# EVALUATION OF OUTLIER FILTERING ALGORITHMS FOR ACCURATE MEASUREMENT OF TRAVEL TIME RELIABILITY INCORPORATING LANE-SPLITTING SITUATIONS

OBADA M. A. ASQOOL

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2022

# EVALUATION OF OUTLIER FILTERING ALGORITHMS FOR ACCURATE MEASUREMENT OF TRAVEL TIME RELIABILITY INCORPORATING LANE-SPLITTING SITUATIONS

**OBADA M. A. ASQOOL**

**DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING SCIENCE**

**FACULTY OF ENGINEERING**
**UNIVERSITY OF MALAYA**
**KUALA LUMPUR**

**2022**

**UNIVERSITY OF MALAYA**

**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: OBADA M. A. ASQOOL

Matric No: 17198995/1 (Old: KGA180041)

Name of Degree: Master of Engineering Science

Title of Dissertation ("this Work"):

Evaluation of Outlier Filtering Algorithms for Accurate Measurement of Travel Time Reliability Incorporating Lane-Splitting Situations

Field of Study: Transportation Engineering (Civil Engineering)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 29-May-2022

Subscribed and solemnly declared before,

Witness's Signature Date:02-06-2022

Name:

Designation:

**EVALUATION OF OUTLIER FILTERING ALGORITHMS FOR ACCURATE MEASUREMENT OF TRAVEL TIME RELIABILITY INCORPORATING LANE-SPLITTING SITUATIONS**

## ABSTRACT

Most developed and developing countries, including Malaysia, experience high traffic congestion, especially in urban areas. When congestion occurs, traffic moves at a lower speed which increases travel time and in consequence, people spend considerable time fulfilling their daily journeys. Measuring Travel Time Reliability (TTR) helps traffic professionals to quantify congestion based on travel time parameter, thus, adopting suitable strategies to mitigate traffic congestion. On the other hand, motorcycles are significantly high in Malaysia and other Association of Southeast Asian Nations (ASEAN) countries. As such, detection of travel time based on the media access control (MAC) address is not straightforward. Raw travel times data does not represent the actual traffic condition of passenger cars since motorcycles travel faster than cars during congestion using the gap between two parallel rows of cars. This situation is called lane-splitting. In the past, many travel time filtration algorithms were established. However, there is a real need to determine which algorithms can produce the most accurate results when considering actual and large datasets from lane-splitting situations. Therefore, this study aims to investigate the best algorithm for data filtration and how to use it to obtain accurate data for measuring TTR. In order to find the best filtration algorithm, two stages were adopted in this study. The first stage was the validation of the performance of the previous algorithms. The assessment was conducted by observing the performance of each algorithm and comparing its performance with other algorithms when lane-splitting data were applied. The second stage was to investigate the sensitivity of the algorithm parameters for different days. To analyse TTR, Travel Time Index (TTI), Planning Time Index (PTI), and Buffer Time Index (BTI) were calculated with respect to the time of day

(TOD), day of week (DOW), holidays, and election days. This study used travel time datasets collected from three routes in Kuala Lumpur via Wi-Fi detectors during May 2018. The results showed that the Jang algorithm was found to have the best performance for two of the three routes, whereas the TransGuide algorithm was the best algorithm for one route. The parameters of the Jang algorithm and TransGuide algorithm were sensitive for different days. Thus, the Jang algorithm and TransGuide algorithm could be used after calibrating their parameters for each day. After the filtration, TTR measures were calculated. The analysis of TTR measures showed that on weekdays and weekends, the three routes suffered from high variability in travel time. On election days and holidays, the road network operates near to free-flow condition for most of the time with low variability in travel time. The findings and contribution of this research are beneficial for transportation companies and authorities that depend on MAC addresses to collect travel time data in Malaysia and ASEAN countries. Furthermore, this research introduced the concept of TTR, demonstrating its importance to Malaysian traffic researchers. Accordingly, Malaysian transportation authorities need to adopt TTR measures in their studies, reports, and guidelines to maintain reliable travel time.

**Keywords:** Travel Time, Travel Time Reliability, Lane-Splitting, MAC Address, Outlier Filtering Algorithm.

## PENILAIAN ALGORITMA PENAPISAN OUTLIER UNTUK KEBOLEHPERCAYAAN PENGUKURAN TEPAT MASA PERJALANAN YANG MENYERTAKAN SITUASI PEMISAHAN LORONG

## ABSTRAK

Sebilangan besar negara maju dan sedang membangun, termasuk Malaysia, mengalami kesesakan lalu lintas yang tinggi, terutamanya di kawasan bandar. Apabila kesesakan berlaku, lalu lintas bergerak pada kelajuan yang lebih rendah lalu meningkatkan masa perjalanan dan akibatnya, orang menghabiskan banyak masa dalam perjalanan harian mereka. Mengukur Kebolehpercayaan Masa Perjalanan (KMP) boleh membantu profesional lalu lintas untuk mengukur kesesakan berdasarkan parameter masa perjalanan, dengan demikian, berupaya menerapkan strategi yang sesuai untuk mengurangkan masalah yang berkaitan dengan kesesakan. Sebaliknya, jumlah motosikal juga sangat tinggi di Malaysia dan negara-negara Persatuan Negara-negara Asia Tenggara (ASEAN) yang lain. Oleh kerana itu, pengesanan masa perjalanan berdasarkan alamat kawalan akses media (MAC) adalah tidak mudah. Data mentah masa perjalanan tidak menunjukkan keadaan lalu lintas sebenar kerana motosikal bergerak lebih pantas daripada kereta semasa kesesakan, di mana mereka menggunakan jurang antara dua barisan kereta selari. Keadaan ini disebut *lane-splitting*. Pada masa lalu, banyak algoritma penapisan masa perjalanan telah dibuat. Walau bagaimanapun, ada keperluan sebenarnya untuk menentukan algoritma mana yang dapat menghasilkan hasil yang paling tepat ketika mempertimbangkan set data yang sebenar dan besar dari situasi *lane-splitting*. Kajian ini bertujuan untuk menyelidik algoritma terbaik untuk penyaringan data dan menggunakannya untuk mendapatkan data yang tepat untuk mengukur KMP. Untuk mencari algoritma penapisan terbaik, dua peringkat telah dilaksanakan. Tahap pertama adalah pengesahan prestasi algoritma sebelumnya. Penilaian dilakukan dengan memerhatikan prestasi setiap algoritma dan membandingkan prestasinya dengan

algoritma lain ketika data *lane-splitting* digunakan. Tahap kedua adalah untuk menyiasat kepekaan parameter algoritma untuk hari yang berbeza. Untuk menganalisis KMP, Indeks Masa Perjalanan (IMP), Indeks Waktu Perancangan (IWP), dan Indeks Waktu Penampan (IWPen) telah dikira dengan mengambil kira waktu dalam hari (WDH), hari dalam seminggu (HDS), hari cuti, dan hari mengundi. Dataset masa perjalanan, pada bulan Mei 2018, dari tiga laluan di Kuala Lumpur yang dikumpulkan oleh pengesan Wi-Fi telah digunakan dalam analisis. Hasil kajian menunjukkan bahawa algoritma Jang didapati mempunyai prestasi terbaik untuk dua dari tiga laluan tersebut. Manakala, algoritma TransGuide adalah algoritma terbaik untuk satu laluan. Parameter algoritma Jang dan algoritma TransGuide adalah sensitif untuk hari yang berbeza. Oleh itu, algoritma Jang dan algoritma TransGuide dapat digunakan setelah menukar parameternya untuk setiap hari. Selepas penapisan, langkah-langkah KMP telah dikira. Analisis langkah-langkah KMP menunjukkan bahawa pada hari kerja dan hujung minggu, ketiga-tiga laluan mengalami kebolehubahan yang tinggi dalam masa perjalanan. Pada hari pilihan raya dan hari cuti, rangkaian jalan raya beroperasi dengan keadaan aliran bebas hampir sepanjang masa dengan kebolehubahan masa perjalanan yang rendah. Hasil kajian ini bermanfaat bagi syarikat pengangkutan dan pihak berkuasa yang bergantung pada alamat MAC untuk mengumpulkan data masa perjalanan di Malaysia dan negara-negara ASEAN. Selanjutnya, penyelidikan ini memperkenalkan konsep KMP dan menunjukkan kepentingannya kepada penyelidik trafik Malaysia. Pihak berkuasa pengangkutan Malaysia harus mengambil langkah-langkah KMP dalam kajian, laporan, dan panduan mereka untuk mengekalkan masa perjalanan yang dipercayai.

**Kata kunci**: Masa Perjalanan, Kebolehpercayaan Masa Perjalanan, *Lane-splitting*, Alamat MAC, Algoritma Penapisan *Outlier*

## ACKNOWLEDGEMENTS

All thanks be to Allah, the Most Merciful and Compassionate, whose incredible benevolence and great guidance helped me to accomplish this dissertation. I send the highest regards and sincere salutation upon his messenger, his family, the companions, and the followers.

Firstly, it is my utmost pleasure to dedicate this work to my dear parents, my family, and my fiancée, who granted me the gift of their unwavering belief in my ability to accomplish this goal: thank you for your support and patience.

My utmost gratitude goes to the very supportive supervisors, Dr Suhana Koting and Dr Ahmad Saifizul Abdullah, for their valuable suggestions, guidance, thorough editing, intuitive comments, and meticulous revisions of my work. The success of this work largely depends on their painstaking efforts.

Finally, I send my appreciation to all my lecturers, staff, and colleagues of the Faculty of Engineering for their contributions toward making this endeavour a reality.

OBADA M. A. ASQOOL, 2022

**TABLE OF CONTENTS**

# LIST OF FIGURES

# LIST OF TABLES

## LIST OF SYMBOLS AND ABBREVIATIONS

| | | |
|---|---|---|
| ASEAN | : | Association of Southeast Asian Nations |
| ATIS | : | Advanced Traffic Information System |
| AVI | : | Automatic vehicle identification |
| BTI | : | Buffer time index |
| DBKL | : | Dewan Bandaraya Kuala Lumpur/Kuala Lumpur City Hall |
| DOW | : | Day of week |
| FHWA | : | Federal highway administration |
| ITSSB | : | Integrated Transportation Solutions Sdn. Bhd. |
| KL | : | Kuala Lumpur |
| KLCC | : | Kuala Lumpur city center |
| LOS | : | Level of service |
| MAC | : | Media access control |
| MAD | : | Mean absolute deviation |
| PTI | : | Planning time index |
| RFID | : | Radio-Frequency identification |
| TCI | : | Traffic congestion index |
| TOD | : | Time of day |
| TTI | : | Travel time index |
| TTR | : | Travel time reliability |
| TTV | : | Travel time variability |
| UCR | : | Urban congestion reports |
| US | : | United States |

# LIST OF APPENDICES

# CHAPTER 1: INTRODUCTION

## 1.1 Background

Transportation has been important throughout the ages, given its direct benefit to humanity and life in general. Transportation has also been closely associated with the civilisational development of societies (Fenelon, 2017). As a result of the importance of transportation and its link to the economy, governments, universities and research centres continue to develop more updated transportation systems and alleviate problems that arise from it. The development strategies have three directions: effective management of existing transportation facilities, building new facilities, and the invention of new modes of transportation, for example, autonomous vehicles.

Nowadays, traffic congestion is one of the most common issues faced by road users and traffic consultants, and given the rising population in most countries worldwide, this problem will continue to become an issue in the future. Wang, Guo, and Yu (2018) stated that to mitigate the congestion problem, the first step is to study the current situation using the traffic congestion index (TCI) to reflect the traffic flow status. After that, traffic engineers can use the TCI for planning and managing road networks. Nowadays, people spend considerable time travelling each day due to increased traffic demand and limited road capacities. The journey's time varies, especially in urban areas, given many factors such as road accidents, weather conditions, changing traffic demands, traffic control, and road types.

While travel time is a key parameter used to evaluate traffic congestion, it is also more understandable for travellers to comprehend than other traffic parameters such as traffic density and traffic flow because travel time represents travellers' experience. The variability in travel time is a major problem for travellers. They would rather travel a route with higher mean travel times and smaller variability than routes with lower mean travel times and larger variability (Lyman & Bertini, 2008; Van Lint & van Zuylen, 2005). Therefore, quantifying the variability suitably helps operators and travellers make better-informed decisions (Xu, Jabari, & Prassas, 2020). However, doing so requires collecting travel time data extensively.

Due to the development of advanced traffic sensing technologies, travel time data can be collected via various data sources, such as the automatic licence plate identification technique, probe vehicles technique, media access control (MAC) address-based technique and Radio-Frequency Identification (RFID) technique (Z. Chen & Fan, 2020; Guo et al., 2019). These technologies assist researchers to conduct studies to investigate variability in travel time.

Each data source is affected by sources of error. For instance, MAC address detectors are affected by (a) en route stops, (b) nonauto observations, and (c) the presence of multiple devices in a single-vehicle (Moghaddam & Hellinga, 2013). These error sources generate outliers in travel time datasets. Some situations generate abnormal data, which, in turn, lead to inaccuracy in travel time datasets. The lane-splitting situation is an example. The outliers and lane-splitting data need to be removed from datasets to obtain accurate travel time data

Lane splitting is defined as “the practice of passing slower moving traffic by riding a motorcycle in the gap between two parallel lanes of traffic heading in the same direction” (Ouellet, 2012). The United States, except for California, prohibits this practice (Rice, Troszak, & Erhardt, 2015). However, it is legal in several European countries (Kurlantzick & Krosner, 2016). In 2014, the state of New South Wales in Australia allowed lane-splitting at a speed less than 30 km/h (Rice et al., 2015). The 30 km/h restriction was imposed for safety concerns (Vanessa Beanland, 2018). Some researchers recommended governments consider legalizing lane-splitting for motorcycles riding in only slow-moving or stopped traffic (V Beanland, Pammer, Sledziowska, & Stone, 2015; Kurlantzick & Krosner, 2016).

Allowing motorcycles to travel freely through traffic reduces the level of congestion. In addition, it encourages people to use motorcycles because the travel time is less than cars, especially at peak periods (Sperley & Pietz, 2010). A study in Paris, France, showed that the average lane-splitting speed of motorcycle riders in stopped traffic on a ring road is 38 km/h (Aupetit, Espié, & Bouaziz, 2015), indicating a big difference in the travel time of lane-splitting motorcycles and passenger vehicles. The frequency of motorcycles lane-splitting decreases as traffic speed increases (Ouellet, 2012). However, there are important safety concerns relating to lane-splitting. Unexpected vehicles opening doors and changing lanes without indicating lead to safety hazards for motorcycles that travel between rows of vehicles (Sperley & Pietz, 2010). In certain countries where the number of motorcycle riders is high, the lane-splitting situation is obvious, affecting the accuracy of travel time measurement, especially during congestion.

## 1.2 Statement of Problem

Kuala Lumpur (KL) is one of many congested capitals globally. In fact, it is the 46th most congested city from 416 cities based on the TomTom traffic index for 2019. This index shows an increase in traffic congestion levels in KL from 36% in 2018 to 37% in 2019 (TomTom International BV, n.d.). Between 2010 and 2014, there was a 28% increment in total motor vehicles in KL, which signifies a rapid increase in traffic demand. The congestion problem will continue to become a more serious issue because the increase in road capacity does not match the rise in traffic demand (Abdullah, Ramli, & Mohamad, 2017). Traffic congestion leads to economic and healthy issues. The estimated total cost of congestion in Malaysia was around RM 13.09 billion per year, where (RM 10.82 billion/year) represents the total wage loss, (RM 1.08 billion/year) represents the total fuel loss, and (RM 1.19 billion/year) represents the total loss on environmental impact (Ministry of works Malaysia, 2019).

Notably, travel time is a significant parameter to assess traffic conditions and congestion on roadways (Martchouk, Mannering, & Bullock, 2011). Furthermore, accurate travel time is pivotal for road operators to evaluate, monitor, and manage the performance of road networks and for road users to plan for their journeys (Liu, Xia, & Phatak, 2020).

In Malaysia, the percentage of motorcycles using our roads is very high. For example, in mid-2017, the number of registered motorcycles was 12,933,042, around 46% of the total registered vehicles (Lee, 2017). The significant difference in the travel time of lane-splitting motorcycles and passenger vehicles means that it is essential to remove lane-splitting observations from the travel time data when developing travel time measures (e.g., average travel time and travel time index) for passenger vehicles. Retaining the lane-splitting observations in the travel time data causes incorrect passenger vehicle travel time measures. Given Malaysia's high percentage of motorcycles, lane-splitting significantly impacts travel time measurement, particularly at peak periods. Thus, exploring the best filtering algorithm is crucial to minimise errors, thus, obtaining accurate travel time measurements.



## 1.3 Research Gap

Malaysia and other ASEAN countries have unique traffic patterns, given the high percentage of motorcycle users on our roads. On the other hand, this traffic pattern does not exist in other developed countries due to the low percentage of motorcycle users travelling on main roads. For this reason, the effectiveness of most published travel time filtration algorithms needs to be verified when deploying actual empirical data from ASEAN countries.

The aim of measuring travel times using MAC addresses is to obtain accurate results of the passenger vehicles. In Malaysia and other ASEAN countries, the error in the datasets is due to the motorcycles data during congestion (lane-splitting data) and the outliers. Because of the significant difference in the travel time of lane-splitting motorcycles and passenger vehicles, it is essential to filter out the lane-splitting data using a filtering algorithm to obtain an accurate travel time pattern for passenger vehicles. Several researchers have developed algorithms for filtering outliers. At present, there is no evaluation of the algorithms using travel time datasets that contain lane-splitting observations. Therefore, it is essential to evaluate the well-known filtering algorithms to identify the best algorithm for filtering outliers and lane-splitting observations.

Many algorithms have been proposed to identify and minimise outliers in travel time datasets (Clark, Grant-Muller, & Chen, 2002; Dion & Rakha, 2006; Jang, 2016; Southwest Research Institute, 1998). However, the proposed algorithms were not validated using empirical data, representing actual lane-splitting situations.

Further to the issue related to the filtration algorithms, it is also important to highlight the concept of travel time reliability (TTR) since it is a relatively recent term in traffic engineering. It has gained much attention due to the development of travel time collection technologies, where accurate and vast travel time data can be gathered. The Federal Highway Administration (FHWA) has adopted a formal definition of TTR, that is, "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day" (Texas Transportation Institute & Cambridge Systematics Inc., 2005). Travel Time Reliability (TTR) measures are "the tools used to evaluate the extent of unreliability or variability in travel time for specific parts of the road network" (Tsolakis, Tan, & Makwasha, 2011). Examples of these measures include the travel time index (TTI), planning time index (PTI), and buffer time index (BTI). These measures are calculated using travel time data.

Having said that, there is no published report for TTR in Malaysia. Unlike Malaysia, most developed countries publish TTR measures reports. For example, the United States (US) Department of Transportation periodically publishes urban mobility and urban congestion reports (UCR) to communicate recent trends in congestion and strategies to manage congestion and improve mobility. The reports contain TTR measures that gauge progress toward meeting transportation system management and operations objectives (Federal Highway Administration (FHWA), 2020). Malaysia could acquire better planning, management, and operations for its transportation system by adopting TTR measures.

## 1.4 Research Questions

The research questions that arise from the problem statement include the following:

1- What is the best algorithm that can filtrate outliers and lane splitting observations?
2- Are parameters of the best algorithms sensitive for different days?
3- Is it possible to measure TTR on Malaysian roads?
4- How do TTR measures vary concerning TOD, DOW, holidays and election days?

## 1.5 Research Objectives

The main objectives of this research include the following:

1- To evaluate the previously established filtration algorithms in terms of accurate representation of the actual situation and the sensitivity of the parameters using travel time data with lane-splitting observations and outliers;
2- To identify the best filtration algorithm with certain calibration procedures when incorporating lane-splitting situations; and

3- To produce accurate travel time reliability measures based on filtered data on selected routes in Kuala Lumpur (KL).

## 1.6 Research Significance

The output from this research is anticipated to mainly benefit traffic consultants, road and local authorities, logistics companies, and road users at large since the research provides valuable insight in calculating and measuring travel time for congestion measure purposes.

So far, TTR measures have not appeared in transportation research in Malaysia and, consequently, are not used by Malaysian road authorities. The potential reasons include the following:

1. Given that TTR had only recently been introduced as a performance measure, it may not have received much attention from the authorities.
2. Special data sources are needed to measure travel time, unlike traffic flow and density data sources.
3. Historically, authorities have focused on construction and maintenance and less operations focused, performance-driven, and aspects associated with user experience.

It is anticipated that this research will help Malaysian authorities to overcome the first two reasons. In order to overcome the first reason, this research introduces the concept of TTR, demonstrating its importance to Malaysian traffic researchers. Also, to overcome the second reason, using MAC address sensors, travel time can be collected and filtered out using the best algorithm as discussed in the results.

## 1.7 Dissertation Structure

This dissertation is organised into five chapters:

**Chapter 1:** Presents the background of the research, statement of the problem, research gap, research questions, research objectives, research significance, and dissertation structure.

**Chapter 2:** Presents the literature review related to the research and includes three main sections: traffic congestion, travel time filtration algorithms, and travel time reliability.

**Chapter 3:** Presents the research methodology flowchart, study area, data collection, data description, and data analysis.

**Chapter 4:** Presents the results and discussion on filtration of travel time data and analysis of travel time reliability measures for the area under study.

**Chapter 5:** Presents a summary of the entire research where the main findings and conclusions are highlighted. In addition, research limitations and recommendations for future research are presented.

# CHAPTER 2: LITERATURE REVIEW

This chapter systematically reviews the literature on traffic congestion, travel time filtration algorithms and travel time reliability (TTR). The review focuses on defining and measuring traffic congestion and traffic problems in ASEAN countries in the traffic congestion section. In the travel time filtration algorithms section, the review focuses on the percentile algorithm, mean absolute deviation algorithm, TransGuide algorithm, Dion and Rakha algorithm, and Jang algorithm. In the TTR section, the review focuses on TTR definitions and measures, factors affecting reliability, recurrent and non-recurrent congestion, travel time variability, and the importance of TTR.

## 2.1 Traffic Congestion

### 2.1.1 Defining and Measuring Traffic Congestion

Traffic congestion is one of the significant problems that threaten our societies leading to time and energy wastage, pollution, stress, and decreased productivity (Bertini, 2006; Rao & Rao, 2012). Although researchers have proposed various definitions of traffic congestion, they have failed to provide a universally accepted definition (Aftabuzzaman, 2007). Many studies have attempted to paraphrase the following definition: “Congestion is a situation in which demand for road space exceeds supply” (Transportation Association of Canada, 2017). While other studies defined traffic congestion based on user’s expectation as in the following definition: Congestion refers to an excess of vehicles on a section of road at a particular time, resulting in slower or much slower speeds than normal or "free flow" speeds. Moreover, the term is related to stopped or stop-and-go traffic (Cambridge Systematics Inc., 2005).

To select effective congestion mitigation measures, the identification of congestion characteristics is a crucial primary step (Rao & Rao, 2012). Thus, for road agencies, the development of congestion measures to alleviate congestion is a high priority task.

However, congestion is not always quantified uniformly since traffic congestion directly affects the speed, travel time, quality of services, delay, travel cost, fuel consumption, and emission of pollutants. These effects are widely considered in the literature to understand the implications of traffic congestion. Based on the aspects affected by congestion, there are seven groups of traffic congestion measures: travel time, TTR, speed, delay, level of service (LOS), cost, and environmental measures (Transportation Association of Canada, 2017).

### 2.1.2 Traffic Congestion and Traffic Problems in ASEAN.

In the ASEAN region, economic growth has led to rapid traffic growth. Since 2000, the motorisation growth has increased rapidly. In ASEAN countries like Vietnam, Indonesia, Myanmar, and Cambodia, between 2000 and 2010, the number of automobiles more than doubled. However, this increment was not accompanied by the corresponding expansion in infrastructure and road services leading to traffic congestion, especially in ASEAN's major cities. (ERIA Study Team, 2010).

Traffic congestion, high rates of traffic accidents, air pollution from transport sources, and inadequate access to transport facilities are serious problems many cities face because of rapidly growing motorisation. Indeed, major cities and capital cities in ASEAN countries suffer from the most rapid increase in motorisation. Two and three-wheeler vehicles are the dominant transport mode in cities like Ho Chih Minh and Hanoi. Here, the central areas of the capitals are highly congested, with the traffic speeds reported to be extremely low during weekday peak hours, resulting in longer travel times. As a result, the cities' liveability and productivity are threatened by the deteriorating urban environment. (ERIA Study Team, 2010).

In developing countries, studies quantifying congestion are limited due to the lack of traffic datasets. Congestion indexes (travel time indexes) have been calculated for 278

Asian cities using trip data from Google Maps. In 2016, about 24 of the cities had populations above 5 million; these cities were the largest. After calculating the congestion index for the largest cities, the most three congested cities were Metro Manila, Kuala Lumpur (KL), and Yangon City, respectively, which fall within the ASEAN region. This means that the mitigation of traffic congestion in the ASEAN region requires greater effort (Asian Development Bank, 2019).

In Malaysia, the total number of registered motorcycles in 2016 was 12.68 million, accounting for 45.9% of the total number of registered road motor vehicles. Across all ASEAN countries, in 2016, the total number of registered motorcycles was 199 million, representing about 76.4% of the total number of registered road motor vehicles (ASEAN Statistics Division, 2018). These statistics imply that motorcycles were the major means of transport in ASEAN. A contributing factor to this figure is because motorcycles are relatively affordable for most people in the ASEAN region. Furthermore, the proportion of youth in the ASEAN population is enormous. However, on the other hand, the motorcycle regulatory framework in ASEAN is underdeveloped. Moreover, the place or position of motorcycles in traffic facilities design and urban transportation planning is not equal to the very high use of motorcycles. Motorcycles were also a significant factor in accidents and traffic congestion (Kitamura, Hayashi, & Yagi, 2018).

Since effective transportation systems are critical for promoting international trade with and within the ASEAN region, it is important to remove the barriers to trade in the land transport sector. The unprecedented commitment of skills and resources and new levels of cooperation are required to remove the barriers in ways that protect the natural environment, promote growth, and meet the complex demands arising from new safety requirements. So, to maintain a sustainable environment, public transport services need to be improved and developed. It is also necessary to improve street furniture and pedestrian facilities and promote bicycles for short-distance travel (ERIA Study Team, 2010).

#### 2.1.2.1 Lane-splitting

In the Asian countries, lane-splitting is a common practice among motorcycle riders (Maulina, Danilasari, Nazhira, & Jufri, 2022). Lane-splitting is defined as "the practice of passing slower moving traffic by riding a motorcycle in the gap between two parallel lanes of traffic heading in the same direction" (Ouellet, 2012). Some researchers differentiate between lane-splitting and lane-filtering, while others use them interchangeably. Those who use them differently consider lane-filtering as the practice of riding a motorcycle between stopped or nearly stopped traffic, and they consider lane-splitting where the traffic is moving (higher than 30 km/h) (V Beanland et al., 2015; Kelly, 2016; Mulvihill et al., 2013).

The United States, except for California, prohibits this practice (Rice et al., 2015). However, it is legal in several European countries (Kurlantzick & Krosner, 2016). In 2014, the state of New South Wales in Australia allowed lane-splitting at a speed less than 30 km/h (Rice et al., 2015). The 30 km/h restriction was imposed for safety concerns (Vanessa Beanland, 2018). Some researchers recommended governments consider legalizing lane-splitting for motorcycles riding in only slow-moving or stopped traffic (V Beanland et al., 2015; Kurlantzick & Krosner, 2016).

Allowing motorcycles to travel freely through traffic reduces the level of congestion. In addition, it encourages people to use motorcycles because the travel time is less than cars, especially at peak periods (Sperley & Pietz, 2010). A study in Paris, France, showed that the average lane-splitting speed of motorcycle riders in stopped traffic on a ring road is 38 km/h (Aupetit et al., 2015), indicating a big difference in the travel time of lane-splitting motorcycles and passenger vehicles. The frequency of motorcycles lane-splitting decreases as traffic speed increases (Ouellet, 2012).

There are safety concerns relating to lane-splitting. Unexpected vehicles opening doors and changing lanes without indicating lead to safety hazards for motorcycles that travel between rows of vehicles (Sperley & Pietz, 2010). However, surprisingly, lane-splitting provides safety benefits to motorcyclists. In California, where lane-splitting is legal, the rate of rear-end collisions between motorcycles and cars is unusually low comparing to the other states that prohibit lane-splitting. In addition, the injuries are less severe if crashes occur during lane-splitting. As such, lane-splitting is safer than waiting in traffic if it is practiced at a low speed (Kurlantzick & Krosner, 2016).

## 2.2 Travel Time Filtration Algorithms

Many algorithms and approaches found in the literature addressed the filtration of outliers in travel time data. Each algorithm has some assumptions and levels of complexity. In this section, the algorithms appropriate for travel time data collected by MAC addresses were reviewed.

### 2.2.1 Percentile Algorithm

This algorithm uses percentiles to define the validity range. Clark et al. (2002) proposed to use the 10th percentile as a lower limit and the 90th percentile as an upper limit. These limits are applied after dividing the times into small equal time windows, usually 5 min or 15 min. This method is simple, easy to be applied, and can be used in real-time applications. However, it does not consider the comparison with the previous time window. Therefore, the filter will not work correctly if the number of outliers is more than the number of true observations, or all the observations are outliers.

### 2.2.2 Mean Absolute Deviation Algorithm

The validity range in this algorithm is defined based on the median and mean absolute deviation ($MAD$). Usually, 5 min or 15 min is appropriate for time windows (Clark et al., 2002).

$$MAD = \frac{\sum_{i=1}^{n} |(JT_i - M_e)|}{n} \tag{1}$$

where, $JT_i$ is the travel time for vehicle $i$; $M_e$ is the median of travel times in the time window, and $n$ is the number of travel time observations in the time window.

The lower limit of the validity range is $M_e - 3\,MAD$ and the upper limit is $M_e + 3\,MAD$.

The method can be readily applied, where it is simple. The usage of this algorithm in real-time applications is also possible. However, it does not consider the comparison with the previous time window. Thus, it has similar shortcomings of the percentile algorithm if the number of outliers is more than the number of true observations, or all observations are outliers.

### 2.2.3 TransGuide Algorithm

The TransGuide algorithm was one of the earliest algorithms proposed by the Southwest Research Institute for automatic vehicle identification (AVI) data. The algorithm defines travel time observation as valid if it rests inside pre-defined travel time limits based on the previous average travel time. These limits represent a validity range (Southwest Research Institute, 1998). The following equations describe this rolling average algorithm:

$$Stt_{AB\,t} = \{t_{Bi} - t_{Ai} | t - t_W \leq t_{Bi} \leq t \ \ and \ \ tt'_{AB\,t}(1 - l_{th}) \leq t_{Bi} - t_{Ai} \leq tt'_{AB\,t}(1 + l_{th})\} \tag{2}$$

$$tt_{AB\ t} = \frac{\sum_{i=1}^{|Stt_{AB\ t}|}(t_{Bi} - t_{Ai})}{|Stt_{AB\ t}|} \tag{3}$$

Eq. (2) defines $Stt_{AB\ t}$, which is a set of valid travel times from point A to point B at time t; these travel times will be used in Eq. (3) to calculate $tt_{AB\ t}$, that is the average travel time for the corresponding set of observations. $t_{Ai}$ represents the detection time of a vehicle i at point A; $t_{Bi}$ represents the detection time of a vehicle i at point B; t is the time at which the travel time estimation occurs; $t_W$ is time window; $tt'_{AB\ t}$ is the previous average travel time from A to B, and $l_{th}$ is the link threshold travel time parameter. The time window $t_W$ and the link threshold travel time $l_{th}$ are the main parameters of the TransGuide algorithm. Here, $t_W$ defines the period of time to be considered when estimating the current average travel time, while $l_{th}$ is used to identify and remove outlier observations. The proposed values for these parameters are 2 min for $t_W$ and 0.2 for $l_{th}$. This means that any travel time observations between a pair of readers that differs by more than 20% from the average travel time associated with the observations made in the previous 2 min would be considered outliers and not considered in the calculation of the current interval average travel time (Dion & Rakha, 2006). This algorithm has a low complexity level. Also, it is relatively easy to understand its mechanism, that is, dividing the time into small windows then making comparisons between the observations in the current window with the average travel time of the previous window. This algorithm can also be applied in real-time applications.

### 2.2.4 Dion and Rakha Algorithm

Dion and Rakha (2006) contended that the TransGuide algorithm could not track abrupt changes in observed travel times with a low sampling rate. So, to address this shortcoming, an enhanced filtering algorithm was proposed, which applies a series of filters to the collected travel times to remove invalid observations, assuming that travel times in a time window follow the lognormal distribution. Furthermore, this algorithm considers any travel time observation that falls outside a validity range defined using the mean and standard deviation as the outlier. The developers of this algorithm proposed several versions as follows:

#### 2.2.4.1 Version 1

$$tts_{AB\ k} = \begin{cases} e^{[(\alpha).\ln(tt_{AB\ k-1})+(1-\alpha).\ln(tts_{AB\ k-1})]} & , \quad n_{vk-1} > 0 \\ tt_{AB\ k-1} & , \quad n_{vk-1} = 0 \end{cases} \tag{4}$$

$$\sigma^2_{stt_{AB}\ k} = \begin{cases} \alpha\,.\left(\sigma^2_{tt_{AB}\ k-1}\right) + (1-\alpha).\ \left(\sigma^2_{stt_{AB}\ k-1}\right), & n_{vk-1} > 1 \\ \sigma^2_{stt_{AB}\ k-1} & , \quad n_{vk-1} = \{0{,}1\} \end{cases} \tag{5}$$

$$\alpha = 1 - \ (1 - \ \beta)^{n_{vk}} \tag{6}$$

$$tt_{AB\min k} = \ e^{\left[\ln(tts_{AB\ k}) - n_{\sigma}.\left(\sigma_{stt_{AB}\ k}\right)\right]} \tag{7}$$

$$tt_{AB\max k} = \ e^{\left[\ln(tts_{AB\ k}) + n_{\sigma}.\left(\sigma_{stt_{AB}\ k}\right)\right]} \tag{8}$$

$$Stt_{AB\ k} = \{t_{Bi} - \ t_{Ai} | t_k - t_{k-1} < t_{Bi} \le t_k \ \ and \ \ tt_{AB\min k} \le t_{Bi} - \ t_{Ai} \le \ tt_{AB\max k}\} \tag{9}$$

$$tt_{AB\ k} = \frac{\sum_{i=1}^{n_{vk}}(t_{Bi} - \ t_{Ai})}{n_{vk}} \tag{10}$$

$$\sigma^2_{tt_{AB}\ k} = \begin{cases} 0 & , \quad n_{vk} = 0 \\ \dfrac{[\ln(t_{Bi} - \ t_{Ai})_k \ - \ln(tts_{AB\ k})]^2}{n_{vk}} & , \quad n_{vk} = 1 \\ \dfrac{\sum_{i=1}^{n_{vk}}[\ln(t_{Bi} - \ t_{Ai})_k \ - \ln(tts_{AB\ k})]^2}{n_{vk} - 1} & , \quad n_{vk} \ge 2 \end{cases} \tag{11}$$

where, $k$ is the time of the end of the current time window. $k - 1$ is the time of the end of the previous time window; $tts_{AB\ k}$ is the predicted smoothed mean of travel times from A to B for time k; $\sigma^2_{stt_{AB}\ k}$ is the predicted smoothed variance of travel times from A to B for time k; $tt_{AB\ k}$ is the mean of travel times from A to B at time k; $\sigma^2_{tt_{AB}\ k}$ is the variance of travel times from A to B at time k; $Stt_{AB\ k}$ is a set of valid travel times from point A to point B at time k; $tt_{AB\min k}$ is the minimum of valid travel time from A to B at time k;

$tt_{AB\max k}$ is the maximum of valid travel time from A to B at time k; $n_{vk}$ is the number of valid travel times from A to B at time k; $\alpha$ is an exponential smoothing factor; $\beta$ is a sensitivity parameter, and $n_{\sigma}$is the number of standard deviations that defines the size of the validity window.

Eq. (4) and Eq. (5) describe how to derive $tts_{AB\ k}$ and $\sigma^2_{stt_{AB}\ k}$ based on the values of the previous sampling interval and $\alpha$. Eq (6) is used to calculate $\alpha$ based on $\beta$ and $n_{vk}$. Eq. (7) and Eq. (8) are used to define the lower and upper limits for the validity range based on the results of Eq. (4), Eq. (5) and $n_{\sigma}$. Eq. (9) uses the results of Eq. (7) and Eq. (8) to determine the valid travel time observation. Eq. (10) and Eq. (11) exhibit how to calculate $tt_{AB\ k}$ and $\sigma^2_{tt_{AB}\ k}$ based on the valid travel time observation, using them in the calculations of the next time window. $\beta$, $n_{\sigma}$ and the size of the time window are user-defined parameters. The recommended values for these parameters are, 0.2 – 0.5 for $\beta$, 2 or 3 for $n_{\sigma}$, 2 min for the size of the time window (Dion & Rakha, 2006).

#### 2.2.4.2 Version 2

The differences between version 2 and version 1 are calculating $\alpha$ and $\sigma^2_{tt_{AB}\ k}$ in Eq. (14) and Eq. (19), whereas other equations are the same.

$$tts_{AB\ k} = \begin{cases} e^{[(\alpha).\ln(tt_{AB\ k-1})+(1-\alpha).\ln(tt_{AB\ k-1})]} & , \quad n_{vk-1} > 0 \\ tt_{AB\ k-1} & , \quad n_{vk-1} = 0 \end{cases} \tag{12}$$

$$\sigma^2_{stt_{AB}\ k} = \begin{cases} \alpha\,.\left(\sigma^2_{tt_{AB}\ k-1}\right) + (1-\alpha).\ \left(\sigma^2_{stt_{AB}\ k-1}\right), & n_{vk-1} > 1 \\ \sigma^2_{stt_{AB}\ k-1} & , \quad n_{vk-1} = \{0,1\} \end{cases} \tag{13}$$

$$\alpha = \begin{cases} 1-\ (1-\ \beta)^{n_{vk}} & , \quad n_a < 3\ and\ \ n_b < 3 \\ \max(0.5, 1-\ (1-\ \beta)^{n_{vk}}) & , \quad n_a \geq 3\ or \quad n_b \geq 3 \end{cases} \tag{14}$$

$$tt_{AB\min k} =\ e^{\left[\ln(tts_{AB\ k})-n_{\sigma}.\left(\sigma_{stt_{AB}\ k}\right)\right]} \tag{15}$$

$$tt_{AB\max k} =\ e^{\left[\ln(tts_{AB\ k})+n_{\sigma}.\left(\sigma_{stt_{AB}\ k}\right)\right]} \tag{16}$$

$$Stt_{AB\ k} = \{t_{Bi} - t_{Ai} | t_k - t_{k-1} < t_{Bi} \le t_k \ and \ tt_{AB\min k} \le t_{Bi} - t_{Ai} \le tt_{AB\max k}\} \tag{17}$$

$$tt_{AB\ k} = \frac{\sum_{i=1}^{n_{vk}} (t_{Bi} - t_{Ai})}{n_{vk}} \tag{18}$$

$$\sigma^2_{tt_{AB\ k}} = \begin{cases} 0 & , \quad n_{vk} = 0 \ and \ n_a < 3 \ and \ n_b < 3 \\ \dfrac{[\ln(t_{Bi} - t_{Ai})_k - \ln(tts_{AB\ k})]^2}{n_{vk}} & , \quad n_{vk} = 1 \ and \ n_a < 3 \ and \ n_b < 3 \\ \dfrac{\sum_{i=1}^{n_{vk}} [\ln(t_{Bi} - t_{Ai})_k - \ln(tts_{AB\ k})]^2}{n_{vk} - 1} & , \quad n_{vk} \ge 2 \ and \ n_a < 3 \ and \ n_b < 3 \\ 0.01 \, . \, (tt_{AB\ k}) & , \quad n_a \ge 3 \ or \quad n_b \ge 3 \end{cases} \tag{19}$$

These changes have been proposed to track sudden variations in traffic conditions. In particular, the amendments enable the algorithm to consider as valid the third of three successive observations outside the validity limits, as long as all three observations are either above or below the validity limits. $n_a$ and $n_b$ have been introduced in this version, $n_a$ is the number of consecutive observations above the validity limits, and $n_b$ is the number of consecutive observations below the validity limits.

However, this algorithm is more complex than other algorithms, given that it has many assumptions that made it complicated. This complexity makes understanding and application more difficult. In addition, the multiplicity of parameters that need to be calibrated makes it impractical. However, it can be used in real-time applications.

### 2.2.5 Jang Algorithm

Jang (2016) proposed a new outlier filtering algorithm consisting of two parts based on the number of observations in the time window. In this algorithm, if the number is less than 3, which is insufficient to generate an effective measure of location, it utilises a validity range based on the previous time window. However, if the number of observations is more than or equal to 3, the current time window's observations are used to determine the validity range. In the second part of this algorithm, given that the median is the best measure of central tendency for skewed variables, the median is adopted to measure location. The minimum sample size for generating an effective median is three observations since the median can detect the discordant value if two travel time observations are true and one is discordant, though the mean may not. The median absolute deviation is utilised to define the validity range. If the number of valid observations is less than the outliers or all the travel times are outliers, the comparison between the median of the current time window and the mean of the previous time window is adopted to address and overcome this issue.

$IF\ n < 3\ then$

$$S_{1AB}(t) = \{k|t - T_w < t_{B,k} \leq t\ \} \cap \left\{m| \left|\frac{t_{B,m} - t_{A,m} - T_{AB}(t - T_w)}{T_{AB}(t - T_w)}\right| \leq \alpha\right\} \tag{20}$$

$$T_{AB}(t) = \frac{\sum_i (t_{B,i} - t_{A,i})}{N(S_{1AB}(t))} \quad , \quad \text{where } \ i \in S_{1AB}(t) \tag{21}$$

$IF\ n \geq 3\ then$

$$M_t = median(t_{B,m} - t_{A,m}) \tag{22}$$

$$M_{AD} = median|(t_{B,m} - t_{A,m}) - M_t| \tag{23}$$

$$IF \left|\frac{M_t - T_{AB}(t - T_w)}{T_{AB}(t - T_w)}\right| \geq \gamma \text{ , then apply (20) and (21)} \tag{24}$$

$$S_{2AB}(t) = \{k|t - T_w < t_{B,k} \leq t\ \} \cap \{m|(M_t - \beta M_{AD}) \leq (t_{B,m} - t_{A,m}) \leq (M_t + \beta M_{AD})\} \tag{25}$$

$$T_{AB}(t) = \frac{\sum_i (t_{B,i} - t_{A,i})}{N(S_{2AB}(t))} , \quad \text{where } \ i \in S_{2AB}(t) \tag{26}$$

where, $T_w$ is a time window or collection interval; $n$ is the number of travel time observations in $T_w$; $S_{1AB}(t)$ and $S_{2AB}(t)$ are the sets of valid travel times from A to B at time $t$; $N\left(S_{1AB}(t)\right)$ and $N\left(S_{2AB}(t)\right)$ are the numbers of valid travel times from A to B at time $t$; $T_{AB}(t)$ is the average travel time of valid observations from A to B at time $t$; $T_{AB}(t - T_w)$ is the average travel time of valid observations from A to B at time $(t - T_w)$; $t_{A,i}$ is the detection time of vehicle $i$ that has a valid travel time at point A; $t_{A,m}$ is the detection time of vehicle $m$ at point A; $t_{B,i}$ is the detection time of vehicle $i$ that has a valid travel time at point B; $t_{B,m}$ is the detection time of vehicle $m$ at point B. $M_t$ is the median of the travel time observations in the current time window. $M_{AD}$ is the median absolute deviation of travel time observations in the current time window. $\alpha$, $\beta$, and $\gamma$ are parameters.

Eq. (20) and Eq. (21) represent the first part of this algorithm. They are applied when the number of observations in the current time window is less than 3. Eq. (20) defines the valid travel time based on the comparison of each travel time observation in the time window with the average travel time of the previous time window $T_{AB}(t - T_w)$. If the absolute difference divided by $T_{AB}(t - T_w)$ exceeds $\alpha$, the observation will be considered as an outlier; otherwise, it will be valid. Eq. (21) is utilised to calculate $T_{AB}(t)$ and to use it as the average travel time of the previous time window for the next time window calculations. Notably, this part is similar to the TransGuide algorithm. Eq. (22) to Eq. (26) are the second part of this algorithm. Eq. (22) and Eq. (23) are employed to calculate $M_t$ and $M_{AD}$ respectively; Eq. (24) is used to deal with the situation that the number of valid observations is less than the outliers or all the travel times are outliers; Eq. (25) determines the valid travel time by defining a validity range based on $M_t$, $M_{AD}$ $and$ $\beta$. Eq. (26) is used to calculate $T_{AB}(t)$. The recommended values for the parameters are 5 min for $T_w$, 0.35 for $\alpha$, 3 for $\beta$, and 0.3 for $\gamma$ (Jang, 2016).

It is noticeable that this algorithm is different from the other filtration algorithms presented in this study. The other algorithms depend on determining the validity range based on either the previous time window or the current time window, while the Jang algorithm consists of both. However, it has a medium level of complexity and can be employed in real-time applications.

## 2.3 Travel Time Reliability

### 2.3.1 Travel Time Reliability Definitions

The literature has many definitions of TTR introduced in early studies. The Federal Highway Administration (FHWA) defines TTR as "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day" (Texas Transportation Institute & Cambridge Systematics Inc., 2005). This definition is commonly found in the literature, where many researchers consider it as a formal definition of TTR (P. Chen, Sun, & Qi, 2017; Z. Chen & Fan, 2019; Lyman & Bertini, 2008).

Traffic researchers addressed the concept of TTR from many aspects. Tsolakis et al. (2011) classified the definitions of TTR into five classifications:

1) Reaching the destination in an acceptable time

- "Percent of trips no longer than expected travel time plus a certain acceptable additional time" (Florida Department of Transportation, 2000).
- "Percent of the same trips by time of day and trip purpose, within a specific range of travel time" (Elefteriadou & Cui, 2007).

2) Consistency and dependability in travel time

- "the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day" (Texas Transportation Institute & Cambridge Systematics Inc., 2005).

3) Variability in journey time

- “Measure of the variability of travel time” (Cambridge Systematics Inc., 1998).
- “The variability of travel times that occur on a facility or a trip over the course of time” (Vandervalk, Louch, Guerre, & Margiotta, 2014).

4) Travelling with non-recurrent congestion

- “A measure of the amount of congestion that users of the transportation system experience at a given time” (Lyman & Bertini, 2008; Tsolakis et al., 2011).

5) Transport system ability and performance

- “The ability of the transport system to provide the expected level of service/quality, upon which users have organized their activities” (Tsolakis et al., 2011).

### 2.3.2 Travel Time Reliability Measures

Travel time reliability (TTR) measures are the tools used to estimate the degree of unreliability or variability in travel time for particular sections of the road network. They can be used to understand how much travel time variability or unreliability road users experience. The different definitions mentioned above of TTR lead to various TTR measures (Tsolakis et al., 2011). As such, many researchers placed much effort in developing TTR measures to advantage road users and transportation professionals (Florida Department of Transportation, 2000; Kaparias, Bell, & Belzner, 2008; Lomax & Margiotta, 2003; Lyman & Bertini, 2008; Texas Transportation Institute & Cambridge Systematics Inc., 2005).

Lomax and Margiotta (2003) classified TTR measures into three types: 1) statistical range measures; 2) buffer time measures; 3) tardy trip indicators. However, there are differences noted between these types on the calculation part but are most evident on the communication side of the issue.

#### 2.3.2.1 Statistical Range Measures

These measures use standard deviation statistics to estimate the variety of transport conditions that passengers might encounter. Usually, they consider the form of an average value plus or minus a value that covers the expectations for 68% to 95% of the trips. Here, 68% and 95% represent 1 and 2 standard deviations on each side of the mean. Typically, this type of measure is characterised by the information given in a relatively unprocessed format and based on concepts familiar to statisticians. Some of the statistical range measures are quite challenging to explain to non-statisticians and individual travellers to make travel decisions (Lomax & Margiotta, 2003).

Three TTR measures were reported based on the standard deviation statistic, that includes the following (Lomax & Margiotta, 2003):

1- Travel time window: is a combination between the standard deviation and the average using a "plus or minus" expression. One standard deviation will represent 68% of the trips.

$$Travel\ Time\ Window = Average\ Travel\ Time \pm Standard\ Deviation \qquad (27)$$

2- Percent variation: is also known as the coefficient of variation, a combination between the standard deviation and the average in a ratio.

$$Percent\ variation = \frac{Standard\ Deviation}{Average\ Travel\ Time} \times 100\% \qquad (28)$$

3- Variability index: is a ratio of the difference in the upper and lower 95% confidence intervals between the peak period and the off-peak period. Usually, the variability index value is greater than 1 because the interval differences in the peak periods are larger than in the off-peak.

$$Variability\ index = \frac{(upper\ 95\%\ confidence\ interval\ value - lower\ 95\%\ confidence\ interval\ value)\ for\ Peak\ period}{(upper\ 95\%\ confidence\ interval\ value - lower\ 95\%\ confidence\ interval\ value) for\ off-Peak\ period} \quad (29)$$

#### 2.3.2.2 Buffer Time Measures

The concept of buffer time may correlate to the way travellers think. Conceptually, travellers endeavour to answer questions such as How far is the destination?; When do I have to arrive?; How bad is the traffic?; How much time do I need to allow for this?; When am I going to leave?. In the time allocation stage, an estimate is made on how much additional time should be required for any uncertainty in travel conditions since incidents, weather, work zones, special event, holiday, and other disruptions affect the decisions of travellers. Buffer time measures indicate the impact of irregular conditions in the amount of extra time that travellers must consider to achieve their destination. (Lomax & Margiotta, 2003). Three buffer time measures are presented below:

1- Buffer time index (BTI): illustrates the extra buffer time (or time cushion) that most travellers add to their average travel time to ensure on-time arrival when scheduling trips. To account for any unforeseen delay, this extra period is added. This index is expressed as a percentage in that as reliability becomes worse, its value increases. For example, a buffer index of 50 per cent means that, for a 10-minute average travel time, a traveller should budget an additional 5 minutes (10 minutes × 50 per cent = 5 minutes) to ensure on-time arrival in most cases. The 5 extra minutes is called the buffer time. BTI is calculated as the difference between the 95th percentile travel time and average travel time, divided by the average travel time. Here, 95th percentile travel time is used to represent a near-worst case scenario for travel time. A simple analogy is that only one weekday per month would account for being late for a commuter or driver who uses a 95th percentile reliability measure. (Texas Transportation Institute & Cambridge Systematics

Inc., 2005). Notably, some references use BTI as an index, not as a percentage (Higatani et al., 2009).

$$BTI = \frac{95th\ percentile\ travel\ time - Average\ travel\ time}{Average\ travel\ time} \times 100\% \quad (30)$$

Or

$$BTI = \frac{95th\ percentile\ travel\ time - Average\ travel\ time}{Average\ travel\ time} \quad (31)$$

2- Planning time index (PTI): reflects the total travel time that should be planned when an adequate buffer time is included. The difference between PTI and BTI is that PTI includes typical delay and unexpected delay, while BTI includes simply unexpected delay. PTI is near-worst case travel time to free-flow travel time. For example, a PTI of 1.80 means that, for a 10-minute trip in free-flow traffic, the total time that should be planned for the trip is 18 minutes (10 minutes × 1.80 = 18 minutes). PTI is useful as it can be directly compared to the TTI (a measure of average congestion) on similar numeric scales. PTI is calculated as the 95th percentile travel time divided by the free-flow travel time (Texas Transportation Institute & Cambridge Systematics Inc., 2005).

$$PTI = \frac{95th\ percentile\ travel\ time}{Free\ flow\ travel\ time} \quad (32)$$

3- Travel time index (TTI): is the ratio of average travel time observed during peak periods compared to free-flow travel time. This measure implies how much longer travel time is during congested conditions relative to free-flow traffic conditions (Culotta, Fang, Habtemichael, & Pape, 2019). Some references consider TTI a measure of congestion and do not consider TTI a measure of reliability (Texas Transportation Institute & Cambridge Systematics Inc., 2005). Notably, TTI can be calculated for any time of the day and not only for peak periods, as shown in Figure 2.1.

$$TTI = \frac{Average\ travel\ time}{Free\ flow\ travel\ time} \tag{33}$$

Figure 2.1 describes the relationship between TTI, BTI, and PTI. TTI represents the average extra time required, as compared to the times of free-flow traffic conditions. BTI signifies the extra time needed above the average travel time, whereas PTI represents the total travel time necessary (Texas Transportation Institute & Cambridge Systematics Inc., 2005).

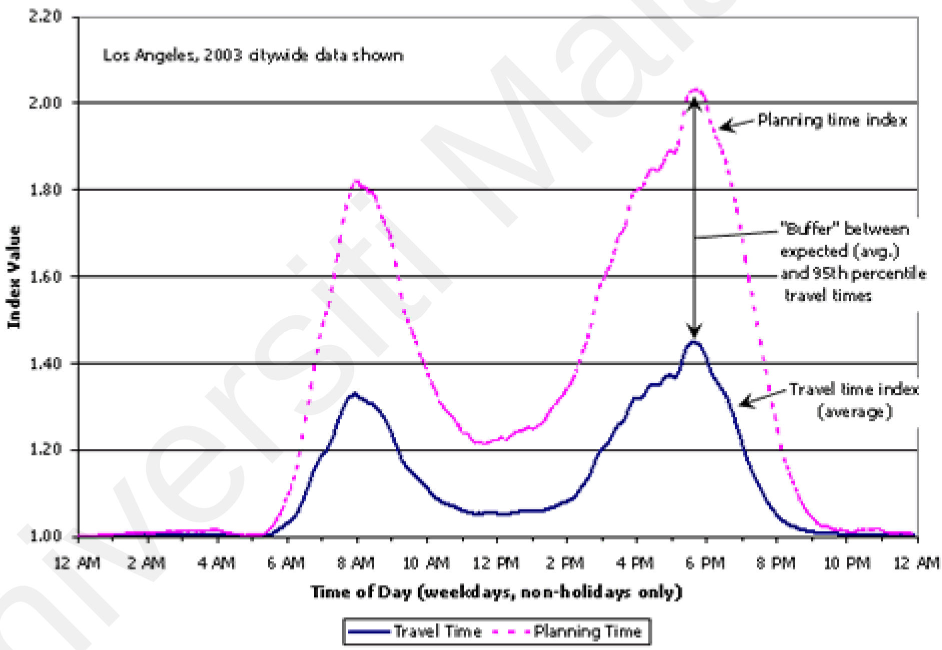



**Figure 2.1: The relationship between the travel time index, the buffer time index, and the planning time index (Texas Transportation Institute & Cambridge Systematics Inc., 2005).**

#### 2.3.2.3 Tardy Trip Indicators

These measures assess how often travellers will unacceptably be late. A threshold to identify an acceptable late arrival time is used in these measures, where the average travel time is not adopted. The time can be increased in minutes above the average, a percentage of the trip time, or some absolute value in minutes. The examples of these measures are reflected below: (Lomax & Margiotta, 2003)

1- Florida Reliability Method: is used to estimate the limit of the acceptable extra travel time range; this measure uses a percentage of the average travel time in the peak and is calculated as follow:

$$\begin{aligned} &\textit{Florida Reliability statistic} \\ &= 100 - (\textit{percent of trips with travel times greater than expected}) \end{aligned} \quad (34)$$

2- On-time arrival: has a concept similar to the Florida method. To indicate the percentage of trip travel times that can be termed reliable, it uses an acceptable "lateness threshold" of some percentage. It can be calculated as follows:

$$\begin{matrix} \textit{On}-\textit{time} \\ \textit{arrival} \end{matrix} = 100 - \begin{pmatrix} \textit{percent of travel rates greater than } 110\% \\ \textit{of the average travel rate} \end{pmatrix} \quad (35)$$

3- Misery Index: describes the negative aspect of trip reliability and can be examined by the average number of minutes that the worst trips exceed the average. It can be calculated as follows:

$$\textit{Misery Index} = \frac{\begin{matrix} \textit{Average of the travel rates for} \\ \textit{the longest } 20\% \textit{ of the trips} \end{matrix} - \begin{matrix} \textit{Average travel rates} \\ \textit{for all trips} \end{matrix}}{\textit{Average travel rate}} \quad (36)$$

### 2.3.3 Factors Affecting Reliability

TTR is affected by many factors (Culotta et al., 2019; Kittelson, Associates, & Program, 2013; Margiotta et al., 2013):

1- Bottlenecks: referring to road sections showing decreased traffic capacity compared to the capacity of upstream road sections. Typical bottlenecks include changes in alignment (e.g., horizontal curves), lane drops, changes in physical road characteristics (e.g., tunnels), presence of weave and merge sections, geometric changes, interchanges, and access points to residential or commercial developments (Culotta et al., 2019).

2- Traffic-control devices: are used to inform, guide, and control traffic flow (both pedestrians and vehicles). Any problems in traffic-control devices lead to disruption in traffic flow, which lead to delays and unreliable travel time. Such problems include the use of inappropriate devices, improper device placement, wrong colour, shape, and size, poor timing devices, and device failure (Culotta et al., 2019).

3- Weather: environmental conditions trigger changes in driver behaviour. Severe weather conditions like fog, snow, and heavy rain, cause drivers to drive more cautiously, slowing down and leaving more space between vehicles to maintain safety, thus reducing roadway throughput (Kittelson et al., 2013; Margiotta et al., 2013).

4- Incidents: traffic incidents are events that disrupt traffic flow, usually due to physical impedance in travel lanes. Vehicular crashes and breakdowns, and debris on the roadway reduce the capacity either by blocking lanes physically or by creating visual distractions causing motorists to slow down, resulting in reduced roadway throughput (Kittelson et al., 2013; Margiotta et al., 2013).

5- Work zones: are areas where roadway construction activities lead to tentative physical changes to the highway environment. Lane closures, lane width reductions, type and duration of work, work intensity, pavement condition, and work zone length are factors that can cause delays in work zones. Short-term work zones tend to have more effect on disrupting traffic than long-term work zones. In long-term work zones, road users tend to become more familiar with the new traffic pattern (Culotta et al., 2019).

6- Travel demand fluctuations: are the daily and seasonal changes in the demand that contribute to increased travel. For example, seasonal changes might be due to holidays, part-year residents moving out or tourists coming in, or school-related traffic during the school year at the beginning and end of the school day (Culotta et al., 2019) .

7- Special events: are a unique case of travel demand fluctuations where traffic flow in the proximity of the event will be radically different from normal patterns. Sometimes, Special events lead to surges in traffic demand that overwhelm the system (Kittelson et al., 2013).

Most Likely, factors that cause fluctuations in either demand or supply have inter-relationships, where they are not independent. For example, extreme weather can be deemed a factor in reducing the road network capacity, also affecting travel demand. Moreover, accidents may increase given this situation (Hojati, 2014).

### 2.3.4 Recurrent and Non-Recurrent Congestion

Congestion can be divided into recurrent and non-recurrent. Recurrent congestion includes delays that are foreseeable in frequency and extent (e.g., peak-hour traffic), whereas in contrast, non-recurrent congestion is caused by unanticipated delays resulting from a temporary decline in road capacity (e.g. blocked lane(s) due to accidents) or sudden increases in demand (e.g. special events). Lack of TTR is related to delays caused by congestion, especially delays caused by non-recurrent events. Consequently, improvements in congestion can also increase the reliability of travel times (Culotta et al., 2019).

In a study conducted in the US to estimate congestion by source, recurrent congestion represented 40%, whereas non-recurrent congestion accounted for 60%. In this study, non-recurrent congestion is higher than recurrent congestion. Estimation of congestion by source for individual highways is instrumental in developing mitigation strategies. Figure 2.2 displays the sources of congestion for the study mentioned above. Bottlenecks, that is, recurrent congestion, represents 40%, while traffic incidents represent the highest source of non-recurrent congestion (Cambridge Systematics Inc., 2005).

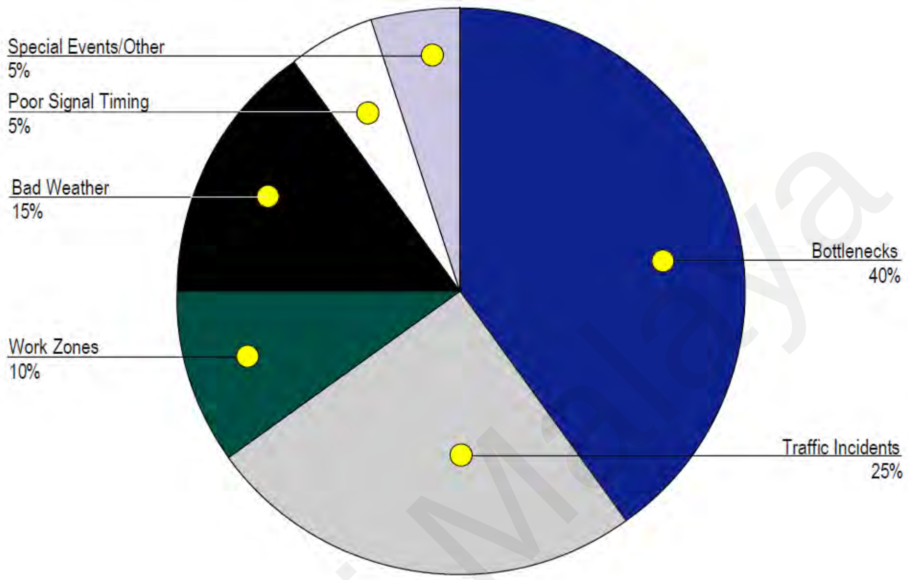



**Figure 2.2: The sources of congestion (Cambridge Systematics Inc., 2005).**

Four components describe the concerns of travellers concerning congestion: duration, extent, intensity, and variation. Duration is the length of time the transportation system is under congestion, measured by congested hours. Extent is the number of vehicles or people affected by congestion. Indeed, it can be measured by person-miles of travel and by per cent, route-miles, or lane-miles of the transportation system affected by congestion. The third component, namely intensity, depicts the severity of the congestion, measured by TTI, PTI and BTI. The variation in the first three components is the fourth component of congestion which can be considered an indicator of the system's reliability. It can be measured by PTI and BTI (Cambridge Systematics Inc., 2008).

### 2.3.5 Travel Time Variability

Travel Time Variability (TTV) relates to the uncertainty in trip journey times. TTV can be considered from three different perspectives: the day-to-day variability, the period-to-period variability, and the vehicle-to-vehicle variability (Noland & Polak, 2002).

The day-to-day (or inter-day) variability shows the variability between similar trips during the same time period on various days. Day-to-day variability is illustrated in Figure 2.3 below. This figure shows the travel time of the same road segment on different working days in a probability distribution function. Travel demand fluctuations, weather conditions, incidents, and driving behaviour can lead to day-to-day variability. (Büchel & Corman, 2020).

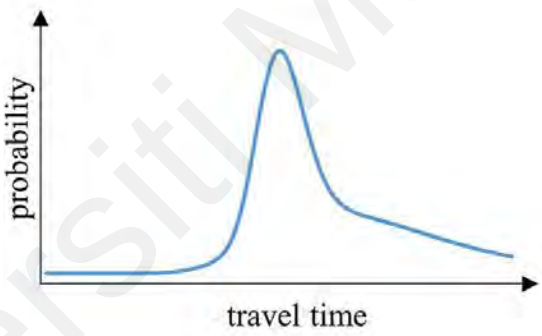



**Figure 2.3: Typical representation explaining of day-to-day variability (Büchel & Corman, 2020).**

Period-to-period variability (also known as inter-period) is the second type of TTV. It explains the variability between vehicles making similar trips at different times on the same day. Figure 2.4 shows a representation of period-to-period variability. In this case, travel times are longer during peak periods than during off-peak periods. Short-term changes in travel demand, weather conditions, and incidents can result in period-to-period variability (Büchel & Corman, 2020).

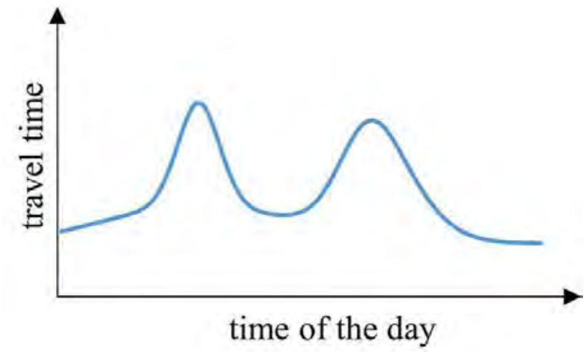



**Figure 2.4: Typical representation explaining of period-to-period variability (Büchel & Corman, 2020).**

Lastly, vehicle-to-vehicle variability (also known as inter-vehicle) is the third type of variability. It illustrates the variability between travel times that different vehicles travel at similar times over the same road segment. Figure 2.5 shows this type of variability. For instance, at 14:00, travel times are around 100 to 150 seconds. This range represents vehicle-to-vehicle variability. Different delay times at traffic signals, types of vehicles, conflict with pedestrians, and differences in driving behaviour can lead to vehicle-to-vehicle variability (Büchel & Corman, 2020).

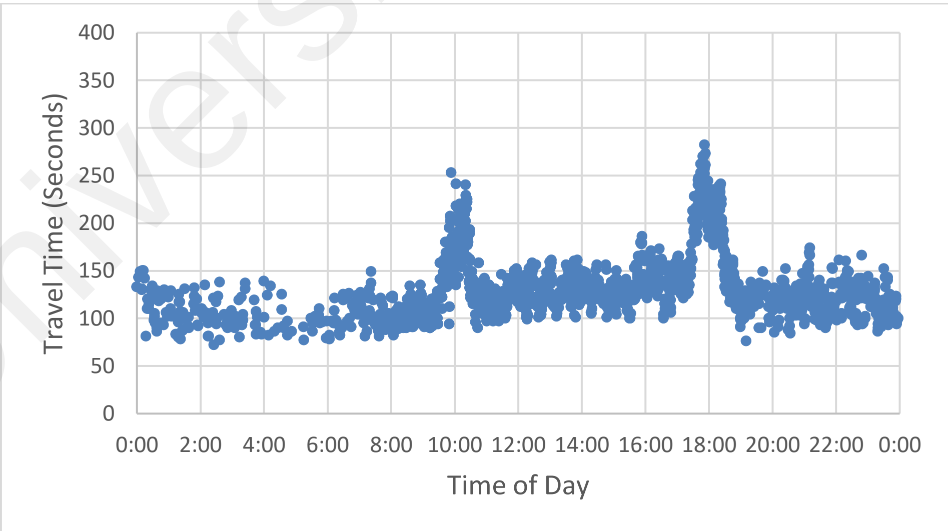



**Figure 2.5: Typical representation explaining of vehicle-to-vehicle variability.**

The terms TTR and TTV are frequently used interchangeably (Fosgerau, Hjorth, Brems, & Fukuda, 2008; Tsolakis et al., 2011). As revealed in the TTR definition section of this study, some researchers consider TTR as a measure of TTV (Cambridge

Systematics Inc., 1998; Fosgerau et al., 2008). However, TTV indicates the variance of travel time over time, while TTR includes variance and predictability (Fosgerau et al., 2008; Hojati, 2014). Travellers and operators prefer to use TTR, where it is more easily conceptualised and understood (Lomax & Margiotta, 2003).

### 2.3.6 Importance of Travel Time Reliability

TTR is important to transportation system users, whether they are vehicle drivers, freight shippers, transit riders, or air travellers. TTR allows personal and business travellers to better utilise their own time. Shippers and freight carriers require predictable travel times to remain competitive. Given the importance of TTR, decision-makers and transportation professionals should consider using it as a key performance measure (Texas Transportation Institute & Cambridge Systematics Inc., 2005).

Many studies have shown that, in general, travellers would rather use the routes with higher mean travel times and smaller TTV in place of routes having lower mean travel times and larger TTV (Lyman & Bertini, 2008; Van Lint & van Zuylen, 2005). Therefore, providing confidence intervals around the average travel time is beneficial to minimise the level of anxiety and stress induced by uncertainty. It will also assist road users in making better decisions regarding departure time and route choice (Hojati, 2014).

TTR is better than simple averages in quantifying the benefits of traffic management and operation activities. For instance, consider a typical before-and-after study that attempts to quantify the benefits of an incident management program. The improvement in average travel time may seem to be modest, as shown in Figure 2.6a. While Figure 2.6b, on the other hand, indicates that reliability measures show a more significant improvement. They demonstrate the effect of improving the worst few days of an unexpected delay (Texas Transportation Institute & Cambridge Systematics Inc., 2005).

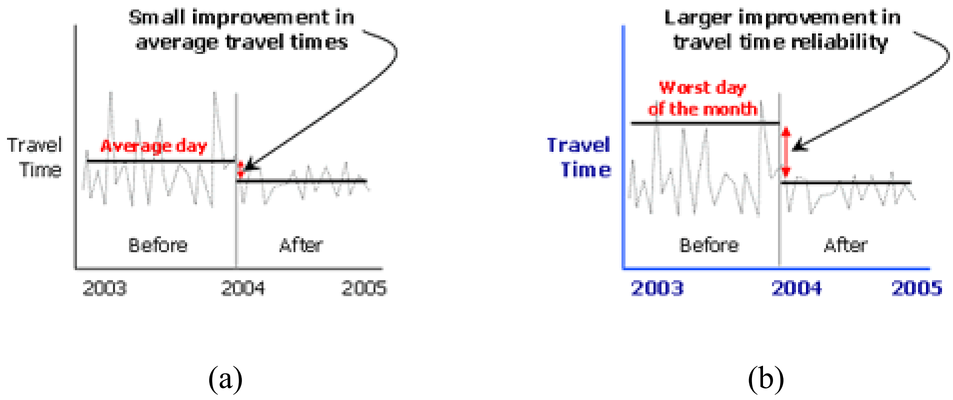



**Figure 2.6: Benefits of an incident management program: (a) improvement in average travel time, (b) improvement in travel time reliability (Texas Transportation Institute & Cambridge Systematics Inc., 2005).**

# CHAPTER 3: METHODOLOGY

This chapter describes the methodology adopted in this research. The methodology employed to achieve the objectives of this research is described in the following sections: research methodology flowchart, study area, data collection, data description, and data analysis.

## 3.1 Research Methodology Flowchart

In order to achieve the research objectives, the appropriate methods and analyses were selected after reviewing the literature relevant to the subject under study. These analyses were carefully organised to ensure consistency between them. So as to facilitate understanding the outlines of the procedures and analyses, a research flowchart was developed, as illustrated in Figure 3.1. Here in this figure, four phases are adopted to achieve the research objectives. The literature review is the first phase to understand the research problem and to identify the research gap. The second phase is collecting the data, where all required data is gathered before commencing the analysis. After that, filtration of travel time data constitutes the third phase consisting of two stages. The first stage is the validation of the previous travel time filtration algorithms, applying these algorithms to travel time data for one day. The second stage is the examination of the sensitivity of the algorithm parameters for different days. The two stages set out to evaluate the previously established filtration algorithms and identify the most appropriate algorithm and parameters able to filter lane-splitting observations and outliers. The fourth phase is the empirical analysis of TTR measures based on filtered data. The results and discussions constitute the fifth phase, which reveals the answers to the research questions. The last phase includes the conclusions and recommendations based on the results of the analysis.

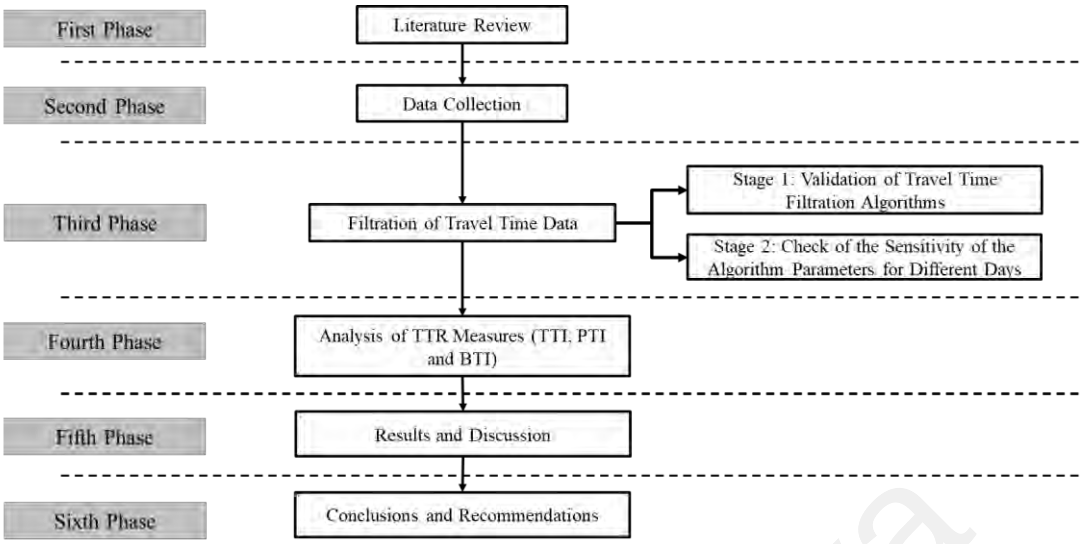



**Figure 3.1: Research Methodology Flowchart**

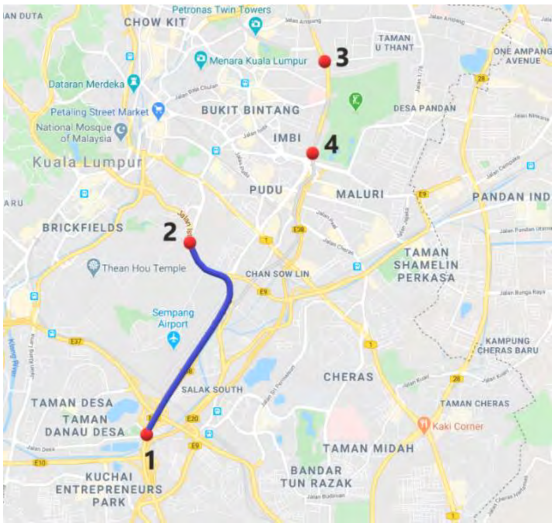
## 3.2 Study Area

An urban roads network at the heart of KL was selected for carrying out the data analysis. The network is located near Kuala Lumpur city centre (KLCC), where numerous skyscrapers such as Petronas Twin Towers and many shopping centres, hotels, and businesses offices are located. Four units of MAC address sensors were installed at this network to measure travel time. The locations of the sensors were as follow:

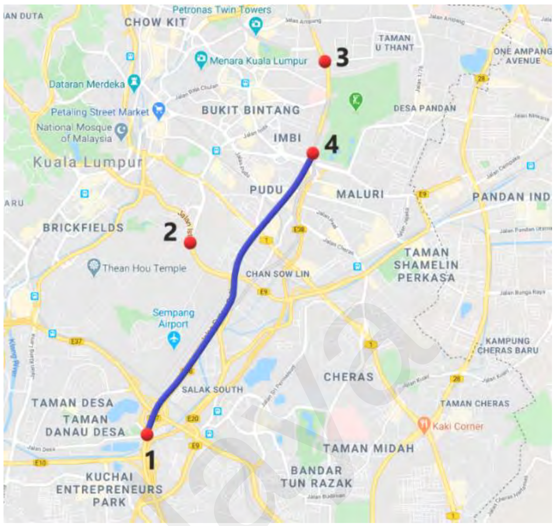
- Sensor 1 at KL-Seremban Highway.
- Sensor 2 at Jalan Istana.
- Sensor 3 at Jalan Tun Razak (near the U.S. Embassy).
- Sensor 4 at Jalan Yew.

These sensors were utilised to collect travel time data from three routes. Figure 3.2 presents the locations of the sensors and the routes. These three routes were selected because they are at the heart of KL. Also, they have different lengths and variation in the number of collected observations.

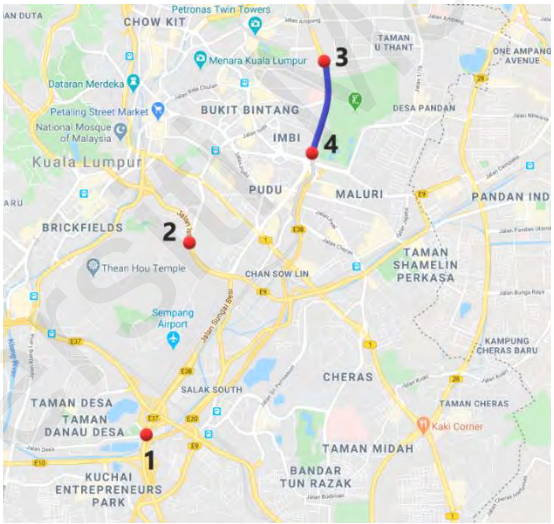
In order to facilitate the discussion in this section of the research, the routes were named, A for the route that reaches between sensor 1 and sensor 2, B for the route that reaches between sensor 1 and sensor 4, and C for the route that reaches between sensor 4 and sensor 3. The routes information is presented in Table 3.1.




(a)




(b)




(c)

**Figure 3.2: Maps of the routes: (a) route A, (b) route B, and (c) route C (Google, n.d.)**

As shown in the above figure, route A consists of two segments; one is part of the KL-Seremban Expressway, and the other is part of Jalan Istana. Route B consists of three segments: a portion of the KL-Seremban Expressway, Jalan Sungai Besi, and Jalan Yew, while the last route consists of one segment, which forms a part of Jalan Tun Razak.

**Table 3.1: Routes information**

| Route | Length (m) | Segments | Connected sensors | No. of travel time observations during May 2018 |
|---|---|---|---|---|
| Route A | 3880 | 2 | 1 and 2 | 89225 |
| Route B | 5690 | 3 | 1 and 4 | 37320 |
| Route C | 1410 | 1 | 4 and 3 | 117114 |

## 3.3 Data Collection

Integrated Transportation Solutions Sdn. Bhd. (ITSSB) collected the traffic data used in this study under the Pilot Project of Advanced Traffic Information System (ATIS) in 2018. The project was conducted in collaboration with the Integrated Transport Information System (ITIS) DBKL. As shown in Figure 3.3, ITSSB developed a system that uses the MAC address to collect traffic data. The system anonymously detects, transmits, records, matches and analyses MAC addresses that smartphones periodically transmit via Wi-Fi to measure TTR. Millions of MAC address data were collected during the pilot project period. The data used in this study was collected during May 2018, consisting of 18 weekdays, eight weekends, three general election days and two public holidays. The analysis was carried out inclusive of these days.


**Figure 3.3: MAC address sensors**

The percentage of mobile devices and vehicles equipped with Wi-Fi technology is vast. The MAC address can be captured using sensors at strategic key points on road networks and transmitted to the back-end server for precise travel time measurement. The sensor at a specified location first logs the time stamp of the mobile device that enters its zone. When another sensor logs the device again at a different location, the difference in time stamps is utilised to estimate the vehicle's travel time between the two locations equipped with that particular mobile device (Abdullah et al., 2017).

## 3.4 Data Description

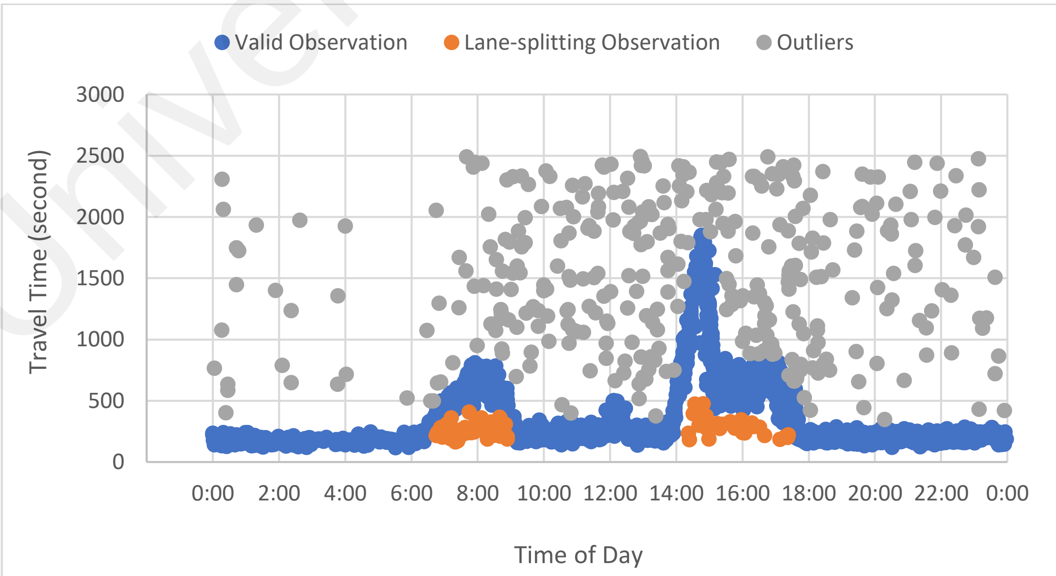
Data description using MAC Address data for lane-splitting, based on the actual situation, incorporating a high percentage of motorcycles have never been reported. This section presents actual MAC Address datasets in Malaysia to show the impact of lane-splitting on travel time patterns. The presented datasets had raw matched MAC address data before any filtration algorithm was applied.

The travel time datasets for the study area were grouped into three categories of observations: valid, outlier, and lane-splitting. Lane-splitting observations represent the motorcycle observations at peak periods. The method adopted to observe lane-splitting was manual classification based on the experience of the researcher as Moghaddam and Hellinga (2014) did in their research. After revising many travel time patterns from different studies the researcher understood how to manually classify the dataset. Valid observations pattern is clear, where usually there are morning peak and evening peak. In addition, lane-splitting observations are clear because they are existed only during peak hours.

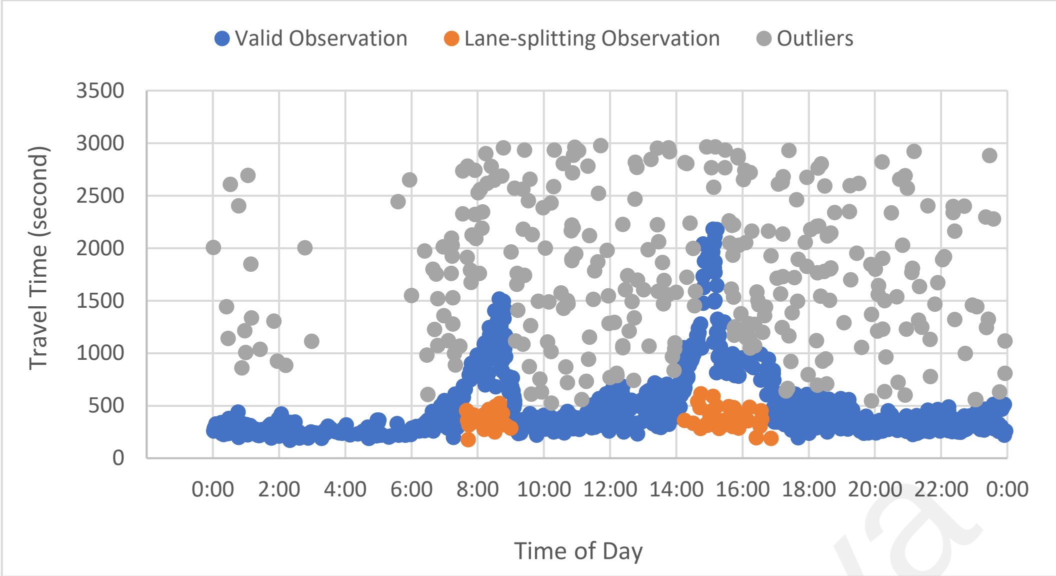
To classify travel time datasets into valid, lane-splitting, and outlier, first, the travel time dataset for a whole day for one route was presented using a graph. Second, using the graph, for each hour, the classification was done. Third, for the second day for the same route the first two steps were repeated. Fourth, after the whole route were classified, the first three steps were repeated for the next route.

Figure 3.4 presents the travel time datasets of routes A, B, and C on 28th May 2018 for 24 hours. In this figure, the blue points are valid observations, whereas the grey points are outliers, and the orange points are lane-splitting observations. Figure 3.4a displays the travel times for route A. The differences between the three categories are evident for this route. Figure 3.4b exhibits the datasets of route B, where the number of valid observations is relatively small compared to outliers and lane-splitting data since this route is longer with numerous intersections. Figure 3.4c demonstrates the travel time observations for route C.

To notice the difference in travel time between passenger vehicles and motorcycles, Tables 3.2, 3.3, and 3.4 present average travel time for passenger vehicles and motorcycles during morning peak period from 8:00 to 9:00. It is clear that the difference between passenger vehicles and motorcycles is very high for all routes.

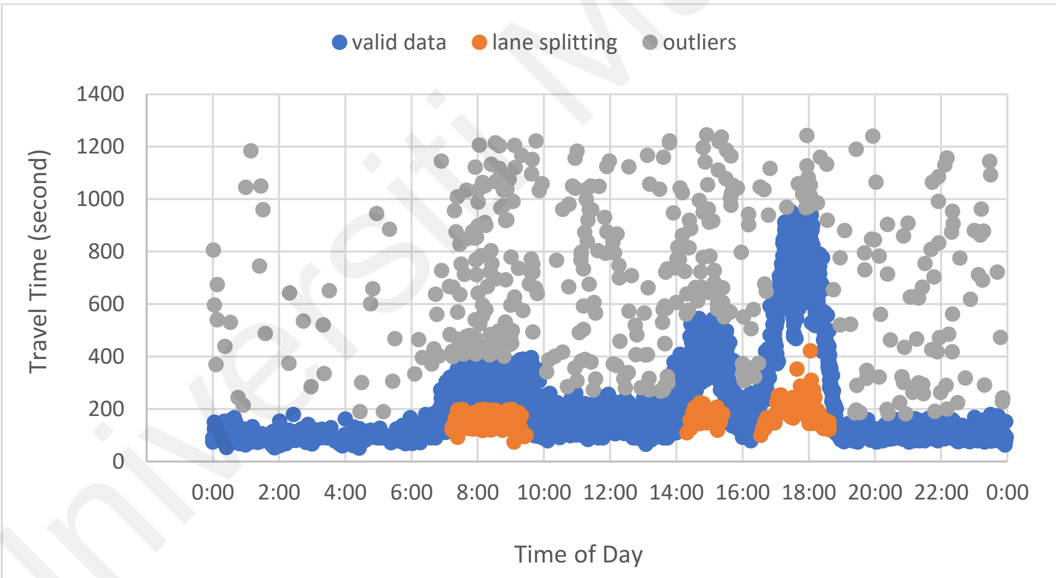



(a)




(b)

**Figure 3.4: Travel time datasets for 28th May 2018: (a) route A, (b) route B, and (c) route C.**




(c)

'Figure 3.4, continued'

**Table 3.2: Average travel time for passenger vehicles and motorcycles for route A**

| Time | Average travel time (Second) | | % Difference |
|---|---|---|---|
| | Passenger vehicles | motorcycles | |
| 8:00 - 8:09 | 697 | 309 | 55.65 |
| 8:10 - 8:19 | 669 | 270 | 59.64 |
| 8:20 - 8:29 | 616 | 273 | 55.59 |
| 8:30 - 8:39 | 571 | 248 | 56.64 |
| 8:40 - 8:49 | 505 | 272 | 46.07 |
| 8:50 - 8:59 | 411 | 229 | 44.34 |

**Table 3.3: Average travel time for Passenger vehicles and motorcycles for route B**

| Time | Average travel time (Second) | | % Difference |
|---|---|---|---|
| | Passenger vehicles | motorcycles | |
| 8:00 - 8:09 | 909 | 381 | 58.09 |
| 8:10 - 8:19 | 950 | 375 | 60.48 |
| 8:20 - 8:29 | 1018 | 423 | 58.46 |
| 8:30 - 8:39 | 1217 | 436 | 64.17 |
| 8:40 - 8:49 | 1203 | 412 | 65.76 |
| 8:50 - 8:59 | 860 | 320 | 62.79 |

**Table 3.4: Average travel time for Passenger vehicles and motorcycles for route C**

| Time | Average travel time (Second) | | % Difference |
|---|---|---|---|
| | Passenger vehicles | motorcycles | |
| 8:00 - 8:09 | 250 | 154 | 38.31 |
| 8:10 - 8:19 | 274 | 151 | 44.89 |
| 8:20 - 8:29 | 265 | 158 | 40.45 |
| 8:30 - 8:39 | 281 | 156 | 44.45 |
| 8:40 - 8:49 | 303 | 154 | 48.98 |
| 8:50 - 8:59 | 297 | 169 | 42.97 |

## 3.5 Data Analysis

### 3.5.1 Filtration of Travel Time Data

To evaluate the performance of the filtration algorithms to detect the outliers in the travel time datasets, the main approach used in the literature was by applying the filtration algorithms to field data and using graphs to present the algorithm performance (S. Chen, Wang, & van Zuylen, 2010; Clark et al., 2002; Dion & Rakha, 2006; Jang, 2016; Wu, Wu, & Rilett, 2020).

Regarding the current study, extensive empirical travel time data from three routes were used to validate the performance of the selected algorithms, as discussed in Chapter 2 for detecting outliers and lane splitting observations. These algorithms included the percentile algorithm, mean absolute deviation algorithm, TransGuide algorithm, Dion and Rakha algorithm (version 1 and version 2), and Jang algorithm. The equations for each algorithm were presented in Chapter 2. Two stages were adopted to evaluate the previously established filtration algorithms and identify each route's most appropriate algorithm and parameters. R software was used to analyse the effectiveness of the algorithms and calculations in this study. R software is an open-source programming language that highly used by statisticians for statistical computing and graphics (Venables, Smith, & R Development Core Team, 2009). It was used in this research because it is excellent for statistical computing and analysis, and it supports various data types.

Stage 1: Validation of the Previous Filtration Algorithm

The travel time datasets from 00:00 to 23:59 on 28th May 2018 for the three routes were used to validate the abovementioned algorithms. This day was selected because it is a weekday, and there are considerable amounts of lane-splitting observations. To reach the best performance for each algorithm, the values of algorithm parameters were calibrated using a trial-and-error method. In order to identify the best algorithm, the assessment was undertaken by observing the performance of each algorithm and comparing its performance with the other algorithms using graphs. In addition, the mean absolute relative error (MARE) was used as a numerical indicator to compare the algorithms' performances.

$$MARE = \frac{1}{n}\sum_{t=1}^{n}\frac{|x(t) - y(t)|}{x(t)} \qquad (37)$$

where n is the number of samples. x(t) is the average travel time from ground truth data (the valid observations in Figure 3.4) at collection interval t (five minutes), and y(t) is the average travel time from a filtering algorithm at the collection interval t (five minutes).

Travel time data collected by MAC addresses can be used as ground truth for intelligent transportation system applications (Haghani, Hamedi, Sadabadi, Young, &

Tarnoff, 2010). In this study, the ground truth was extracted manually as Moghaddam and Hellinga (2014) did in their study.

Stage 2: Sensitivity Analysis of the Algorithm Parameters

The best-selected algorithm for each route was applied on other ten-day datasets to check the sensitivity of the algorithm parameters for different days. After that, for the days that showed unaccepted results, the calibration of the parameters was performed to check the capability of the best algorithm to filter data from all days. The calibration of parameters was carried out using a trial-and-error method. The mean absolute relative error (MARE) was used as a numerical indicator to compare the algorithms' performance before and after calibration. The ground truth used to calculate the MARE in this stage is presented in Appendix A.

### 3.5.2 Analysis of Travel Time Reliability Measures

Following filtration of the datasets, TTI, PTI, and BTI were adopted as TTR measures in this study. The following equations were used to calculate these measures with respect to TOD, DOW, holidays, and election days (Culotta et al., 2019; Texas Transportation Institute & Cambridge Systematics Inc., 2005). Accordingly, TTI, PTI, and BTI were selected for several reasons; first, they are used by numerous authorities worldwide, and second, the concepts of these measures are easily understood than other TTR measures. TOD, DOW, holidays, and election days were selected because each category has different travel time pattern, and the researchers usually calculate travel time reliability measures for each category separately. Travel time datasets for the entire month (May 2018) were utilised for this analysis.

$$\text{TTI} = \frac{\text{Average travel time}}{\text{Free flow travel time}} \tag{38}$$

$$\text{PTI} = \frac{\text{95th percentile travel time}}{\text{Free flow travel time}} \tag{39}$$

$$\text{BTI} = \frac{\text{95th percentile travel time} - \text{Average travel time}}{\text{Average travel time}} \tag{40}$$

$$\text{Free flow travel time} = \frac{\text{length of route}}{\text{Free flow speed}} \tag{41}$$

The free-flow speed is defined as the 85th percentile speed during overnight hours between 10:00 p.m. and 5 a.m. (Z. Chen & Fan, 2020; Fan & Gong, 2017; Florida Department of Transportation, 2011; Schrank, Eisele, Lomax, & Bak, 2015).

It is also important to mention that travel time data were aggregated at 5 minutes intervals by calculating the average of the observations to avoid fluctuations in travel time (Yang & Wu, 2016).

# CHAPTER 4: RESULTS AND DISCUSSION

This chapter presents the results from analysing the data. As discussed in Chapter 3 (the methodology chapter), data analysis consisted of two main stages: (1) filtration of travel time data and (2) analysis of TTR measures (TTI, PTI and BTI) for the area under study. In this chapter, the first section presents the filtration results of travel time data and includes validating the previous filtration algorithm and checking the sensitivity of the algorithm parameters for different days. The second section discusses the results of the analysis of TTR measures (TTI, PTI and BTI) for the study area.

## 4.1 Filtration of Travel Time Data

### 4.1.1 Validation of the Previous Filtration Algorithm

The outlier detection algorithms presented in the literature review chapter were applied for the travel time datasets for routes A, B, and C. These included the percentile algorithm, mean absolute deviation algorithm, TransGuide algorithm, Dion and Rakha algorithm and, Jang algorithm. The travel time dataset from 00:00 to 23:59 during 28th May 2018 was used in the validation.

#### 4.1.1.1 Validation of the Previous Filtration Algorithm for Route A

Figure 4.1 shows the performance of the filtration algorithms for route A. Figure 4.1a presents the valid data after applying the percentile algorithm using the 25th percentile as the lower limit and the 75th percentile as the upper limit. The algorithm detects lane-splitting observations and most outliers but removed a significant number of valid observations. In contrast, Clark et al. (2002) proposed using the 10th percentile as a lower limit and the 90th percentile as an upper limit. However, these limits did not give a good performance. The mean absolute deviation algorithm with a validity range $M_e \pm 3\ MAD$ proposed by Clark et al. (2002) was not able to remove the lane-splitting data at the morning peak period and evening peak period and failed to detect many outliers. Thus, the validity range was modified to $M_e \pm 0.8\ MAD$. This modification had a positive effect on the performance of the algorithm, as shown in Figure 4.1b, but it removed a significant number of valid observations.

Figure 4.1c shows the performance of the TransGuide algorithm with $l_{th}$ = 0.5, however, it is not 0.2 as proposed by (Southwest Research Institute, 1998). This algorithm does not have the ability to detect lane-splitting data at the onset of the morning peak period and in the middle of the evening peak period. Figure 4.1d reveals the behaviour of Dion and Rakha's version 1 algorithm. The values of the parameters that gave the best results are $\beta$ = 0.5, $n_\sigma$ = 2.5, and size of the time window = 5 minutes. However, the algorithm cannot eliminate a few lane-splitting observations at the morning and evening peak periods and failed to detect many outliers at the evening peak period. Dion and Rakha's version 2 algorithm performance is presented in Figure 4.1e. Here, the adopted values of parameters are $\beta$ = 0.5, $n_\sigma$ = 2.5, size of the time window = 5 minutes, n skips = 10. Notably, the performance of version 2 is worse than version 1. Jang (2016) proposed using the parameters as follow: $\alpha$ = 0.35, $\beta$ = 3, and $\gamma$ = 0.3 in his algorithm. However, these values did not provide good results. Thus, modifications were proposed, where $\alpha$ = 1, $\beta$ = 1.5, and $\gamma$ = 0.3. The performance of the modified parameters algorithm is promising, as reflected in Figure 4.1f. The best algorithm for route A is the Jang algorithm.

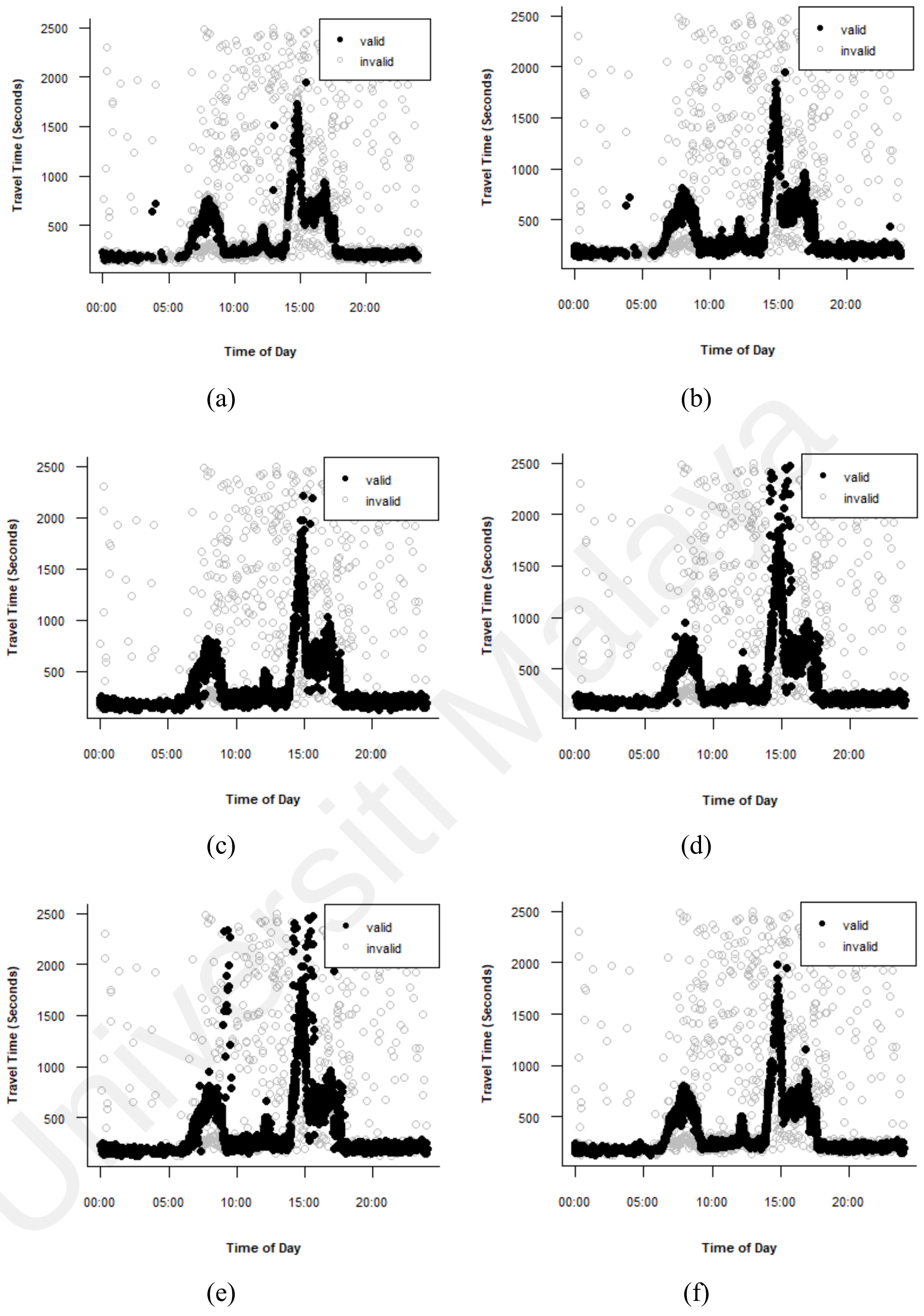
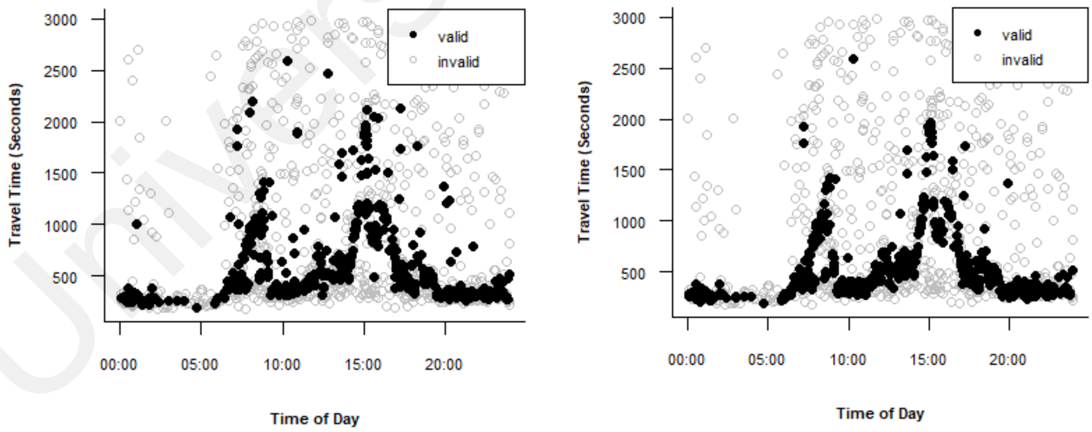



**Figure 4.1: Performance of filtration algorithms of route A: (a) Percentile algorithm, (b) Mean absolute deviation, (c) TransGuide algorithm, (d) Version 1 Dion and Rakha algorithm, (e) Version 2 Dion and Rakha algorithm, and (f) Jang algorithm**

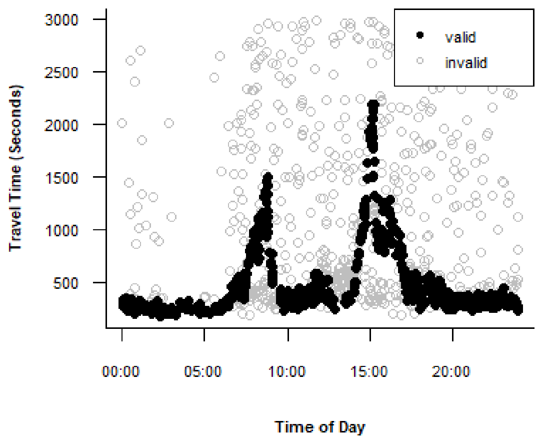
#### 4.1.1.2 Validation of the Previous Filtration Algorithm for Route B

The performances of the algorithms for route B are presented in Figure 4.2. Figure 4.2a illustrates applying the percentile test using the 25th percentile as a lower limit and the 75th percentile as an upper limit. However, the result is not good. The mean absolute deviation algorithm with a validity range $M_e \pm 0.3\ MAD$ cannot detect all lane-splitting observations at the morning peak period, as shown in Figure 4.2b. Figure 4.2c displays the result using the TransGuide algorithm with $l_{th}$ = 0.3. Noticeablly, the performance of this algorithm is the best for route B, where it efficiently detects the lane-splitting data and outliers. Figure 4.2d and Figure 4.2e display the behaviour of Dion and Rakha's version 1 and version 2 algorithms, respectively. Here, both algorithms exhibited a bad performance to detect outliers. Figure 4.2f illustrates the behaviour of the Jang algorithm using the following values of the parameters $\alpha$ = 0.3, $\beta$ = 1, and $\gamma$ = 0.3. However, the algorithm cannot remove all of the lane-splitting observations at the morning peak period, as a significant number of valid travel time observations are removed.

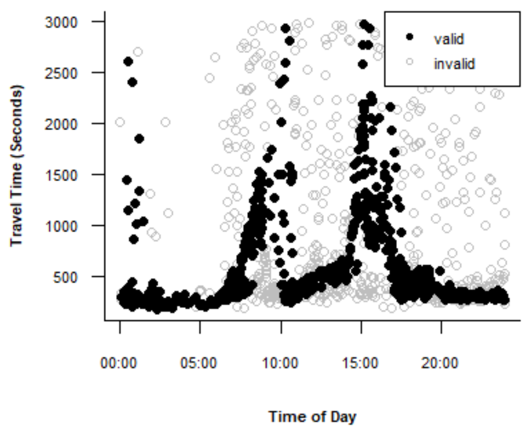



(a) (b)

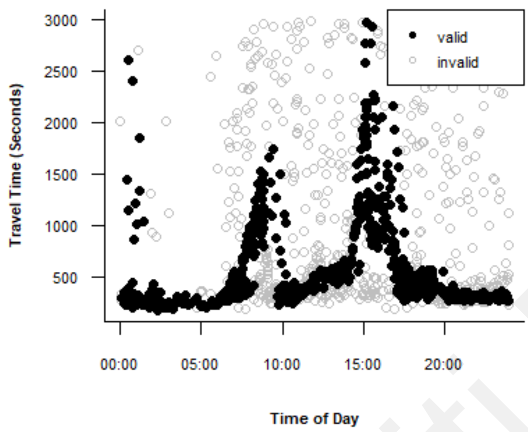
**Figure 4.2: Performance of filtration algorithms of route B: (a) Percentile algorithm, (b) Mean absolute deviation, (c) TransGuide algorithm, (d) Version 1 Dion and Rakha algorithm, (e) Version 2 Dion and Rakha algorithm, and (f) Jang algorithm**



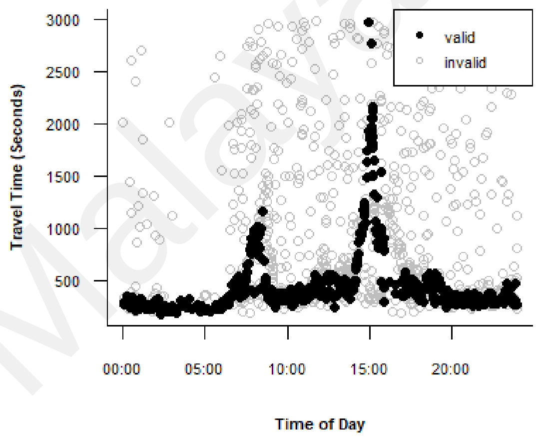

(c)




(d)




(e)




(f)

‘Figure 4.2, continued’

#### 4.1.1.3 Validation of the Previous Filtration Algorithm for Route C

The results for route C are shown in Figure 4.3. Figure 4.3a shows the result of applying the percentile algorithm using the 25th percentile as a lower limit and the 75th percentile as the upper limit. The algorithm showed a good performance detecting lane-splitting data and outliers but removed a significant number of valid observations. Figure 4.3b shows that the mean absolute deviation algorithm with a validity range of $M_e \pm 0.5\,MAD$ showed good performance detecting lane-splitting data and outliers but removed a significant number of valid observations. Figure 4.3c shows that the TransGuide algorithm with $l_{th}$ = 0.6 does not have the ability to detect the splitting lane data. Dion and Rakha’s version 1 and version 2 algorithms failed to remove lane-splitting data at the morning peak and were unable to detect a considerable number of outliers, as shown in Figure 4.3d and Figure 4.3e, respectively. Figure 4.3f exhibits the performance of Jang algorithm with $\alpha = 0.5$, $\beta = 1$, and $\gamma = 0.3$. The performance of this algorithm is excellent for detecting lane-splitting observations and outliers. The best algorithm for route C is the Jang algorithm.

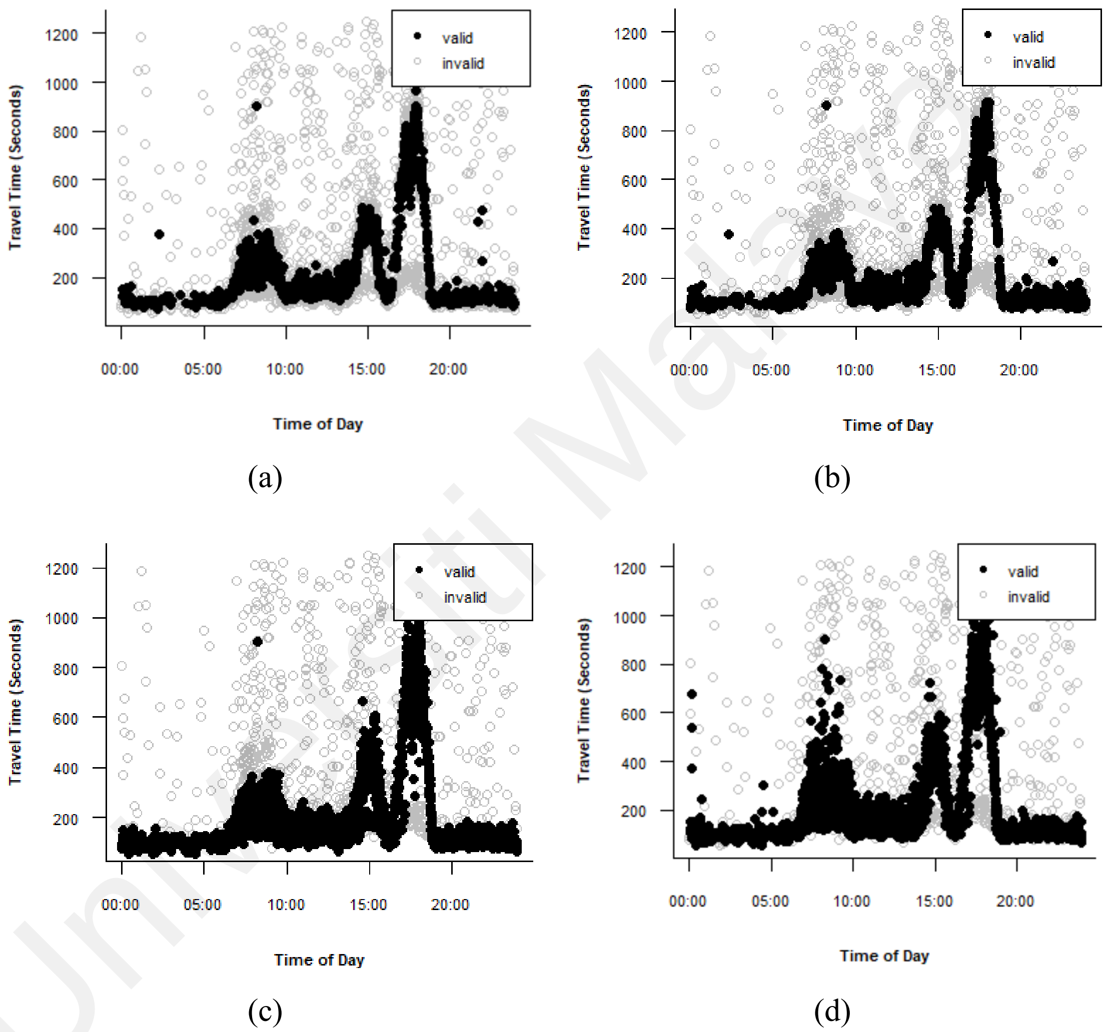
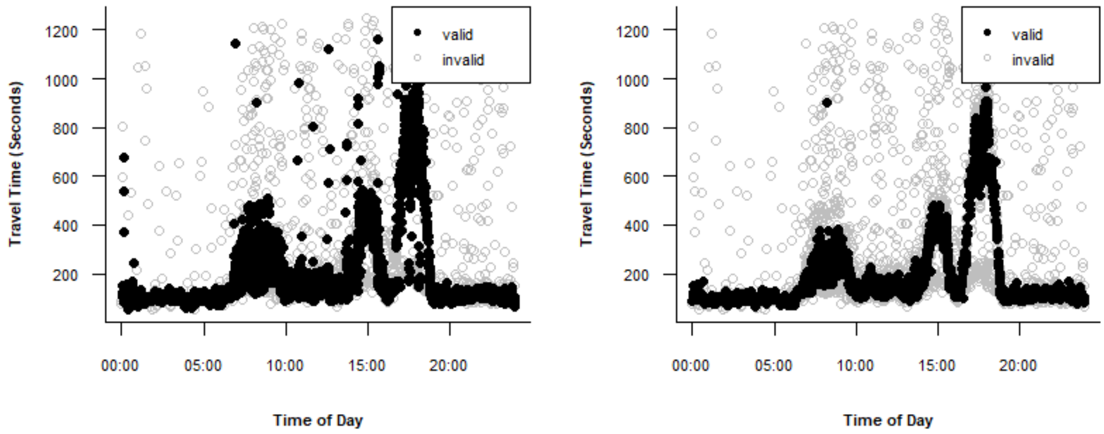



**Figure 4.3: Performance of filtration algorithms of route C: (a) Percentile algorithm, (b) Mean absolute deviation, (c) TransGuide algorithm, (d) Version 1 Dion and Rakha algorithm, (e) Version 2 Dion and Rakha algorithm, and (f) Jang algorithm**




(e) (f)

'Figure 4.3, continued'

Table 4.1 presents the MARE values for the three routes for 28th May 2018. The algorithm with the smallest MARE value indicates the best performance, which means that this algorithm has minimum error relative to the ground truth data. The table shows that the Jang algorithm is the best algorithm for Routes A and C, while the TransGuide algorithm is the best for Route B. This finding affirmed the conclusions drawn from the discussion of Figures 4.1-4.3.

**Table 4.1: MARE values for the three routes for 28th May**

| Filtering Algorithm | Mean absolute relative error (MARE) | | |
| --- | --- | --- | --- |
| | Route A | Route B | Route C |
| Percentile algorithm | 0.132 | 0.392 | 0.158 |
| Mean absolute deviation algorithm | 0.123 | 0.386 | 0.140 |
| TransGuide algorithm | 0.050 | 0.129 | 0.078 |
| Dion and Rakha algorithm Version 1 | 0.049 | 0.313 | 0.082 |
| Dion and Rakha algorithm Version 2 | 0.052 | 0.226 | 0.083 |
| Jang algorithm | 0.028 | 0.153 | 0.050 |

### 4.1.2 Check of the Sensitivity of the Algorithm Parameters for Different Days

Travel time datasets from ten days are used to check whether there is a need to calibrate the parameters of the best algorithm for each day to obtain the best performance or whether the parameters used in the previous section are suitable for all days. These days are the 2nd, 5th, 8th, 11th, 14th, 17th, 20th, 23rd, 26th, and 29th of May 2018. It was found that the 2nd, 8th, 14th, 17th, and 23rd of May were weekdays. While, 5th, 20th, and 26th of May were weekends. The 11th of May was an election day, and the 29th of May was a holiday. However, only the algorithm that gave the best performance for each route is tested in this section.

When any of the days mentioned above show unaccepted performance, it indicates that the algorithm's parameters are sensitive for different days. Thus, calibration of the algorithm’s parameters is required for the days showing unaccepted performance. Calibration means modifying the values of the algorithm’s parameters using the trial-and-error method to obtain the algorithm's best performance. This step aims to check the ability of the algorithm to filter data from all days. For example, if an algorithm has three user-defined parameters X, Y, and Z. Usually, the developers of the filtration algorithms propose values for each parameter. First, the proposed values are used. If the performance is acceptable, adjusting the values will be not necessary to carry out, while if the performance is unacceptable, the parameters need to be calibrated. Parameter X is the first parameter to be calibrated by increasing and decreasing the value until reaching to the best performance. After that, parameter Y is the second parameter to be calibrated by increasing and decreasing the value until reaching to the best performance. Finally, parameter Z is the last parameter to be calibrated by increasing and decreasing the value until it reaches the best performance condition.

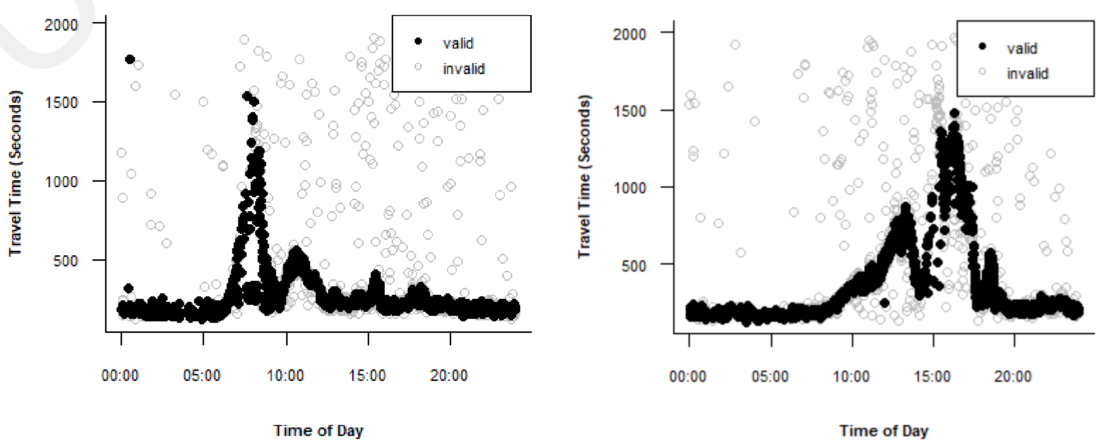
#### 4.1.2.1 Check of the Sensitivity of the Algorithm Parameters for Different Days for Route A

The Jang algorithm with $\alpha = 1$, $\beta = 1.5$, $\gamma = 0.3$, and $t_W = 5$ exhibited the best performance for route A for the 28th May travel time dataset. The performance of the Jang algorithm with these parameters is applied for the ten days. Figure 4.4 shows the performance of the Jang algorithm for route A for the days mentioned above. Five datasets have poor performance (refer to Figures 4.4a, 4.4c, 4.4e, 4.4f, and 4.4h) because considerable amounts of lane-splitting observations during morning peak period remained after filtration. Interestingly, all these figures represent weekdays. Thus, the parameters of the Jang algorithm for route A are sensitive for datasets from different days. Therefore, it is crucial to calibrate the parameters for each day to gain acceptable results.

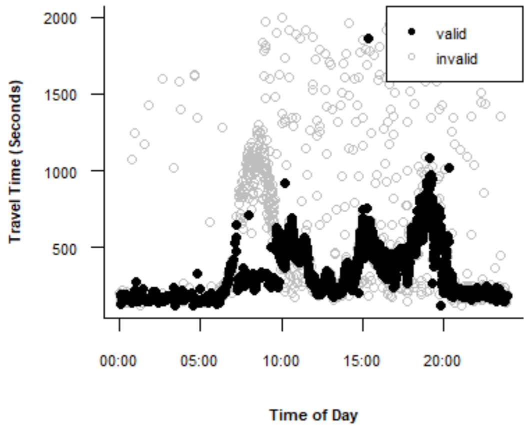
Employing the trial-and-error method, the parameters of the Jang algorithm for the datasets showing unacceptable results in Figure 4.4 are calibrated. Figure 4.5 displays the performance of the Jang algorithm before and after calibration of the parameters. Undoubtedly, the performance of the algorithm after calibration outperforms the performance of the algorithm before calibration. This indicates that the travel time dataset for each day needs to be calibrated.

The Jang algorithm has four parameters: $\alpha$, $\beta$, $\gamma$, and $t_W$. Table 4.2 shows the values of the parameters after calibration for the days showing unacceptable performance before calibration for route A. Here, $\alpha$, $\beta$, and $t_W$ are sensitive, while $\gamma$ is insensitive. Indeed, four of five days have the same parameters. This indicates that, after calibration, there are two different new parameters sets, namely $\alpha$ = 0.5, $\beta$ = 0.5, $\gamma$ = 0.3, and $t_W$ = 1 are for the 8th of May. While, another parameters set, namely, $\alpha$ = 0.5, $\beta$ = 0.5, $\gamma$ = 0.3, and $t_W$ = 1, is for the 2nd, 14th, 17th, and 23rd of May. $t_W$ is different for just one day. This means that the sensitivity of $t_W$ is less than $\alpha$ and $\beta$.

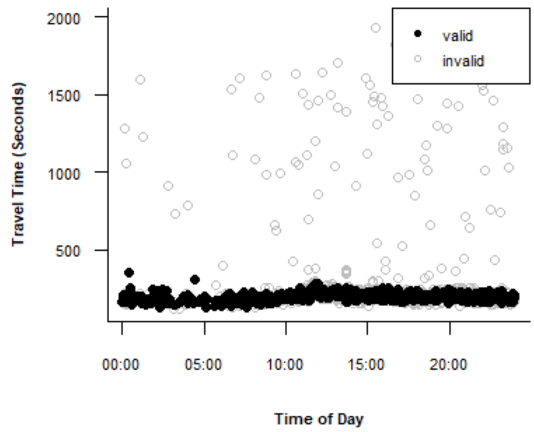





(a) (b)

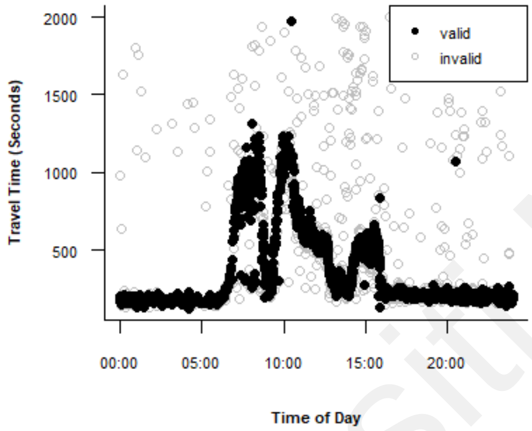



(c)

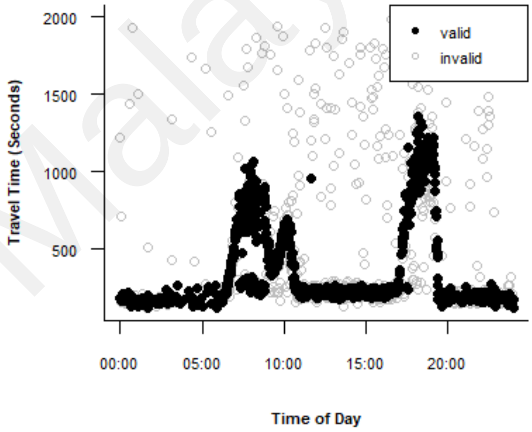



(d)

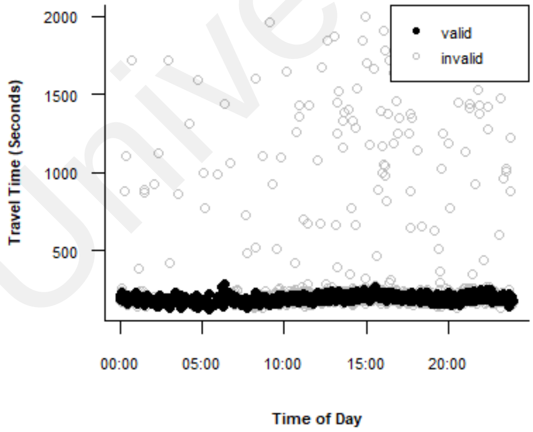
**Figure 4.4: Performance of Jang algorithm of route A for ten days: (a) 2nd May, (b) 5th May (c) 8th May, (d) 11th May, (e) 14th May, (f) 17th May (g) 20th May, and (h) 23rd May**



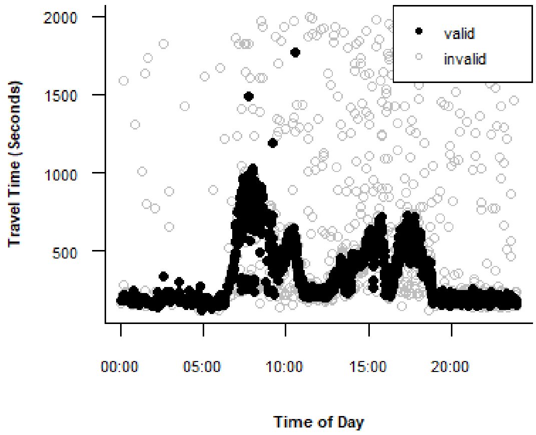

(e)



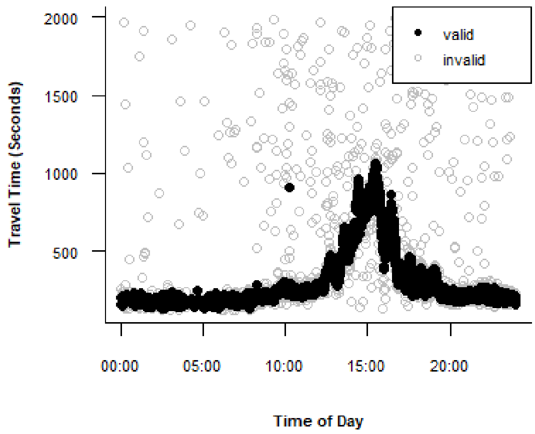

(f)



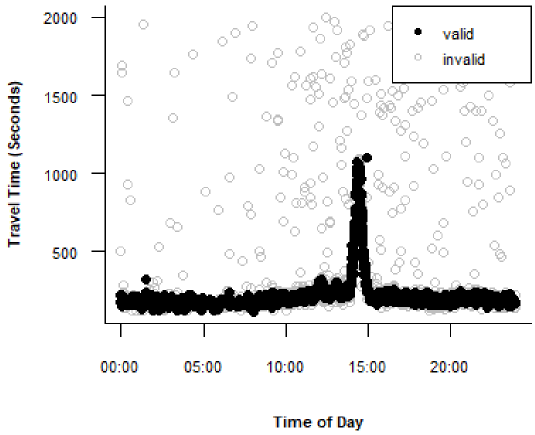

(g)




(h)

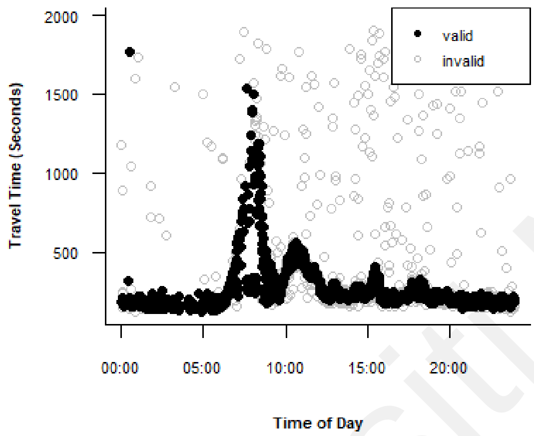



(i)

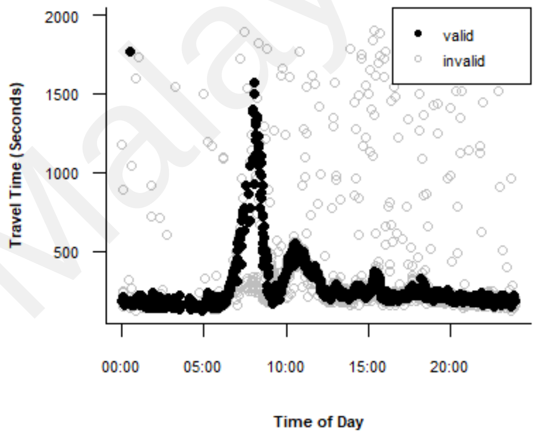



(j)

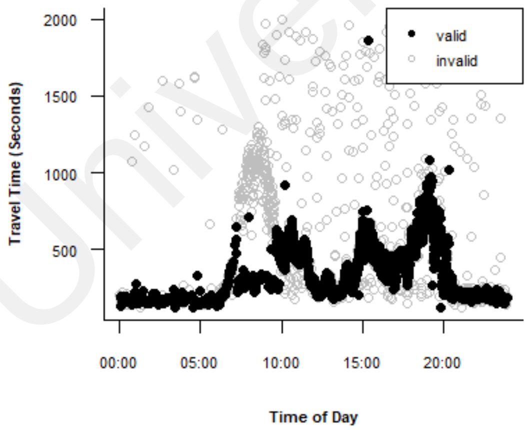
'Figure 4.4, continued'



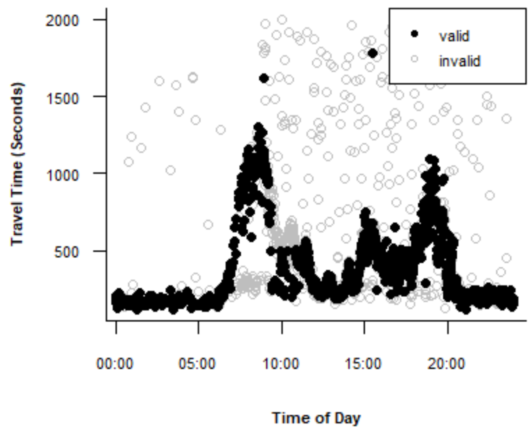

(a)



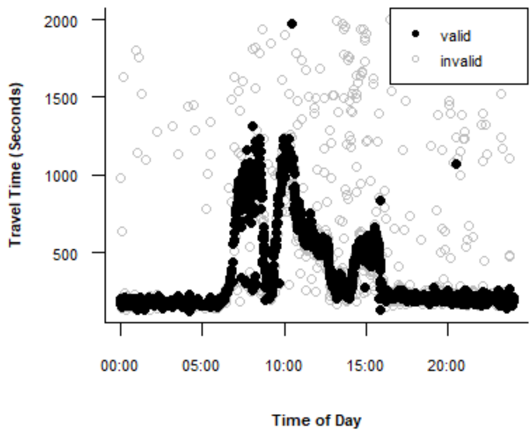

(b)



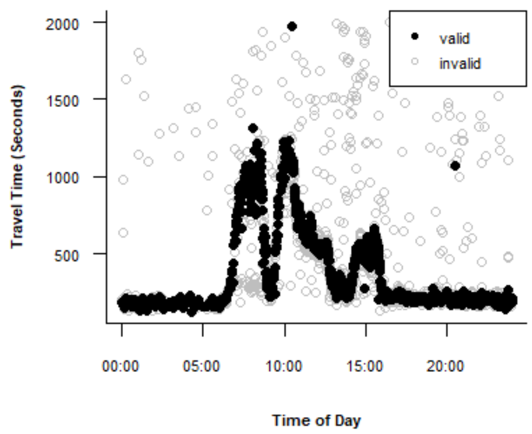

(c)




(d)

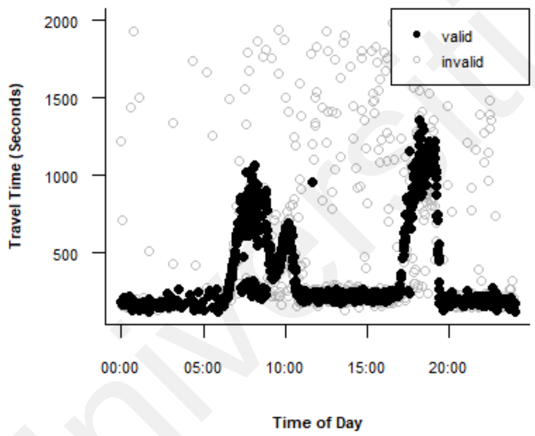



(e)

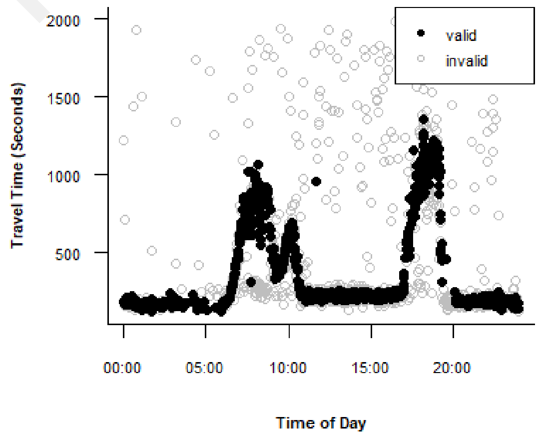



(f)

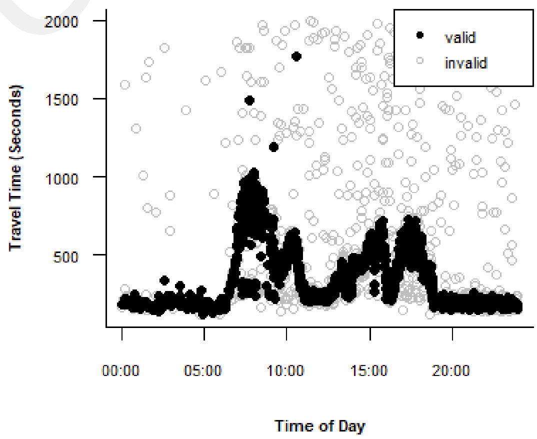
**Figure 4.5: Performance of Jang algorithm of route A for some days before and after calibration: (a) 2nd May – before calibration, (b) 2nd May – after calibration, (c) 8th May – before calibration, (d) 8th May – after calibration (e) 14th May – before calibration, (f) 14th May – after calibration, (g) 17th May – before calibration, (h) 17th May – after calibration, (i) 23rd May – before calibration, and (j) 23rd May – after calibration**

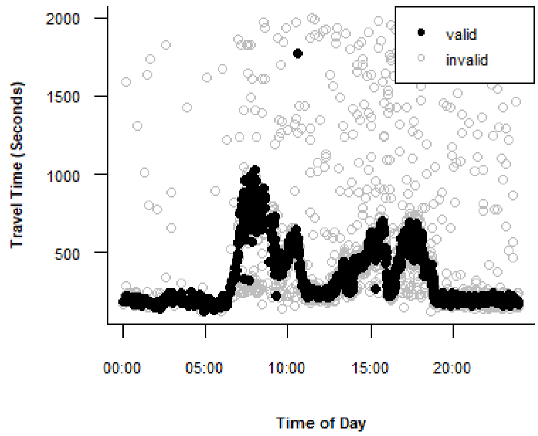



(g)




(h)



(i) (j)

'Figure 4.5, continued'

**Table 4.2: Values of the parameters after calibration for route A**

| Date | $\alpha$ | $\beta$ | $\gamma$ | $t_W$ |
|---|---|---|---|---|
| 2nd May | 0.5* | 1* | 0.3 | 5 |
| 8th May | 0.5* | 0.5* | 0.3 | 1* |
| 14th May | 0.5* | 1* | 0.3 | 5 |
| 17th May | 0.5* | 1* | 0.3 | 5 |
| 23rd May | 0.5* | 1* | 0.3 | 5 |
| Value before calibration | 1 | 1.5 | 0.3 | 5 |

*Note: * means different from the value before calibration*

Table 4.3 presents the MARE values for route A before and after calibration for the days that need to be calibrated. The MARE values for entire day after calibration were less than before calibration for all days, indicating that the calibration of the parameters improved the performance of Jang algorithm. For the morning peak period, the amounts of lane-splitting observations before calibration were considerable for all days as shown in Figures 4.5a, 4.5c, 4.5e, 4.5g, and 4.5i. As such, the MARE values for 8:00 - 9:00 before calibration were high as presented in Table 4.3. The MARE values for 8:00 - 9:00 after calibration were much less than before calibration for all days. This indicates that the calibration of the parameters highly improved the performance of Jang algorithm during this period.

**Table 4.3: MARE values for route A**

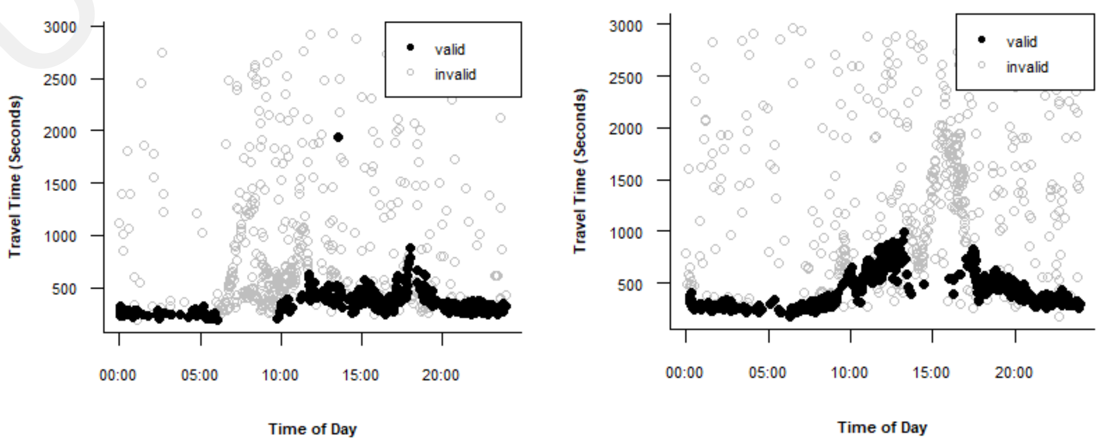
| Date | Mean absolute relative error (MARE) | | | |
|---|---|---|---|---|
| | Entire Day | | 8:00 - 9:00 | |
| | Before calibration | After calibration | Before calibration | After calibration |
| 2nd May | 0.051 | 0.040 | 0.319 | 0.069 |
| 8th May | 0.100 | 0.063 | 0.779 | 0.027 |
| 14th May | 0.040 | 0.033 | 0.205 | 0.031 |
| 17th May | 0.050 | 0.039 | 0.181 | 0.038 |
| 23rd May | 0.037 | 0.029 | 0.207 | 0.042 |

#### 4.1.2.2 Check of the Sensitivity of the Algorithm Parameters for Different Days for Route B

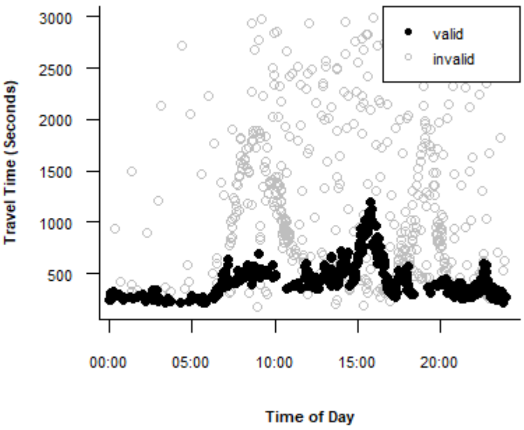
The TransGuide algorithm with $l_{th}$ = 0.3, and $t_W$ = 5 was the best algorithm for route B for the dataset of 28th May. The ten days, as mentioned earlier, are filtered using the TransGuide algorithm with $l_{th}$ = 0.3, and $t_W$ = 5. Figure 4.6 presents the performance of the TransGuide algorithm for route B for the ten days. Figures 4.6a, 4.6b, 4.6c, 4.6e, 4.6h, and 4.6i show the datasets with poor performance. These six figures represent the data of 4 weekdays and 2 weekends. Indeed, the parameters of the TransGuide algorithm for route B are sensitive for datasets from different days. Thus, the algorithm needs to be calibrated for each day. Figure 4.7 illustrates the performance of the days having unacceptable results using the TransGuide algorithm with $l_{th}$ = 0.3, and $t_W$ = 5. In addition, it presents the performance of these days after calibration of the parameters. Thus, the calibrated parameters positively affect the performance of the algorithm.

Table 4.4 shows the values of the parameters after calibration for the days that showed unacceptable performance before calibration for route B. Here, both parameters, $l_{th}$ and $t_W$ are sensitive. After calibration, there are five different new parameters sets. $l_{th}$ = 0.5, and $t_W$ = 5 are the best parameters for the 2nd and 5th of May. In contrast, each day for the rest of the days has a different parameter set.

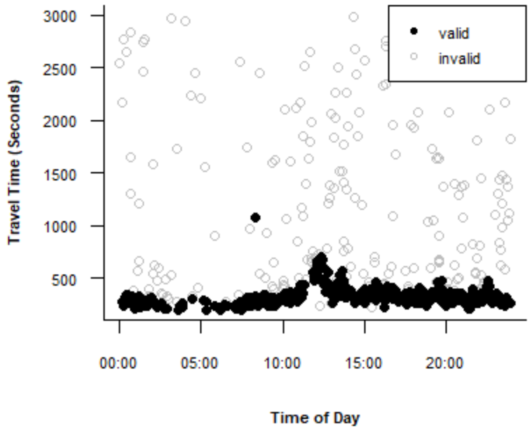





(a) (b)

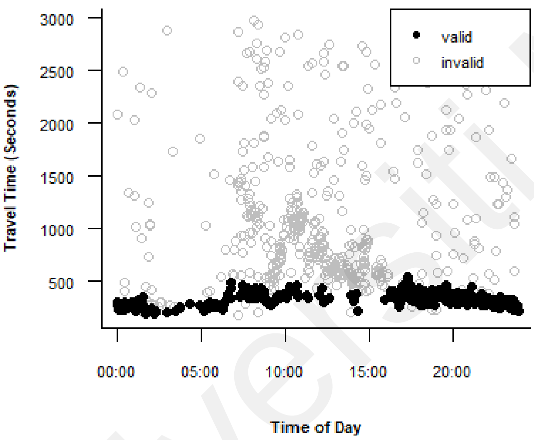



(c)

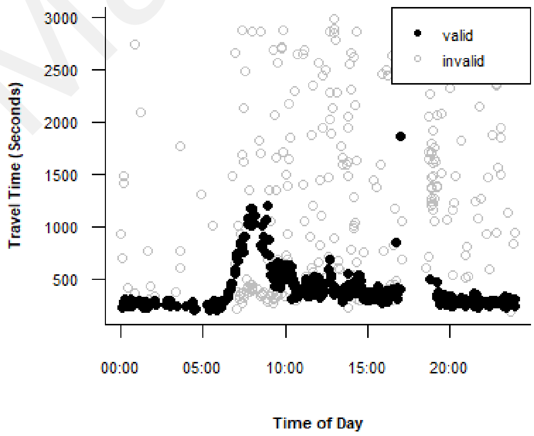



(d)

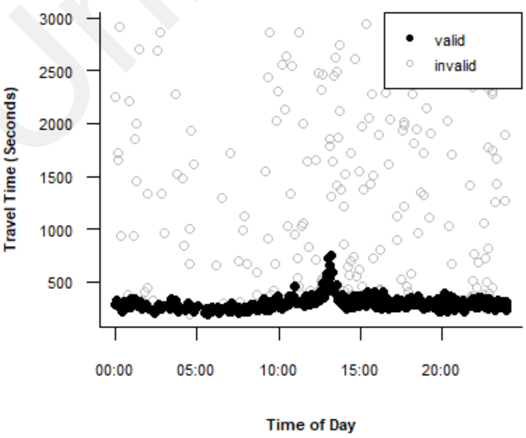
**Figure 4.6: Performance of TransGuide algorithm of route B for ten days: (a) 2nd May, (b) 5th May, (c) 8th May, (d) 11th May, (e) 14th May, (f) 17th May, (g) 20th May, (h) 23rd May, (i) 26th May, and (j) 29th May**



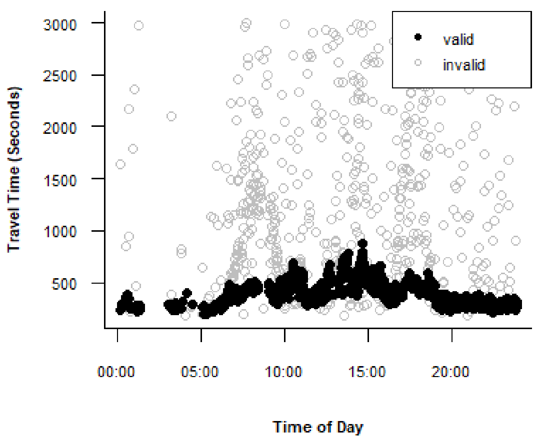

(e)



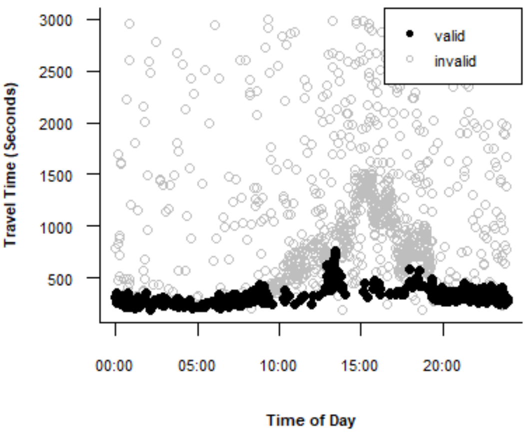

(f)



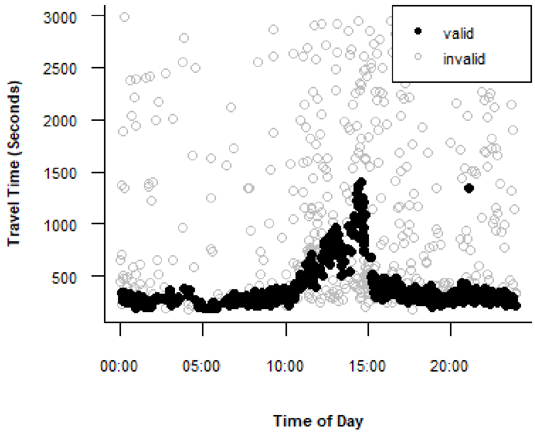

(g)




(h)

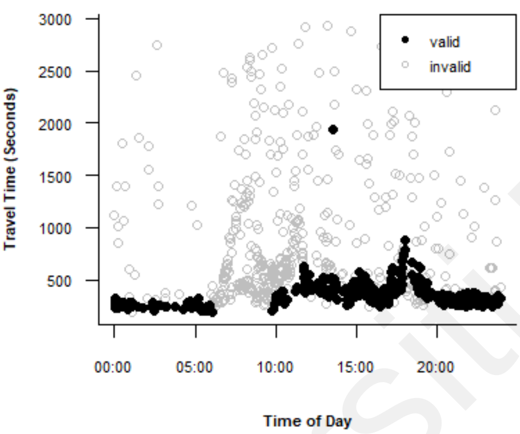



(i)

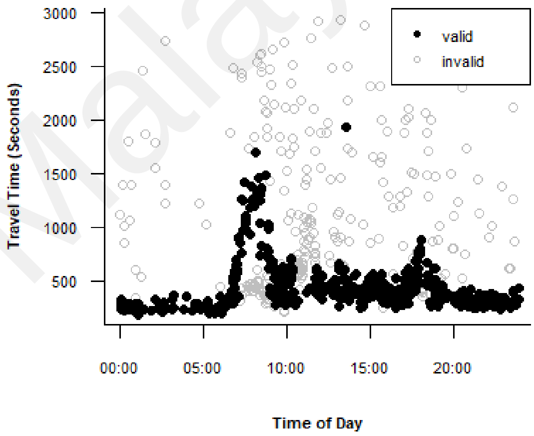



(j)

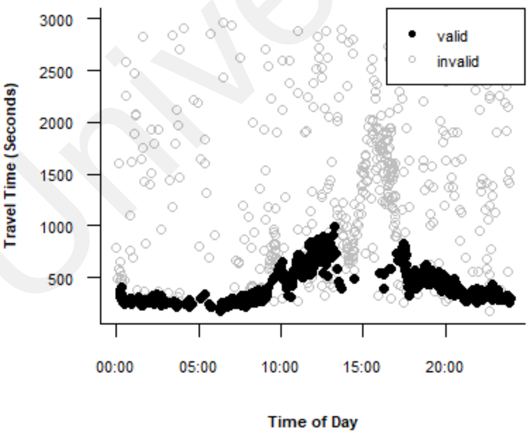
'Figure 4.6, continued'



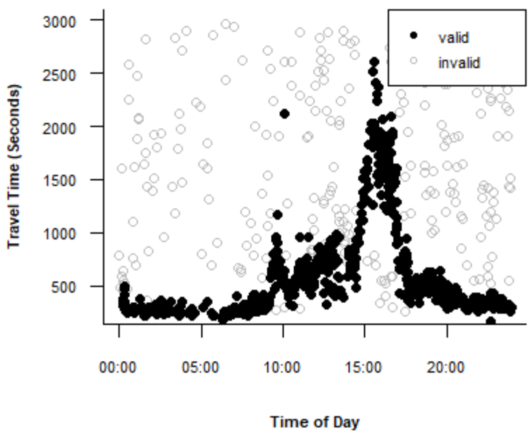

(a)



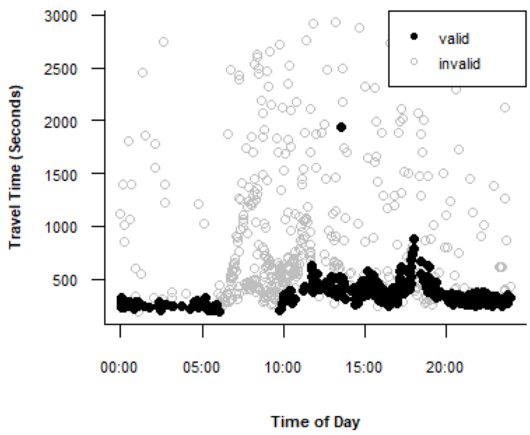

(b)



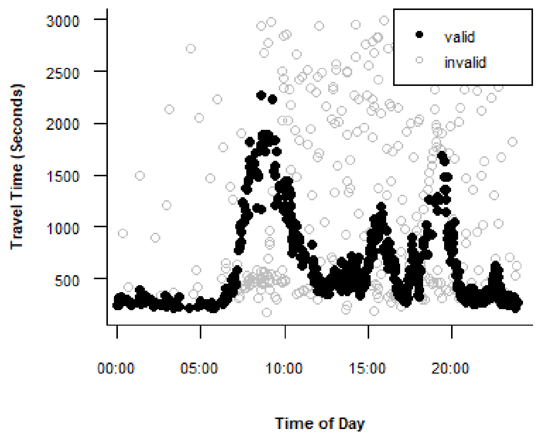

(c)




(d)



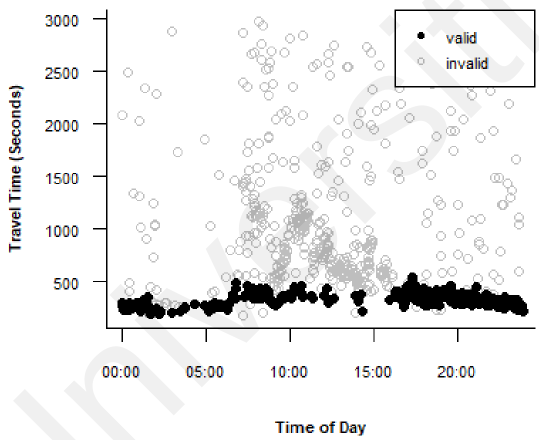

(e)



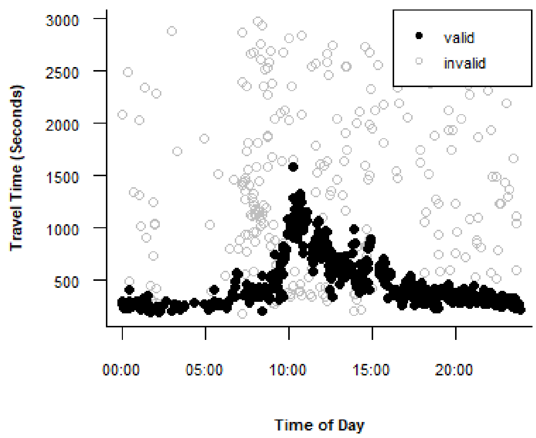

(f)

**Figure 4.7: Performance of TransGuide algorithm of route B for some days before and after calibration: (a) 2nd May – before calibration, (b) 2nd May – after calibration, (c) 5th May – before calibration, (d) 5th May – after calibration, (e) 8th May – before calibration, (f) 8th May – after calibration, (g) 14th May – before calibration, (h) 14th May – after calibration, (i) 23rd May – before calibration, (j) 23rd May – after calibration, (k) 26th May – before calibration , and (l) 26th May – after calibration.**

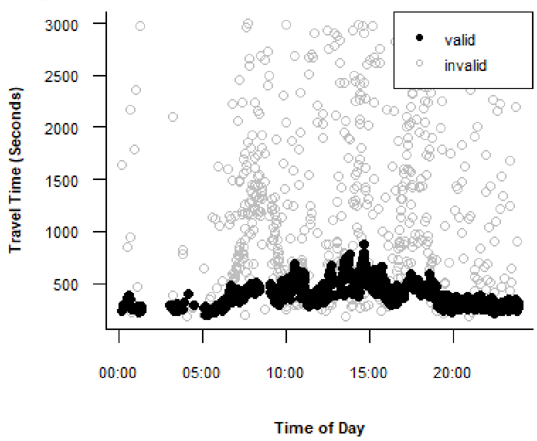


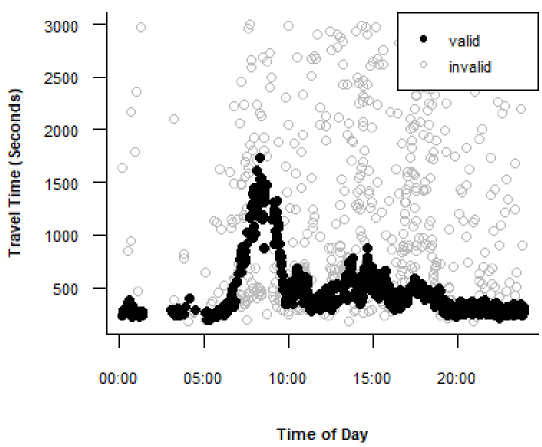



(g)

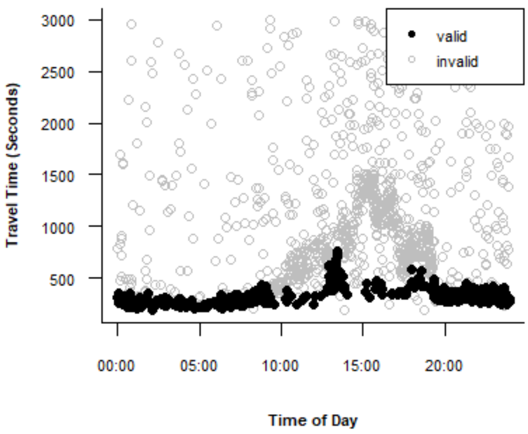



(h)

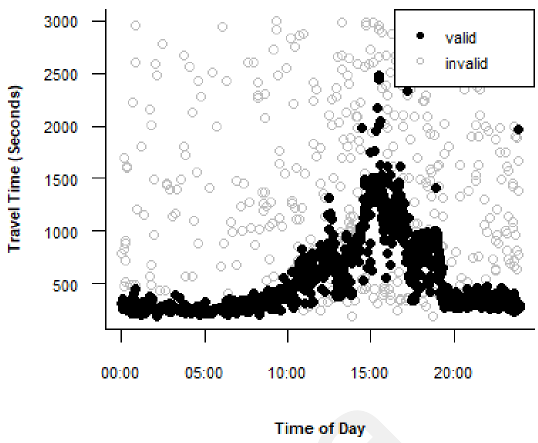


(i) (j)




(k)




(l)

'Figure 4.7, continued'

**Table 4.4: Values of the parameters after calibration for route B**

| Date | $l_{th}$ | $t_W$ |
|---|---|---|
| 2nd May | 0.5* | 5 |
| 5th May | 0.5* | 5 |
| 8th May | 0.35* | 4* |
| 14th May | 0.45* | 5 |
| 23rd May | 0.3 | 6* |
| 26th May | 0.5* | 4* |
| Value before calibration | 0.3 | 5 |

*Note: * means different from the value before calibration*

Table 4.5 presents the MARE for route B for the days that need to be calibrated. The values of MARE for all days after calibration were less than before calibration, indicating that the calibration of the parameters improved the performance of TransGuide algorithm.

**Table 4.5: MARE values for route B**

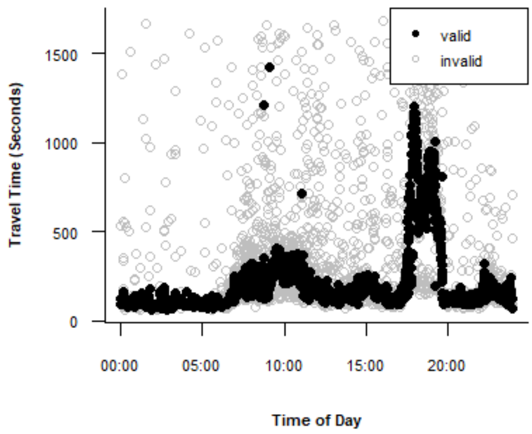
| Date | Mean absolute relative error (MARE) | |
|---|---|---|
| | Before Calibration | After Calibration |

| | | |
|---|---|---|
| 2nd May | 0.318 | 0.081 |
| 5th May | 0.227 | 0.036 |
| 8th May | 0.301 | 0.067 |
| 14th May | 0.329 | 0.030 |
| 23rd May | 0.137 | 0.069 |
| 26th May | 0.357 | 0.034 |

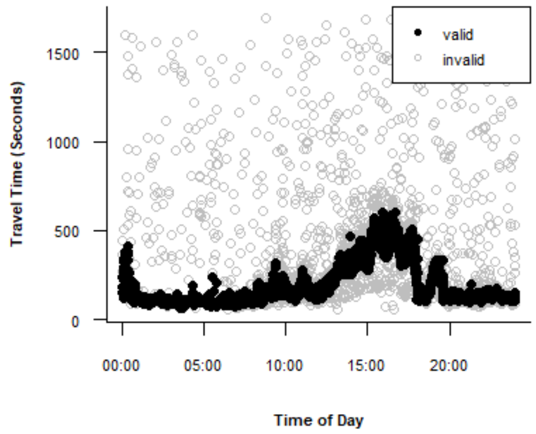
#### 4.1.2.3 Check of the Sensitivity of the Algorithm Parameters for Different Days for Route C

For route C, the Jang algorithm with $\alpha$ = 1, $\beta$ = 1, $\gamma$ = 0.3, and $t_W$ = 5 was the best algorithm for the 28th May travel time dataset. The Jang algorithm with these parameters is applied for the ten days. Figure 4.8 illustrates the performance of the Jang algorithm on route C for the days mentioned above. On the other hand, Figure 4.8h shows the poor performance for the data for 23rd May because considerable lane-splitting observations during evening peak period remained after filtration. So, given that one dataset showed unacceptable performance, the parameters of the Jang algorithm for route C are sensitive and needed to be calibrated for each day to gain accurate performance. Figure 4.9 shows the Jang algorithm's performance before calibration of the parameters and after calibration for 23rd May. Here, the performance after the calibration is highly improved.

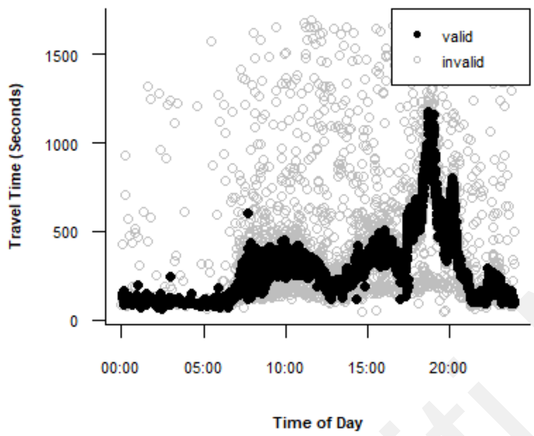
Table 4.6 shows the values of Jang algorithms' parameters after calibration for the day showing unacceptable performance before calibration for route C. Here, only $\alpha$ is sensitive, while other parameters are insensitive. For this route, only one parameter of one day is sensitive. Therefore, the sensitivity of route C is less than the other routes.



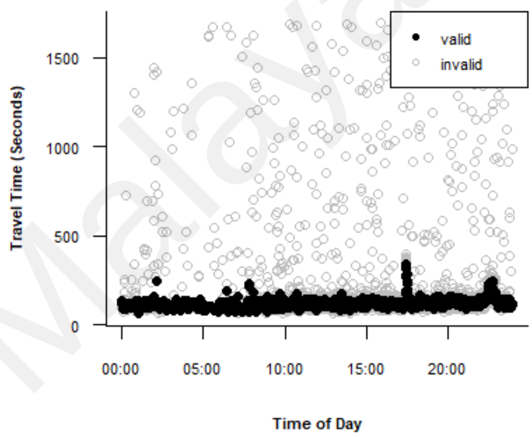

(a)




(b)

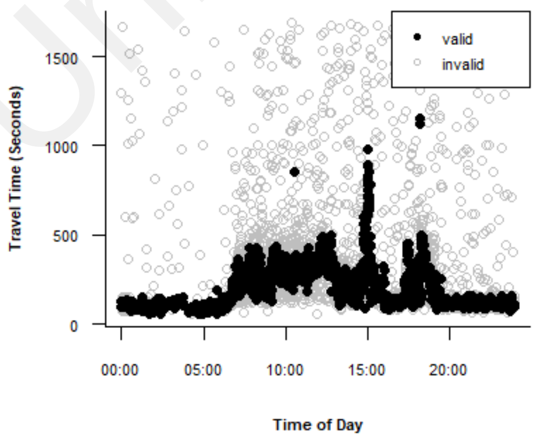



(c)

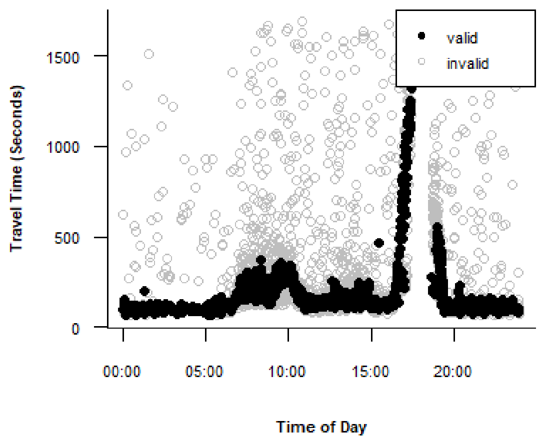



(d)

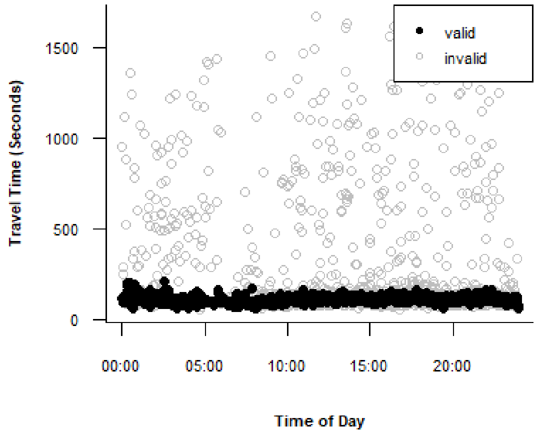
**Figure 4.8: Performance of Jang algorithm of route C for ten days: (a) 2nd May, (b) 5th May, (c) 8th May, (d) 11th May, (e) 14th May, (f) 17th May, (g) 20th May, (h) 23rd May, (i) 26th May, and (j) 29th May**



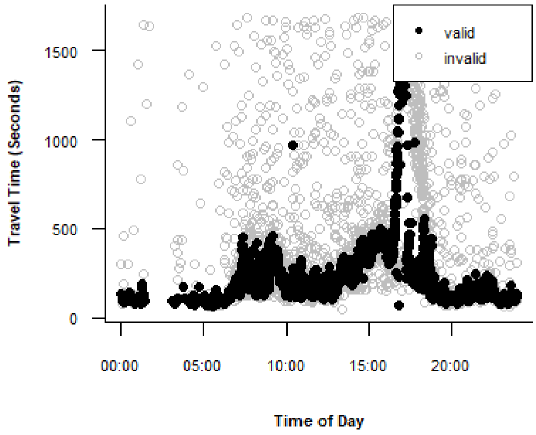

(e)



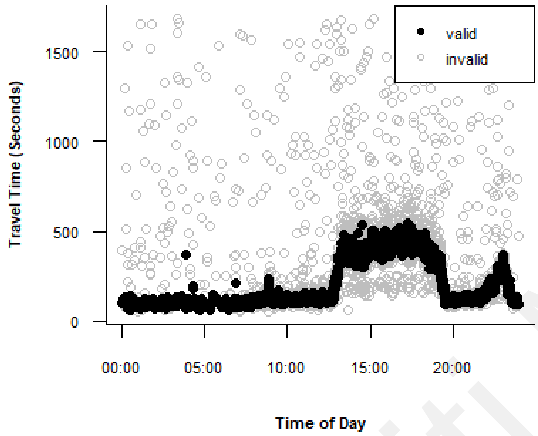

(f)



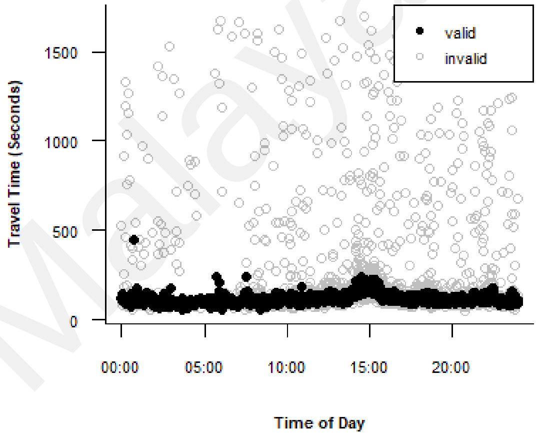

(g)




(h)

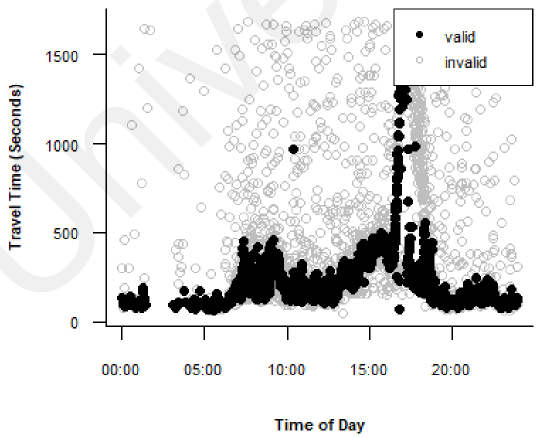



(i)

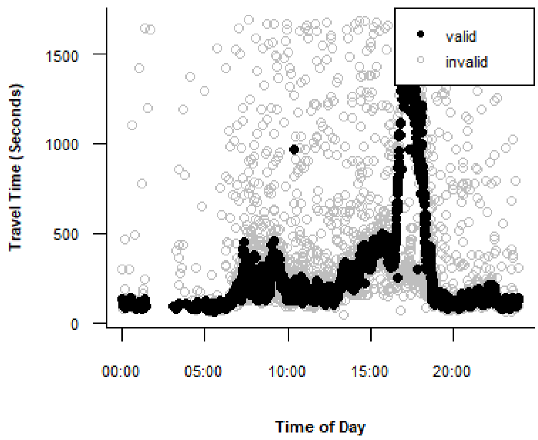



(j)

'Figure 4.8, continued'




(a)




(b)

**Figure 4.9: Performance of Jang algorithm of route C for 23rd May before and after calibration: (a) before calibration, and (b) after calibration**

**Table 4.6: Values of the parameters after calibration for route C**

| Date | $\alpha$ | $\beta$ | $\gamma$ | $t_W$ |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 23rd May | 0.5* | 1 | 0.3 | 5 |
| Value before calibration | 1 | 1 | 0.3 | 5 |

*Note: * means different from the value before calibration*

Table 4.7 presents the MARE for route C before and after calibration for 23rd May. The MARE value after calibration was less than before calibration, indicating that the calibration of Jang algorithm parameters improved the performance of Jang algorithm. For the evening peak period, the number of lane-splitting observations before calibration was considerable as shown in Figure 4.9a. As such, the MARE value for 17:00 - 18:00 before calibration (0.544) was very high as presented in Table 4.7. The MARE value for 17:00 - 18:00 after calibration (0.039) was much less than before calibration (0.544). This indicates that the calibration of the parameters highly improved the performance of Jang algorithm during this period.

**Table 4.7: MARE values for route C**

| Date | Mean absolute relative error (MARE) | | | |
|---|---|---|---|---|
| | Entire Day | | 17:00 - 18:00 | |
| | Before calibration | After calibration | Before calibration | After calibration |
| 23rd May | 0.104 | 0.077 | 0.544 | 0.039 |

### 4.1.3 Summary of Filtration of Travel Time Data

The summary of the evaluation of filtration algorithms and the sensitivity of the algorithms' parameters are depicted in Table 4.8. The Jang algorithm is the best for routes A and C, while the TransGuide algorithm is the best for route B. In order to compare the routes concerning the sensitivity of the algorithm's parameters, the number of days with unacceptable performance before calibration and the number of new parameter sets after calibration are used. Noticeably, route C is less sensitive, where only one day showed unacceptable performance before calibration. In contrast, route B is more sensitive since it has the highest number of days with unacceptable performance before calibration and the highest number of new parameters sets after calibration. Lastly, given that routes A and C are less sensitive than route B, it can be concluded that the Jang algorithm is less sensitive than the TransGuide algorithm.

**Table 4.8: Summary of filtration of travel time analysis**

| Route | Best algorithm | No. of days with unacceptable performance before calibration | No. of parameters sets after calibration |
|---|---|---|---|
| A | Jang | 5 | 2 |
| B | TransGuide | 6 | 5 |
| C | Jang | 1 | 1 |

The route length is the distance between the two Wi-Fi sensors at the start and the end of the route. In order to test the relationship between the distance between the sensors and the number of travel time observations, Pearson correlation coefficient, effect size, and coefficient of determination are calculated. Table 4.9 shows that the effect size (d) is compatible with the Coefficient of Determination ($R^2$) since $R^2$ (0.93) and d (-7.47) are very large based on Cohen's standard (Cohen, 2013). In addition, the slope of the trend line in Figure 4.10 and the sign of d are negative. Thus, there is a very large negative correlation between the distance between the sensors and the number of observations. This indicates that the distance between the sensors has to be shortened to obtain more observations. As depicted in Table 4.8 and Figure 4.10, it can be concluded that an increment in the number of observations makes the Jang algorithm the best filtration algorithm since the Jang algorithm was the best algorithm for routes A and C, having more observations than for route B.

**Table 4.9: Results of correlation test**

| Pearson correlation coefficient (r) | Effect Size (d) | Coefficient of Determination ($R^2$) |
|---|---|---|
| -0.966 | -7.47 | 0.93 |

Regarding the sensitivity of the algorithms' parameters, route C is less sensitive than the other routes, having the highest number of observations. Route B is more sensitive, having the lowest number of observations. However, route A is more sensitive than route C and less sensitive than route B. Route A has fewer observations than route C and more than route B. Thus, it can be concluded that the algorithms' parameters' sensitivity is in inverse proportion to the number of observations. Given the inverse proportion between the distance between the sensors (or the length of the route) and the number of observations, the sensitivity of the algorithms' parameters is directly proportional to the distance between the sensors.

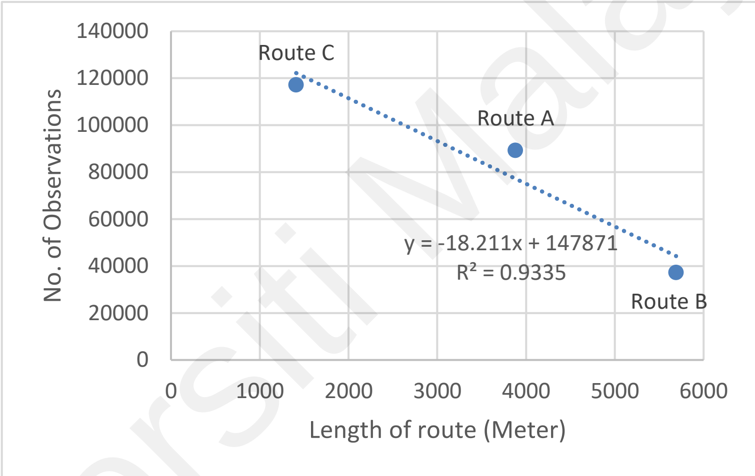



**Figure 4.10: The relationship between the length of routes and the number of observations**

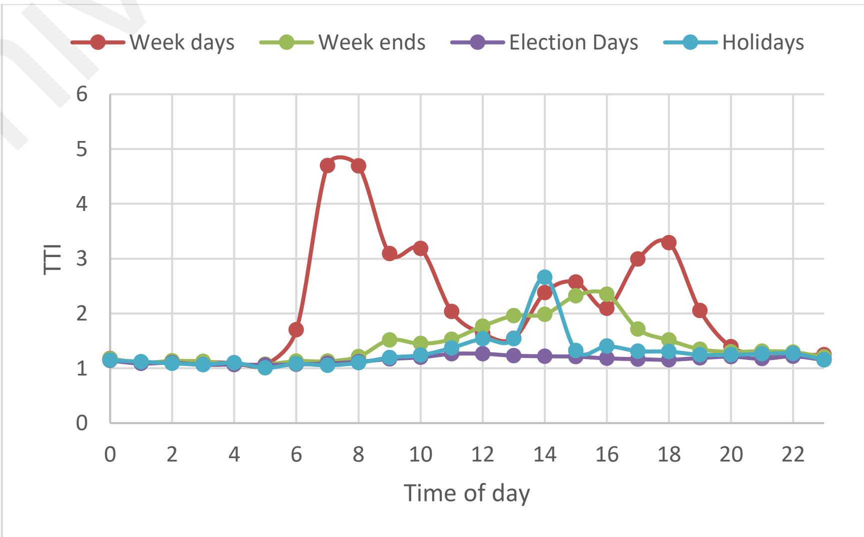
## 4.2 Analysis of Travel Time Reliability Measures for Weekdays, Weekends, Election Days, and Holidays

There were 18 weekdays, 8 weekends, 3 general election days and 2 holidays during May 2018. The analysis was undertaken inclusive of these days.

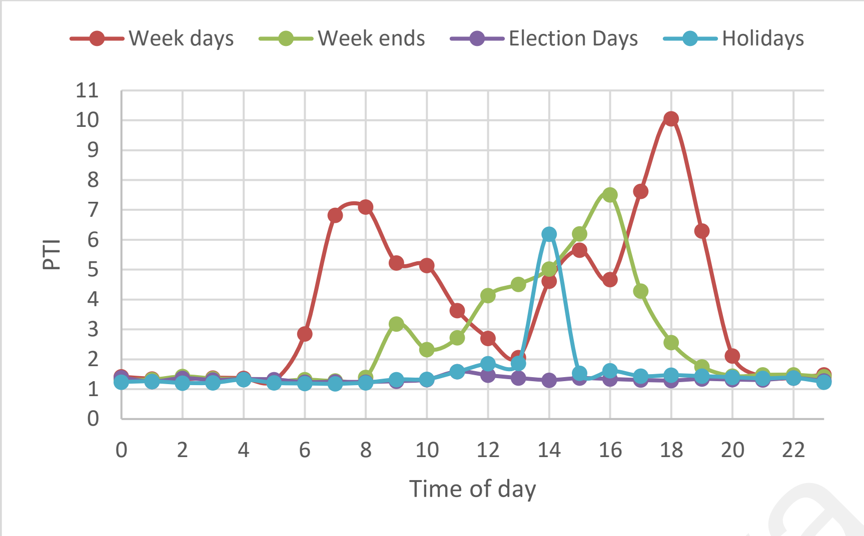
### 4.2.1 Travel Time Reliability Measures Patterns for Route A

Figure 4.11a presents the patterns of TTI for route A. On weekdays, there are morning and evening peaks. Here, TTI is about 4.8 in the morning peak and 3.2 in the evening peak. Weekends and holidays have one peak in the afternoon, and the pattern of election days is close to free-flow travel time. Figure 4.11b below shows the patterns of PTI for route A. For weekdays, PTI is about 10 in the evening peak and about 7 in the morning peak. As such, TTI is higher in the morning peak, and PTI is higher in the evening. This indicates that the traffic congestion in the morning peak period is more stable than traffic congestion in the evening peak period. The peaks in PTI patterns are similar to the peak in TTI. Figure 4.11c displays the patterns of BTI for route A. For weekdays, BTI is about 0.5 in the morning peak period and about 2 in the evening peak period. This measure confirms that the traffic congestion in the morning peak period is more stable than traffic congestion in the evening peak period. The BTI pattern of weekends shows that there is a high variation in travel time during the daytime. From 12:00 to 16:00, the values of BTI for weekends are higher than the values of BTI for weekdays. This means that from 12:00 to 16:00, the variation in travel time during weekends is higher than the variation in travel time during weekdays.

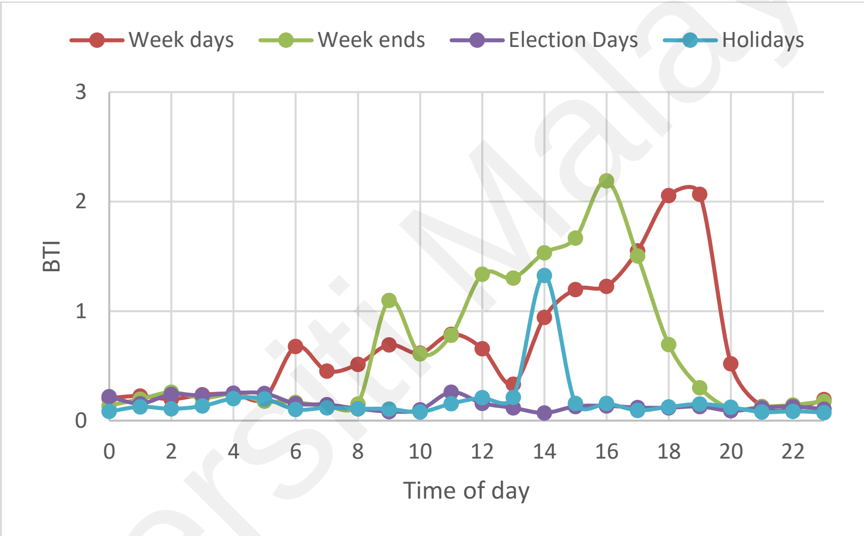
Notably, the difference between TTI and PTI is increased at the peak period compared to the off-peak period, as shown in Figure 4.12. This figure presents TTI and PTI patterns of weekdays for route A. This implies that TTV at the peak period is higher than the variability of travel time at the off-peak period.




(a)




(b)




(c)

**Figure 4.11: TTR measures for route A: (a) TTI, (b) PTI, and (c) BTI**

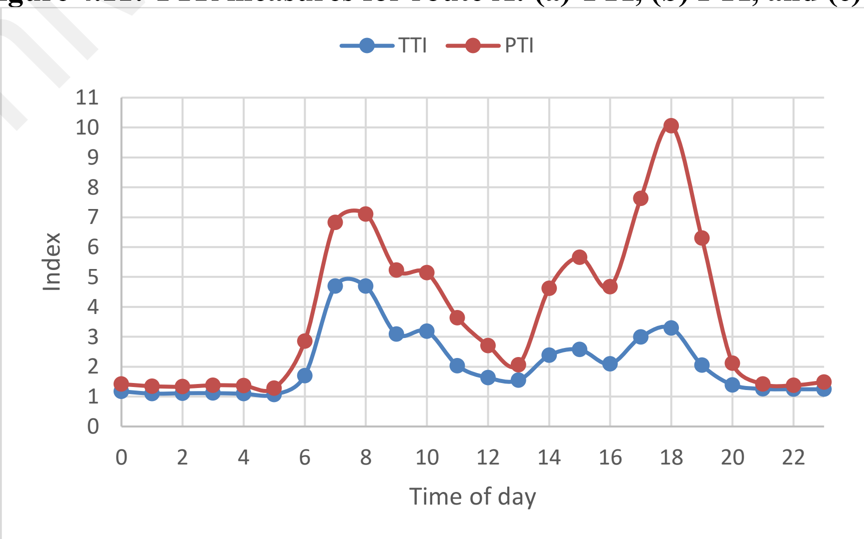
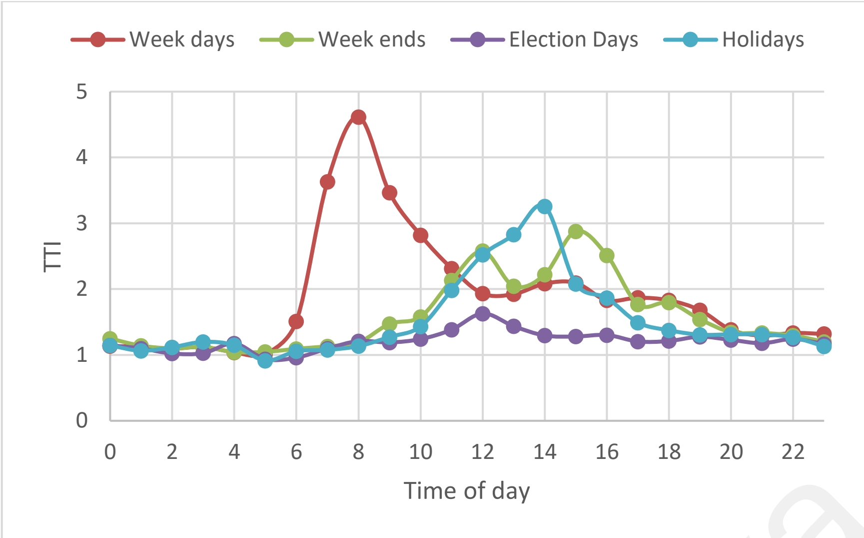



**Figure 4.12: TTI and PTI patterns of weekdays for route A**

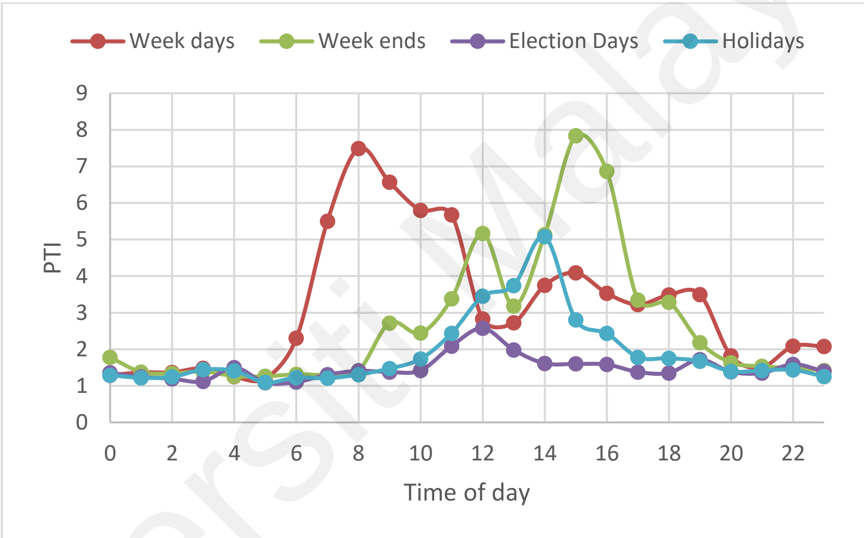
### 4.2.2 Travel Time Reliability Measures Patterns for Route B

Figure 4.13a presents the patterns of TTI for route B. The weekdays have a single peak, which is in the morning. This indicates that there is no congestion in the evening. Notably, there are two peaks for weekends, the first peak at 12:00 and the second peak at 15:00. The peak for holidays is at 14:00, and the pattern for election days is close to free-flow travel time. Figure 4.13b displays the patterns of PTI for route B. For the weekdays, the values of PTI from 10:00 to 11:00 are high. It could be due to the instability of the length of congestion hours in the morning. For the weekends, the value of PTI at the second peak is higher than the first peak indicating that the uncertainty at the second peak is higher than the uncertainty at the first peak. Figure 4.13c shows the patterns of BTI for route B. Here, the peaks times of BTI are different from the peaks times of TTI. For the weekdays, the highest value of BTI is 1.45 at 11:00, which is less than the highest value for weekends. On the other hand, the variation of travel time during election days and holidays is low.

Figure 4.14 presents the TTI and PTI patterns of weekdays for route B. PTI pattern is not smoothed as TTI pattern. TTI pattern shows that there is only one peak period, that is the morning peak. However, PTI pattern shows morning peak period and evening peak period from 14:00 to 19:00.

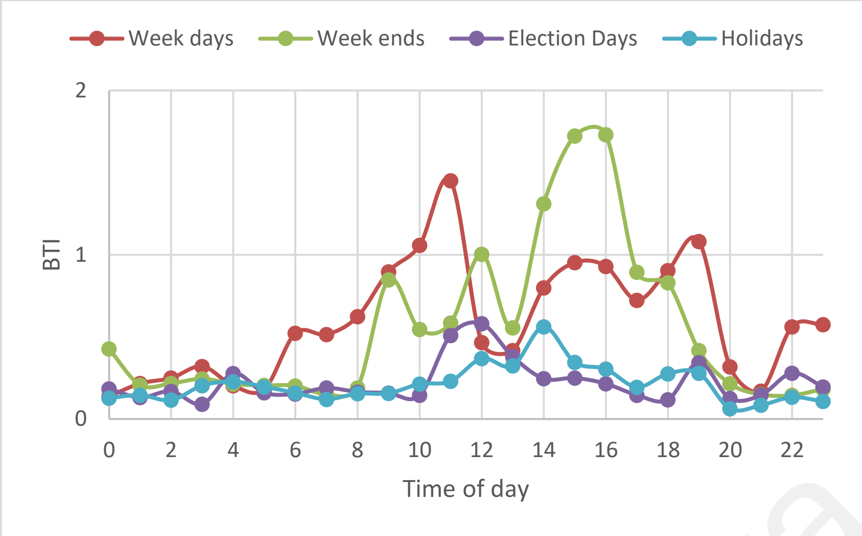



(a)




(b)

**Figure 4.13: TTR measures for route B: (a) TTI, (b) PTI, and (c) BTI**




(c)

'Figure 4.13, continued'

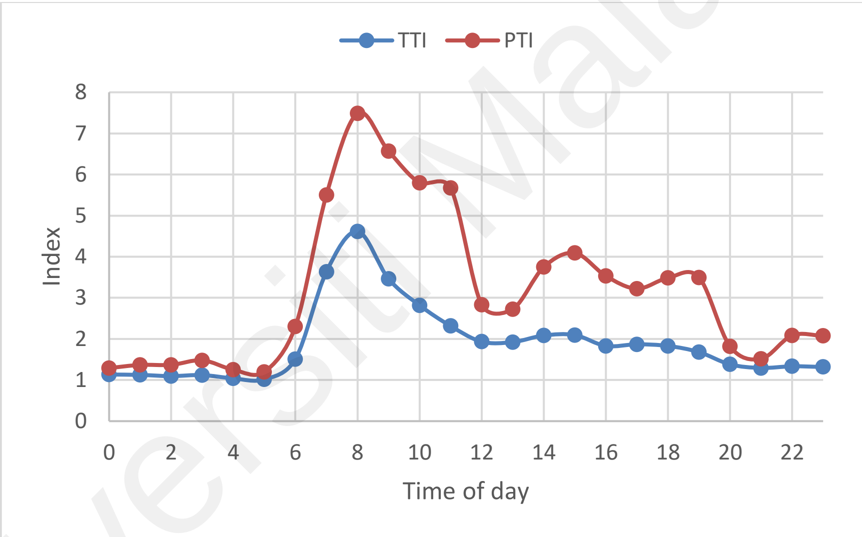
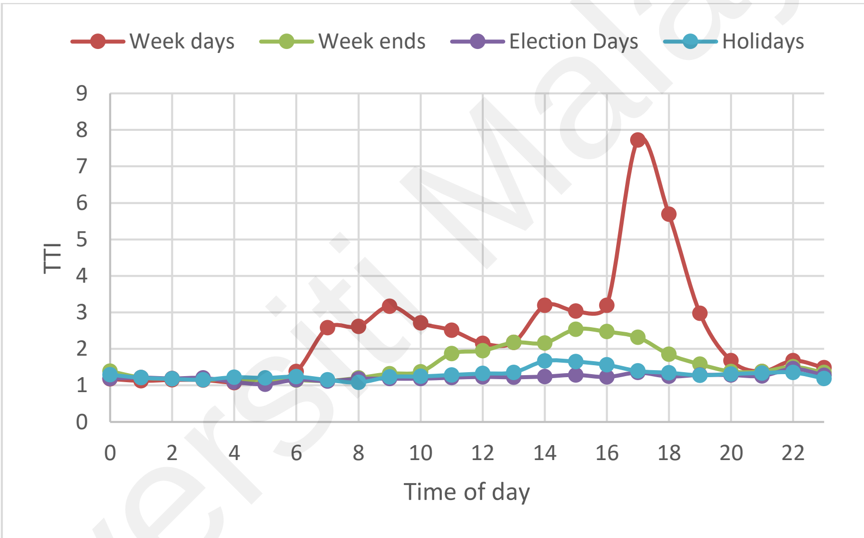



**Figure 4.14: TTI and PTI patterns of weekdays for route B**

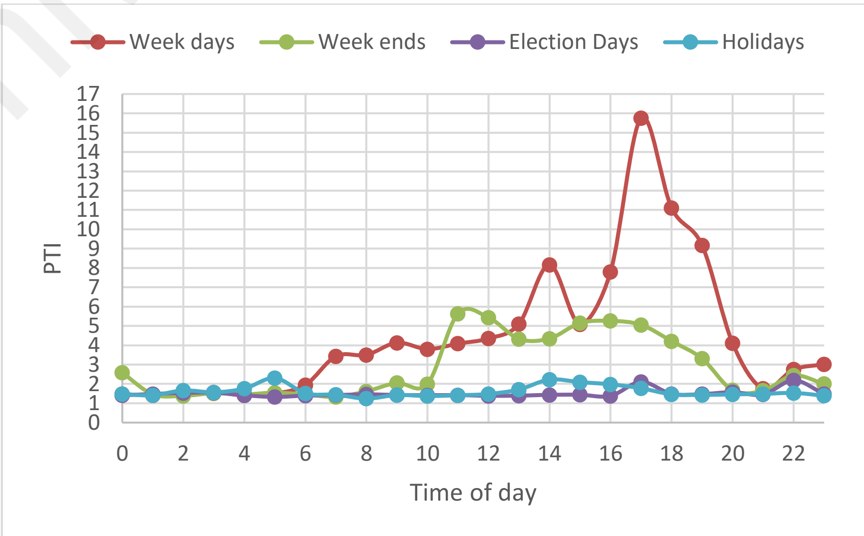
### 4.2.3 Travel Time Reliability Measures Patterns for Route C

Figure 4.15a displays the patterns of TTI for route C. For weekdays, the value of TTI at the evening peak is very high than the morning peak. This means that the congestion in the evening is more severe than in the morning. TTI is about 8 in the evening. This implies that the average travel time is eight times the free flow of travel time. It can be seen that the peak of weekdays is at 15:00. On election days and holidays, the traffic condition is near to the free-flow condition. Figure 4.15b shows PTI patterns for route C; for weekdays, the value of PTI at the evening peak is 16. This value is very high, and the value of TTI at the same peak is also very high. Thus, it implies that the level of variability in travel time and the congestion level were high at this peak. For weekends, from 11:00 to 18:00, the values of PTI are between 4 and 6. Figure 4.15c displays the patterns of BTI for route C. For weekends, the variability in the morning is low, increasing in the afternoon and the evening. For weekdays, the variability is very high in the afternoon.

Figure 4.16 presents TTI and PTI patterns of weekdays for route C. The difference between TTI and PTI is increased at the evening peak period. This implies that, for route C, TTV at the evening peak period is higher than the variability of travel time at the off-peak and morning peak.

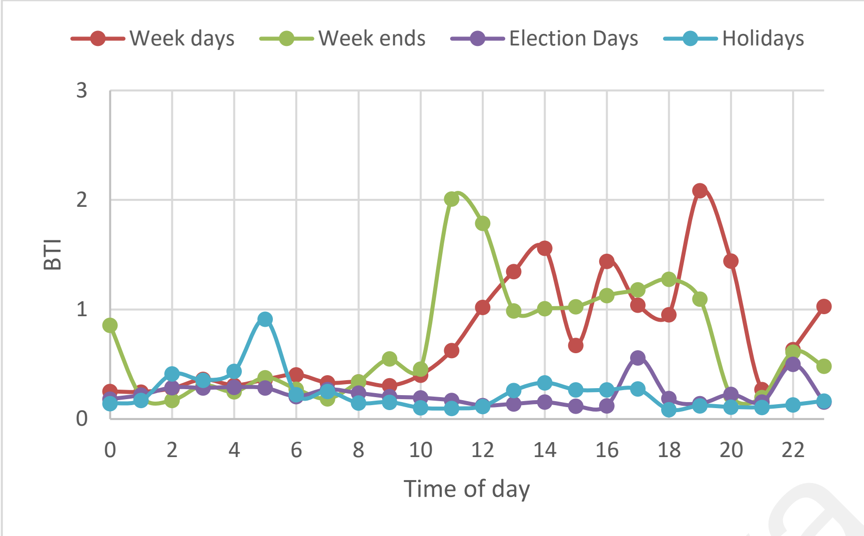



(a)




(b)

**Figure 4.15: TTR measures for route C: (a) TTI, (b) PTI, and (c) BTI**




(c)

‘Figure 4.15, continued’

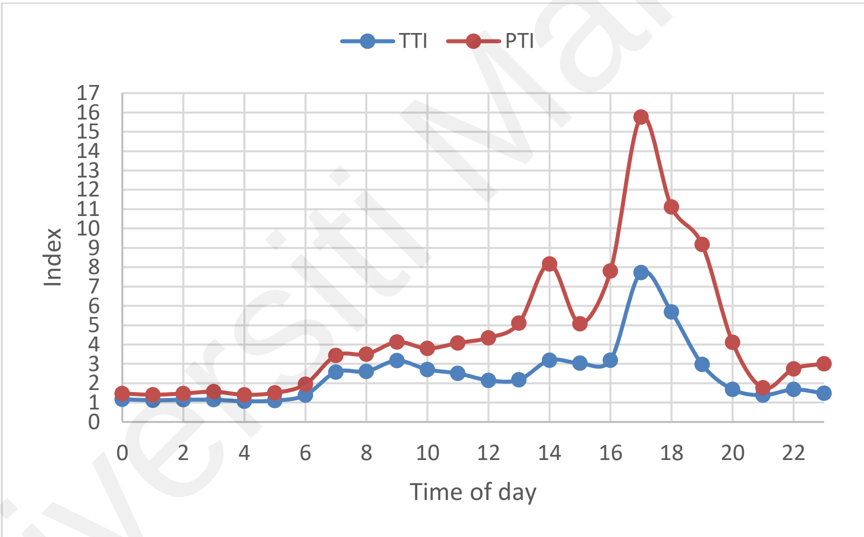



**Figure 4.16: TTI and PTI patterns of weekdays for route C**

### 4.2.4 Congestion Ranking of Days and Routes Using Travel Time Reliability Measures

In Table 4.10, TTR measures were calculated with respect to the day of week (DOW). Holidays and election days were excluded from this analysis. Average TTI, average PTI, and average BTI indicated the average values of hours. Ranking 1 means that the corresponding day has the least value of the TTR measure, which is the most reliable day compared to the other days. The most reliable day suggests the least travel time variable day. The most reliable day for routes A, B, and C is Sunday since the average TTI, average

PTI, and average BTI are the least compared to the other days. The least reliable day is different for each route. For route A, the least reliable day is Monday. Whereas for route B, the least reliable day is Tuesday, and for route C, the least reliable day is Friday. It is noticeable that all of the least reliable days are weekdays. However, weekends are more reliable compared to weekdays.

For route A, the ranking of average TTI for Monday is 6, whereas the ranking of average PTI and average BTI for Monday is 7. This shows inconsistency between TTR measures due to the different formulation of the associated equations. Each TTR measure describes a different aspect of congestion. Mainly, TTI describes typical delays, while PTI describes typical delays and unexpected delays, and BTI describes unexpected delays.

**Table 4.10: Ranking of weekdays and weekends based on TTR measures**

| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| Route A | | | | | | | |
| Average TTI | 2.21 | 2.04 | 1.87 | 1.96 | 2.32 | 1.67 | 1.22 |
| Ranking | 6 | 5 | 3 | 4 | 7 | 2 | 1 |
| Average PTI | 4.12 | 3.03 | 2.74 | 3.02 | 3.59 | 2.74 | 1.38 |
| Ranking | 7 | 5 | 3 | 4 | 6 | 2 | 1 |
| Average BTI | 0.70 | 0.45 | 0.41 | 0.46 | 0.46 | 0.48 | 0.13 |
| Ranking | 7 | 3 | 2 | 5 | 4 | 6 | 1 |
| Route B | | | | | | | |
| Average TTI | 1.82 | 2.22 | 1.86 | 1.75 | 1.82 | 1.81 | 1.34 |
| Ranking | 4 | 7 | 6 | 2 | 5 | 3 | 1 |
| Average PTI | 2.98 | 3.41 | 2.76 | 2.21 | 2.61 | 2.72 | 1.79 |
| Ranking | 6 | 7 | 5 | 2 | 3 | 4 | 1 |
| Average BTI | 0.51 | 0.49 | 0.40 | 0.23 | 0.40 | 0.40 | 0.28 |
| Ranking | 7 | 6 | 5 | 1 | 4 | 3 | 2 |
| Route C | | | | | | | |
| Average TTI | 2.37 | 2.43 | 2.40 | 2.25 | 2.75 | 1.81 | 1.35 |
| Ranking | 4 | 6 | 5 | 3 | 7 | 2 | 1 |
| Average PTI | 4.06 | 3.98 | 4.11 | 3.71 | 4.91 | 2.76 | 2.05 |
| Ranking | 5 | 4 | 6 | 3 | 7 | 2 | 1 |
| Average BTI | 0.60 | 0.54 | 0.55 | 0.56 | 0.57 | 0.47 | 0.41 |
| Ranking | 7 | 3 | 4 | 5 | 6 | 2 | 1 |

Tables 4.11, 4.12, and 4.13 present TTR measures during morning and evening peak periods for weekdays. The analysis considered the morning peak period from 7:00 to 10:00, while the evening peak period considered periods between 16:00 to 19:00. Here, route C is the most reliable route at the morning peak; however, route B is the most reliable route at the evening peak. On the other hand, it is difficult to determine which route is least reliable because the rankings of TTI, PTI, and BTI are different. For example, at the morning peak, route A is the least reliable route based on average TTI, while route B is the least reliable route based on average PTI and average BTI. This is because each TTR measure describes a different aspect of congestion. Mainly, TTI describes typical delays, while PTI describes typical delays and unexpected delays, and BTI describes unexpected delays.

**Table 4.11: Ranking of the routes based on TTI**

| | Average TTI during Morning Peak | Ranking | Average TTI during Evening Peak | Ranking |
|---|---|---|---|---|
| Route A | 3.92 | 3 | 2.61 | 2 |
| Route B | 3.63 | 2 | 1.80 | 1 |
| Route C | 2.77 | 1 | 4.89 | 3 |

**Table 4.12: Ranking of the routes based on PTI**

| | Average PTI during Morning Peak | Ranking | Average PTI during Evening Peak | Ranking |
|---|---|---|---|---|
| Route A | 6.08 | 2 | 7.16 | 2 |
| Route B | 6.34 | 3 | 3.43 | 1 |
| Route C | 3.72 | 1 | 10.96 | 3 |

**Table 4.13: Ranking of the routes based on BTI**

| | Average BTI during Morning Peak | Ranking | Average BTI during Evening Peak | Ranking |
|---|---|---|---|---|
| Route A | 0.57 | 2 | 1.73 | 3 |
| Route B | 0.77 | 3 | 0.91 | 1 |
| Route C | 0.34 | 1 | 1.38 | 2 |

# CHAPTER 5: CONCLUSION AND RECOMMENDATION

## 5.1 Summary of Research

Travel time reliability (TTR) is an important measure of traffic congestion. It is significant for road users, while TTR is important to manage and operate road networks for transportation agencies. One of the main advantages of travel time is that it describes the experience of road users. This allows road users to understand travel time better than other traffic parameters like traffic volume and density. Also, providing road users with accurate travel times will help them make more informed decisions regarding departure times.

While travel time data can be collected using various techniques, each technique has its own source of errors, generating outliers in the travel time dataset. Also, given lane-splitting situations, collecting travel time using MAC addresses needs to be handled carefully. As mentioned earlier in this study, the percentage of motorcycles in Malaysia and ASEAN is exceptionally high. Therefore, lane-splitting has a significant effect on travel time patterns. Outliers and lane-splitting observations have to be removed from travel time datasets to obtain accurate travel time measurement. In the past, many travel time filtration algorithms were utilised to filter outliers. However, there is a real need to determine which algorithm can generate the most precise results when considering actual and large datasets from lane-splitting situations. This study investigated the best algorithm for data filtration to obtain accurate data for measuring TTR. The objectives of this study were accomplished by employing an appropriate methodology.

Based on the findings of this study, the following conclusions are presented:

- Jang algorithm and TransGuide algorithm effectively filtrated the outliers and lane splitting data. The Jang algorithm is the best algorithm for routes A and C, while the TransGuide algorithm is the best algorithm for route B.

- The parameters of the Jang algorithm and TransGuide algorithm are sensitive for different days. Accordingly, this indicates that both algorithms can be used after the calibration process has been undertaken.
- It is indicated that routes C and A were less sensitive than route B. Therefore, it can be concluded that the Jang algorithm is less sensitive than the TransGuide algorithm.
- After comparing the three routes, an increase in the distance between sensors (route length) led to a decrease in detected travel time observations.
- It can be concluded that an increase in the number of observations makes the Jang algorithm the best filtration algorithm. The Jang algorithm was the best algorithm for routes A and C, having more observations than route B.
- It can be concluded that the sensitivity of the algorithms' parameters is in inverse proportion to the number of travel time observations.
- On weekdays and weekends, the three routes suffered from high variability in travel time. In addition, the peak hours are different for weekdays and weekends.
- On election days and holidays, the road network, except for one route, operates near free-flow conditions for most of the day with low variability in travel time.
- Each TTR measure describes a different aspect of congestion. Primarily, TTI describes typical delays, while PTI describes typical delays and unexpected delays, while BTI describes delays. When these measures are used for the ranking of routes, there is the probability that the results are not precisely the same for each measure since each one describes a different aspect of congestion.

## 5.2 Contribution of Research

The study's findings indicated that the travel time dataset collected by MAC addresses can be filtered out, thus obtaining accurate travel time patterns. This research will be beneficial for traffic consultant companies and authorities that depend on MAC address to collect travel time data in Malaysia and other countries that have lane-splitting situations on their roads. Even though there are many published filtering algorithms so far, the usage of the algorithms is completely depending on the characteristics of the travel time data. Specifically for this research, among the main five prominent algorithms that have been tested, it is found that two of them can produce promising results. Even though the best two algorithms can filter out the outliers and lane-splitting observations, but they required a lot of improvement to reduce the sensitivity issue.

Regarding TTR, the findings indicated that TTR measures can be calculated in Malaysia, and the study area suffered from high variability in travel time on weekdays and weekends. This will be beneficial for Malaysian traffic authorities since that show them the importance of TTR and how TTR describes the variability in traffic.

## 5.3 Research Limitations

The limitations of this study include the following:

- Due to Covid-19, the researcher could not collect recent traffic data due to Malaysia's full and partial lockdowns. Undoubtedly traffic is highly affected by these lockdowns. This was the main reason why the data in this research was archived data of the actual traffic condition of May 2018.
- The other traffic parameters, such as traffic volume and density, were not collected in the study area during May 2018. Therefore, the relationships between travel time and the other traffic parameters could not be quantified.

## 5.4 Future Research

This research addressed how to filter out travel time data on Malaysian roads, which commonly experienced lane-splitting situations and how to use filtered data to calculate

TTR measures. In order to extend this study, there are several opportunities for further research to be undertaken as follow:

- As the established algorithms are sensitive, a new insensitive outlier filtrating algorithm is recommended. This will help traffic professionals to avoid the calibration process for each day.
- The calculation of TTR measures for KL using data from many years will be beneficial to understand the variability in travel time in KL. This will also help the Malaysian authorities to develop suitable congestion mitigation strategies.
- The sources of congestion include bottlenecks, traffic incidents, weather conditions, work zones, poor signal timing, and special events. Estimating congestion by source should be further studied in the context of Malaysia to understand the extent of recurrent congestion and non-recurrent congestion. No doubt, this will be very useful in developing mitigation strategies.

# REFERENCES

Abdullah, A. S., Ramli, R., & Mohamad, F. F. (2017). Real-time and Automated Traffic Data Inventory and Monitoring System (TDIM). In S. Yusoff (Ed.), *University of Malaya Living Labs: Transforming Research into Action* (Vol. 1, pp. 113-124). Kuala Lumpur, Malaysia: Sustainability Science Research Cluster.

Aftabuzzaman, M. (2007). *Measuring traffic congestion-a critical review.* Paper presented at the 30th Australasian Transport Research Forum.

ASEAN Statistics Division. (2018). Transports Statistics. Retrieved from https://data.aseanstats.org/

Asian Development Bank. (2019). *Asian development outlook 2019 update: fostering growth and inclusion in Asia's cities*. Retrieved from https://www.adb.org/publications/asian-development-outlook-2019-update

Aupetit, S., Espié, S., & Bouaziz, S. (2015). Naturalistic study of riders' behaviour in lane-splitting situations. *Cognition, Technology & Work, 17*(2), 301-313.

Beanland, V. (2018). *Evaluation of the ACT Motorcycle Lane Filtering Trial*. Retrieved from https://www.justice.act.gov.au/sites/default/files/2019-09/Lane_Filtering_Evaluation_Report.pdf

Beanland, V., Pammer, K., Sledziowska, M., & Stone, A. (2015). *Drivers' attitudes and knowledge regarding motorcycle lane filtering practices.* Paper presented at the Proceedings of the 2015 Australasian Road Safety Conference.

Bertini, R. L. (2006). *You are the traffic jam: an examination of congestion measures.* Paper presented at the The 85th annual meeting of transportation research board.

Büchel, B., & Corman, F. (2020). Review on Statistical Modeling of Travel Time Variability for Road-Based Public Transport. *Frontiers in Built Environment, 6*.

Cambridge Systematics Inc. (1998). *NCHRP Report 399: Multimodal corridor and capacity analysis manual*. Retrieved from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_399.pdf

Cambridge Systematics Inc. (2005). *Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation*. Retrieved from https://ops.fhwa.dot.gov/congestion_report/

Cambridge Systematics Inc. (2008). *Cost-effective performance measures for travel time delay, variation, and reliability* (Vol. 618): Transportation Research Board.

Chen, P., Sun, J., & Qi, H. (2017). Estimation of delay variability at signalized intersections for urban arterial performance evaluation. *Journal of Intelligent Transportation Systems, 21*(2), 94-110.

Chen, S., Wang, W., & van Zuylen, H. (2010). A comparison of outlier detection algorithms for ITS data. *Expert Systems with Applications, 37*(2), 1169-1178.

Chen, Z., & Fan, W. (2019). Data analytics approach for travel time reliability pattern analysis and prediction. *Journal of Modern Transportation, 27*(4), 250-265.

Chen, Z., & Fan, W. D. (2020). Analyzing travel time distribution based on different travel time reliability patterns using probe vehicle data. *International Journal of Transportation Science and Technology, 9*(1), 64-75.

Clark, S. D., Grant-Muller, S., & Chen, H. (2002). Cleaning of matched license plate data. *Transportation research record, 1804*(1), 1-7.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Academic press.

Culotta, K., Fang, V., Habtemichael, F., & Pape, D. (2019). *Does Travel Time Reliability Matter?* (Report No. FHWA-HOP-19-062). Retrieved from Federal Highway Administration website: https://ops.fhwa.dot.gov/publications/fhwahop19062/fhwahop19062.pdf

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

Elefteriadou, L., & Cui, X. (2007). *A framework for defining and estimating travel time reliability*. Paper presented at the Transportation Research Board 86th Annual Meeting.

ERIA Study Team. (2010). Current status of ASEAN transport sector. *ASEAN Strategic Transport Plan 2011-2015*, 3-1.

Fan, W., & Gong, L. (2017). *Developing a systematic approach to improving bottleneck analysis in North Carolina*. Retrieved from https://connect.ncdot.gov/projects/research/RNAProjDocs/2016-10%20Final%20Report.pdf

Federal Highway Administration (FHWA). (2020). Urban Congestion Trends and Related Reports. Retrieved from https://ops.fhwa.dot.gov/perf_measurement/reliability_reports.htm

Fenelon, K. G. (2017). *The economics of road transport* (Vol. 9): Routledge.

Florida Department of Transportation. (2000). The Florida Reliability Method in Florida's Mobility Performance Measures Program. *Florida Department of Transportation*.

Florida Department of Transportation. (2011). *SIS bottleneck study (Technical Memorandum No. 2—Methodology to identify bottlenecks)*. Retrieved from https://www.fdot.gov/docs/default-source/planning/systems/programs/mspi/pdf/Executive-Summary-letter-2-15-13.pdf

Fosgerau, M., Hjorth, K., Brems, C., & Fukuda, D. (2008). *Travel time variability: Definition and valuation*. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.1202&rep=rep1&type=pdf

Google. (n.d.). Kuala Lumpur Roads. Retrieved from https://www.google.com/maps/@3.1310283,101.7046947,14.32z

Guo, J., Li, C., Qin, X., Huang, W., Wei, Y., & Cao, J. (2019). Analyzing distributions for travel time data collected using radio frequency identification technique in urban road networks. *Science China Technological Sciences, 62*(1), 106-120.

Haghani, A., Hamedi, M., Sadabadi, K. F., Young, S., & Tarnoff, P. (2010). Data collection of freeway travel time ground truth with bluetooth sensors. *Transportation Research Record, 2160*(1), 60-68.

Higatani, A., Kitazawa, T., Tanabe, J., Suga, Y., Sekhar, R., & Asakura, Y. (2009). Empirical analysis of travel time reliability measures in Hanshin expressway network. *Journal of Intelligent Transportation Systems, 13*(1), 28-38.

Hojati, A. T. (2014). *Modelling the impact of traffic incidents on travel time reliability.* (PhD Thesis). The University of Queensland, Australia. Retrieved from https://doi.org/10.14264/uql.2014.492

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

Kaparias, I., Bell, M. G., & Belzner, H. (2008). A new measure of travel time reliability for in-vehicle navigation systems. *Journal of Intelligent Transportation Systems, 12*(4), 202-211.

Kelly, T. (2016). *Lane Filtering Submission*. Retrieved from

Kitamura, Y., Hayashi, M., & Yagi, E. (2018). Traffic problems in Southeast Asia featuring the case of Cambodia's traffic accidents involving motorcycles. *IATSS research, 42*(4), 163-170.

Kittelson, Associates, & Program, S. S. H. R. (2013). *Evaluating Alternative Operations Strategies to Improve Travel Time Reliability*: Transportation Research Board.

Kurlantzick, A., & Krosner, B. (2016). *Motorcycle Lane Splitting: A Literature Review*. Retrieved from http://ridetowork.org/files/docs/MotorcycleLaneSplittingLiteratureReviewRev1.0.pdf

Lee, J. (2017). Vehicle registrations in Malaysia hit 28.2 million units. Retrieved from https://paultan.org/2017/10/03/vehicle-registrations-in-malaysia-hit-28-2-million-units/

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

Lomax, T., & Margiotta, R. (2003). *Selecting travel reliability measures*. Retrieved from https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2003-3.pdf

Lyman, K., & Bertini, R. L. (2008). Using travel time reliability measures to improve regional transportation planning and operations. *Transportation Research Record, 2046*(1), 1-10.

Margiotta, R. A., Lomax, T. J., Hallenbeck, M. E., Dowling, R. G., Skabardonis, A., Turner, S., & Systematics, C. (2013). *Analytical procedures for determining the impacts of reliability mitigation strategies* (0309129265). Retrieved from https://www.nap.edu/catalog/22806/analytical-procedures-for-determining-the-impacts-of-reliability-mitigation-strategies

Martchouk, M., Mannering, F., & Bullock, D. (2011). Analysis of freeway travel time variability using Bluetooth detection. *Journal of transportation engineering, 137*(10), 697-704.

Maulina, D., Danilasari, K. R., Nazhira, F., & Jufri, S. S. (2022). Why riders perform risky riding behavior in Jakarta: The effects of hazardous situations and gender on risk perception. *Psychological Research on Urban Society, 1*(1), 12.

Ministry of works Malaysia. (2019). *Malaysian ITS Blueprint (2019-2023)*. Retrieved from Ministry of works Malaysia website: http://www.kkr.gov.my/public/Malaysian%20ITS%20Blueprint.pdf

Moghaddam, S. S., & Hellinga, B. (2013). Evaluating the performance of algorithms for the detection of travel time outliers. *Transportation research record, 2338*(1), 67-77.

Moghaddam, S. S., & Hellinga, B. (2014). Algorithm for detecting outliers in Bluetooth data in real time. *Transportation Research Record, 2442*(1), 129-139.

Mulvihill, C., Salmon, P., Filtness, A., Lenne, M., Walker, G., Cornelissen, M., & Young, K. (2013). *Lane filtering and situation awareness in motorcyclists: An on-road proof of concept study*. Paper presented at the Proceedings of the 2013 Australasian Road Safety Research, Policing and Education Conference.

Noland, R. B., & Polak, J. W. (2002). Travel time variability: a review of theoretical and empirical issues. *Transport reviews, 22*(1), 39-54.

Ouellet, J. V. (2012). *Lane splitting on California freeways*. Paper presented at the Proceedings of the Transportation Research Board 91st Annual Meeting.

Rao, A. M., & Rao, K. R. (2012). Measuring urban traffic congestion-a review. *International Journal for Traffic & Transport Engineering, 2*(4).

Rice, T., Troszak, L., & Erhardt, T. (2015). *Motorcycle lane-splitting and safety in California*. Retrieved from https://www.carbuyer.com.sg/cb-content/uploads/2015/06/Motorcycle-Lane-Splitting-and-Safety-2015-1.pdf

Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). *2015 urban mobility scorecard*. Retrieved from https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-scorecard-2015-wappx.pdf

Southwest Research Institute. (1998). *Automatic vehicle identification model deployment initiative - system design document*. Retrieved from Bureau of Transportation Statistics, U.S. Department of Transportation website: https://rosap.ntl.bts.gov/view/dot/48800

Sperley, M., & Pietz, A. J. (2010). *Motorcycle lane-sharing: literature review*. Retrieved from https://www.oregon.gov/ODOT/Programs/ResearchDocuments/Motorcycle_Lane_Sharing.pdf

Texas Transportation Institute , & Cambridge Systematics Inc. (2005). *Travel Time Reliability: Making It There On Time, All The Time* (Report No. FHWA-HOP-06-070). Retrieved from Federal Highway Administration website: https://ops.fhwa.dot.gov/publications/tt_reliability/brochure/ttr_brochure.pdf

TomTom International BV. (n.d.). TomTom traffic index. Retrieved from https://www.tomtom.com/en_gb/traffic-index/ranking

Transportation Association of Canada. (2017). *Primer on Defining and Measuring Urban Congestion* Retrieved from Transportation Association of Canada Website: https://www.tac-atc.ca/sites/default/files/site/doc/Bookstore/defining-measuring-congestion.pdf

Tsolakis, D., Tan, F., & Makwasha, T. (2011). *Valuation of travel time reliability: a review of current practice* (1921709960). Retrieved from https://www.onlinepublications.austroads.com.au/items/AP-R391-11

Van Lint, J., & van Zuylen, H. J. (2005). Monitoring and predicting freeway travel time reliability: Using width and skew of day-to-day travel time distribution. *Transportation Research Record, 1917*(1), 54-62.

Vandervalk, A., Louch, H., Guerre, J., & Margiotta, R. (2014). *Incorporating Reliability Performance Measures Into the Transportation Planning and Programming Processes: Technical Reference* (SHRP 2 Report S2-L05-RR-3,). Retrieved from https://www.nap.edu/catalog/22594/incorporating-reliability-performance-measures-into-the-transportation-planning-and-programming-processes-technical-reference

Venables, W. N., Smith, D. M., & R Development Core Team. (2009). An introduction to R. In: Citeseer.

Wang, W.-X., Guo, R.-J., & Yu, J. (2018). Research on road traffic congestion index based on comprehensive parameters: Taking Dalian city as an example. *Advances in Mechanical Engineering, 10*(6), 1687814018781482.

Wu, Z., Wu, Z., & Rilett, L. R. (2020). Innovative nonparametric method for data outlier filtering. *Transportation research record, 2674*(10), 167-176.

Xu, Z., Jabari, S. E., & Prassas, E. (2020). Applying Finite Mixture Models to New York City Travel Times. *Journal of Transportation Engineering, Part A: Systems, 146*(5), 05020001.

Yang, S., & Wu, Y.-J. (2016). Mixture models for fitting freeway travel time distributions and measuring travel time reliability. *Transportation Research Record, 2594*(1), 95-106.