Advani, Chintan, Thakkar, Smit Bharat, Arkatkar, Shriniwas, & Bhaskar, Ashish
(2020)
Performance Evaluation of Urban Arterial Network Using Wi-Fi Sensors Under Heterogeneous Traffic Conditions.
In Ulengin, Fusun, Patil, Gopal R., Ozaydin, Ozay, Tavasszy, Lorant, & Verman, Ashish (Eds.) *Recent Advances and Emerging Issues in Transport Research: Proceedings of the World Conference on Transport Research - WCTR 2019.*
Elsevier BV, Netherlands, pp. 1022-1037.

https://doi.org/10.1016/j.trpro.2020.08.130

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 48 (2020) 1022–1037

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Performance Evaluation of Urban Arterial Network Using Wi-Fi Sensors Under Heterogeneous Traffic Conditions

Chintan Advani[a], Smit Thakkar[b], Shriniwas Arkatkar[c*], Ashish Bhaskar[d]

[a]Masters student , Civil Engineering Department, SVNIT, Surat-395007, Gujarat, India E - mail: chintanadvani@gmail.com
[b]Masters Student, Civil Engineering Department, SVNIT, Surat-395007, Gujarat, India E - mail: thakkarsmit11@gmail.com*
[c]Assistant Professor, Civil Engineering Department, SVNIT, Surat-395007, Gujarat, India E - mail: sarkatkar@gmail.com
[d]Senior Lecturer, Department of Civil Engineering, QUT, Brisbane,Australia E - mail:ashish.bhaskar@qut.edu.au

Abstract

Traffic data collection is one of the major issue faced by the researches as well as government authorities for the study of traffic movement behaviour. Conventional methods of data collection under mixed traffic condition do not yield promising results and cannot be completely relied upon. Thus, this study aims to assess the potential use of Wi-Fi based sensors for the performance evaluation of arterial closed network and formation of O-D matrix in Centre Business District area of Surat city with help of Wi-Fi based data collection technique. This paper expounds the essentiality for the use of Wi-Fi data as an outlining parameter in the performance evaluation of the study network, which is unlike the conventional examinations that signifies the use of Bluetooth observations for evaluating the network operation based on the travel time reliability measures. The data provides exceptional results than the other hackneyed techniques, which results in less sample size and accuracy. Further, the network development was executed with the deployment of four sensors at critical confluences. This was useful for scrutinizing the traffic movements along the arterial network.

The sensors used here work on the principle of recurring detections of MAC ids for the devices along with their analogous time stamps for the duration in which they remain within the detection range. Thus, re-identification of the same moving vehicle along various network points will help to determine the travel time where the difference in the time stamps of certain MAC ids moving among several junction points were exercised for the determination of travel time and development of O-D matrix along all possible routes. The Wi-Fi data obtained from the sensors in Surat city network corresponds to the critical routes for various hours of the day. This leads to development of primary and secondary matrix comprising of total and cut through traffic, which are useful for multiple aspects. The trial results show the successful penetration (unique detections to vehicles) ratio ranging from 20-to30percentage after significant results from various statistical analysis test as well as considerations from literature studies. The ultimate outcomes will justify the substantiation of results obtained from sensors along with assessments with other performance-measuring device.

Keywords: Wi-Fi/Bluetooth Sensors;Travel Time; Penetration Rate; Arterial Network.

*Corresponding author. Tel: +91840252777
E-mail address: sarkatkar@ced.svnit.ac.in

2352-1465
10.1016/j.trpro.2020.08.130

1. Introduction

Upgrades in expectations for everyday comforts of the general population and expanded financial exercises in urban territories have prompted a relating surge in rush hour gridlock crosswise over existing urban transport systems. Subsequently, there is a requisite for examining the performance of this network under altered traffic conditions. However, due to the heterogeneity of the class of vehicles under mixed traffic conditions monitoring the traffic is an emergent issue in any urban network and to mitigate the needs of this burgeoning traffic issue for any urban arterial road is extremely important for the smooth functioning of any urban region. Travel time is one of the earliest and most imperative measures to assess the performance of an arterial or freeway system. Vehicles on urban systems are not conserved because of mid-connect sources and sinks, for example, stopping or side roads, and the vast majority of the urban connections are shared by various vehicle categories, for example, autos, cars, bicycles, and people on foot and so forth. Every one of these qualities add to the complexities in travel time estimation on urban systems.
The customary strategies' have inadequacies for travel time estimation and Origin-Destination (OD) study and the city's pattern to impact the most recent techniques of traffic data collection and travel pattern, the Wi-Fi innovation could be a favored other option to give a bigger sample size of pertinent movement information at a lower cost than that is conceivable with the ordinary techniques. Along these lines, Wi-Fi Technology has turned out to be progressively far reaching lately to conduct travel time and Origin-Destination (OD) study. The utilization of this technology is anyway still genuinely novel in OD data collection for urban streets where the roadway systems and movement attributes are exceptionally multifaceted for Indian conditions. This paper abridges a recently-completed study in which four Wi-Fi detectors are deployed and used to appraise travel time and OD data for an unpredictably organized CBD street system of Surat City, Gujarat-India.

The Wi-Fi sensor is an electronic communication protocol that uses a unique electronic identifier in every device, called the Media Access Control (MAC) address. The uniqueness of the MAC deliver makes it conceivable to track Wi-Fi enabled gadgets on vehicles at an upstream point out and re-distinguish these gadgets at a downstream point. The time difference between the two perceptions can be utilized to appraise travel time. In spite of the fact that Wi-Fi/Bluetooth innovation has been seen to be prepared to do such an application, the exactness and dependability of movement time estimation utilizing this innovation require facilitating examination.

The MAC address is a one of a kind 48-bit identifier for the gadget. For example, '01:23:45:67:89: ab' this deliver given to a gadget by the maker, is separated into six octets (units of eight bits), where the initial three ('01:23:45') are the Organizationally Unique Identifier (OUI) that speak to the producer. The last three octets ('67:89: ab') are appointed by the maker in any capacity they want to make the gadget one of a kind.

On the off chance that one device is distinguished by the MAC address at two destinations and the correct examples of time those recognizable pieces of proof occurred are known, the time it took the device to venture to every part of the separation between the two locales and the speed at which it did as such can be figured. Aggregating data of different gadgets would give an average travel time or stream speed between those destinations.

2. Literature review

The prominence of travel time reliability analysis is gaining importance with the rampant rise in vehicular population. Effort to enhance this evaluation by means of electronic detection methods is one of the areas that is being exploited lately. A review of the literature has shown that there are various methods in practice for this purpose, like, video graphic analysis, GPS databased evaluation, by means of loop detectors etc.(Carrion & Levinson, 2013). Travel speed estimated from loops especially from the stop-line of signalized intersections is not representative of the space mean speed along the corridor. GPS based travel speed measures from third party (such as Google) lacks the spatial temporal coverage needed for a seamless travel time estimation. Video graphic analysis on the other hand demand high work force, as it involves the data extraction compilation. It is in the light of these scenarios, that we decided to explore the possibilities of using Bluetooth/Wi-Fi sensors(Park, Saeedi, Kim, & Porter, 2016), which, as per literature, could provide robust detection of vehicles (Abbott-Jard, Shah, & Bhaskar, 2013; Bhaskar & Chung, 2013; Tahmasseby & Engineering, 2015), for this purpose.

A framework to model the Traffic and Communication Simulation was proposed in a study (Bhaskar & Chung, 2013), which investigated the modelled temporal errors from the Bluetooth Media Access Control Scanner (BMS) data and thereafter, the accuracy and reliability of travel time estimations from BMS data, which was a sign of promise of the abilities of Bluetooth Scanners as a means of traffic analysis. Another study (Michau et al., 2017) heralded Bluetooth as a not-fully-explored technology in the management of urban transport networks, and encouraged researchers and practitioners to take a more cautious look at what is currently understood as a mature technology for monitoring travellers in urban environments. Few methods to overcome the noise present in the data were also elaborated by the authors. The effect of antenna characteristics on Wi-Fi data in terms of travel-time determination for non-motorised vehicles and on foot commuters were studied in another work (Abedi, Bhaskar, & Chung, 2013), which also evaluated the effects of variations in detections with different antenna gains in order to recommend most suitable set up obtaining better accuracy. A study pertaining to Indian conditions (Mathew, Devi, Bullock, & Sharma, 2016) to capture and evaluate the travel time data and determine travel time reliability parameters using Bluetooth based sensors along two different routes in Chennai city, India. Based on the data collected over a one-month span, the authors perceived that use of Bluetooth technology has the prospective to provide precise travel time estimations for urban arterial roads in India.

The general observations based on literature was that Bluetooth could perform consistently in measuring travel time reliability. On the other hand, the usage of Wi-Fi, which can help obtain a similar database, was observed to be one of the less-explored technologies in this domain. It is also worth noting that, even though Bluetooth is gaining popularity in India, as a data collection method because of its cost-effectiveness and ease of installation, studies pertaining to this area were not widely observed in Indian conditions. A study of paper provide significant insights into the strengths and weaknesses of the Wi-Fi technology (Blogg, Semler, Hingorani, & Troutbeck, 2010) where the Bluetooth detections data are compared to both Video and Automated Number Plate Recognition (ANPR) O-D data. The research focuses on determining the upper sample size as well as station optimization for the various location. Further he author also works on the OD determination for two different approaches based on MAC to volume detection rates. The case study (Carpenter, Fowler, & Adler, 2012) provides noteworthy understanding into the efficacy of Bluetooth data for Origin-Destination studies. The paper shows that it is likely to use Bluetooth data on intricate study corridor with multiple competing routes. A successful placement of sensor depends on thoughtful structural design and device orientation. The paper summaries three method of device placement that provides a trade-off between Origin-Destination data and travel speed information. In all studies, precaution was taken to ensure that the Bluetooth sensors are located at optimal location to intercept traffic and signals on or after side streets. The matrix generated in the paper centres on practical and competent mechanism for route explicit trip tables. The study on (Thogulava, Antonovs, Bingham, Hill, & Hanchett, 2015) describes the processes adopted in deriving matrices from Bluetooth traffic surveys to compare with Worcester transport model. The paper adds knowledge to harness the data available from various technologies such as Bluetooth to develop robust traffic matrices and models. Statistical techniques adopted in cleaning of Bluetooth records and obtaining robust travel times is critical to obtaining reliable trip matrices for traffic models. The path mapping process described capitalizes on techniques available with current technology to lead to a step-change in approach to transport modelling by using observed travel paths as the basis of obtaining matrices. With increasing penetration of Bluetooth technology in vehicles and personal gadgets, the proportion of paths that can be detected compared to link volumes only increases, eventually providing the possibility of having models built based on travel paths supplemented by matrices and trip generation only for forecasting purposes. The study in for a single arterial network has stated the variation of penetration rate with respect to volume for urban arterial network as well travel time reliability for the same. Although there are no specific guidelines mentioned in the paper regarding cut through traffic, it can be used as a platform for determining the reliability of penetration rate for the section. Moreover, expressway study in (Advani, Thakkar, Shah, Arkatkar, & Bhaskar, 2019) supports the study as approached in earlier paper with the only difference that the performance of intercity expressway is determined instead of urban network but the approach is beneficial related to our study.

3. Encouragement of Wi-Fi over Bluetooth

The significance for the use of Wi-Fi data over Bluetooth data can only concluded if the Wi-Fi data set is capable to prove its reliability over Bluetooth. The reliability of both the types of detection can be measured based on some descriptive statistics, explaining the deviation in mean and variance of the dataset. In regard to this, t-test for two-sampled mean was carried out between Wi-Fi and Bluetooth detections over video graphic volume count to explain the variation in the mean, and F-test for two sampled variances was determined for the same as indicated in the Table 1 below.

Table 1: T-test between the dataset of observed volume and Wi-Fi detections

T-Test: Paired Two Sample for Means	*Vehicle volume*	*Wi-Fi detections*
Mean	95.6	45.08333333
Variance	96.10847458	94.38276836
Observations	60	60
Pearson Correlation	0.212839373	
Hypothesized Mean Difference	0	
df	59	
t Stat	31.95499195	
P(T<=t) one-tail	3.00693E-39	
t Critical one-tail	1.671093032	
P(T<=t) two-tail	6.01386E-39	
t Critical two-tail	2.000995378	

The Table (1) shows that there is a significant difference in the mean value of the two dataset. Sufficient number of observation are available to justify the results obtained. Both, for the one-tail and two-tail tests, at a confidence interval of 95%, the probability value is beyond 0.05, which indicates a significant variation in the mean value for the two data sets.

Table 2: T-test between the dataset of observed volume and Bluetooth

t-Test: Paired Two Sample for Means	Vehicle volume	Bluetooth detections
Mean	95.6	2.066666667
Variance	96.10847458	1.995480226
Observations	60	60
Pearson Correlation	-0.25628318	
Hypothesized Mean Difference	0	
df	59	
t Stat	70.63655376	
P(T<=t) one-tail	5.16297E-59	
t Critical one-tail	1.671093032	
P(T<=t) two-tail	1.03259E-58	
t Critical two-tail	2.000995378	

Similar results as observed for Wi-Fi detection is observed for Bluetooth device. The results from Bluetooth is much vague and unjustified, and therefore cannot be used for obtaining results of the field volume due to insufficient sample size and non-representation of field conditions. Similarly, the results of the F-test between vehicle volumes along with Wi-Fi and Bluetooth detection for unequal variance are stated below.

Table 3: F-test between the dataset of observed volume and Wi-Fi for unequal variance

F-Test Two-Sample for Variances	*Vehicle volume*	*Wi-Fi detections*
Mean	95.6	45.08333333
Variance	96.10847458	94.38276836
Observations	60	60

df	59	59
F	1.018284124	
P(F<=f) one-tail	0.472378372	
F Critical one-tail	1.539956607	

Table 4: T-test between the dataset of observed volume and Bluetooth for unequal variance

F-Test Two-Sample for Variances	*Vehicle volume*	*Bluetooth detections*
Mean	95.6	2.066667
Variance	96.10847458	1.99548
Observations	60	60
df	59	59
F	48.16308041	
P(F<=f) one-tail	2.11446E-34	
F Critical one-tail	1.539956607	

Here, the F-test ultimately explains the importance of Wi-Fi results over Bluetooth results. For the Wi-Fi detections the variance observed is nearly same as that of the vehicular volume which indicate that there is no significant variance difference among the two and hence the probability for one tail test is 0.47 which is greater than 0.05 and holds true for confidence interval of 95%, but the same is not true for the Bluetooth detections as significant deviation is observed in both the variance as well as the mean value. Thus, from all the above analysis, it is obvious that for Wi-Fi ids as the mean value is different and thus this indicates that a direct use of Wi-Fi detection data do not resemble the actual field conditions but the variance is same, and hence a relation can be established in expanding the Wi-Fi unique detections to the observed volume.

4. Deployment of sensors

Thoughtful consideration was made before placement of the sensors to ensure that the fruitful data is collected which can serve for multiple purpose. Usually, literature review suggests that optimal deployment location should be determined based on purpose for which it is deployed. For e.g., if the primary aim of the study is to determine travel time for any particular direction, then sensors should be placed adjacent to the road such that minimum noisy data is obtained and elevation which can assure suitable range to capture all the W-Fi enabled devices passing through that point in any particular direction are captured should be selected. However, our study was targeted on multiple aspects i.e. determining travel time as well generation of O-D matrix, so that overall network performance can be evaluated. Sometimes, it may also be possible that the corridor specific feature will prevent the optimal layout from being implemented.

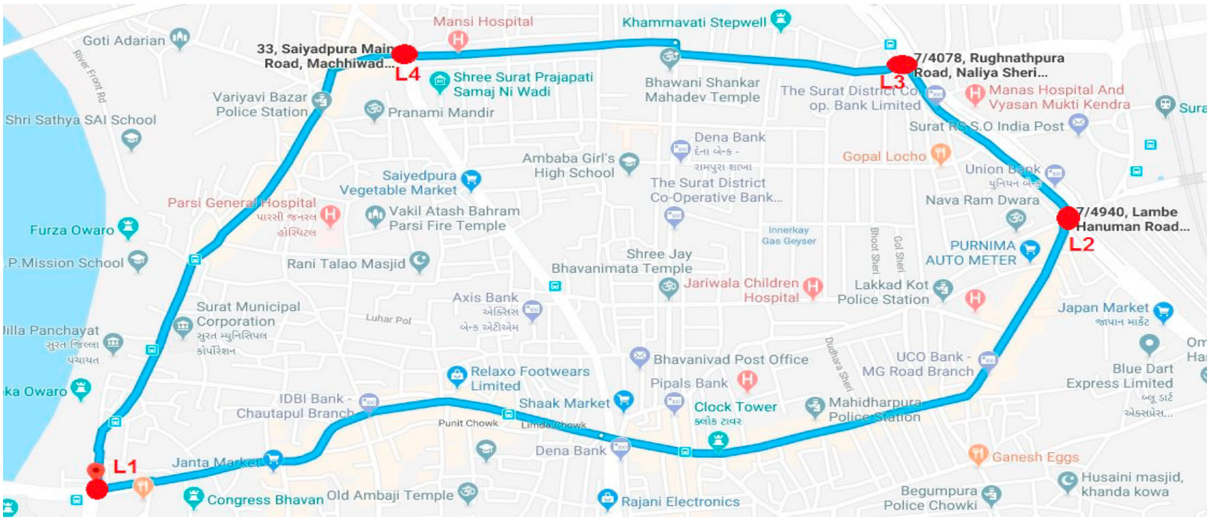


Figure 1: Central Business District network for Surat City

The Figure 1 shows an urban arterial closed network with sensors at four critical confluences and with the aim to capture maximum possible detections with utmost accuracy. The locations L1 to L4, marked in the Figure 1 are the intersection and the sensors are placed at the centre of the confluence for capturing traffic movements from all the directions. Thus the entire network consists of four major links with some minor secondary links and all the links are usually saturated for the entire day as the network is a part of highly dense commercial zone as well one of the primary route towards the railway station of the city. The main aim of using a hybrid approach here is to provide sufficient level of details for the O-D movement, as well as providing travel speed information for some specific movement in the network.

Sensors were located at maximum distance of 50 meters of traffic travelling in all the directions and had a clear line of sight for optimal data collection. These considerations are made to reduce the effect of physical barriers like bridge, road and other barriers on capturing rate of the sensors. In addition, attempts were made to minimize the collection of wayward data collected from the pedestrians or other cross streets roads. Along with the sensors, video cameras were installed sensors to determine the capturing rate (penetration rate) at various locations.

5. Data Collection

The entire team identified the major travel movements in the study corridor through visual inspection. The location chosen helped to develop O-D matrices as well as travel speed profile from the data at various level of aggregation for different time intervals of the day. Sensors were deployed over a period of 9 hours from morning 10 A.M to evening 7 P.M on 16th February 2018. The sensors were enabled with the battery power having configuration of 12 V and 5 A output, which can be, suffice for the sensor to operate for the mentioned hours. The sensors were enabled with 5-dBi strength antenna having capturing zone up to 100 meters if placed at suitable elevation on open ground without obstructions. Details regarding the data collected at all the locations is mentioned in the Table (5) below.

Table 5: Observed dataset collected from the Wi-Fi sensors

Station	L1	L2	L3	L4
TIME	08:20-19:05	09:00-19:20	09:15-19:10	10:05-19:38

TOTAL DETECTIONS	54,076	2,11,045	1,22,349	85,578
UNIQUE DETECTION	20,624	48,015	31,170	19,545

Table 5 shows the variation of detection of the unique ids at all the locations for various hours of the day. It is ideally considered that the trend of the detections usually represent the traffic conditions prevailing on the field.

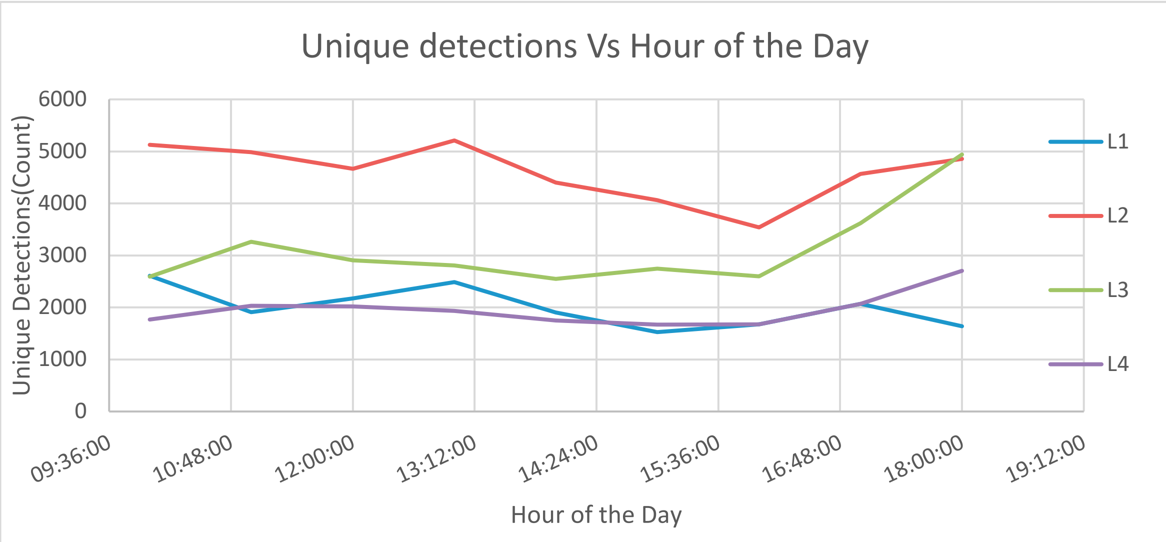


Figure 2: Graphical Representation of unique detection over time for all the locations

From the Figure 2 it is clear that for each of the location the trend obtained is completely different compared to any other location. This indicates the heterogeneity of demand at various locations based on their land use pattern and type of intersections where the sensors are placed.

6. Data Analysis

Data from the sensors was analysed using analysis carried out in excel along with statistical software like Mini-Tab, SPSS along with some additional tools. Three major objective of the analysis were:

1. Summarizing the specific travel movement patterns by development of O-D matrix along various links for the mentioned study period.
2. Developing travel speed profiles for the links on which the O-D matrix is determined.
3. Determining approach to predict cut-through (non-stopping) traffic from matching movements among links.

6.1 Developing O-D matrix

An O-D matrix will represent the spatial interaction of all the vehicles for various segments in the space. Here as the four sensors are placed a 4 X 4 matrix will be generated such that each row will represent the origin from which the trip started and the column represents the desired destination towards which the vehicle travelled. The matrix is formed entirely based on the matching of the unique data at several points of the network. The data can be also validated if the volumes on all the links are calculated manually or by the conventional techniques like licence plate or video graphic extraction method.

The distinction between the standard matrix and the matrix that we have considered for the study is that in this study the O-D matrix is route specific i.e. identification of the various routes adopted to reach a common destination point from same origin is studied here. Further, the obtained detections on various routes can be expanded to the link volume on the same routes by multiplying it with some suitable expansion factor. However, for our study, currently the matrix is developed only based on the matching detections directly obtained from the coding techniques [13, 14].

6.1.1 Generation of O-D matrix

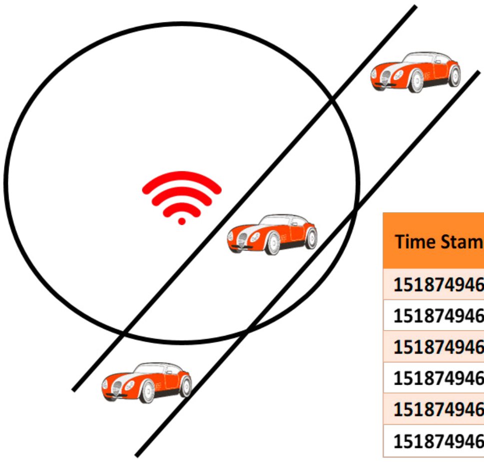
The raw data available from the sensors was used to determine the route specific OD matrix. Here the deployment of sensors at different locations corresponds to different period. Hence, the time duration for which the data set was common at all the points i.e. from 10:15 to 18:45 is considered for all the analysis. For the generation of the matrix a two-step process is required, which includes removal of erroneous data further carried by determination of sequential link trips. Thus, these unique trips provide information for the movements on links. This generated matrix is further reduced based on primary and secondary destination, which includes the determination of these unique ids on detection for more than one stations that is intermediate stations.

STEP1: DATA CLEANING

The raw data consist of 473048 no of data set retrieved from all the four sensors for the period of 8-hour duration of which the total unique detection are 119354 which comprises of around 25% of the total detections. This huge data set need to be filtered and thus processed before they can be used to determine the link volumes. For determination of O-D matrix, it is necessary that the id should be detected at more than one points on the network. This is because the id that is detected only once at any station may be a static id or the id corresponding to the pedestrians, which is of no use for our current study.

Thus, the first step includes the detection of the unique id at all the stations, which includes removal of duplicates at that station. The duplicate detections are a result of the slow moving vehicle or the stoppage of vehicle near the sensor, which results into multiple detection for the period in which it remains within the detection range. Thus, for determination of the O-D matrix only one of the detection value along with it corresponding timestamp must be considered for travel time determination. The literature review suggests that depending on the type of purpose of the study the timestamp should be considered. For example, if one desires to find the period for which the device was within the detection range then it becomes necessary to determine both the first as well as last time stamp for the same sensor. For our study, it is necessary to know the matching between the two points so here the travel time between two stations is considered based on first time stamp at which the id was detected at each sensor. A hybrid approach may be considered for future study in which an average time stamp may be considered based on all the reads detected in a given time window to predict the accuracy of prediction of travel time.

Figure 3 below shows that an id, which was present at location L1 for some duration within the detection range, had multiple detection. Thus, out of multiple detections, only the first detection is considered here and the rest are deleted with the help of filtering technique. Finally, in this way all the multiple detections at each location are removed keeping only the first timestamp of detection and hence the total no of detections are considerably reduced which is shown in Table 5.



Time Stamp	Wi-Fi MAC Id	Signal Strength	Location Id	Data Cleaning
1518749465	da:a1:19:05:9a:95	-58	265831	Accepted
1518749465	da:a1:19:05:9a:95	-63	265831	Rejected
1518749467	da:a1:19:df:0f:a6	-63	265831	Accepted
1518749467	da:a1:19:df:0f:a6	-65	265831	Rejected
1518749468	ac:c3:3a:30:2a:1a	-53	265831	Accepted
1518749468	5c:99:60:0e:1e:96	-64	265831	Accepted

Figure 3: Example of removal of duplicate MAC ids

STEP 2: TRIP MATRIX

After cleaning the above data set, the remaining ids are required to form the matching matrix. Here the first matrix includes all the ids matching between other stations. Thus, the matrix consists of the breakage of the trips i.e. a single trip here is included as two different trips. For example, let an id A travels from point L1 to L3 via L2 then for the primary matrix two trips are considered here i.e. Trip 1: L1 → L2 and Trip 2 L2→L3. The primary matrix is necessary to forecast the link volumes between all the inter links prevailing among the various origins and the destinations for any observed network. Moreover, for further studies related to route choice analysis it is necessary to determine the distribution of the trips in any particular hour with respect to travel time in upstream as well as downstream direction, which is ultimately helpful in analysing the overall directional performance of the links. In addition, it is necessary to note that for the primary matrix total matchings are considered irrespective of the time taken to reach at the destination. Thus, the matrix includes all categories of vehicles and all kinds of trips, which may or may not be the desired origin and destination based on the purpose of the study.

Table 6: Primary O-D matrix

O/D	1	2	3	4
1	-	1145	922	797
2	1084	-	2160	1063
3	677	1510	-	1017
4	536	872	977	-

Once the primary matrix is created now it becomes necessary to reduce the first matrix based on some optimal travel time to form secondary matrix, which will remove the errant dataset, included non-useful categories (stopping vehicles) of trips and thus including only through vehicular trip, which would yield necessary results.

Thus, the following criteria are considered for formation of the secondary matrix:

1. Only unique ids are considered for trip formation.
2. Maximum time difference between links for determination of through traffic is determined based on the optimum values determined by the whisker box plots for all the links and the corresponding histograms of travel time, which are indicated latterly.

3. Round trips are not specified differently in the matrix.

6.1.2 Arterial link travel time using Wi-Fi records

For the removal of extreme records, a filter will be required for removal of scattering data to maintain the homogeneity in the travel time. This for short duration data, filtering can be carried out by segmenting the entire data into small intervals of 30 minutes or 1-hour duration and then applying corresponding percentile value depending on the confidence interval required for the data. Usually the previous research is done by consideration of some arbitrary value for the link and by this way, the outliers are removed. However, for our study as the data set is adequate and the duration of study is for 9 hours, histograms are plotted for the entire day and then the maximum probability with which a vehicle may travel on a given link is determined. This value is then considered as the optimum bar for the separation of outliers from the handy data. Thus, link L1-L2 is selected as an example to represent the overall distribution acting over all other link of the network. Similar trends are developed for all other links and distributions are checked.

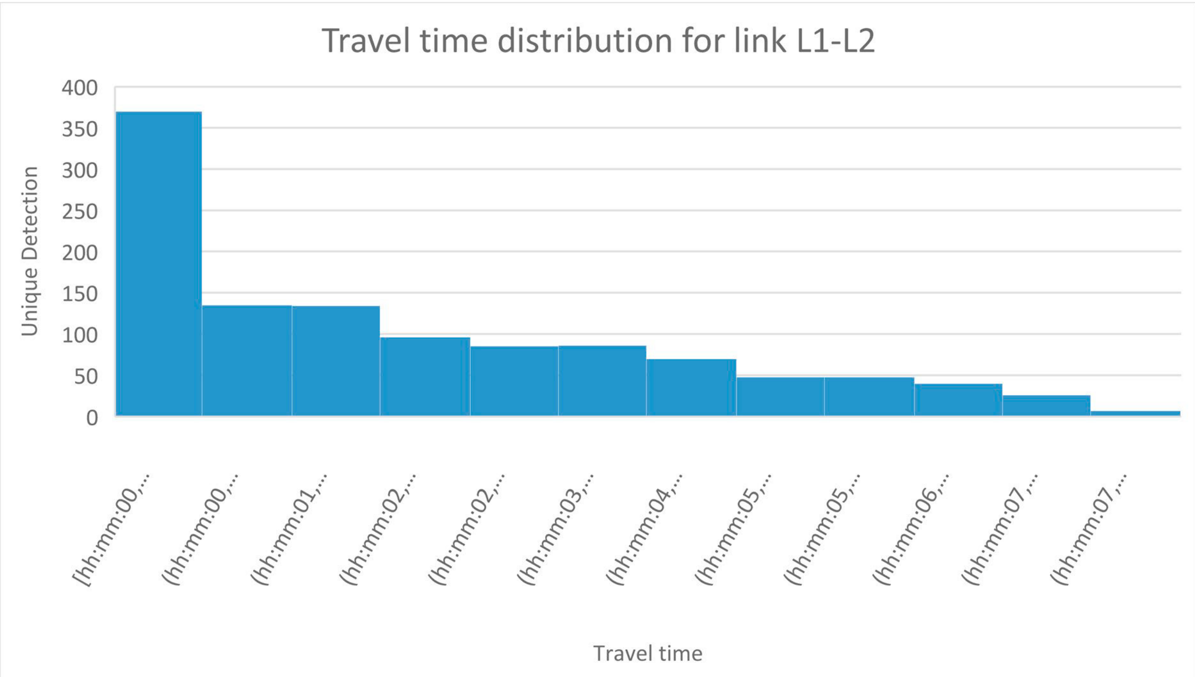


Figure 4: Travel time distribution for vehicles travelling on link L1-L2

Figure 4 shows the distribution of the unique detections with corresponding class time interval for a particular link. Similar kind of distributions were plotted for all the links for the optimal travel time, as one of the important aim of the research was to determine the through traffic. For precise analysis, it is also necessary to determine the travel time variation on both ways for the same link, as it is not necessary that homogeneous traffic conditions prevail in both the directions. Thus, precautions have been taken to bifurcate the optimal travel time on both the ways unlike previous study with some arbitrary assumed values.

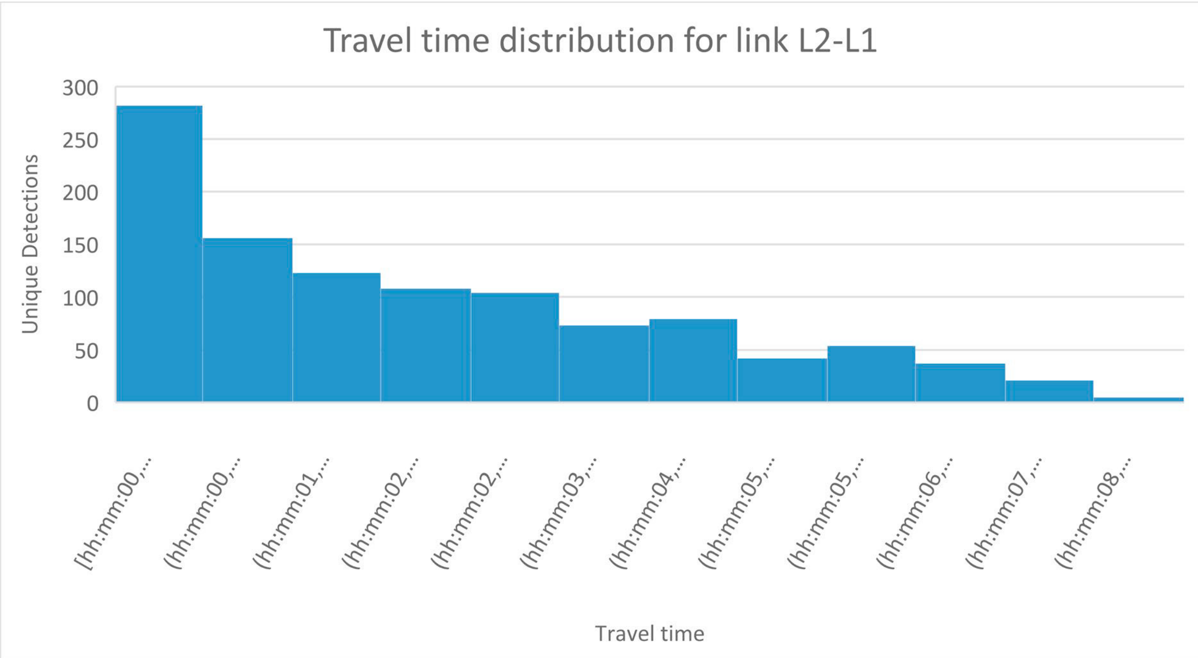


Figure 5: Travel time distribution for vehicles travelling on link L2-L1

Figure 5 indicates the distribution of travel time in downward direction is nearly same as that of the observed detections in upstream direction. The analysis shows that for the studied link, distribution is nearly same in both upstream as well as downstream direction.

Here it is necessary to determine the optimum values for considering the trips because for urban arterial network especially for corridors like the one considered in our study which is highly dense the optimum travel time are greatly influenced by the jam conditions and thus if the travel time is underestimated it will result into less number of through trips which may not be necessarily true and vice versa when the same is overestimated. The optimum time should also be determined such that the influence of cyclist, pedestrians and other non-considerable modes are not included which are alternate modes and may yield highly biased results. Thus, different criteria were considered before determining the optimal value for the travel time and thus it varies based on link for which trips are considered.

6.2 Reliability of mined-class interval travel time

After formation of histograms as well as analysing the movement behaviour for all the links and thereby selecting the optimal range, the specious data values are removed from all the links. This will form a reduced data set containing only through moving vehicles. Further, it is necessary to determine if there are any outlier in travel time range considered for the through traffic. Owing to this, whisker box plots are presented for various links as shown in Figure 6. The Figure clearly states that all the values are within the considerable limit containing no outliers for all major links. This proves the reliability of the methodology considered here for the bifurcation of the through traffic. For all the links considered here, the mean value of travel time is higher to that of the median value.

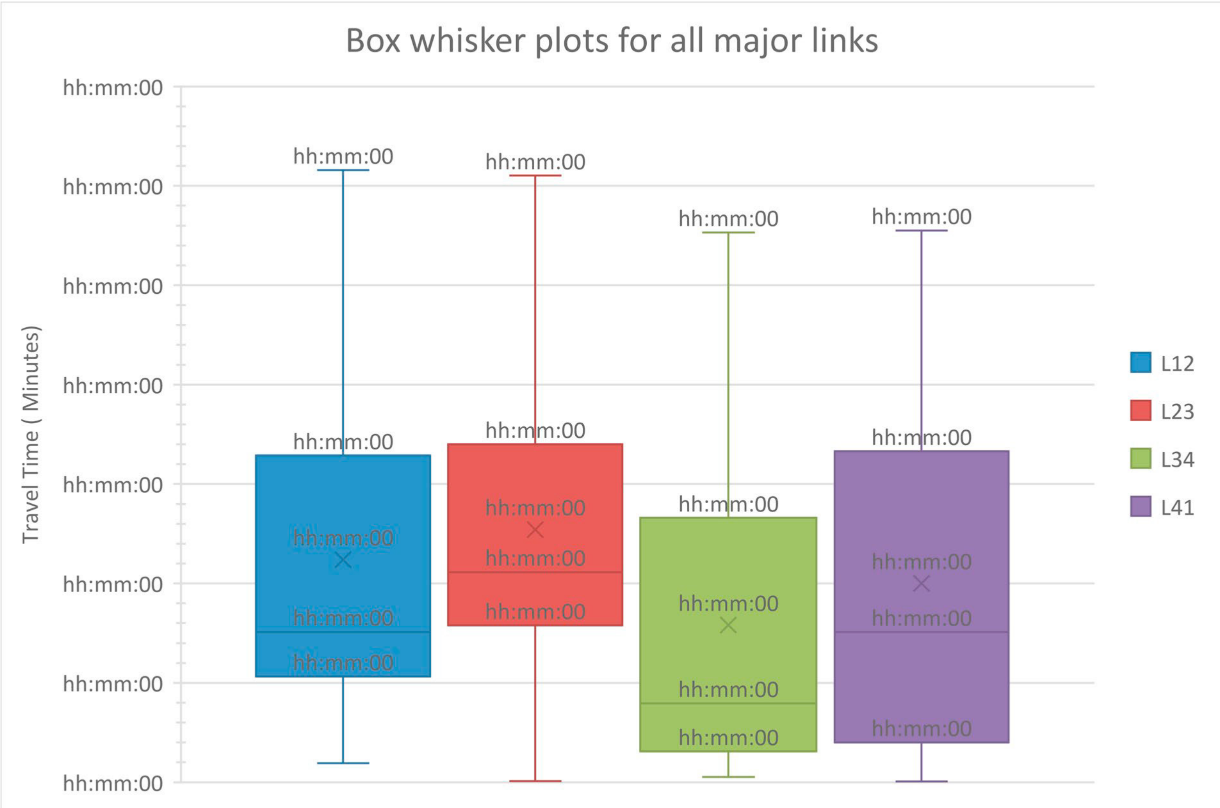


Figure 6: Whisker box plot of cut through travel time for all major routes

Once the reliability of the travel time for through traffic is assured, it becomes necessary to know what kind of distribution the through moving vehicles follows. Therefore, the data set of two different networks were plotted in easy fit software. Easy Fit Professional software was utilized for distribution fitting in this investigation. The product runs the information through 60 likelihood dispersions and furnishes the list of distributions alongside the best fitting curve utilizing the Kolmogorov-Smirnov (KS) Test. Travel time can be clarified utilizing three measurable distributions – Burr, Generalized Extreme Value (GEV) and Pearson 6. In such manner, distribution fitting has been completed at whole length of the investigation period.

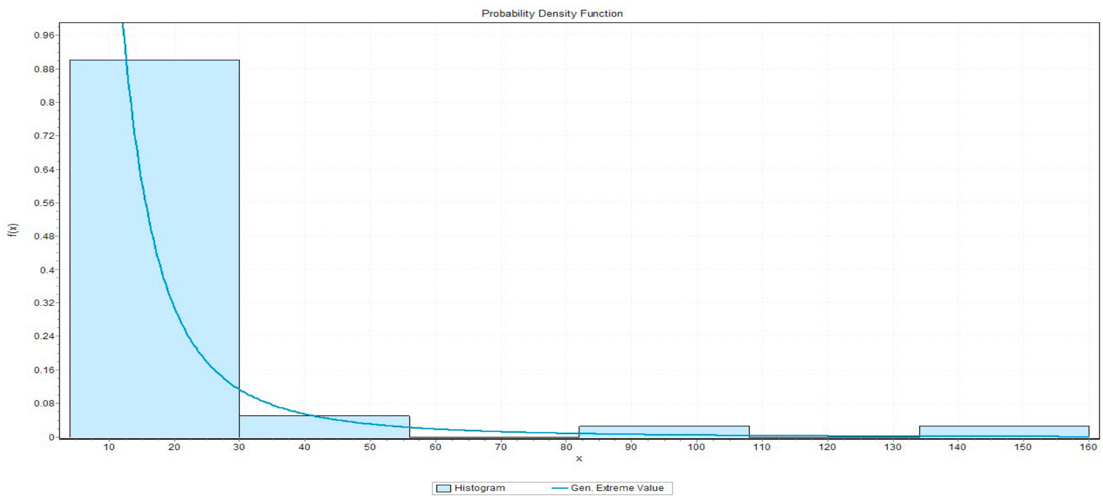


Figure 7: Burrs Distribution for mined class interval on link L1-L2

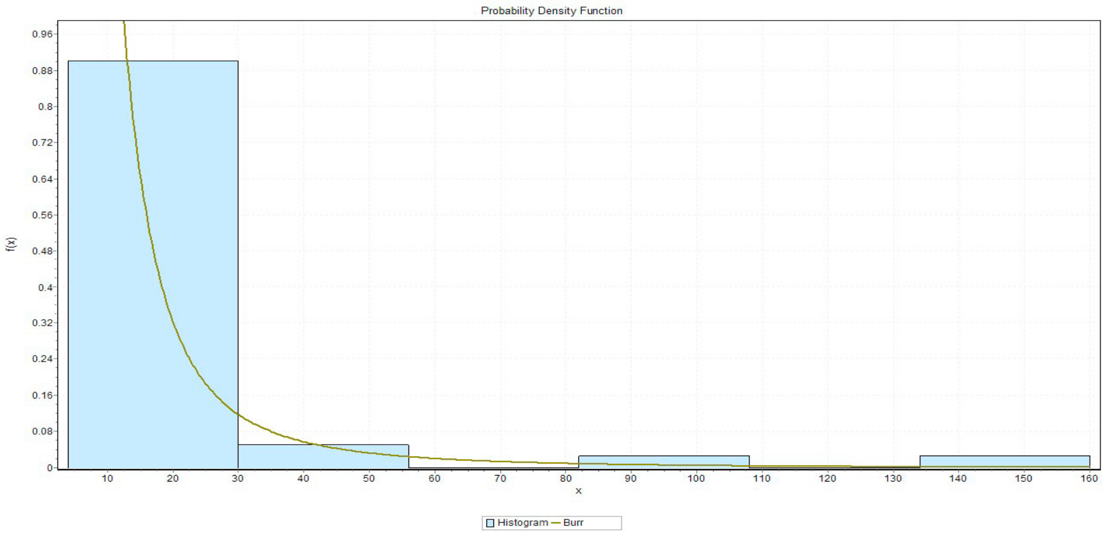


Figure 8: GEV Distribution for mined class interval on link L2-L1

The Figure 7 above shows the plot for the GEV distribution and Figure 8 shows the Burrs distribution; both of them are in the top best distribution rankings. Burrs distribution is better compared to GEV for most of the links so the distribution of same links is stated above.

6.3 Goodness of Fit Details

The goodness of fit is determined based on Kolmogorov – Smirnov test (K-S test) which is a non-parametric test of one-dimensional probability distribution, which is used to compare a sample with some reference probability distribution. Thus, it is a special case of normal distribution where the samples are compared with a normal standard distribution.

Table 7: Goodness of fit for Burrs distribution based on K-S test

Burr					
Kolmogorov-Smirnov					
Sample Size	40				
Statistic	0.09934				
P-Value	0.78841				
Rank	3				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.16547	0.18913	0.21012	0.23494	0.25205
Reject?	No	No	No	No	No

Table 8: Goodness of fit for GEV distribution based on K-S test

Gen. Extreme Value
Kolmogorov-Smirnov

Sample Size	40				
Statistic	0.13075				
P-Value	0.46226				
Rank	10				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.16547	0.18913	0.21012	0.23494	0.25205
Reject?	No	No	No	No	No

In addition, the details for goodness of fit indicates that for all α value from 0.2 to 0.01, the critical value are within the accepted limit and thus are accepted for all confidence intervals. Moreover, the critical values are same for both the above stated distribution as per Kolmogorov-Smirnov analysis indicating that any of the distribution can be considered for the analysis. After assuring the reliability, the primary O-D matrix is reduced to form the secondary matrix for through traffic as stated below.

Table 9: Secondary O-D matrix for through moving vehicles

O/D	**1**	**2**	**3**	**4**
1	-	374	232	176
2	282	-	672	276
3	165	420	-	288
4	222	117	248	-

It is ideally true that if the time distribution and whisker plots are reliable for the through traffic then it should also be possible to form a model that can determine the proportion of through traffic based on the total matchings between two points. In addition, if this data is observed for several days on various links then it can be extrapolated and hence prediction can be carried out.

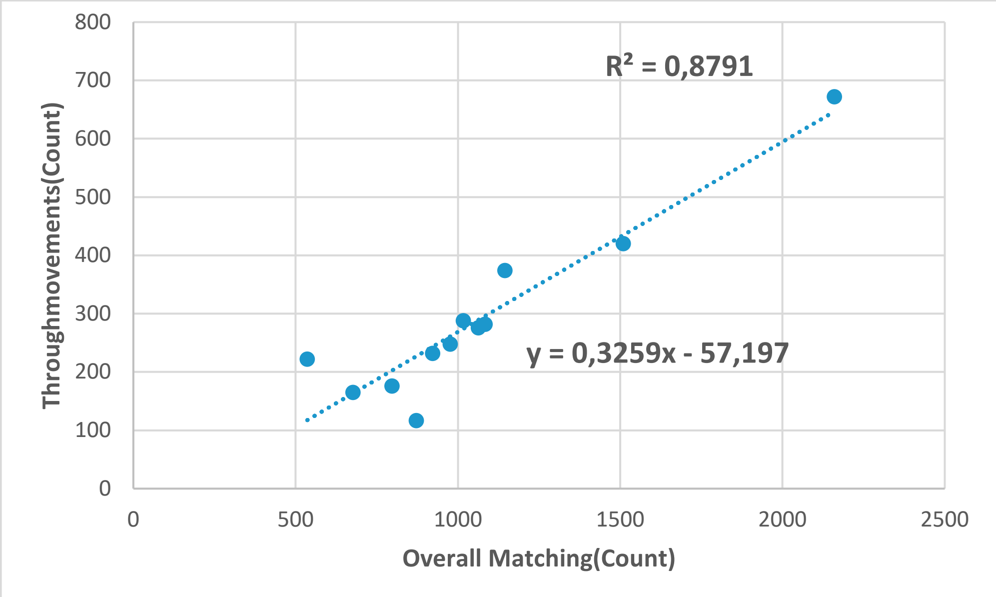


Figure 9: Model determination to predict through movements based on overall matching

The Figure (9) above indicates a plot between the through movements with respect to overall matching between two points. It can be clearly concluded that with increase in the overall matching between two points the through traffic gradually increases. In addition, it should be noted that the function possesses a linear trend with high regression value. Along with this, model was also check for major statistical test and showed competing results for all of them. Such a high R^2 value assures the accuracy of data collection and corresponding adopted techniques for data cleaning and filtration.

6.4 Expansion Factor Determination

In addition, to provide the details regarding the link trips based on MAC ids, the matching detections can also be converted into link volumes by multiplying with some corresponding expansion factors. For our study, attempt was made to determine the expansion factor by placing video cameras at locations where the sensors were already present. However, due to presence of heavy traffic and limitation of the cameras, it was not possible to capture volume from all the directions and thus determination of expansion factor is not taken into consideration for the present study. Another approach can also be carried by combining the annual average daily traffic volume and its comparison with the video graphic volume as well Wi-Fi id volume to yield more robust forecasting of the link volumes.

7. Summary and Conclusion

The study provides a significant insight into the use of Wi-Fi based sensors for development of O-D matrix as well as development of the speed profiles on various links of the study network. The work shows that the new ITS based Wi-Fi technology yields satisfactory results for evaluation of such complex networks without compromising the efficiency of the results. It is quite lucid that the efficiency of the results is largely affected by the placement location of the sensors. The paper outlines the methodology for formation of O-D matrix by matching the MAC ids for the devices along with their analogous time stamps for the duration in which they remain within the detection range. For convenience, the first time stamp was considered while matching the ids between two stations. Here the development of route based O-D matrix is carried out with the development of algorithm, which was used for data cleaning and filtration in the R software. The route based O-D matrix would be useful to forecast the link volumes by multiplying the obtained matching trips with some expansion factors. However, before forecasting the link volume it becomes necessary to reduce the primary matrix to form O-D matrix based on through traffic for which the speed profile is necessary on various links. Literature review have considered the use of some arbitrary value to determine the through traffic on the links based on general logic considering the distance and average speed on the road.

Thus, the study attempts to eliminate this random assumption by determining the distribution of travel time followed by the vehicles on any specific link for the study period based on which the optimal value of travel time of through traffic was determined. Further, to determine the accuracy of this result the distribution followed by the assumed through traffic vehicles was observed with the help of easy fit software that runs through 60 different models to determine the best fit. The software shows that Burrs and GEV are the best fitting models for most of the links. These distributions were checked for various confidence interval and the model showed acceptance for all these confidence intervals. Thus, after the acceptance of these models whisker box plots were also made to determine if any link presented outlier data so that the through traffic travel time can be updated. Ultimately, secondary matrix consisting of only through traffic was thereby determined which can finally be expanded to link volumes. Lastly, an attempt was made to determine the model that can be used to predict the through traffic from the matching detection among links, which would be very fruitful for further applications. Linear model was observed based on affirmative results from major statistical tests. As more studies will be carried out, new methodology with higher accuracy can be developed with precise forecasting methods and many unexplored applications will be inspected.

References

Abbott-Jard, M., Shah, H., & Bhaskar, A. (2013). *Empirical evaluation of Bluetooth and Wifi scanning for road transport.* Paper presented at the Australasian Transport Research Forum (ATRF), 36th.

Abedi, N., Bhaskar, A., & Chung, E. (2013). Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement.

Advani, C., Thakkar, S., Shah, S., Arkatkar, S., & Bhaskar, A. (2019). Exploring Performance of Intercity Expressway Using ITS-Based Technology. In *Innovative Research in Transportation Infrastructure* (pp. 13-22): Springer.

Bhaskar, A., & Chung, E. J. T. R. P. C. E. T. (2013). Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. *37*, 42-72.

Blogg, M., Semler, C., Hingorani, M., & Troutbeck, R. (2010). *Travel time and origin-destination data collection using Bluetooth MAC address readers.* Paper presented at the Australasian transport research forum.

Carpenter, C., Fowler, M., & Adler, T. J. J. T. R. R. (2012). Generating Route-Specific Origin–Destination Tables Using Bluetooth Technology. *2308*(1), 96-102.

Carrion, C., & Levinson, D. J. T. R. P. C. E. T. (2013). Valuation of travel time reliability from a GPS-based experimental design. *35*, 305-323.

Mathew, J. K., Devi, V. L., Bullock, D. M., & Sharma, A. J. T. R. P. (2016). Investigation of the use of Bluetooth sensors for travel time studies under Indian conditions. *17*, 213-222.

Michau, G., Nantes, A., Bhaskar, A., Chung, E., Abry, P., & Borgnat, P. J. I. T. o. I. T. S. (2017). Bluetooth data in an urban context: Retrieving vehicle trajectories. *18*(9), 2377-2386.

Park, S., Saeedi, A., Kim, D. S., & Porter, J. D. J. J. o. T. E. (2016). Measuring intersection performance from bluetooth-based data utilized for travel time data collection. *142*(5), 04016014.

Tahmasseby, S. J. J. o. T., & Engineering, T. (2015). Traffic data: Bluetooth sensors vs. crowdsourcing—A comparative study to calculate travel time reliability in calgary, alberta, canada. *3*, 63-79.

Thogulava, H., Antonovs, V., Bingham, S., Hill, M., & Hanchett, K. (2015). *Developing Origin-destination Matrices Using Bluetooth Data for Strategic Transport Models-Worcester Case Study.* Paper presented at the European Transport Conference 2015Association for European Transport (AET).