USE OF BLUETOOTH DATA FOR MONITORING URBAN TRANSPORTATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

PINAR KARATAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
CIVIL ENGINEERING

JULY 2023

Approval of the thesis:

USE OF BLUETOOTH DATA FOR MONITORING URBAN TRANSPORTATION

submitted by **PINAR KARATAŞ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Civil Engineering, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Erdem Canbay
Head of the Department, **Civil Engineering, METU** ________________

Prof. Dr. Hediye Tüydeş Yaman
Supervisor, **Civil Engineering, METU** ________________

Examining Committee Members:

Prof. Dr. Hediye Tüydeş Yaman
Civil Engineering, METU ________________

Prof. Dr. Pınar Karagöz
Computer Engineering, METU ________________

Assoc. Prof. Dr. Hande Işık Öztürk
Civil Engineering, METU ________________

Assoc. Prof. Dr. Murat Özen
Civil Engineering, Mersin University ________________

Asst. Prof. Dr. Funda Türe Kibar
Civil Engineering, Baskent University ________________

Date: 18.07.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name: Karataş, Pınar

Signature:

ABSTRACT

USE OF BLUETOOTH DATA FOR MONITORING URBAN TRANSPORTATION

Karatas, Pinar
Doctor of Philosophy, Civil Engineering
Supervisor: Prof. Dr. Hediye Tuydes Yaman

July 2023, 111 pages

Due to their low cost, many urban regions are being accessorized by Bluetooth readers (BTR), which can scan any Bluetooth (BT) enabled devices around them. When BT-enabled devices are monitored over an analysis time or across BTRs, traffic state (in terms of speed or travel time) can be estimated in a time-dependent manner. Estimation of these average travel times for urban corridors can further allow generation of input data for Variable Message Signs (VMSs) while average speed for corridor segments helps determination of Level of Service (LOS) values. However, BT data has also some uncertainties and complexities, which need preprocessing and systematic clean-up approach to eliminate outlier values due to either a) very long times caused by slow travel or stopped durations or b) detours often observed in open urban networks. This study focused on development of a BT data analysis framework to eliminate the travel outliers on urban arterials which may stem from vehicles taking detours/subtours or waiting. The cleaned-up data is used to obtain travel time probability dsitrubitions for road links between consecutive BTRs, leading to estimation of travel time confidence intervals. The statistical representation of average link travel time finally allow estimation of urban corridor travel times. Numerical results were obtained from anlaysis of BTR network of 7 intersections along a study corrdior in Mersin. Determination of a 50 seconds threshold for the rescan time allowed estimation of mid-to-mid travel times between 7 BTRs. Use of 4km/hr slow movement filter first eliminated the majority of the extreme values in the travel time data. A second stage of interquartile-range data clean-up further improved the accuracy of BT-based travel time method. Validations was performed via comparisons with the Floating Car Data (FCD) for the same corridor.

Keywords: Bluetooth, Urban Transportation, Data Analysis Procedure, Travel Time, Time Dependent

ÖZ

KENTSEL ULAŞIM DEĞERLENDİRMELERİNDE BLUETOOTH VERİSİNİN KULLANIMI

Karataş, Pınar
Doktora, İnşaat Mühendisliği
Tez Yöneticisi: Prof. Dr. Hediye Tüydeş Yaman

Temmuz 2023, 111 sayfa

Düşük maliyetleri nedeniyle birçok kentsel bölge, çevrelerindeki aktif Bluetooth (BT) özellikli cihazları tarayabilen Bluetooth okuyucular (BTO) ile donatılmaktadır. BT özellikli cihazlar bir analiz süresi boyunca veya BTO'lar arasında izlendiğinde, trafik durumunu (hız veya seyahat süresi açısından) zamana bağlı bir şekilde tahmin edebilmektedirler. Kentsel koridorlar için ortalama seyahat sürelerinin tahmin edilmesi, Değişken Mesaj Sistemleri (DMS'ler) için veri sağlarken; koridor segmentleri için ortalama hız, Hizmet Seviyesi (LOS)' nin belirlenmesine yardımcı olmaktadır. Bununla birlikte, BT verileri, a) yavaş hareketler veya beklemelerden kaynaklanan çok uzun süreler veya b) açık kentsel ağlarda sıklıkla gözlenen sapmalar nedeniyle ön analiz ve sistematik bir yaklaşımla aykırı değerleri ortadan kaldırmayı gerektiren birçok belirsizliğe ve karmaşıklığa sahiptir. Bu çalışma, şehir içi arterlerde araçların yoldan sapma/alternatif yollardan gitme veya beklemesinden kaynaklanabilecek seyahat aykırı değerlerini ortadan kaldırmak için bir BT veri analizi çerçevesinin geliştirilmesine odaklanmıştır. Temizlenmiş veriler, ardışık BTO'lar arasındaki yol kesimleri için seyahat süresi olasılık dağılımlarını elde etmek

için kullanılır ve seyahat süresi güven aralıklarının tahmin edilmesine olanak sağlar. Ortalama kesim seyahat süresinin istatistiksel değerlendirmesi, nihayetinde kentsel koridor seyahat sürelerinin tahmin edilmesini sağlar. Sayısal sonuçlar, Mersin'de bir çalışma koridoru boyunca 7 kavşaktan oluşan BTO ağının analizinden elde edilmiştir. Yeniden tarama süresi için 50 saniyelik bir eşiğin belirlenmesi, 7 BTR arasındaki orta-orta seyahat sürelerinin tahmin edilmesini sağlamıştır. İlk olarak 4 km/saat yavaş hareket filtresinin kullanılması seyahat süresi verilerindeki uç değerlerin çoğunu ortadan kaldırmıştır. İkinci bir aşama olan çeyrekler arası (IQR) veri temizliği, BT tabanlı seyahat süresi yönteminin doğruluğunu daha da artırmıştır. Doğrulamalar, aynı koridor için Hareketli Araç Verileri (FCD) ile karşılaştırma yoluyla gerçekleştirilmiştir.

Anahtar Kelimeler: Bluetooth, Kentsel Ulaşım, Veri Analiz Yöntemi, Seyahat Süresi, Zamana Bağlı

To my son, Uras…

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor Assoc. Prof. Dr. Hediye Tüydeş Yaman for her endless support, valuable guidance, and patience throughout this study.

Besides my advisor, I would like to thank the rest of my thesis monitoring committee members; Assoc. Prof. Dr. Hande Işık Öztürk and Assoc. Prof. Dr. Murat Özen and other jury members Prof. Dr. Pınar Karagöz, and Asst. Prof. Dr. Funda Türe Kibar.

The study was conducted as a part of a TUBITAK project of Integrated Systems and System Design (ISSD) company at METU Teknokent. I also would like to thank ISSD company for the data and technical support during the data collection and analysis period. My sincere thanks to Beyhan İpekyüz, Kerimcan Ürel and Fulya Özsan for the wonderful teamwork during the analysis process.

I would like to thank Meltem Tangüler, Burhan Alam, Gülçin Dalkıç and Beyhan İpekyüz for their friendship, moral support and all the fun we had together during this long period of my life.

I sincerely thank my family; Mehmet Karataş, Güllü Karataş and my dear brothers, Kaan and Cihan for their endless support and confidence during my life. Last but not least, I would like to thank my son, Uras and my husband İsmail Sevinen for being with me and putting up with me throughout my thesis process.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

BLE	: Bluetooth Low Energy
BT	: Bluetooth
BTR	: Bluetooth Reader
BTR_ID	: Bluetooth Reader Name
DIAC	: Device Identification and Authentication Code
FCD	: Floating Car Data
GPS	: Global Positioning System
IQR	: Inter Quartile Range
MAC	: Media Access Control
MAC_ID	: Media Access Control Address
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MPE	: Mean Percentage Error
MSE	: Mean Square Error
OD	: Origin-Destination
TT	: Travel Time
USA	: United States of America

LIST OF SYMBOLS

SYMBOLS

T_{inq}	:	Inquiry Interval for BTR
$t_{m;n}$	:	n^{th} Reading of vehicle m
Δt	:	Rescan time
$\delta t_{m;i}$	:	Stay (dwell) time of vehicle m within the capture zone of BTR i
τ	:	Time interval (15-min, 30-min, 60-min, etc.)
$\Omega_{m,i,j}^{BT,F2F}$	:	First to first travel time of vehicle m, between BTR i to j
$\Omega_{m,i,j}^{BT,L2L}$	:	Last to last travel time of vehicle m, between BTR i to j
$\Omega_{m,i,j}^{BT,M2M}$	:	Mid to mid travel time of vehicle m, between BTR i to j
$\Omega_{m,i,j}^{BT,F2L}$	:	First to last travel time of vehicle m, between BTR i to j
$\Omega_{m,i,j}^{BT,L2F}$	:	Last to first travel time of vehicle m, between BTR i to j
$\Omega_{m,i,j,\tau}^{BT}$	:	Travel time of vehicle m, between BTR i to j within time interval τ
$L_{i,j}^{BT}$	:	Length between BTR i to j
$\overline{\Omega}_{i,j,\tau}^{BT}$	:	Average travel times between BTR i to j within time interval τ
$\bar{u}_{i,j,\tau}^{BT}$	:	Average BT speed between BTR i to j within time interval τ
$\bar{u}_{i,j,\tau}^{FCD}$	:	Average FCD speed between BTR i to j within time interval τ

CHAPTER 1

INTRODUCTION

Real time data collection in traffic engineering is extremely an important and highlighted topic for better traffic corridor management and optimization. In the literature, various data collection methods have been studied including magnetic loops, road tube counters, piezo sensors, radars, Floating Car Data (FCD), Wi-Fi, GPS/cell-phone tracking, AVL (Automatic Vehicle Location) systems used to estimate the link/ corridor travel time, average speed or density and Origin-Destination (OD) matrices (Antoniou et. al., 2010). Bluetooth data has recently become a popular data source for traffic studies and it is also considered as a part of Intelligent Transportation Systems (ITS) concept.

1.1 Bluetooth Technology and Data

The working principle of the Bluetooth relies on capturing specific Media Access Control address (MAC_ID) of any Bluetooth device by a Bluetooth reader (BTR). When the BTR locatios are fixed and MAC_IDs are unique, the analysis of the Bluetooth data enables the monitoring of movement of Bluetooth equipped vehicles (or travelers) in a network; which can later be used to estimate speed, or travel time.

Bluetooth data collected at the BTR locatioıns is defined as a low-cost, high reliability, accurate and continuous data source. Ahmed et al. (2008) stated the most essential benefits of the Bluetooth data as i) cost-effectiveness, ii) flexibility, iii) extensibility and iv) being widely-adopted. This technology has become quite common and is observed in 5% to 10% of vehicles on the roadway (Day et. al., 2010). Furthermore, it is also used by passengers in vehicles.

Bluetooth data has also been widely used in the literature for estimating travel time and origin-destination (OD) in transportation studies. The specific MAC_ID value enables the estimation of travel time information, which is a useful data source that comes to the forefront in the decision-making process of users when initiating many trips. In addition, OD estimation is a very difficult and comprehensive field could contribute to infrastructure and network planning issues, especially due to the open traffic network in the cities. Average speeds on links also enabled to capture the incident/congestion locations, or queue lengths near the signalized intersections, besides Level of Service (LOS) estimation. It is an effective data set for decision makers because the speed evaluations reveal important results, particularly on the effective use and performance of road capacity. For a long-term data, travel time information provides the determination of the corridor characteristics especially for urban roads for a specific time or time-dependent manner. Many other studies are being conducted to investigate the use of Bluetooth technology in other transportation-related fields such as incident detection, arterial signal coordination, passenger waiting times, route choice estimation, pedestrian and bicycle travel, parking management systems, and so on.

1.2 Challenges of the Bluetooth Data

However, this technology is still relatively unknown among transportation engineers, has a variety of characteristics that have a direct and significant impact on data quality for transportation studies. There is no information about the exact location where the reading was captured during data collection for this reason, how to calculate the stay time period within the BTR capture zone and how to include it in the travel time calculation should be examined in detail. Besides these, there isnot any specific information on whether the captured MAC_ID is a vehicle or a device inside a vehicle or even any other stationary device within the capture zone. Although there are many studies on Bluetooth technology features that some changeable features have a direct effect on the collected data; the effect of such parameters on transportion has not been fully investigated.

In the process of applying technology in our country, many of our cities are equipped with BTR devices to bring out real-time travel time information to drivers by variable message signs (VMS). With these developments demanded by municipalities, travel time estimation issues with Bluetooth data have become more popular in our country. However, a thorough data pre-analysis methodology is required to ideftify the characteristics of the BT data in the analysis process and also technological Bluetooth features. It is crucial to understand exactly what kind of information can be obtained using Bluetooth, how the data can be used in urban transportation and what are the meaning of the results determined from the data. At the same time, it is important to determine the data properties originating from the BTR device in detail and to collect them with the same standardization throughout the country in terms of creating the ITS architecture. Within the scope of the preliminary analysis, to minimize the errors that will occur in the data collection and clean up the data, it is necessary to determine a flexible and comprehensive method that can be updated depending on the time and regional differences.

1.3 Scope and Objective of the study

The main purpose of this thesis is to determine the usage of Bluetooth data in urban transportation, the conceptual definitions, the contributions it provides and the constraints that arise. It is crucial to first thoroughly and accurately assess the resources/ collected data in order to make an appropriate estimate or use of them, as well as to understand how to solve possible limitations. To determine what kind of information BTR data provides and how this information can be used in the forecasting process, the collected data should be examined under many different situations and extreme conditions have been determined.

When data collection errors or specific features are evaluated and eliminated; a new, flexible and comprehensive method will be introduced to estimate travel time. At this stage, each unique MAC_ID in the data set is subjected to a data cleaning process in order to be grouped as "possible moving vehicle" considering its movement in the network. To provide qualified and reliable travel time estimation all the process is conducted on the possible moving vehicles. By using this methodology, corridor travel times are calculated and accuracy of the estimations are conducted with respect to FCD data source. This possible link travel times can later use to estimate time-dependent urban arterial travel time (TD_UrbArt_TT) and speed (TD_UrbArt_S) to inform the user on expected travel times of possible destinations by VMS and to capture possible emergency or traffic situations by speed estimation. During these studies, an interim study was conducted to evaluate the effect of the inquiry interval parameter, which is stated to have an effect on data quality in studies in which Bluetooth technology was examined in the literature, and an update was made in this device feature, which is presented as one of the important contributions of the study.

1.4 Structure of the Thesis

Chapter 2 mainly presents the characteristics and structure of Bluetooth data and background of the study by explaining the usage of Bluetooth in transportation studies. Chapter 3 represents the conceptual definitions and methods discussed in the literature and highlights the possible limitations with a sample conceptual data set. For a comprehensive evaluation of the methodology proposed in Chapter 4, case study locations and data are presented in detail. The determination of the threshold values required for the solution of these extreme cases and the examination of the effect of these values on the required calculations. The contribution of the data analysis methodology presented is discussed by small case examples. To examine its contribution, link travel time and speed values for BTR pairs on the selected corridor are calculated at 1-hour and 2-hour time intervals for 4 different days and two directions. These results are compared with FCD data and an expected link travel time interval was determined in 95% confidence interval for selected 6 BTR pairs. This link travel time estimation is later used to estimate corridor travel time.

As an additional information and detail study, Chapter 5 presents the impact of the Bluetooth data collection feature, the inquiry interval, on transportation data. The thesis concludes with the conclusions and further recommendations presented in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Bluetooth Data and Structure

Bluetooth® is a low-power wireless connectivity technology used to stream audio, transfer data and broadcast information between devices (https://www.bluetooth.com). This technology was invented by telecom vendor Ericsson in 1994 and uses a 2.4 GHz short-range radio frequency spectrum. Bluetooth data collection depends on scanning its capture zone and collect all Media Access Control (MAC) addresses of available devices (Figure 3.2). Capture zone differs according to Bluetooth reader type, which categorized as Class 1 to Class 3 as provided in Table 2.1 and it can be increased by additional antennas to receivers. For each Bluetooth enabled device, a unique MAC address exists in the form of "AB:CD:**:**:**" which consists of 48-bit electronic identifier.

Table 2.1 Bluetooth Reader Class Classification

Class	Transmission Power	Range
Class 1	100 mW (20dBm)	100m
Class 2	2.5 mW (4dBm)	10m
Class 3	1 mW (0dBm)	1m

Bluetooth is a promising technology for short-range, low-power wireless communications, but its device discovery procedure can be slow and inefficient

(Jiang et. al., 2004). The device discovery procedure is used to establish a connection between two Bluetooth devices, and it involves several steps, including inquiry, paging, and frequency hopping. The inquiry process is used to discover nearby Bluetooth devices, and it can take a long time due to the large number of possible Bluetooth devices and the limited number of inquiry channels. To sum up, inquiry is the capture capability of a BTR, which has two parameters, one is changeable (inquiry interval) and one is standard (inquiry scan interval) for any specific BTR (Table 2.2). When inquiry interval is determined as a value, rescan time of collected data can be moved up to inquiry scan interval times higher than this value. For smaller inquiry intervals, data will be increased because of the increased number of readings. Paging process is used to establish a connection with a specific Bluetooth device, and it can also take a long time due to the frequency-matching delay. The frequency hopping process is used to avoid interference with other wireless devices and can further increase the delay. In a similar manner, Thamrin and Sahib (2009) reflected that the connection process can be affected by the synchronization problem, which causes device and discovery delay. To optimize Bluetooth connection time, the paper proposes three possible changes - reducing the random backoff parameter or eliminating it, using a single frequency train instead of two, or combining both ideas.

Table 2.2 Timing Parameters of Inquiry and Inquiry Scan (Jiang et. al., 2004)

Parameter	**Description**	**Recommended value**
$\mathbf{T_{inquiry}}$	inquiry interval	60s
$\mathbf{T_{w\ inquiry}}$	inquiry window length	10.24 s
$\mathbf{T_{inquiryscan}}$	inquiry scan interval	1.28 s
$\mathbf{T_{w\ inquiryscan}}$	inquiry scan window length	10ms
$\mathbf{T_{train}}$	length of a train	10ms
$\mathbf{N_{inquiry}}$	train repetition number	≥256

In the literature, the impact of inquiry is examined to ascertain how this variable parameter affects the data acquired data quality (Zaruba and Chlamtac, 2003, Peterson et. al., 2006 and Ramli and Hasbullah, 2010). Zaruba and Chlamtac (2003) demonstrated through simulations that their accelerated inquiry approach outperforms the original Bluetooth inquiry scheme significantly in terms of discovery times. Results revealed better discovery times with accelerated inquiry process according to original conditions. The dynamics of the Bluetooth discovery process involve the inquiry and inquiry scan substates (Peterson et. al., 2006). Optimizing inquiry time in Bluetooth-enabled devices involves reducing the inquiry substate duration to minimize the time spent in the inquiry substate, which can lead to a significant reduction in power consumption and an increase in throughput. The study provides valuable insights into the Bluetooth discovery process and can help to improve the performance of Bluetooth-enabled devices and emphasized that low inquiry times reduce power requirements and the interaction between closer BTRs by increased data. Ramli and Hasbullah (2010) proposed an algorithm to improve the efficiency of the inquiry process while conserving power and presents a mathematical model and simulation results based on modification in its parameters.

The format of the collected data may alter depending on the trademark of the Bluetooth reader, but the substance logically remains the same, which includes mainly Read Date and Time information, MAC address, Device type, etc. Within this main information, MAC address enables to monitor and track devices, and capture time enables to investigate travel time information, direction of movement, etc.

2.2 Bluetooth Use in Traffic Studies

As Bluetooth is one of the low-power, low-cost, short-range wireless communication systems, it is mostly used for in-vehicle applications to provide a wireless network for integrating mobile electronic devices with the vehicle (Chen and Chen, 2005). This establishes an active connection between the vehicle and any BTR located in the road network, allowing vehicle movement to be monitored as well. As an early attempt, Murphy et. al. (2002) focused on the feasibility of using Bluetooth technology for short-term ad hoc connections between moving vehicles and investigated potential ways to modify the inquiry process to decrease connection setup times. Results revealed that Bluetooth has limitations in range and bandwidth, but it is suitable for use in a highly mobile environment with proper modifications especially to the inquiry process.

Barceló et. al. (2010a) provided a comprehensive overview of the potential of Bluetooth technology for improving traffic management and forecasting in transportation systems. Authors explained how Bluetooth sensors work, what kind of data they capture, and how this data can be used to estimate time-dependent O-D matrices and short-term travel time forecasts. However, it is concluded that more research is needed to develop more accurate and reliable methods for expanding Bluetooth samples. Bluetooth data can be used in transportation for a variety of purposes, such as preventing accidents, planning routes, collecting tolls, and more (Barrales-Guadarrama, 2010). Bluetooth data can also be used to provide real-time information to drivers about traffic conditions, road closures, and other important information. Additionally, Bluetooth systems can be used to collect tolls electronically, reducing the need for physical toll booths and improving traffic flow. With a general perspective, arterial performance measures; average speed, travel time and OD estimation, were investigated by Bluetooth MAC data for 27 days on a 2.5 mi suburban signalized arterial in Oregon, Portland (Quayle et. al., 2010). To evaluate general reliability (technology evaluation), traditional GPS floating car method was employed, and results revealed that larger data sets of Bluetooth could effectively capture the performance characteristics. Evaluation of before-and-after signal timing performance measurement reflected that Bluetooth could capture the changes as a low-cost technology. In the study outliers can take the form of pass-by (discontinuous) trips along a corridor or non-auto trips along a corridor, which tend to be longer in duration than continuous auto trips. While these longer trips or outliers may not be meaningful for travel time estimations, they can provide valuable information on the types of corridor trips and even the value of progressing traffic between MAC reader stations. The study suggests that outliers should not be discarded from the data set but should be screened in relation to the objective of the study. Overall, the study makes the case that the automated, low-cost nature of Bluetooth devices may completely change how data is gathered and used in the transportation industry. Likewise, a methodology was developed to distinguish vehicular movements and travel characteristics (specifically OD and corridor travel time estimation) from Bluetooth data by a case study of Ankara, Turkey (Yucel et. al., 2013; Yucel, 2015). Despite low penetration rates in the study, Bluetooth data was found effective in travel time estimation for an urban corridor, but for OD estimation. To estimate reliable OD matrices, there is a need to compare the data with any other supportive data source and it should be improved with further studies for validation purposes. In the following years, with a thesis study on estimation of travel time in intelligent transportation systems with the use of Bluetooth sensor data in Turkey, the main trends in traffic behavior were observed and the Bluetooth activation, whether there is more than one active Bluetooth device in the vehicle, etc. situations have been studied (Koçak, 2021). Within the scope of another thesis, the data obtained from the Bluetooth-based traffic monitoring system established in the city center of Konya was verified by comparing with traffic density, travel time estimates and other traffic data sources (Soykök, 2021). It has been concluded that Bluetooth-based traffic monitoring systems can be used as an effective tool in traffic management and that the predictions made should be verified with another traffic data source (Yandex Traffic, Google Maps, etc.) to test the reliability of Bluetooth data.

Porter et al. (2013) presented experimental results on antenna characterization in the context of using Bluetooth technology to evaluate the performance of transportation services. The study uses a single Bluetooth reader to characterize five distinct types of antennas and two Bluetooth readers to obtain trip time samples using the five different antenna types. The use of Bluetooth technology in building a sensing platform to collect traffic data such as the actual number of vehicles, vehicles' speed, vehicle's positions, queue length, and lane blockages through tracking vehicles at urban signalized intersections and streets are discussed by investigating the inquiry process in Bluetooth specifications (Mohrehkesh, and Nadeem, 2014). As a result, the standard inquiry process, which can take up to 10 seconds to receive a response from a slave, was studied theoretically as a way to improve it.

Later, Diaz et. al. (2015) focused on the effect of Bluetooth technology on travel time estimation, analyzed the specific Bluetooth technology features that have an impact on travel time estimation. A thorough technique was developed to address problems like intrinsic errors, multiple detections, and outliers to interpret the data collected by a Bluetooth traffic monitoring device and produce a reliable travel time estimate. This methodology emphasized the difficulties of removing outliers and the requirement for the algorithm to adapt to the actual traffic circumstances while still offering highly reliable travel time forecasts with a 5-minute resolution. The functionality of two advanced Bluetooth devices in combination with traditional Bluetooth technology and compares the data gathered with benchmark data sets (from manual counts, radar data, and floating car data (FCD)) to assess Bluetooth technology's advancements in terms of travel time and segment speed (Cotten et. al. 2020). The results demonstrated that compared to Bluetooth Demodulator (BT DM), Bluetooth Low Energy (BLE) achieved much higher matched rates.

In addition to the benefits of the Bluetooth data source in the field of transportation, it should be considered that it also causes many difficulties and limitations in the data analysis process. Some of the main limitations mentioned by Yildirimoglu (2021) are listed as; i) missed detection rates, which can happen when Bluetooth devices are not recognized by the sensors because of elements like limited battery life or weak signal, ii) alternative routes between sensor zones that make impossible to determine trip durations and courses for vehicles with any degree of accuracy. iii) low quality of the estimations by multi-mode traffic assessments, and direct measurement approaches due to low sample rates and anomalous journey time observations.

Besides these general focus studies with Bluetooth data, there are numerous studies focusing on one side to traffic monitoring such as travel time estimation, OD estimation, incident detection, arterial signal coordination, pedestrian travel studies, etc. as categorized in Table 2.3 respectively. These studies will be discussed in detail under specific subtitles.

2.2.1 Bluetooth for Travel Time Estimation

It can be clearly seen that Bluetooth data are very successful on travel time estimation studies especially for segment level, which can be proved with case studies in literature. If any MAC addresses are simultaneously logged at multiple locations, the unique MAC addresses can be matched, and the difference in time stamps can be used to estimate the travel time (Wasson et. al., 2008). The use of time-stamped MAC addresses from Bluetooth devices was investigated to estimate freeway and arterial travel times in Indianapolis, USA. Directional travel times for freeway and corridor travel times for both freeway and arterial corridor analyzed. Results revealed that arterial data have larger variance due to the impact of signals and the noise, introduced when motorists briefly divert from the network.

For quantifiable travel mobility metrics for a rural interstate highway in Indiana, travel time profiles were collected by 3 Bluetooth receivers for 12-week period (Haseman et. al., 2010). Researchers focused on traffic volumes, travel times and hours of delay which exceeds 10 min for study periods and investigated the effects of crashes and opening/closing lane situations. “Travel time delay” was found to be a meaningful data source to validate volume and capacity threshold in which queuing conditions occur.

Table 2.3 Major Bluetooth Data Based Studies in Literature

Topic	Studies	
Travel Time Estimation	• Wasson et. al., 2008 • Haseman et. al., 2010 • Haghani et. al., 2010 • Tsubota et. al., 2011 • Martchouk et. al., 2011 • Van Boxel et. al., 2011 • Aliari and Haghani., 2012 • Saeedi et. al., 2013 • Moghaddam and Hellinga, 2014 • Namaki Araghi et. al., 2016	• Park et. al., 2016 • Mathew et. al., 2016 • Erkan and Hastemoğlu, 2016 • Remias et. al., 2017 • Liu et. al., 2020 • Civcik and Kocak, 2020 • Ghavidel et. al., 2021 • Yildirimoglu, 2021 • Jedwanna and Boonsiripant, 2022 • Altintasi et.al., 2022
Origin-Desitination Estimation	• Barceló et. al., 2010a • Blogg et. al.. 2010 • Barceló et. al., 2010b • Carpenter et. al., 2012	• Chitturi et. al., 2014 • Michau et. al., 2014 • Montero et. al., 2015 • Dunlap et. al., 2016
Incident Detection	• Yu et. al., 2015 • Salem et. al., 2015 • Margreiter, 2016	• Karatsoli et. al., 2017 • Mercader and Haddad, 2020
Other Studies	**Arterial Signal Coordination**	• Day et. al., 2010
	Passenger Waiting Time	• Bullock et. al., 2010 • Remias et. al., 2013
	Trip Behaviour	• Crawford et. al., 2018 • Remias et. al., 2013
	Route Choice	• Hainen et. al., 2011 • Garrido-Valenzuela et.al., 2020
	Demand Estimation	
	Mobile Ticketing	• Cipriani et.al., 2021
	Parking Management Systems	• Ferreira et.al., 2020 • Chien et. al., 2020
	Safety Analysis	• Yuan et. al., 2018
	Pedestrian Travel	• Malinovskiy et. al., 2012 • Yoshimura et. al., 2017
	Bicycle Travel	• Abedi et. al., 2015 • Ryeng et. al.,2016
	Public Transit Ridership	• Pu et. al., 2021

Haghani et. al. (2010) discussed the use of Bluetooth sensors as a new and effective means of data collection for measuring the quality of real-time travel time provided by traffic surveillance systems. The time difference of the ID matches provides a measure of travel time and space mean speed based on the distance between the successive stations. In the study, a four-step offline filtering algorithm which was designed to extract ground truth from the pool of Bluetooth observations was used starting to identify and discard outliers among single observations in each time interval.

Arterial traffic congestion analysis using Bluetooth data was examined to investigate link travel time (direct measurement of travel time between pairs of scanners) and duration change (time spent by Bluetooth devices to pass through the detection range of Bluetooth scanners) depending on the congestion level (Tsubota et. al., 2011). The findings highlighted the characteristics of Duration and address future research needs to make use of this important data source, such as the need for further development of filters to eliminate biases in Bluetooth samples and for more reliable estimation of travel time and Duration.

The travel-time data was collected for two weeks (by deducting the time stamp at the downstream station from the upstream one when the same MAC address is recorded at both upstream and downstream station), and intervehicle and inter-period variability was analyzed along two segments of I-69 in Indianapolis (Martchouk et. al., 2011). The study provided a framework for collecting and analyzing travel-time data using Bluetooth probe data and supplementing it with corresponding remote traffic microwave sensors data which has great potential to be able to study and eventually predict, with reasonable accuracy, average travel times and travel-time variability.

Van Boxel et. al. (2011) suggested using Bluetooth technology to collect travel times and vehicle speeds in an original and anonymous manner. However, it is mentioned that there are certain potential errors that must be addressed and fixed. Results revealed that the methodology is evaluated for both Interstate highways and urban arterial corridors and is found to be effective at identifying Bluetooth outliers and capable of working in a real-time environment. Aliari and Haghani (2012) proposed a validation technique including planning and choosing road network portions for evaluation, deploying sensors, and processing recorded data for travel time data gathering in transportation systems using Bluetooth sensors. Bluetooth sensors are a useful tool for gathering accurate travel time data, but also have some typical causes of error and limitations.

Bluetooth data was used to collect accurate and precise travel time data between signalized intersections and received signal strength indicator (RSSI) to improve accuracy of intersection- to- intersection travel time samples on a busy arterial road in Tigard, Oregon (Saeedi et. al., 2013). 5 consecutive signalized intersections were donated by Bluetooth data collection units (DCUs). The study emphasized the features of Bluetooth data as multiple MAC addresses detected in a fixed time, multiple detection of single MAC address, different detection of single MAC address by different addresses and same timestamp of different MAC addresses. In this study first-to-first travel time calculation was used to investigate travel time estimation and GPS-based floating car data was to validate it. Furthermore, for travel time evaluation, four different methods (fist-to-first, last-to-last, average-to-average and RSSI based methods) mentioned in the literature were compared, and t-test analysis results reflected that RSSI-based method was significantly differ from other three.

With a parallel focus, Bluetooth data was collected to analyze prediction of near-future travel times on signalized intersections using k nearest neighbor technique by identification of historical data on a 3.1 km arterial roadway in Ontario, Canada (Moghaddam and Hellinga, 2014). For real-time arterial travel time prediction, the most important technical problems defined as presence of outliers and inconsistency in the order of available travel time data, should be detected. For model calibration, field data was employed through optimization of model parameters, and performance of the model was investigated by a separate sample of field data. Although this model has good capabilities, it should be further improved by employment of time series model besides k-nn pattern, extended to predict along a route consisting of multiple segments and should be applied another roadway sections.

Namaki Araghi et. al. (2016) focused on estimation of mode-specific travel time with Bluetooth sensors, by detecting device type and radio signal strength indication (RSSI), located on different road sections; highway/motorway, bicycles paths and arterial road. The accuracy of the proposed method was evaluated by manual deciphering of video recordings and travel times obtained by automated number plate recognition (ANPR) and a clustering method. For mode classification by Bluetooth data, two methods were employed: static cutoff and clustering, by defining upper and lower limit of travel time for each mode and grouping of data according to relevance consecutively. Results reflected that the proposed method has almost same accuracy level with ANPR under mixed traffic conditions with a Mean Absolute Percentage Error (MAPE) of 18% and 17%, respectively.

In a similar manner, estimation of intersection performance using time/ duration data from Bluetooth- based Data Collection Units (DCUs) and a method to reduce variability of this data were introduced for consecutive 5 signalized intersections (1.61 km distance between each intersection) along a high-volume urban arterial in Tigard, Oregon (Park et. al., 2016). Duration time was defined by multiple detection of same MAC address within the road length covered by the DCU's antenna and used for estimation of control delay. The strength of BT device signals was measured by Received Signal Strength Indicator (RSSI), which also used to reduce variability of duration data due to signal strength differences and generate "modified duration data". Researchers suggest using performance measures as control delay and approach delay with wireless communication technology data for intersection performance estimation.

Reliability and evaluation of travel time data for two alternative routes were investigated in Chennai, India (Mathew et. al., 2016). The corridors were 6.1 and 9.5 km, respectively, and donated by 4 different BMS. The penetration rates of two corridor were 7.11% and 10.04%, however the match rates were 3.93% and 7.69%, respectively. The most captured device was cell phone (62 and 72% for each corridor). To investigate travel time reliability of BT, some of the reliability measures were analyzed which were Mean, Standard Deviation, Travel time index, Buffer time index and planning time index. Within these measures planning time index reflected biased results by congestion, while mean, standard deviation and travel time index predicted the reliability accurately when compared to others. A follow up study, feasibility of BT probe vehicle technology was determined at a typical corridor in Chenneai, India to expand BT usage (Remias et. al., 2017). The 2.4 km section of an urban corridor was donated by 5 Bluetooth Monitoring Stations (BMSs) for a week. In the concept of the study, penetration rate analysis, device type detection, different BT sensor performance and travel time analysis of each direction was determined along the study area. Results reflected that penetration rates differed from 0.65 to 10% according to BT device type and time of day (peak vs off-peak), and the highest percentage of device were cell phones (%59). Travel time determination showed that there was a serious pedestrian movement along the corridor, which caused outliers, and significant multi-modal traffic with bicycles, buses, motorcycles, etc. However, this analysis could not reflect any information about any movement between BT data collection methods, even they stop at a part and continue movement, and this should be further evaluated using different algorithms such as delay and reliability analysis. The applicability of Bluetooth was investigated at Bogazici Bosporus has only one entry and exit point with heterogeneous traffic in Istanbul, Turkey (Erkan and Hastemoğlu, 2016). The study performed penetration rate and vehicle classification in addition to stream travel time method analysis. Two BT locations were selected with a distance of 850 m, within 1560 m section length, and also recorded on video during two peak periods of a day. During the study period, the penetration rate was found to be 5 % and 91 % of the vehicles were Light Motor Vehicles or Two Wheelers. The result reflected that estimation of stream travel time by such a limited data was challenging but established a linear relationship between speeds of different classes. However, this study should further be investigated by different data collection methods.

Liu et. al. (2020) investigated the accuracy of Bluetooth technology in complex urban traffic environments to provide accurate travel time information. To examine and quantify the effects of numerous detection issues and the noise in Bluetooth travel time estimations on the accuracy of average Bluetooth travel times, five different Bluetooth travel time-matching approaches i)First - First (F-F), ii)First - Last (F-L) Last – iii) First (L-F), iv) Last - Last (L-L), and v)Average - Average (A-A) were used. Large travel time outliers are identified and removed using the adopted Kalman filtering algorithm, but small estimates, such as zero-second estimates, cannot be identified as effectively because the differences between them and normal estimates are typically too small to activate the algorithm.

By exhibiting penetration rate analysis, class recognition, and flow trip time estimation, Civcik and Kocak (2020) examined the possible use of Bluetooth data as a traffic sensor. According to the investigation, Bluetooth is a practical and cost-effective travel time estimation technique for heterogeneous traffic situations. A striking result of the study is that as the distance between Bluetooth stations increases, the measurement error decreases.

An examination of travel-time variability using Wi-Fi and Bluetooth data was conducted by Ghavidel et. al. (2021) over a 12.5 km section of Tehran, Iran's Resalat Highway utilizing data from 12 Wi-Fi and Bluetooth sensors. When compared to other models in the literature, the authors suggest a new conditional random-effects model that depicts time-varying travel time as a continuous phenomenon and better fits real data. According to the study, there is a significant correlation between the mean and standard deviation of travel times as well as between these two variables and the time of day.

Yildirimoglu (2021) examined the difficulties and constraints associated with using Bluetooth data sets in transportation studies for calculating travel times and vehicle paths. A joint method was proposed for simultaneously inferring vehicle paths and travel times using Bluetooth records and presented with a case study in Brisbane, Australia. The suggested approach performs noticeably better than a naive model where travel times are calculated by direct matching, according to results confirmed against travel time data.

A data-driven approach was adopted to investigate the use of Bluetooth technology for traffic monitoring and travel time forecasting in urban areas (Carrese et. al., 2021). Various statistical models are tested within a methodology framework in which the Bluetooth data is filtered, cleaned, and fused for a more accurate prediction of path travel times. Like mobile Floating Car Data (FCD), Bluetooth technology provides a low-cost extension of the spatial coverage of traffic data while preserving the high sample rates typical of stationary systems. This makes it possible to gather and analyze a greater variety of traffic data, which can then be utilized to implement sophisticated forecasting models and techniques to enhance traffic status estimation.

To estimate traffic information including journey time, link speed, and origin-destination estimations, Jedwanna and Boonsiripant (2022) analyzed the application of Bluetooth technology in transportation. The study also examined the effect of vehicle speeds on Bluetooth detection performance and created a data-processing approach to predict travel time based on Bluetooth transactional data. It was discovered that Bluetooth performance improved with a reduction in vehicle speed and that the mean absolute percentage deviation shrank with an increase in penetration rate.

The distribution of travel time information evaluated from different perspectives in which BT-based traffic data were obtained from five consecutive signalized intersections located in Mersin, Turkey during morning peak hours of 07:30-09:30 for the two weekdays (Altintasi et. al., 2022). The findings showed that, with proper sampling rates, the data had a great deal of effectiveness in calculating travel times and monitoring urban traffic. However, the filtering process needs to be carefully managed to separate motorized movements from non-motorized ones.

2.2.2 Bluetooth for O-D Estimation

Origin-Destination (OD) matrix estimation is inherently underdetermined, meaning that there are many possible OD matrices that could explain the observed traffic patterns from traffic counts and automated vehicle identification data (Van der Zijpp. 1997). This makes it important to use additional information, such as link volumes, travel time data or any kind of data collection systems that will provide more intensive information to help constrain the problem. Therefore, the Bluetooth data collection method can provide vehicle-specific information which makes it easy to use OD estimation.

Bluetooth and Wi-Fi detection quality for vehicles were investigated to forecast short-term travel time and estimate time-dependent OD matrices at an urban freeway in Barcelona with 11 entry and 12 exit ramps (Barceló et. al., 2010a). Sensors were located at each entry and exit ramps and in the main corridor immediately after each ramp. In the concept of the study three formulations were introduced for OD pattern, OD flows and deviations from historic OD flows as state variables proposing ad hoc linear Kalman filtering approaches. Results of simulation data proved that proposed approach was suitable for uncongested and congested conditions, however initializing could be critical for some situations, so further research is necessary. The study continued with the ad hoc Kalman filtering approach to explore Bluetooth data quality for travel time forecasting and time-dependent OD matrices by a case study at Barcelona, Spain (Barceló et. al., 2010b). The data was collected for two months as a historical database and filtered out together with real time detection taken for forecasting algorithm. Results of simulation data proved that proposed approach was suitable for uncongested and congested conditions, however initializing could be critical for some situations. Also, it is suggested that an adaptive time- dependent interval length scheme should be included in the analysis.

Passive Bluetooth MAC data, gathered from vehicle and motorist's mobile devices, was examined to collect OD data by comparing it with video and Automated Number Plate Recognition (ANPR) (Blogg et. al., 2010). To understand applicability of BT data, some critical factors, penetration rate, capture rate and MAC noise, were investigated by a case study at 13 km segment motorway in Brisbane, Australia. The travel time was simply estimated by calculating the time difference between the detection of the MAC address at two different readers, which showed promising results and a cost-effective alternative to traditional methods. The case study reflected a 46 % MAC-to-Volume ratio with lower vehicle occupancy, which was an indicator of capture strength. When compared the ANPR and MAC data, percentage of daily OD pairs reflected similar dispersion, and found available. Despite limitations, MAC data can be used for OD estimation as a cost-effective way. In another study, an analytical approach was introduced for O-D estimation for a relatively urban and compact study corridor in Jacksonville, Florida (Carpenter et. al, 2012). Researchers located 14 BT devices at the study area to capture all entry and exit points, 7 of which were located at main arterial to investigate travel time and volume. In the study BT device alignment, data cleaning and trip itinerary generation methods were highlighted. As a cleanup procedure, the MAC_ID's which captured only by one BT device was removed from data set, and for duplicate reads by a single BT device was decreased to one reading which was the latest reading on each BT device. Also, the maximum time difference to stop a trip was defined as 30 minutes for study area. They concluded that each project needs special characterization and BT sensor locating should be optimal for OD evaluation.

To forecast journey times and estimate time-dependent O-D matrices, Chitturi et. al. (2014) tested the performance of Bluetooth-based data using Kalman filtering techniques. A method for using Bluetooth data from a 15-mile corridor in Jacksonville, Florida was proposed to create route-specific O-D tables. The results confirmed the use of Bluetooth technology and aerial photography to conduct O-D studies for collecting traffic data, but they also highlighted the wide variation in Bluetooth detection rates and suggested further research to examine the properties of Bluetooth data and how site characteristics, traffic characteristics, and detector placement affect the sampling rate of individual O-D pairs.

The use of Bluetooth technology was investigated to create dynamic origin-destination matrices and collect data on travel patterns presenting a promising approach for using Bluetooth data to improve transportation planning (Michau et. al. 2014). The difficulties associated with using Bluetooth data were highlighted in the study including problems with data filtering and correction methods and the potential for Bluetooth sensors to miss some nearby Bluetooth-enabled vehicles.

With a slight change in focus, in a transit network, time-sliced OD trip matrices were estimated with counts and travel time data for 3 public transportation line in Spain (Montero et. al., 2015). Travel time information was evaluated by equipped transit-stops from passenger Bluetooth device. The number of transit trips between OD pairs/stops analyzed in a real-time linear Kalman filtering approach. The three key design factors, penetration rate, number and location of detectors and quality of historic time-sliced OD matrix were highlighted which were critical features on the proposed model. The method was first applied in a toy network and simulated data with evaluation of the accuracy on average, however further effort should be investigated on a-medium-sized network, or a traffic network for performance evaluation of the method.

Dunlap et al. (2016) estimated passenger origin and destination information for transit lines using Bluetooth sensor technologies. Bluetooth sensors installed on buses were used to collect data, which was then processed to remove unnecessary detections using filters. The study discovered that the Bluetooth data was trustworthy for determining the origins and destinations of passengers, however a longer data gathering period is advised to get significant results.

2.2.3 Incident Detection

The estimated travel times that are updated on variable message signs using Bluetooth-based systems were used as inputs for incident detection algorithms (Yu et. al., 2015). The usage of Bluetooth technology has simplified the collection and analysis of travel time data, which can increase the precision and effectiveness of incident detection and traffic management in transportation systems. DriveBlue, a real-time system that uses Bluetooth adapters to predict traffic problems and notify authorities, was used by Salem et al. (2015). DriveBlue keeps track of all the moving vehicles on the road, classifies them, examines their behavior, and alerts users to any unusual changes. In tests, DriveBlue demonstrated a detection accuracy of about 80% for moving vehicles.

Bluetooth data was used for automatic incident detection in Northern Bavaria, Germany, with a focus of segment travel time investigation with Bluetooth data (Margreiter, 2016). Bluetooth data was validated and proved the feasibility in usage. Similarly, Karatsoli et. al. (2017) studied on a 15 km corridor for Automatic Incident Detection in Munich. Due to limited number of incidents during study period, the study investigated Bluetooth data collection quality in VISSIM models, where erformance measures were compared for each incident type and location and suggested that there is a need for further evaluation. To improve traffic condition determination and dynamic net control in incident scenarios, Mercader and Haddad (2020) introduced a novel method for calculating travel times on freeways utilizing Bluetooth technology. Low installation and maintenance costs, a low failure rate, and sectional data rather than local data are some benefits of using Bluetooth sensors. The operation principle of this sensing technology is also extremely straightforward because it is based on the detection and re-identification of visible Bluetooth devices on board cars. Bluetooth technology has also developed into mature sensing technology in ITS.

To evaluate arterial signal coordination, signal offsets were used as signal- timing parameter which can be analyzed by two methods, Purdue coordination diagram (PCD) and vehicle reidentification by MAC matching by Bluetooth (Day et. al., 2010). These two methods were investigated by a case study at an arterial corridor in Noblesville, Indiana. From a totally different point of view, Bluetooth data applicability was investigated for passenger waiting times of airport security, security checkpoints, etc. (Bullock et. al., 2010; Remias et. al., 2013). Bluetooth technology has been used to record transit time through security checkpoints and could give managers information for the most efficient distribution of limited screening resources at both the local airport level and the national level by capturing the MAC addresses of discoverable Bluetooth devices carried by passengers. (Bullock et. al., 2010). The use of Bluetooth-enabled devices to gather probing data and sample the amount of time required for passengers to move from the nonsterile to the sterile side of an airport facility was discussed by Remias et al. (2013). To assess and enhance airport operations, data on passenger travel times through airport security checkpoints can be collected using Bluetooth technology. Using Bluetooth data, Crawford et al. (2018) established road user types based on repeated trip behavior. Insights on traveler demands, flexibility, and network knowledge are provided by the resulting road user classes and subclasses, which can be utilized to inform policy development, economic or behavioral models, and other forms of data gathering.

Using field observations of Bluetooth probe vehicles, Hainen et al. (2011) showed that Bluetooth data was successfully used to capture route choice for detours and assess travel times on alternate routes as a meaningful and low-cost data source. Garrido-Valenzuela et.al. (2020) proposed a methodology to infer the most likely route used by a vehicle between two successive detections, despite the lack of perfect information due to missed Bluetooth detections. Cipriani et al. (2021) investigated the potential of new monitoring systems to enhance origin-destination traffic demand estimation and forecast, and proposed a novel method to estimate traffic demand by utilizing the Bluetooth data that is currently available, which provides additional information about paths and incidence matrices. Using Bluetooth technology, traffic spatial data may be collected, allowing for extended coverage of the road network as opposed to being restricted to stationary locations. The necessity for effective MAC address filtering and the constrained coverage area of Bluetooth detectors are the two main drawbacks to using Bluetooth data for traffic demand estimation.

As part of the deployment of a check-in/be-out system for mobile ticketing in urban passenger transport, Ferreira et al. (2020) evaluated the possibility of Bluetooth Low

Energy (BLE) beacons for tracking passengers' trips from the beginning to the end. According to the results, BLE technology can be used for mobile ticketing in urban passenger transportation; however, physical obstacles found in transportation networks, such as automobiles, steal infrastructure, and people, and potential interference from other Bluetooth-enabled devices can impair the efficiency of BLE beacons. The cost of putting Bluetooth beacons was reported as quite low, making it a more cost-effective option for on-street parking management. Bluetooth beacons can also identify the presence of automobiles in parking places as well as their identities (Chien et al., 2020). Yuan et al. (2018) used data from numerous sources, including Bluetooth, weather, and adaptive signal control datasets, to study the links between crash occurrence on urban arterials and real-time traffic, signal timing, and weather features. The Bluetooth total sample rate employed in this studywas 6.05%, which is greater than the prior studies.

Bluetooth data was investigated for pedestrian travel, but because of low penetration rates of pedestrian activity this method became less feasible when compared to vehicles (Malinovskiy et. al., 2012). Similarly, Yoshimura et. al. (2017) focused on pedestrian behavioral characteristics at shopping environment in the historical center of Barcelona, Spain for one month period with respect to before-during-after discount sales. This study had lots of shortcomings, one of which was low penetration rates of pedestrian activity, as expected. Like pedestrian activity, Bluetooth data was investigated for bicycle activity (Abedi et. al., 2015; Ryeng et. al.,2016). Ryeng et. al. (2016) focused on efficiency of Bluetooth and Wi-Fi sensors for bicycle speeds evaluation, and result reflected besides both methods reflect accurate results, Bluetooth method were more accurate. More specifically, Abedi et. al. (2015) focused on the effects of antenna characteristics on Bluetooth data in terms of travel time estimation for both pedestrian and bicycle users. Results reflected that, optimal set up of small and big antennas can increase the accuracy of travel time estimation.

Pu et al. 2021 proposed a method for monitoring real-time transit ridership flow and origin-destination information by passively sensing mobile devices of passengers. Passengers' mobile devices can be passively sensed by the suggested system for monitoring transit ridership flow using Wi-Fi and Bluetooth mobile devices, making it more efficient and cost-effective.

2.3 Quality of Bluetooth Estimation Methods

The investigation of quality and reliability of Bluetooth data, penetration rate, and validation/ verification methods was also studied in literature, also. A study focused on the accuracy of travel time estimation using Bluetooth technology was investigated, by four different estimators named Min-BT, Max-BT, Med-BT and Avg-BT as an outlier detection logic (Araghi et. al., 2012). Results reflected that Min-BT and Med-BT logic reflect more accurate travel time estimations as compared to the other two. As a follow up study, researchers focus on reliability of Bluetooth technology for travel time estimation, which is defined as the percentage of devices captured per trio during the experiment (Araghi et. al., 2015). This study revealed that devices have short-range antennas detect the devices closer to receiver, so provide more accurate travel time estimation, however when the detection zone became smaller, the penetration rates became smaller. Moghaddam and Hellinga (2013) concentrated on measurement error in arterial travel times with Bluetooth detectors and examined the magnitude of errors in detection time and travel time measurement. This travel time measurements were modeled with multiple regression and standard deviation analysis and reflected reliable outcomes.

In Bluetooth data samples, outliers are extreme and easy to recognize, but some outliers may be subtle and need closer inspection, as discussed in detail. Defining an outlier threshold may require adjusting for particular road segments, and smaller thresholds may exclude valid data. Therefore, the cutoff for finding outliers may change based on the particular data set and the study's context. Table 2.4 summarized some of the applied outlier elimination techniques in the literature.

Table 2.4 The studies used filtering during Bluetooth data analysis.

Study	Topic	Filtering Method
Barceló et. al., 2010a	Travel Time & Dynamic OD Estimation (Freeways)	"Mahala Nobis distance" technique; separation between each data point and the data sets means while accounting for covariance, outliers spotted using predetermined threshold distance
Haghani et. al., 2010	Travel time & space mean speed estimation	“four-step offline filtering” approach.
Quayle et. al., 2010	Arterial Performance Measures	“Moving standard deviation”; only data points above the mean are filtered
Van Boxel et. al., 2011	Real-time outlier detection (Interstate highways & urban arterials)	“Using standardized residuals”; outside of a specified range on the graphs by using before-and-after plots to visually identify outliers’ effectiveness is assessed Shapiro-Wilk statistic.
Tsubota et. al., 2011	Arterial Traffic Congestion	“The unrealistic travel time filter “and “the through movement filter”; unrealistic trip time filter removes records more than 30 minutes to travel
Martchouk et. al., 2011	Freeway Travel Time Variability	“Median filter “(replaces each data point with the median of its neighboring points)
Araghi et. al., 2012	Accuracy of Travel Time Estimation on Motorway	“Min-BT, Max-BT, Med-BT and Avg-BT” as an outlier detection logic
Aliari and Haghani, 2012	Travel time on freeway and arterial segments	"point-by-point method." the reported speed falls within the 95% confidence band of the observed speeds
Porter et. al., 2013	Travel time	“travel time threshold” “30 seconds” less than the smallest reasonable travel time and greater than a vehicle traveling at the speed limit; “2 minute” maximum time difference between the average group times
Namaki Araghi et. al., 2016	Travel Time Estimation (arterial)	“mode-specific algorithm and cluster analysis”
Vinagre Díaz et. al., 2016	Travel Time Estimation (Freeways)	“Device Identification and Authentication Code (DIAC)”; filter out non-vehicle devices.
Civcik and Kocak, 2020	Travel Time Prediction	Travel time threshold: 150 second
Liu et. al., 2020	Travel time (arterial)	“Kalman filtering”
Ghavidel et. al., 2021	Highway Travel-Time Variability Analysis	“Maximum Likelihood Estimation” travel times that were 20% greater or less than the expected travel time excluded from analysis-
Carrese et. al. (2021)	Travel Time Forecast (urban arterial)	the maximum and minimum travel time, “filtering around the median or the moda, and the boxplot method”
Altintasi et. al., 2022	Travel Time Evaluation (urban arterial)	Interquartile range (IQR) method
Jedwanna and Boonsiripant, 2022	Travel time estimation (toll road)	Removing duplicate records, removing records with missing data, and removing records with unrealistic travel times (<=60mins)

CHAPTER 3

USAGE OF BLUETOOTH DATA IN TRANSPORTATION AND CONCEPTUAL DEFINITIONS

Bluetooth (BT) data has a complex structure with lots of variabilities and uncertainties, needs serious preliminary analysis and data cleaning operations. Therefore, it is critical to understand the data characteristics and how to use this data in transportation area as travel time and speed in urban corridors. BT does not provide exact point information due to its data feature. Instead, it gives reading information from one or more points within the coverage area. The exact information is only the location points of BTR devices and the shortest corridor distance between them. The cross-BTR (x-BTR) time can be calculated as captured time difference between BTR pairs by using the collected readings with a simple logic. However, since the readings are not taken from a single/ fixed point, it is necessary to integrate the stay time information within the capture zone of each BTR into these calculations. In summary, the most critical component for the estimation of urban arterial travel time is determination of possible moving vehicle travel time between BTR devices. In this section, it will be explained what kind of information is provided by BT data, what kind of outputs can be obtained and what are the limitations should be overcome with these data technology.

3.1 BT Data Structure and Reading Process

BTRs are located on the road sections or intersections as seen in Figure 3.1 in order to be used in the transportation area. The located BTR devices are on the 1st class Bluetooth type, covering a capture zone of 100 m diameter in this study, symbolized by the yellow circle area in the figure. BTR scans any device with an active Bluetooth capability and saves a set of information that includes the capture time (Discovery time), location (BTR_ID), specific MAC_ID, and device type. This group of information is called as "Reading", as seen in any row in the example data set in Figure 3.2. Discovery Time reflects the EPOC capture time, which is a universal time stamp including both date and time information. MAC_ID is the unique information for each device which is further coded as shown in the figure for security concerns. Device Type indicates whether the captured device is a vehicle or not with a default brand march information.

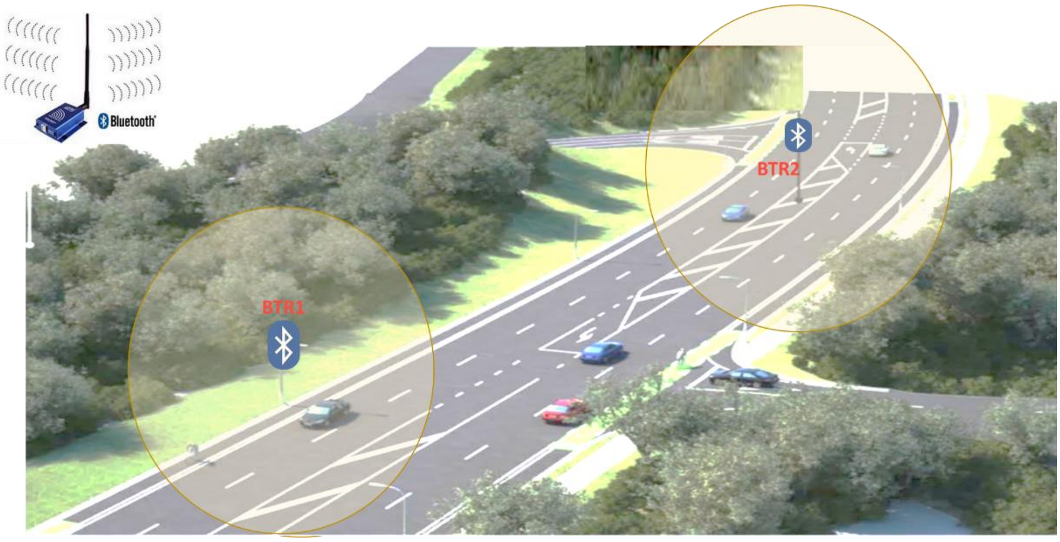
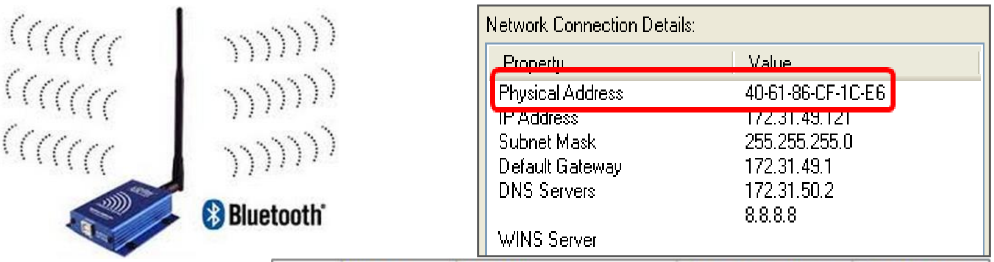


Figure 3.1. Bluetooth Data Collection Procedure.

A MAC_ID can belong to a vehicle or any other type of device, such as a cell phone, a television, an internet receiver, and so on. Besides, every unique MAC_ID can be captured once or multiple times or it can leave without being captured at all. The information obtained from the device type could lead to the classification of this data as a vehicle or not (non-vehicle). However, some non-vehicle devices, such as cell phones or headphones, behave as if they were vehicles because they are the ones traveling inside a vehicle. Moreover, some of these devices can be captured by the same BTR for the entire reading duration or a very long portion of it.



	BTRID	MacID	discoveryTime	deviceType
1	3367	x_04FE31ACFEF5	1562014803	0
2	3367	x_3021DA311DC4	1562014815	240408
3	3368	x_30C3D97E7424	1562014804	340408
4	3370	x_BC307EB3D3F1	1562014802	0
5	3370	x_AC7A4D5E98C0	1562014803	340408

Figure 3.2 Bluetooth Reader and Data Structure of any Device

To summarize; a MAC_ID can appear in the following six different ways:

- It can be a vehicle; read by:
    - Multiple times (in a long/ short period),
    - Only one time,
    - None
- It can be any other device have active BT enabled device having MAC_ID:
    - Travelling inside a vehicle acting as a vehicle,
    - A device carried by a pedestrian/ cyclist
    - Stationary near a BTR (like an audio system inside a store).

3.2 BT Data Analysis Framework

It is very important to examine and classify the captured MAC_ID features in detail in order to distinguish these general collection features of BT data and possible extreme cases. Within the scope of the preliminary data analysis methodology, all day data collected from all active BTRs was examined as reoccurrence time, regardless of which BTR they captured (Figure 3.3). After the calculated reoccurrence times are distinguished according to captured BTR locations, whether captured in a single location (rescan time) or captured by multiple BTR locations (cross-BTR time).

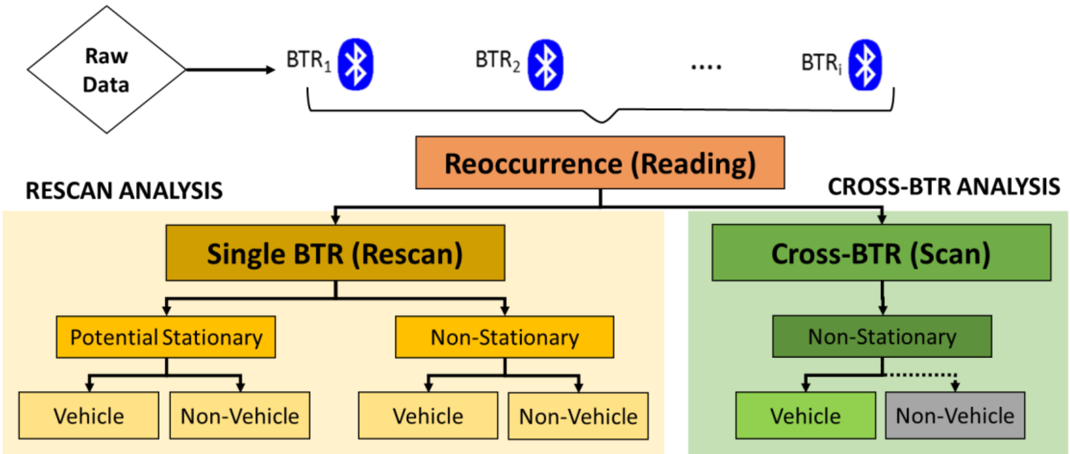


Figure 3.3. Preliminary Data Analysis Framework

At first glance, considering the purpose of using BT data in transportation, it is required to examine the collected unique MAC_IDs by grouping them as "stationary" or "non-stationary", considering that the target group, that needs to be studied, is moving MAC_IDs. MAC_IDs captured by more than one BTR can easily be said to be non-stationary, with one minor exception (captured by only two BTRs and being stationary at the intersection of these two devices capture zone). However, more information is required to determine whether a MAC_ID is stationary when captured by one single BTR. In order to determine the “potential stationary” among the MAC addresses captured by a single BTR there needs to be investigated according to predefined conditions such as staying in the same device capture zone for 2 hours and/or having more than 200. Similarly, MAC addresses that stay in range for a short time and/or have few reads can be defined as potentially non-stationary.

While all non-stationary and potential stationary MAC addresses can have vehicle device type code, they can also be obtained from non-vehicle devices. As a result, MAC_IDs are classified into four groups: i) non-stationary non-vehicle, ii) non-stationary vehicle, iii) stationary non-vehicle, and iv) stationary vehicle. Within these groups, non-stationary non-vehicle and non-stationary vehicle MAC_IDs can be retrieved from the data and used in cross-BTR data analysis without any hesitation. However, determining the rescan time threshold is critical to calculate the stay time within the BTR capture zone. The rescan time distributions and properties calculated for MAC specific, taken from all active BTRs for the entire days, must be examined before selecting any data group. The threshold to be determined should be too short to go anywhere and turn back and long enough not to cause reading disruptions. Because of these two constraints, data must be investigated by analyzing individual device (Single-BTR) and inter device (Cross-BTR) readings separately. The time difference between BTRs is calculated as cross-BTR time without considering the time spent within the capture zone as stay time. However, it is not the right approach to use this value as travel time. Since the location in which BTR devices are placed is the middle of the capture zone; it would be more logical to calculate the possible moving vehicle travel time between BTR devices by including half of both BTR stay time.

All these logical inferences and simple calculations need to be examined in detail by considering transportation dynamics, traffic and Bluetooth data characteristics. Given that data-based features can differ from region to region and even depending on weather conditions, estimating trip times blindly may result in unexpected mistakes in the research results.

3.3 Bluetooth (BT) Data Analysis- Conceptual Definitions with Example Data Set

To better illustrate the data characteristics, limitations and possible outcomes, a conceptual map for an example data set was prepared as shown in Figure 3.4.

According to the figure, there are 4 different MAC_IDs captured by one or two BTRs, represented by blue, yellow, orange and green stars as MAC_ID 1, 2, 3 and 4 respectively. "$t_{m;n}$" represents nth reading of vehicle m as "Reading ID", all of which belongs to any row in data set as reading time (Discovery time). All, readings are tabulated in Table 3.1, which is sorted first by MAC_ID and then by discovery time at the second level. Although the device type information gives information about whether the MAC_ID is a vehicle or not, it is aimed to classify each MAC_ID by analyzing data such as the number of BTRs captured, the number of readings and the total possible stay time as "stationary" and "non-stationary".

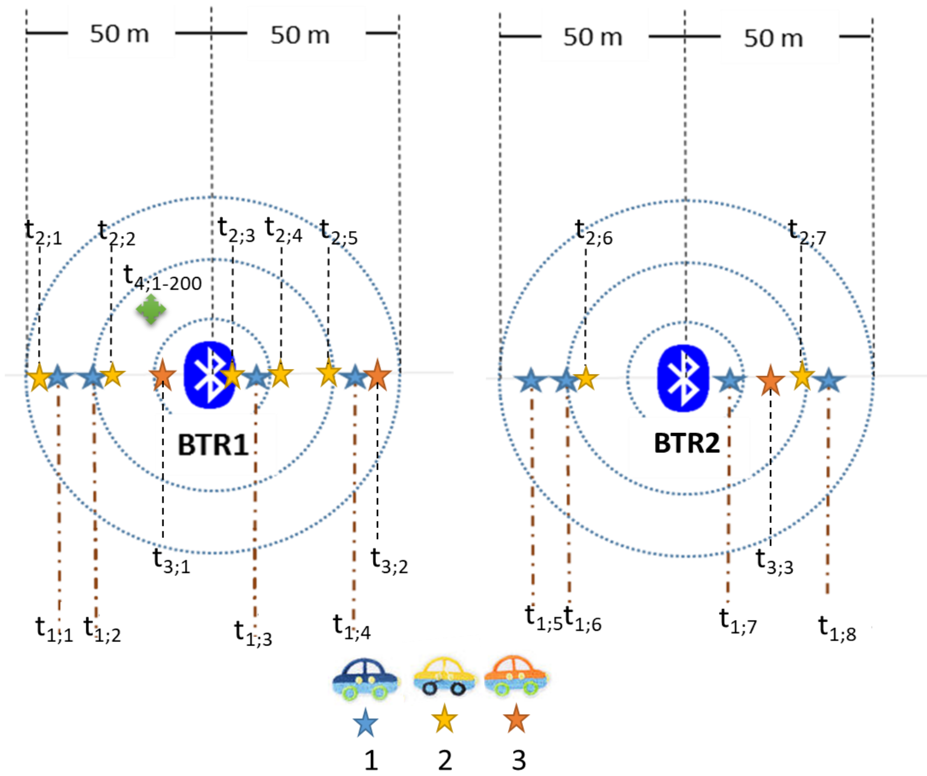


Figure 3.4. A Schematic Representation of BT readings at two consecutive BTRs

Table 3.1 BTR Data for the Sample case in Figure 3.4

Reading ID	BT_ID	MAC_ID	Discovery Time	Device Type	Δt
$t_{1;1}$	BTR1	MAC1	1549918801	240414	NA
$t_{1;2}$	BTR1	MAC1	1549918802	240414	1
$t_{1;3}$	BTR1	MAC1	1549918805	240414	3
$t_{1;4}$	BTR1	MAC1	1549918808	240414	3
$t_{1;5}$	BTR2	MAC1	1549918843	240414	35
$t_{1;6}$	BTR2	MAC1	1549918898	240414	55
$t_{1;7}$	BTR2	MAC1	1549919082	240414	184
$t_{1;8}$	BTR2	MAC1	1549919087	240414	5
$t_{2;1}$	BTR1	MAC2	1549918801	580204	NA
$t_{2;2}$	BTR1	MAC2	1549918803	580204	2
$t_{2;3}$	BTR1	MAC2	1549918804	580204	1
$t_{2;4}$	BTR1	MAC2	1549918806	580204	2
$t_{2;5}$	BTR1	MAC2	1549918806	580204	0
$t_{2;6}$	BTR2	MAC2	1549918809	580204	3
$t_{2;7}$	BTR2	MAC2	1549918813	580204	4
$t_{3;1}$	BTR1	MAC3	1549918803	0	NA
$t_{3;2}$	BTR1	MAC3	1549918809	0	6
$t_{3;3}$	BTR2	MAC3	1549919458	0	649
$t_{4;1}$	BTR1	MAC4	1549918801	340408	NA
$t_{4;2}$	BTR1	MAC4	1549918803	340408	2
$t_{4;.....}$	...	...	...	...	...
$t_{4;200}$	BTR1	MAC4	1549948707	340408	29904

The concept of "Reoccurrence Time" simply stands for the time difference between two captures of the same MAC_ID by any BTR. For example, according to Figure 3.4; reoccurrence time can be found by the discovery time difference between every reading of the same blue car which captured by any BTR (i.e. "$t_{1;2}$- $t_{1;1}$", "$t_{1;3}$- $t_{1;2}$", "$t_{1;4}$- $t_{1;3}$", "$t_{1;5}$- $t_{1;4}$", "$t_{1;6}$- $t_{1;5}$", "$t_{1;7}$- $t_{1;6}$" or "$t_{1;8}$- $t_{1;7}$"), respectively. Reoccurrence can be divided into two groups according to location information; rescan and cross-BTR. "Rescan time" (Δt) is the time difference between two captures of a MAC_ID by the same BTR; which is any time difference of first car within capture zone of BTR 1 (i.e. "$t_{1;2}$- $t_{1;1}$", "$t_{1;3}$- $t_{1;2}$", "$t_{1;4}$- $t_{1;3}$") and capture zone of BTR 2 (i.e. "$t_{1;6}$- $t_{1;5}$", "$t_{1;7}$- $t_{1;6}$" or "$t_{1;8}$- $t_{1;7}$") (Equation 3.1). "Cross-BTR (x-BTR) time" is the time difference between the last reading of the BTR 1 and the first reading of the BTR 2 (i.e., "$t_{1;5}$- $t_{1;4}$") (Equation 3.2).

$\forall\ vehicle\ m;$

$$\begin{bmatrix} t_{m;1} \\ t_{m;2} \\ t_{m;3} \\ \cdots \\ t_{m;p} \end{bmatrix} \Longmapsto \delta t_{m;i}\ and\ \begin{bmatrix} t_{m;(p+1)} \\ t_{m;(p+2)} \\ t_{m;(p+3)} \\ \cdots \\ t_{m;r} \end{bmatrix} \Longmapsto \delta t_{m;j}$$

$$\Delta t = t_{i,m,n} - t_{i,m,(n-1)} \qquad [3.1]$$

$$\Omega^{BT}_{m,i,j} = \mathrm{t_{m;(p+1)}} - \mathrm{t_{m;p}} \qquad [3.2]$$

The green stars in the figure represent a stationary non-vehicle MAC_ID, without any confusion which has been active for 29906 seconds (8.5 hours) and received over 200 readings and not captured by any other BTR in the network. Blue (MAC1), yellow (MAC2) and orange (MAC3) stars most likely represent vehicles or devices that traveled inside a vehicle and were captured by both BTRs. The vehicle represented by the yellow star moved through the coverage area of the two BTR devices in a regular motion, when all the readings are considered, the reoccurrence time values are quite short. MAC2 passed without stopping or making unexpected movements within range of the BTR pair. In addition, while the vehicle represented by the blue star exhibits a regular movement within the scope of the first BTR, a very high value (184 seconds) stands out in the rescan time by the second BTR. Given the general differences between other readings, this value is not expected and should be studied in detail. This can reveal the possibility that the vehicle has left the coverage area and returned or interrupted for a while for a technical reason, such as an object/vehicle interfering with the signal. As a result, it is vital to determine whether these types of values will be used while assessing the vehicle's movement or not. Similarly, the vehicle represented by the orange star was captured only once by the second BTR; twice by the first BTR. Rescan time with BTR1 revealed a small value while cross-BTR time between BTR1 and BTR2 seemed high when compared to other vehicles movement. This result should be investigated but if the time of the day for these 3 vehicles are different, this result can be acceptable considering the possibility that the result that gives a long time coincides with the traffic density. As can be seen, due to the fact that BT data presents many uncertainties and variable results, and when the variability of traffic is added to this, the correct interpretation of the results is of great importance.

3.3.1 Stay Time Calculation

It should be taken into consideration that when considering the movement of the vehicle along any corridor, this movement is not just between BTR devices but also inside the BTR coverage area. In particular, the waiting times at the intersection should be included in the analysis of the data collected with the BTRs placed at the signalized intersections where the traffic lights are located. Stay time ($\delta t_{i;m}$) is calculated as the time difference between the last reading and the first reading directly within the scope of the same BTR device in the most basic approach (Equation 3.3).

$$\delta t_{i;m} = t_{i,m,p} - t_{i,m,1} \quad [3.3a]$$

$$\delta t_{j;m} = t_{j,m,r} - t_{j,m,1} \quad [3.3b]$$

The stay time value is indicated in the Figure 3.5 by the movement of the MAC1 vehicle, which is indicated by a blue star on the example conceptual map. The vehicle was read 4 times by both BTR devices, but the stay time on the 1st device was calculated as 7 seconds and the second device stay time was calculated as 244 seconds (Equations 3.4 and 3.5). Considering that the vehicle moves between two BTRs in the same time interval and it is thought that it should pass with a similar behavior, it should be examined whether these two different stay time values are meaningful.

$$\delta t_{1;1} = \text{stay time}_{1\_BTR1} = t_{1;4} - t_{1;1} = 7 \text{ sec} \quad [3.4]$$

$$\delta t_{1;2} = \text{stay time}_{1\_BTR2} = t_{1;8} - t_{1;5} = 244 \text{ sec} \quad [3.5]$$

In order to calculate the stay time correctly, it is very important to determine the data collection and rescan frequency characteristics of the BTR device exactly. At this point, the inquiry interval value, which is a changeable device feature, should be examined and revised if necessary. Although there are many studies on this subject as discussed in the literature, there is no detailed study on the effect of inquiry time on the data to be used in the field of transportation.

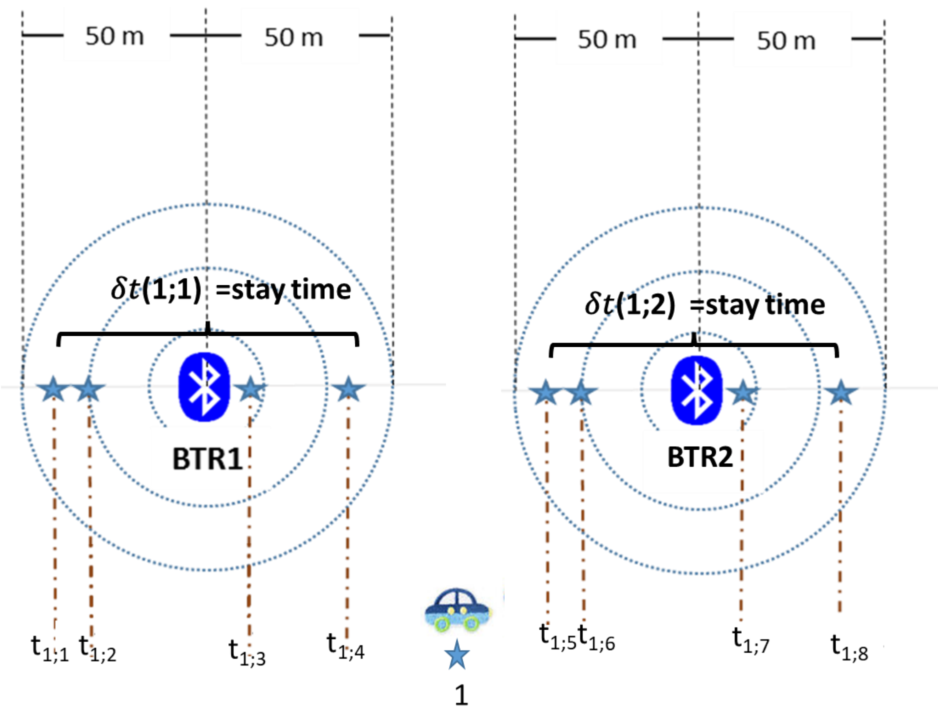


Figure 3.5. A Schematic Representation of BT readings for Stay Time Calculation

3.3.1 Travel Time Calculation

"Cross-BTR (travel time)" will be simply determined as the time difference of any MAC_ID between two different BTRs (Equation 3.2). However, this value only refers to the time elapsed between 2 BTR devices. There is not any commonly accepted method for whether the stay time in the Bluetooth capture zone will be included in this calculated travel time or how it will be included.

In the literature vehicle travel time can be calculated by five different methods: i) first to first (F2F), ii) last to last (L2L), iii) mid to mid (M2M), iv) first to last (F2L), v) last to first (L2F). These concepts are calculated as follows depend on the capture order of unique MAC_IDs:

i) last to first travel time is the time difference between the last reading of BTR i and the first reading of BTR j;

$$\Omega_{m,i,j}^{BT,L2F} = t_{m;(p+1)} - t_{m;p} \quad [3.6]$$

ii) first to first travel time is the time difference between the first reading of BTR i and the first reading of BTR j;

$$\Omega_{m,i,j}^{BT,F2F} = t_{m;(p+1)} - t_{m;1} \quad [3.7]$$

iii) last to last travel time is the time difference between the last reading of BTR i and the last reading of BTR j;

$$\Omega_{m,i,j}^{BT,L2L} = t_{m;r} - t_{m;p} \quad [3.8]$$

iv) first to last travel time is the time difference between the first reading of BTR i and the last reading of BTR j;

$$\Omega_{m,i,j}^{BT,F2L} = t_{m;r} - t_{m;1} \quad [3.9]$$

v) mid to mid travel time is the summation of "the time difference between the last reading of BTR i and the first reading of BTR j" and "half of the stay times for both BTRs";

$$\Omega_{m,i,j}^{BT,M2M} = t_{m;(p+1)} - t_{m;p} + \frac{(\delta t_{m;i}) + (\delta t_{m;j})}{2} \quad [3.10]$$

The travel time calculation methods are indicated in the Figure 3.6 by the movement of the MAC1 vehicle, which is indicated by a blue star on the example conceptual map, to better illustration. Travel time values calculated with 5 different methods for the MAC1 device are displayed in Equation 3.11 to 3.15 ranging between 35 seconds through 286 seconds. Detailed studies need to be done on which of these five methods to use, which varies considerably depending on how the stay time value is included with the same data set.

$$\Omega_{1,1,2}^{BT,L2F} = t_{1;5} - t_{1;4} = 35 \text{ sec} \quad [3.11]$$

$$\Omega_{1,1,2}^{BT,F2F} = t_{1;5} - t_{1;1} = 42\ sec \quad [3.12]$$

$$\Omega_{1,1,2}^{BT,L2L} = t_{1;8} - t_{1;4} = 279sec \quad [3.13]$$

$$\Omega_{1,1,2}^{BT,F2L} = t_{1;8} - t_{1;1} = 286\ second \quad [3.14]$$

$$\Omega_{1,1,2}^{BT,M2M} = t_{1;5} - t_{1;4} \frac{(\delta t_{1;1})+(\delta t_{1;2})}{2} = 160.5\ sec \quad [3.15]$$

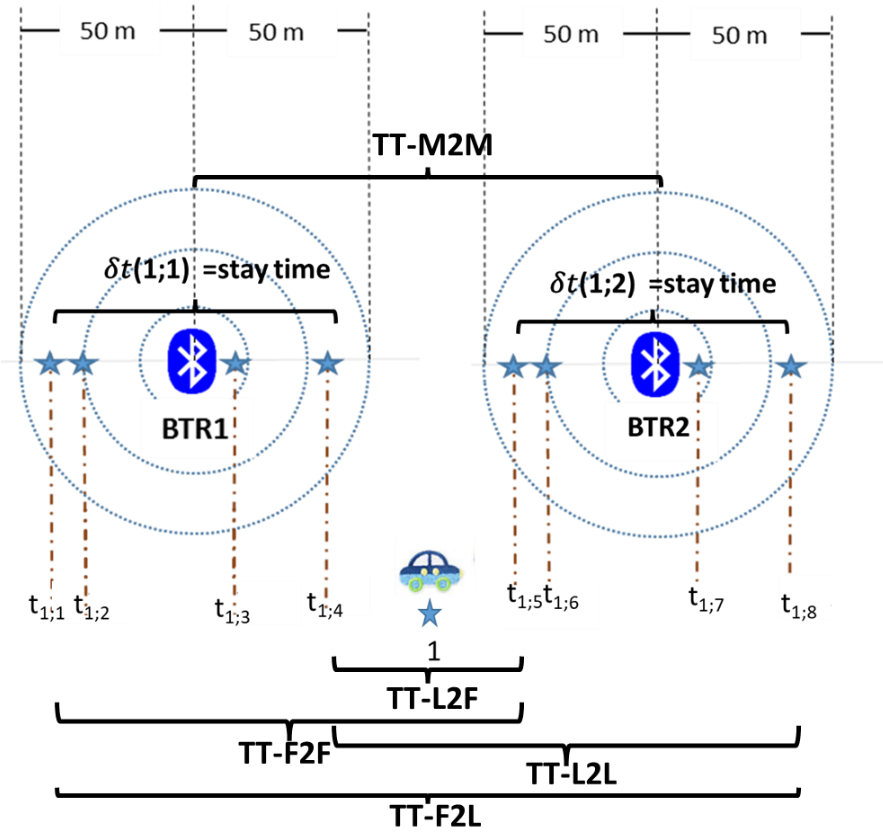


Figure 3.6. A Schematic Representation of BT Data for Travel Time Calculation

In addition to the travel times calculated with different methods for the same device, the travel time values calculated with the same method for more than one MAC_ID at the same time intervals should also be a part of the travel time estimation process based on BTR pair (link). For this purpose, the travel times calculated by the M2M method for the MAC1, MAC 2 and MAC3 vehicles in Figure 3.4 are shown in

Equation 3.16 to 3.18. Considering that the 3 calculated travel time values differ significantly, it is necessary to work on the reason for this difference and how to eliminate it. It should be taken into account that this difference may be due to the vehicle behavior, as well as the collection characteristics of BT data and simple technological errors.

$$\Omega_{1,1,2}^{BT,M2M} = t_{1;5} - t_{1;4} \frac{(\delta t_{1;1})+(\delta t_{1;2})}{2} = 160.5\, sec \quad [3.16]$$

$$\Omega_{2,1,2}^{BT,M2M} = t_{2;6} - t_{2;5} \frac{(\delta t_{2;1})+(\delta t_{2;2})}{2} = 7.5\, sec \quad [3.17]$$

$$\Omega_{3,1,2}^{BT,M2M} = t_{3;3} - t_{3;2} \frac{(\delta t_{3;1})+(\delta t_{3;2})}{2} = 652\, sec \quad [3.18]$$

3.3.2 Movement Vector

To fully analyze the behavior of each MAC_ID in the network on the analysis day, the creation of a "movement vector", in which data such as stay time and travel time are summarized, will form an important basis for corridor travel time estimation. The process starts by sorting the data first by MAC_ID and then by time as in the example data set (Table 3.1). Starting with a different column being formed for each MAC_ID, an algorithm is established when producing the movement vector. For each column, the MAC_ID information is followed by the Device type information so that it can be used when necessary. Following this two information, below lines are added respectively for each captured BTR:

- BTR_ID,
- BTR reading time,
- The number of rescans by the BTR,

- The time elapsed between the first reading and the last reading of the BTR (stay time),
- The time elapsed between the last reading time by the previous BTR and the first reading time of the current BTR device (ttime)

Flexibility is provided in the algorithm; i) for stay time calculation according to the rescan threshold value to be determined and ii) the use of five travel time calculation methods in the literature. To better understand the algorithm working principle and calculations, the example corridor where the 3 BTR device located is in Figure 3.7. The movement vector for this created sample conceptual map is seen in Table 3.2 which was prepared according to above systematic. In this example, however, no thresholds for the rescan time were considered, as there was no detailed data pre-analysis as to whether it was necessary. For the travel time calculation method, since there will be a change in the stay time calculation according to the threshold determination, calculations were made according to the L2F method.

In the movement vector, the first 2 lines show the MAC properties, and each subsequent group of 5 shows the data calculated within the scope of each BTR captured. For easier understanding of the sample data, a color code is used for each BTR in the figure; and this color is also reflected in the corresponding rows of the table. For example, since the BTR3 device coded with green cannot captured by the MAC3 device, there is no information in the green lines of the MAC3 column.

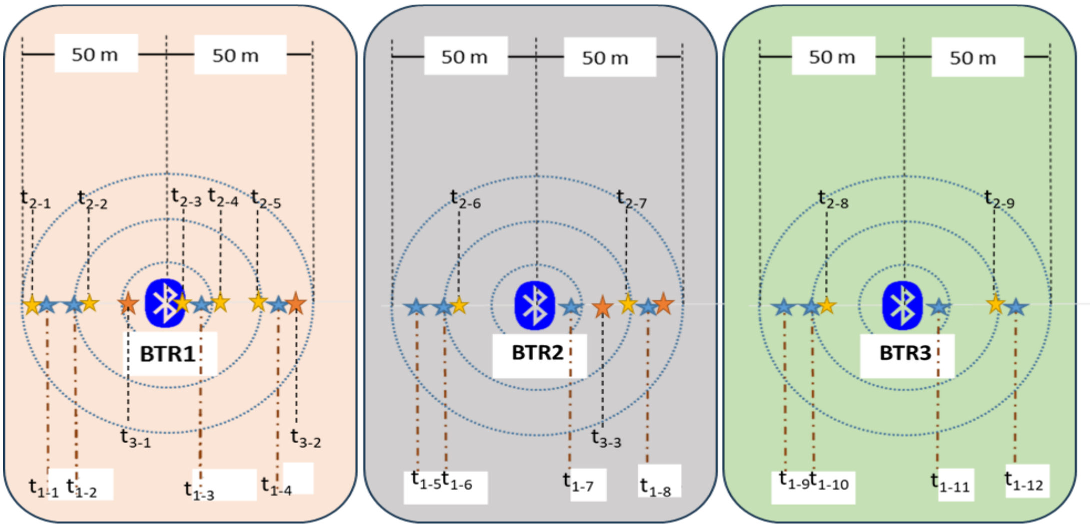


Figure 3.7. A Schematic Representation of BT data over Multiple BTRs for Movement Vector Estimation

Table 3.2. Movement Vector for the Sample Data in Figure 3.7

MAC_ID	**MAC1**	**MAC2**	**MAC3**
Device Type	**1**	**0**	**0**
BTR_ID	BTR1	BTR1	BTR1
Epoch Time	1549918801	1549918801	1549918803
nRescan	3	4	1
Stay Time	7	5	6
Ttime (L2F)	NA	NA	NA
BTR_ID	BTR2	BTR2	BTR2
Time	1549918843	1549918809	1549919458
nRescan	3	1	NA
Stay Time	**244**	4	NA
Ttime (L2F)	35	3	**649**
BTR_ID	BTR3	BTR3	
Time	1549919108	1549918978	
nRescan	3	1	
Stay Time	24	20	
Ttime (L2F)	21	**165**	

3.4 Urban Arterial Travel Time (TD_UrbArt_TT) Estimation

Time dependent urban arterial travel time information provides notable information to road users for travel planning. However, since the collected data are quite uncertain and include many extreme situations, serious preliminary analyzes should be made in the process of these analyzes.

Readings of each captured unique MAC_ID of each vehicle m, needs to be used for possible moving vehicle travel time determinations. Calculation methods are discussed in detail in the following sections, all of which need further investigation on extreme value determination and outlier elimination. The clean-up process will be detailed in the following sections considering the determination of “possible moving vehicle travel time” for BTR pairs. The possible moving vehicle travel times for captured MAC_IDs are averaged after the cleaning process to determine the time dependent urban arterial travel times for each BTR pair (as link travel time) according to time interval (Equation 3.19).

$$\overline{\Omega}_{i,j,\tau}^{BT} = Avg(\Omega_{i,j,m,\tau}^{BT}) \qquad [3.19]$$

3.5 Urban Arterial Speed (TD_UrbArt_Speed) Estimation

The time dependent urban arterial speed estimation is valuable information for decision makers in the determination of Level of Service (LOS). The time dependent urban arterial travel times calculated between BTR pairs needs to be converted into speed data using the exact distance between pairs. The distances between BTR pairs are calculated in R using shortest path algorithm and exact BTR coordination ($L_{i,j}^{BT}$). Average speed is calculated based on time intervals for each BTR using Equation [3.20].

$$\bar{u}^{BT}_{i,j,\tau} = \frac{L^{BT}_{i,j}}{\bar{\Omega}^{BT}_{i,j,\tau}} \quad [3.20]$$

3.6 Speed Validation with FCD Data

Average moving vehicle travel speed is calculated using moving travel time (remained ones after outlier detection) and length between BTR pairs based on time intervals for each BTR pair. Speed value is important in determination of LOS. It is important to examine the accuracy of the BT data by calculating the error with the average speed values calculated for each BTR pair with the FCD data calculated at the same intervals from the same perspective. Performance measurement will be determined with below equations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\bar{u}^{FCD}_{i,j,\tau} - \bar{u}^{BT}_{i,j,\tau}| \quad [3.21]$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(|\bar{u}^{FCD}_{i,j,\tau} - \bar{u}^{BT}_{i,j,\tau}|\right)^2} \quad [3.22]$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\bar{u}^{FCD}_{i,j,\tau} - \bar{u}^{BT}_{i,j,\tau}}{\bar{u}^{FCD}_{i,j,\tau}}\right| * 100 \quad [3.23]$$

3.7 Probabilistic Cross BTR- OD Travel Time Estimation

In addition, cross-BTR time will be used to evaluate the absence of any device readings in consecutive BTR readings in a probabilistic manner. For this, the distribution of all travel times seen in each BTR pair is examined. Statistical values such as mean and standard deviation values are calculated, and the distribution of the values are analyzed. In statistics, summation of multiple variables will fit in a Normal Distribution, so for total travel time from the first BTR to the last one (path travel time) will be checked with summation of each BTR pair travel time (link travel time).

It can be assumed that your vehicle passed through the absent BTR region according to this total expected travel time under defined the confidence interval. Example probabilistic cross-BTR investigation is depicted in Figure 3.8.

$t_{\text{BTR01-BTR06}} = t_1 + t_2 + t_3 + t_4 + t_5$

$t_1 \sim N(\overline{t_1}, s_1)$, $t_2 \sim N(\overline{t_2}, s_2)$, $t_3 \sim N(\overline{t_3}, s_3)$, $t_4 \sim N(\overline{t_4}, s_4)$, $t_5 \sim N(\overline{t_5}, s_5)$

$\overline{t_{\text{BTR01-BTR06}}} = \overline{t_1} + \overline{t_2} + \overline{t_3} + \overline{t_4} + \overline{t_5}$

$s^2_{\text{BTR01-BTR06}} = s_1^2 + s_2^2 + s_3^2 + s_4^2 + s_5^2$

$$t_{\text{BTR01-BTR06}} \sim N(\overline{t_{\text{BTR01-BTR06}}}, s_{\text{BTR01-BTR06}}),$$

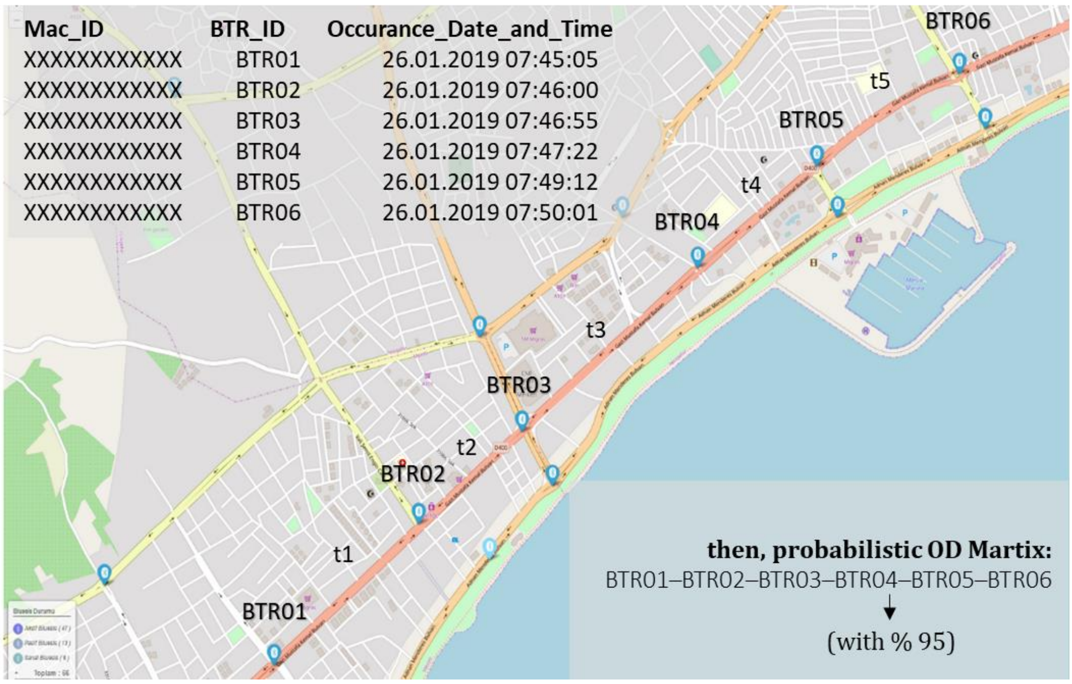


Figure 3.8. Probabilistic Cross-BTR OD Travel Time Estimation

$t_{\propto,lower} \leq \mathrm{t}_{\mathrm{BTR01-BTR06}} \leq t_{\propto,upper}$

$\overline{t_{\mathrm{BTR01-BTR06}}} - z_{\propto} . s_{\mathrm{BTR01-BTR06}} \leq \mathrm{t}_{\mathrm{BTR01-BTR06}} \leq \overline{t_{\mathrm{BTR01-BTR06}}} + z_{\propto} . s_{\mathrm{BTR01-BTR06}}$

$z_{\propto} = 1.96$ → 95% Confidence Interval

$$\overline{t_{\mathrm{BTR01-BTR06}}} - 1.96 * s_{\mathrm{BTR01-BTR06}} \leq \mathrm{t}_{\mathrm{BTR01-BTR063}} \leq \overline{t_{\mathrm{BTR01-BTR06}}} + 1.96 * s_{\mathrm{BBTR01-BTR06}}$$

3.8 A General BT-based Urban Arterial Travel Time Analysis Framework

The general BT-based urban arterial travel time analysis framework is presented in Figure 3.9. Considering that data collection and BTR device specifications made as preliminary analysis; the process continues on working with the raw data specific to each MAC_ID and the captured BTR. The recsan threshold value obtained from the preliminary analysis of the raw data (which will later discussed in the following sections with case study) directly contributes to the stay time calculation process and indirectly to the travel time. While the calculated stay time and travel time values are compiled with the movement vector in order to be used effectively in following studies, the filtering process of the travel time data is started to work. Filtering needs to be applied in two stages at selected time intervals; first, slow movements are eliminated and then basic statistical outlier determination is made using IQR. In the study this new point of view to filter the data in two stage is a newly proposed method for outlier detection despite using a basic statistical method in the second stage. Remaining M2M link travel times are expected to be from “possible moving vehicle” which will further be validated by FCD data. After this stage of the methodology, link travel time is estimated for each BTR pair with a time-dependent or user-flexible perspective. To estimate link travel time, it is important to investigate the distribution of “possible moving vehicle” travel times within selected data. This evaluation includes determining the statistical distribution of the data and determining the expected travel time range for the link according to the characteristics of this distribution.

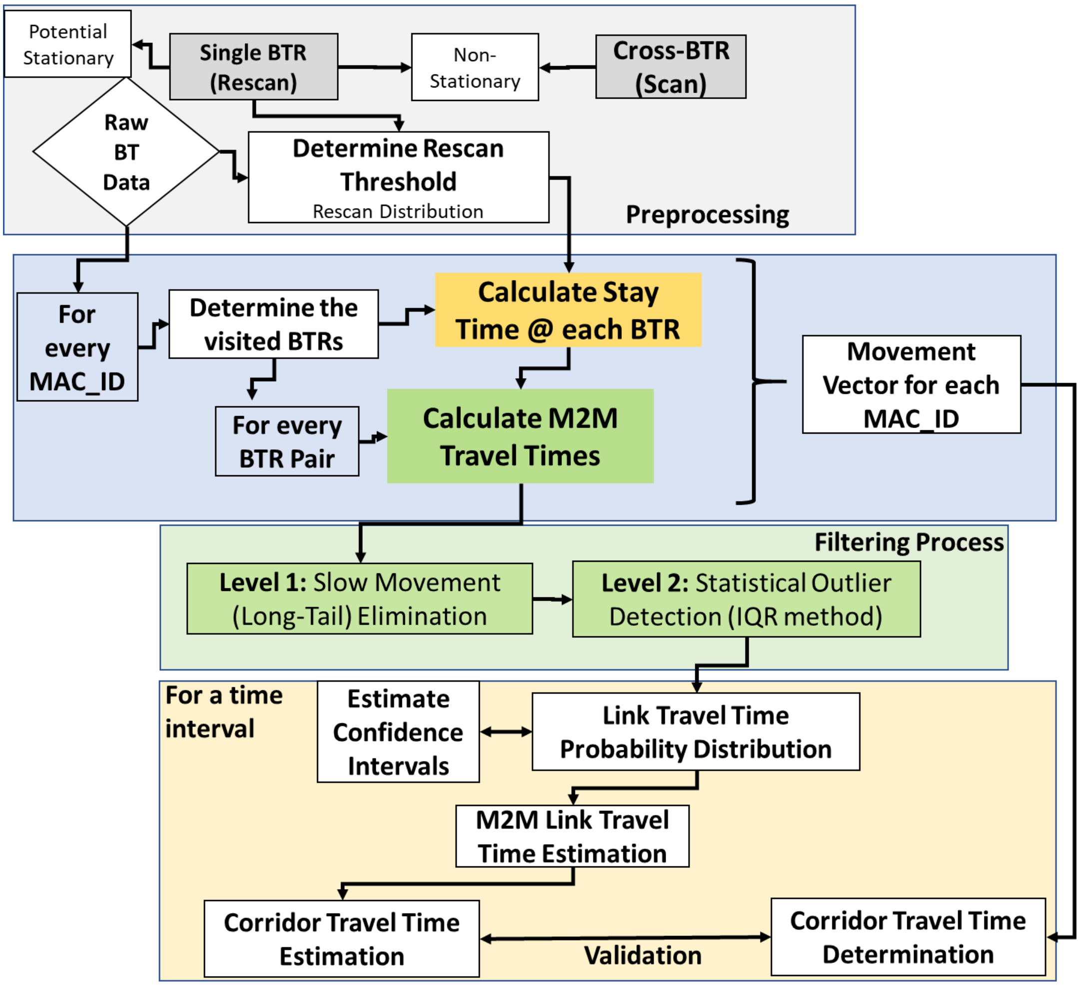


Figure 3.9. A General BT Data Analysis Framework

CHAPTER 4

URBAN ARTERIAL TRAVEL CHARACTERISTICS OF MERSIN

4.1 Study Corridor

Mersin is a Mediterranean city located on the southern coast of Turkey (Figure 4.1). One of the largest seaports in Turkey is located in Mersin, which makes the city a significant commercial center. Public transportation in Mersin is provided by municipal buses, private public buses, and minibuses. The rail system has not yet been brought to the city. Therefore, only roads are allowed for transportation within the city; as no other forms of transportation, such as rail or water, are available.

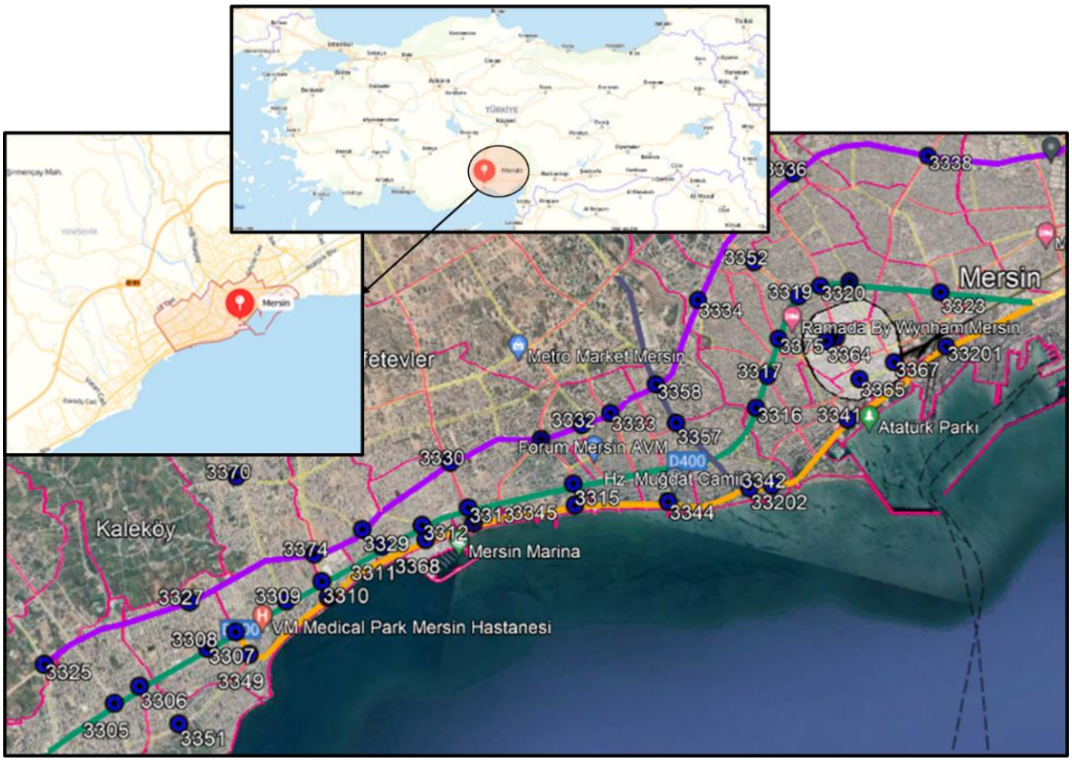


Figure 4.1. Existing 48 Bluetooth locations in Mersin

In Mersin, 48 Bluetooth readers (BTR) were strategically placed throughout the city, primarily placed on three main arterials that run parallel to the sea and carry most of the city's traffic, to collect real-time data. There is no traffic congestion in the city, except for unexpected situations such as wrong parking on the roadside or an accident. Instead of network-based evaluations during the process involving the development of data processing method, a corridor is chosen to understand the usage of BT data in urban studies. One of the 3 parallel main arterials called Gazi Mustafa Kemal Boulevard (D-400) is selected as study corridor which has the highest capture rates and traffic flow considering the characteristics of the corridor and the variability of the distances between the devices (Figure 4.2). This is a two-lane corridor with a speed limit of 70 km/h that connects the city in an east-west direction. There are six consecutive signalized junctions along its 3850 meters in length. On the corridor there are numerous public transportation lines including dolmuş and public transit. The Entry-Exit Matrix of Study Corridor for September 6, 7, 13 and 14 are summarized in Table 4.1. Although there is a serious density between consecutive BTR pairs in matrix, it is seen that the situation of being detected by the next BTR without being caught in between is at a level that needs further evaluations.

Figure 4.2. Mersin Study Corridor

Table 4.1**.** Cross BTR Corridor Entry-Exit Matrix of Study Corridor for September 6, 7, 13 and 14 2022.

	September 6, 2022 (Tuesday)							September 7, 2022 (Wednesday)						
	BTR07	BTR08	BTR09	BTR10	BTR11	BTR12	BTR13	BTR07	BTR08	BTR09	BTR10	BTR11	BTR12	BTR13
BTR07	0	640	188	115	50	25	7	0	614	172	106	56	26	8
BTR08	630	0	532	94	30	10	2	611	0	487	97	22	12	2
BTR09	62	647	0	638	81	29	6	63	630	0	604	94	21	7
BTR10	29	170	657	0	629	88	15	33	175	624	0	617	91	19
BTR11	6	29	57	719	0	778	58	11	25	44	664	0	792	63
BTR12	6	14	23	138	783	0	468	8	18	20	127	788	0	424
BTR13	8	9	10	57	136	738	0	4	7	15	52	116	703	0
	September 13, 2022 (Tuesday)							September 14, 2022 (Wednesday)						
	BTR07	BTR08	BTR09	BTR10	BTR11	BTR12	BTR13	BTR07	BTR08	BTR09	BTR10	BTR11	BTR12	BTR13
BTR07	0	677	183	118	43	31	3	0	644	179	110	59	16	8
BTR08	662	0	495	71	22	11	2	544	0	457	77	19	3	5
BTR09	62	665	0	650	78	31	9	51	643	0	665	73	23	11
BTR10	40	164	643	0	629	79	18	45	154	601	0	602	77	17
BTR11	6	25	49	675	0	735	44	8	28	39	660	0	772	58
BTR12	3	18	11	132	763	0	436	6	15	20	105	714	0	418
BTR13	6	11	18	60	128	727	0	6	13	12	64	133	725	0

4.2 Data Collection Overview

Data is collected from the same days of the week in different time periods considering whether BT data can be used in urban transportation, and also determining the effect of external factors such as weather, education periods, etc (see Table 4.2). To capture the weekdays routine traffic, the days of the week are selected as Tuesday and Wednesday, excluding the high weekly movement days of Monday and Friday when compared to others. Simultaneously, data were collected with special attention in different seasons; when schools are open and closed; and even the first week of fall semester and the following week. The data generally belongs to the year 2019, when the thesis work started in Mersin, is further aimed to examine the change in 3 years, supported by the data of the first week and the following week of 2022 fall semester. The data collection process is also supported with FCD data to be used as ground truth on the same days of 2022.

Table 4.2 Data Collection Periods for Mersin BTR Network

Week	Date		Notes
	Tuesday	Wednesday	
W1	---	15 May 19 (48sec)	Before study Spring Semester
W2	21 May 19 (48sec)	22 May 19 (48sec)	Before study Spring Semester
W3*	**28 May 19 (48sec)**	**29 May 19 (Test Day)**	**Test Week for Inquiry Interval**
W4	2 July 19 (10sec)	3 July 19 (10sec)	After study Summer Break
W5	9 Jul 19 (10sec)	10 July 19 (10sec)	After study Summer Break
W6**	10 Sep 19 (10sec)	11 Sep 19 (10sec)	After study Start of Fall Semester
W7***	17 Sep 19 (10sec)	18 Sep 19 (10sec)	After study Fall Semester
W8	5 Nov 19 (10sec)	6 Nov19 (10sec)	After study Fall Semester
W9	12 Nov 19 (10sec)	13 Nov19 (10sec)	After study Fall Semester
W10**	6 Sep 22 (10sec)	7 Sep 22 (10sec)	After study Start of Fall Semester
W11***	13 Sep 22 (10sec)	14 Sep 22 (10sec)	After study Fall Semester

*During the test varying inquiry times of 10 secs, 20 secs, 30 secs and 48 secs were tried based on the test schedule.
**Start of Fall semester in September for 2019 and 2022
***Second week of Fall semester in September for 2019 and 2022

Following the preliminary analyses of data obtained at various times and locations, the 29th of May 2019 was chosen as the test day to determine the inquiry interval within the scope of the thesis and to investigate the effect of this determination on the quality of the collected data. Within the scope of the thesis, data were collected with a 48-second inquiry interval at the beginning of the study by default, while it was set to 10 seconds after this study.

4.3 Descriptive Statistics

Descriptive statistics for analyzed dates are summarized in Table 4.3. In the table "nReading" represents the total number of readings from all active BTRs during data collection period, while "nUniqueMAC" represents the total number of unique MAC_IDs. The data was collected between 06:00 and 21:00 except for 13 November, which was possibly due to a systematic error. Even though there are 48 BTRs in the Mersin network, all devices are active for only one day (6 November) out of 21 days of data. While the number of BTRs varies between 45 and 48 active devices for 2019; the number of active devices decreased to 34 in 2022.

Table 4.3 Collected Data Descriptive Statistic

Date	nBTRID	nReading	nUniqueMAC	nReading/MAC	nReading/BTR
15.May.19[a]	47	291349	25799	11.3	6198.9
21.May.19[a]	47	302987	26680	11.4	6446.5
22.May.19[a]	47	247851	24590	10.1	5273.4
28.May.19[a]	47	317330	26889	11.8	6751.7
29.May.19[b]	47	360036	26847	13.4	7660,3
2.July.19	47	459117	26731	17.2	9768.4
3.July.19	47	443666	26595	16.7	9439.7
9.July.19	47	427923	26695	16.0	9104.7
10.July.19	47	428998	26525	16.2	9127.6
10.Sept.19	47	464276	29325	15.8	9878.2
11.Sept.19	47	462894	28806	16.1	9848.8
17.Sept.19	47	467201	29557	15.8	9940,4
18.Sept.19	47	466566	29005	16.1	9926.9
5.Nov.19	46	479680	30189	15.9	10427.8
6.Nov.19	48	587321	30233	19.4	12235.9
12.Nov.19	45	477878	30324	15.8	10619.5
13.Nov.19[c]	45	334918	26520	12.6	7442.6
6.Sept.22	34	670753	43765	15.3	19728.0
7.Sept.22	34	692717	43837	15.8	20374.0
13.Sept.22	34	666083	43998	15.1	19590,7
14.Sept.22	34	668534	43820	15.3	19662.8

[a] *Inquiry Interval is 48 second*
[b] *Test Day, Inquiry Interval is 10, 20, 30 and 48 second*
[c] *Time interval is 10:32-20:59*

In the table, the percentage of readings per MAC_ID (nReading/MAC) is calculated using the total number of readings and the number of unique MAC_IDs. This value increased slightly with decreasing the inquiry interval but remained very close when the inquiry interval was equal to 10 seconds despite the date change. Furthermore, because the number of active BTRs varies greatly from day to day, the percentage of readings per BTR (nReading/BTR) is calculated using the total number of readings and the number of active BTRs to examine the change of BTR reading. The percentage of readings per MAC_ID is in a general upward trend during the data days. This increase has reached a very high level by 2022 despite the decrease in the number of active BTRs. With this result, it can be concluded that the BT capture rate from vehicles or non-vehicles has increased over time, which supports the efficient usability of BT data for urban traffic.

4.4 Mersin Data Analysis Framework

A case study application was made on the corridor determined in Mersin in order to discuss the application process and outputs to propose a comprehensive method based on the general data analysis framework discussed in Chapter 3. In the base framework the usage of Bluetooth data in urban transportation is discussed in detail, but as highligted there is a need for further evaluation to determine the threshold and propose a framework for Mersin. To that end, the framework prepared in order to compile the whole stages of the process and summarize its usage within the scope of the case study after preprocessing is presented in Figure 4.3. In this framework rescan time threshold is determined as 50 seconds, and slow movement elimination is determined as 4 km/h according to preprocessing of the case study data. These threshold values can differ according to BTR data collection characteristics and region specific situation. Determination process for these threshold will be further evaluated in the following sections.

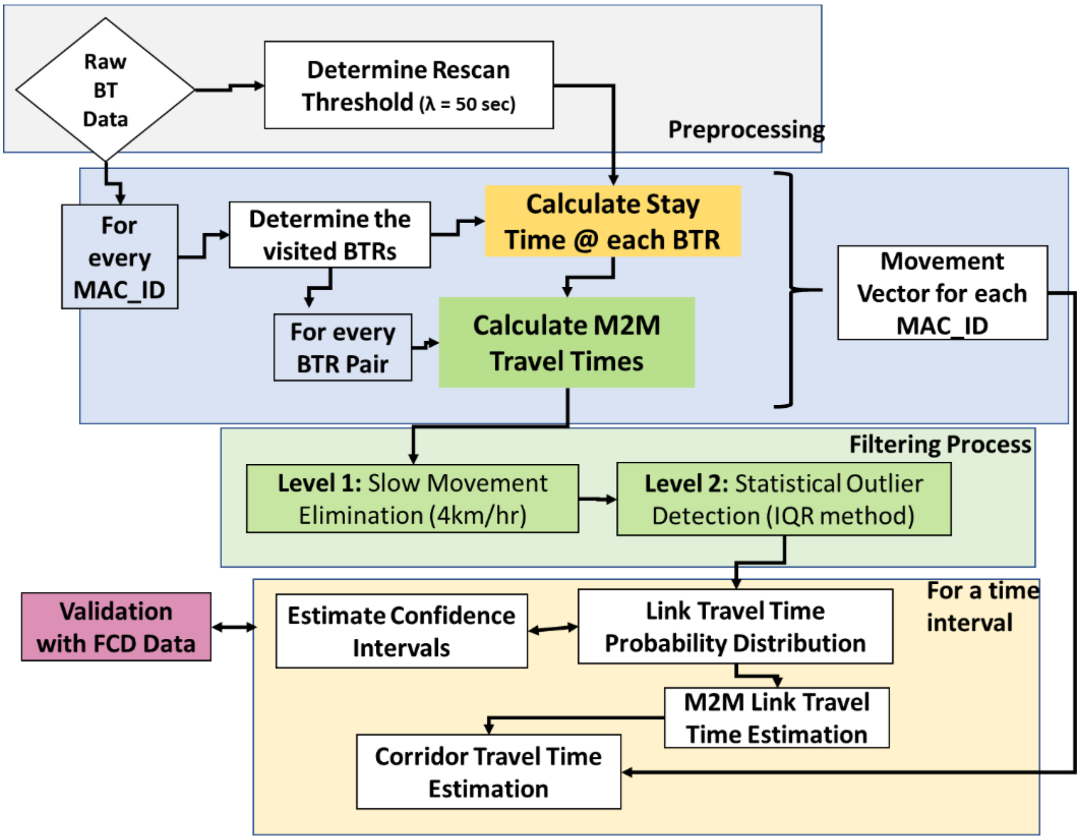


Figure 4.3. Data Analysis Framework for Mersin.

4.4.1 Determination of Rescan Threshold

The frequency of daily rescan times should be investigated to evaluate peak value of the distribution and general pattern of rescan times. Within the scope of the thesis, because of examining the daily rescan time distributions, a special study was conducted for the inquiry interval time, which will be explained in detail in the following sections, based on the BTR reading characteristics. As a result of this study, the inquiry interval value is changed to 10 seconds. After that, all rescan time distribution graphs analyzed for 4 different days show a single peak value in the region of approximately 12.5-13 seconds (Figure 4.4). This value is approximately 1.28 times the specified 10 second inquiry interval value (inquiry scan interval), meaning that even though the BTR did not receive any signals, it scans the coverage area every 12.8 seconds to capture active MAC_IDs. It is clearly seen that the rescan time is generally less than 30 seconds, with the new inquiry interval determined within the scope of the thesis.

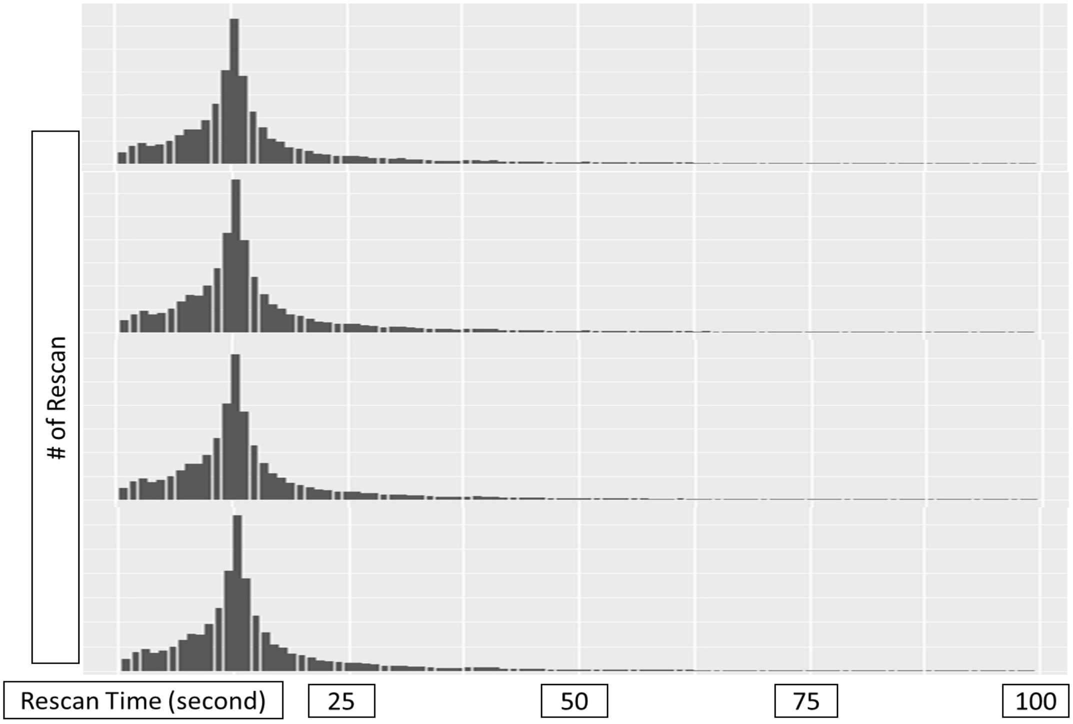


Figure 4.4. 4-Day Rescan Time Frequency Distribution

These distributions reveal that a 50 second rescan time threshold (λ) can be safely determined in the scope of this study. The stay time value is calculated by summing the reading times sequentially captured by the same BTR for the same MAC_ID (rescan) if it is less than this threshold value. When there is a rescan time value greater than this value:

- if it is the first BTR from which the MAC_ID is captured, the readings before the time greater than 50 seconds are ignored,

- if it is not the first BTR point, the movement up to the last reading of 50 seconds before is interrupted at that point and after this value a new movement is considered to start from the first reading.

4.4.2 Effect of Determined Rescan Threshold on Stay Time Calculation

During the Bluetooth data analysis process, to examine the effect of the determined rescan time threshold value on the stay time calculation, we can consider the MAC1 device in the Sample Conceptual Map, the readings of which are tabulated in Table 4.4. Readings greater than the 50 second threshold, highlighted in red, are ignored according to the algorithm, assuming that the device is captured by only 2 BTRs. It is highly probable that 2 values greater than 50 seconds are seen for 2 consecutive readings and then read by the same device after 5 seconds meaning that the rescan values may be extended due to a technical reason or another obstacle in front of it or the vehicle made an unpredictable movement.

Table 4.4 Reading Data for MAC1 from Sample in Figure 3.4

Reading ID	BT_ID	MAC_ID	Discovery Time	Device Type	Δt
$t_{1;1}$	BTR1	MAC1	1549918801	240414	NA
$t_{1;2}$	BTR1	MAC1	1549918802	240414	1
$t_{1;3}$	BTR1	MAC1	1549918805	240414	3
$t_{1;4}$	BTR1	MAC1	1549918808	240414	3
$t_{1;5}$	BTR2	MAC1	1549918843	240414	35
$t_{1;6}$	**BTR2**	**MAC1**	**1549918898**	**240414**	**55**
$t_{1;7}$	**BTR2**	**MAC1**	**1549919082**	**240414**	**184**
$t_{1;8}$	**BTR2**	**MAC1**	**1549919087**	**240414**	**5**

After these 3 readings are ignored, the stay time values calculated as 7 seconds and 244 seconds in Equation 3.4 and 3.5 are recalculated as 7 seconds and 0 seconds. This revision is summarized in Table 4.5 with the movement vector of MAC1 before and after threshold. This example shows that the accuracy of the calculations made with the threshold value determined for the rescan time significantly increased the data analysis process.

Table 4.5 Movement for MAC1 from Sample Conceptual Map before and after Rescan time threshold.

MAC_ID	**MAC1 (Before)**		**MAC1 (After)**
Device Type	**1**		**1**
BTR_ID	BTR1		BTR1
Epoch Time	1549918801		1549918801
nRescan	3		3
Stay Time	7	**λ =50 second**	7
Ttime (L2F)	NA		NA
BTR_ID	BTR2		BTR2
Time	1549918843		1549918843
nRescan	3		**NA**
Stay Time	**244**		**NA**
Ttime (L2F)	35		35

4.5 Determination of Possible Moving Vehicle Travel Time

Within the scope of the thesis, it is aimed to determine possible moving vehicles in link travel time calculation in order to eliminate unexpected situations such as the technical errors caused by Bluetooth data collection and captured non-vehicle MAC_IDs (pedestrian, cyclist, etc.). Within the scope of this determination, firstly, the effect of the rescan threshold defined in the previous section on the travel time calculation will be explained with an example and its contribution will be stated. In addition, it is important to determine the efficiency and accuracy of the travel time calculation methods mentioned in the literature and to examine their use within the scope of the methodology. After all these general evaluations, the outlier detection methodology proposed in this thesis will be presented in this section and its accuracy will be evaluated with FCD data.

4.5.1 Effect of Determined Rescan Threshold on Travel Time Calculation

It was emphasized that the stay time values calculated for BTRs cause differences in the travel time results according to calculation methods explained in Section 3.3.1. After excluding the last 3 readings in Table 4.4, since no stay time is calculated in the 2nd BTR capture zone, the change that occurred for Equation 3.11 to 3.15, is given below:

$\Omega_{1,1,2}^{BT,L2F} = t_{1;5} - t_{1;4}$=35 sec → $t_{1;5} - t_{1;4}$=35 sec

$\Omega_{1,1,2}^{BT,F2F} = t_{1;5} - t_{1;1}$=42 sec → $t_{1;5} - t_{1;1}$=42 sec

$\Omega_{1,1,2}^{BT,L2L} = t_{1;8} - t_{1;4}$=279 sec → $t_{1;5} - t_{1;4}$=35 sec

$\Omega_{1,1,2}^{BT,F2L} = t_{1;8} - t_{1;1}$=286 sec → $t_{1;5} - t_{1;1}$=42 sec

$\Omega_{1,1,2}^{BT,M2M} = t_{1;5} - t_{1;4} \frac{(\delta t_{1;1})+(\delta t_{1;2})}{2}$=160.5 sec → $t_{1;5} - t_{1;4} + \frac{(\delta t_{1;1})+(\delta t_{1;2})}{2}$=38.5 sec

When the changes are examined, it is seen that the results became very close after the determined threshold, in contrast to the serious differences between the calculation methods without any limit values. With these results, it is clearly seen how important the threshold determination for the rescan time is in the data analysis process. In all the calculations made after this stage within the scope of the thesis, stay time was calculated according to the threshold value determined for rescan time, and this threshold was also taken into account in the movement vector algorithm.

4.5.2 Effect of Travel Time Calculation Methods

It was observed that after the threshold determined for the stay time calculation in the previous section and the difference between the travel times calculated with the 5 methods in the literature decreased significantly. However, in order to examine the difference between these calculation methods, the travel time values calculated with five methods for the selected BTR pair are represented in the Figure 4.5 according to the time of day. While it is noteworthy that there are very extreme values in the calculated travel times in the figure, there is no significant difference between the methods. Since travel time values are mostly under 1 hour, the graphic axis is limited to 3600 seconds for travel time and zoomed in to examine whether there is a difference between the calculation methods, and it was seen that there was no significant difference between the methods. These examinations and evaluations were also made and compared for many different BTR pairs within the network; no significant difference was observed between the methods in any of the graphs.

In the light of this result, considering that the distance between the exact points where the BTR devices are located is known and considering that it would be appropriate to include the stay time value of both BTRs in the calculation, the mid-to-mid (M2M) travel time calculation method is used in the data analysis process.

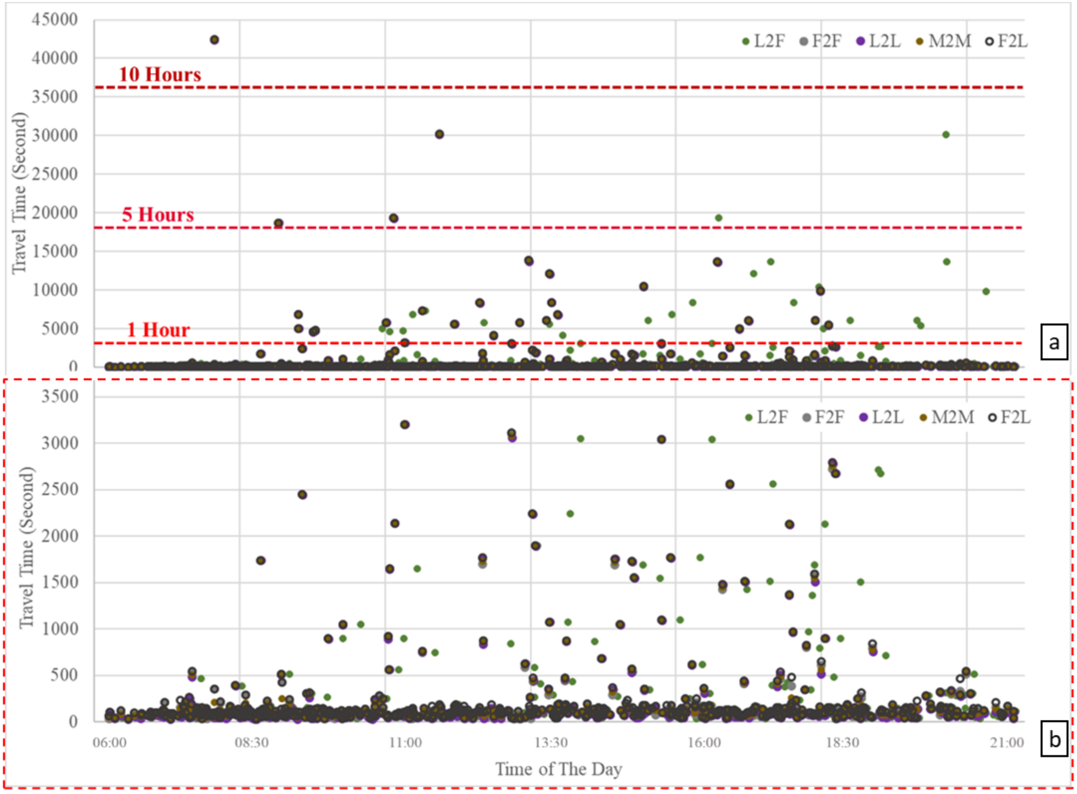


Figure 4.5. Daily travel time distribution of different calculation methods for a selected BTR Pair a) all data and b) truncated values at 1 hour (3600 second)

4.5.3 M2M Travel Time Calculation Methods

The calculated daily M2M travel time distribution is seen in Figure 4.6, reflecting both all the results and truncated at 3600 seconds. It can be seen from the figure that there are serious extreme values that may reflect pedestrian/bicycle or other unpredictable movements. Therefore, it is very important to determine the outlier detection method for this data. When the distribution graph is cut at 1000 seconds, it can be easily seen that majority of the values are below 5 minute (300 seconds), even the travel times over 300 seconds may be the possible outliers. Since the travel time varies depending on the changes in the traffic situation during the day, the Outlier detection application will be made specifically for 15-minute, 30-minute, 1 hour and 2-hour intervals of the day (between 6:00 to 21:00), respectively.

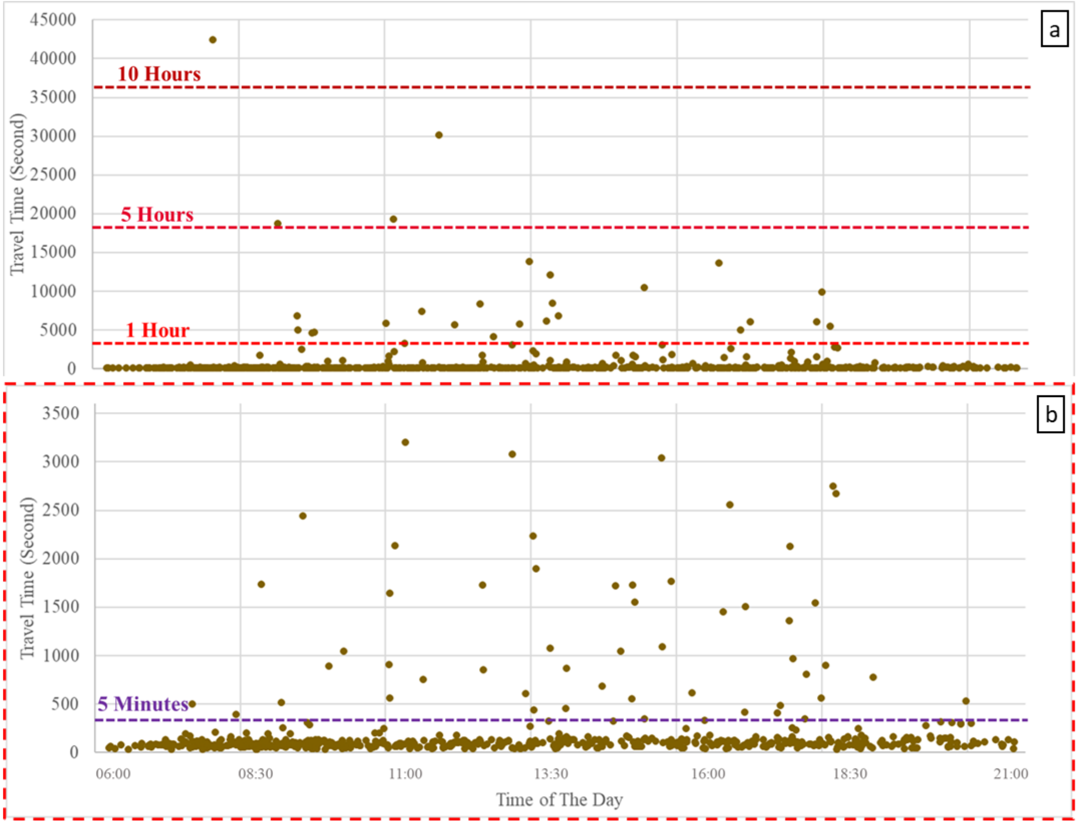


Figure 4.6. Daily M2M travel time distribution from BTRi to BTRj a) all data and b) truncated values at 3600 second.

Daily M2M travel time frequencies for selected 6 consecutive BTR pairs for 4 days respectively with dashed lines and 4-day total with continuous line color coded according to BTR pair is represented in Figure 4.7. While creating the distributions, the histogram intervals were determined as variables in accordance with the data characteristics and fixed as a singular value after a certain upper value. Each BTR pair on the same corridor has a different length, and any significant variation is identified when day or BTR pair comparisons are made. Although some BTR pairs have positive kurtosis for distribution and some have normal kurtosis, all groups generally show positive skewness. The frequency chart given as an example has been examined for many corridors and time intervals and reveals a rather small sample after the interval, which usually corresponds to 255 seconds.

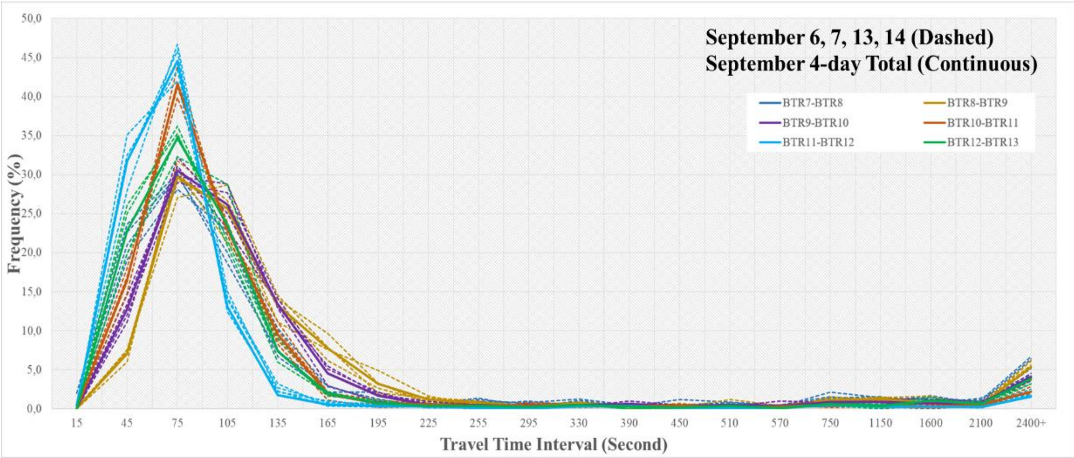


Figure 4.7. Daily M2M travel time frequencies for 6 BTR pairs

In order to examine whether the travel time varies according to the time of day and to evaluate the effects of possible differences on the outlier detection method, the travel time frequency graphs calculated at 2-hour intervals for selected 3 BTR pairs over the corridor are given in the Figure 4.8. Since the distances between the 3 selected BTR pairs are different, it is noteworthy that they have similar distributions, although there are small differences in the travel time with maximum frequency. More importantly, no significant hourly variability was detected because of the analyses made by dividing the day into 2 hours. However, it is noteworthy that there are sudden changes in travel time frequencies in these distributions, which are prepared in 2-hour intervals, since there is a serious decrease in the number of samples.

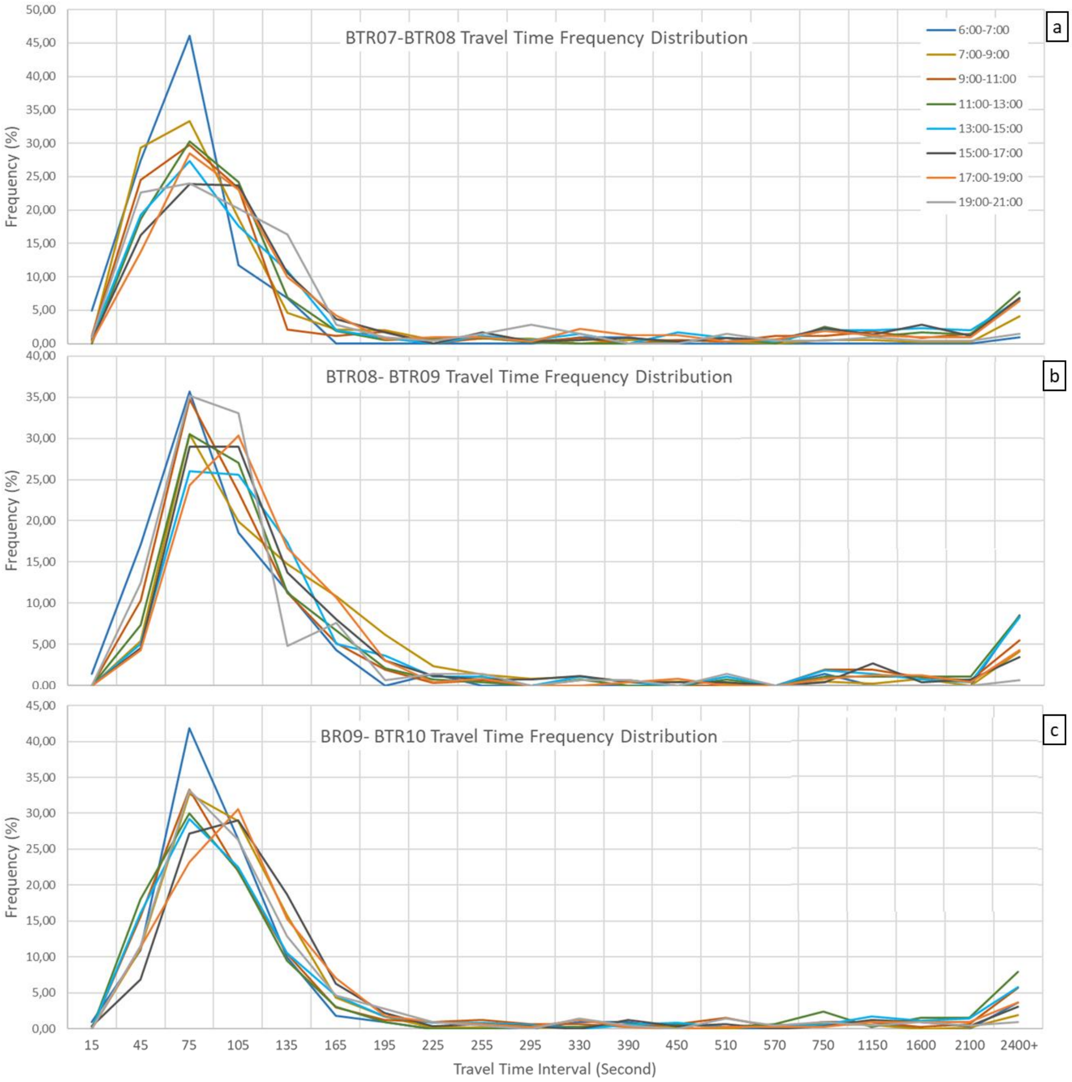


Figure 4.8. M2M travel time frequencies for 2-hour interval for a) BTR07-BTR08, b) BTR08-BTR09 and c) BTR09-BTR10 pair

4.5.4 Outlier Detection Method for M2M Travel Time Calculation Methods

To see the possible scenarios in detail specific to the interval; it is aimed to clean the data using the most frequently used statistical outlier detection method (IQR method) in the literature. All captured movements for each BTR pair and for each time interval are sorted from smallest to largest. The important statistical points are determined in Figure 4.9; the minimum value (min/Q0), the first quartile range (Q1), mean (m/Q2), the third quartile range (Q3), maximum value (max/Q4), Interquartile range (IQR=Q3-Q1). Since the captured MAC_IDs are not only vehicles but also include pedestrians, etc., it is seen that there are statistically serious deviations when all data are examined without outlier detection by any method. So, data needs to be cleaned up statistically between defined lower bound (LB) and upper bound (UB) (Ranga Suri et. al., 2019). The lower and upper limit of outlier detection can be determined by very simple statistical methods that are commonly used, such as;

- IQR Method: LB= Q1-1.5*IQR and UB= Q3+1.5*IQR
- Mid-50 Method: LB= Q1 and UB= Q3

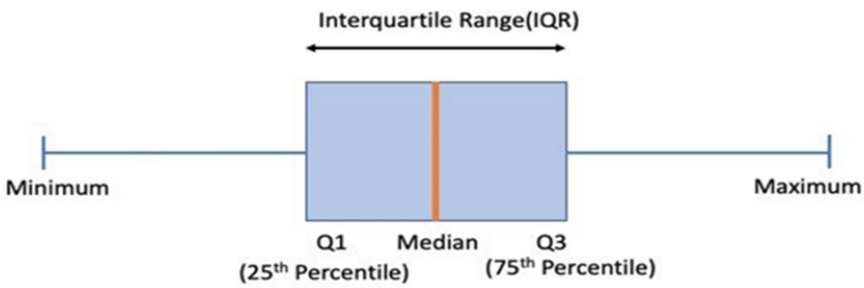


Figure 4.9. A sample box plot indicating important statistical points.

In order to evaluate the capacity of the most commonly used statistical outlier detection method with calculated M2M travel time data, the intervals between 6:00-9:30 in the morning for the selected BTR pair are summarized in Table 4.6. According to the table, data cleaning processes to be performed at 30-minute time intervals using presented the outlier detection methods mentioned above reveal some problematic situations. For example, for the first time interval (6:00-6:30) there are only 5 MAC_ID make a movement on this link and results varies in a big range. The two methods are applied respectively, and both did not eliminate 444 second travel time (which reflects a 5.56 km/h speed) on the link possibly due to the extreme value of 11136 second. Similarly, due to 2 extreme values such as 6518.5 and 37258 in the last time interval (9:00-9:30), values such as 626, 750 and 756 that were clearly outliers in the data could not be cleaned with basic statistical methods. However, as in the 5th interval (8:00-8:30), very high values of 12633.5 and 33343.0 seconds can be cleared by both methods, revealing similar average travel times.

Table 4.6. Cross BTR Travel Times for selected BTR Pair under 30-min intervals between 6:00-9:30.

Length: 680 m	**6:00-6:30**	**6:30-7:00**	**7:00-7:30**	**7:30-8:00**	**8:00-8:30**	**8:30-9:00**	**9:00-9:30**
N	5	3	11	19	22	23	20
Min TT	73.50	63.00	48.50	41.00	56.00	55.00	51.00
TT-Q1	82.00	101.00	66.00	63.00	82.13	86.00	75.75
Median-TT	121.00	139.00	88.00	75.50	119.00	125.00	106.00
TT-Q3	444.00	172.25	141.00	102.00	151.63	174.25	410.38
Max-TT	11136.00	205.50	312.50	38003.50	33343.00	37911.50	37258.0
IQR	362.00	71.25	75.00	39.00	69.50	88.25	334.63
TT-LB	-461.00	-5.88	-46.50	4.50	-22.13	-46.38	-426.19
TT-UB	987.00	279.13	253.50	160,50	255.88	306.63	912.31
std dev.	4902.02	71.30	75.66	8698.72	7451.61	8059.78	8333.48
AvgTT	**2371.30**	**135.83**	**116.14**	**2082.63**	**2190,68**	**2403.76**	**2376.23**
AvgTT_Method1	**180,13**	**135.83**	**96.50**	**76.69**	**110,93**	**118.11**	**208.22**
AvgTT_Method1	215.67	139.00	95.60	78.83	114.85	127.64	130.65

Time Interval	**M2M Cross-BTR Times**
6:00-6:30	73.5-82-121-**444**-**11136**
6:30-7:00	63-139-205.5
7:00-7:30	48.5-62.5-64-68-72.5-88-117.5-132-150-162-**312.5**
7:30-8:00	**41**-53-53.5-54-61-65-66-66-73-75.5-88-89-92-95-109-146-161-178.5-**38003.5**
8:00-8:30	56-58-63-65.5-73.5-80,5-87-89-96-114-117-121-125-126-127.5-146-153.5-156.5-166.5-197-**12633.5-33343.0**
8:30-9:00	55-57.5-58-71-74-79-93-93.5-97-109-124.5-125-127-147.5-153-161.5-173-175.5-270-397-**4564**
9:00-9:30	51-52,5-65-69-72-77-85-86-95,5-100-112-136,5-137-139-**338,5-626-750-756**-**6518,5-37258**

The outlier detection process for this 3-example problematic time interval data was conducted with two methods respectively to show the constraints in detail (Table 4.7). According to this table, even if the data is cleaned up according to the two methods, there are still some values in the interval that can be counted as outliers. Although the results are relatively acceptable for data cleaned with the Mid-50 method, consideration should be given to the use of this method as it causes a significant decrease in the number of data samples in the range.

As it can be briefly but strikingly explained with this sample data set, cleaning BT transportation data with basic statistical methods has serious margins of error despite many successful studies in the literature. Although most of these studies in the literature have achieved successful results because these methods were applied in a short time interval (i.e. 2 hours only) or in limited road sections with less unexpected movement, that is, in conditions where the margin of error is relatively low. Within the scope of the study, it can create too many errors in the case of application in small time intervals such as 15-30 minutes or in open networks with any kind of movements can be seen. It was also seen that a 2-stage application as a new method gave more accurate results as data cleaning and analysis of many days with different BTR pairs with different characteristics and opposite directions of them and at different intervals such as 15 minutes, 30 minutes, 1 hour and 60 minutes.

Table 4.7. Outlier detection of Cross BTR Travel Times for selected BTR Pair under 3 example 30-min intervals with a) IQR method and b) Mid-50 method

	Filtered with IQR method			**Filtered with Mid-50 Method**		
Length: 680 m	**6:00-6:30**	**8:00-8:30**	**9:00-9:30**	**6:00-6:30**	**8:00-8:30**	**9:00-9:30**
N	4	20	18	3	11	10
Min TT	73.50	56.00	51.00	82.00	87.00	77.00
TT-Q1	79.88	78.75	73.25	101.50	105.00	88.38
Median-TT	101.50	115.50	97.75	121.00	121.00	106.00
TT-Q3	201.75	132.13	138.50	282.50	126.75	136.88
Max-TT	444.00	197.00	756.00	444.00	153.50	338.50
IQR	121.88	53.38	65.25	181.00	21.75	48.50
TT-LB	-102.94	-1.31	-24.63	-170.00	72.38	15.63
TT-UB	384.56	212.19	236.38	554.00	159.38	209.63
std dev	177.13	39.90	75.66	198.70	21.35	76.61
Avg Speed	**13.59**	**22.07**	**11.76**	**11.35**	**20.68**	**18.74**
AvgTT	180.13	110.93	208.22	215.67	118.36	130.65

	Time Interval	**M2M Cross-BTR Times**
IQR method	**6:00-6:30**	73.5-82-121-**444**-~~11136~~
	8:00-8:30	56-58-63-65.5-73.5-80.5-87-89-96-114-117-121-125-126-127.5-146-153.5-156.5-166.5-197-~~12633.5-33343.0~~
	9:00-9:30	51-52,5-65-69-72-77-85-86-95,5-100-112-136,5-137-139-**338.5-626-750-756**-~~6518.5-37258~~
Mid-50 Method	**6:00-6:30**	~~73.5~~-82-121-**444**-~~11136~~
	8:00-8:30	~~56-58-63-65.5-73.5-80,5~~-87-89-96-114-117-121-125-126-127.5-146-153.5-~~156.5-166.5-197-12633.5-33343.0~~
	9:00-9:30	~~51-52.5-65-69-72~~-77-85-86-95,5-100-112-136,5-137-139-**338.5**-~~626-750-756-6518.5-37258~~

To sum up, proposed outlier detection methodology is seen in Figure 4.10, starting with a slow-movement (long-tail) elimination and continues with the commonly used statistical method (IQR) to eliminate both the slowest and fastest vehicles without decreasing the sample much. The contribution of this new method proposed in this thesis is that it carries out a 2-stage process, although a common outlier detection method is included. In stage 1 the vehicles moving slower than 4 km/h were eliminated from the data. In the literature 4-5 km/h shows the traffic congestion; by this information lower than 4km/h most probably reflects the unexpected movements when the possible moving vehicle needs to be determined. In the second stage, after elimination of the long tails represented in Figure 4.7 and Figure 4.8, data seemed more closer to Normal distribution of which needs further detection with IQR method.

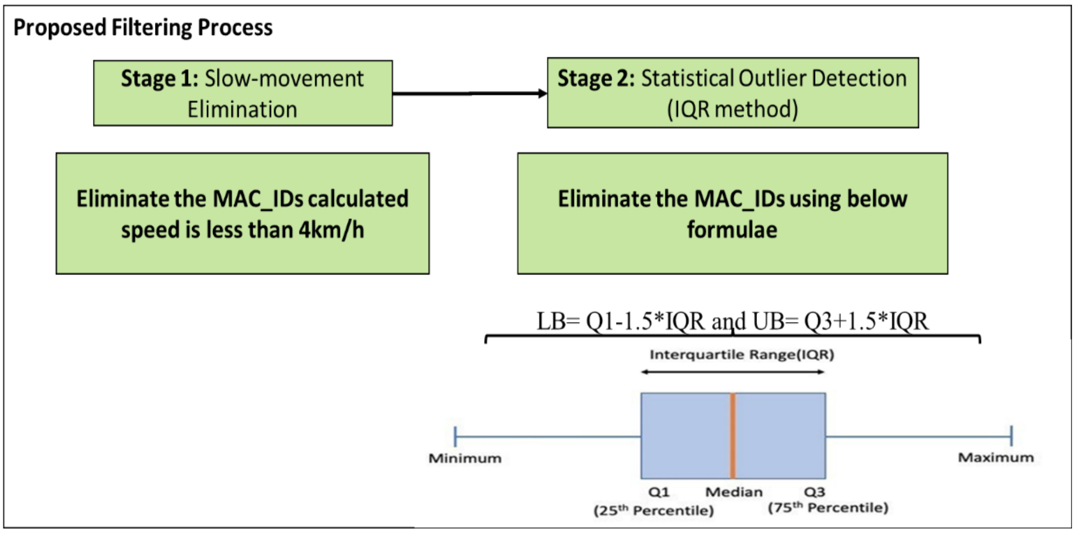


Figure 4.10. Proposed Outlier Detection Methodology

The proposed outlier detection process for the same 3-example problematic time interval data is conducted to show the effect of the methodology (Table 4.8). At the first stage the vehicles travelling lower than 4 km/h (which correspondence to 612 second travel time for this BTR Pair) are eliminated from the data. In this stage nearly all the extreme values can be eliminated from the data in the first run. In the second stage remaining travel times are further cleaned-up according to the IQR method. Since the extreme values of the first stage were seriously eliminated, only two outliers were removed in the second level cleaning. The fact that the average speed is over 20km/h in all 3-time intervals determined according to the cleaned data reveals a very acceptable result considering that there is no traffic jam in the region.

Table 4.8. Example of Proposed Outlier detection of Cross BTR Travel Times for above selected BTR Pair under 3 example 30-min intervals.

	Stage 1 (Eliminate lower than 4km/h)			**Stage 2 IQR method**		
Length: 680 m	**6:00-6:30**	**8:00-8:30**	**9:00-9:30**	**6:00-6:30**	**8:00-8:30**	**9:00-9:30**
N	4	20	15	3	20	14
Min TT	73.50	56.00	51.00	73.50	56.00	51.00
TT-Q1	79.88	78.75	70.50	77.75	78.75	69.75
Median-TT	101.50	115.50	86.00	82.00	115.50	85.50
TT-Q3	201.75	132.13	124.25	101.50	132.13	109.00
Max-TT	444.00	197.00	338.50	121.00	197.00	139.00
IQR	121.88	53.38	53.75	23.75	53.38	39.25
TT-LB	-102.94	-1.31	-10.13	42.13	-1.31	10.88
TT-UB	384.56	212.19	204.88	137.13	212.19	167.88
std dev.	177.13	39.90	70.18	25.33	39.90	30.24
Avg Speed	13.59	22.07	22.72	26.56	22.07	26.83
AvgTT_	180.13	110.93	107.73	92.17	110.93	91.25

	Time Interval	**M2M Cross-BTR Times**
Stage 1	**6:00-6:30**	73.5-82-121-444-~~11136~~
	8:00-8:30	56-58-63-65.5-73.5-80.5-87-89-96-114-117-121-125-126-127.5-146-153.5-156.5-166.5-197-~~12633.5-33343.0~~
	9:00-9:30	51-52.5-65-69-72-77-85-86-95.5-100-112-136.5-137-139-338.5-~~626-750-756-6518.5-37258~~
Stage 2	**6:00-6:30**	73.5-82-121-~~444~~
	8:00-8:30	56-58-63-65.5-73.5-80.5-87-89-96-114-117-121-125-126-127.5-146-153.5-156.5-166.5-197
	9:00-9:30	51-52.5-65-69-72-77-85-86-95.5-100-112-136.5-137-139-~~338.5~~

4.5.5 Accuracy of Proposed Outlier Detection Method with respect to FCD Data

The data cleaned as "possible moving vehicle" after outlier detection with basic statistical analyzes and proposed method is compared with FCD data for the same BTR pair on the same days within selected time intervals. The Mean Absolute Error

(MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) of 3 different outlier detection method with FCD data within 4 different time intervals through the day is presented in Table 4.9. The results revealed that when the proposed method is compared with the other two methods, the lowest error margins are observed when the day is monitored at 1-hour intervals and cleaned with the new method. In general, it is noteworthy that the margin of error is higher when the time interval is short in calculations made with Bluetooth data. The main reason for this is considered to be that the number of MAC_IDs captured in a short time interval is less and even minor problems or unexpected situations can seriously change the calculations.

Table 4.9 Error comparison of BTR07- BTR08 Pair on average comparative fused FCD vs. BT speed over aggregated time cleaned up with a) proposed outlier detection method, b) IQR method and c) Mid-50 method for

		MAE (km/h)	**RMSE (km/h)**	**MAPE (%)**
a) Proposed Outlier Detection Method	15-minute interval	11.19	12.28	70.28
	30-minute interval	20.26	20.59	125.22
	1-hour interval	***10.92***	***11.40***	***63.86***
	2-hour interval	11.51	12.14	65.58
b) IQR Method	15-minute interval	12.78	14.03	131.38
	30-minute interval	21.67	22.27	162.86
	1-hour interval	11.50	11.95	72.68
	2-hour interval	11.75	12.31	67.69
c) Mid-50 Method	15-minute interval	13.92	16.10	114.65
	30-minute interval	22.32	22.80	168.25
	1-hour interval	12.49	12.93	82.64
	2-hour interval	13.02	13.50	82.28

In order to determine which time interval can be used to examine the change of time dependent nature of traffic with Bluetooth data, the distribution of 1-day possible moving vehicle travel time values of the same BTR pair was examined by considering FCD speed data at 4 different time intervals (Figure 4.11). In the graphics the critical statistical values for travel times analyzed for 15-minute, 30-minute, 1-hour and 2-hour intervals throughout the day and the BT and FCD speed data calculated against these values are depicted, respectively. The red and black lines shown in bold and straight in the graph show the speed change on the right axis of the graph for Bluetooth and FCD data, respectively. The left axis, on the other hand, refers to the travel time values given as thin and dashed in seconds. When the results are compared, although FCD data always has higher speeds than BT data, the differences are reduced somehow during peak hour traffic. Considering that the speed axes are not equal due to representational constraints, the interval value at which the FCD and BT speed data are closest to each other is 15-minutes. But at the same time, this interval is one of the intervals that causes a lot of errors in the data cleaning process, since it provides the least data sample.

For this reason, the average comparative fused FCD and BT speed error comparisons for 4 days (September 6, 7, 13 and 14) and 2 different time intervals (1 hour and 2 hours) over the aggregated time cleaned up with proposed outlier detection method for 6 BTR pairs are presented in the Table 4.10. According to the table, although the MAPE value was determined as inaccurate for the BTR07-BTR08 pair for 2 days, the predictions made for the other BTR pairs and days were determined as accurate/reasonable with respect to the FCD data. Especially for the BTR pair (BTR10-BTR11) where the distance between each other is the longest, the accuracy between FCD and BT speed values is calculated as highly accurate forecast for all days and intervals.

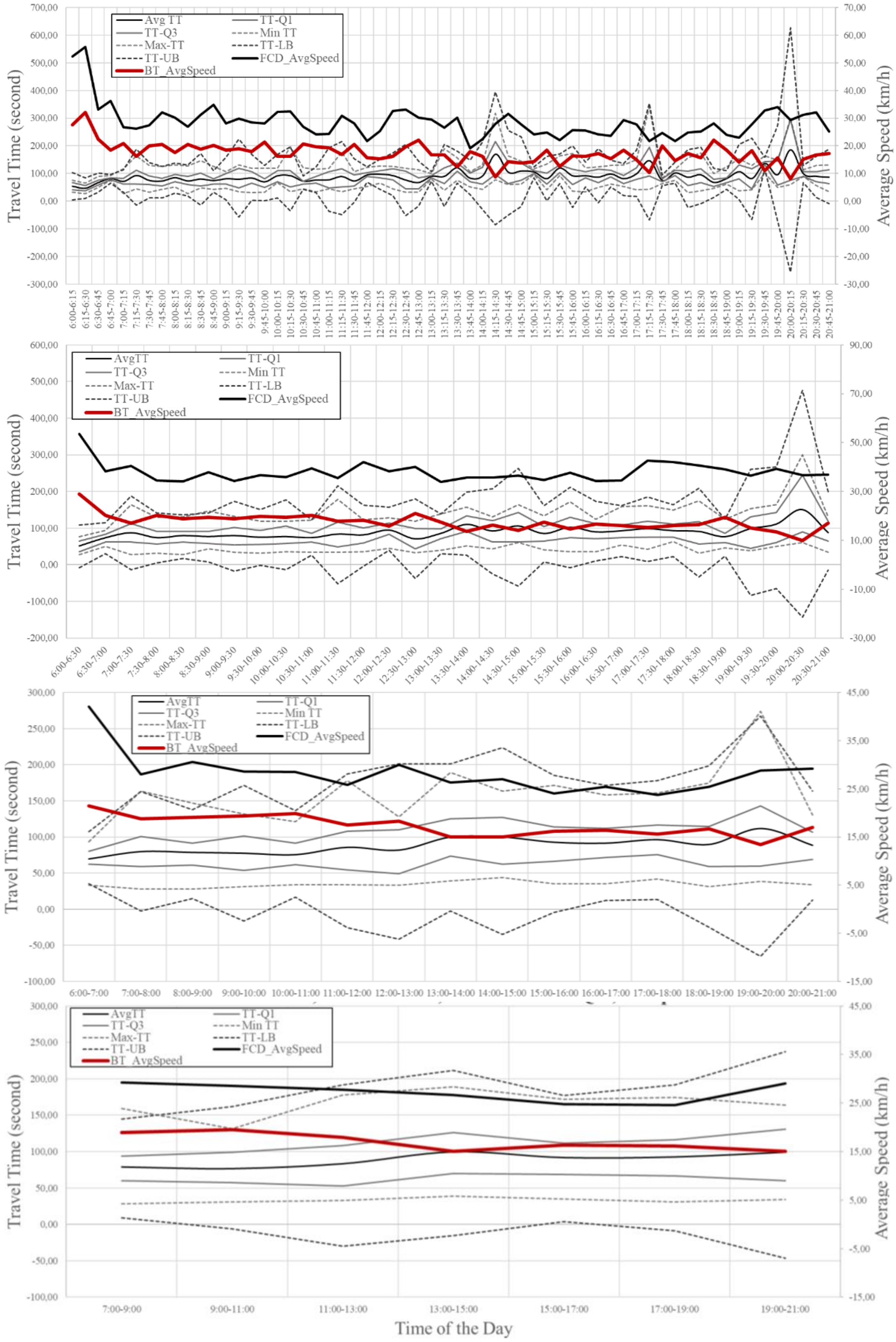


Figure 4.11. Daily cleaned Travel time change of BTR07- BTR08 Pair on September 6, 2022, in a)15-minute, b)30-minute, c)1-hour and d) 2-hour interval.

Table 4.10 Error comparison on average comparative fused FCD vs. BT speed over aggregated time periods (1 hour and 2-hour periods)

		1 hour			**2 hours**		
		MAE (km/h)	**RMSE (km/h)**	**MAPE (%)**	**MAE (km/h)**	**RMSE (km/h)**	**MAPE (%)**
September 6, 2022	BTR07-BTR08	**10.92**	**11.40**	**63.86**	**10.21**	**10.37**	**61.23**
	BTR08-BTR09	**2.60**	**3.63**	9.12	**1.78**	**1.97**	6.43
	BTR09-BTR10	**6.25**	**6.92**	30.21	**5.55**	**5.65**	27.35
	BTR10-BTR11	**2.79**	**3.57**	7.20	**1.85**	**2.15**	4.78
	BTR11-BTR12	**2.84**	**3.70**	9.25	**2.39**	**3.03**	7.80
	BTR12-BTR13	**4.12**	**5.07**	14.33	**2.81**	**3.37**	9.73
September 7, 2022	BTR07-BTR08	**8.03**	**8.97**	**45.26**	**7.56**	**7.87**	**42.71**
	BTR08-BTR09	**2.58**	**3.89**	9.01	**1.37**	**1.63**	5.20
	BTR09-BTR10	**7.34**	**7.98**	37.28	**6.78**	**6.87**	34.66
	BTR10-BTR11	**3.55**	**4.23**	8.70	**2.42**	**2.82**	6.09
	BTR11-BTR12	**3.47**	**4.49**	10.97	**1.93**	**2.20**	6.17
	BTR12-BTR13	**3.39**	**4.57**	11.67	**4.76**	**5.04**	17.92
September 13, 2022	BTR07-BTR08	**6.02**	**6.45**	32.61	**5.71**	**5.90**	31.49
	BTR08-BTR09	**2.94**	**4.00**	10.01	**3.04**	**3.31**	10.89
	BTR09-BTR10	**5.01**	**5.57**	26.03	**4.80**	**5.03**	24.66
	BTR10-BTR11	**4.21**	**4.87**	10.25	**3.55**	**4.18**	8.96
	BTR11-BTR12	**2.46**	**2.95**	8.32	**2.26**	**2.56**	7.68
	BTR12-BTR13	**4.07**	**4.84**	14.35	**3.41**	**3.74**	12.29
September 14, 2022	BTR07-BTR08	**6.08**	**7.00**	34.17	**5.86**	**6.18**	33.38
	BTR08-BTR09	**6.08**	**7.00**	34.17	**3.21**	**3.88**	13.14
	BTR09-BTR10	**5.42**	**5.99**	28.00	**5.02**	**5.33**	26.07
	BTR10-BTR11	**3.28**	**3.90**	8.29	**2.87**	**3.38**	7.44
	BTR11-BTR12	**1.73**	**2.61**	5.68	**0.93**	**1.28**	3.12
	BTR12-BTR13	**2.44**	**3.39**	8.82	**1.46**	**2.04**	5.40

4.6 Urban Arterial Cross BTR Travel Time Estimation

The corridor travel times were investigated in a link-based approach, to combine the links as a corridor due to the data availability for pairwise evaluations. The possible moving vehicle link travel time information is investigated after introduced outlier detection method to determine an average travel time and expected travel time interval for each BTR pair and summarized in Table 4.11 for the whole day. To estimate corridor travel time as the summation of link travel time, the data characteristics and distribution of each link travel time must be determined statistically. To check the normality of each link with both the cleaned data and the all data, the distribution of link travel times are presented for 6 BTR pair in Figure 4.12 and Figure 4.13. When the distributions are examined, it is seen that the data get very close to the normal distribution curve with the cleaning process. However, it is not the right approach to complete the normality check process of the data by only visually examining the distributions.

Normality check evaluation was made in SPSS environment for all BTR pairs on the corridor and the results are presented in the Table 4.12. by applying Kolmogorov-Smirnov and Shapiro-Wilk methods. The Kolmogorov-Smirnov test and the Shapiro-Wilk test are both tests for normality. In Kolmogorov-Smirnov test, the cumulative distribution function of the data is compared to the cumulative distribution function of a normal distribution. A significant p-value and the small test statistic in this test imply that there is insufficient evidence to reject the null hypothesis that the data came from a normal distribution. According to the Shapiro-Wilk test, there isn't enough evidence to reject the null hypothesis of normality for them with the small test statistic (close to 1) and a significant p-value.

Table 4.11. Descriptive Statistics of Daily Corridor Cross-BTR (between BTR07 to BTR13) Travel Time Investigation

	BTR07-BTR08	**BTR08-BTR09**	**BTR09-BTR10**	**BTR10-BTR11**	**BTR11-BTR12**	**BTR12-BTR13**
N	537	458	570	572	710	419
Mean	**86.57**	**102.27**	**92.33**	**85.60**	**69.52**	**79.27**
Std dev.	32.43	32.98	30.15	26.13	19.61	26.69
Minimum	28	36.5	12.5	22	23.5	36
Q1	61.5	78.5	70	66.375	54	58.75
Median	83	95.5	91	81	68.5	76.5
Q3	109.5	120.75	111.875	101	82	96.75
Maximum	189	197.5	181	162	126.5	157.5
IQR	48	42.25	41.875	34.625	28	38
Skewness	.447	.701	.367	.659	.437	.678
Kurtosis	-.356	.057	-.348	-.08	-.143	-.044
95% Confidence Interval for Mean:						
Lower Bound	83.82	99.24	89.85	83.74	68.07	76.71
Upper Bound	89.31	105.29	94.81	84.48	70.96	81.83

When the results presented in the table are examined, it is seen that the results with both test methods have a significant p value for all BTR pairs. At the same time, the values that should be small according to the Kolmogorov-Smirnov test are small and the values that should be close to 1 are observed to be 0.96 and above for the Shapiro-Wilk test. In the light of these results, there is not any strong evidence to reject the null hypothesis of normality for each BTR pair link travel times. Examination of SPSS normality check and frequency distributions revealed similar results, however, it is important to examine the skewness and kurtosis values, which give important clues about the distribution of the data mentioned in the literature.

Table 4.12 Normality Check of Daily Corridor Cross-BTR (between BTR 3309 to 3313) Travel Time Investigation

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistics	df	Sig.	Statistics	df	Sig.
BTR07-BTR08	.068	537	< .001	.976	537	< .001
BTR08-BTR09	.098	458	< .001	.959	458	< .001
BTR09-BTR10	.052	570	.001	.983	570	< .001
BTR10-BTR11	.080	572	< .001	.962	572	< .001
BTR11-BTR12	.044	710	.002	.983	710	< .001
BTR12-BTR13	.077	419	< .001	.958	419	< .001

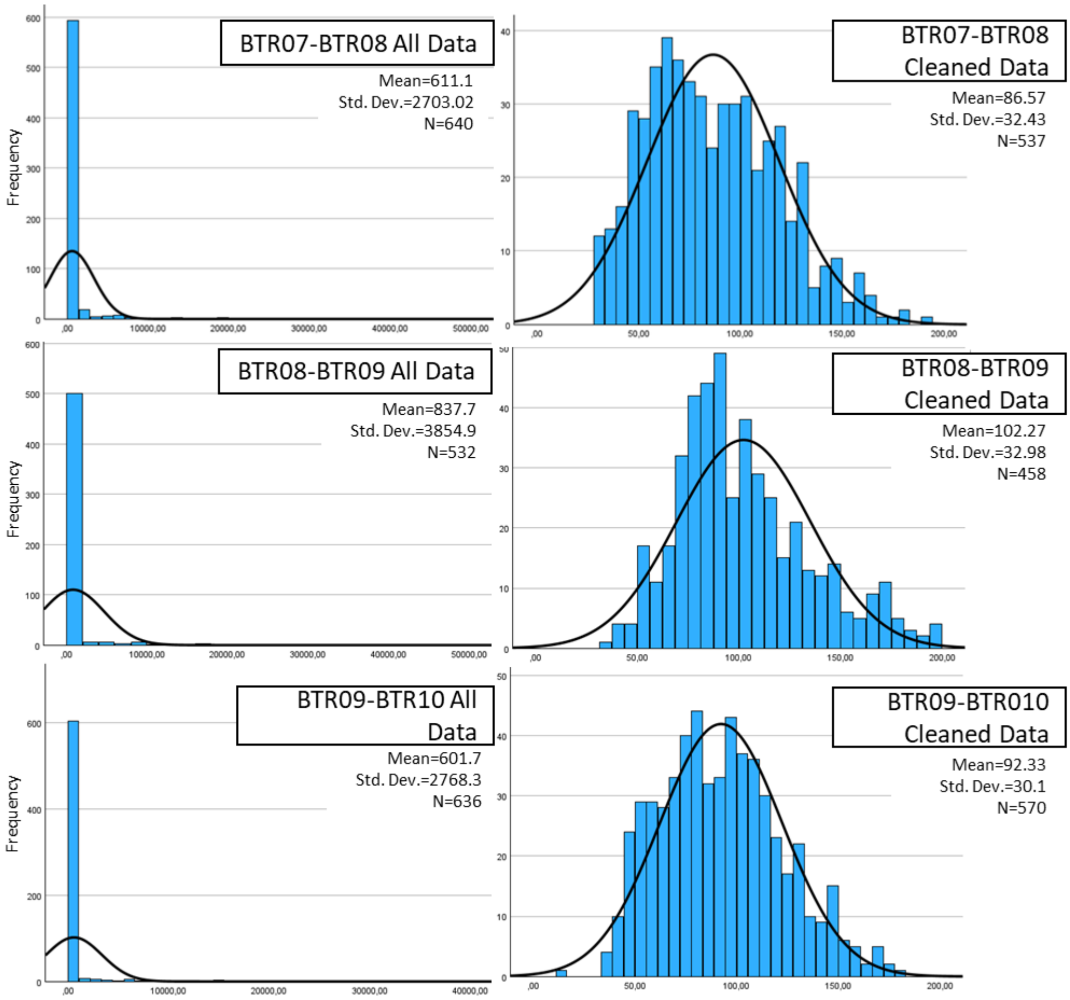


Figure 4.12. Daily Distribution of Corridor Cross-BTR (between BTR07 to BTR10) Travel Time Investigation for all data and cleaned data respectively.

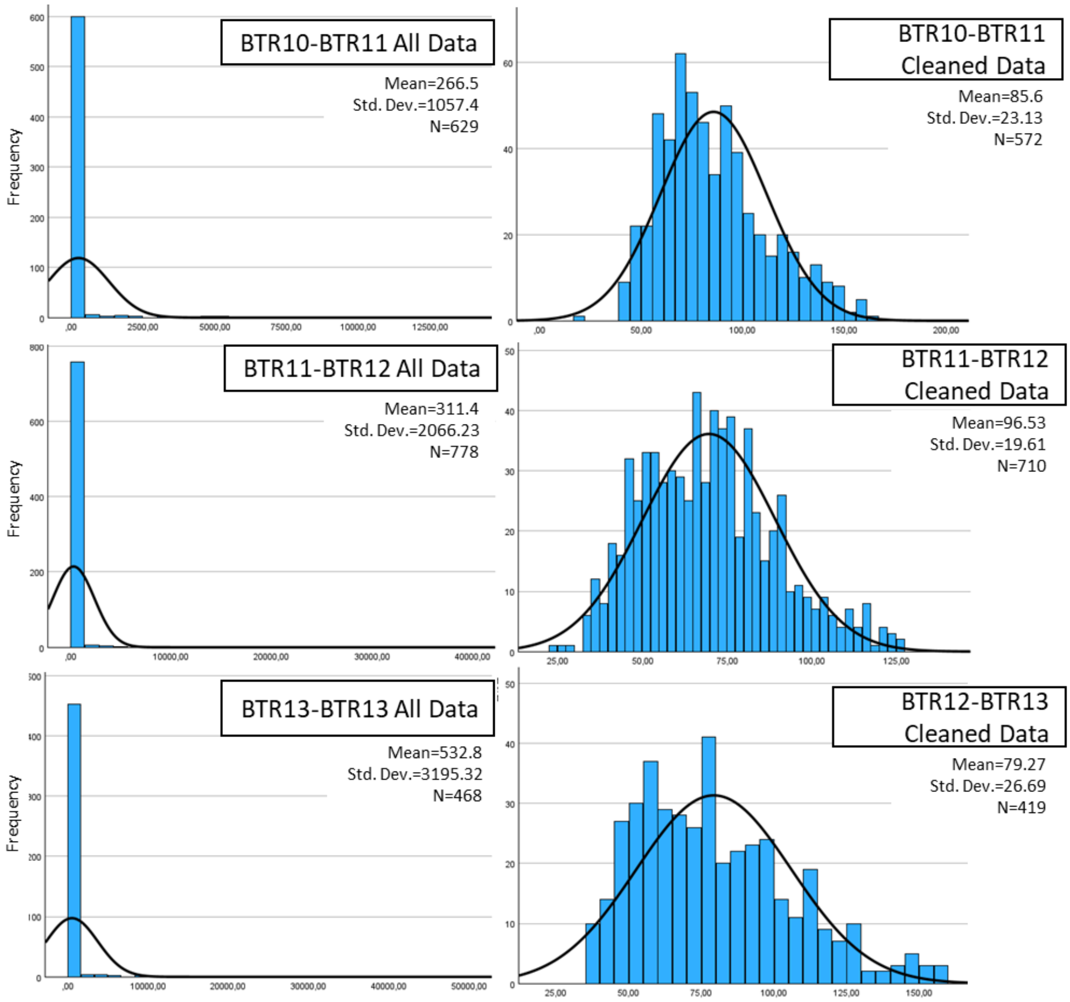


Figure 4.13. Daily Distribution of Corridor Cross-BTR (between BTR10 to BTR13) Travel Time Investigation for all data and cleaned data respectively.

When compared to a normal distribution, the data distribution's asymmetry is measured by skewness, and its heaviness in the tails is measured by kurtosis. According to the results summarized in Table 4.11, skewness values range between 0.367 to 0.701, all of which suggests slight to moderate level right-skewed distribution, meaning the tail on the right side of the distribution is longer or more pronounced. The tails of a distribution with a kurtosis value of -0.356 are somewhat platykurtic, meaning they are lighter or heavier than those of a normal distribution. A virtually mesokurtic distribution with a kurtosis value of 0.057 indicates that the tails are substantially identical to those of a normal distribution. In conclusion, there isn't substantial evidence to imply that the data significantly deviates from a normal distribution based on the findings of the Kolmogorov-Smirnov and Shapiro-Wilk tests, considering the skewness and kurtosis values, or the results of the skewness and kurtosis values. It is important to keep in mind that the skewness and kurtosis values show minor deviations from complete normality, pointing to a distribution that is somehow right skewed and platykurtic.

4.7 Urban Arterial Corridor Travel Time Estimation

Besides this link-based estimation, the 221 MAC_IDs are captured that consecutively passed through the 7 BTR locations along the corridor is analyzed to compare the possible actual movement on the selected corridor. The descriptive statistics of this captured MAC_IDs are summarized in Table 4.13. At the same time, the determined important statistical parameters for MAC_IDs that are captured consecutively along the corridor and determined within the scope of link-based estimation on the corridor are shown in Figure 4.14.

Table 4.13. Descriptive Statistics of Captured Moving Vehicles Travel Time of the Study Corridor (from BTR07 to BTR13)

	BTR07-BTR08	**BTR08-BTR09**	**BTR09-BTR10**	**BTR10-BTR11**	**BTR11-BTR12**	**BTR12-BTR13**
N	221	221	221	221	221	221
Mean	**69.38**	**81.67**	**72.88**	**75.61**	**59.15**	**75.93**
Std dev.	31.50	35.02	27.57	26.94	19.73	25.40
Minimum	15	32	29	20	28	30
Q1	47	54	50	54	44	55
Median	63	75	70	70	55	73
Q3	86	100	89	94	72	93
Maximum	195	272	161	190	129	155
IQR	39	46	39	40	28	38

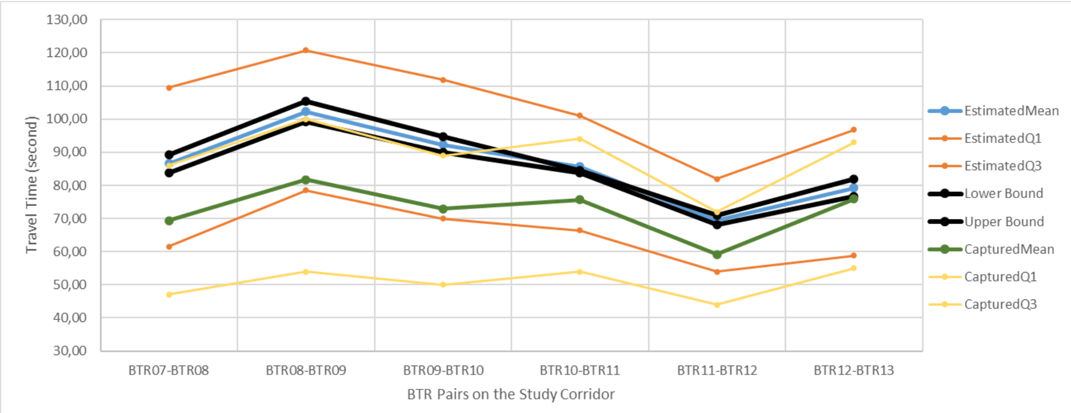


Figure 4.14. Important statistical parameters for the pairwise used in link-based estimations and captured on the corridor consecutively MAC_IDs.

The estimated link-based travel times are within a very narrow limit because of which the consecutively captured MAC_ID travel times located below this interval but within the Q1-Q3 interval. At the same time, the time-dependent corridor mobility with consecutively captured MAC_ID during the day is examined in Figure 4.15 which was prepared by using the first captured travel times obtained from these reading. As reflected in daily cleaned travel time change graphs for different time intervals, this graph also did not reveal any daily differences which was supported by low congestion levels in the study corridor and even in Mersin.

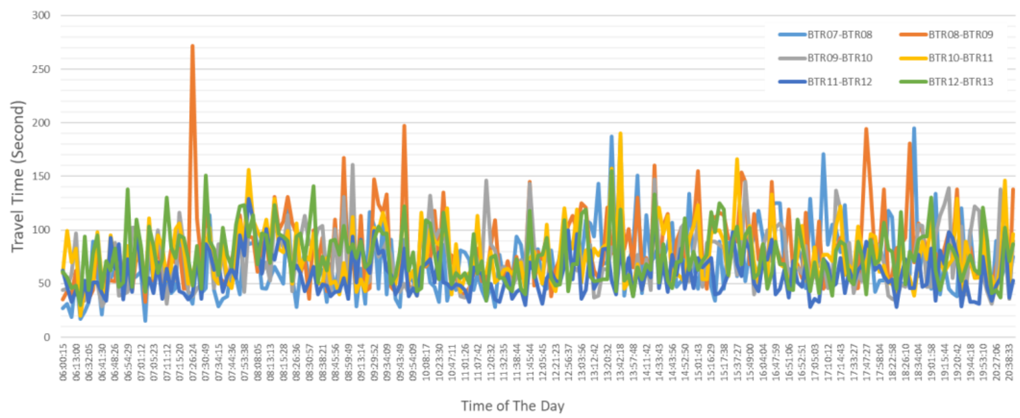


Figure 4.15. Distribution of Corridor Travel Times based from Moving Vehicles (BTR07 to BTR13).

Considering that the link travel distributions fit the normal distribution for each BTR pair located on the corridor, and no changes were observed in the duration of sleep during the day, the time dependent perspective was not included in this study. The urban arterial travel time estimation process for the study day in study is summarized with the formulations below, with the view that the sum of the data that fits the normal distribution also fits the normal distribution in the Methodology section. The estimated travel time interval for the corridor was calculated at the 95% confidence interval as (401.60; 629.49). Total average travel time for MAC_IDs captured by 7 devices consecutively is 434.62 seconds, which fits within the estimated range.

$t_{\text{BTR07-BTR13}} = t_1 + t_2 + t_3 + t_4 + t_5 + t_6$

$t_{\text{BTR07-BTR09}} = t_1 \sim N\ (86.57, 32.43)$ $t_{\text{BTR08-BTR09}} = t_2 \sim N\ (102.27, 32.98)$

$t_{\text{BTR09-BTR10}} = t_1 \sim N\ (92.33, 30.15)$ $t_{\text{BTR10-BTR11}} = t_2 \sim N\ (85.60, 26.13)$

$t_{\text{BTR10-BTR11}} = t_1 \sim N\ (69.52, 19.61)$ $t_{\text{BTR10-BTR11}} = t_2 \sim N\ (79.27, 26.69)$

$\overline{t_{\text{BTR07}-\text{BTR13}}} = \overline{t_1} + \overline{t_2} + \overline{t_3} + \overline{t_4} + \overline{t_5} + \overline{t_6} = 515.55$

$s^2_{\text{BTR07}-\text{BTR13}} = s_1^2 + s_2^2 + s_3^2 + s_4^2 + s_5^2 + s_6^2 = 4828.07$

$$t_{\text{BTR07-BTR13}} \sim N\ (\overline{t_{\text{BTR8}-\text{BTR13}}}, s_{\text{BTR8-BTR13}}) \rightarrow N\ (515.55, 69.48)$$

$t_{\propto,lower} \leq t_{\text{BTR07}-\text{BTR13}} \leq t_{\propto,upper}$

$\overline{t_{\text{BTR7}-\text{BTR13}}} - z_{\propto} * s_{\text{BTR7}-\text{BTR13}} \leq t_{\text{BTR7}-\text{BTR13}} \leq \overline{t_{\text{BTR7}-\text{BTR13}}} + z_{\propto} * s_{\text{BTR7}-\text{BTR13}}$

$z_{\propto} = 1.96 \rightarrow$ 95% Confidence Interval

$$515.55 - 1.96 * 69.48 \leq t_{\text{BTR8}-\text{BTR13}} \leq 515.55 + 1.96 * 69.48$$

95% CI for mean: $379.37 \leq t_{\text{BTR8}-\text{BTR13}} \leq 651.73$

CHAPTER 5

INQUIRY INTERVAL ANALYSIS FOR RESCAN TIME

The analysis began with the data collection period, and preliminary results revealed that there is a second peak in the rescan distribution when the inquiry interval is 48 seconds. Given this outstanding result and the literature review, it was decided to set up a test day for the inquiry interval. The following three periods are used for single BTR analysis:

- Before Analysis with when Tinq= 48 sec
- Test Day Analysis with when Tinq=10/20/30/48 sec
- After Analysis with when Tinq= 10 sec

5.1 Before Period Analysis

Rescan time frequency graphs for the before period are investigated and depicted in Figure 5.1 for selected 3-day. The majority of calculated rescan times produce a peak between 15 and 18 seconds, but there is also a second peak around 60 seconds due to the inquiry interval (Tinq=48 second*1.28 inquiry scan interval= 61.44 second). Inquiry interval was arranged to scan capture area in every 48 seconds according to default feature of the BT device, which created a second peak in the 60 second region. This second peak value has revealed quite high and misleading results to set a threshold when calculating the stay time in the BTR coverage area.

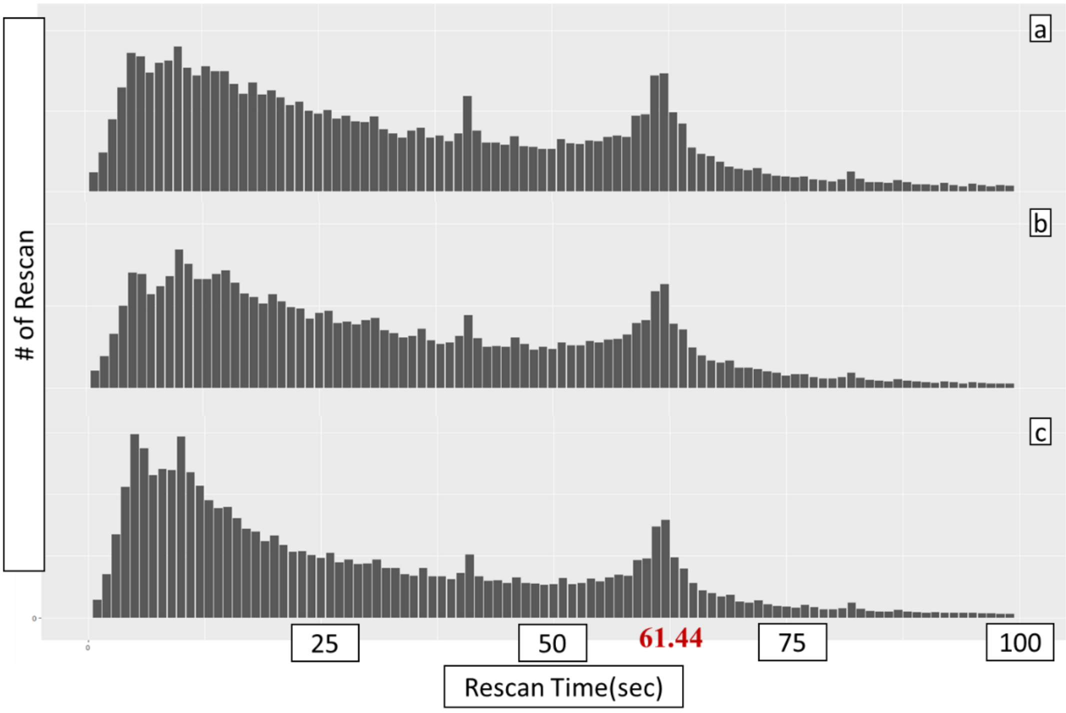


Figure 5.1. Rescan Time Frequency of a)15 May, b)22 May and c) 28 May when Inquiry Interval is 48 second

5.2 Test Day Analysis

In the test day, 3 different inquiry interval was determined, and applied to Mersin city network for whole day (29 May 2019, Wednesday) but for different time intervals. Time intervals defined to capture peak hour differentiation as in Table 5.1. Time code information is specified in the table as “x_y where x represents the hour and y represents minute-based interval ranges from 0 to 3 (0=00:00-00:15; 1=00:15-00:30; 2=00:30-00:45; 3=00:45 -01:00). During the morning peak, which was predicted to be between 7:00 and 9:00 a.m., data was collected for at least 30 minutes for three possible inquiry intervals (10, 20 and 30 seconds). However, because the device settings were manually controlled, a 5-minute extension occurred in the interval predicted as 07:00-08:00, and this continued by reflecting increasingly on other intervals. If more than one inquiry interval is used in the data collection intervals, these intervals are summarized in the table as the "Mix Inquiry Time".

Table 5.1. Test Day Inquiry Interval Protocol

Data Collection Periods	**T_{inq}**	**Time Codes**	**Mix Inquiry Time***
06:00 - 06:30	10 sec	6_0; 6_1	
06:30 - 07:00	20 sec	6_2; 6_3	
07:00 - 08:05	30 sec	7 (full)	
08:05 - 08:35	20 sec	8_0; 8_1	**8_0 (**T_{inq} =30 sec & 20 sec)
08:35 - 10:12	10 sec	8_2_8_3; 9(full); 10_0	**8_2 (**T_{inq} =20 sec & 10 sec) **10_0 (**T_{inq} =10 sec & 20 sec)
10:12 - 11:12	20 sec	10_1; 10_2; 10_3; 11_0	**11_0 (**T_{inq} =20 sec & 30 sec)
11:12 - 12:30	30 sec	11_1; 11_2; 11_3; 12_0; 12_1	
12:30 – 13:00	20 sec	12_2; 12_3	
13:00 – 14:30	10 sec	13(full); 14_0; 14_1	
14:30 – 15:30	20 sec	14_2; 14_3; 15_0; 15_1	
15:30 – 17:30	30 sec	15_2; 15_3; 16(full) 17_0; 17_1	
17:30 – 18:30	20 sec	17_2; 17_3; 18_0; 18_1	
18:30 – 19: 30	10 sec	18_2; 18_3; 19_0; 19_1	
19:30 – 21:00	48 sec	19_2; 19_3; 20(full)	

* Mix inquiry time is shown with both inquiry color coding in the bar diagrams.

The results revealed that each inquiry interval produces a peak that is 1.28 times greater than the chosen inquiry interval (see Figure 5.2). Figure 5.2 depicts the frequency of each rescan time, with peaks at approximately 13 seconds for 10 second inquiry intervals, 26 seconds for 20 second inquiry intervals, and 38 seconds for 30 second inquiry intervals. Figure also depicts box plot graphs for the inquiry intervals, demonstrating that the lower the inquiry interval, the smaller the variation in the first and third quartiles.

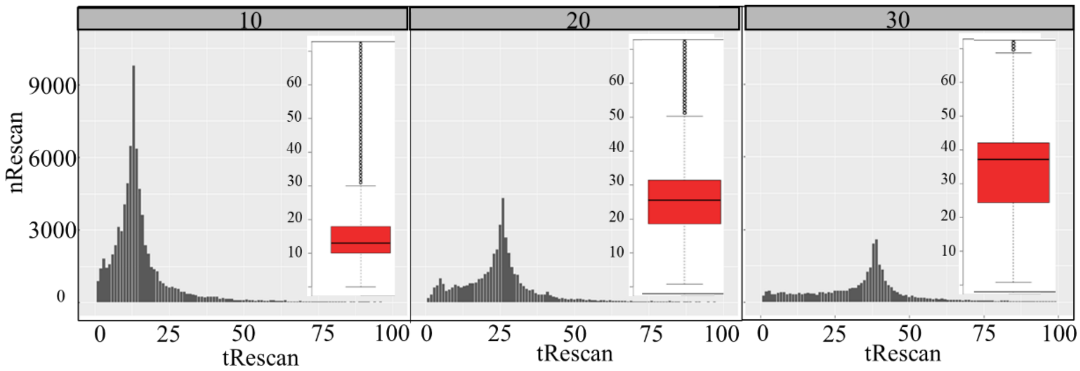


Figure 5.2. Rescan time distribution and box plot graphs for Inquiry intervals.

For an urban area, any device that travels with a speed above 50 km/h can pass through the capture zone before BTR scan for 25 % of the time, which can decrease the success of analysis when the inquiry interval is equal to 48 seconds. In addition, although there is a significant peak value in the inquiry interval values determined as 20 and 30 seconds; it is seen that 50% of the data is distributed over a wide range, as can be seen in the box plot graph, especially for 30 seconds.

5.3 Selection of Inquiry Interval

Real time analysis of a data process is directly affected by data size and storage. Decreasing inquiry interval, means to much scan of any MAC_ID and this creates an increase in data storage. The increased data size, on the other hand, must be tolerated because the increased readings will provide a significant advantage in terms of further investigation of vehicle behavior. For these purposes, the increase of data size should be investigated for each inquiry interval and summarized in Table 4.3. The number of readings in the table increased for all previous days and the test day, but there was no significant difference in the number of unique MAC. However, because the possible inquiry interval values change throughout the day, a numerical comparison of the increase in data is useful for providing a general idea, even if it does not reflect the actual case. So, to better understand the differentiation, the number of rescans and unique MAC_IDs are graphically depicted in 15-minute intervals comparing before (15 May) and the test day (29 May) in Figure 5.3.

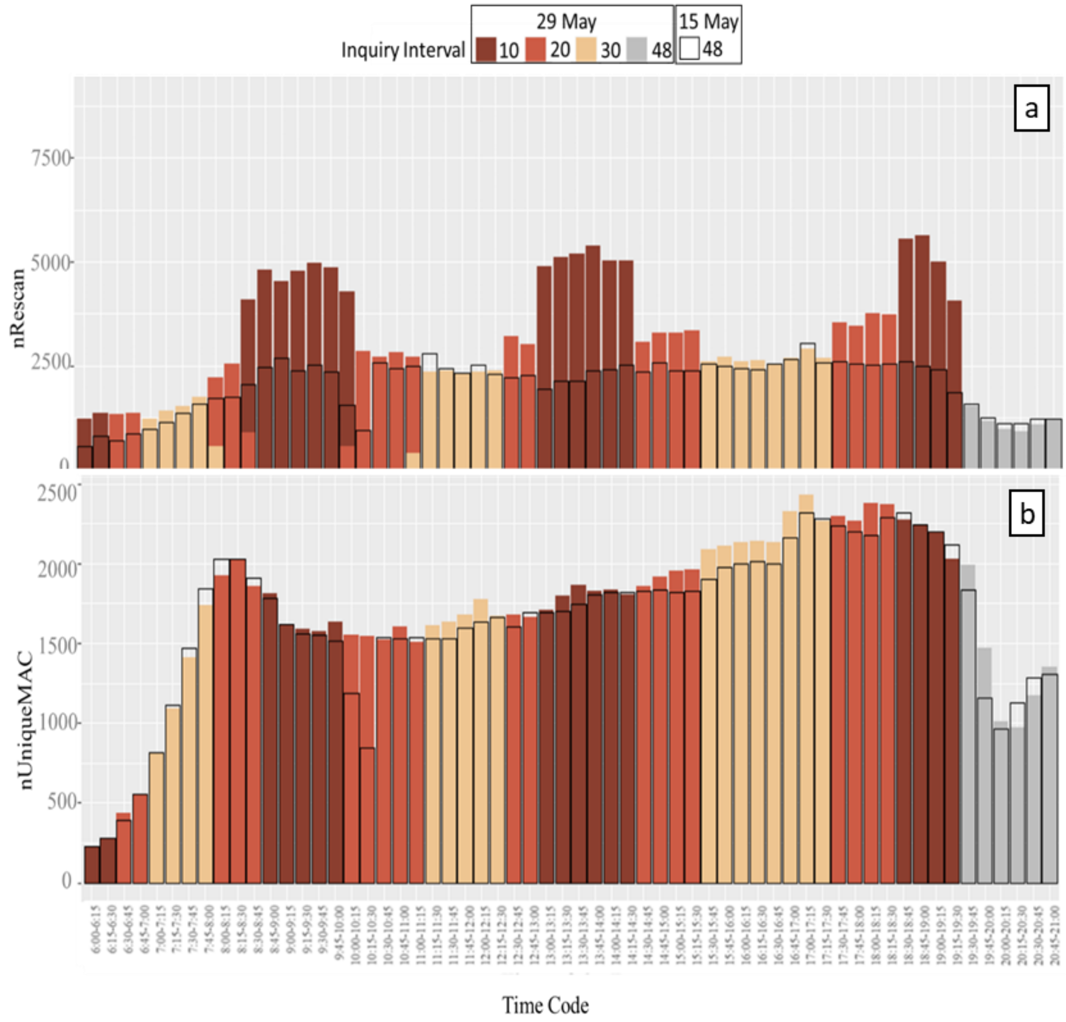


Figure 5.3. 15-min Interval number of a) Rescan and b) Unique MAC differences between 29 May and 15 May

In the graphics, color bars represent inquiry intervals on test days, while ghost bars reflect the comparison day. Results revealed that for 15-minute intervals throughout the day, number of readings nearly doubles when the inquiry interval is 10 seconds when compared to 48 seconds for same interval of the selected days. However, while there is no significant increase in the number of readings when the inquiry interval is set for 30 seconds, this increase level is relatively low when 20 seconds is set. Despite a remarkable increase in the number of Readings in the graph, no significant results could be drawn because of the change in the inquiry interval in the number of unique MAC_IDs captured. However, because the rescan frequency graphs in Figure 5.2 show a spread rather than a high frequency peak value for 20 and 30 seconds, it is important to examine the increase that the change will cause in the data and data analysis process. For this purpose, device status percentages are calculated and summarized in the Table 5.2, with the analysis performed by discarding the number of readings collected from potential stationary MAC_IDs in the data using the method described within the scope of the methodology.

According to the findings, 15 to 20% of the readings are from potential stationary vehicles, and this percentage increases in lockstep with the total number of readings. Removing this data group in real-time/ time dependent travel time estimation analyzes will tolerate approximately half of the increase in the data collected. Considering the analysis and evaluations, the value of the inquiry interval is set at 10 seconds. As a result of this decision, it was decided to monitor real increases in the data and analysis process. The increased data should also be tested for the experimentally established travel estimation algorithm studies for Mersin, as well as its effect on real-time analysis. The first solution to the potential increase is to reduce the MAC_ID readings by removing the Stationary group from the data.

Table 5.2. Distribution of MAC_IDs according to device status

Date	nReading	NonStat NonVeh (%)	NonStat Veh (%)	StatNon Veh (%)	Stat Veh (%)
15.May.19	291349	42.64	41.62	14.86	0.87
21.May.19	302987	39.96	42.91	16.95	0.18
22.May.19	247851	39.81	41.57	18.53	0.10
28.May.19	317330	38.09	41.22	20,69	0.00
29.May.19	360036	40,01	40,53	19.24	0.22
2.Tem.19	459117	39.36	36.99	23.38	0.26
3.Tem.19	443666	40.33	36.98	22.64	0.05
9.Tem.19	427923	42.58	38.34	18.74	0.35
10.Tem.19	428998	43.00	37.20	19.74	0.06
10.Eyl.19	464276	45.87	38.70	15.28	0.15
11.Eyl.19	462894	44.36	37.66	17.76	0.22
17.Eyl.19	467201	44.68	38.31	16.21	0.80
18.Eyl.19	466566	45.71	38.07	15.87	0.35
5.Kas.19	479680	46.60	36.35	16.77	0.28
6.Kas.19	587321	37.86	29.34	15.46	0.31
12.Kas.19	477878	47.17	36.49	16.21	0.13
13.Kas.19	334918	47.39	35.27	16.91	0.43
6.Eyl.22	670753	55.34	22.91	21.60	0.15
7.Eyl.22	692717	56.22	22.52	21.05	0.20
13.Eyl.22	666083	57.70	22.90	19.32	0.08
14.Eyl.22	668534	57.62	22.56	19.78	0.05

5.4 After Period Analysis

Rescan time frequency graphs are calculated for the after period of 3-day data represented in Figure 5.4. All graphs show a single and exact peak around 13 seconds, which reduces the error rate of BTR rescan. In addition, it is noteworthy that the rescan time value over 50 seconds is quite low. It can be concluded that the selected 10-second inquiry interval value is very effective in terms of rescan threshold for stay time calculation.

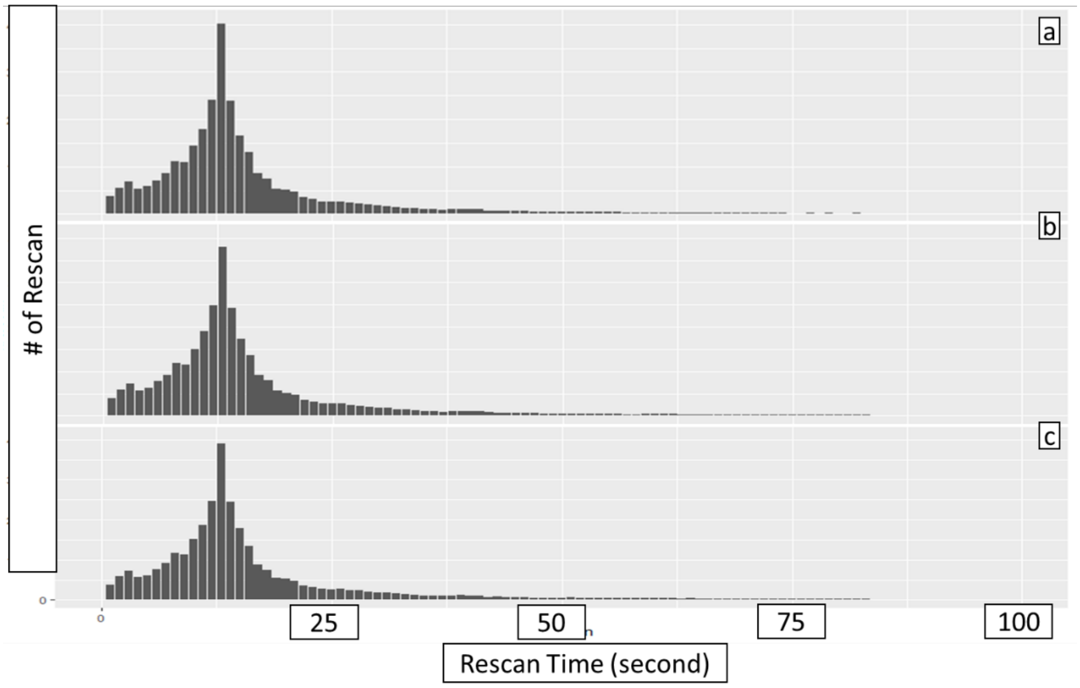


Figure 5.4. Rescan Time Frequency of a)3 July, b)10 July and c)18 September when Inquiry Interval is 10 second

The actual change in BTR data collection capacity can be observed by comparing the data collected with 48 seconds before the test day with the data collected with 10 seconds in the after studies. For this purpose, the graphs of the number of readings and the unique MAC_ID number captured between May 15 and July 2 are prepared and shown in the Figure 5.5. The number of readings nearly doubled during the 10-second inquiry interval period on test day versus before study comparison. However, slightly different results are noticeable for the number of unique MAC_ID captured. There is a remarkably high increase in the number of unique MAC_IDs especially in time intervals where the number of readings increased significantly.

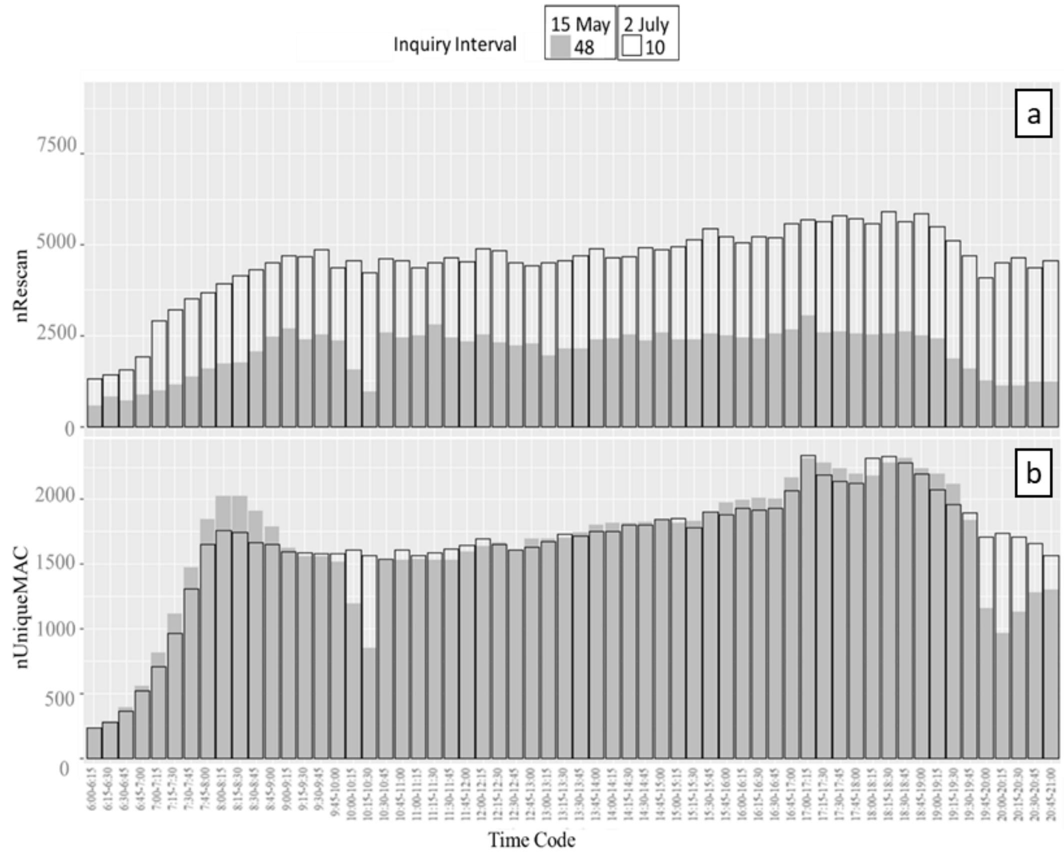


Figure 5.5. 15-min Interval number of a) Rescan and b) Unique MAC differences between 15 May and 3 July

CHAPTER 6

CONCLUSIONS AND FURTHER RECOMMENDATIONS

6.1 General Overview and Conclusion

For developing countries, traffic data collection technologies such as Bluetooth hold a special importance as it requires a little investment cost. But the quality of the traffic estimations is crucial and must be addressed to provide reliable traffic information. The technology of BTR devices directly affects the quality of data and reliability and accuracy. To rely on collected data and to minimize system related errors, changeable parameters such as inquiry interval should be investigated in detail. However, every change in BTR reading process leads to differentiations in data and storage.

BT data, which is notable for allowing the tracking of a specific vehicle movement in a corridor or network, has many uncertainties and complexities. There is no precise information in the data about the MAC_ID obtained; it could be a device belonging to a vehicle, carried in the vehicle, carried by a pedestrian, or in the coverage area in a completely steady state. Although the fixed location information of BTR devices and the distance information between each other are clear, the exact location of the MAC_ID with respect to reading time is not available in the collected data. At the same time, because the BTR capture zone varies depending on environmental conditions, other device/vehicle effects in the region, and even weather conditions, whether the MAC ID is captured by the BTR and how many times it is captured may vary. Even for a BTR couple with an intersecting capture zone, a stationary device at the intersection may present a situation as it is moving. Because the use of this system in an urban network provides many movement alternatives with a limited number of BTR devices, some unexpected situations can be observed such as a long travel time for a BTR pair relatively close to each other.

These longer unexpected travel times may reflect a MAC ID, which indicates that the vehicle has traveled and passed a route that is not within the scope of the BTR network, or it may represent a pedestrian moving entirely on the route.

Considering all these circumstances, preliminary analyses were conducted for the BTR inquiry interval value to facilitate the analysis process, and this value was changed to 10 seconds. Although this change results in a significant increase in data, it will allow for a more in-depth examination of the device's behavior as the reading frequency increases. Thus, when calculating the stay time value in the BTR capture zone, a Rescan value of 50 seconds will be considered as a threshold, and it will be concluded that this device does not stay in range continuously and leaves the system for a while. Although this result does not always indicate that the device has left the system, it is always better to be on the safe side when analyzing data.

Movement between BTRs is calculated by the difference between the last reading on the first device and the first reading on the second device as cross-BTR time (L2F); it is corrected by integrating the stay time value in each BTR capture zone. While examining the movement between BTRs, potential stationary MAC IDs were removed from the data to the size of the data storage. Each BTR travel time value obtained was statistically analyzed by grouping it into 15-minute, 30-minute, 1-hour and 2-hour intervals throughout the day (from 6:00 to 21:00). When the distributions of these travel times are examined by group, it was concluded that the data requires careful outlier detection. The basic statistical outlier detection methods were first applied to the data for different time intervals; however due to the numerous limitations such as extreme travel time values or small sample sizes within the selected intervals, is reduced the success of these methods. For this reason, a two-stage-outlier detection method is proposed to determine "possible moving vehicle". This method depends on the elimination of slow-movement vehicles in the first stage, continuing with the commonly accepted statistical method (IQR method). Accuracy of average "possible moving vehicle" speed values are later investigated with respect to FCD data for the same BTR pair on the same days. When the results are compared, although FCD data has higher speeds than BT data, the differences are closer to each other during peak hour traffic. In addition, with the increase in the distance between BTR devices, it is observed that BT speeds are closer to FCD speeds.

For time dependent urban arterial travel time estimation, stay time threshold is applied and M2M cross-BTR time are cleaned for estimations. Average link travel time and a possible interval were determined by using possible moving vehicle travel time values at 2-hour intervals for each BTR pair on the corridor. These values were compared with the travel times obtained from the MAC_IDs that consecutively passed through the 7 BTR point along the corridor. This study shows that major patterns in travel time estimation can be captured using Bluetooth data on urban corridors. For the corridor selected within the scope of the study, the travel time within the 95% confidence interval was determined as (379.37-651.73) seconds. However, for corridors such as urban corridors, where significant differences in travel times are expected to occur with instantaneous changes, this range is estimated as a high range. Despite this limitation, it is possible to make a meaningful estimation for the corridor at such a serious confidence interval and this resource is very cheap and effective.

6.2 Recommendations for Future Studies

With the proposed data cleaning and analysis method in the thesis, limitations in the data collection process and complexity in the collected data have been solved to a great extent. However, this defined inquiry interval is specific to BTR device characteristic which needs to be controlled for every BTR device. It is very important to ensure standardization and consistency in the application of this system, which provides cheap and effective data especially in many cities are donating with BTR devices. The proposed outlier detection method can be applied to any kind or BTR network, due the first threshold depends on the speed which can be calculated for a pair with any distance. As the limitations and possible solutions of this method, which is proposed for speed estimation in urban corridors, are examined very well within the scope of this study, it will be very easy to integrate it into real-time data analysis and estimation processes. The use of BT data will provide significant contributions and conveniences in travel forecasts, especially in access-controlled road sections such as tunnels.

There are still other issues to discuss for Bluetooth based data use in traffic such as archiving and reliability in real-time estimation, which requires special data filtering and analytics methods. In fact, it is foreseen that effective studies can be carried out on passenger waiting times at airports, bus stops or train stations with MAC_IDs, which are not used in the travel time estimation process due considered as stationary and extracted from the data set within the scope of this study.

REFERENCES

Abedi, N., Bhaskar, A., Chung, E. and Miska, M. (2015). “Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses.” *Transportation. Research Part C,* 60, 124-141. (10.1016/j.trc.2015.08.010)

Ahmed, H., El-Darieby, M., Morgan, Y. and Abdulhai, B. (2008) "A Wireless Mesh Network-Based Platform for ITS," VTC Spring 2008 - IEEE Vehicular Technology Conference, Marina Bay, Singapore, pp. 3047-3051, doi: 10.1109/VETECS.2008.329.

Aliari, Y. and Haghani, A. (2012). “Bluetooth Sensor Data and Ground Truth Testing of Reported Travel Times.” *Transportation Research Record,* 2308, 167-172. (10.3141/2308-18).

Altintasi, O., Balci, M.E. and Ozen, M. (2022). “Urban arterial travel time evaluation using Bluetooth Data.” *Politeknik Dergisi*, 25(2), 767-773. (10.2339/politeknik.784652).

Antoniou, C., Balakrishna, R. and Koutsopoulos, H. (2010) “Exploiting emerging data collection technologies for dynamic traffic management applications.” *Proceedings of World Conference on Transport Research (WCTR).*

Araghi, B.N., Pedersen, K.S., Christensen, L.T., Krishnan, R. and Lahrmann, H. (2012). “Accuracy of travel time estimation using Bluetooth technology: Case study limfjord tunnel Aalborg.” *Proc., 19th Intelligent Transport Systems World Congress,* Vienna, Austria.

Araghi, B.N., Hammershøj Olesen, J., Krishnan, R., Tørholm Christensen, L. and Lahrmann, H. (2015). “Reliability of Bluetooth Technology for Travel Time Estimation.” Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 19(3), 240-255. (10.1080/15472450.2013.856727)

Barceló, J., Montero, L., Marqués, L. and Carmona, C. (2010a). “Travel Time Forecasting and Dynamic Origin–Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring." *Transportation Research Record*, 2175, 19-27. (10.3141/2175-03)

Barceló, J., Montero, L., Marqués L. and Carmona, C. (2010b). "A Kalman-Filter Approach for Dynamic OD Estimation in Corridors Based On Bluetooth and Wi-Fi Data Collection." Proc., 12th World Conference on Transport Research, Lisbon, Portugal.

Barrales-Guadarrama, R., Rodríguez-Rodríguez, M. E., Barrales-Guadarrama, V. R. and Mocholí-Salcedo, A. (2010) "A Bluetooth Development Platform for Wireless Sensors," IEEE Electronics, Robotics and Automotive Mechanics Conference, Cuernavaca, Mexico, 2010, pp. 213-219, doi: 10.1109/CERMA.2010.34.

Blogg, M., Semler, C., Hingorani, M. and Troutbeck, R. (2010). "Travel Time and Origin-Destination Data Collection using Bluetooth MAC Address Readers." *Proc., 33rd Australasian Transport Research Forum,* Canberra, Australia.

Bullock, D. M., Haseman, R., Wasson, J.S. and Spitler, R. (2010). "Automated Measurement of Wait Times at Airport Security- Deployment at Indianapolis International Airport, Indiana." *Transportation Research Record*, 2177, 60-68. (10.3141/2177-08)

Carpenter, C., Fowler, M. and Adler, T. J. (2012). "Generating Route-Specific Origin–Destination Tables Using Bluetooth Technology." *Transportation Research Record*, 2308, 96-102. (10.3141/2308-10)

Carrese, S., Cipriani, E., Crisalli, U., Gemma, A.and Mannini, L. (2021) "Bluetooth Traffic Data for Urban Travel Time Forecast", Transportation Research Procedia, 52, pp. 236-243 (10.1016/j.trpro.2021.01.027).

Chen, Y., and Chen, L. (2005) "Using Bluetooth wireless technology in vehicles," IEEE International Conference on Vehicular Electronics and Safety, 2005., Shaanxi, China, 2005, pp. 344-347, doi: 10.1109/ICVES.2005.1563670.

Chien, C.F., Chen, H.T. and Lin, C.Y. (2020). “A Low-Cost On-Street Parking Management System Based on Bluetooth Beacons.” *Sensors,* 20. 4559. (10.3390/s20164559)

Chitturi, M.V, Shaw, J.W, Campbell IV, J.R. and Noyce, D.A. (2014). “Validation of Origin-Destination Data from Bluetooth Reidentification and Aerial Observation.” *Transportation Research Record*, 2430, 116-123. (10.3141/2430-12).

Cipriani, E., Gemma, A., Mannini, L., Carrese, S., and Crisalli, U. (2021) “Traffic demand estimation using path information from Bluetooth data”, Transportation Research Part C: Emerging Technologies, 133. (10.1016/j.trc.2021.103443).

Civcik, L. and Kocak, S. (2020). “Travel Time Prediction with Bluetooth Sensor Data in Intelligent Traffic System (ITS).” *European Journal of Science and Technology,* Special Issue, 522-529. (10.31590/ejoset.829619)

Cotten, D., Codjoe, J. and Loker, M. (2020). “Evaluating Advancements in Bluetooth Technology for Travel Time and Segment Speed Studies.” *Transportation Research Record*, 2674(4), 193-204. (10.1177/0361198120911931)

Crawford, F., Watling, D.P. and Connors, R.D. (2018). “Identifying road user classes based on repeated trip behavior using Bluetooth data.” *Transportation Research Part A*, 113. (10.1016/j.tra.2018.03.027)

Day, C. M., Haseman, R, Premachandra, H., Brennan, T. M., Wasson, J. S., Sturdevant, J.R. and Bullock, D. M. (2010). “Evaluation of Arterial Signal Coordination- Methodologies for Visualizing High-Resolution Event Data and Measuring Travel Time.” *Transportation Research Record*, 2192, 37-49. (10.3141/2192-04)

Diaz, J.J.V., Gonzalez, A.B.R and Wilby, M. R. (2015) "Bluetooth traffic monitoring systems for travel time estimation of freeways," *IEEE Transactions* *on Intelligent Transportation Systems*, 17(1), 123-132. (10.1109/TITS.2015.2459013)

Dunlap, M., Li, Z., Henrickson, K. and Wang, Y. (2016). "Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit." *Transportation Research Record*, 2595, 11-17. (10.3141/2595-02)

Erkan, İ.and Hastemoglu, H. (2016). "Bluetooth as a traffic sensor for stream travel time estimation under Bogazici Bosporus conditions in Turkey." *Journal of Modern Transportation*, 24(3), 207-214. (10.1007/s40534-016-0101-y)

Ferreira, M.C., Dias T.G. and e Cunha, J.F. (2020). "Is Bluetooth Low Energy feasible for mobile ticketing in urban passenger transport?" *Transportation Research Interdisciplinary Perspectives*, 5, 100120. (10.1016/j.trip.2020.100120)

Garrido-Valenzuela, F., Raveau, S. and Herrera, J. C. (2020) "Bayesian Route Choice Inference to Address Missed Bluetooth Detection," *IEEE Transactions on Intelligent Transportation Systems*, 23(3), 1173-1187. (10.1109/TITS.2020.3028804)

Ghavidel, M., Khademi, N., Samani, E. B., and Kieu, L. M. (2021). "A Random Effects Model for Travel-Time Variability Analysis Using Wi-Fi and Bluetooth Data." Journal of Transportation Engineering, Part A: Systems, 148(2). (10.1061/JTEPBS.000062)

Hainen, A. M., Wasson, J.S., Hubbard, S. M. L., Remias, S. M., Farnsworth, G. D. and Bullock, D. M. (2011). "Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles." *Transportation Research Record*, 2256, 43-50. (10.3141/2256-06)

Haghani, A., Hamedi, M., Sadabadi, K. F., Young, S., and Tarnoff, P. (2010). "Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors". *Transportation Research Record*, 2160(1), 60–68. https://doi.org/10.3141/2160-07

Haseman, R. J., Wasson, J. S. and Bullock, D. M. (2010). "Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking." Transportation Research Record, 2169, 40-53. (10.3141/2169-05)

Jedwanna, K., and Boonsiripant, S. (2022). "Evaluation of Bluetooth Detectors in Travel Time Estimation." *Sustainability,* 14, 4591.

Jiang, J.-R., Lin, B.-R., & Tseng, Y.-C. (2004). "Analysis of Bluetooth device discovery and some speedup mechanisms." *Journal of the Institute of Electrical Engineering, 11*(4), 301-310.

Karatsoli, M., Margreiter, M. and Spangler, M. (2017). "Bluetooth-based travel times for automatic incident detection – A systematic description of the characteristics for traffic management purposes." *Transportation Research Procedia*, 24, 204-211. (10.1016/j.trpro.2017.05.109)

Koçak, S. (2021). "Akıllı Trafik Sistemlerinde (ITS), Bluetooth Sensor Verileri Yardımıyla Seyahat Süresi Tahmini Gerçekleştirme.", MSc Thesis, Konya Technical University, Konya, Turkey.

Liu, Y., Xia, J.C. and Phatak, A. (2020). "Evaluating the Accuracy of Bluetooth-Based Travel Time on Arterial Roads: A Case Study of Perth, Western Australia." *Journal of Advanced Transportation*. (10.1155/2020/9541234)

Malinovskiy, Y., Saunier, N. And Wang, Y. (2012). "Analysis of Pedestrian Travel with Static Bluetooth Sensors." *Transportation Research Record*, 2299, 137-149. (10.3141/2299-15)

Margreiter, M. (2016). "Automatic Incident Detection Based on Bluetooth Detection in Northern Bavaria." Transportation Research Procedia, 15, 525-536. (10.1016/j.trpro.2016.06.044)

Martchouk, M., Mannering, F. and Bullock, D. (2011). "Analysis of Freeway Travel Time Variability Using Bluetooth Detection." *Journal of Transportation Engineering*, 137 (10). (10.1061/(ASCE)TE.1943-5436.0000253)

Mathew, J. K., Devi, L. V., Bullock, D. M. and Sharma, A. (2016). “Investigation of the Use of Bluetooth Sensors for travel Time Studies under Indian Conditions.” *Transportation Research Procedia*, 17, 213-222. (10.1016/j.trpro.2016.11.077)

Mercader, P. and Haddad, J. (2020). “Automatic incident detection on freeways based on Bluetooth traffic monitoring.” *Accident Analysis and Prevention,* 146. (10. 1016/j.aap.2020.105703)

Michau, G., Nantes, A., Chung, E., Abry, P. and Borgnat, P. (2014). “Retrieving Dynamic Origin-Destination Matrices from Bluetooth Data.” *Transportation Research Board (TRB) 93rd Annual Meeting*, Washington D.C.

Moghaddam, S. S. and Hellinga, B. (2013). “Quantifying Measurement Error in Arterial Travel Times Measured by Bluetooth Detectors.” *Transportation Research Record*, 2395, 111-122. (10.3141/2395-13)

Moghaddam, S.S. and Hellinga, B. (2014). “Real-Time Prediction of Arterial Roadway Travel Times Using Data Collected by Bluetooth Detectors.” *Transportation Research Record,* 2442, 117-128. (10.3141/2442-13)

Mohrehkesh, S. and Nadeem, T. (2014) "Optimized inquiry mechanisms for Bluetooth devices," 2014 IEEE International Conference on Communications (ICC), Sydney, NSW, Australia, pp. 421-427, doi:10.1109/ICC.2014.6883355.

Montero, L., Codina, E. and Barceló, J. (2015). “Dynamic OD transit matrix estimation: formulation and model-building environment.” Advances in Intelligent Systems and Computing, 1089, 347-353. (10.1007/978-3-319-08422-0_51)

Murphy, P., Welsh, E. and Frants, J. F. (2002) “Using Bluetooth for Short-Term Ad Hoc Connections Between Moving Vehicles: A Feasibility Study”. Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367), Birmingham, AL, USA, 2002, pp. 414-418 vol.1, doi: 10.1109/VTC.2002.1002746.

Namaki Araghi, B., Krishnan, R. and Lahrmann, H. (2016). “Mode-Specific Travel Time Estimation Using Bluetooth Technology.” J*ournal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 20(3), 219-228. (10.1080/15472450.2015.1052906)

Park, S., Saeedi, A., Kim, D.S. and Porter, J.D. (2016). “Measuring intersection performance from Bluetooth-based data utilized for travel time data collection.” *Journal of Transportation Engineering*, 142(5). (10.1061/(ASCE)TE.1943-5436.0000836)

Peterson, B. S. , Baldwin, R. O. and Kharoufeh, J. P. (2006) "Bluetooth Inquiry Time Characterization and Selection," *IEEE Transactions on Mobile Computing*, 5(9), 1173-1187. (10.1109/TMC.2006.125)

Porter, J. D:, Kim, D. S. , Magaña, M. E., Poocharoen, P. and & Gutierrez Arriaga, C. A. (2013) “Antenna Characterization for Bluetooth-Based Travel Time Data Collection”, *Journal of Intelligent Transportation Systems*, 17:2, 142-151, (10.1080/15472450.2012.696452).

Pu, Z., Zhu, M., Li, W., Cui, Z., Guo, X. and Wang, Y. (2021) "Monitoring Public Transit Ridership Flow by Passively Sensing Wi-Fi and Bluetooth Mobile Devices" *IEEE Internet of Things Journal*, 8(1). (10.1109/JIOT.2020.3007373)

Quayle, S. M., Koonce, P., DePencier, D. and Bullock, D. M. (2010). “Arterial Performance Measures with Media Access Control Readers-Portland, Oregon, Pilot Study.” *Transportation Research Record*, 2192, 185-193. (10.3141/2192-18)

Ramli, S. S. M. and Hasbullah, H. (2010) "An improved inquiry procedure in device discovery Bluetooth network," 2010 International Symposium on Information Technology, Kuala Lumpur, Malaysia, pp. 771-776, doi: 10.1109/ITSIM.2010.5561549.

Ranga Suri, N.N.R., Murty, M. and Athithan, N. (2019) "Research Issues in Outlier Detection" Outlier Detection: Techniques and Applications. pp. 29-51. Springer International Publishing, Cham

Remias, S.M., Hainen, A.M. and Bullock, D.M. (2013). “Leveraging Probe Data to Assess Security Checkpoint Wait Times.” *Transportation Research Record*, 2325, 63-75. (10.3141/2325-07).

Remias, S.M., Hainen, A.M., Mathew, J.K, Vanajakshi, L., Sharma, A. and Bullock, D.M. (2017). “Travel Time Observations Using Bluetooth MAC Address Matching: A Case Study on The Rajiv Gandhi Roadway: Chennai, India.” *Joint Transportation Research Program*, Purdue University, West Lafayette, Indiana. (10.5703/1288284316505)

Ryeng, E. O., Haugen, T., Gronlund, H. and Overa, S. B. (2016). “Evaluating Bluetooth and Wi-Fi sensors as a tool for collecting bicycle speed at varying gradients.” Transportation Research Procedia, 14, 2289-2296.

Saeedi, A., Park, S., Kim, D. S. and Porter, D. (2013). “Improving Accuracy and Precision of Travel Time Samples Collected at Signalized Arterial Roads with Bluetooth Sensors” *Transportation Research Record*, 2380, 90-98. (10.3141/2380-10)

Salem, A., Nadeem, T., Cetin, M. and El-Tawab, S. (2015). “DriveBlue: Traffic Incident Prediction through Single Site Bluetooth.” *Proc., 18th IEEE International Conference on Intelligent Transportation Systems*, Gran Canaria; Spain. (10.1109/ITSC.2015.123).

Soykök, E. (2021). “Trafik Yönetiminde Bluetooth Teknolojisi Kullanimi.”, MSc Thesis, Necmettin Erbakan University, Konya, Turkey.

Thamrin, T. and Sahib, S. (2009) "The Inquiry and Page Procedure in Bluetooth Connection," 2009 International Conference of Soft Computing and Pattern Recognition, Malacca, Malaysia, pp. 218-222, doi: 10.1109/SoCPaR.2009.52.

Tsubota, T., Bhaskar, A., Chung, E. and Billot, R. (2011). “Arterial traffic congestion analysis using Bluetooth Duration data.” *Proceedings of the 34th Australasian Transport Research Forum.* The Planning and Transport Research Centre (PATREC), http://www.patrec.org/atrf.aspx, pp. 1-14.

Van Boxel, D., Schneider, W. H., and 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. https://doi.org/10.3141/2256-08

Van der Zijpp, N.J. (1997)"Dynamic OD-Matrix Estimation from Traffic Counts and Automated Vehicle Identification Data" *Transportation Research Record*, 1607. (10.3141/1607-13)

Vinagre Díaz, J. J., Rodríguez González, A. B. and Wilby, M. R. (2016) "Bluetooth Traffic Monitoring Systems for Travel Time Estimation on Freeways," in *IEEE Transactions on Intelligent Transportation Systems*, 17(1), 123-132. (10.1109/TITS.2015.2459013)

Wasson, J.S., Sturdevant, J.R. and Bullock, D.M. (2008). "Real-time travel time estimates using media access control address matching." *Institute of Transportation Engineers (ITE) Journal*, 78(6), 20-23.

Yildirimoglu, M. (2021). "Joint Estimation of Paths and Travel Times from Bluetooth Observations." *Transportmetrica B: Transport Dynamics*, 9(1), 324-342. (10.1080/21680566.2020.1852128)

Yu, W., Park, S., Kim, D.S. and Ko, S.S. (2015). "An Arterial Incident Detection Procedure Utilizing Real-Time Vehicle Reidentification Travel Time Data." *Journal of Intelligent Transportation Systems*, 19(4), 370-384. (10.1080/15472450.2014.972762)

Yuan, J., Abdel-Aty, M., Wang, L., Lee, J., Yu, R. and Wang, X. (2018). "Utilizing Bluetooth and adaptive signal control data for real-time safety analysis on urban arterials." *Transportation Research Part C*, 97, 114-127. (10.1016/j.trc.2018.10.009)

Yucel, S. (2015). "Use Of Bluetooth Technology For Traffic Analysis in Urban Road Networks.", MSc Thesis, Middle East Technical University, Ankara, Turkey.

Yucel, S., Tuydes-Yaman, H., Altintasi, O. and Ozen, M. (2013). "Determination of Vehicular Travel Patterns in an Urban Location Using Bluetooth Technology." *Proc., ITS America 23rd Annual Meeting and Expo*, Nashville, Tennessee.

Yoshimura, Y., Amini, A., Sobolevsky, S., Blat, J. and Ratti, C. (2017). "Analysis of pedestrian behaviors through non-invasive Bluetooth monitoring." *Applied Geography*, 81. 43-51. (10.1016/j.apgeog.2017.02.002)

Zaruba, G. V. and Chlamtac, I. (2003) "Accelerating Bluetooth inquiry for personal area networks." *GLOBECOM '03. IEEE Global Telecommunications Conference (IEEE Cat. No.03CH37489)*, San Francisco, CA, 2, 702-706. (10.1109/GLOCOM.2003.1258329)

(n.d.). *Learn About Bluetooth*. Bluetooth. https://www.bluetooth.com/

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Karataş, Pınar

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Civil Engineering	2015
BS	Çukurova University Civil Engineering	2011
High School	Halil Çiftçi Anatolian High School, Ceyhan, Adana	2006

WORK EXPERIENCE

Year	Place	Enrollment
2022-Present	Avensa Consulting SRL	Junior Expert
2013-2022	METU Civil Engineering	Research Assistant
2011-2013	Çukurova University Civil Engineering	Research Assistant

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

Journal

1. Tuydes-Yaman H., Karatas-Sevinen P., Oncu, Z.P. and Dalkic G., (2022) "Towards a New Walking Evaluation Approach: Power of Surveys and Route-based Evaluations in GIS Environment. " Kocaeli Journal of Science and Engineering, Vol. 5(2), pp. 212-226. Doi: 10.34088/kojose.1004404

2. Karatas-Sevinen P.,Tuydes-Yaman H., (2018) "Variability in Sidewalk Pedestrian Level of Service Measures and Rating." Journal of Urban Planning and Development, 144(4), 4018042.Doi: 10.1061/(ASCE)UP.1943-5444.0000483

Book Chapter

1. Engineering Tools and Solutions for Sustainable Transportation Planning, Chapter Title:Evaluation of Walkability and Pedestrian Level of Service (2017), Hediye Tuydes-Yaman, Pinar Karatas-Sevinen, IGI Global, Editor: Hermann Knoflacher, Ebru V. Ocalir-Akunal, Issue:1, Page Number 374, ISBN:9781522521167, English (Academic Book)

International Conferences

1. Selim Dundar, Hediye Tuydes-Yaman, Beyhan Ipekyuz, Pinar Karatas-Sevinen, Egecan Emre Huner, Evren Gungor (2018). Planning for a Successful Development of Intelligent Transportation Systems (ITS) in Turkey. 1st International Conference on Intelligent Transportation Systems-BANU-ITSC'18

2. Gulcin Dalkic, Pinar Karatas-Sevinen, Hediye Tuydes-Yaman (2018). Commute Modal Preferences of Middle East Technical University (METU) Students. 13th Internatıonal Congress on Advances in Civil Engineering

3. Dundar Selim,Evren Gungor,Huner Egecan Emre, Pinar Karatas-Sevinen, Hediye Tuydes-Yaman, Ipekyuz Beyhan (2018). An Evaluatıon Index for Monitoring Intelligent Transportation Systems (ITS) Development. ACE2018 - 13th Internatıonal Congress on Avances in Civil Engineering

4. Hediye Tuydes-Yaman, Bahar Oz, Pinar Karatas-Sevinen, Gulcin Dalkic (2018). Barriers to Walking among College Students. Transportation Research Board (TRB) 97th Annual Meeting

5. Hediye Tuydes-Yaman, Pinar Karatas-Sevinen, Zeynep Pinar Tagmat, Gulcin Dalkic (2017). Evaluation of Walkability Among METU Students via Traditional Survey Versus Route-Based Data. Transportation Research Board 96th Annual Meeting

6. Pinar Karatas, Hediye Tuydes-Yaman (2016). A Pairwise Comparison of Different Pedestrian Level of Service PLOS Ratıngs. Transportation Research Board (TRB) 95th Annual Meeting

7. Hediye Tuydes-Yaman, Pinar Karatas, Oruc Altintasi (2015). Lessons Learnt From METU Campus Walkability Evaluations. Transportation Research Board (TRB) 94th Annual Meeting

8. Hediye Tuydes-Yaman, Oruc Altintasi, Pinar Karatas (2014). Evaluating Pedestrian Level of Service at Middle East Technical University METU Campus. 11th International Congress on Advancess in Civil Engineering (ACE)

National Conferences

1. Ezgi Kundakci, Pinar Karatas, Gulcin Dalkic, Basar Ozbilen, Hediye Tuydes-Yaman (2015). Sürdürülebilir Ulaşım Ve Kampüs Trafik Güvenliği ODTÜ Öğrenci Bakışı. Sürdürülebilir Ulasim için Ulusal Yol ve Trafik Güvenliği Kongresi

2. Gulcin Dalkic, Basar Ozbilen, Pinar Karatas, Ezgi Kundakci, Hediye Tuydes-Yaman (2015). Sürdürülebilir Ulaşım İçin Yerleşke İçinde Bisiklet Kullanımı ODTÜ Kampüsü Örneği. TMMOB İMO 7. Kentsel Altyapı Sempozyumu

3. Hediye Tuydes-Yaman, Pinar Karatas (2014). ODTÜ Yerleşkesinde Bluetooth Uygulamaları İle Yaya Hareketlerinin İncelenmesi. 1. Karayolu Akıllı Ulaşım Sempozyumu ve Sergisi

4. Hediye Tuydes-Yaman, Pinar Karatas (2013). Yol Kenarı Elektronik Panolarının Trafik Güvenliğine Etkisi. 4.Karayolu Trafik Güvenliği Sempozyumu ve Sergisi