Doctoral Thesis

Origin-Destination (OD) Estimation of the Small Bus and Paratransit Passenger Using Wi-Fi Scanner

2020 年 3 月

ARIEF HIDAYAT

ACKNOWLEDGEMENT

I want to address my deepest gratitude to all who have been involved and contributed to the completion of this thesis. Firstly, I give the highest thanks to Allah SWT, the all mighty God for protection, ability and healthy to complete my thesis.

I am so grateful to the LPDP-DIKTI (BUDI) Indonesia scholarship, and Urban and Transportation Planning Laboratory, Department of Civil Engineering, Tokyo University of Science for making it possible for me to study in Japan.

My special and heartily thanks to my supervisor, Professor Shintaro Terabe, who helped, encouraged and directed as a student during my study in Japan. It is with him supervision that this work came into existence. I want to say thank to all of professors in Urban and Transportation Planning Laboratory who helps me during the study in Tokyo University of Science. Further, my highly appreciate belong to all the professor as the reviewers of my thesis.

I am also profoundly thankful for my lab-mates and surveyors. Their names cannot be disclosed, but I want to acknowledge and appreciate their help and transparency during my research. Their information helped me complete this thesis.

I am also so thankful to Universitas Teknologi Sulawesi, who permitted me to study advance in Japan. This thesis is dedicated to my wife, kids, father/father in law, mother/mother in law who encouraged me and prayed for me throughout the time of my study.

Sincerely,

Arief Hidayat

ABSTRACT

This study attempts to use a wireless fidelity (Wi-Fi) scanner to estimate the origin–destination (OD) data from passenger transit services. The author used Wi-Fi equipment in the probe request mode to capture and collect the media access control (MAC) addresses of passenger devices as the transit service traveled along its route. Hence, the objectives of this study were to gather travel data based on Wi-Fi use with a relatively inexpensive tool and to interpret the transit behavior using this data.

This research was conducted on two different transit modes in different case study areas (a small city in a developed country and a city in a developing country). The small city case study was conducted in a small bus and the developing country case considered the paratransit mode. These two modes are totally different, owing to the characteristic areas of study. The first experiment was in Obuse, Nagano prefecture, Japan and the second was in Makassar, South Sulawesi province, Indonesia. These are small cities, each of which has its own characteristics. Obuse is a tourist area with less traffic around the city and there are still many agricultural areas interspersed with urban areas. In terms of transportation, visitors use small buses for travel. Regarding the second experiment, Makassar city the fifth-largest city in Indonesia. It has heavy traffic congestion and mixed land uses.

For the bus transit mode, this thesis presents a new data cleaning procedure to characterize bus passenger volume and travel trends using a combination of MAC address and global positioning system (GPS) data. The approach proposed and developed in this study can produce outputs, such as an OD matrix and passenger volume for a bus route section. A comparison between passenger volumes obtained from the Wi-Fi data processing procedure and data obtained using the ground truth procedure indicates that the number of passengers determined using the Wi-Fi data acquisition and processing procedure is less than the number of passengers

determined using the ground truth procedure. For the paratransit mode, a cleaning procedure is developed to clean raw Wi-Fi data from non-passenger data, match ground truth and Wi-Fi data, and construct an OD or boarding–alighting dataset based on Wi-Fi estimation.

This research also describes road crowdedness and travel speed stability or travel speed reliability. The road crowdedness is based on the Wi-Fi data or MAC addresses. The main aspects of this subsection are the MAC address and vehicle speed. The vehicle speed was calculated from GPS log data. The author showed that the total number of MAC addresses is related to vehicle speed. That is, if the speed is lower, the total number of MAC addresses will increase, and vice versa. The level of service is based on a travel speed stability. Travel speed stability was calculated from the average and standard deviation of the speed. If travel speed stability index approaches 0, it means that the speed of the road not reliable.

The contribution of this research is very useful for small cities in developed countries or cities in developing countries. The new procedure can be adjusted for small bus and paratransit modes. Moreover, this research procedure for determining small bus and paratransit passengers could help the government to select an economical survey method. The crowdedness and travel time stability index could be considered as an “image” about the characteristic of the small city. Stakeholders and decision-makers can determine the optimal policy based on travel speed stability analysis without traffic volume ground truth data.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1. Background

Currently, the development of transportation in metropolitan cities is very rapid. Owing to the recent developments of technology for metropolitan cities, the use of technology for transportation surveys is also advancing.

Data about movements in transit systems in a city or region is becoming increasingly important. For example, it is important to know the number of passengers boarding or alighting in order to implement effective strategies for alleviating congestion and improving the quality of the transportation service.

In this case, the collection of transportation data is very important for investigating the current transit supply and demand in order to formulate better transit strategies in the future. The data of interest for this application are commonly called origin–destination (OD) data, which reflect a person's movements from one place to another.

However, traditional methods to record travel data rely on manual counts by on-board observers on each mode. The process of manual counting is very time-intensive, has a high cost, and has the potential for significant errors. In suburban cities or small cities, OD data retrieval and collection are still entirely conventional. Conventional transportation surveys still use a manual surveying method to count the number of vehicles, passenger, pedestrian, or cyclist.

The conventional methods of data collection to estimate OD and passenger volume include questionnaires, manual counting, and interviews (Tolouei et al., 2017). A manual survey involves manual counting by an individual within a specified period. Manual surveys require many surveyors, carry high cost, and are not suitable compared to surveying developments in the big data era for example questionnaires, interviews, and manual counting. A small city and adequate funding are reasons to continue carrying out conventional methods in terms of OD data collection.

Methods for retrieving OD data using technology have been developed. Newly developed technologies, such as smart cards, have been applied in many urban centers around the world. However, these technologies still require high cost, a fixed location, and a prominent place for equipment installation.

Currently, smartcards, phone signals, wireless systems, and image processing techniques are used in OD surveys. However, technologies used for transportation surveys all have distinct shortcomings and advantages. Loop detectors are widely used in developed countries to survey large volumes of vehicles. However, they cannot quantify the number of passengers in a vehicle and surveying technologies that can quantify the number of passengers in a vehicle are expensive. Smartcards, sensor systems, image processing techniques, and wireless devices are useful for surveying vehicles and passengers but like other tools, this equipment must also operate in a static state. Therefore, a technology is needed that is cheaper, easier, and can operate for a long time.

However, both conventional and technological methods for collecting transportation data have biases and errors. Therefore a comparison between these sources of data provides validity to improve the hypothesis (Wismans et al., 2018).

In most small cities, suburban areas, countryside, or cities of developing countries, very high technology is not used due to its high cost and the relatively low movement levels. Therefore, cheap and flexible technology to retrieve trip information data is needed that can accommodate small or medium cities or suburban areas, as well as countryside areas or cities of developing countries. In small cities, suburban areas, and cities of developing countries, the typical transportation modes are small buses (Wu and Deng, 2013) or paratransit (Pojani and

Stead, 2015).

Nowadays, cheap technologies to be used include wireless fidelity (Wi-Fi) and Bluetooth (Andión et al., 2018). Wi-Fi technology has developed significantly everywhere in developed and developing countries. Wi-Fi technology it is ubiquitous, whether it is implemented in a smartphone, tablet, laptop, vehicle, building, or other Wi-Fi-enabled devices (Pattanusorn et al., 2016). These are useful not only in computer science applications but also for transportation.

Wi-Fi technology has developed significantly and is relatively new for trip information surveys or transportation research. The development of this technology for data acquisition and processing is rapid. Wi-Fi is more convenient for the long-term survey periods. Wi-Fi obtains media access control (MAC) addresses with high and low frequency from Wi-Fi or Bluetooth-enabled devices. A MAC address is a unique device identifier that consists of numbers and letters, which is specific to a device with Wi-Fi or Bluetooth capability (Cunche, 2014; Dunlap et al., 2016; Musa and Eriksson, 2012; Vanhoef et al., 2016).

Wi-Fi is one of the communication technologies that is most widely used today for smartphones, laptops, tablets, and other devices that are currently in high demand around the world. Therefore, Wi-Fi can be used as a tool to observe movements. Consequently, Wi-Fi is one of the most useful options for obtaining travel data. Moreover, for transportation services, Wi-Fi could be used for analyzing the crowdedness of a location, such as for the prevention of human stampedes, for traffic redirection, or for remotely reporting the status of a location, with respect to the street condition or street crowdedness (Araico, 2017).

Based on this background, the objectives of this thesis are to gather travel data based on Wi-Fi technology with an inexpensive tool and to interpret the transit behavior into OD, passenger volume, or travel data. Beyond these objectives, the author aims to determine the crowdedness and travel speed stability of the street for transportation service purposes by using Wi-Fi technology.

1.2. Research Objective

This thesis attempts to use Wi-Fi technology to obtain OD data from passengers using transit services. The author used Wi-Fi equipment in probe request mode to capture and collect the MAC addresses of passenger devices while the transit mode was traveling along its route. Additionally, location technology such as global positioning system (GPS) was used in this research, which conformed with the Wi-Fi equipment.

The general objective of this research is to understand and analyze Wi-Fi and GPS data. More specifically, the objectives of this study are as follows:

1. To associate or combine Wi-Fi and GPS log data.
2. To estimate passenger volume, OD matrix, or travel data based on Wi-Fi and GPS log data, which requires a new cleaning procedure.
3. To validate the OD estimation by comparing it with ground truth data.
4. To estimate road crowdedness and travel speed stability from Wi-Fi technology.

1.3. Research Novelty

The novel contribution of this research is a method to interpret Wi-Fi data as passengers. In more detail, the novel aspects of this study are as follows:

1. This research proposes a new cleaning method to transform Wi-Fi log data into estimated passenger levels and an OD matrix.
2. The experiment tool presented in this thesis used a mobile Wi-Fi scanner instead of a static scanner, enabling in-vehicle applications.
3. This research conducted for small bus and paratransit transportation modes, which is applicable for small cities or cities in developing countries.
4. The procedures to determine the OD estimation are based on associated Wi-Fi and

GPS log data.

5. This research also verified the road crowdedness and travel speed stability based on the Wi-Fi analysis and travel speed stability index from GPS log analysis.

1.4. Limitations and Assumptions

There are some limitations and assumptions of this study, as follows:

1. The results of the field survey obtained a large dataset, which may include significant erroneous date. The data include passengers and nonpassengers. Nonpassengers may be in the form of pedestrians, vehicles, or buildings, which were recorded because they were near the bus and acquired by the Wi-Fi scanner. Such errors and nonpassengers must be cleared during the passenger verification process in this thesis. However, even in the data retrieval process, certain passenger data may not be recorded if the passengers turned off their Wi-Fi devices. Therefore, there are two possible errors in the proposed methodology: (1) nondetection of passengers without smartphones and (2) erroneous detection of nonpassengers with smartphones (Fukuda et al., 2017).
2. Another limitation is the consideration of the number of Wi-Fi devices held by passengers. The author assumed that each passenger carries only one Wi-Fi device. Accordingly, each detected device can be uniquely associated with one onboard passenger. However, some passengers may carry multiple Wi-Fi devices. This introduces another source of error when attempting to estimate the number of passengers.

CHAPTER 2
LITERATURE REVIEW

2.1. MAC Address

In the principle method, multiple sensors are used to record the unique Bluetooth or Wi-Fi MAC address for each wireless communication device (Dunlap et al., 2016; A. Petre et al., 2016). A MAC address is a unique device identifier that consists of particular numbers and letters, which is specific to a device with Wi-Fi capability (Cunche, 2014; Dunlap et al., 2016; Musa and Eriksson, 2012; Vanhoef et al., 2016). Wi-Fi systems capture MAC address data from Wi-Fi device users (Abedi et al., 2015; Dunlap et al., 2016; Jackson et al., 2014). The MAC address is constant for particular time on each device that is Wi-Fi-enabled. In addition, one MAC address cannot be assigned to two devices (Asija, 2016; Hidayat et al., 2017, 2018; Sapiezynski et al., 2015; Shiravi et al., 2016).

A Wi-Fi MAC address can be used to identify a mobile device, and it can be used to determine the location of a mobile device when it is combined with the received signal strength at multiple locations (Xu et al., 2013). The use of Android is also widely applied to detect pedestrian movements. Most smartphones, which have Wi-Fi functionality, usually send probe request mode to connect to a Wi-Fi access point (depending on the device). A probe request frame includes the MAC address to analyze the pedestrian flow (Fukuzaki et al., 2014). MAC address data were traced to determine the position of pedestrians with a probabilistic method, yielding a set of candidate lists of destinations, including the probability of each list of destinations being the true one (Hamacher et al., 2010).

In another method, a penetration ratio is calculated by combining tracking and counting data from Wi-Fi signals. This is the ratio between the number of counts and the number of tracks

(J. van den Heuvel et al., 2016). Pedestrian data are estimated with a system to detect anonymous MAC addresses of devices at short distances at fixed locations (Jackson et al., 2014) and the performance of a BT–Wi-Fi system to detect anonymous MAC addresses of devices at short distances at fixed locations is evaluated (Lesani and Moreno, 2016).

2.2. Why Wi-Fi? What about Bluetooth?

There are two cheap technologies that could be used for transportation. Wi-Fi and Bluetooth are new technologies for transportation. The author prefers using Wi-Fi technology because of its high range of approximately 100–300 m and the ability to capture two channels of different frequencies, 2.4 and 5 GHz (Abbott-jard et al., 2013; Kabir and Khan, 2010). Bluetooth is intended for portable products at short ranges, for use anywhere that at least two Bluetooth devices are within a 10-m range (Chhabra, 2013; Man, 2002); such devices have limited data transfer rates (Man, 2002) and limited battery power. Consequently, it offers very low power consumption and, in some cases, will not measurably affect battery life. On the contrary, Wi-Fi is designed for a longer-range connection and supports devices with portable and substantial power supplies (Pothuganti and Chitneni, 2014).

Based on low-power research, Bluetooth networks are attractive for indoor positioning due to the easy installation, low power consumption, and long cycle of the beacons. Regarding power consumption, the benefits of Bluetooth are impressive. Regarding accuracy, however, the advantages are not so obvious. Even with maximal signal strength, the experiments did not achieve comparable coverage to that of Wi-Fi (Lindemann et al., 2016). In conclusion, the author prefers to use Wi-Fi rather than Bluetooth. However, the author researched cleaning, analyzing, and distinguishing data based on several Bluetooth studies. Table 1 shows the differences between Bluetooth and Wi-Fi based on the standard parameter (Kabir and Khan, 2010).

Table 1. Differences between Bluetooth and Wi-Fi

Standard	Bluetooth	Wi-Fi
IEEE Specification	802.15.1	802.11a/b/g
Frequency Band	2.4 GHz	2.4 GHz, 5 GHz
Max Data Rate	3 Mbps	54 Mbps
Packet Length	1024 Bytes, 8–27 Bytes (BTLE)	1024 Bytes
Distance Coverage	10 m	100 m
Battery Life	Regular charging	Hourly charging
Nominal TX Power	0-10 dBm	15–20 dBm
Number of RF Channel	79	14,23
Bandwidth	1 MHz	22 MHz
Modulation Scheme	GFSK	BPSK, QPSK, COFDM, CCK, M-QAM
Coexistence Mechanism	AFH	Dynamic Frequency Selection, Transmit Power Control
Basic Cell	Piconet	BSS
Extended Cell	Scattered	ESS
Max Number of Nodes	8	2007
Encryption	E0 Stream Cipher	RC4 stream cipher (WEP), AES block cipher
Authentication	Shared Secret	WPA2 (802.11.i)
Data protection	16-bit CRC	32-bit CRC
Protocol Complexity	High	Medium
Cost	Medium	High

2.3. Infrastructure Mode and Probe Request Mode

To understand the equipment employed in this research, we present an explanation of the infrastructure and probe request modes of Wi-Fi. Wi-Fi can operate in infrastructure mode or probe request mode. For this study, the Wi-Fi equipment or Wi-Fi scanner is operated in probe request mode.

Wi-Fi technology is based on IEEE 802.11 standards (including 802.11a, 8.02.11b, 802.11g, and 802.11n) (Cisco, 2008; Najafi et al., 2014). The most common mode of operation for 802.11 is called infrastructure mode, in which mobile devices or wireless access points communicate with other wireless access points and wired networks (typically Ethernet).

In infrastructure mode, networks communicate with each other through access points. Most wireless devices, such as smartphones, tablets, routers, and laptops, are configured for infrastructure mode. A smartphone can be identified by its international mobile equipment identity (IMEI) number or MAC address. Smartphone Wi-Fi is designed to periodically transmit probe requests to identify known access points (Matte, 2017; Yaik et al., 2016). Probe requests are active scans by mobile devices (Sun et al., 2017; Verbree et al., 2013), the content of which includes the sender's MAC address (Musa and Eriksson, 2012).

A Wi-Fi scanner in probe request mode can be used to collect MAC address data from devices. With an appropriate Python script installed on the scanner, it can collect probe requests automatically (Fukuda et al., 2017). Several previous studies have used Wi-Fi scanners for bus passengers (Dunlap et al., 2016; Hidayat et al., 2018), paratransit passengers (Fukuda et al., 2017), and used pedestrians (Fukuzaki et al., 2014; Hidayat et al., 2017; A.-C. Petre et al., 2017; Schauer et al., 2014). A Wi-Fi scanner can load all MAC address data into a log file. This tool can access MAC addresses without connecting to the internet and is capable of passively scanning devices to collect data. A Wi-Fi scanner operating in probe request mode can also be used to obtain MAC addresses from devices that are operating in infrastructure mode.

2.4. Wi-Fi on Bus and Paratransit

To understand Wi-Fi in the context of transportation surveys, Bluetooth and Wi-Fi technologies are reviewed below. The previous applications of these technologies and their reported results are reviewed to understand the purpose of Wi-Fi for OD surveys and to build on these technologies to further advance transportation survey capabilities. The process captures and reads the MAC address installed on the vehicle and estimates the time taken to travel from the starting point to the end point using static Wi-Fi devices placed at several locations. The data are interpreted as vehicles following previous research that used the indicators of more mileage and longer travel time. This system has been applied to a large bus (Dunlap et al., 2016), a small bus in a small city (Hidayat et al., 2018), and pedestrian tracking (Abedi et al., 2013; Musa and Eriksson, 2012; A.-C. Petre et al., 2017).

Previously, a study related to the on-bus use of Wi-Fi calculated the OD bus passenger matrix (Dunlap et al., 2016). The study attempted to compare the results of Wi-Fi and Bluetooth usage and develop a raw data filtering procedure (Wi-Fi and Bluetooth data) into an estimation of passenger numbers. Similarly, in research on transit data in Obuse, Japan, passenger estimation is predicted using a speed indicator as well as the MAC address positional filtering procedure (Hidayat et al., 2018).

Previous research on passenger behavior using on-bus Wi-Fi data revealed that MAC addresses that do not change location on the same bus could be associated with passengers (Jiang et al., 2016). Data filtering is essential for identifying passengers and nonpassengers. This research applies a filtering process to interpret and verify the raw nonpassenger Wi-Fi data obtained in the experimental process (Fukuda, Kobayashi, Nakanishi and Suga, 2017; Oransirikul, Nishide, Piumarta and Takada, 2014) to measure bus passenger loads by monitoring Wi-Fi transmissions from mobile devices, thereby revealing insights into the patterns of travel as well as pick-ups and drop-offs.

Research using Wi-Fi equipment for paratransit had been done before with the possibility of using Wi-Fi-based passive passenger monitoring to collect spatial information about passengers, including boarding and alighting information (Fukuda et al., 2017). The difference between the study by Fukuda et al. (2017) and the present research is in the data matching analysis. Fukuda et al. (2017) proposed two methods of matching data, which were based on time and location gaps (Wi-Fi and ground truth data). On the contrary, the present Wi-Fi research considers matching each MAC address (first and last appearance) with the ground truth data (boarding and alighting) after reducing them by time and location distance thresholds. Several innovative studies based on travel surveys use Wi-Fi to understand travel behavior or transportation.

2.5. Learning How to Clean Wi-Fi Data Based on Wi-Fi and Bluetooth Nonpassenger Experiment

Recently, there has been much research on methods of collecting nonpassenger data. Research using Wi-Fi and Bluetooth as tools for counting nonmotorized travel users confirms this (Böhm, Ryeng and Haugen, 2016; Malinovskiy et al., 2012; Nishide and Takada, 2013; Poucin, Farooq and Patterson, 2016). There are significant benefits and challenges to the use of Wi-Fi and Bluetooth data for the analysis of spatiotemporal dynamics of human movement, crowd MAC data collection and monitoring (Abedi, Bhaskar and Chung, 2013), and combining data from both (Wi-Fi and Bluetooth) sensor types (Wi-Fi and Bluetooth), which yields useful insights into pedestrian dynamics (van den Heuvel, Ton and Hermansen, 2016). Previous research has also used Wi-Fi and Bluetooth data of pedestrians inside terminals to detect moving pedestrians and their behavioral patterns within the terminal, and thus to create an O–D motion matrix (Shlayan, Kurkcu and Ozbay, 2016).

Similar research on Wi-Fi systems for traffic monitoring focused on pedestrians, vehicles, and bicycles (Jackson, Lesani and Moreno, 2014). The study placed Wi-Fi scanners on streetlights to capture MAC address data, identifying and estimating numbers of pedestrians, bicycles, and vehicles from the speed of each travelling MAC address. Initially, pedestrians were identified by filtering for MAC addresses that could be captured across a distance of 100 m (Malinovskiy, Saunier and Wang, 2012). One study detected vehicles using Wi-Fi and Bluetooth devices installed in a car (Ahmed, El-Darieby, Morgan and Abdulhai, 2008) and another detected vehicles with a focus on travel time (Mai, Kusakabe, Suga and Oguchi, 2017).

Wi-Fi with Android applications is also widely used to detect pedestrian movements. Although it depends on the device, most smartphones usually send probe request frames to connect to a Wi-Fi access point, which include the MAC address (Fukuzaki, Nishio, Mochizuki and Murao, 2014). In this study, MAC address data were traced to determine the position of the pedestrian with a probabilistic method consisting of a set of candidate lists of destinations, with the probability of each record of targets being the true one (Hamacher, Heller and Ruzika, 2010).

Finally, the research uses Wi-Fi devices paired permanently in strategic locations (Lesani and Moreno, 2016), and the use of software installed on smartphones; it should be noted that most pedestrians do not make use of this software (Shlayan et al., 2016).

There are several studies on using Bluetooth and Wi-Fi to detect pedestrians. Bluetooth is used to measure pedestrian activity with a detection rate of 2%, which is lower than that of Wi-Fi. Wi-Fi data is noisier, owing to the higher detection rate, but can it also be more informative (Lesani and Moreno, 2016). Previous research on pedestrian activity demonstrated that Bluetooth could be used to determine pedestrian flow and density, but is less accurate compared to ground truth data (Schauer et al., 2014). However, the results were shown to vary depending on the location and time of the survey. Filtering data to isolate pedestrian activity uses the calculated speed to discriminate between pedestrians, cyclists, and cars (Abedi et al., 2015). Another investigation applied a waiting time filter as a Bluetooth data filter to determine pedestrian flow more accurately (Kurkcu and Ozbay, 2017).

2.6. Wi-Fi, Time, and Speed

In intersection estimation research, there is a relationship between Wi-Fi data and travel time. Such intersection estimation research seeks to confirm the accuracy of Bluetooth and Wi-Fi data on urban roads against reliable travel time results (Shiravi et al., 2016). The research that detects human movement uses high-frequency Wi-Fi in association with GPS to identify the position of the MAC address or access points (Sapiezynski et al., 2015).

The use of Wi-Fi and Bluetooth in public terminal transportation has also been applied in a high and wide frequency band to capture MAC addresses, such that the travel behavior of pedestrians can be identified and understood in terms of seconds and minutes (Shlayan et al., 2016). This public terminal transportation research considers high-frequency Wi-Fi data compared with Bluetooth data. The reliability of travel time detected using Bluetooth has been investigated to identify the ability of Bluetooth to detect MAC addresses (Araghi et al., 2015).

In such studies, data processing was conducted by dividing the detection zone and detection time. Another empirical evaluation of Wi-Fi was conducted on road transport. The method employed was to perform "exit to exit" detection using a data-filtering procedure with time as the main variable (Abbott-jard et al., 2013). "Exit to exit" detection is intended to identify the "beginning" and "end" of each MAC address's identification time. Other travel research using static equipment has been conducted for bicycle users with time and speed filtering to confirm the penetration rate of Wi-Fi data (Böhm et al., 2016; Ryeng et al., 2016). Research on the time travel estimation process is important in confirming the accuracy of Wi-Fi data (Hidayat et al., 2018).

Bluetooth applications for determining vehicle travel time can be used to detect abnormal traffic. The index characteristic for travel time used a "min," "max," "average," and "medium" scale (Namaki Araghi et al., 2015). Other research focuses on OD in urban areas. This approach uses MAC address data and calculates the travel time and speed of vehicles based on the "start" and "end" of a recorded MAC address (Khliefat and Shatnawi, 2017). The use of Bluetooth can also be applied to road surveys, and previous investigations have demonstrated the importance of the MAC address data cleaning process for the investigation of transport activity on roads using "entrance" and "exit" specifications (Abbott-jard et al., 2013). This process records the position of a MAC address at the beginning (entrance) and end (exit) of a probe by a Bluetooth scanner tool. This method also uses a timestamp parameter to calculate the travel time of each vehicle, either in a public terminal space or on the highway (Shlayan et al., 2016).

Bluetooth and Wi-Fi research are strongly associated with time and speed variables for static and mobile scanners. The filtering process, based on time variables, is vital for determining whether Bluetooth is useful for a transportation survey (Abedi et al., 2015; Erkan and Hastemoglu, 2016; Filgueiras et al., 2014; Young, 2012). Time analysis is used to determine the OD matrix based on travel time (J.Barceló et al., 2010). Other investigations (Namaki Araghi et al., 2015; Porter et al., 2013; Pourhassan, 2016) described the calculated travel time between multiple Bluetooth and Wi-Fi antenna sensors.

Furthermore, Araghi and colleagues used a combination of methods to analyze travel time (Araghi et al., 2015). A previous investigation conducted by Romancyshyn and colleagues described the use of Bluetooth to determine travel time; delays in urban areas demonstrated that weather significantly affects travel time, and that Bluetooth could be used to monitor vehicle activity accurately (Romancyshyn, 2016). The difference between Bluetooth and Wi-Fi has also been examined and it was determined that Wi-Fi has a more extensive operating range of 10–100 m (Ferro and Potorti, 2005). Moreover, Wi-Fi can operate using two frequency bands (2.4 and 5 GHz), whereas Bluetooth can only operate at 2.4 GHz.

2.7. Geographic Information System and Wi-Fi

Geographic information system (GIS) software illustrates spatial data distributions. In the case of transportation, this is accomplished using the open-source Quantum GIS software. There are several studies related to the use of GIS for transportation and travel time. The distribution of spatialized Wi-Fi data can be identified if the data have specific XY coordinates captured through the Wi-Fi scanner and GPS (Feng and Liu, 2012; Odiyo, 2014). GIS is also more efficient for wireless data deployment in the development of trade and services or urban planning (Aldasouqi and Salameh, 2014). Furthermore, research on travel diary data has been conducted based on travel time using GIS. In such research, the analyses use “starting” and “ending” person–trip data in combination with spatiotemporal data (Yu and Shaw, 2004).

An analysis of “day-to-day” variations in travel time using GPS connected to a notebook PC can capture vehicle movement over multiple days. One study describes the tracking of vehicle movements using a GPS device based on travel time and travel speed (Ohmori et al., 2002). Path GIS research uses spatial trajectory analysis on all of the GIS data points obtained for vehicle movements (Zambrano et al., 2016). More specifically, by examining the logical path of tourist movements using GPS data from tourism spots, the travel time data for each tourist (Meng-Lung Lin, Chien-Min Chu and Tsai, 2009) or the fastest and shortest travel times can be classified using GIS modeling (Abousaeidi et al., 2016; Ilayaraja, 2013).

CHAPTER 3
METHOD

3.1. Overview of Field Experiment

This research was conducted on two different transit modes, which were in two different case study areas, which were a small city in a developed country and a large city in a developing country. For these case study areas, the experiments were conducted in a small bus and in paratransit mode, respectively.

These two modes are totally different because of the characteristic area of study. The first experiment was in Obuse, Nagano prefecture, Japan, and the second was in Makassar, South Sulawesi province, Indonesia.

These areas are a small city and a large city, which have their own characteristics. Obuse is tourism area with less traffic around the city and there are still many agricultural areas interspersed with urban areas. For transportation, this city uses small buses to distribute its visitors. The second experiment was in Makassar, which is one of the major cities in Indonesia, South Sulawesi province. Makassar city is large city with heavy traffic and mixed land use. For transportation, the majority of people use a paratransit mode, which is called *petepete*.

3.2. Small Bus Field Experiment

For the small bus experiment, the study location is the town of Obuse in the Kamitakai District in the Nagano Prefecture of Japan. Obuse is one of the top tourist destinations in Japan. It attracts 764,000 visitors every year. Obuse is unique for its chestnut processing industry and as a historic city with a variety of tourist attractions. The survey was done during the years 2016, 2017, and 2018. In 2016 and 2017, the survey was conducted in one day which are 30 October

2016 and 22 October 2017. In 2018, it was conducted for two days (13–14 October 2018). In October 2016, the town had an estimated population of 10,698 and a population density of 560 people per square kilometer. Its total area is 19.12 km^2.

Obuse has a circulating shuttle bus called the "Romango." The Romango is a hop on–hop off bus, which means that the bus stops at each location for only a few minutes and allows passengers to travel around the city with as single tour ticket for an entire a day (Figure 1). The fare for a day-trip ticket is 3 USD, which is equivalent to approximately 300 Japanese yen. Obuse operates two buses following a circular route every Saturday and Sunday, which connect to nine bus stops. Visitors can reach several tourist attractions by using this bus.

Figure 1. Obuse Romango bus

The Romango bus is a medium-sized bus with a maximum capacity of 25 passengers. The bus route has nine stops (Table 2) along its route (Figure 2) and seven route circulations based on the timetable. The route circulations are referred to as circulation numbers (CNs), which depend on the bus service time. The Romango bus traverses seven CNs in its circulation from bus stop one (BS1) to bus stop nine (BS9) and passes nine bus stops: BS1 (Obuse Highway

Oasis Park), BS2 (Obuse Station), BS3 (Hokusai Museum), BS4 (Obuse Museum), BS5 (Matsumura Town Parking), BS6 (Obuse Hot Springs), BS7 (Floral Garden), BS8 (Jyokoji Temple), and BS9 (Ganshoin Temple). The circular route is approximately 15 km in length and the overall route length is 8.8 km. The longest segment is 2.7 km from BS1 to BS2 and the shortest is 0.3 km from BS4 to BS5. The buses depart every 30 min from 9:50 AM to 5:50 PM. The round-trip travel time is 50 min. Bus no. 1 operates from 9:50 AM to 5:10 PM and bus no. 2 operates from 10:20 AM to 5:50 PM.

Table 2. Obuse bus stops

Number of Bus Stop	Bus Stop Name/Famous Destination Near Bus Stop	Photographs
1	Obuse Highway Oasis Park	
2	Obuse Station	
3	Hokusai Museum	
4	Obuse Museum	
5	Matsumura Town Parking	
6	Obuse Hot Springs	

Number of Bus Stop	Bus Stop Name/Famous Destination Near Bus Stop	Photographs
7	Floral Garden	
8	Jyokoji Temple	
9	Ganshoin Temple	

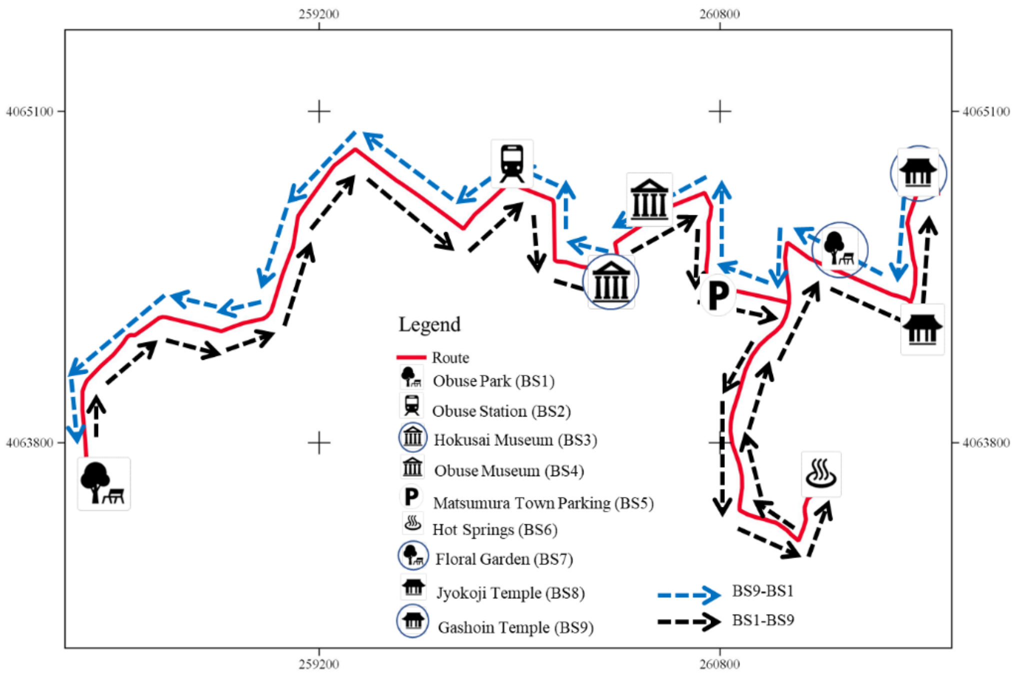


Figure 2. Bus route

The passenger/tourist demographics in Obuse area are generally in the adult-to-senior age range. Tourists in Obuse come not only from within Nagano prefecture, but also from other prefectures and other countries.

3.3. Small Bus Wi-Fi Scanner Installation

The measured number of boarding or alighting passengers was collected from the Wi-Fi scanner and compared with the manually counted results. The Wi-Fi scanner, which was developed by the author's laboratory professor as a prototype, scans for mobile devices that emit management packets every five seconds. The Wi-Fi scanner equipment captures MAC addresses from devices such as smartphones, laptops, tablets, computers, and other Wi-Fi-enabled devices. A MAC address is a unique code, specific to a Wi-Fi-enabled device, and does not contain any personal information. The Wi-Fi scanner equipment includes an antenna, a GPS receiver, and a mobile battery (Figure 3).

This scanner uses a Raspberry Pi minicomputer as a controller. The Raspberry Pi is a single-board computer running a quad-core CPU at 900 MHz. It is powered by a 30,000 mAh portable battery, with up to a 12-hour battery life. This system has 1 GB RAM, a USB port, four-pole stereo output, video port, and an HDMI port output. The Raspberry Pi also includes a high-frequency BU 353 GPS tracking device and a microSD slot, which is used for loading the operating system and storing data. This system was built to make a Wi-Fi scanner to operate in probe request mode to record MAC addresses from devices running in infrastructure mode.

The Wi-Fi scanner was placed in the bus and positioned in proximity to the bus driver (Figure 4). The scanner has an approximate range of 100–300 m and detects bus passengers' and surrounding pedestrians' Wi-Fi devices, as well as Wi-Fi devices located within buildings and vehicles (Figure 3). Upon concluding the field survey, the Wi-Fi scanner was turned off, and the MAC addresses were downloaded for further analysis. The MAC addresses are referred to as raw data that will be used to estimate passenger information.

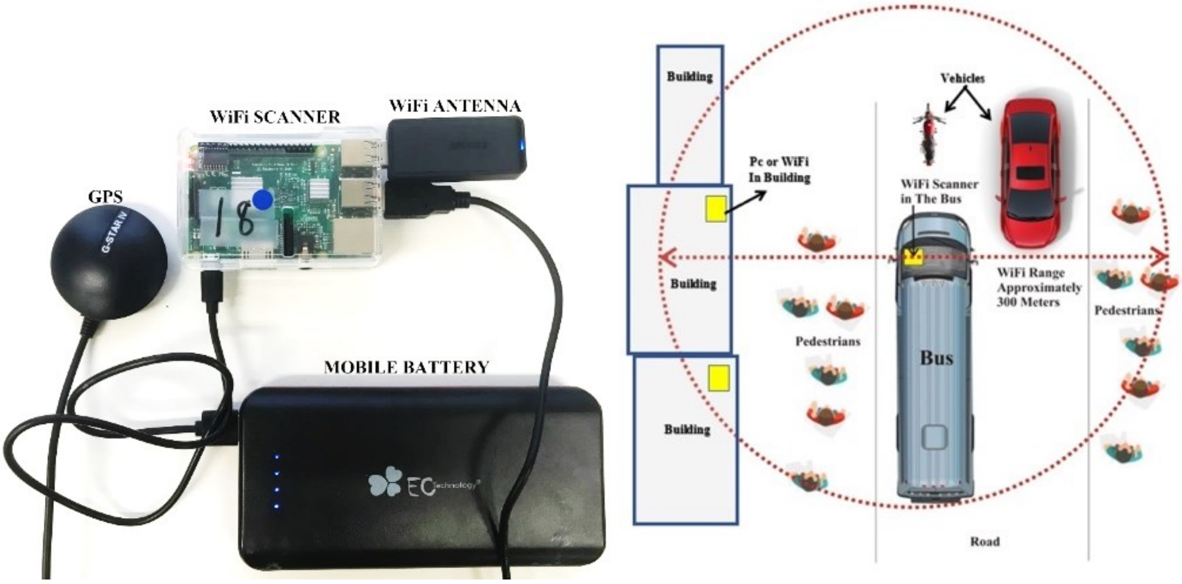


Figure 3. One set of Wi-Fi scanner equipment and illustration of Wi-Fi range

Figure 4. Wi-Fi scanners installed on buses No. 1 and No. 2

3.4. Paratransit Field Experiment

This study was conducted in Makassar, one of the major cities in Indonesia. Makassar is the capital city of South Sulawesi Province in Indonesia. It has a population of approximately 1,500,000 people and an area of approximately 175.77 km^2 (Figure 6). Makassar has a paratransit system called *petepete* (Figure 5).

The *petepete* fleet in Makassar is comprised of approximately 6,000 units with 18 routes.

The *petepete* include medium-sized buses as well as minibuses that can accommodate approximately 10–12 people per vehicle.

The *petepete* have a clear route but do not have nodes or fixed stops, because the stops depend on passenger requirements. The *petepete* system does not have a timetable. As explained in the introduction, *petepete* have paratransit characteristics in general.

Figure 5. *Petepete* overview

This experimental test was conducted on five routes (Table 3 and Figure 7). The route codes are B, D, E, G, and H. These routes were chosen according to the route length; number of *petepete*; diversity of land use; and primary, secondary, and local level road coverage. The routes cover areas of mixed land use, such as residential, traditional markets, business, industrial, and shopping centers. The five routes are the Cendrawasih, Ikip, Antang, Panampu, and Daya routes. The length of the routes extended 23.3 km, and the time required for one circuit was approximately 60–80 min. Data retrieval was carried out on August 23–27, 2018, from 09:00 to 17:00.

Table 3. *Petepete* routes, unit, and length of route

Code Route	Route	Unit	Length of Route (Km)
A	Makassar Mall -BTN Minasa Upa	165	12.1
B	Pasar Butung - Cendrawasih Trm. Malengkeri	421	12.4
C	Makassar Mall - Tallo	220	7.4
D	Makassar Mall Terminal. Regional Daya - Perumnas Sudiang	802	23.3
E	Makassar Mall - UNM - Perumnas Panakkukang	379	11.5
F	Makassar Mall - Veteran Trm. Malengkeri	286	10.4
G	Makassar Mall - Ir. Sutami/Toll - Trm. Regional Daya	348	20.1
H	Makassar Mall - Perumnas Antang	329	15.5
I	Makassar Mall - STIKI - Borong	299	9.3
J	Makassar Mall - Pa'Baeng2 - Perumnas Panakkukang	200	10.2
S	Makassar Mall - BTP	221	14.8
B1	Trm. Malengkeri - Cendrawasih - Kampus Unhas	146	24
C1	Tallo - Kampus Unhas	36	20.4
E1	Perumnas Pankkukang - UNM - Kampus Unhas	149	19.5
F1	Trm. Malengkeri - Vetran - Kampus Unhas	53	20.3
R1	Pasar Baru - Ujung Tanah - Kampus Unhas	2	20.1
W	BTP - Terminal Daya - SMA Negeri 6	5	9.6
	Total	4113	260.9

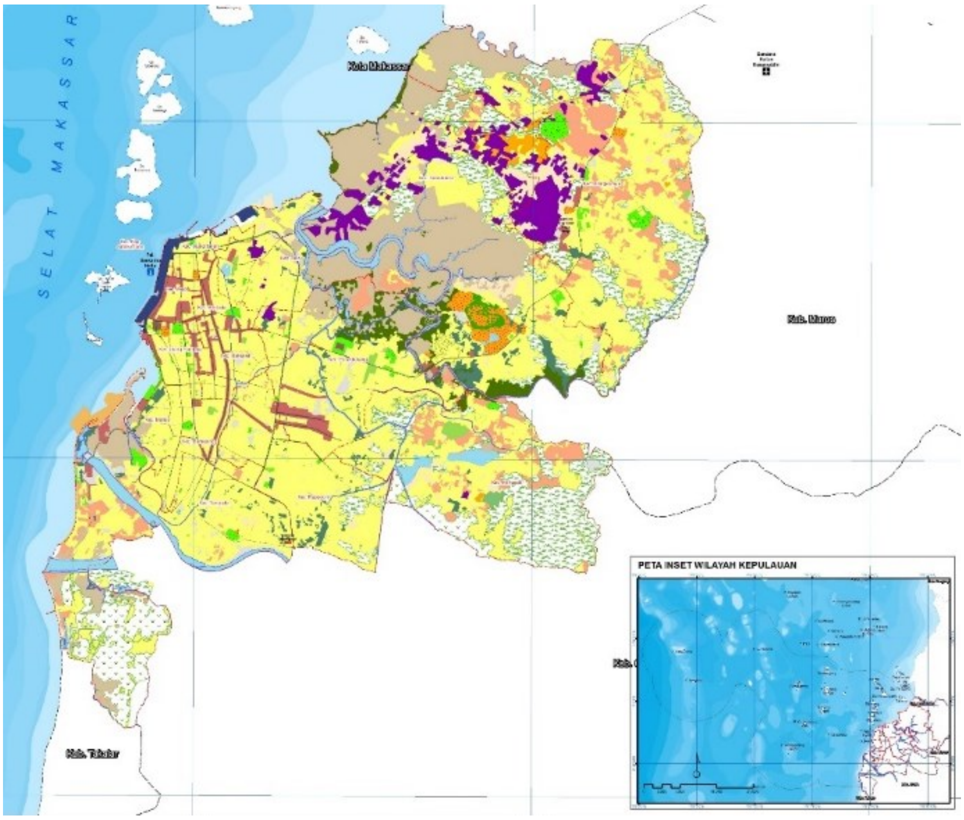
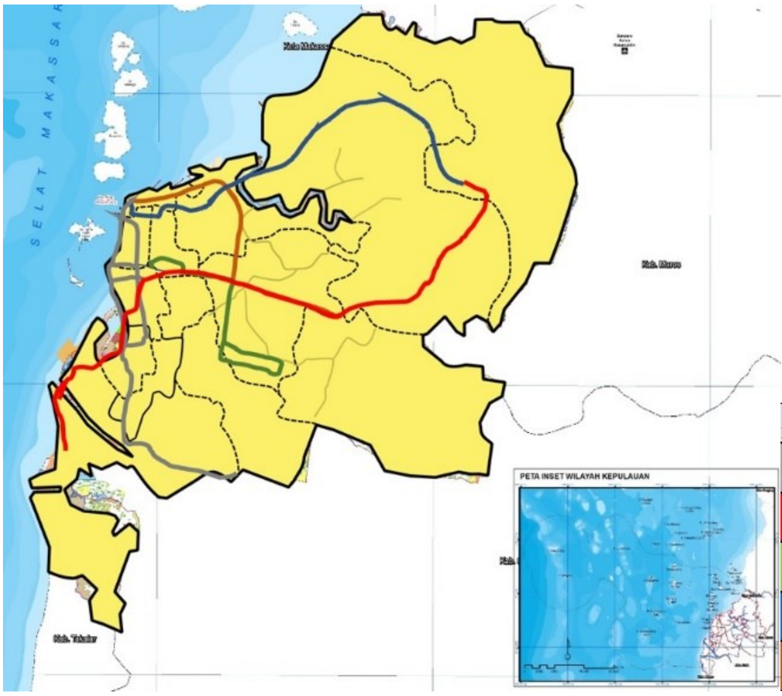


Figure 6. Makassar land use map



CODE ROUTE	ROUTE
B	Pasar Butung - Cendrawasih Trm. Malengkeri (Cendrawasih)
D	Makassar Mall Terminal. Regional Daya - Perumnas Sudiang (Daya)
E	Makassar Mall - UNM - Perumnas Panakkukang (Ikip)
G	Makassar Mall - Ir. Sutami/Toll - Trm. Regional Daya (Panampu)
H	Makassar Mall - Perumnas Antang (Antang)

Figure 7. Route field experiment

3.5. Paratransit Field Survey Scenario

3.5.1. First Scenario

This survey scenario used five routes (Figure 8). Each route had four *petepete* units, and each vehicle was followed by a surveyor and a Wi-Fi scanner. This survey was conducted on August 23, 24, 26, and 27 of 2018 from 09:00 until 17:00.

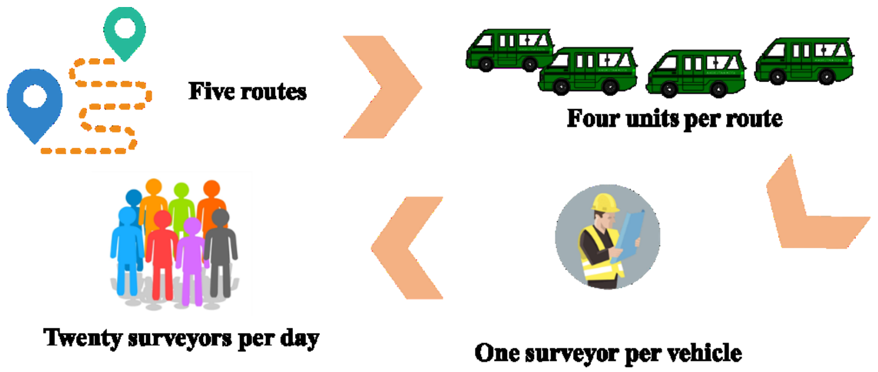


Figure 8. First scenario of paratransit survey

3.5.2. Second Scenario

This survey scenario used only one route (Figure 9). The route had 20 *petepete* units, and each vehicle was followed by a surveyor and a Wi-Fi scanner. This survey was conducted on August 25, 2018 from 09:00 until 17:00.

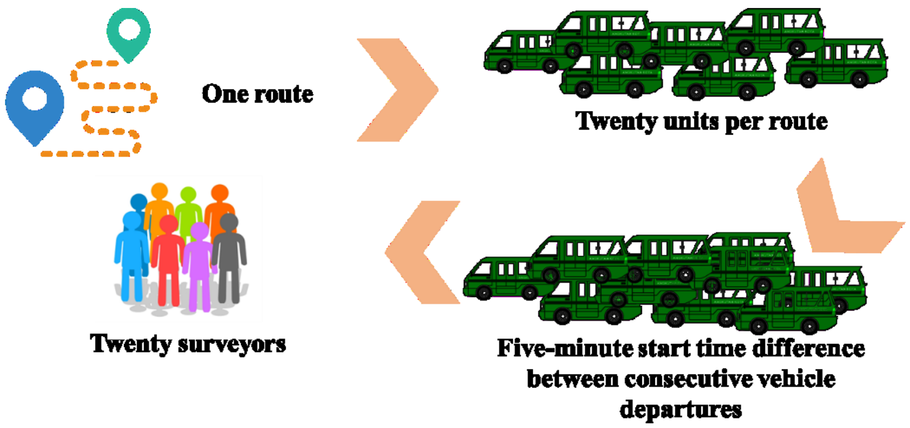


Figure 9. Second scenario of paratransit survey time

3.6. Wi-Fi Scanner Equipment and Installation for Paratransit Survey

The Wi-Fi scanner equipment used in this study is equivalent to that used in the Obuse survey, and included an antenna, a GPS receiver, and a mobile battery (Figure 10). The Wi-Fi scanner was placed in the *petepete* and positioned in proximity to the driver. The surveyor guarded the Wi-Fi equipment along with recording ground truth data. After field testing inside the *petepete,* the Wi-Fi scanner captured approximately 5,700,000 MAC addresses. This scanner provided two datasets, i.e., Wi-Fi log data and GPS log data.

Figure 10. Wi-Fi scanner equipment

3.7. Small Bus and Paratransit Ground Truth Surveys

OD ground truth data was collected to validate the results of the Wi-Fi data analysis. Each vehicle had one surveyor to capture the ground truth data. The surveyor recorded the boarding and alighting of passengers from 09:00 to 17:00 for the Obuse Romango buses and from 08:00 to 18:00 for the *petepete*. The surveyor oversaw recording the passenger conditions and keeping the Wi-Fi equipment functioning properly and carried a package of survey tools in the *petepete* to record data. The form completed by the surveyor included the passenger ID, boarding and alighting time, and the place names of origin and destination. Figure 11 shows photographs of surveyor recording by manual counting inside a bus (Figure 11 a) and *petepete* (Figure 11 b).

a) Obuse ground truth survey

b) Petepete ground truth survey

Figure 11. Ground truth survey

CHAPTER 4
SMALL BUS APPLICATION

4.1. Introduction

The main objective of this part is to estimate small bus passengers based on Wi-Fi scanner survey data from Obuse. This part describes a new data cleaning procedure used to characterize bus passenger volume and OD data using a combination of MAC address and GPS data. This research used a Wi-Fi scanner to detect MAC addresses of individual bus passengers.

The Wi-Fi scanner can perform a probe request mode to capture MAC addresses from mobile devices or other Wi-Fi-enabled modalities without connecting to the internet. The approach developed in the proposed study can output passenger volumes for different bus route sections.

Several cleaning rules were applied to filter the log data to include only passengers based on the nature of the user's behavior: (1) the passenger boards the bus at a bus stop, (2) the passenger alights at another bus stop, and (3) the passenger can be detected between these two stops. The time stamps and longitude and latitude of the devices are also recorded by the GPS antenna and the logger integrated into the Wi-Fi scanner. A bus stop visited by the bus was identified by matching the recorded longitude and latitude with a map of the bus route.

The comparison results of passenger volumes and OD obtained from the Wi-Fi data and the ground truth indicates that the number of passengers determined using the Wi-Fi data is less than that obtained from the ground truth survey. Therefore, the new cleaning procedure is effective to clean Wi-Fi raw data into passenger data.

4.2. Processing Data

The processing data was produced in two ways. The first is by simple estimation (Figure 12) and the second is a specific value alternative (Figure 13). The results of the survey include both Wi-Fi and GPS log data. The Wi-Fi log contains the time and MAC address data, and the GPS log contains the time and coordinate (position) data. The Wi-Fi and GPS logs obtained from the Wi-Fi scanner were distinct.

The data processing procedure combines the Wi-Fi and GPS logs, converts the coordinate system to Universal Transverse Mercator (UTM), calculates the speed, bus stop zone, bus route circulation, and finally the passenger volume.

Two data processing schemes were applied to the small bus. The first (Figure 12) was used to determine the estimated passenger volume and OD, which used the 2016–2017 data. The second (Figure 13) was used to decide the threshold, which used the 2018 data.

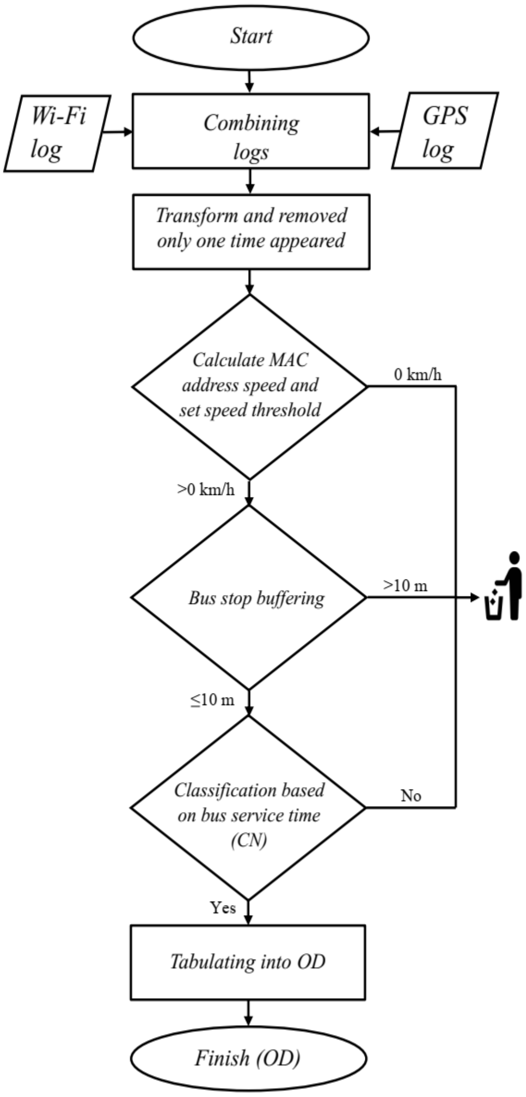


Figure 12. Wi-Fi procedure flowchart

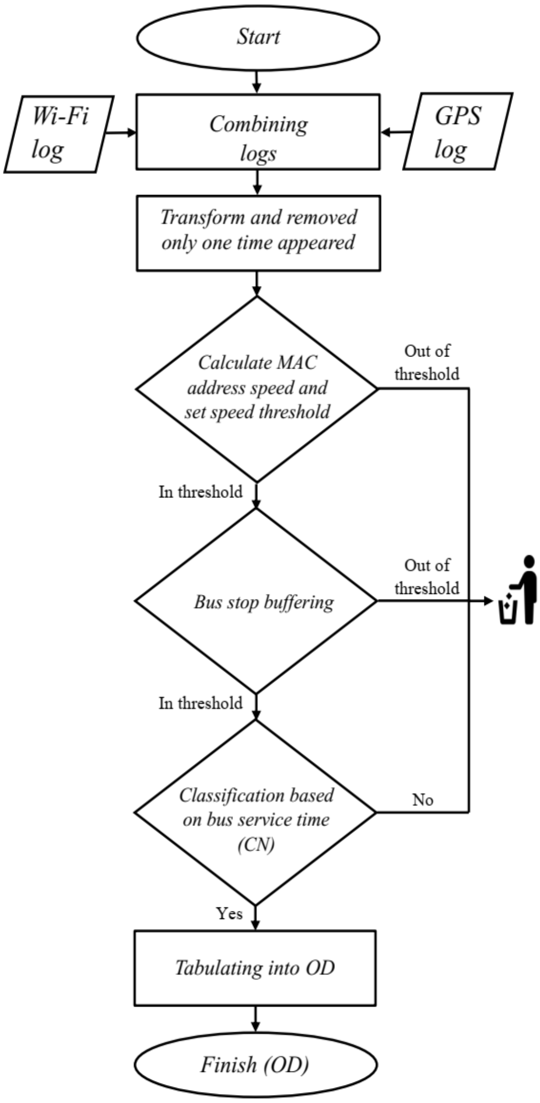


Figure 13. Wi-Fi procedure flowchart with deciding threshold

4.2.1. Wi-Fi and GPS Logs

The first step of the data processing procedure combines the Wi-Fi and GPS log data. Merging these two logs allows the location coordinates of each MAC address captured by the Wi-Fi scanner to be determined.

The merging technique uses the pandas concat module from the pandas data-analysis Python library (Pandas, 2019). The time column data in the GPS log were used as an index for merging the two datasets. Merged data were displayed on the GPS log, because GPS log data include time column data reported in units of data per second. The time column on the merged log provides information regarding the amount of time elapsed between individual MAC address pings throughout a passenger's bus trip.

The coordinate system of this data is decimal degree. The data are then converted from decimal degree coordinates to UTM (meters system), enabling the calculation of distance and speed for each MAC address. The processing procedure is accomplished using the geopandas Python module.

After the coordinate data have been converted, the converted data are input into the QGIS software package. On a QGIS map, the Wi-Fi data points can be observed and are in alignment with the bus route.

4.2.2. Speed Calculation

In this research, two types of procedures were conducted. The first experiment used a direct speed estimation process. The direct speed implied the MAC address speed was faster than the bus speed. The direct speed was calculated based on the location where the MAC address was initially detected. This process did not consider GPS log data that was acquired between MAC address pings. Therefore, the MAC address speed estimation was faster compared to the bus speed (Figure 14).

Furthermore, the direct speed calculated using MAC address pings was higher than the speed calculated using the ground truth approach (Hidayat et al., 2018). Therefore, it was necessary to improve the calculation procedure. The improved quantity was called the route speed (the second experiment). The route speed method calculated the MAC address speed based on the route recorded in the GPS log. This improved calculation implied that the MAC speed was the same as the bus speed.

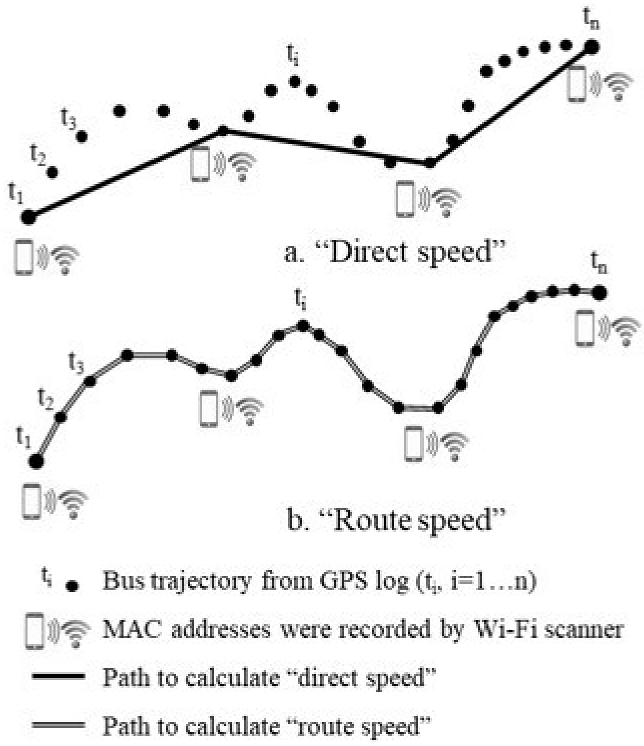


Figure 14. "Direct speed" and "route speed"

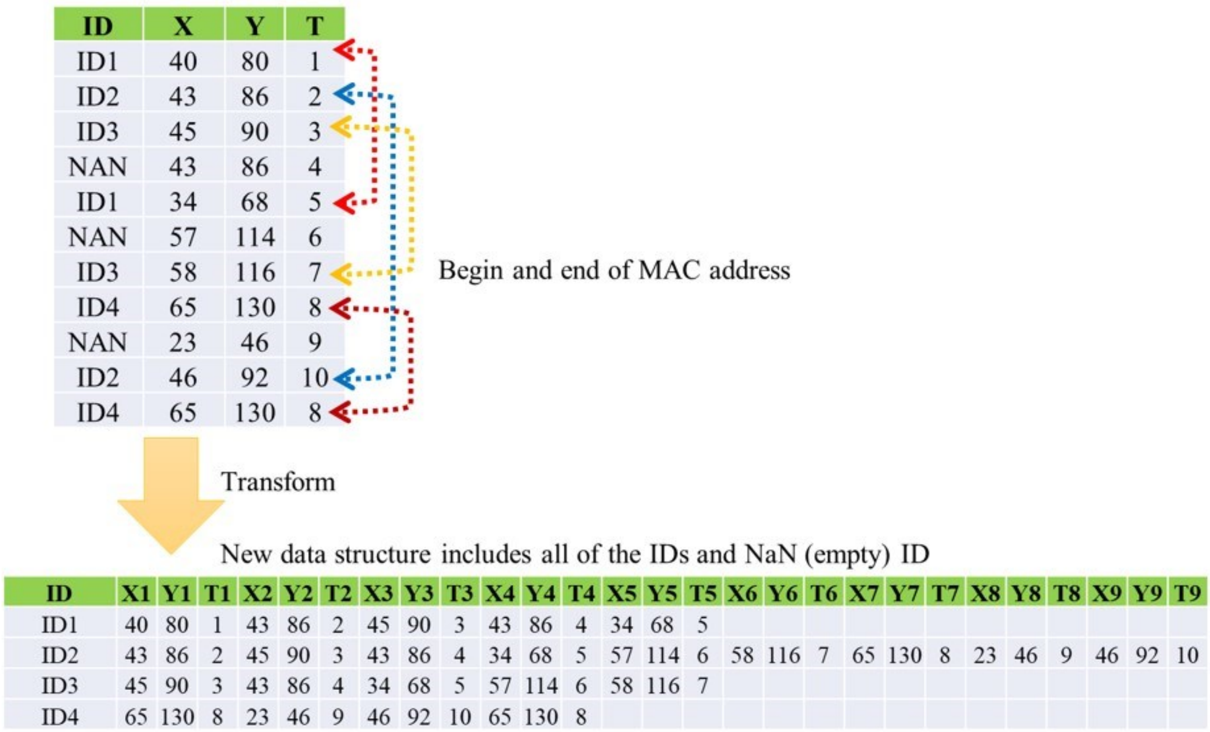
This route speed analysis can fit the speed calculated using the MAC address and GPS log data to the speed of the bus. Figure 14 shows the new procedure for analyzing Wi-Fi data based on the bus route, which is referred to as the route speed.

The route speed is the MAC address speed calculated based on the bus route, which is recorded in the GPS log. The GPS log significantly affects the calculation of the MAC address speed. Among the MAC address data, there is also time data that is not included with the MAC address and could be used to analyze the speed. This tracks the MAC address based on the route as well as the time. On this basis, the author considers the route speed as the best way to calculate the MAC address speed.

Before calculating the route speed, the date must be transformed. This data transformation procedure, shown in Figure 15, aims to identify and sort the data between the beginning and end of the MAC address detection. The analysis includes the time and coordinates from the

GPS log data that do not have MAC address information (Not a Number (NaN) data).

This transformation populates the GPS log data between MAC address appearances. The data transformation was an advanced analytical step for the Wi-Fi data cleaning procedure, which was accomplished using the pandas iloc dataframe Python module. In Figure 15, ID, X, Y, and T represent the MAC address, longitude, latitude, and time, respectively.



ID	X	Y	T
ID1	40	80	1
ID2	43	86	2
ID3	45	90	3
NAN	43	86	4
ID1	34	68	5
NAN	57	114	6
ID3	58	116	7
ID4	65	130	8
NAN	23	46	9
ID2	46	92	10
ID4	65	130	8

ID	X1	Y1	T1	X2	Y2	T2	X3	Y3	T3	X4	Y4	T4	X5	Y5	T5	X6	Y6	T6	X7	Y7	T7	X8	Y8	T8	X9	Y9	T9
ID1	40	80	1	43	86	2	45	90	3	43	86	4	34	68	5												
ID2	43	86	2	45	90	3	43	86	4	34	68	5	57	114	6	58	116	7	65	130	8	23	46	9	46	92	10
ID3	45	90	3	43	86	4	34	68	5	57	114	6	58	116	7												
ID4	65	130	8	23	46	9	46	92	10	65	130	8															

Figure 15. Example data transformation

The speed of each MAC address is intended for cleaning data based on the average speed of each MAC address. An analysis of the MAC address speed is essential for equating the MAC address speed to the bus speed. The equation for calculating the MAC address speed is shown below.

$$V_i = \frac{\sqrt{(X_{i+1}-X_i)^2+(Y_{i+1}-Y_i)^2}}{T_{i+1}-T_i}. \tag{1}$$

where:

i = Time frame of data acquisition

V_i = Speed at the i-th time frame

X_i = Longitude for the i-th MAC address

Y_i = Latitude for the i-th MAC address

T_i = Time for the i-th MAC address

4.2.3. Moved Categorization

This data processing step categorizes each MAC address measurement as “moved” or “stopped.” The speed threshold for categorizing a MAC address as “moved” is 1 km/h. Therefore, if a MAC address has moved faster than 1 km/h, then the MAC address is categorized as “moved,” and if the speed of a MAC address is measured below the 1 km/h threshold, the MAC address is categorized as “stopped.” The 1 km/h threshold is based on the average of the MAC address and bus speeds. This research was conducted by trial and error. If the threshold is set less than 1 km/h, the “moved” categorization is overestimated, and if the threshold is set greater than 2 km/h, it is underestimated.

This research used a preliminary threshold of > 1 km/h to justify the “moved” and “stopped” categorization. Thresholds are used to view the change trends of MAC address data. The threshold is applied as 1 km/h because it considers the bus speed. The 1 km/h threshold is sufficiently moderate to be used for further analysis because the Obuse Romango bus is a hop on–hop off bus and the route is quite narrow. These factors make the Obuse Romango bus slower than a common bus. If the threshold is not set, all the unique MAC addresses are considered as passengers and nonpassengers (moved and stopped classification). If the threshold is set as 1 km/h for the moved classification, then speeds in the range $0 < Vi < 1$ km/h are classified as stopped and those satisfying 1 km/h $\leq Vi$ are classified as moved.

4.2.4. Bus Stop Zoning, Time Spent at Intersection, and Analysis of MAC Addresses in the Vicinity of Bus Stops

In the bus stop zoning processing step, data are buffered around each bus stop using QGIS. The buffering zone radius surrounding a bus stop is approximately 10 m.

Based on the 2018 data, the distance of raw data is mostly near or around the bus stop. The distance between raw data is 0–50 m. Figure 16 and Figure 17 show the data from 13 and 14 October 2018, in terms of the distance between MAC addresses and bus stops.

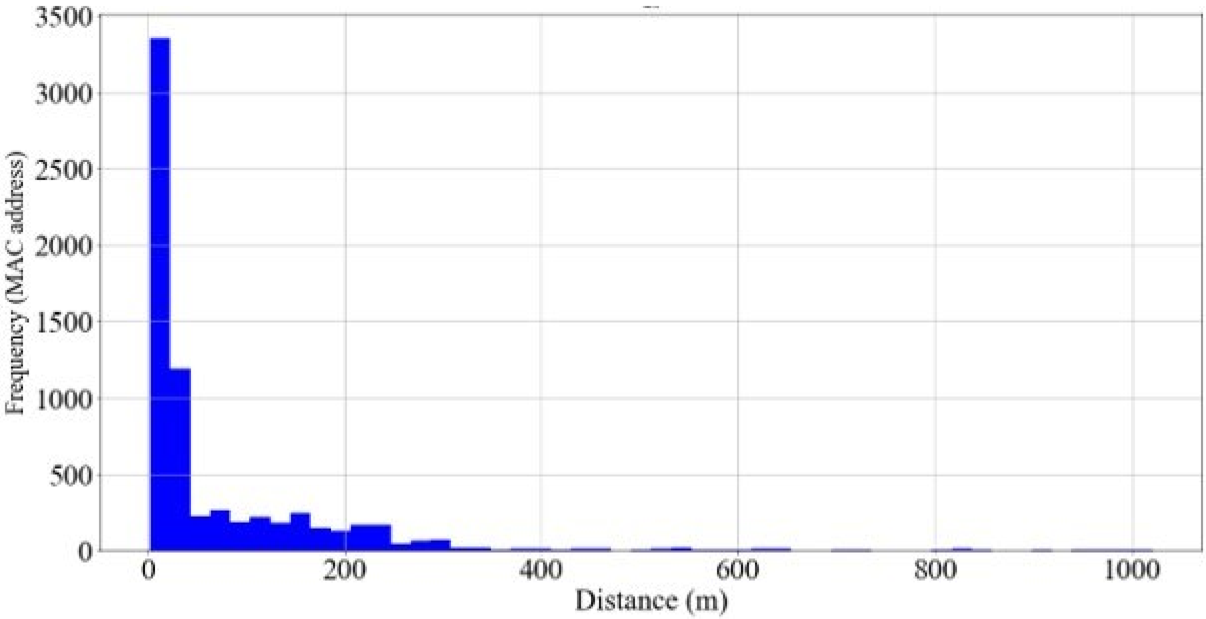


Figure 16. Distance between MAC addresses and bus stops on 13 Oct 2018

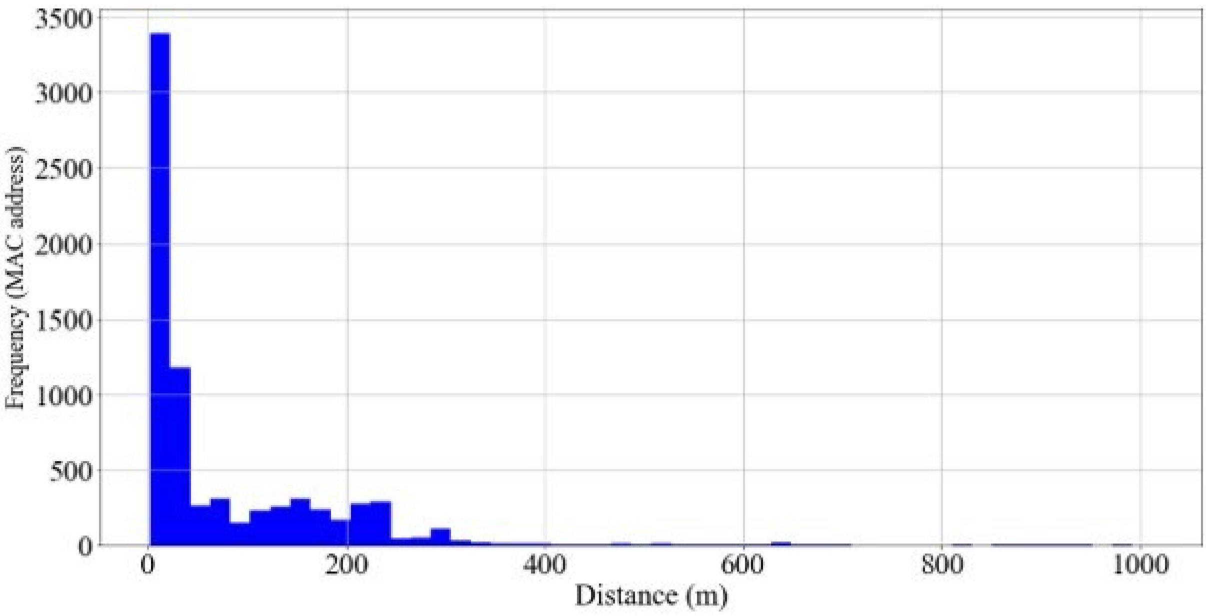


Figure 17. Distance between MAC addresses and bus stops on 14 Oct 2018

The aim here is to determine how far people queue or stand around the bus stop. However, if the radius is too large, it will include people not only around the bus stop also outside the bus stop. If it is too small, it will be inappropriate for data around the bus stop.

This zone, surrounding the bus stop, is used for intersection analysis in conjunction with the time data from the GPS log. This zoning procedure analyzes the time of bus movement recorded at the bus stop or around the bus stop.

The intersection is used to determine the presence of a MAC address at the bus stop. The GPS time log data and the intersection around the bus stop are used as the basic data for classifying the time data of a MAC address that is either at the bus stop or outside of the bus stop. This processing step is used to determine passenger volume for each MAC address. This analysis uses Python's classify function to classify the time within the buffer zone.

4.2.5. Classification Based on Route Circulation

In this final processing step, the data from the previous processing step were re-analyzed based on route circulation. The Romango bus has seven route circulations in one day, and MAC addresses that are detected in a circulation can be interpreted as traveling passengers. MAC addresses that are not in circulation can be eliminated from this dataset. This processing step uses Python's time classified program analysis.

4.3. Passenger Volume

4.3.1. Direct Speed Estimation Result

For the direct speed estimation, the results obtained after data processing generally show significant differences among the trends in the number of passengers in trip segments BS1-BS2-BS3, BS3-BS9-BS3, and BS3-BS2-BS1 and CNs. A CN is a circulation number that indicates the bus service route. The number of passengers for BS1-BS2-BS3 and BS3-BS2-BS1 is high, as these routes connect to the entrance of Obuse and provide mobility toward Obuse Station (BS2). These data may be overestimated owing to the relatively high incidence of nonpassenger data being detected from pedestrians or vehicles around the bus. Congestion and high traffic volume on the road caused the Wi-Fi scanners to detect or capture nonpassengers. On the contrary, the BS3-BS9-BS3 segment data were found to be relatively stable in every circulation. According to the data, the change in passenger volume between circulations was low; this is because the high travel volume in the morning is directed toward Obuse station. In the second through the fifth circulations, there was a higher level of passenger travel because many people travel in the time period of 10:00–15:00 and the number of passengers fell as the afternoon began.

- *CN1–6 (before noon)*

The estimated OD data shown in Figure 18 are based on CN1 to CN6 data from the morning until noon. The pattern of tourist travel is dominated by the museums (BS3, BS4), hot spring (BS6), garden (BS7), and historical places (BS8, BS9). Based on the data, the variation was high for segments S1-2 to S2-3 and from S3-2 to S2-1, and there was high passenger activity on BS2 and BS1. Medium variation occurred in segments from S3-4 to S6-9 and from S9-5 to S4-2, indicating stable movement. There was low variation in S6-9, indicating that the tourist attraction destinations, such as temples, are long distances away from the station.

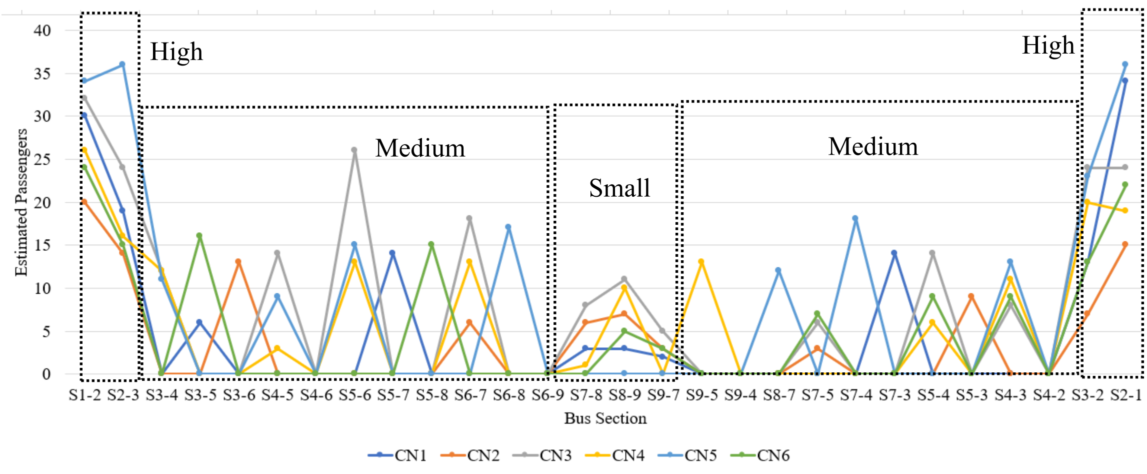


Figure 18. Estimated number of passengers for travel data from CN1–6 (AM)

- *CN7–14 (after noon)*

The estimated OD data in Figure 19 for CN7–14 in the afternoon show almost the same pattern as that in the morning. The relative movement in segments from S3-4 to S4-2 are consistent with the morning data, showing that bus users start to spread evenly throughout the bus stops to reach the tourist destinations. However, compared with the morning travel data, the movement in segments from S1-2 to S3-4 and from S4-2 to S2-1 were higher in the afternoon owing to the end of the visiting periods.

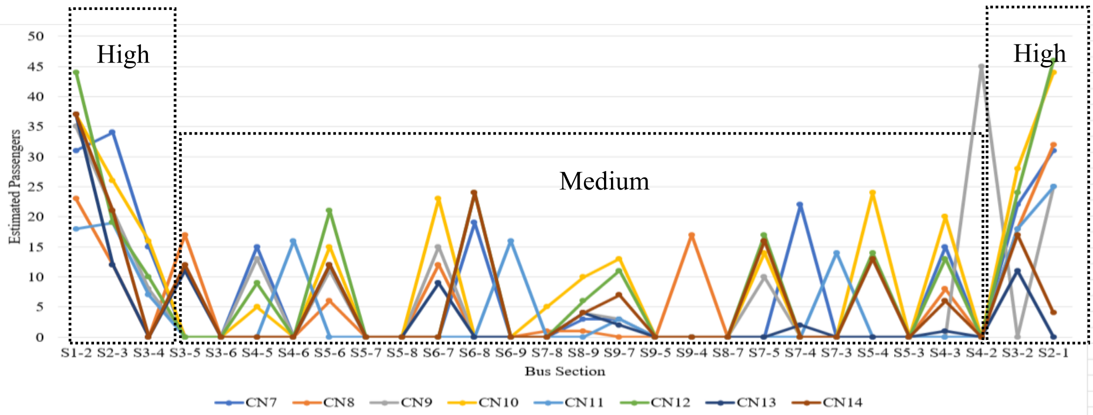


Figure 19. Estimated number of passengers for travel data CN7–14 (PM)

4.3.2. Route Speed Estimation Result

The passenger volume is the total number of passengers onboard the bus between two bus stops. Figure 20 shows passenger volume results, indicating the estimated passenger volume in each section and each CN. For each CN, the bus circulates around the route from BS1 to BS9 and returns to BS1 but skipping BS8 and BS6.

A section (S) is a section between two bus stops. For example, S1-2 refers to the section between BS1 and BS2. Figure 20 shows the travel trend for each CN and section.

After the improved calculation (route speed), for the 2016 data, Figure 20 shows a small number of passengers in CN1, CN4, and CN7 because these represent the bus time service in the morning, at lunch time, and the end of the bus service time. CN1 is the first bus starting time and CN14 is the last time bus operating time, which is around 17:00. This is the time many tourism spots cease operation and people return home. Accordingly, the volume is lower than that of other CNs. CN4 is also lower, because it is in afternoon, and people tend not to use the bus during lunch or when resting after lunch. Therefore, the number of people going to tourism spots is lower than during other CNs. On the contrary, CN2, CN3, CN5, and CN6 show a slightly high number of passengers, because they are around 11:00–16:00, when many people travel to tourism spots by bus, and these CN have a high number of passengers around the

Obuse city center. CN3 and CN5 show quite similar patterns because of on-peak hours. These two CNs have more passengers compared to the other CNs, owing to the high volumes of passenger movement in the morning, between 9:00 and 11:00, compared to afternoon travel.

In terms of the difference between sections, S3-4 and S4-5 show the highest volume because they are around city center. S9-7 and S2-1 show the lowest volume because they are outside the city center. With respect to the places, S1-2, S2-3, S3-4, show S4-5 a very high number of passengers. BS1 is Obuse park (BS1), which is the entrance of Obuse city. BS2 is Obuse station and BS3–BS5 are city center of Obuse, which is a very high-density location. Bus stops BS6–BS9 are very far from the city center and the tourist spots are a temple and hot spring.

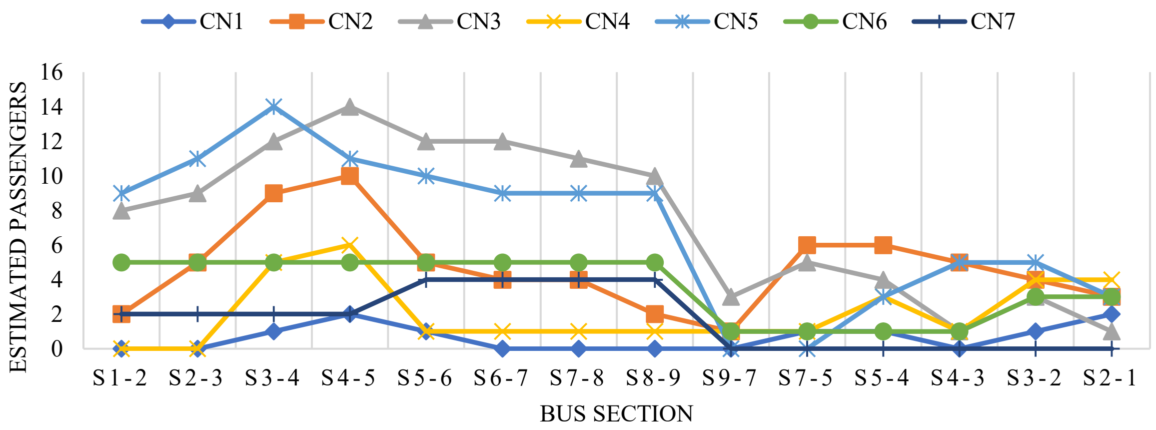


Figure 20. Passenger volume for sections (2016)

Figure 21 demonstrates that the data showed different travel trends. The highest passenger volumes were observed in sections S2-3, S3-4, S4-5, S5-6, S6-7, S9-7, S7-5, S5-4, and S4-3, whereas average passenger volumes were observed in sections S1-2, S6-7, S7-8, S8-9, and M2-1. Additionally, the route circulations with the highest volume of passengers were CN1, CN4, and CN5. Considering the time, CN2, CN4, and CN6 shows the high passenger volume because these CNs are at 11:00, 13:00, and 16:00, which are the best times to travel from one spot to another.

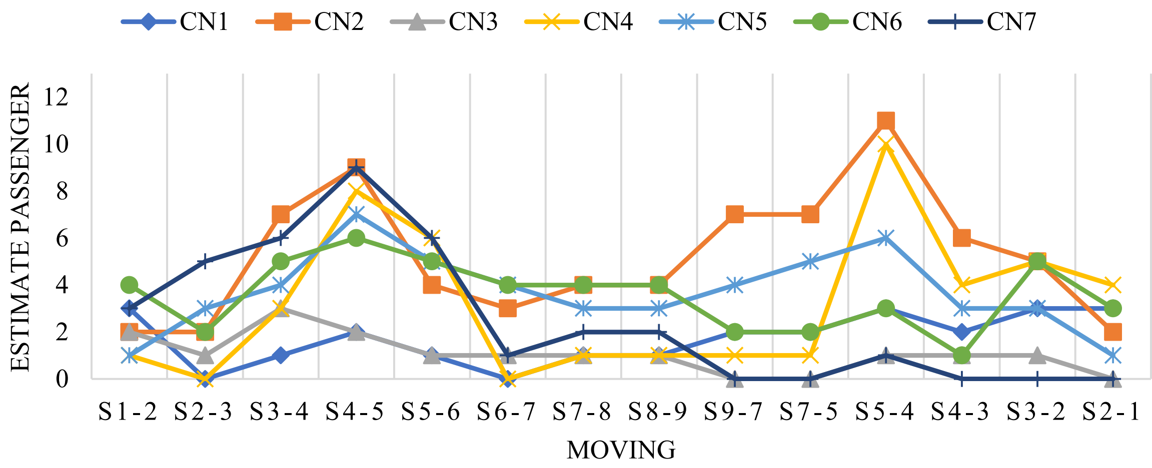
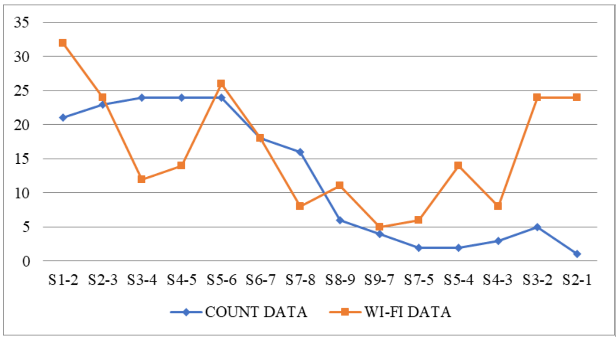


Figure 21. Passenger volume for sections (2017)

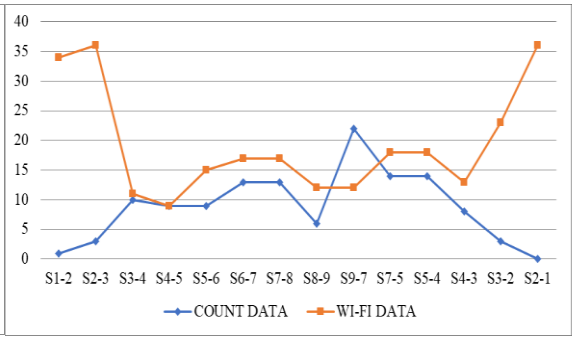
4.3.3. Direct Speed Comparison Result

For the first result, the observer count data was then compared with the simultaneously collected results of the proposed Wi-Fi-based method. The observer data include the number of passengers boarding and alighting at each bus stop. The manual count data were only available for CN3, CN5, CN8, CN10, and CN12.

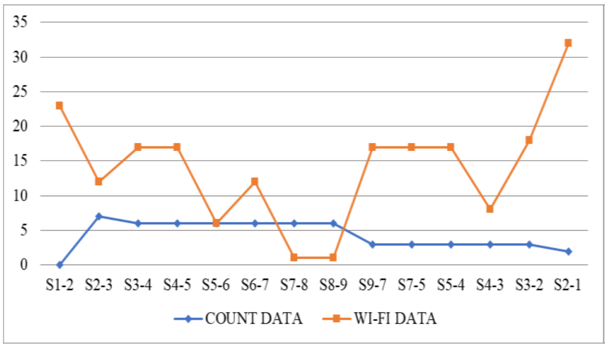
Figure 22 shows the difference between the manually counted (observer) data and Wi-Fi data in five circulations: CN3, CN5, CN8, CN10, and CN12. In general, the results show reliable Wi-Fi data compared with the observer data for CN3 in sections S2-3, S5-6 to S7-5, and S4-3; for CN5 in segments from S3-4 to S4-3; for CN8 in segments from S5-6 to S8-9; for CN10 in segments S4-5 and S7-8; and for CN12 in sections S3-4, S4-5, and S8-9. A striking difference between the manually counted and Wi-Fi data, indicating unreliable data, was found for the other segments listed above.



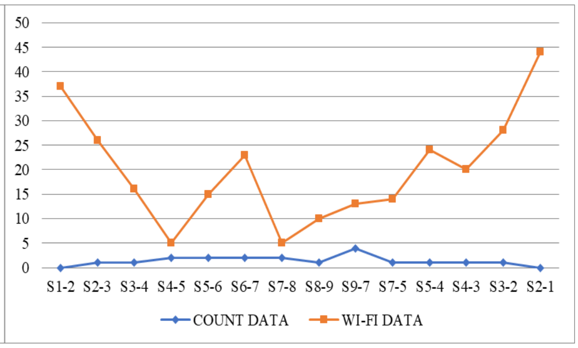
a) CN3



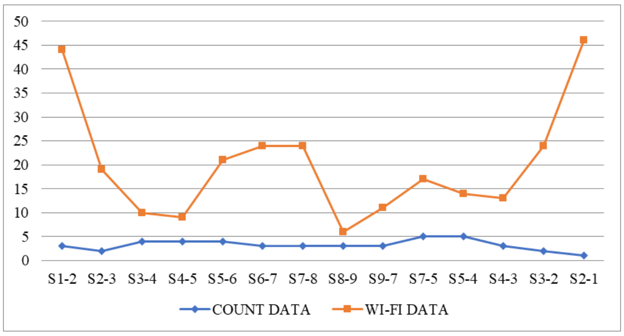
b) CN5



c) CN8



d) CN10



e) CN12

Figure 22. Comparison between the Wi-Fi data and observer data

The data were analyzed based on a correlation analysis to determine which Wi-Fi data were related and which were unrelated in each circulation by comparison with the observer data. The correlation analysis of the Wi-Fi data compared with the observer data for CN3 yielded an R-value of 0.390, indicating a low (positive) correlation. The CN5, CN8, CN10, and CN12 data had R-values of 0.671, 0.639, 0.633, and 0.530, all of which indicate relatively high (positive) correlations. Based on the correlation analysis, the observer data and Wi-Fi data generally have a correlation line indicating they are related, because there was such a small difference between the two datasets.

However, if the data are split based on the movement between each bus stop, there is a small difference between the observer and Wi-Fi data that can be considered reliable. All the circulation data for BS1 and BS2 were associated with a high traffic volume, as it is the main route in Obuse; this is because the main activity center is Obuse station (BS2) and the tourist entrance is at Obuse park (BS1). CN3, CN5, CN8, CN10, and CN12 at BS4, BS5, BS6, and BS7 is a daytime route that most pedestrians tend to walk in order to see some of the interesting places along the route. To assess the details of each circulation and each bus stop, the observer data and Wi-Fi data were analyzed according to the ratio of Wi-Fi over estimated (RWOE) to the observer data. The values of RWOE greater than 8 indicate unreliable route segments. The threshold of RWOE > 8 was used based on the author's assumption that if the data difference was in the range 1–7, it was acceptable. The RWOE analysis was employed to illustrate the differences between the Wi-Fi data (MAC address) and the traffic counting. According to Table 4, CN3, CN5, and CN8 had more reliable data than C10 and C12. For more details, please refer to Table 4.

Table 4. Comparison of observer and Wi-Fi data for each trip segments in five circulations

S	**CN3**			**CN5**			**CN8**		
	MCD	**WD**	**RWOE**	**MCD**	**WD**	**RWOE**	**MCD**	**WD**	**RWOE**
S1-2	21	32	11.0	1	34	33.0	0	23	23.0
S2-3	23	24	1.0	3	36	33.0	7	12	5.0
S3-4	24	12	-12.0	10	11	1.0	6	17	11.0
S4-5	24	14	-10.0	9	9	0.0	6	17	11.0
S5-6	24	26	2.0	9	15	6.0	6	6	0.0
S6-7	18	18	0.0	13	17	4.0	6	12	6.0
S7-8	16	8	-8.0	13	17	4.0	6	1	-5.0
S8-9	6	11	5.0	6	12	6.0	6	1	-5.0

S	CN3			CN5			CN8		
	MCD	**WD**	**RWOE**	**MCD**	**WD**	**RWOE**	**MCD**	**WD**	**RWOE**
S9-7	4	5	1.0	22	12	-10.0	3	17	14.0
S7-5	2	6	4.0	14	18	4.0	3	17	14.0
S5-4	2	14	12.0	14	18	4.0	3	17	14.0
S4-3	3	8	5.0	8	13	5.0	3	8	5.0
S3-2	5	24	19.0	3	23	20.0	3	18	15.0
S2-1	1	24	23.0	0	36	36.0	2	32	30.0

S	CN10			CN12		
	MCD	**WD**	**RWOE**	**MCD**	**WD**	**RWOE**
S1-2	0	37	37.0	3	44	41.0
S2-3	1	26	25.0	2	19	17.0
S3-4	1	16	15.0	4	10	6.0
S4-5	2	5	3.0	4	9	5.0
S5-6	2	15	13.0	4	21	17.0
S6-7	2	23	21.0	3	24	21.0
S7-8	2	5	3.0	3	24	21.0
S8-9	1	10	9.0	3	6	3.0
S9-7	4	13	9.0	3	11	8.0
S7-5	1	14	13.0	5	17	12.0
S5-4	1	24	23.0	5	14	9.0
S4-3	1	20	19.0	3	13	10.0
S3-2	1	28	27.0	2	24	22.0
S2-1	0	44	44.0	1	46	45.0

S : Bus section
CN : Circulation number
MCD : Manual count data
WD : Wi-Fi data
RWOE : Ratio wi-fi over estimated

4.3.4. Route Speed Comparison between Estimated and Ground Truth

- *2016- and 2017-year data*

This section presents the comparison between estimated passenger volume for sections and ground truth volume based on the route speed analysis for the 2016 and 2017 data. The ground truth data refers to the manual counting survey inside the bus. The author tried to count the passengers manually in CN2 and CN3. In Figure 23, only the total number of passengers in CN2 and CN3 was appeared to compare the estimated passenger and ground truth data.

Figure 23 demonstrates that the method calculates fewer passengers than the ground truth. The calculated number of passengers determined using the estimation method closely resembles the tendency of the ground truth. The speed calculation based on the route speed is effective to clean the Wi-Fi data, because the speed considers the GPS log data along the bus route.

Section S2-1 is the only section where the number of passengers identified using the method is greater than the number of passengers from the ground truth. The differences between the ground truth data and the estimated value show the extent of the underestimation. Sections S1-2, S2-3, S3-4, S5-6, S6-7, S7-8, and S9-7 have large underestimations (more than 10 passengers). Section S4-5, S8-9, S7-5, S5-4, S4-3, S3-2, and S2-1 have medium gaps (less than 10 passengers). These underestimations occur because for the speed calculation procedure, the author assumed that not all passengers may have a Wi-Fi-enabled device and not all the passengers turn on their Wi-Fi device.

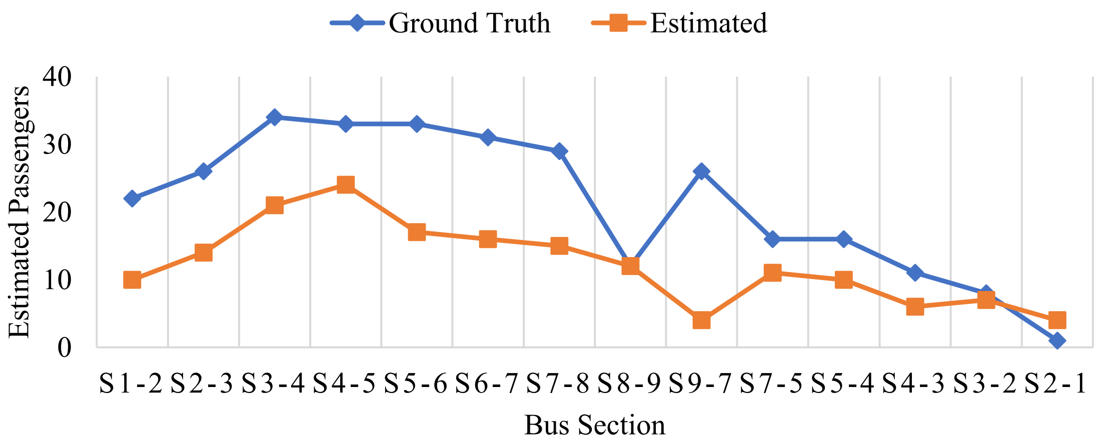


Figure 23. Comparing ground truth and estimated passenger volume for sections (2016)

There is still a bias between the Wi-Fi estimated and ground truth data. The trend from Figure 23 reveals that estimated passenger procedure results are closer to the ground truth data for the 2016 data. The correlation results show that the estimated passenger procedure and ground truth have a high correlation, with a factor of 0.78. This research method gives yields similar passenger counts to the ground truth data, and an accurate estimate could be achieved after determining a correcting ratio. This is an excellent opportunity for further research to improve the analysis.

Figure 24 demonstrates that the Wi-Fi procedure calculates fewer passengers than the ground truth method for the 2017 data. The number of estimated passengers in sections S2-3 to S4-3 is less than the number of passengers determined using the ground truth procedure. The number of passengers determined using the Wi-Fi method shows a similar trend to the ground truth data. The passenger volume determined using the Wi-Fi method is higher than the ground truth data in sections S1-2, S3-2, and S2-1, whereas the passenger volume determined using the Wi-Fi method is less than the ground truth in sections S2-3 and S4-3. There is still a bias between the direct and route speed procedure and the ground truth data. The result must be underestimated, because not all passengers use Wi-Fi devices or turn on the Wi-Fi. The trend from Figure 24 reveals that the route speed procedure result tendency is quite similar to the ground truth data.

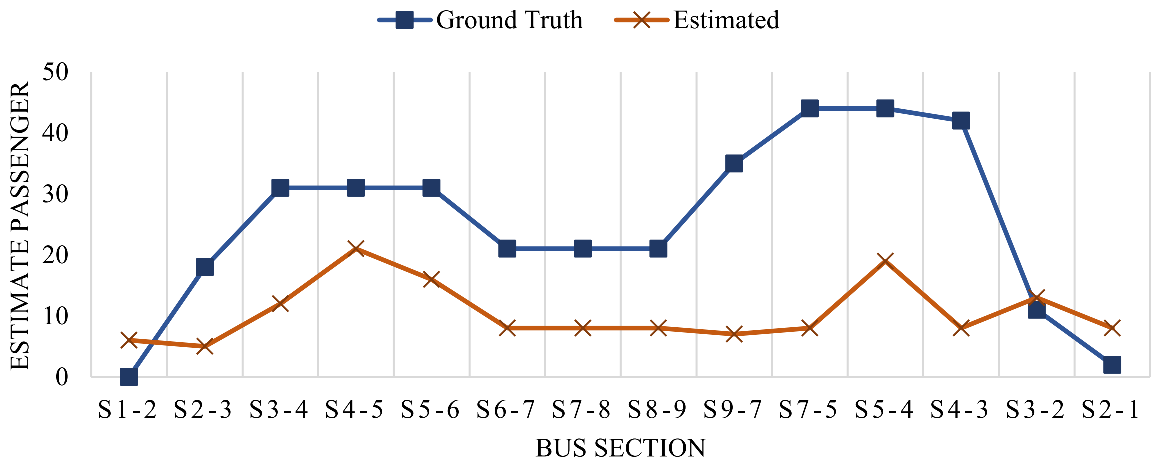


Figure 24. Comparing ground truth and estimated passenger volume for sections (2017)

However, because of the underestimated passenger volume, a magnification ratio is required to render the estimated passenger volume very close to the ground truth data.

- *2018-year data*

The results for 2018 are shown in Figure 25. There are four lines in the figure, which are the ground truth data (GT) and estimated Wi-Fi for 13–14 October 2018. For 13 October 2018, the ground truth data and Wi-Fi estimate have a similar tendency passenger flow pattern. The high passenger volume occurs in S2-3 and S3-2 and the low passenger volume is in S8-9. Similar to 13 October 2018, the 14 October 2018 data show the same passenger flow pattern.

The passenger volume estimated from Wi-Fi data is quite similar to the ground truth data. The estimated volume shows the same pattern as the ground truth data. These data can be transformed using magnification ratio to make the estimated close to the ground truth data.

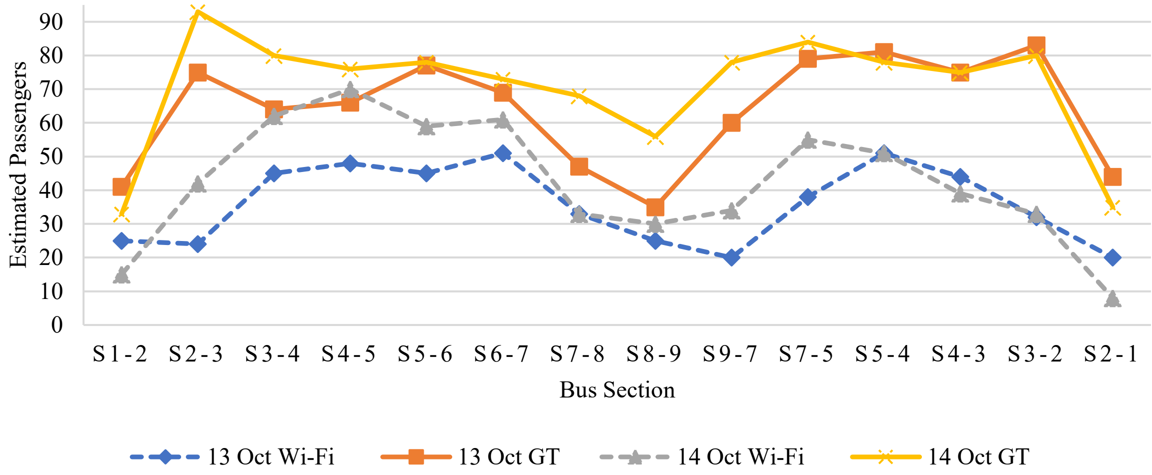


Figure 25. Comparing ground truth and estimate passenger volume for sections (2018)

4.3.5. Magnification Ratio

A magnification ratio was applied to the estimated passenger volume to make it close to the ground truth data. In this case, the author used a ratio of 50% for the 2016, 2017 and 2018 data. The results are shown in Figure 26 and Figure 27.

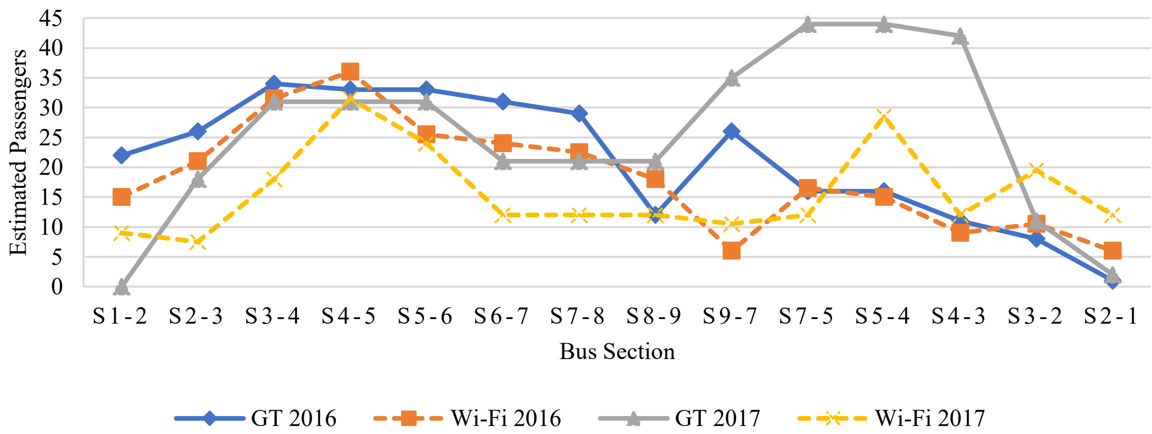


Figure 26. Comparing ground truth and estimated passenger volume for sections (2016 and 2017) with magnification ratio 50%

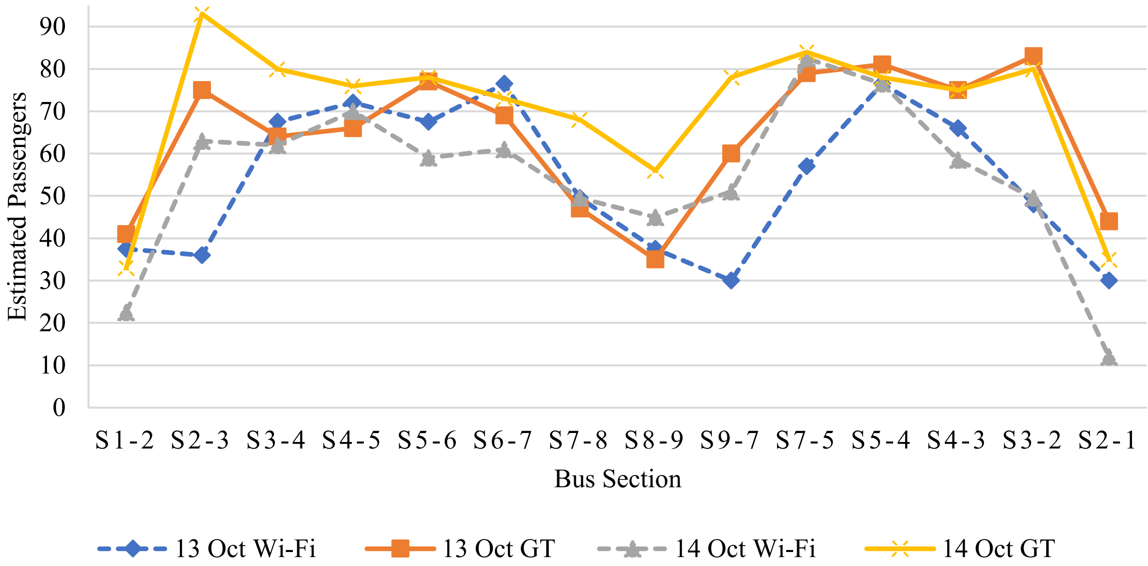


Figure 27. Comparing ground truth, estimate passenger volume for sections (2018) year with magnification ratio 50%

For the 2016 data, the correlation factor is 0.77 and for the 2017 data, the correlation factor is 0.37. The correlation is relatively high, and the tendency of the estimated passenger volume is similar to that of the ground truth data. For the 2018 data, on 13 October 2018, the correlation is 0.57 and on 14 October 2018, the correlation is 0.86. This means that the correlation between the ground truth and estimated passenger volume is high, and they show a similar pattern.

4.4. Origin–Destination

4.4.1. Estimation Result

The OD matrix data (Figure 28) is the result of the previously described data analysis and processing procedures conducted on raw data. The data cleaning procedure was able to produce OD matrix data for the field surveys conducted in 2016 and 2017 (Figure 28). In 2016, significant passenger movements were observed between BS1-BS5, BS1-BS9, BS2-BS8, BS3-

BS2, BS3-BS4, BS3-6, BS4-BS5, BS7-4, BS7-5, and BS7-6. These bus stops are near tourist attractions located within Obuse. The increased frequency of visitors or passenger movements is correlated to tourist attractions in the city center of Obuse. For the field survey conducted in 2017, the OD data is more prominent between BS1-BS2, BS1-BS9, BS4-BS5, BS4-BS6, BS5-BS4, and BS7-BS9. The chord diagram also reveals contrasting bus passenger movement patterns between the two field surveys conducted in 2016 and 2017.

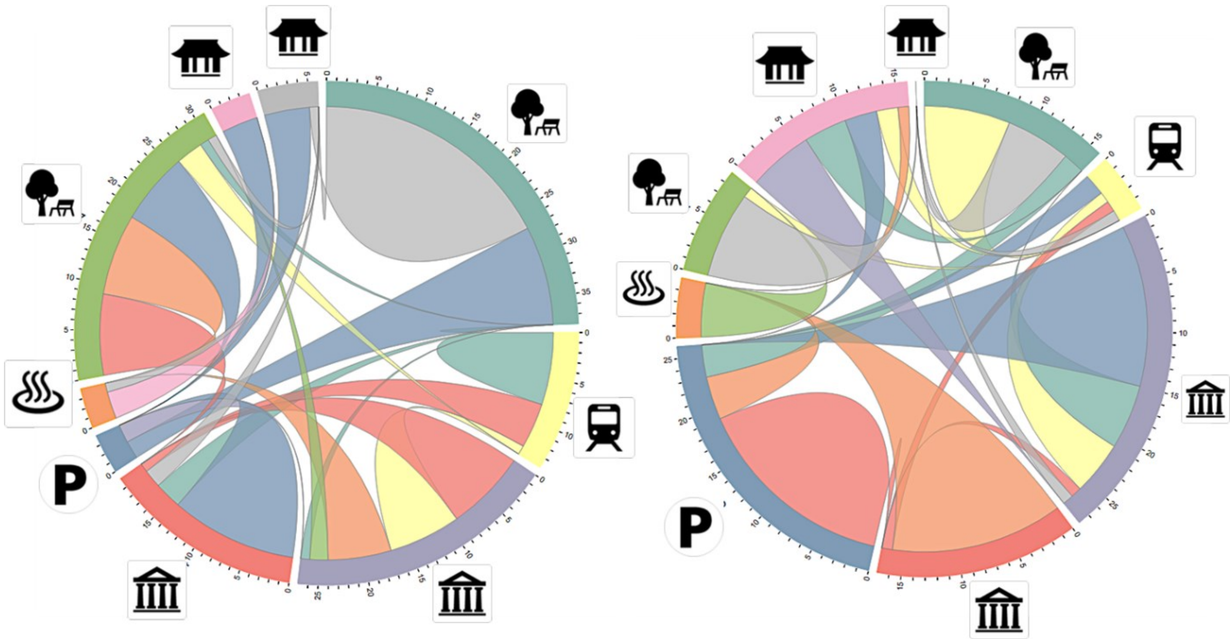


Figure 28. OD matrix results for field surveys conducted in 2016 and 2017

4.4.2. Correlation and Threshold Justification

For each dataset, we apply 21 speed thresholds, which range from greater that 0 km/h to greater than 2 km/h, in increments of 0.1 km/h. For the bus stop buffer, we apply 50 values which are 10 m to 500 m in increments of 10 m. We process the data and construct estimated OD or boarding–alighting matrices for the 13 October and 14 October data. After processing the data, there are 1050 sets of estimated OD data. These estimated OD data are compared with the ground truth data and the correlation values are calculated.

For the 13 October 2018 data (Figure 29), the highest correlation between the estimated OD and the ground truth OD is approximately 0.7 for the speed >1.9 and 2.0 km/h with a buffer of 40 m. For the 14 October 2018 data (Figure 30), the highest correlation between the estimated OD and the ground truth OD is approximately 0.5 for the speed >1 with a buffer of 40 m. Other speed thresholds and bus stop buffer values yield lower correlations. This indicates that many nonpassenger (noise) data are still included in the estimated passenger volume.

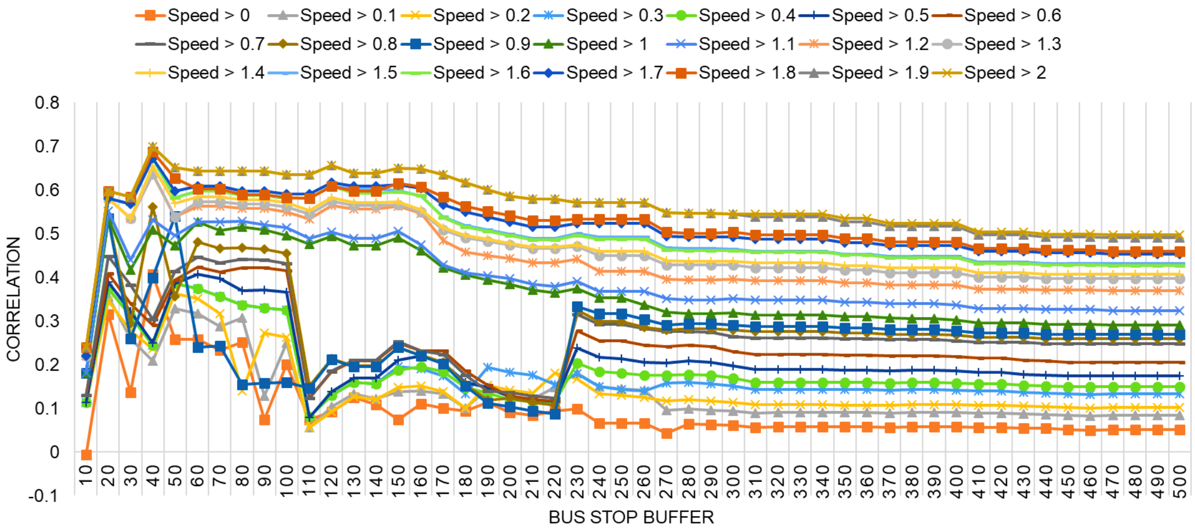


Figure 29. Correlation value between estimated OD and ground truth OD (13 Oct 2018)

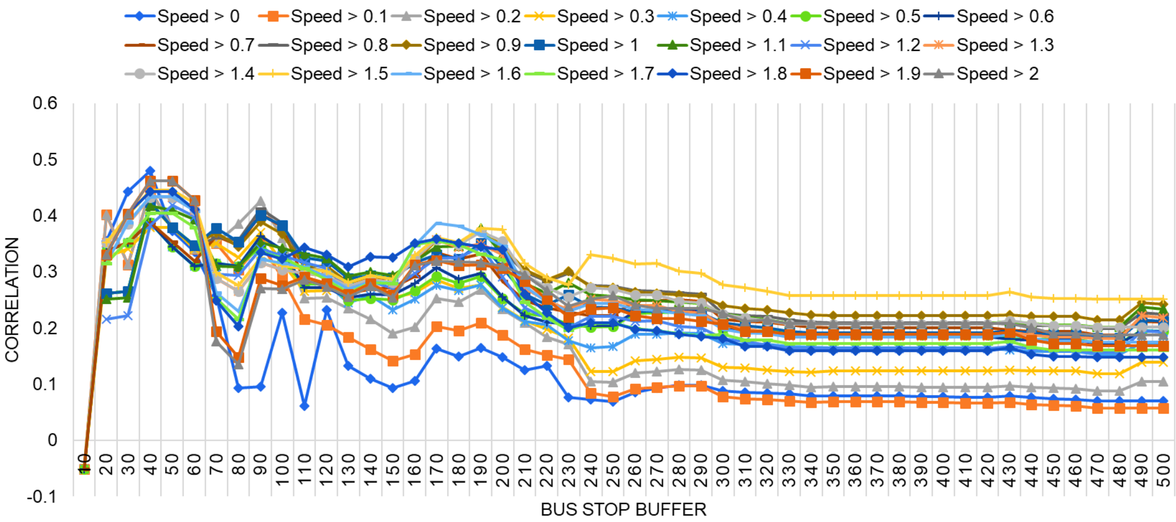


Figure 30. Correlation value between estimated OD and ground truth OD (14 Oct 2018)

4.5. Summary

This section summarizes the results from the small bus Wi-Fi application. A new processing or cleaning procedure was applied to clean the Wi-Fi data. The OD and passenger volume could be tabulated after the cleaning procedure. For the speed analysis, the direct speed result is overestimated compared to the route speed result. Naturally, not all the passengers use Wi-Fi devices or active the Wi-Fi function of their devices, which means that the passenger volume is underestimated. Therefore, the author applied a magnification ratio (50%) to make the estimated passenger close to the ground truth data. However, the estimated passenger volume and ground truth data show a similar trend. The estimated passenger is lower than the ground truth because of certain factors, such as the Wi-Fi on/off function and crowded locations.

Time, distance, and speed are important parameters to justify the moving MAC address. Bus stop buffering was applied to justify the nearest MAC address to the bus stop. The specific thresholds yielding the maximum correlations are as follows: for 13 October 2018, the threshold speed >1.9 km/h and the bus stop buffer is 40 m; for 14 October 2018, the threshold speed >1.0 km/h and the bus stop buffer is 40 m.

The magnification ratio and threshold are depending on cases. It is not universal value. It is more data needed to make a decision value. This research can measure passenger trip patterns. The procedure can be used by transportation planners, officers, and street planners to plan transit systems.

CHAPTER 5
PARATRANSIT APPLICATION

5.1. Introduction

The paratransit service has characteristics such as no stopping spot, no timetable, and irregular patterns, because stopping depends on the driver or passenger demand. This study uses Wi-Fi scanner equipment to capture MAC addresses of the paratransit passengers in Makassar, Indonesia.

The objectives of this chapter are 1) to produce a procedure to clean Wi-Fi raw data and remove nonpassenger data, 2) to match between ground truth data and Wi-Fi data, 3) to apply cleaning process based on dB (decibel) data, and 4) to construct OD or boarding–alighting data based on Wi-Fi estimation. Figure 31 and Figure 32 illustrate the matching and dB cleaning processes, respectively.

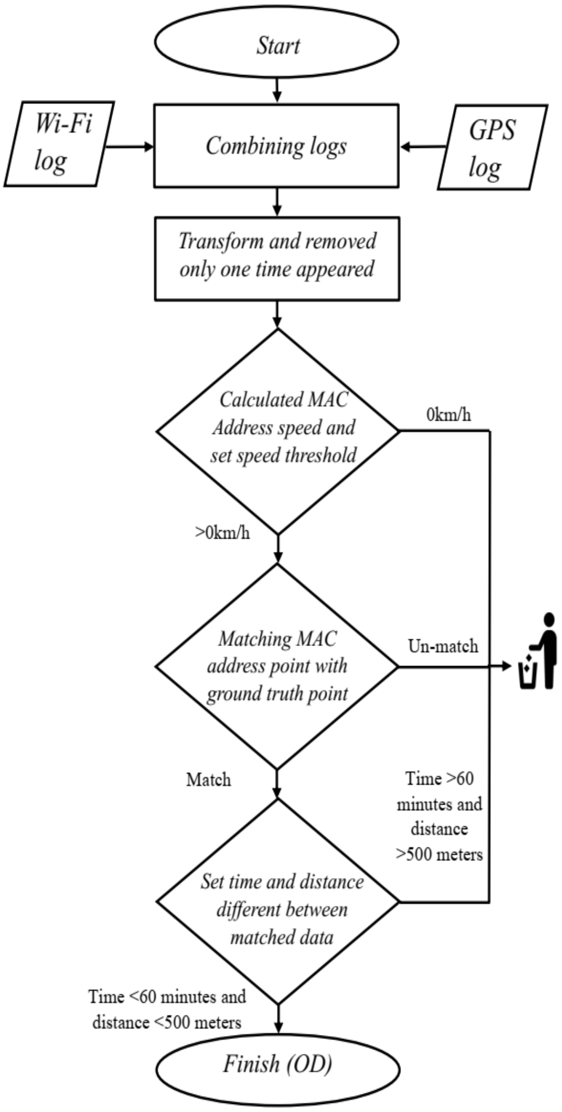


Figure 31. Flowchart processing data for matching process

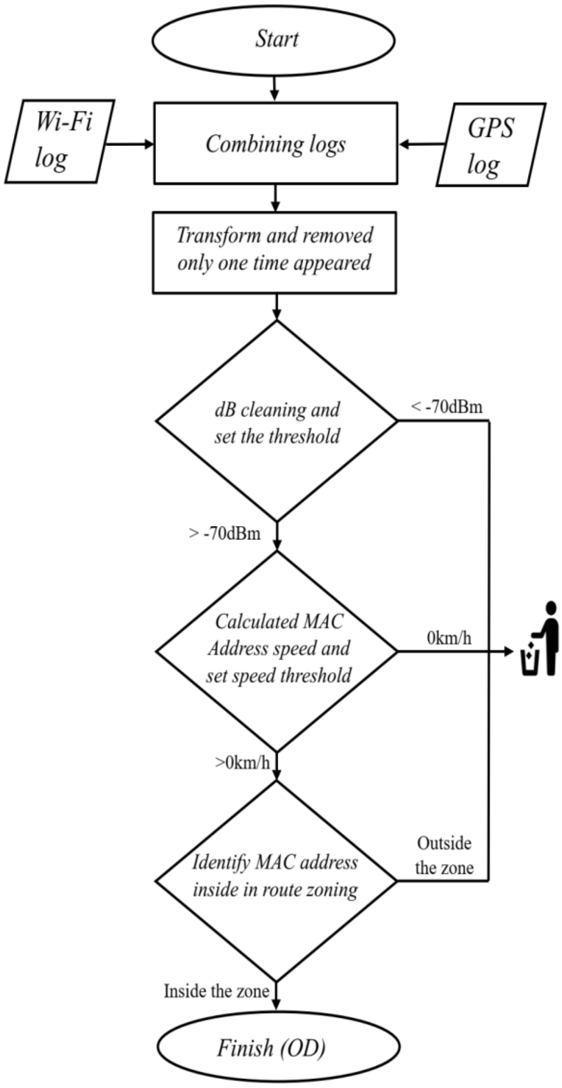


Figure 32. Flowchart for dB cleaning

5.2. Processing Data

5.2.1. Combining Log Data and dB Cleaning

This first step was to combine two raw datasets, i.e., the GPS and Wi-Fi log data. This step gave the MAC address log data its location coordinates. The GPS log data contain the following data structures: time, longitude (X) coordinate, and latitude (Y) coordinate. The Wi-Fi log data contain the following data structures: time and MAC address. The GPS dataset was the primary dataset in this merge, because it captured time and coordinate data once per second along the *petepete* route.

Figure 33 presents a visualization of the merge or combination method. It shows one mobile device MAC address on a *petepete* moving from (X_1, Y_1) at t_1 to (X_n, Y_n) at t_n, and the MAC address appears several times. However, between each MAC address appearance, the GPS data are logged every second (X_i, Y_i, t_i). The index used to combine these two datasets was the time variable in the GPS log data (Figure 34).

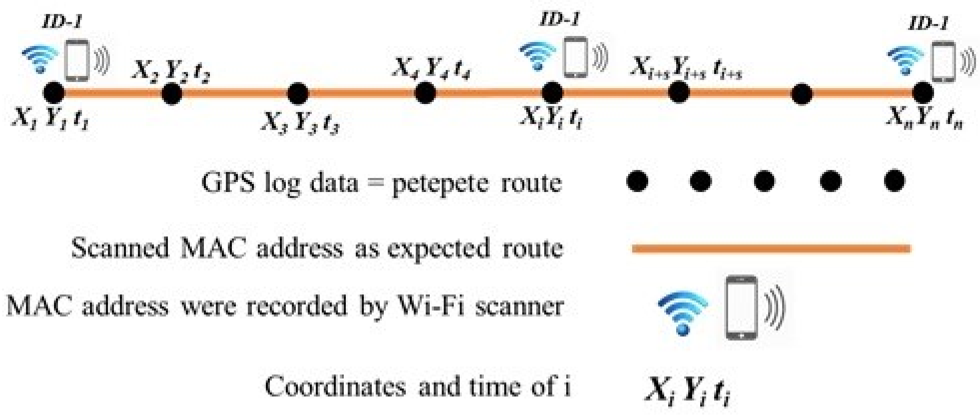


Figure 33. Visualization of the combination method

Merge Key

Time	Longitude (X)	Latitude (Y)
10:56:03	778932.6	9438371
10:56:04	778932.6	9438371
10:56:05	778921.5	9438360
10:56:06	778910.4	9438360
....	....	....

GPS log

Merge

Time	MAC
10:56:03	ced8448xxx
10:56:06	723404fxxx
......	......

WiFi log

Result

Time	MAC	Longitude (X)	Latitude (Y)
10:56:03	ced8448xxx	778932.6	9438370.58
10:56:04	NaN	778932.6	9438370.58
10:56:05	NaN	778921.5	9438359.56
10:56:06	723404fxxx	778910.4	9438359.6
....	......	....	....

GPSWiFi.csv

Figure 34. Merge of the GPS and Wi-Fi logs

After combining the Wi-Fi log and GPS log, there are two possibilities: 1) Some Mac address were occasionally captured by the Wi-Fi scanner at the same time or different time; 2) another possibility is no MAC address, i.e., not a number (NaN), which is captured by the Wi-Fi scanner at the time.

Figure 35 shows the raw data on a map. These points were based on X and Y coordinates. QGIS was used to create the map.

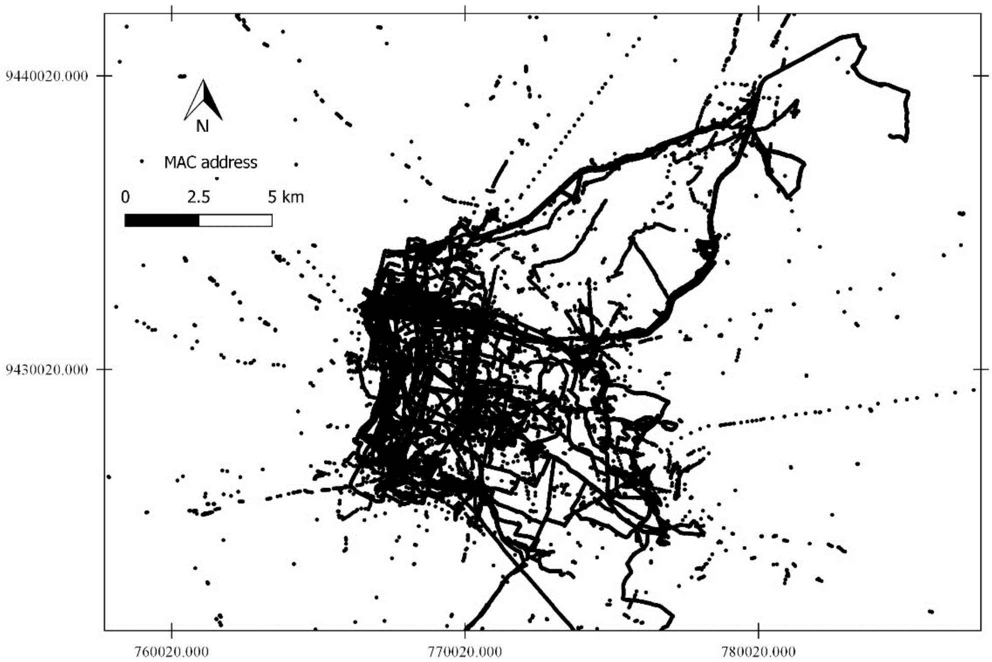


Figure 35. MAC addresses on a map

After combining, the data were cleaned based on the dB column data. This dB-based cleaning of the raw data is important, because many weak dB data are included in the research survey data. Figure 36 shows the distribution of dB raw data, which are mostly in the weak range of Wi-Fi devices. This is approximately −70 to −100 dBm. The weak range of Wi-Fi devices indicates nonpassenger data or noise data. Based on Table 5, it is better to use devices with Wi-Fi dB strength greater than −67 dBm. Therefore, the raw data were cleaned by purging the data with dB below than −70 dBm.

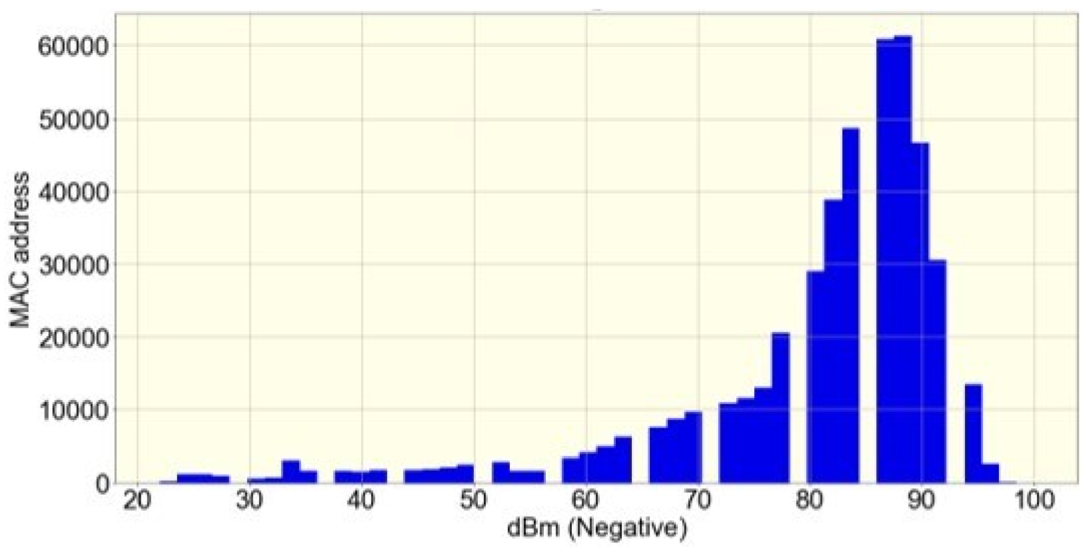


Figure 36. Distribution of dB raw data

Table 5. dB strength (Metageek, 2018)

Signal Strength	Annotation	Explanation	Required for
−30 dBm	Amazing	Maximum achievable signal strength. The client can only be a few feet from the AP to achieve this. Not typical or desirable in the real world.	N/A
−67 dBm	Very Good	Minimum signal strength for applications that require very reliable, timely delivery of data packets.	VoIP /Vo Wi-Fi, streaming video
−70 dBm	Okay	Minimum signal strength for reliable packet delivery.	Email, web
−80 dBm	Not Good	Minimum signal strength for basic connectivity. Packet delivery may be unreliable.	N/A
−90 dBm	Unusable	Approaching or drowning in the noise floor. Any functionality is highly unlikely.	N/A

This helps to adequately calculate the MAC address speed based on the route recorded in the GPS logs. The two datasets were merged and cleaned by using the Python Jupyter Notebook software package. After the merging and cleaning process, the resulting data file was labeled as GPSWi-Fi.csv.

5.2.2. Adding Coordinate Positions to Ground Truth Data

The second step was to add the longitude (X) and latitude (Y) from the GPS to the ground truth data based on the time of boarding and alighting. This step enhances the ground truth with location or coordinates. The OD ground truth data contain the following data structures: time, annotation (boarding and alighting), and passenger ID.

Passenger ID is an identification code that is given by the surveyor to identify each passenger anonymously. For example, in “18Device-5G,” “18” is a unique passenger number, “Device-5” represents Wi-Fi scanner number 5, and “G” is a route code for the *petepete* route.

The XY data were merged with the ground truth data using the time column as the merging index. The results of the merger are shown in Figure 37.

Merge Key

Time	Origin	Destination	Annotation	Passenger ID
9:53:13	Elementary School	Gas Station	Boarding	18Device-5G
9:54:20	Hotel	Market	Boarding	19Device-5G
----	----	----	----	----

OD Ground Truth

Time	Longitude	Latitude
9:53:13	772909.7	9435849
9:54:20	773410.7	9436312
----	----	----

GPS log

Result

Time	Origin	Destination	Annotation	Passenger ID	Longitude	Latitude
9:53:13	Elementary School	Gas Station	Boarding	18Device-5G	772909.7	9435849
9:54:20	Hotel	Market	Boarding	19Device-5G	773410.7	9436312
----	----	----	----	----	----	----

GTGPS.csv

Figure 37. Combination of ground truth data and GPS log

The Python Jupyter Notebook software package was used for this process of merging the two datasets. The resulting data file for this step was labeled as GTGPS.csv. Figure 38 shows ground truth data on a map, as they include the coordinate data.

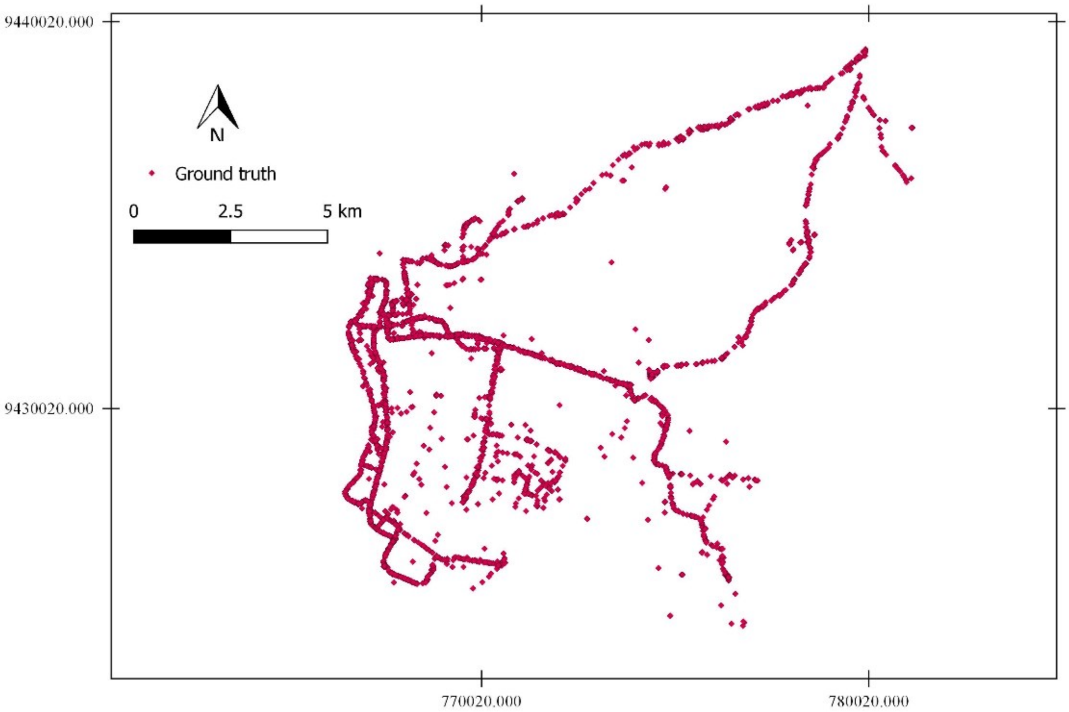


Figure 38. MAC addresses on a map

5.2.3. Wi-Fi Speed Calculation and "Moved" Categorization

The third step was to calculate the MAC address speed to classify the "moved" category. Before calculating the speed, the GPSWi-Fi.csv data had to be transformed. This analysis included time and coordinate data that did not have MAC address data (NaN).

This transformation considered the GPS log data between the appearance of MAC addresses. The data transformation in Figure 39 aimed to identify and sort the data between the beginning and end of MAC address detection.

Time	MAC	Longitude (X)	Latitude (Y)
10:56:03	ced8448xxx	778932.6	9438370.58
10:56:04	NaN	778932.6	9438370.58
10:56:05	NaN	778921.5	9438359.56
10:56:06	723404fxxx	778910.4	9438359.6
10:56:07	ced8448xxx	778910.4	9438359.6
10:56:08	NaN	778921.6	9438359.6
10:56:09	723404fxxx	778921.6	9438359.6
......	....	....	....

Result

MAC	X1	Y1	T1	X2	Y2	T2	X3	Y3	T3	X4	Y4	T4	X5	Y5	T5
ced8448xxx	779	9438	10:56:03	779	9438	10:56:04	779	9438	10:56:05	779	9438	10:56:06	779	9438	10:56:07
723404fxxx	779	9438	10:56:06	779	9438	10:56:07	779	9438	10:56:08	779	9438	10:56:09			

Figure 39. Example of data transformation (GPSWi-Fi_transform.csv)

This transformation was useful before calculating the MAC address speed, because the MAC addresses were sorted by X, Y, and T (between the first and the last appearance of the MAC address). The visualization of this speed calculation is presented in Figure 40. In Figure 40, ID is a MAC address, “X” is the longitude, “Y” is the latitude, and “T” is the time. Data transformation was an advanced analytical step for the Wi-Fi data cleaning procedure, which was accomplished using the pandas iloc dataframe Python module. The transformation dataset was labeled GPSWi-Fi_transform.csv.

After transforming the dataset, the next procedure was calculating the speed. An analysis of the MAC address speed was essential for equating the MAC address speed to the *petepete* speed (Equation 1). This procedure used GPSWi-Fi_transform.csv data. After speed calculation, the average of the speed for each MAC address was calculated. The equation used for calculating the MAC address speed is shown below.

$$V_i = \frac{\sqrt{(X_{i+1} - X_i)^2 + (Y_{i+1} - Y_i)^2}}{t_{i+1} - t_i} \qquad (2)$$

where:

i = Time frame of data acquisition

V_i = Speed of MAC address at the i-th time frame

X_i = Longitude of MAC address at the i-th time frame

Y_i = Latitude of MAC address at the i-th time frame

t_i = Time at the i-th time frame

The average of the speed is given by:

$$\bar{V} = \frac{N}{\sum \frac{1}{V_i}} \tag{3}$$

where:

$\bar{V}$ = Average of the speed

N = Total number of time frame (speed)

V_i = Speed at the i-th time frame

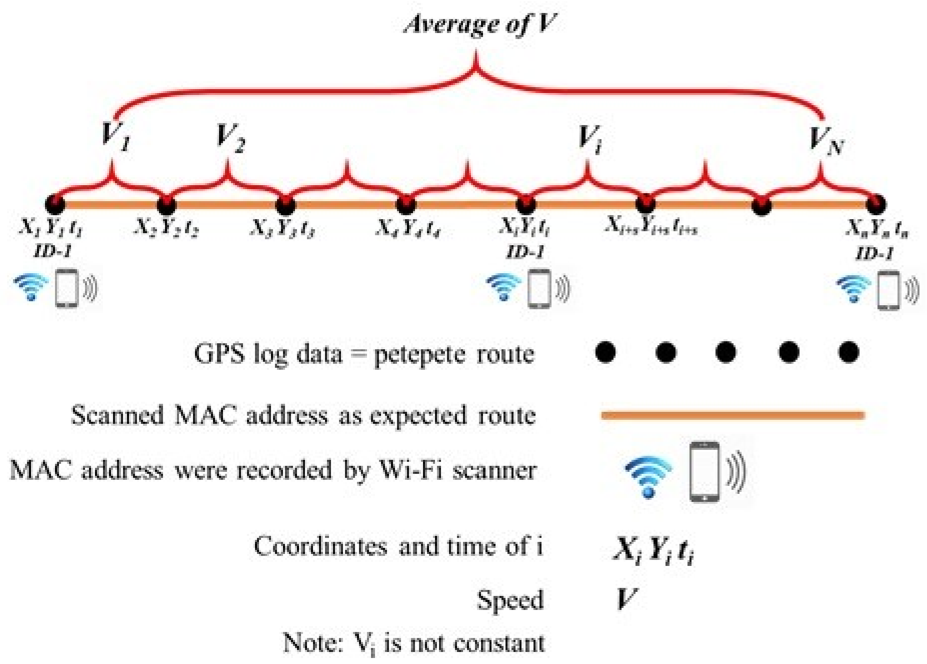


Figure 40. Speed calculation visualization

The next process was to categorize each MAC address as "moved" or "stopped" based on the average speed. The threshold for classifying a MAC address as "moved" was an average speed above 0 km/h. Therefore, if a MAC address had an average speed below 0 km/h, then the MAC address was categorized as "stopped." "Stopped" was defined as a nonpassenger or someone or something with a Wi-Fi device near the *petepete* when the *petepete* was operating along the route. The GPSWi-Fi_transform.csv data was sorted and assigned as "moved" and "stopped." MAC addresses in the "stopped" category could be deleted. This step used the Python Jupyter Notebook software package. The results of this analysis were labeled as Wi-Fimoved.csv.

5.2.4. Assign "First" and "Last" Appearance of MAC Address

The fourth step was to assign the "first" and "last" appearances of a MAC address. The Wi-Fimoved.csv data were used in this processing step for attaching and sorting the first and the last MAC address appearances. For each MAC address, this procedure preserved the first and the last times the address appeared (Figure 41).

MAC	Time	Average Speed (km/h)	X	Y
0005985xxx	11:40:19	1.2	773755	9436620
0005985xxx	11:45:01	1.2	773756	9436621
0005985xxx	11:59:21	1.2	773757	9436622
0005985xxx	12:02:03	1.2	773758	9436623
0005985xxx	12:10:19	1.2	773759	9436623
....	....	....	....	....

Result

MAC	Time	Annotation	Average Speed (km/h)	X	Y
0005985xxx	11:40:19	First	1.2	773755	9436620
0005985xxx	12:10:19	Last	1.2	773759	9436623
....	....	....	....	....	....

WiFi Moved_firstlast.csv

Figure 41. Assigning "first" and "last" MAC address

This step used the Python Jupyter Notebook software package. The data file resulting from this process was labeled as Wi-Fi Moved_firstlast.csv.

5.2.5. Distance between Wi-Fi Data and Ground Truth Data

The fifth step was to calculate the distance between the point of the MAC address and the boarding or alighting location of the real passenger. Let A stand for a dataset that contains the position of the MAC address from Wi-Fi Moved_firstlast.csv, and B stand for a dataset that includes the boarding or alighting point of the real passenger from GTGPS.csv. Every combination of points between A and B was calculated.

This analysis used the QGIS software package. GIS software is useful for analyzing the spatial relationship between two datasets. For this step, a nearest neighbor analysis was applied (QGIS, 2019). This analysis uses a tool in QGIS, which is called the distance matrix. The result provides the distance (in meters) between the MAC address and each passenger ID point (Figure 42).

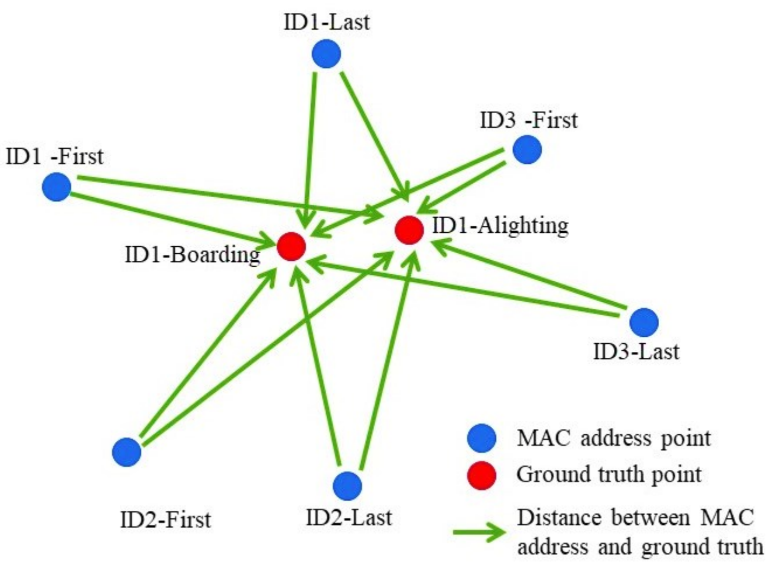


Figure 42. Visualization of the distance matrix

The results are in Table 6, whereas the visualization is shown in Figure 42. InputID is the MAC address, and TargetID is the ground truth data. This analysis considered the distance between every combination of InputID and TargetID. The results of this step were as labeled distance_matrix.csv.

Table 6. Distance matrix result (distance_matrix.csv)

InputID (MAC)	TargetID (Ground Truth Data)	Distance (m)
0165fdexxx	18Device-5G	1144.27
0165fdexxx	19Device-5G	462.63
0165fdexxx	20Device-5G	1977.47
0165fdexxx	21Device-5G	2577.16
….	….	….

However, distance_matrix.csv did not include other information. It therefore needed to be merged back again. The dataset distance_matrix.csv was combined with the Wi-Fi Moved_firstlast.csv and GTGPS.csv datasets. After joining, the results were labeled as Wi-Fi Moved_firstlast_distance.csv and GTGPS_distance.csv. The Python Jupyter Notebook software package was used in this step. The results of the merge are shown in Table 7. These two datasets were merged again with the Wi-FiGT.csv dataset.

Table 7. Results of assigning “first” and “last” appearance Wi-FiGT.csv

MAC	Annotation MAC	Passenger ID	Annotation Passenger ID	Distance (m)
0165fdexxx	First	17Device-5G	Boarding	89.15
0165fdexxx	Last	17Device-5G	Alighting	59.15
081d7fexxx	First	40Device-5G	Alighting	200.15
081d7fexxx	Last	40Device-5G	Boarding	430.15
14db969xxx	First	12Device-5G	Boarding	300.11
1c0f362xxx	Last	71Device-5G	Alighting	121.32
….	….	….	….	….

5.2.6. Matching Data

The sixth step aimed to match between the two conditions of MAC address annotation (first and last) and ground truth annotation (boarding and alighting). This analysis matched the paired MAC addresses and passenger IDs based on the annotation. The dataset used for this process was Wi-FiGT.csv. To create the matching pairs, the following conditions were used:

```
If MAC address = first and passenger ID = boarding
    return matched as boarding
        if MAC address = last and Passenger ID = alighting
            return matched as alighting
                if MAC address = first and Passenger ID = alighting
                    return Unmatched
                        if MAC address = last and Passenger ID = boarding
                            return Unmatched
                                else: unpaired
```

These conditions were applied to eliminate nonmatches between the Wi-Fi and ground truth data. The system accepted the condition if the MAC annotation First is matched with the Passenger ID Boarding (ground truth data) annotation, and the MAC annotation Last is matched with the Passenger ID Alighting annotation as boarding or alighting. If the labels of MAC the address and passenger ID are not consistent with each other, they will be determined as unmatched. If there are no pairs (only boarding or alighting remains), it will be determined as unpaired. The logical structure of this analysis is illustrated in Figure 43. After the data were matched, the unmatched and unpaired data were not necessary and could be eliminated.

MAC	Annotation MAC	Passenger ID	Annotation Passenger ID	Distance (m)	Match Annotation
0165fdexxx	First	17Device-5G	Boarding	89.15	Boarding
0165fdexxx	Last	17Device-5G	Alighting	59.15	Alighting
081d7fexxx	First	40Device-5G	Alighting	200.15	Unmatched
081d7fexxx	Last	40Device-5G	Boarding	430.15	Unmatched
14db969xxx	First	12Device-5G	Boarding	300.11	Unpaired
1c0f362xxx	Last	71Device-5G	Alighting	121.32	Unpaired
----	----	----	----	----	----
				Eliminate	

Result

MAC	Annotation MAC	Passenger ID	Annotation Passenger ID	Distance (m)	Match Annotation
1c0f362xxx	first	17Device-5G	Boarding	89.15	Boarding
1c0f362xxx	Last	17Device-5G	Alighting	59.15	Alighting
----	----	----	----	----	----

WiFiGT_matching.csv

Figure 43. Matching data process

Jupyter Notebook was used for this analysis. The result of this matching was labeled as Wi-FiGT_matching.csv.

5.2.7. Reduction by Time and Distance Difference

The seventh step aimed to eliminate or reduce the matched dataset by the time and distance threshold. This step was the final process and used Wi-FiGT_matching.csv as input data. There were still many matched data, of MAC addresses and passenger IDs, which did not represent real passengers. Because the previous step did not consider the time and location but only the annotation of the MAC address and passenger ID, this step is essential to classify the MAC addresses matching with the ground truth data by distance and time.

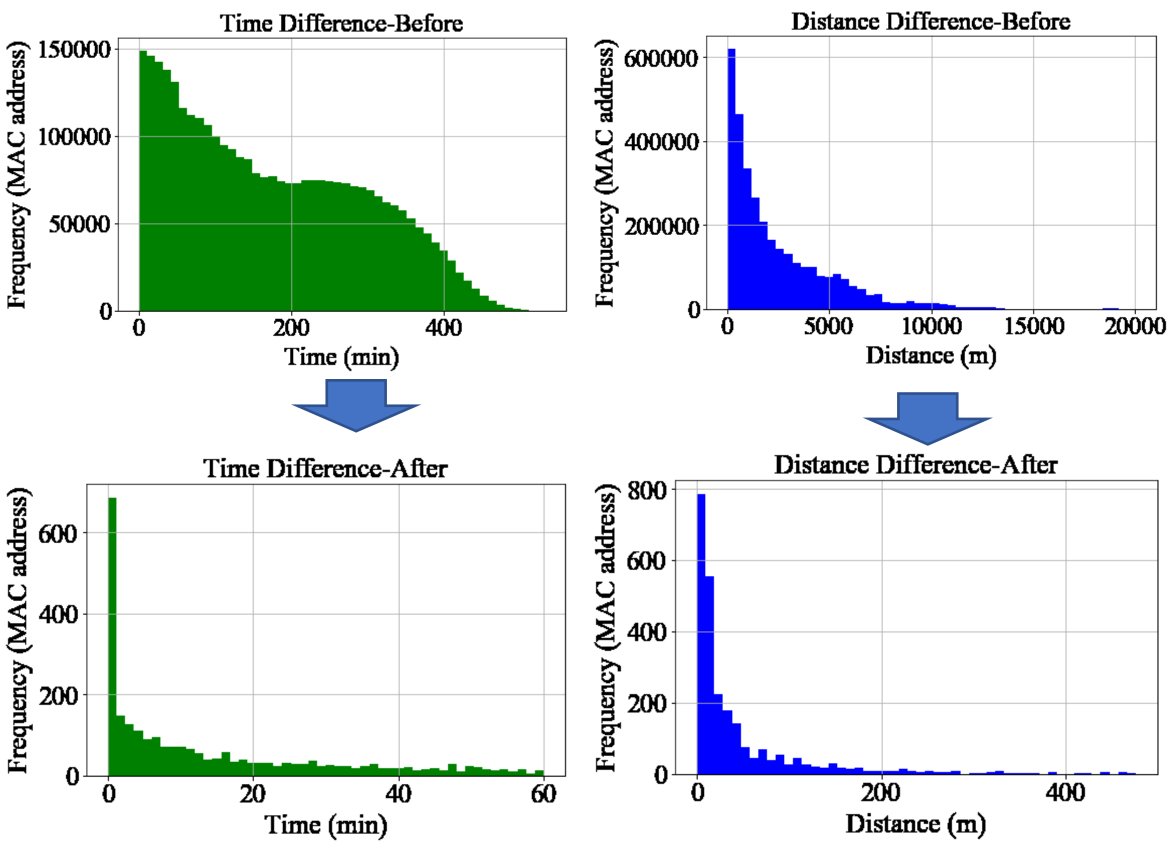
The time difference is the difference (in minutes) between the MAC address data points and ground truth data points. Figure 44 (a) (top) shows the histogram of the time difference before this step. The distance difference is the difference (in meters) between the MAC address data points and ground truth data points. Figure 44 (b) (top) shows the histogram of the distance difference before this step. There are many matched but nonpassenger MAC addresses.

In order to reduce the number of MAC addresses, or to delete the matched but nonpassenger MAC addresses, the thresholds for both the time difference and distance difference should be selected appropriately.

The threshold for the time difference is the longest time difference between a MAC address and passenger ID. This means that the MAC address was recorded at the time of the threshold before or after the passenger was boarding or alighting. If the time difference is longer, a MAC address which is long before and after the time of boarding or alighting will be included, which will cause overestimation by several passengers. However, if the time difference is shorter, some MAC addresses that belong to real passengers will be excluded, which will cause underestimation by several passengers.

The threshold for distance difference is the farthest distance between the MAC address and passenger ID. This means that the MAC address was recorded at the threshold distance from the location where the passenger was boarding or alighting. If the distance difference is longer, a MAC address which is farther away from the position of boarding or alighting will be included it will cause overestimation by a few passengers. However, if the distance difference is shorter, some MAC addresses that belong to real passengers will be excluded, which will cause underestimation in the number of passengers. After applying these thresholds appropriately and simultaneously, unrealistic matched MAC addresses and passengers IDs are removed.

The threshold for the time difference is set to <60 min, which is taken from the travel time required for the *petepete* to complete one trip from end to end. The threshold for distance is set to <400 m, which was reported as the farthest distance the Wi-Fi equipment can reach to capture the MAC address of a device. Previous research reported that MAC address and ground truth data for boarding and alighting considered a Wi-Fi sensor range of 500 m (Fukuda et al., 2017). Figure 44 shows the frequency before and applying the thresholds. This step used Jupyter Notebook.



a) Time difference b) Distance difference

Figure 44. Histogram of time and distance difference before and after cutting off by thresholds

5.2.8. Buffer Route Zone

For dB cleaning, we construct a zone around the *petepe*te. This buffer zoning is used to identify the MAC address and ground truth data, which facilitates obtaining the OD table. A buffer zone of 100 m around the *petepete* route was applied. This divides into the route into 62 zones, as shown in Figure 45.

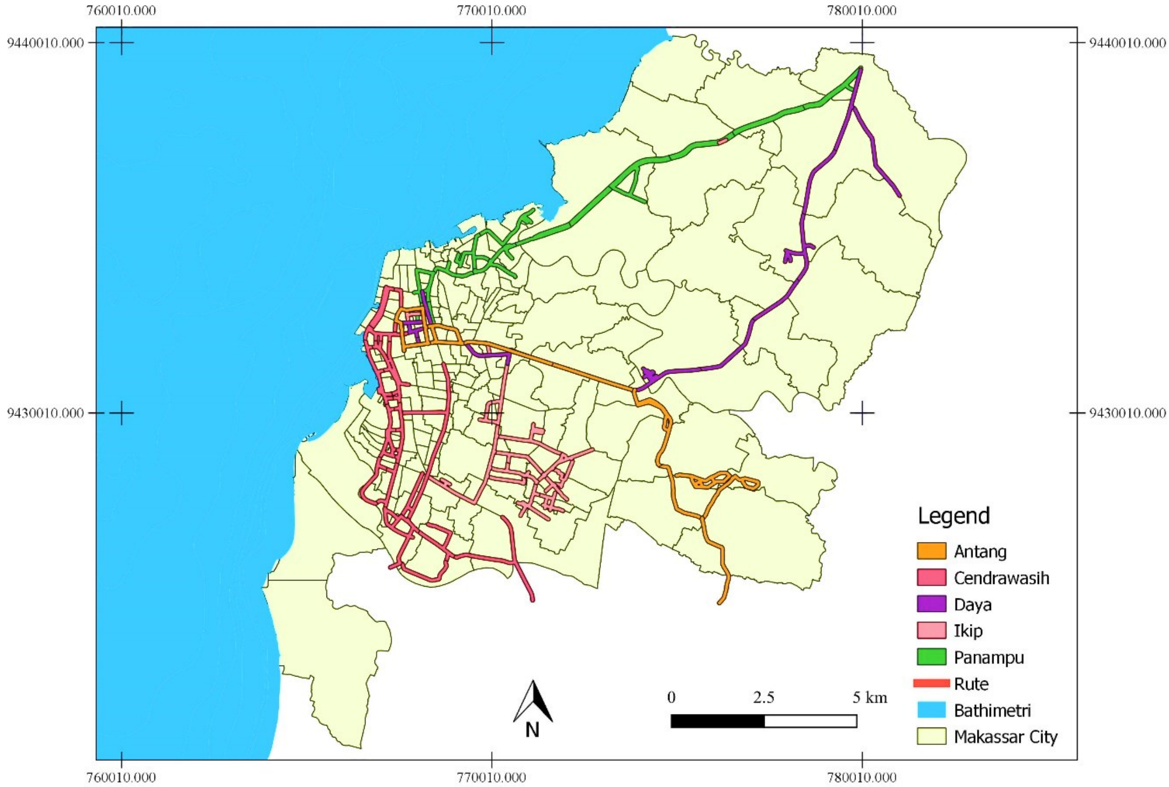


Figure 45. Zoning around the route

5.3. Estimation Result

5.3.1. Matching Process Result

This research acquired approximately 700,000 unique MAC addresses, and the ground truth data included 2925 passengers for the five survey days. After processing, 1211 MAC addresses were found to match with the ground truth data. Therefore, 99.82% of the Wi-Fi data was cleaned, or only 0.18% data was left to identify the passengers. The ratio of the ground truth data to the cleaned data was 1:2. Therefore, only 1211 of the 2925 passengers, or 41.4% of the observations from the onboard survey produced a match between the cleaned MAC address and the ground truth data. Figure 46 shows intersection (matched data) between the raw data and the ground truth data.

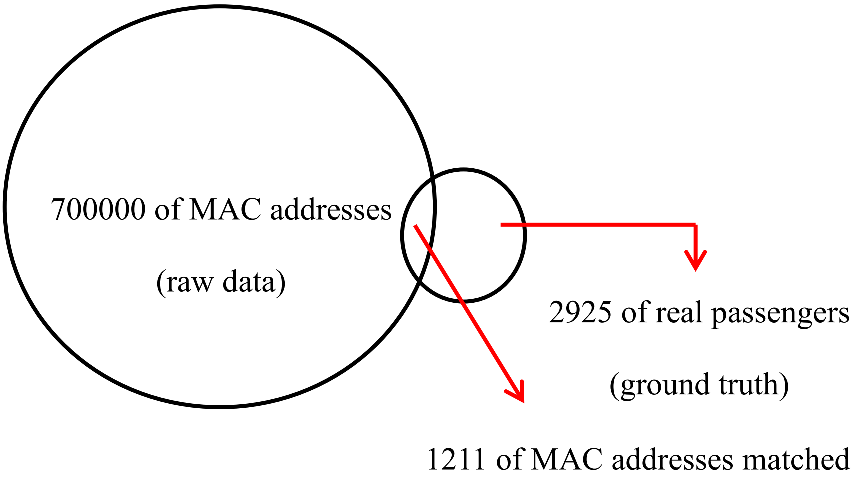


Figure 46. Data intersection

In the device results, Figure 47 shows matched data between Wi-Fi matched results and the ground truth in 20 devices that were used. There is a difference between these datasets. The Wi-Fi results are lower than the ground truth data. Some devices after cleaning show similar relative patterns with the ground truth. Only devices 15 and 23 show very high differences in the number of passengers, which are 145 and 164 passengers, respectively.

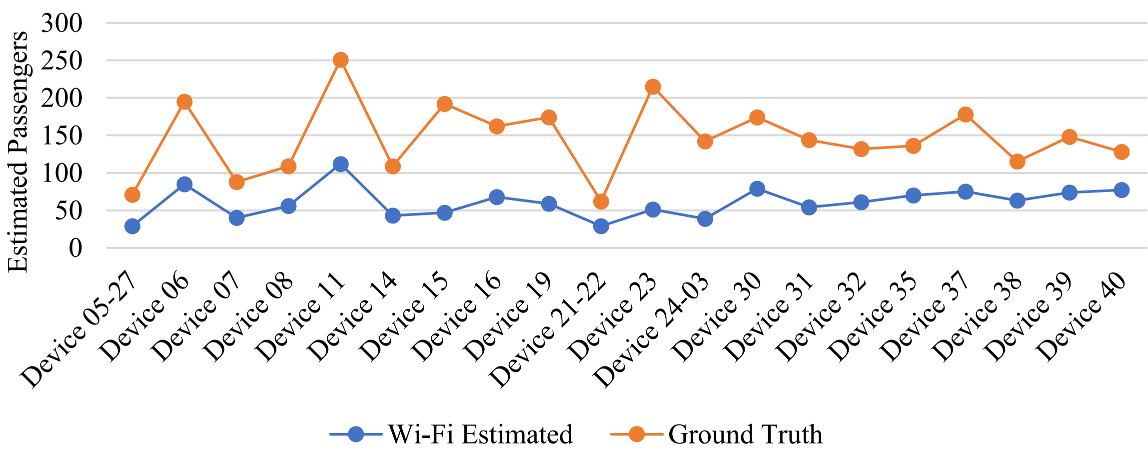


Figure 47. Wi-Fi matched results and ground truth data based on each device

There is still a bias between the Wi-Fi matched results and the ground truth data. The trend from Figure 47 reveals that the route speed procedure result is related to the ground truth data. A correlation analysis reveals that the Wi-Fi result ground truth and route speed procedure have a high correlation, with a factor of approximately 0.70. This means that the Wi-Fi results are correlated to the ground truth data.

This study was applied on five routes, which are Antang, Cendrawasih, Daya, Ikip, and Panampu. In terms of the differences among survey days (Figure 48), August 27, 2018 shows a high number of passengers because it was in the beginning of the week (Monday). August 25–26, 2018 showed lower results than August 23 and 27, 2018, because it was at the end of the week. Many passengers used paratransit on weekdays, such as students and workers. However, at the end of the week, many *petepete*s were dominated by the nonstudent users. These graphs show a similar tendency between the Wi-Fi matched results and the ground truth data.

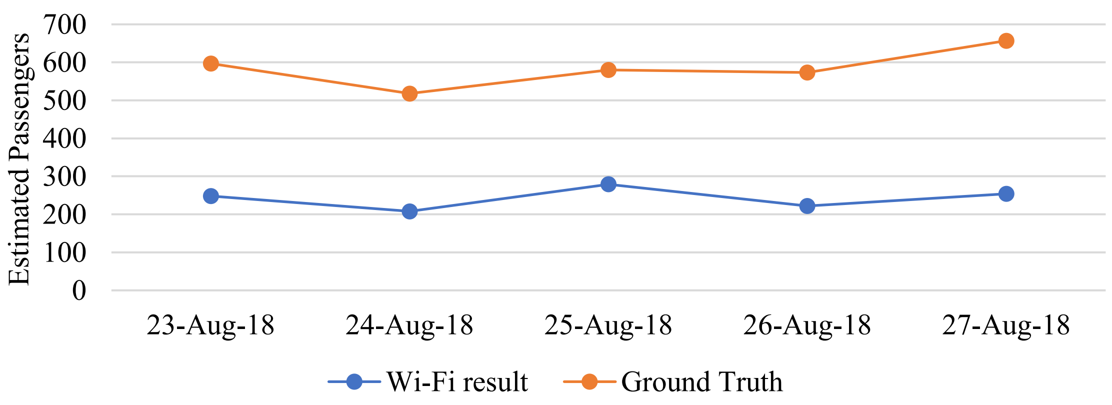


Figure 48. Wi-Fi matched results and ground truth data based on the date of the survey

In terms of the survey route difference (Figure 49), Cendrawasih route shows the highest number of passengers, because this route has a dynamic pattern of land use such as residential, shopping centers, education facility, offices, health facilities, and bus station. The lowest number of the passengers was observed for Panampu route from the Wi-Fi estimation and Ikip route from the ground truth data. Based on these route conditions, the Wi-Fi matched results and ground truth data show a similar pattern.

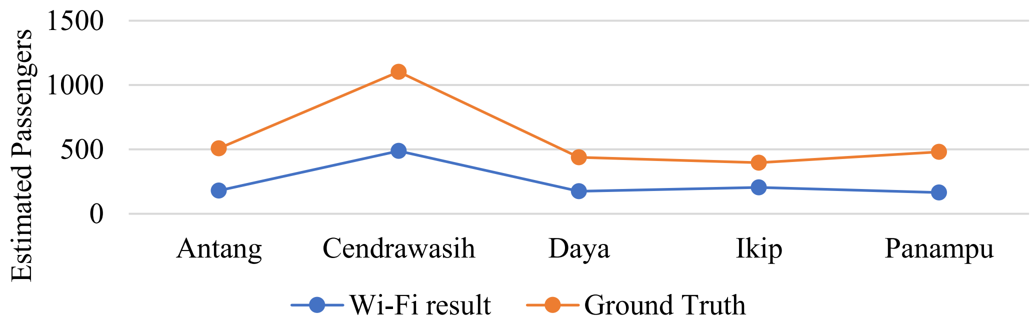


Figure 49. Wi-Fi matched results and ground truth data based on the route

Figure 50 shows the difference between before and after processing. The map on the left side is before the cleaning of the raw ground truth data and Wi-Fi data, and there are still many Wi-Fi data points on the map, with approximately 700,000 addresses. On the right side, there are 1211 Wi-Fi data matched with ground truth data after cleaning. It means 605 MAC addresses that is matching with the ground truth. After the cleaning process, the MAC addresses show the same pattern as ground truth data, and some MAC address points match with ground truth points. Figure 51 presents the Wi-Fi matched results, which show the first and last appearances of MAC addresses, and ground truth data, which indicate boarding and alighting. This OD shows the trip information of paratransit passengers, which can be used for demand estimation or other modeling for future paratransit work.

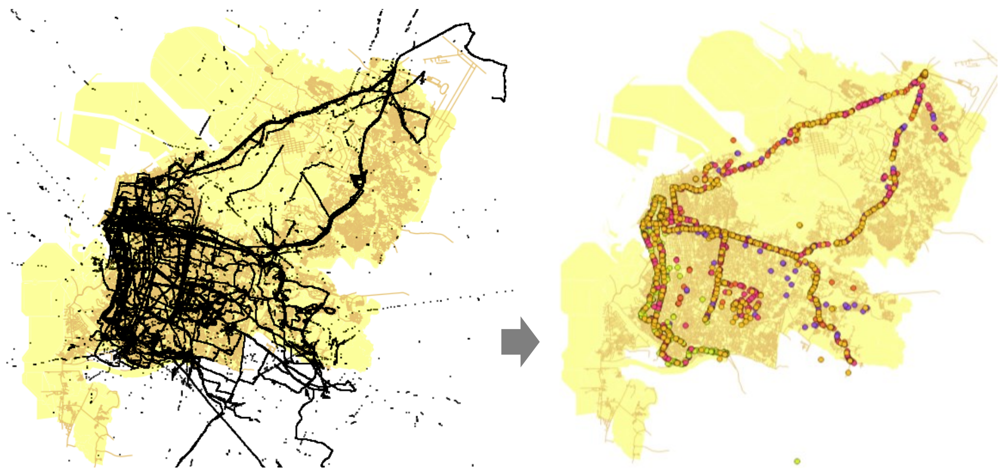


Figure 50. Map points before and after the cleaning process

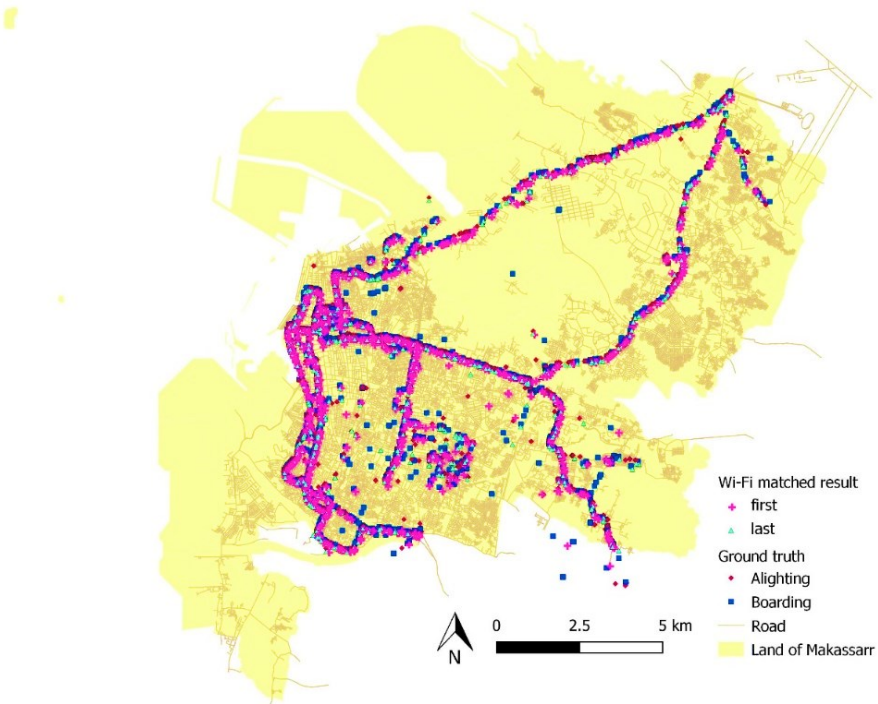


Figure 51. Origin–destination map of MAC addresses after cleaning and matching process

5.3.2. dB Cleaning

The first step is a matching process, which is necessary to use the ground truth data. What if there is no ground truth data, and only Wi-Fi data? This section aims to analyze the Wi-Fi data without considering the ground truth data. Ground truth is only used to compare the data. This result is based on the dB cleaning data. The dB data represents the Wi-Fi signal strength.

We consider that MAC addresses with dB data > −70 dBm have sufficient signal strength to use in the data processing. After the cleaning procedure, we obtained the OD matrix data for each date and each route. We compared to the ground truth data and calculated the correlation between the estimated OD and ground truth.

Table 8 and Figure 52 show the correlation between the date and route. A high correlation was obtained on all days for Antang (0.85) and on 27 August 2018 for Antang (0.79). Cendrawasih and Daya route have a high correlation on 23 to 27 August 2018. The results in this table show the accuracy between all devices on all dates and routes.

Table 8. Correlation value based on route and date

Date	Route				
	Antang	Cendrawasih	Ikip	Daya	Panampu
23-Aug-18	0.74	0.07	0.48	0.68	0.37
24-Aug-18	0.62	0.15	0.50	0.74	0.39
26-Aug-18	0.30	0.36	0.15	0.66	0.66
27-Aug-18	0.79	0.74	0.43	0.75	0.25
All	0.85	0.75	0.58	0.71	0.64
25-Aug-18		0.69			

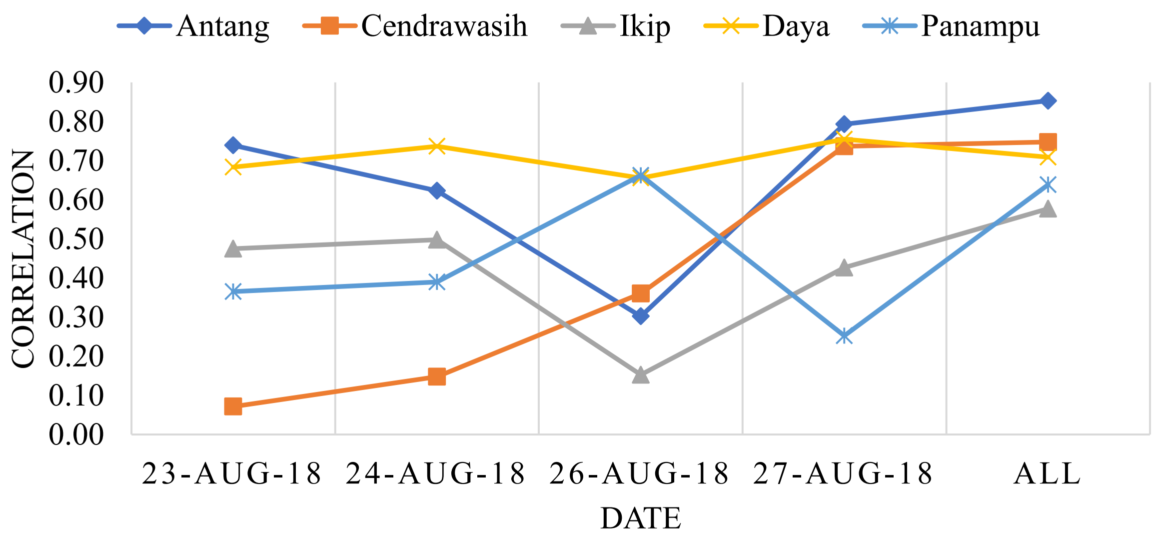


Figure 52. Correlation value based on route and date

For each device accuracy (Table 9), we analyze each device and calculate the correlation based on the date, route, and devices. A high correlation was obtained for the Antang, Cendrawasih, Daya, and Panampu routes. We obtained low results for some devices, because the vehicles travelled off the route for a specific time. For example, for devices 14, 16, 39, and 40, the vehicles were rented and for another device, many motorcycles and bicycles were travelling on the routes.

Table 9. Correlation for each device

Date	Device by route																		
	Antang				Cendrawasih				Daya				Ikip				Panampu		
	16	24-03	31	37	6	14	19	32	8	15	38	39	21-22	30	35	40	*05-27	11	23
23-Aug-18	0.01	0.81	0.53	0.43	0.16	0.03	0.15	0.64	0.16	0.79	0.21	-0.04	No data	0.26	0.18	-0.02	No data	0.18	0.19
24-Aug-18	-0.10	0.63	0.47	0.52	0.22	No Data	0.13	-0.02	0.01	0.81	No data	-0.03	No data	0.45	0.37	-0.01	No data	0.01	0.14
26-Aug-18	0.02	0.00	0.42	No Data	0.52	-0.03	0.29	0.12	0.18	0.64	0.02	0.43	0.08	0.26	-0.02	0.11	No data	0.05	0.73
27-Aug-18	0.23	0.42	0.61	0.56	0.66	0.39	0.60	0.05	0.76	0.76	0.26	-0.04	0.45	0.48	-0.03	0.25	0.15	0.11	0.33
Date	Cendrawasih																		
25-Aug-18	0.15	0.22	No data	0.61	0.62	0.28	-0.03	-0.03	0.47	0.26	0.58	0.07	0.33	0.25	no data	-0.03	0.62	0.11	0.65

5.4. Summary

This section summarizes the results from the paratransit Wi-Fi application. For the matching processing section, this method could identify the MAC addresses that belong to real passengers. Based on the results, a matching rate of 41.4% was achieved between the MAC addresses and ground truth data. The correlation value is 0.70.

For the dB cleaning section, dB data could be used to clean the MAC address data without matching with the ground truth data. Small results were obtained for some devices because of the vehicle travelling out of the route, such as device 14,16, 39, and 40 because the vehicle was rented and for another device, many motorcycles and bicycles were travelling on the routes. The correlation value from dB cleaning is high if we analyze by date and route. However, if we analyze by devices, many parameters must still be used to improve the accuracy.

CHAPTER 6

ROAD CROWDEDNESS AND TRAVEL SPEED STABILITY

6.1. Small Bus Application

Transportation service considerations for traffic flow always include road crowdedness, with speed as an important parameter. The author used Wi-Fi equipment to collect MAC address data and calculate the crowdedness of the road based on the MAC addresses and speed of the small bus and paratransit mode. In this part, the crowdedness is determined by comparing the speed of the vehicle and the total MAC addresses in a certain time. The time interval between one point (XY) and another point is one minute. In every one-minute interval, the speed and the total MAC addresses were analyzed.

Figure 53 shows several datasets that were needed for this analysis, which are the link of the route, GPS log, and Wi-Fi log. The GPS log is used to determine the bus speed with time interval every 30 seconds during the bus service time.

The correlation analysis was used to justify the relationship among all the data. The GPS log is analyzed to estimate the average and standard deviation of the speed. The travel speed stability or travel speed reliability, represented by the level of the vehicle speed on the road, can be estimated by using the average speed divided by the standard deviation for each link of the road along the route. This is called the travel speed stability index and represents the speed reliable level of the road.

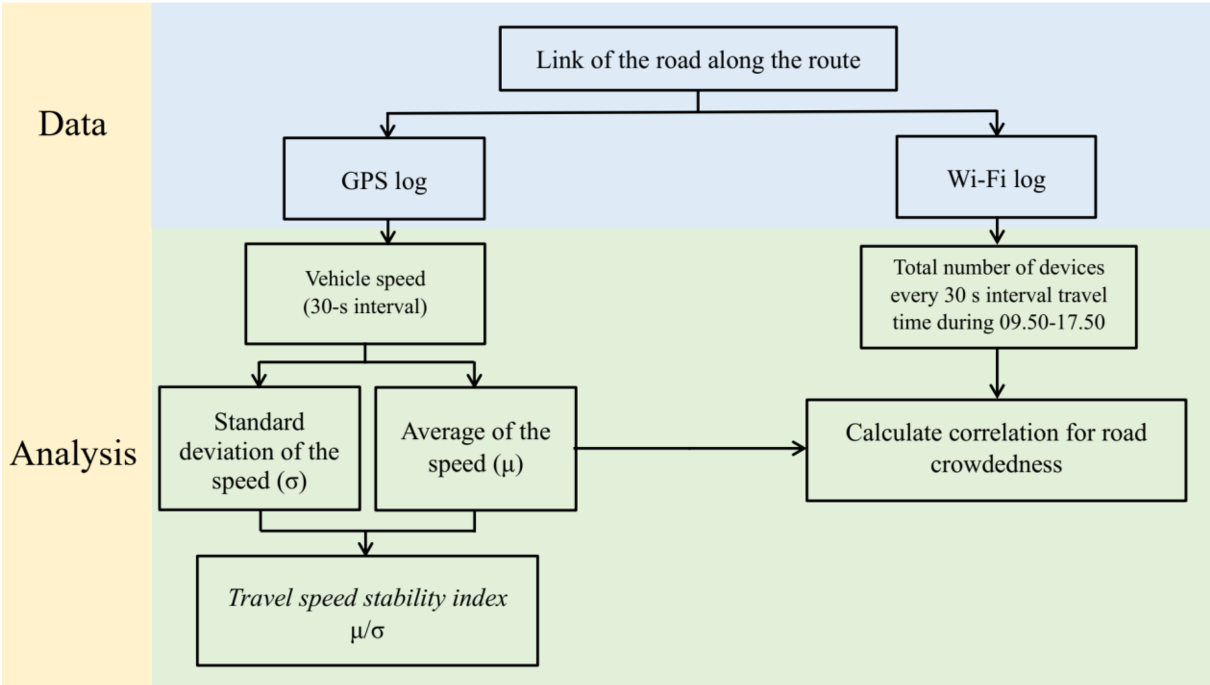


Figure 53. Road crowdedness and travel speed stability index for small bus

There are two hypotheses. The first is that if the speed decreases, the total number of MAC addresses increases and vice versa.

The small bus system in Obuse had two buses on 13–14 October 2018, which were bus no. 1 and no. 2. In Figure 54 to Figure 57, the results show that the speed line and number of device line are contradicting each other similar statement with the hypotheses.

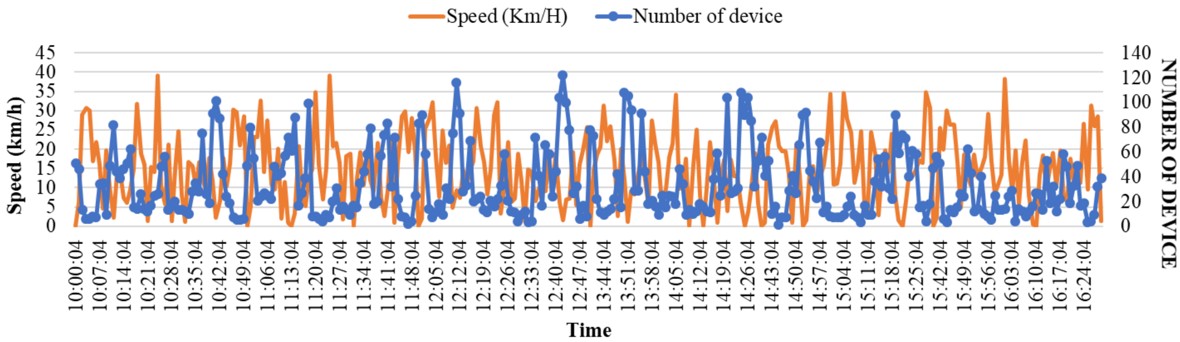


Figure 54. Bus no. 1 on 13 October 2018

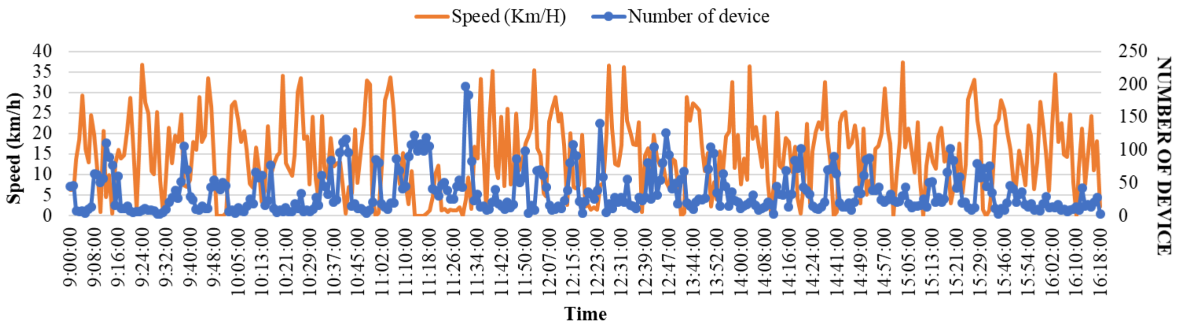


Figure 55. Bus no. 1 on 14 October 2018

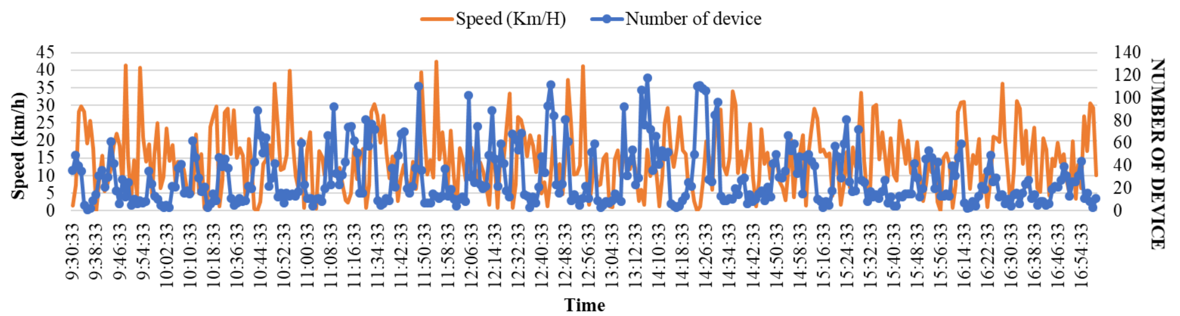


Figure 56. Bus no. 2 on 13 October 2018

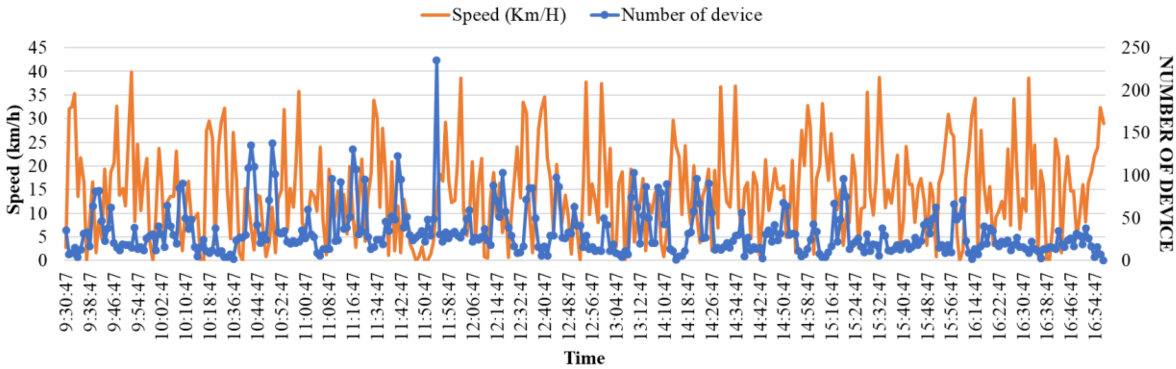
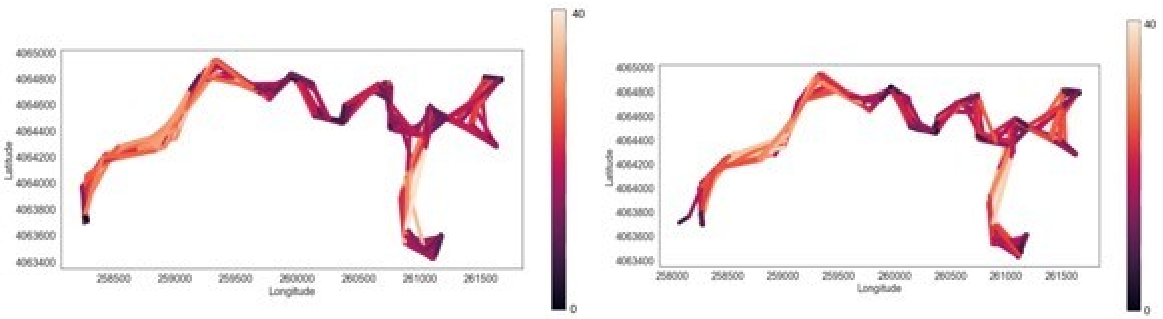
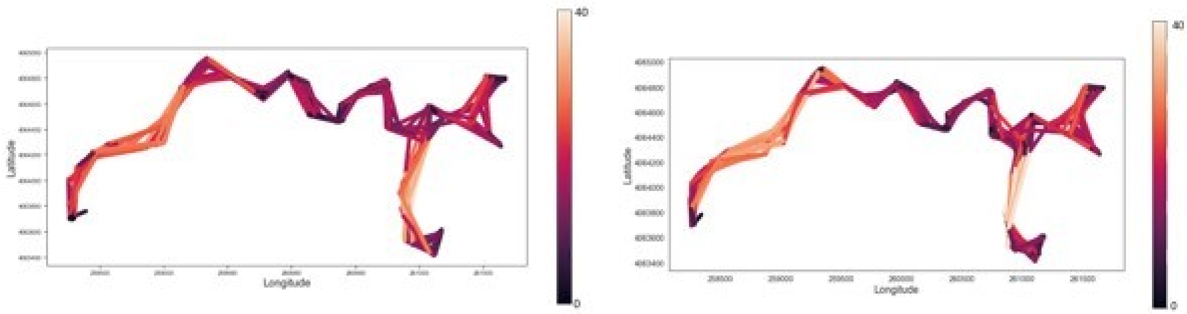


Figure 57. Bus no. 2 on 14 October 2018

In Figure 58, there are some crowded places along the bus route, such as BS1-BS2, BS5-BS6, BS2-BS1, and BS8-BS9. These crowded places were justified by using speed data.



a) Speed bus no. 1 on 13–14 October 2018 (unit: km/h)



a) Speed bus no. 2 on 13–14 October 2018 (unit: km/h)

Figure 58. Speed change along the small bus route

From the correlation analysis, the correlation values are high. For bus no. 1 on 13–14

October 2018, the values are −0.629 and −0.531 (Figure 59). For bus no. 2 on 13–14 October 2018, the values are −0.387 and −0.411 (Figure 60). These values are negative, which means that the speed decreases and the total number of MAC addresses increases, or vice versa.

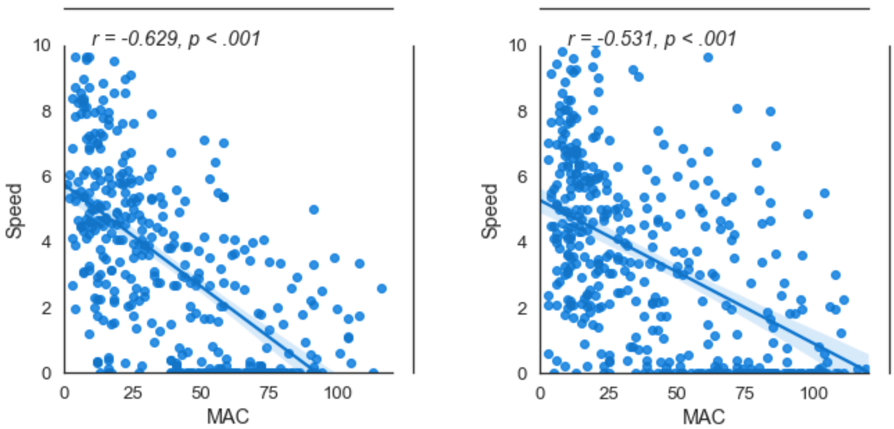


Figure 59. Correlation values for bus no. 1 on 13–14 October 2018

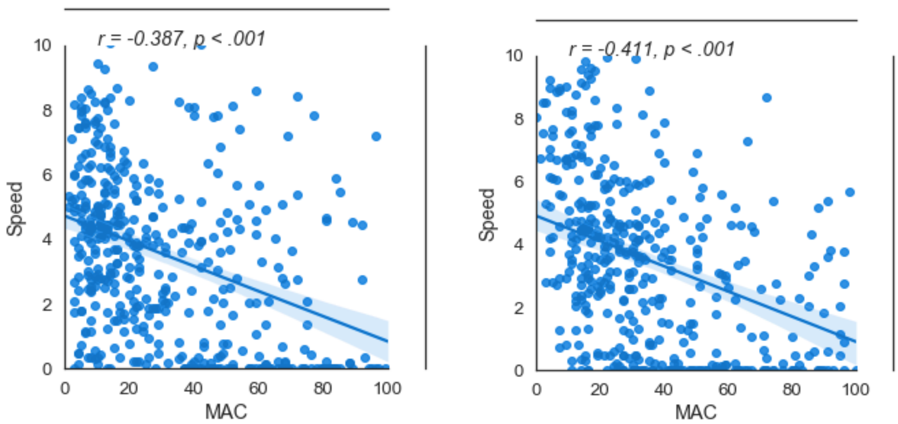


Figure 60. Correlation values for bus no. 1 on 13–14 October 2018

After analyzing the travel speed stability, it can be determined as the speed level of the road. Figure 61 shows that some places have low travel speed stability values close to 0. This means that these links of the road are not reliable for travel speed. The travel speed stability means about travel speed reliable or not. If the link of the road has sometime fast speed or slow speed, it shows about the speed situation of the road. if the speed index approaches 0, it means the travel speed is not reliable or not stable. Based on speed index, we implemented two levels to justify the travel speed stability level based on the value of speed index, which were high reliable, reliable, average, and not reliable.

Transportation officers can determine a plan from this information, such as by changing the timetable for particular times or special events.

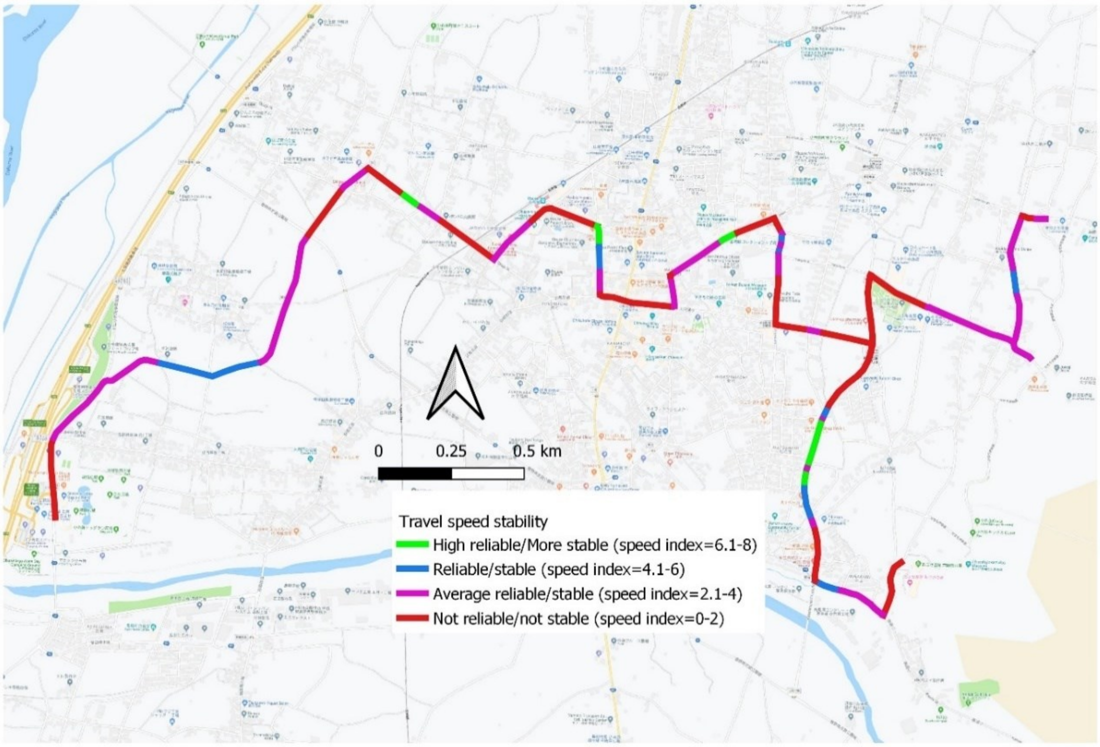


Figure 61. Travel speed stability level for Obuse case study

6.2. Paratransit Application

This section aims to 1) identify the relation between vehicle speed and MAC address data to understand road crowdedness, and 2) justify the travel speed stability of the road route based on the travel speed stability index.

Figure 62 shows several data that were needed for the analysis in this section, which are the link of the route, GPS log, Wi-Fi log, total lane of road, road type, direction, traffic signal, and land use condition. The link of the route is each link on the road for all the route. The GPS log is used to determine the *petepete* speed with time interval of every 30 seconds during the *petepete* service time. The Wi-Fi log represents total number of MAC addresses that are counted by our equipment every 30 seconds. The total lane refers to the number of lanes for each link of the road. The road type, direction, traffic signal, and land use are dummy data. For road type, it is defined into three road types, which are primary road, collector road, and local road. Direction is defined into two categories, which are one and two directions. Traffic signal is defined into two categories, which are with and without signal. Land use is defined into eight types, which are shops, residential, offices, religion building, education facility, health facility, bus terminal, and harbor. All the data were found from the survey and secondary data from the government.

A correlation analysis was used to justify the relationships among all the data. The GPS log was analyzed to estimate the average and standard deviation of the speed. The travel speed stability, in terms of the road, can be estimated by using average of the speed divided by the standard deviation for each link of the road along the route.

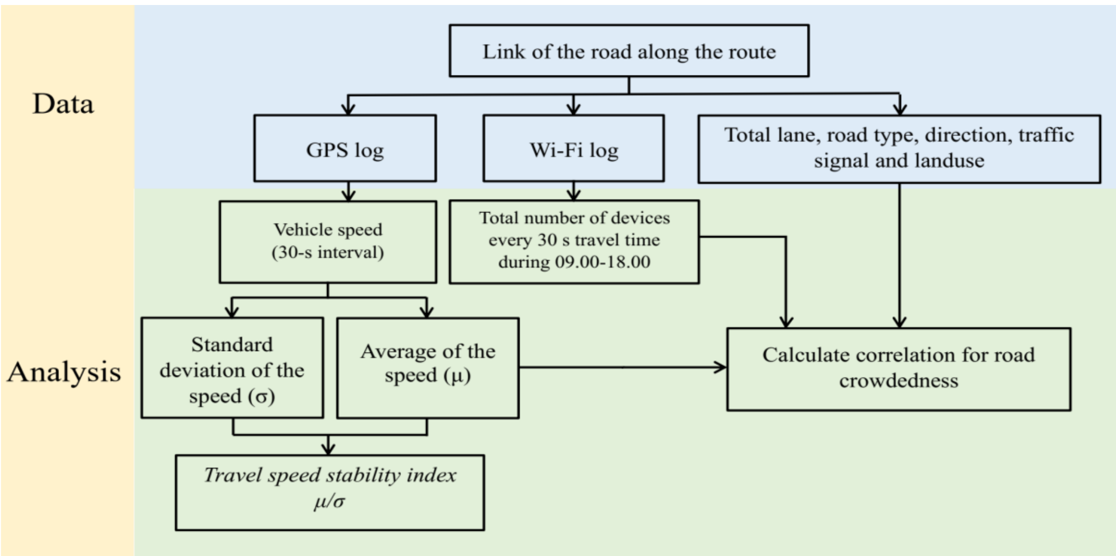


Figure 62. Road crowdedness and travel speed stability level procedure for paratransit

In this research, a link is justified from node to node of the road for all roads inside the route. Each link contains all the data and it was calculated as the speed index. Figure 63 shows an example of the link and it contains many MAC address data, GPS log, building (land use), and road data around the blue link. The analysis to combine all the data used QGIS spatial analysis. The RI estimation used Python software.

Figure 63. Justify the link

Figure 64 shows the total MAC addresses (number of Wi-Fi devices) and average of the speed for each link. According to this figure, the total number of MAC addresses is related to the average speed. This means that if the vehicle becomes slower, the total number of MAC addresses will increase and if the vehicle becomes faster, the total number of MAC addresses will increase.

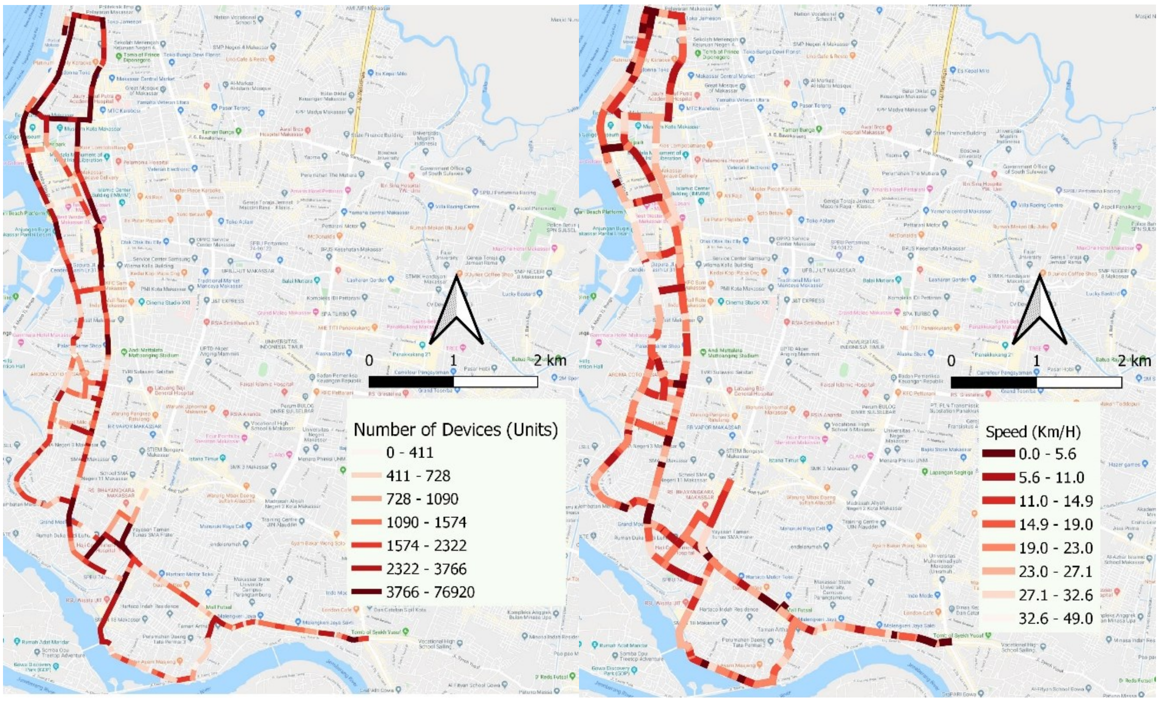


Figure 64. Total MAC addresses and speed relation

A correlation analysis shows that the speed and the total number of MAC addresses for all scanner are related. The correlation factor is approximately −0.60. This means that there is a correlation between these two parameters. The hypothesis "if the vehicle becomes slower, then the total number of MAC addresses will increase and if the vehicle becomes faster, then the total number of MAC addresses will decrease" is accepted. In Figure 65, it proves the hypotheses.

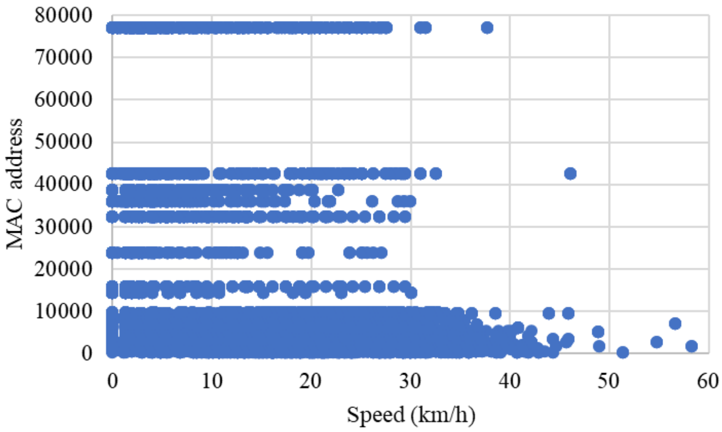


Figure 65. Correlation value between total MAC addresses and speed

The correlation between speed and signal is shown in Figure 66. The correlation factor with signal is negative and that without signal is positive. Based on the figure, if the road has a signal, the speed decreases and if the road does not have a signal, the speed increases.

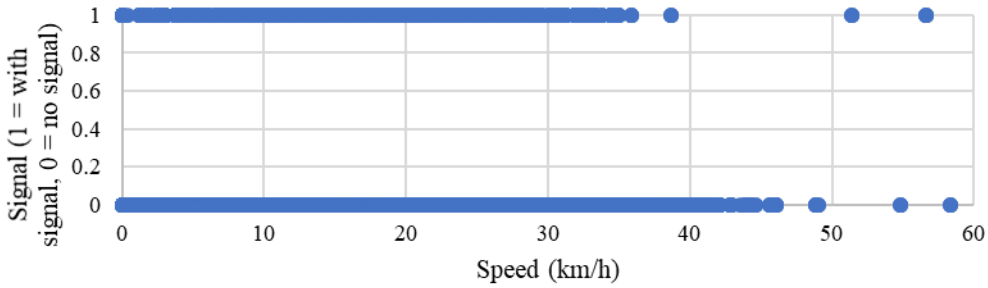


Figure 66. Correlation value between signal and speed

The correlation value between speed and total lane is shown in Figure 67. The total lane parameter affects the speed. If the total lane is higher, then the speed is faster, which means the traffic capacity of the road is higher.

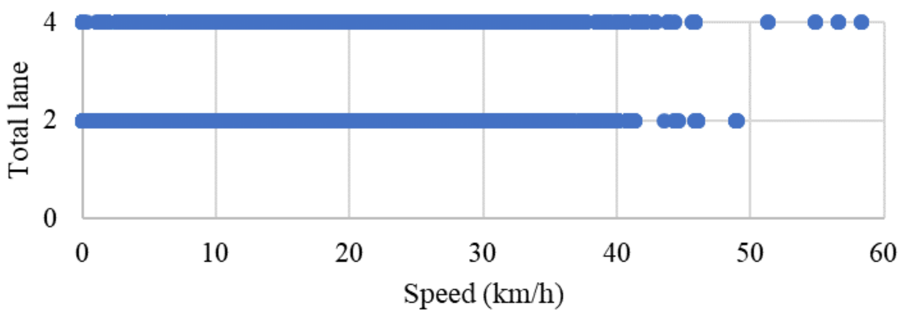
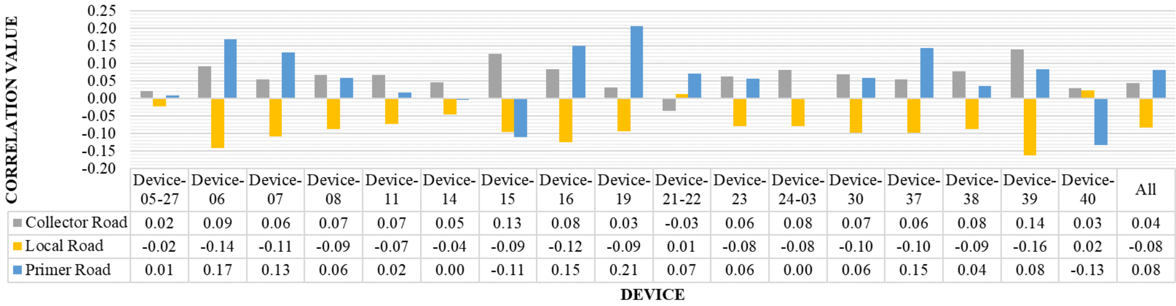


Figure 67. Correlation value between total lane and speed

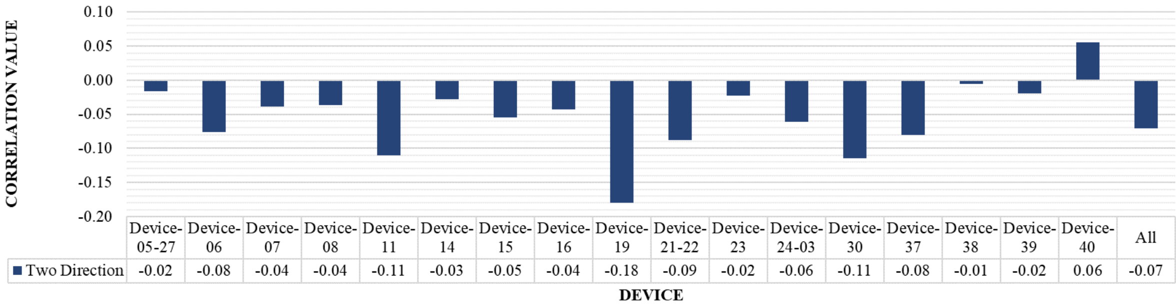
Figure 68 shows the correlation between speed and road type. Overall, there are many differences between road type and devices. The highest correlation is on device 16 on a primer road.



	Device-05-27	Device-06	Device-07	Device-08	Device-11	Device-14	Device-15	Device-16	Device-19	Device-21-22	Device-23	Device-24-03	Device-30	Device-37	Device-38	Device-39	Device-40	All
Collector Road	0.02	0.09	0.06	0.07	0.07	0.05	0.13	0.08	0.03	-0.03	0.06	0.08	0.07	0.06	0.08	0.14	0.03	0.04
Local Road	-0.02	-0.14	-0.11	-0.09	-0.07	-0.04	-0.09	-0.12	-0.09	0.01	-0.08	-0.08	-0.10	-0.10	-0.09	-0.16	0.02	-0.08
Primer Road	0.01	0.17	0.13	0.06	0.02	0.00	-0.11	0.15	0.21	0.07	0.06	0.00	0.06	0.15	0.04	0.08	-0.13	0.08

Figure 68. Correlation value between road type and speed

As with the signal data, the type of direction affects the speed. One direction or one way increases the speed. Otherwise, if there are two directions, it makes the vehicle slower, as shown in Figure 69.



	Device-05-27	Device-06	Device-07	Device-08	Device-11	Device-14	Device-15	Device-16	Device-19	Device-21-22	Device-23	Device-24-03	Device-30	Device-37	Device-38	Device-39	Device-40	All
Two Direction	-0.02	-0.08	-0.04	-0.04	-0.11	-0.03	-0.05	-0.04	-0.18	-0.09	-0.02	-0.06	-0.11	-0.08	-0.01	-0.02	0.06	-0.07

Figure 69. Correlation value between direction and speed

Figure 70 shows the correlation between speed and land use. The land use affects the speed, such as harbor, bus terminal, religion building, offices, shops, and facility. Overall, these land uses make the speed slower rather than faster. The highest correlation (−0.25) was on device 16 with a bus terminal.

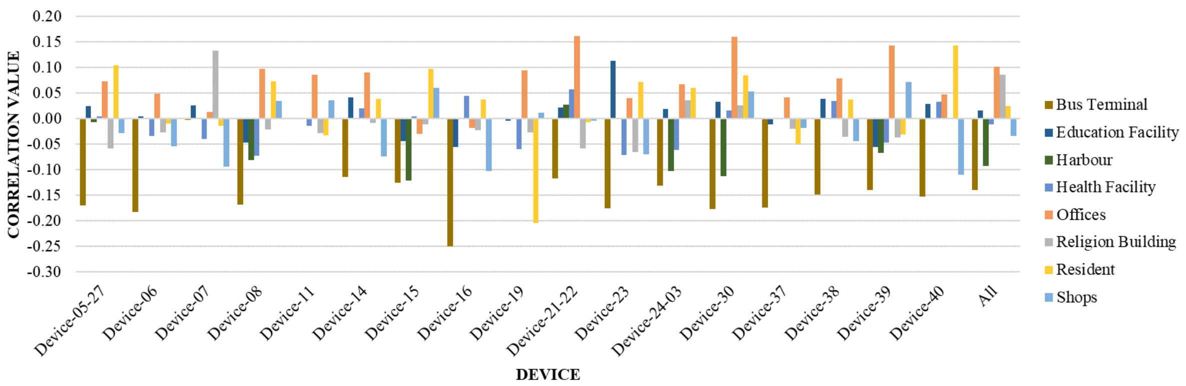


Figure 70. Correlation value between land use and speed

After analyzing the travel speed stability index, it can be determined as the reliability level of the road. Figure 71 shows that some places have low speed index close to 0. it means the travel speed is not reliable or not stable. According to the author's experience and the 2018 survey, these places (marked in black) are very crowded places such as *Karebosi* park, terminal, harbor, shops, and some offices. These crowded places made the travel speed that became not reliable. Based on speed index, we implemented two levels to justify the travel speed stability level based on the value of speed index, which were high reliable, reliable, average, and not reliable.

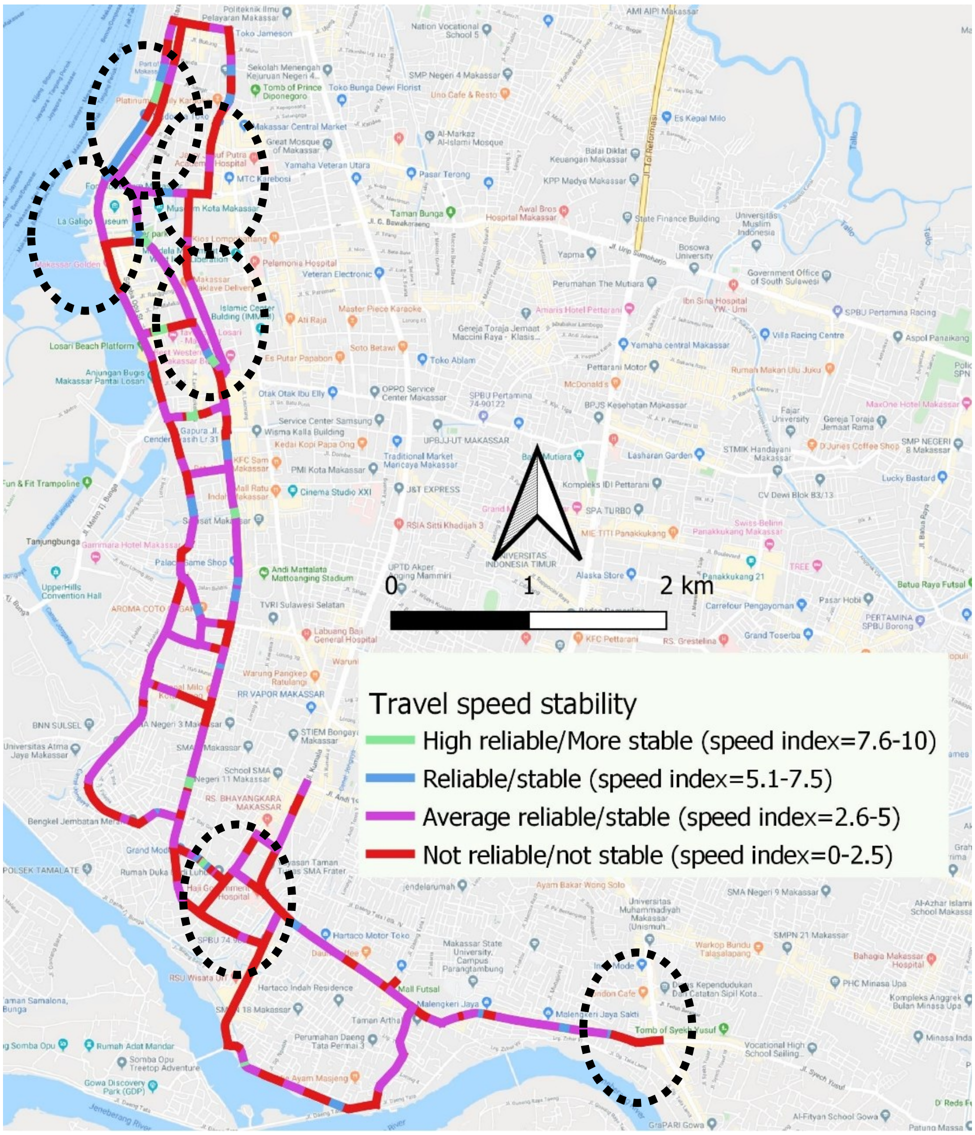


Figure 71. Travel speed stability as the speed reliability for Makassar case study

6.3. Summary

For the road crowdedness, the relation between vehicle speed and MAC address data is to justify the road as crowded or not. The correlation value is negative, which means that the total number of MAC addresses increases as the speed decreases. The travel speed stability index is the average speed divided by the standard deviation for each link or segment of the road. Based on the index, there are many segments of the road with reliability levels. Transportation officers could determine a plan from this information, such as changing the timetable for particular time or special event.

As with Obuse, based on the travel speed stability index, there many roads with high travel speed stability levels. According to the author's experience and the 2018 survey, the places (marked in black) are very crowded places, such as *Karebosi* park, terminal, harbor, shops, and some offices. Based on the speed index, the author implemented four levels to justify the travel speed stability level based on the value of the speed index.

CHAPTER 7
CONCLUSION

The author developed a survey method, cleaning method, and analysis method to estimate the passenger volume and OD matrix based on data collected with a Wi-Fi scanner. The analysis and interpretation of the Wi-Fi data yielded information regarding passenger volume as well as information regarding passengers' travel between destinations. The investigation concerns the application of a new method for collecting and processing data to determine the OD matrix and passenger volumes for small bus route sections and OD for paratransit *petepete*.

The results of the investigation for the small bus can be summarized as follows: a new processing procedure was developed to transform data into OD and passenger volumes for each section, which determines passenger OD information and the number of passengers boarding and alighting a bus at a specific bus stop. The results of the comparison between data obtained using the Wi-Fi procedure and observed data indicate similar travel patterns, and the number of passengers determined using the Wi-Fi procedure is generally less than the number of passengers determined using the ground truth method.

For the speed calculation, this research applied the route speed method for cleaning Wi-Fi data rather than the direct speed. The route speed method interpolates from GPS time-stamped data, so that it uses not only MAC address stamp data but also the GPS log, which is based on the actual vehicle route. With the route speed method, the MAC address speed was found to be the same as the actual vehicle speed. The number of estimated passengers or OD found by the route speed method was lower than that from the ground truth data. This result reflects the undercounting caused by the presence of passengers without Wi-Fi devices.

This research produced findings for *petepete*, which included a new OD measurement procedure by matching Wi-Fi data and ground truth data. A procedure was proposed to clean Wi-Fi raw data into passenger OD data. This cleaning process successfully reduced noise in Wi-Fi data, because the Wi-Fi scanner detected the MAC addresses associated not only with *petepete* passengers but also nonpassengers, such as pedestrians, cars, or motorcycles, as well as of devices located in buildings along the streets.

This study provided a new method called the matching process for paratransit. The matching process is an essential part of making precise estimations, which matched between the Wi-Fi cleaned data and ground truth data. Based on the cleaning procedure, only 0.18% are identified as the passengers from the Wi-Fi raw data. The matching process produced 41.4% similarity between the Wi-Fi data and ground truth data, i.e., the Wi-Fi data successfully identified 41.4% of passengers from the ground truth data.

Two thresholds were used to clean Wi-Fi data, which are the time and distance thresholds. The threshold section is the last part of the cleaning procedure. Distance and time thresholds can be used to exclude unrelated noise data. The threshold process produced cleaned data, which are more similar to real conditions based on the time and distance between the Wi-Fi cleaned data and ground truth. The distance between the Wi-Fi cleaned data and ground truth data is not a significant problem, because there are few matched data which have longer distance difference and shorter time difference. However, the time difference between the Wi-Fi cleaned data and ground truth data is a substantial problem, because the time difference must not be too long. Some matched data contain shorter distance difference and longer time difference. This is because the *petepete* go through the same streets at different times of the day.

This research also analyzed crowdedness and level of service of the road based on Wi-Fi equipment. According to the results, there are correlations between vehicle speed and the total number of MAC addresses, and MAC data can be used to understand the road crowdedness. The negative correlation means that if the vehicle speed decreases then the total number of

MAC addresses increases.

A Wi-Fi scanner is used in this research to determine passenger information. Beyond that, the Wi-Fi scanner can analyze street conditions or travel speed stability levels. This includes information about travel time, travel speed, and standard deviation/variance. Analyzing these aspects, the street condition that link in good service or bad service. The travel speed stability levels can be determined based on travel speed stability index, which are from the standard deviation and average of the speed of each link of the road. Based on the travel speed stability level information, transport planners, stakeholders, government, or street planners can improve the street conditions to improve the travel time and travel speed.

REFERENCES

Abbott-jard, M., Shah, H., & Bhaskar, A. (2013). Empirical evaluation of Bluetooth and Wifi scanning for road transport. *Proceedings of the Australasian Transport Research Forum*, (October), 1–14. Retrieved from http://atrf.info/papers/2013/2013_abbott-jard_shah_bhaskar.pdf

Abedi, N., Bhaskar, A., & Chung, E. (2013). Bluetooth and Wi-Fi MAC Address Based Crowd Data Collection and Monitoring : Benefits , Challenges and Enhancement. *Australasian Transport Research Forum 2013 Proceedings 2*, 1–17. Retrieved from https://eprints.qut.edu.au/71808/

Abedi, N., Bhaskar, A., Chung, E., & 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: Emerging Technologies*, *60*, 124–141. https://doi.org/10.1016/j.trc.2015.08.010

Abousaeidi, M., Fauzi, R., & Muhamad, R. (2016). Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. *Saudi Journal of Biological Sciences*, *23*(5), 555–564. https://doi.org/10.1016/j.sjbs.2015.06.004

Ahmed, H., El-Darieby, M., Morgan, Y., & Abdulhai, B. (2008). A Wireless Mesh Network-Based Platform for ITS. In *VTC Spring 2008 - IEEE Vehicular Technology Conference* (pp. 3047–3051). IEEE. https://doi.org/10.1109/VETECS.2008.329

Aldasouqi, I., & Salameh, W. A. (2014). Using GIS in Designing and Deploying Wireless Network in City Plans. *International Journal of Computer Networks (IJCN)*, *6*(4), 66–75.

Andión, J., Navarro, J. M., López, G., Álvarez-Campana, M., & Dueñas, J. C. (2018). Smart Behavioral Analytics over a Low-Cost IoT Wi-Fi Tracking Real Deployment. *Wireless Communications and Mobile Computing*, *2018*, 1–24.

https://doi.org/10.1155/2018/3136471

Araghi, B. N., Hammershøj Olesen, J., Krishnan, R., Tørholm Christensen, L., & Lahrmann, H. (2015). Reliability of Bluetooth Technology for Travel Time Estimation. *Journal of Intelligent Transportation Systems*, *19*(3), 240–255. https://doi.org/10.1080/15472450.2013.856727

Araico, A. O. (2017). Detection of the crowdedness of a place sensing the devices in the area. *Thesis: Faculty of Electrical Engineering, Mathematics & Computer Science*, *February*, 1–62. Retrieved from https://www.utwente.nl/en/eemcs/dacs/assignments/completed/master/reports/thesis_alejandro_ozaita.pdf

Asija, M. (2016). MAC Address. *IRA-International Journal of Technology & Engineering*, *03*(01), 41–60. Retrieved from https://research-advances.org/index.php/IRAJTE/article/view/125/138

Böhm, M. F., Ryeng, E. O., & Haugen, T. (2016). Evaluating the Usage of Wi-Fi and Bluetooth Based Sensors for Pedestrian Counting in Urban Areas. *European Transport Conference*, 24p-24p. Retrieved from https://trid.trb.org/view/1442384

Chhabra, N. (2013). Comparative Analysis of Different Wireless Technologies. *International Journal Of Scientific Research In Network Security & Communication*, *1*(5), 13–17. Retrieved from http://www.isroset.org/pub_paper/IJSRNSC/ISROSET-IJSRNSC-00117.pdf

Cisco. (2008). 802.11 Security Summary. *Wireless and Network Security Integration Solution Design Guide*, *Chapter 3*, 1–20. Retrieved from https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Mobility/secwlandg20/sw2dg/ch3_2_SPMb.html

Cunche, M. (2014). I know your MAC address: targeted tracking of individual using Wi-Fi.

Journal of Computer Virology and Hacking Techniques, *10*(4), 219–227. https://doi.org/10.1007/s11416-013-0196-1

Dunlap, M., Li, Z., Henrickson, K., & Wang, Y. (2016). Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit. *Transportation Research Record: Journal of the Transportation Research Board*, *2595*(2595), 11–17. https://doi.org/10.3141/2595-02

Erkan, İ., & 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. https://doi.org/10.1007/s40534-016-0101-y

Feng, J., & Liu, Y. (2012). Wifi-based indoor navigation with mobile GIS and speech recognition. *IJCSI International Journal of Computer Science Issues*, *9*(6), 256–263. Retrieved from http://www.ijcsi.org/papers/IJCSI-9-6-2-256-263.pdf

Ferro, E., & Potorti, F. (2005). Bluetooth and wi-fi wireless protocols: a survey and a comparison. *IEEE Wireless Communications*, *12*(1), 12–26. https://doi.org/10.1109/MWC.2005.1404569

Filgueiras, J., Rossetti, R. J. F., Kokkinogenis, Z., Ferreira, M., Olaverri-Monreal, C., Paiva, M., … Gabriel, J. (2014). Sensing Bluetooth Mobility Data: Potentials and Applications. *Advances in Intelligent Systems and Computing*, *262*, 419–431. https://doi.org/10.1007/978-3-319-04630-3_31

Fukuda, D., Kobayashi, H., Nakanishi, W., Suga, Y., Sriroongvikrai, K., & Choocharukul, K. (2017). Estimation of Paratransit Passenger Boarding / Alighting Locations Using Wi-Fi based Monitoring : Results of Field Testing in Krabi City, Thailand. *Journal of the Eastern Asia Society for Transportation Studies*, *12*, 2151–2169. https://doi.org/10.11175/easts.12.2151

Fukuzaki, Y., Mochizuki, M., Murao, K., Nishio, N., Mochizuki, M., & Murao, K. (2014). A pedestrian flow analysis system using Wi-Fi packet sensors to a real environment. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, 721–730. https://doi.org/10.1145/2638728.2641312

Hamacher, H. W., Heller, S., & Ruzika, S. (2010). *Pedestrian and Evacuation Dynamics*. *PED 2010 Conference Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04504-2

Heuvel, T. and H. (2016). Traffic and Granular Flow '15, (October), 1033–1043. https://doi.org/10.1007/978-3-319-33482-0

Heuvel, J. van den, Ton, D., & Hermansen, K. (2016). Advances in Measuring Pedestrians at Dutch Train Stations Using Bluetooth, WiFi and Infrared Technology. In V. L. Knoop & W. Daamen (Eds.), *Traffic and Granular Flow '15* (pp. 11–18). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33482-0_2

Hidayat, A., Terabe, S., & Yaginuma, H. (2017). Mapping of MAC Address with Moving WiFi Scanner. *International Journal of Artificial Intelligence Research*, *1*(2), 34. https://doi.org/10.29099/ijair.v1i2.27

Hidayat, A., Terabe, S., & Yaginuma, H. (2018). WiFi Scanner Technologies for Obtaining Travel Data about Circulator Bus Passengers: Case Study in Obuse, Nagano Prefecture, Japan. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(45), 45–54. https://doi.org/10.1177/0361198118776153

Ilayaraja, K. (2013). Road network analysis in Neyveli Township , Cuddalore District by using Quantum GIS. *Indian Journal of Computer Science and Engineering (IJCSE)*, *4*(1), 56–61.

J.Barceló, L.Montero, Marqués, L., & Carmona, C. (2010). Travel Time Forecasting and Dynamic OD Estimation in Freeways Based on Bluetooth Traffic Monitoring. *89th*

Transportation Research Board 2010 Annual Meeting, (November 2009), 18. Retrieved from http://amonline.trb.org

Jackson, S., Lesani, A., & Moreno, L. F. (2014). Towards a WIFI system for traffic monitoring in different transportation facilities. *Transportation Research Board, 93rd Annual Meeting*, (January), 1–19. Retrieved from https://trid.trb.org/view/1290003

Jiang, M., Fan, X., Zhang, F., Xu, C., Mao, H., & Liu, R. (2016). Characterizing On-Bus WiFi Passenger Behaviors by Approximate Search and Cluster Analysis. *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, 31–36. https://doi.org/10.1109/CCBD.2016.017

Kabir, M. A., & Khan, A. M. A. (2010). *Comparison Among Short Range Wireless Networks: Bluetooth, Zigbee and Wi-Fi*.

Khliefat, I., & Shatnawi, I. (2017). An Optimization of Bluetooth Sensor Locations for Origin-Destination in an Urban Network. *Journal of Transportation Technologies*, *07*(04), 367–375. https://doi.org/10.4236/jtts.2017.74024

Kurkcu, A., & Ozbay, K. (2017). Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, *2644*(August), 72–82. https://doi.org/10.3141/2644-09

Lesani, A., & Moreno, L. F. (2016). Development and Testing of a Real-Time WiFi-Bluetooth System for Pedestrian Network Monitoring and Data Extrapolation. *Transportation Research Board, 95th Annual Meeting*, (January), 1–13. Retrieved from https://trid.trb.org/view/1394097

Lindemann, A., Schnor, B., Sohre, J., & Vogel, P. (2016). Indoor Positioning: A Comparison of WiFi and Bluetooth Low Energy for Region Monitoring. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies* (Vol. 5, pp. 314–321). SCITEPRESS - Science and and Technology Publications.

https://doi.org/10.5220/0005704603140321

Mai, V. H., Kusakabe, T., Suga, Y., & Oguchi, T. (2017). Travel Time Estimation in Mixed Traffic Using Wi-Fi Detector Based Data. *ISCTSC 11th International Conference on Transport Survey Methods*. Retrieved from http://www.isctsc2017.ca/workshops/isctsc2017_workshop_b3.pdf

Malinovskiy, Y., Saunier, N., & Wang, Y. (2012). Analysis of Pedestrian Travel with Static Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, *2299*, 137–149. https://doi.org/10.3141/2299-15

Man, B. M. (2002). Bluetooth and Wi-Fi. *Socket Communications*, (March), 1–13. Retrieved from http://www1.frm.utn.edu.ar/teleinformatica/docs/TechBriefWireless.pdf

Matte, C. (2017). Wi-Fi Tracking: Fingerprinting Attacks and Counter-Measures. *DoctorateThesis Université de Lyon*, 1–165. Retrieved from https://hal.archives-ouvertes.fr/tel-01659783

Meng-Lung Lin, Chien-Min Chu, C.-H., & Tsai, C.-C. C. and C.-Y. C. (2009). Geovisualization of Tourist Activity Travel Patterns Using 3D GIS: An empirical study of Tamsui, Taiwan. *International Journal of Civil and Architectural Engineering*, *3*(12), 394–398. Retrieved from https://waset.org/Publication/geovisualization-of-tourist-activity-travel-patterns-using-3d-gis-an-empirical-study-of-tamsui-taiwan/733

Metageek. (2018). Understanding WiFi Signal Strength. Retrieved from https://www.metageek.com/training/resources/wifi-signal-strength-basics.html

Musa, A. B. M., & Eriksson, J. (2012). Tracking unmodified smartphones using wi-fi monitors. *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems - SenSys '12*, 281. https://doi.org/10.1145/2426656.2426685

Najafi, P., Georgiou, A., & Shachneva, D. (2014). Privacy Leaks from Wi-Fi Probing. *Report*, (March), 1–19. Retrieved from https://andreasgeo.com/wpcontent/uploads/2014/06/Privacy-Leaks-from-Wi-Fi-Probing.pdf

Namaki Araghi, B., Skoven Pedersen, K., Tørholm Christensen, L., Krishnan, R., & Lahrmann, H. (2015). Accuracy of Travel Time Estimation Using Bluetooth Technology: Case Study Limfjord Tunnel Aalborg. *International Journal of Intelligent Transportation Systems Research*, *13*(3), 166–191. https://doi.org/10.1007/s13177-014-0094-z

Nishide, R., & Takada, H. (2013). Detecting Pedestrian Flows on a Mobile Ad Hoc Network and Issues with Trends and Feasible Applications. *International Journal on Advances in Networks and Services*, *6*(1), 108–117. Retrieved from http://www.iariajournals.org/networks_and_services/tocv6n12.html

Odiyo, M. O. (2014). *Use of GIS in Mapping the Spatial Distribution and Security of Wi-Fi Networks*. Geospatial and Space Technology.

Ohmori, N., Muromachi, Y., Harata, N., & Ohta, K. (2002). Analysis of Day-to-Day Variations of Travel Time Using GPS and GIS. *Proceedings of the Conference on Traffic and Transportation Studies, ICTTS*, 1–8.

Oransirikul, T., Nishide, R., Piumarta, I., & Takada, H. (2014). Measuring Bus Passenger Load by Monitoring Wi-Fi Transmissions from Mobile Devices. *Procedia Technology*, *18*, 120–125. https://doi.org/10.1016/j.protcy.2014.11.023

Pandas. (2019). Python Data Analysis Library 0.25.3 Version. *Python Data Analysis Toolkit*. Retrieved from https://pandas.pydata.org/

Pattanusorn, W., Nilkhamhang, I., Kittipiyakul, S., Ekkachai, K., & Takahashi, A. (2016). Passenger estimation system using Wi-Fi probe request. In *2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)* (Vol. 2016, pp. 67–72). IEEE. https://doi.org/10.1109/ICTEmSys.2016.7467124

Petre, A.-C., Chilipirea, C., Baratchi, M., Dobre, C., & van Steen, M. (2017). WiFi Tracking of Pedestrian Behavior. In *Smart Sensors Networks* (pp. 309–337). Elsevier.

https://doi.org/10.1016/B978-0-12-809859-2.00018-8

Petre, A., Chilipirea, C., & Baratchi, M. (2016). WiFi tracking of pedestrian behavior Principles of WiFi tracking, 1–24.

Pojani, D., & Stead, D. (2015). Sustainable Urban Transport in the Developing World: Beyond Megacities. *Sustainability*, *7*(6), 7784–7805. https://doi.org/10.3390/su7067784

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

Pothuganti, K., & Chitneni, A. (2014). A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi. *Advance in Electronic and Electric Engineering*, *4*(6), 655–662. Retrieved from http://eee.guc.edu.eg/Announcements/Comparaitive_Wireless_Standards.pdf

Poucin, G., Farooq, B., & Patterson, Z. (2016). Pedestrian Activity Pattern Mining in WiFi-Network Connection Data. *Transportation Research Board, 95th Annual Meeting*, (August 2015), 20p. Retrieved from https://trid.trb.org/view/1394184

Pourhassan, K. (2016). How Reliable Is the Bluetooth-Based Origin-Destination Data ? *The Conference of Australian Institutes of Transport Research (CAITR)*, 1–12.

QGIS. (2019). Nearest Neighbor Analysis. Retrieved from https://www.qgistutorials.com/en/docs/nearest_neighbor_analysis.html

Romancyshyn, T. (2016). *The Use of Bluetooth and Smartphone GPS Data to Investigate Factors Associated to Variations of Travel Times and Delays*. McGill University.

Ryeng, E. O., Haugen, T., Grønlund, H., & Overå, S. B. (2016). Evaluating Bluetooth and Wi-Fi Sensors as a Tool for Collecting Bicycle Speed at Varying Gradients. *Transportation Research Procedia*, *14*(2352), 2289–2296. https://doi.org/10.1016/j.trpro.2016.05.245

Sapiezynski, P., Stopczynski, A., Gatej, R., & Lehmann, S. (2015). Tracking Human Mobility

Using WiFi Signals. *PLOS ONE*, *10*(7), e0130824. https://doi.org/10.1371/journal.pone.0130824

Schauer, L., Werner, M., & Marcus, P. (2014). Estimating Crowd Densities and Pedestrian Flows Using Wi-Fi and Bluetooth. *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 171–177. https://doi.org/10.4108/icst.mobiquitous.2014.257870

Shiravi, S., Hossain, K., Fu, L., & Ghods, A. (2016). Evaluation of Using Wifi Signals to Estimate Intersection Travel Time. *Proceedings, Annual Conference - Canadian Society for Civil Engineering (2016) 4*, (August), 1–9.

Shlayan, N., Kurkcu, A., & Ozbay, K. (2016). Exploring pedestrian Bluetooth and WiFi detection at public transportation terminals. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 229–234. https://doi.org/10.1109/ITSC.2016.7795559

Sun, L., Chen, S., Zheng, Z., & Xu, L. (2017). Mobile Device Passive Localization Based on IEEE 802.11 Probe Request Frames. *Mobile Information Systems*, *2017*, 1–10. https://doi.org/10.1155/2017/7821585

Tolouei, R., Psarras, S., & Prince, R. (2017). Origin-Destination Trip Matrix Development: Conventional Methods versus Mobile Phone Data. *Transportation Research Procedia*, *26*, 39–52. https://doi.org/10.1016/j.trpro.2017.07.007

Vanhoef, M., Matte, C., Cunche, M., Cardoso, L. S., & Piessens, F. (2016). Why MAC Address Randomization is not Enough. *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security - ASIA CCS '16*, 413–424. https://doi.org/10.1145/2897845.2897883

Verbree, E., Zlatanova, S., van Winden, K. B. A., van der Laan, E. B., Makri, A., Taizhou, L., & Haojun, A. (2013). To localise or to be localised with WiFi in the Hubei museum? *ISPRS*

- *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-4/W4*(4W4), 31–35. https://doi.org/10.5194/isprsarchives-XL-4-W4-31-2013

Wismans, L. J. J., Friso, K., Rijsdijk, J., de Graaf, S. W., & Keij, J. (2018). Improving A Priori Demand Estimates Transport Models using Mobile Phone Data: A Rotterdam-Region Case. *Journal of Urban Technology*, *25*(2), 63–83. https://doi.org/10.1080/10630732.2018.1442075

Wu, D., & Deng, W. (2013). Characteristics Analysis of Urban Traffic in Small and Medium-Sized Cities in Context of Mechanization--with Huai'an as an Example. *Procedia - Social and Behavioral Sciences*, *96*(Cictp), 2240–2247. https://doi.org/10.1016/j.sbspro.2013.08.253

Xu, Z., Sandrasegaran, K., Kong, X., Zhu, X., Zhao, J., Hu, B., & Chung Lin, C.-. (2013). Pedestrain Monitoring System using Wi-Fi Technology And RSSI Based Localization. *International Journal of Wireless & Mobile Networks*, *5*(4), 17–34. https://doi.org/10.5121/ijwmn.2013.5402

Yaik, O. B., Wai, K. Z., K.T.Tan, I., & Sheng, O. B. (2016). Measuring the Accuracy of Crowd Counting using Wi-Fi Probe-Request-Frame Counting Technique. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *8*(2), 79–81. Retrieved from http://journal.utem.edu.my/index.php/jtec/article/view/961

Young, S. E. (2012). Bluetooth Traffic Detectors for Use as Permanently Installed Travel Time Instrument. *Report No.MD-12-SP90984D UNIVERSITY OF MARYLAND*, 15. Retrieved from http://roads.maryland.gov/OPR_Research/MD-12-SP909B4D-Bluetooth-Traffic-Detectors_Report.pdf

Yu, H., & Shaw, S. (2004). Representing and visualizing travel diary data: a spatio-temporal GIS approach. *2004 ESRI International User Conference*, (April), 1–13. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.6247&rep=rep1&type=pdf

Zambrano, G. R., Tepán, F. M., Jaime, A. P., & Guevara, J. M. (2016). Visualization and analysis of vehicle paths with QGis and Weka. *International Journal of Innovation and Applied Studies*, *18*(4), 9324. Retrieved from http://www.ijias.issr-journals.org/