})) \tag{16}$$

or

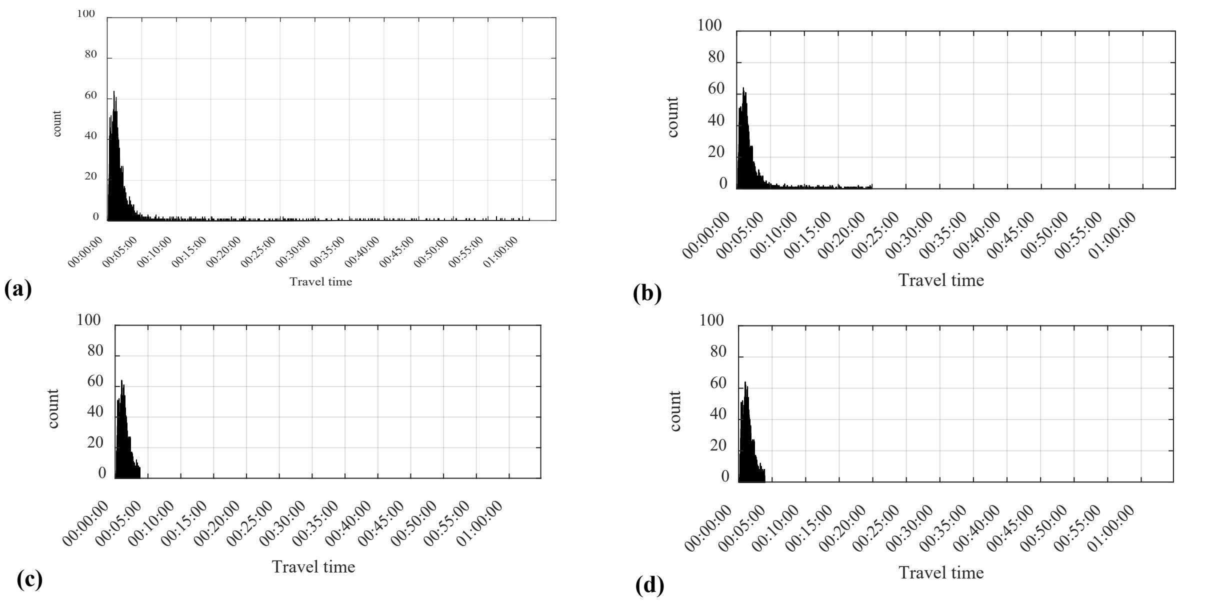
(17)

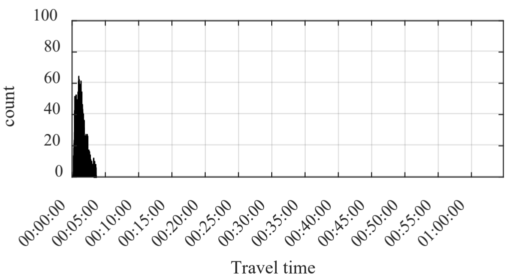
$$Wl = \max(\min(x[x \geq H1 - 3 * MC * MAD]), \min(x)) \quad (18)$$
$$Wu = \min(\max(x[x \leq H3 + 3 * MC * MAD]), \max(x))$$

In this study, we partitioned the travel time matrix into four distinct time periods, namely: AM (7 to 9), DT (9 to 16), PM (16 to 19), and NT (19 to 7). These time intervals have been chosen due to the relatively consistent and stable traffic and congestion patterns observed within each of these periods, a division that is also compatible with transport management practices in Adelaide. This approach serves to enhance the accuracy of the outlier elimination method while aligning with local transportation dynamics.

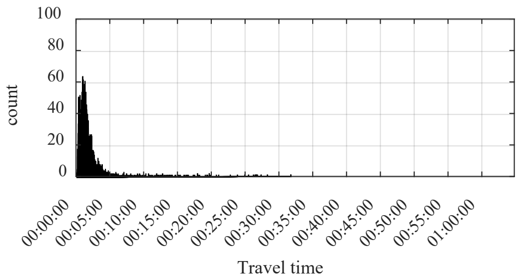
3.7 Stop at signal issue

A statistical method for removing travel time outliers for a pair of sensors could exclude valid trips that stopped at a red light. Removing the valid trip sections from small sample data can be minimized by increasing the upper bound of travel time obtained from the outlier elimination approach by at least one signal cycle time. The most likely scenario, according to our method (Last-Last) for selecting the multi-captured MAC-ID, is stopping at no more than one red light because in this method each vehicle's departure time at downstream and upstream are considered. The adjusted boxplot consists of a box extending from the lower to the upper hinge, with a vertical line at the median, and whiskers that extend to the smallest and largest observations within 1.5 times the medcouple-adjusted median absolute deviation from the lower and upper hinges, respectively. The upper hinge increases by 90 seconds (one signal cycle) and any data outside of the whiskers are considered outliers and removed. In this study, we evaluated a variety of outlier elimination methods for removing errors and anomalies from our data, see review above. These methods included Mean, MAD, Boxplot, Medcouple and Moving Average method which are commonly used in data processing. Our goal was to determine the efficacy of each technique in enhancing the accuracy and reliability of our data analysis. To compare the methods, we evaluated several indicators, such as the mean, median and standard deviation before and after removing outliers. We also represented the overall influence of outlier elimination on our analysis's results in Figure 2.





(e)



(f)

Figure 2: Comparison of the outlier elimination methods on travel time data. (a) Original data. (b) Mean method. (c) MAD method. (d) IQR method. (e) Medcouple method. (f) Moving average method.

By comparing these indicators and visualisations, we were able to identify the most effective outlier elimination method for our data and research question.

Time period / Method	AM			DT			PM			NT		
	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
Original data	2.24	1.67	3.06	1.46	0.95	3.86	1.78	1.30	2.77	1.08	0.85	2.10
Mean	2.01	1.65	1.37	1.10	0.94	0.91	1.61	1.30	1.27	0.97	0.83	0.76
MAD	1.96	1.64	1.29	1.02	0.92	0.55	1.42	1.22	0.89	0.85	0.82	0.43
IOR	1.97	1.65	1.30	1.01	0.93	0.55	1.43	1.23	0.89	0.84	0.80	0.42
Medcouple	1.90	1.60	1.21	1.03	0.92	0.57	1.42	1.21	0.88	0.90	0.82	0.54
Moving average	1.97	1.64	1.30	1.07	0.93	0.82	1.55	1.26	1.11	0.95	0.80	0.71

Table 1. result indicators of various outlier elimination methods for travel time (minutes)

Considering that the effectiveness of outlier elimination techniques relies on the specific attributes of the analyzed data, after thoroughly examining the data in this study and validating it using both real and simulated data from the MASTEM model – a model calibrated with survey and real data – we determined that the Adjusted Turkey method offers the optimal performance in eliminating outliers. Moreover, we noticed that the visual representations of the data created through this method were the most logically consistent and provided valuable insights.

4. Conclusion

This study provides a comprehensive examination of the potential of Bluetooth technology in travel time estimation, as well as the effectiveness of different outlier elimination methods to improve the quality of Bluetooth-based data. After reviewing the existing literature on using Bluetooth data in transport and various outlier elimination methods, and based on our analysis, we concluded that the Adjusted Turkey method produced the lowest Standard Deviation was the most effective at removing outliers and reducing variability in the data specially during peak hours.

While the technique of eliminating outliers shows potential in improving accuracy by effectively removing instances of non-vehicular travel, like those involving pedestrians and bicycles due to their relatively slower speeds compared to vehicles, in order to refine the accuracy of estimated travel times using Bluetooth data, a comprehensive investigation of mode detection techniques becomes crucial, and this research trajectory is distinctly defined for forthcoming investigations.

5. References

Advani, C., Bhaskar, A., Haque, M. M., & Cholette, M. E. (2021). STATER: Slit-Based Trajectory Reconstruction for Dense Urban Network With Overlapping Bluetooth Scanning Zones. IEEE Transactions on Intelligent Transportation Systems, 23(7), 8316-8326.

Barcelö, J., Montero, L., Marqués, L., & Carmona, C. (2010). Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. Transportation research record, 2175(1), 19-27.

Behara, K., Bhaskar, A., & Chung, E. (2018). Novel approach for OD estimation based on observed turning proportions and Bluetooth structural information: Proof of the concept. Australasian Transport Research Forum (ATRF) 2018 Proceedings, Proceedings (pp. 1-5). Australasian Transport Research Forum (ATRF).

Bhaskar, A., & Chung, E. (2013). Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. Transportation Research Part C: Emerging Technologies, 37, 42-72.

Bhaskar, A., Kieu, L. M., Qu, M., Nantes, A., Miska, M., & Chung, E. (2015). Is Bus Overrepresented in Bluetooth MAC Scanner data? Is MAC-ID Really Unique? International Journal of Intelligent Transportation Systems Research, 13(2), 119-130.

Brys, G., Hubert, M., & Struyf, A. (2003). A comparison of some new measures of skewness. Developments in robust statistics: international conference on robust statistics 2001, (pp. 98-113). Physica-Verlag HD.

Brys, G., Hubert, M., & Struyf, A. (2004). A robust measure of skewness. Journal of Computational and Graphical Statistics, 13(4), 996-1017.

Carrese, S., Cipriani, E., Crisalli, U., Gemma, A., & Mannini, L. (2021). Bluetooth Traffic Data for Urban Travel Time Forecast. Transportation Research Procedia, 52, 236-243.

Dion, F., & Rakha, H. (2006). Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. Transportation Research Part B: Methodological, 40(9), 745-766.

Haghani, A., Hamedi, M., Sadabadi, K. F., Young, S., & Tarnoff, P. (2010). Freeway travel time ground truth data collection using bluetooth sensors. Transportation Research Record, Journal of Transportation Research Board, 2160, 60-68.

Jang, J. (2016). Outlier filtering algorithm for travel time estimation using dedicated short-range communications probes on rural highways. IET Intelligent Transport Systems, 10(6), 453-460.

Jedwanna, K., & Boonsiripant, S. (2022). Evaluation of Bluetooth detectors in travel time estimation. Sustainability, 14(8), 4591.

Khedher, M. B. B., & Yun, D. G. (2017). Study on Enhancement of TRANSGUIDE Outlier Filter Method under Unstable Traffic Flow for Reliable Travel Time Estimation-Focus on Dedicated Short Range Communications Probes. Journal of the Korea Academia-Industrial Cooperation Society, 18(3), 249-257.

Khedher, M. B. B., Yun, D. G., & Sung, J. G. (2021). Statistical Method for Real-Time Detection of Travel Time Outliers. Journal of the Korea Academia-Industrial cooperation Society, 22(10), 1-10.

Kottayil, S. S., Tsoleridis, P., Rossa, K., Connors, R., & Fox, C. (2020). Investigation of driver route choice behaviour using bluetooth data. Transportation Research Procedia, 48, 632-645.

Kulkarni, A. R., Kumar, N., & Rao, K. R. (2023). Efficacy of Bluetooth-Based Data Collection for Road Traffic Analysis and Visualization Using Big Data Analytics. Big Data Mining and Analytics, 6(2), 139-153.

Mathew, J. K., Day, C. M., Li, H., & Bullock, D. M. (2017). Curating Automatic Vehicle Location Data to Compare the Performance of Outlier Filtering Methods. JTRP Affiliated Reports, Paper 41.

Michau, G., Nantes, A., & Chung, E. (2013). Towards the retrieval of accurate OD matrices from Bluetooth data: lessons learned from 2 years of data. 36th Australasian Transport Research Forum, ATRF.

Moghaddam, S. S., & Hellinga, B. (2014). Algorithm for detecting outliers in Bluetooth data in real time. Transportation research record, 2442(1), 129-139.
Mouskos, K. C., Niver, E., Pignataro, L. J., Lee, S., Antonious, N., & Papadopoulos, L. (1998). Transmit system evaluation.
Remias, S. M., Hainen, A. M., Mathew, J. K., Vanajakshi, L., Sharma, A., & Bullock, D. M. (2017). Travel time observations using bluetooth mac address matching: a case study on the Rajiv Gandhi roadway: Chennai, India.
Saeedi, A., Park, S., Kim, D. S., & Porter, J. D. (2013). Improving accuracy and precision of travel time samples collected at signalized arterial roads with bluetooth sensors. Transportation research record, 2380(1), 90-98.
Sichani, M., Arnold, R., & Bubendorfer, K. (2022). Finding the Signal: Near Real-time Data Analysis for Urban Traffic Monitoring on a Distributed Bluetooth Sensor Network. 2022 IEEE 18th International Conference on e-Science (e-Science), (pp. 171-181). IEEE.
Sw, R. (1998). Automatic vehicle identification model deployment initiative—system design document. Report prepared for Trans-Guide. Texas Department of Transportation, Southwest Research Institute, San Antonio.
Van Boxel, D., Schneider IV, W. H., & Bakula, C. (2011). Innovative real-time methodology for detecting travel time outliers on interstate highways and urban arterials. Transportation research record, 2256(1), 60-67.
Vickich, M. (2001). Personal e-mail communications. In: June-August.
Wu, Z., Wu, Z., & Rilett, L. R. (2020). Innovative nonparametric method for data outlier filtering. Transportation research record, 2674(10), 167-176.