Using Bluetooth to Estimate Traffic Metrics for Traffic Management Applications

Emmanuel Gbenga Ayodele

(BSc, MSc)

Doctor of Philosophy

School of Civil Engineering and Geosciences,

Newcastle University, Newcastle upon Tyne, UK

November 2017

ABSTRACT

‘Bluetooth’ is a technology that can be integrated into Intelligent Transport Systems (ITS) to facilitate smarter and enhanced traffic monitoring and management to reduce congestion. The current research focus on Bluetooth is principally on journey time management. However, the applicability and viability of Bluetooth potential in problematic urban areas remains unknown. Besides the generic problem of unavailability of processing algorithms, there is gap in knowledge regarding the variability and errors in Bluetooth-derived metrics. These unknown errors usually cause uncertainty about the conclusions drawn from the data. Therefore, a novel Bluetooth-based vehicle detection and Traffic Flow Origin-destination Speed and Travel-time (TRAFOST) model was developed to estimate and analyse key traffic metrics. This research utilised Bluetooth data and other independently measured traffic data collected principally from three study sites in Greater Manchester, UK. The Bluetooth sensors at these locations generated vehicle detection rates (7-16%) that varied temporally and spatially, based on the comparison with flows from ATC (Automatic Traffic Counters) and SCOOT (Split Cycle Offset Optimisation Technique) detectors. Performance evaluation of the estimation showed temporal consistency and accuracy at a high level of confidence (i.e. 95%) based on criteria such as Mean Absolute Deviation (MAD) - (0.031 – 0.147), Root Mean Square Error (RMSE) - (0.041 – 0.195), Mean Absolute Percentage Error (MAPE) - (0.822 – 4.917) and Kullback-Leibler divergence (KL-D) (0.004 – 0.044). This outcome provides evidence of reliability in the results as well as justification for further investigation of Bluetooth applications in ITS. However, the resulting accuracy depends significantly on sample size, network characteristics, and traffic flow regimes. The Bluetooth approach has enabled a deeper understanding of traffic flow regimes and spatio-temporal variations within the Greater Manchester Networks than is possible using conventional traffic data such as from SCOOT. Therefore, the application of Bluetooth technology in ITS to enhance traffic management to reduce congestion is a viable proposition and is recommended.

Dedication

I dedicate this PhD thesis to the *Lord God Almighty, the Alpha and Omega, Jehovah Great I am*, for His goodness and mercies and the grace granted to complete this doctoral degree.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the people that have contributed to the success of this great work. Firstly, to my supervisors: Dr Neil Thorpe and Prof Philip Blythe for their boundless confidence in my ability throughout this remarkable journey of my life. Many thanks for your unquantifiable support.

Secondly, I would like to thank Prof Philip Moore, Prof Margaret C. Bell, and Dr Fabio Galatioto for their initial contributions to the research at different stages. And to the University authority that mediated over the challenges in the programme: Dr Stuart Edwards, Dr Bryn Jones, Dr Neil Gray, Prof Jon Mills and Ms Nicola Houghton. I want to say a big thank you for your precious time spent and the great understanding shown in handling the matter.

Thirdly, my appreciation goes to all the people that have contributed to the enhancement of my training and research development especially Mr David McGeeney for your input on the statistical aspect of my research and Drs Colin Gillespie and Graeme Hill on the R-programming. Also thanks to all the members of staff and friends from the Transport Operations Research Group (TORG).

For encouragement and moral support, I would like to express my appreciation to the entire staff of the Department of Surveying and Geoinformatics, University of Lagos. Also to all my Pastors home and abroad, MFM Newcastle Brethren, Dr Ibukun Adewale, RSST and SAC Newcastle University. Time will not allow me to mention everybody. Nevertheless, thanks to Ms Melissa Ware (The School Secretary) and Mr Graham Patterson of CeG Computing.

For the provision of data for this research, I would like to express my appreciation to: Mr. Peter Jones of Mouchel/2020Liverpool; Mr Timothy Morris, Richard Dolphin and David Atkin of Transport for Greater Manchester (TfGM); Mr Darren Butterfield formerly of TfGM; Stephen Craig of Transport Scotland and Mr Andrew Haysey of Gateshead City Council. Without your cooperation and understanding, the completion of this research might have been near-impossible.

As they say, ‘when you see a lizard dancing on the road, the drummer must be in the bush’. My coming to study for a PhD in the UK on a full-time basis would have been impossible without the support of my sponsors. Big thanks to Petroleum Technology Development Fund (PTDF) – for the award of a PhD Scholarship for 4 years; The University of Lagos – for the study leave with pay; Surveyors Council of Nigeria (SURCON) – for the additional support. The funding provided and the study leave enabled me to concentrate fully on my studies.

Before I round up, I would like to thank my dear wife and children, Adesola, Pipeloluwa and Ohunoluwa Ayodele for everything you are to me. You guys have been wonderful and your love has contributed to the progress achieved thus far! To the blessed memory of my beloved mother, Yetunde Abigail; it is a pity you had gone so soon. Finally, to God be the glory for all His wonderful works in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

Acronyms	**Definition/Meaning**
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ANPR	Automatic Number Plate Recognition
ARIMA	Autoregressive Integrated Moving Average
ATCs	Automatic Traffic Counters
BT	Bluetooth
CAGR	Compound Annual Growth Rate
CV	Coefficient of Variation
DfT	Department for Transport
DoT	Department of Transport
DSS	Decision Support Systems
EDA	Exploratory Data Analysis
FCD	Floating Car Data
FHWA	Federal Highway Administration
GB	Gigabytes
GCCM	Global Connected Car Market
GDP	Gross Domestic Product
GHz	Gigahertz
GLM	General Linear Model
GLS	General Least Squares
GMN	Greater Manchester Network
GNSS	Global Navigation Satellite Systems
HSD	Honestly Significant Difference
ILD	Inductive Loop Detector
IMTD	Independently Measured Traffic Data
ITS	Intelligent Transport Systems
Kbps (Kbit/s)	Kilobit per second
KF	Kalman Filtering technique
KL-D	Kullback-Leibler Distance (Kullback-Leibler Divergence)
LIDAR	Light Detection and Ranging
LOS	Level of Service
LTE	Long Term Evolution
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
MB	Megabytes
Mbps (Mbit/s)	Megabit per second
MHz	Megahertz
MPPM	Most Probable Predictive Model
MPV	Most Probable Value
mW	Milliwatts
NE	North-East
NFC	Near Field Communications
NW	North-West

Acronyms	Definition/Meaning
OCR	Optical Character Recognition
O-D	Origin-Destination
PAN	Personal Area Network
PCA	Principal Component Analysis
R	Statistical and programming language
RF	Radio Frequency
RFID	Radio Frequency Identification
RMSE	Root Mean Square Error
RTA	Roads and Traffic Authority of New South Wales
SatNavs	Satellite Navigation Systems
SCATS	Sydney Coordinated Adaptive Traffic System
SCOOT	Split, Cycle and Offset Optimisation Technique
SE	South-East
S/N	Serial Number
SPSS	Statistical Package for the Social Sciences
SW	South-West
TfGM	Transport for Greater Manchester
TMS	Traffic Management Systems
TRAFOST	Traffic Flow Origin-destination Speed and Travel-time estimation model
UAVs	Unmanned Aerial Vehicles
UIDAHO	University of Idaho
UMCATT	University of Maryland-Center for Advanced Transportation Technology
UTC	Urban Traffic Control
UTMC	Urban Traffic Management and Control
UWE	University of the West of England
Veh	Vehicles
VMS	Variable Message Signs
WiFi	Wireless Fidelity
WSN	Wireless Sensor Networks

Chapter 1. Introduction

1.1 Introduction and Background to the Research

Traffic congestion poses many challenges to road transportation due to the ever-growing population and increased levels of private car use around the world (Miles and Chen, 2004). Traditionally, the challenges of traffic congestion have been managed by increasing road capacity (Chowdhury and Sadek, 2003). Traffic management systems provide an improvement to road congestion through surveillance, optimisation of subsystems (such as traffic signals), and control on highways and local roads (Diebold, 1995; Miles and Chen, 2004). However, despite the traffic management systems in place and major expenditure on new road infrastructures, congestion problems continue to rise, leading to different challenges for health and the economy (Chen and Miles, 1999; Miles and Chen, 2004). In the UK, SCOOT-UTC (Split Cycle Offset Optimisation Technique – Urban Traffic Control) has been widely implemented to manage traffic (SCOOT-UTC, 2011). Traffic data collected by these systems from diverse sources are processed and managed to carry out different strategies to optimise the flow of traffic in order to reduce congestion (Hounsell *et al.*, 2009). While the traditional management systems continue to develop, they are expensive in terms of both procurement and maintenance, and SCOOT is restricted to signalised junctions (Leduc, 2008). Intelligent Transport Systems (ITS) are the integration of transportation systems with a variety of tools (such as software and communications technologies) and are widely used today for enhanced services such as efficiency and safety (Chowdhury and Sadek, 2003; Kosta *et al.*, 2011; Kindleysides, 2014). Through ITS, the traditional solutions to transportation problems can be enhanced or substituted (Chowdhury and Sadek, 2003). However, ITS are data hungry and depend on different streams of measurements to provide useful information to end users (Dalgleish and Hoose, 2009). The current technology-based systems which include Global Navigation Satellite Systems (GNSS) provide a more dynamic and comprehensive solution than is possible using traditional systems (Leduc, 2008). Currently, there is already a wide-spread use of GNSS commonly referred to as satellite navigation (SatNav) for transport applications. While the use of SatNav has received wide acceptance, it is limited by a number of factors such as satellites' geometry, and multi-path effect, particularly in urban areas (Misra and Enge, 2006; Trimble, 2007; Meng *et al.*, 2008). Exploring the potential of other technological options such as wireless communications provides further opportunities to enhance the existing systems using low-cost sensors. Wireless communication technologies such as 'Bluetooth' provide the prospect of gathering key traffic information (such as O-D matrix that has been expensive and difficult to acquire in the past) anywhere across the networks. Wireless technology is cost-effective, accurate, pervasive, easy to deploy and maintain, and low-power (Srinivasan, 2011). Blythe (2006) highlighted the importance of wireless technology in the areas of road user charging, pervasive environmental monitoring, congestion control and fleet management. This is echoed in the Foresight Project on Intelligent Infrastructure Systems (IIS) that sought to address how science and technology could bring intelligence into the infrastructure over the next 50 years (Foresight, 2006). Therefore, exploring the potential benefits of Bluetooth for traffic metrics estimation could contribute to achieving this aim.

Consequently, this research explores the use of Bluetooth sensors for vehicular traffic detection and metrics estimation in urban areas within the context of the applicability of the Bluetooth approach to enhancing traffic management systems in order to reduce congestion. The assessment was conducted through the analysis of data collected from a total of three UK study areas (Birtley, Liverpool and Greater Manchester). Data from Bluetooth sensors and other Independently Measured Traffic Data (IMTD) were used. A novel Bluetooth-based processing and analysis technique (TRAFOST), developed and implemented in this research has helped to accomplish this investigation. Methods of analysis include both quantitative and exploratory data analysis such as time series, correlation, and Principal Component Analysis (PCA).

1.2 Context of the Research

Addressing the transportation problems of congestion from traffic monitoring and management perspectives requires a more complete and efficient solution than is currently available. Bluetooth is considered a technology with the possibility to enhance current systems. Many devices such as mobile phones, laptops and in-vehicle gadgets have Bluetooth embedded in them to exchange data or communicate with one another over short distances without requiring physical contact (Bluetooth, 2012). Literature demonstrates that research into the use of Bluetooth in ITS is still a novel area and thus requires further understanding of the approach, usability and limitations to fully exploit its potential. This research considers these gaps and the applicability of the Bluetooth approach to vehicular traffic sensing and metrics estimation to enhance management systems in order to reduce traffic congestion.

Traffic congestion can be defined in terms of demand-capacity and delay in travel time. Based on demand-capacity, it is the delay caused by one vehicle to others, or when demand exceeds capacity (Thomson, 1978). In terms of travel time, it is the delay in excess of what normally occurs under light or free-flow travel conditions (Lomax *et al.*, 1997). On the other hand, a delay is the amount of extra time spent in congestion over the ideal or free-flow travel time (Camsys and Texas Transportation Institute, 2004). Traffic congestion is generally classified as either recurrent or non-recurrent (Chowdhury and Sadek, 2003). It is usually caused by factors such as bottlenecks (the largest source of congestion and traffic incidents) including crashes and vehicle breakdowns that cause about 25% of congestion problems (DoT, 2012). Congestion problems affect the economy with a detrimental effect on human health and the environment, and thus there have been calls for improvement in road network efficiency (WHO, 2005; Ayodele *et al.*, 2014). Greater Manchester (the main study area in this research) which is the second largest conurbation in the UK after London, is not an exception. Economically, the annual congestion costs in the UK could rise to as much as £22 billion by 2025 (Scullion, 2011). The Eddington report outlines the challenges of congestion, climate change and sustainability (Eddington, 2006). Meanwhile, an efficient transport system has a ripple effect on the economy, such as saving around £2.5 billion for a 5% reduction in travel time for all business travel on the roads – some 0.2 per cent of the UK GDP (Eddington, 2006). Schrank *et al.* (2012) gave a brief summary of the problem of congestion highlighting the massive waste in time, fuel and money. In 2011 in the US alone, fuel wastage was estimated to be 2.88 billion gallons; total delay as 5.52 billion hours, while delay per commuter was 38 hours, making a total cost of $121.2 billion per year (Schrank *et al.*, 2012). The 2009 report shows that the cost is more than $80 billion a year in the US (Srinivasan, 2011). The reality is that an ever increasing population worldwide calls for increased awareness of the importance of cutting-edge research to achieve a smarter and more sustainable environment (Conservation, 2012; Darey, 2012). Therefore, establishing a balance in the road networks through operational efficiency becomes imperative to meet the present challenges. By embracing innovative solutions, this balance in traffic management can be achieved without necessarily investing in building new infrastructures. Bluetooth possesses the potential to enhance the existing systems to reduce congestion and time spent in traffic.

Bluetooth can be used to gather information concerning traffic patterns and to raise awareness of suitable alternatives such as park and ride, or car sharing options. The traffic information collected can be displayed through Variable Message Signs (VMS) or relayed through in-vehicle (IV) technologies to improve efficiency. However, to derive the maximum benefits from the technology, policy changes must be at the heart of future transport guidelines. This change in policy will include support for low-cost technological options. Thereby leading to maintaining a balance in the development of techniques that manage travel demand more efficiently, while upholding an individual's right to freedom of movement (Thorpe, 2005). Weigelt *et al.* (1973, page 2) also stated that 'the need to attain a balance between city planning and its traffic is the key problem of the urban transportation policy during a transition phase from a city without any private automobiles to a city with a high degree of automobile saturation'. For transport engineers and planners, the obvious problem is that availability of timely and accurate data remains a fundamental challenge in attaining this balance.

Interestingly, the availability of Bluetooth technology is increasing not only in electronic devices and mobile phones but also in vehicles. Exploring the potential of the technology in this way to enhance road network efficiency might constitute a cutting-edge solution to traffic congestion problems. Meanwhile, before economic or environmental benefits can be realised fully, understanding the patterns of movement and regularity of trips made by people is essential. The availability of such information will allow traffic management systems to respond better to inform network users of alternative routes and modes. This information has been difficult and expensive to acquire in the past, however Bluetooth offers the opportunity to address this challenge at little cost. Consequently, this research also seeks to investigate the use of Bluetooth data to enhance reliable reconstruction of traffic patterns and trends, which have hitherto been under-investigated. This contribution to knowledge further implies a step towards realising smarter future transport systems, leading to a more sustainable, efficient, and clean road network.

Using Bluetooth technology, two technological challenges are addressed. The first is the monitoring of movements (or passage) of traffic across specific known points in the network. The second is the management of the computational intensity of processing large volumes of data (tens of gigabytes) arising from day-to-day onsite monitoring of the passage of traffic to derive useful information. Bluetooth sensors developed by TDC Systems were used to meet the first requirement, while an appropriate model was developed to address the second challenge. Consequently, there is a need for research to gain a fundamental understanding of these two components (deployment of Bluetooth for traffic detection and the processing and analysis of the acquired data). To this end, an appropriate Bluetooth-based model termed TRAFOST

(Traffic Flow Origin-destination Speed and Travel-time) was developed in this research to process, data mine, and estimate traffic metrics to explore potential applications in traffic management.

By harnessing the opportunities offered by this technology in this way, potentially Bluetooth may take over some of the functionalities of the traditional and more expensive monitoring systems such as the ANPR (Automatic Number Plate Recognition) cameras and inductive loop detectors. The motivation to demonstrate the value for money of using a low-cost Bluetooth sensor started in 2011. Peter Jones led the request from Mouchel/2020Liverpool on this project (Jones, 2011). While qualitative assessment and a literature review suggest that this is a possibility, the need for improved knowledge of statistically reliable results is required to justify the viability of the proposition. Hence, the motivation for this research is to improve the efficiency of the current systems to enhance traffic management using low-cost sensors. This can be achieved by exploring the reliability of the high resolution and timely data provided by Bluetooth to derive traffic metrics such as O-D matrix, link-flow, travel time and speed. Despite the recent rise in publications on the use of Bluetooth for traffic monitoring and other related applications, it is still in a state of continuous evolution. This evolution makes research into potential applications of Bluetooth in ITS an area of enormous potential.

Implementing Bluetooth to improve traffic management has some limitations that include the privacy issue, low vehicle counts (i.e. inability to measure the actual traffic flow), and difficulty in differentiating between modes during congestion. However, it is argued that the enormous potential possessed by the technology far outweighs its limitations particularly in the context of low-cost decision support systems (DSS) for traffic management. In this research due process was followed to ensure respect for the privacy rights of people in compliance with Data Protection Acts (Data Protection Commissioner, 2003). This process includes obtaining ethical approval from Newcastle University.

Also, encrypted data were used in this research to avoid associating any captured device to a particular owner or vehicle. Therefore, this research is neither for surveillance nor aimed at identifying or tracking any particular individual or vehicle. Rather, it seeks answer to the reliability and sufficiency of the accuracy of Bluetooth data to estimate traffic metrics for traffic management applications to reduce congestion. The next section considers the research problems and challenges.

1.3 Research Problems and Challenges

From the literature review presented extensively in Chapter 2, it is evident that there remains a lot to be done regarding Bluetooth applications in traffic management. Besides, the heterogenous sources (vehicles and other modes of transport) of Bluetooth data collection as well as the possibility for duplicate records, there is the generic problem of unavailability of algorithms showing systematic analysis procedure for traffic metrics estimation. Also, the fact that Bluetooth usage is increasing and its estimate is a sample of the total vehicular traffic means a need for a continued study to correctly determine the detection rate required for calibration. In addition, the need for a periodic calibration to ensure reliable detection rate also constitutes a challenge on the use of Bluetooth data for traffic management applications. Overall, the results of the current research on the use of Bluetooth for traffic monitoring and management, which is principally in the area of travel time analysis show that there is the need for further studies (Araghi *et al*., 2015; Barceló *et al*., 2013; Bhaskar *et al*., 2014). These problems need to be addressed to optimally exploit the potential of Bluetooth for traffic management. The next section considers the aim and objectives of the research following the research problems and challenges identified.

1.4 Aim and Objectives of the Research

This research aim is to investigate the reliability and the sufficiency of the accuracy of Bluetooth data to estimate traffic metrics for traffic management applications to reduce congestion.

The specific objectives to achieve this aim are:

i. To carry out a critical review of literature on the application of Bluetooth technology in traffic monitoring and management, and to consider other technological options for road traffic monitoring;

ii. To develop a Bluetooth-based data processing procedure (a model) to derive link-flow, travel time, speed and origin-destination matrix;

iii. To carry out data collection in selected study sites consisting of Liverpool, Birtley and Manchester, and apply the model on a short-term basis to investigate the potential of Bluetooth-derived traffic metrics;

iv. To examine the performance of the model (TRAFOST) developed in Objective ii and the consistency of Bluetooth-derived traffic metrics on a long-term basis, for accuracy and reliability through validation against diverse independent measures of traffic and statistical modelling;

v. To analyse the variability in Bluetooth-derived traffic metrics to enable concrete deductions and sound inference based on the analysis of year 2013 data from the Greater Manchester Network (GMN); and

vi. To interpret the results and make deductions from the research findings in a wider context of applicability and viability in traffic management, and make recommendations for Bluetooth traffic monitoring and metrics estimation.

1.5 Contents of the Main Chapters

This thesis is organised into eight main chapters as follows.

- Chapter 2 critically explores and reviews the available literature relating to the application of Bluetooth traffic sensing and metrics estimation in ITS with the view to enhancing traffic management systems. The review includes the applications of Bluetooth to derive important traffic metrics, a description of other technological options, and policy issues that include privacy, safety and pollution. The literature review highlights a number of key issues with the Bluetooth approach to traffic metrics estimation and application in traffic management. These issues relate to methodology; reliability and validity of the data based on a comparative analysis with independent measurements as against simulation; variability and errors arising from the data over-time, particularly in the problematic urban areas; the growth and detection rates of Bluetooth; and the wider knowledge of the viability of the Bluetooth approach in traffic management.

- Chapter 3 presents the research methodology which includes a novel Bluetooth-based estimation and analysis procedure (TRAFOST), used in this research. TRAFOST was developed to ensure automation, reproducibility and transferability in the Bluetooth approach to traffic metrics estimation. The discussion in this section includes primarily the research design, methods of Bluetooth data cleansing, and the estimation and validation methods of the traffic metrics. The research design describes the research objectives, methods of accomplishment and the expected results. The data cleaning section considers consistency, reliability, representativeness, multiple detection, and outliers. The traffic metrics estimation and validation methods conclude the discussion of this chapter.

- Chapter 4 describes the Bluetooth data collection and preliminary investigation over the three pilot study areas (Liverpool, Birtley and

Manchester – consisting of Wigan, Stockport and Trafford) considered in this research. The Liverpool pilot study presents primarily the results of data quality assessment. Based on the methodology developed and described in Chapter 3, the Birtley pilot study presents the results of the evaluation of Bluetooth data at a micro scale to understand performance and limitation of Bluetooth. The short-term Manchester pilot study builds on the Birtley and Liverpool pilot studies to establish transferability in exploring the potential of Bluetooth.

- Chapter 5 builds on the preliminary investigation of the study sites to establish two key things. Firstly, the assessment of the reliability of TRAFOST. Secondly, the assessment of the validity and reliability of the results obtained in the long-term study by employing different validation techniques to ensure the maintenance of the concept of fit for purpose. Through this understanding, the practicality of both the Bluetooth data and TRAFOST developed in this research is established.

- Chapter 6 presents the detailed description of the variability that may affect any conclusion drawn on Bluetooth-derived traffic metrics. Different temporal dimensions were considered in this exercise such as measurement over hours, days and months to explore temporal consistency. This chapter presents Bluetooth data collected over a period of one year (2013) within the GMN study site which were processed and analysed for this purpose. The computation of detection rates was through the comparisons of Bluetooth and IMT-derived flows collected over the same period in the study locations.

- Chapters 7 and 8 present the results and interpretation of the Bluetooth-estimated traffic metrics in the wider context to understand the added value obtainable from the use of the technology for traffic monitoring and management purposes. Primarily, these two chapters explore the interpretation and application of four different Bluetooth-derived metrics (link-flow, travel time, speed, and O-D matrix) in traffic management to enhance intelligent decisions.

- Chapter 9 summarises the main outcomes of the research, and the implication of the ideas developed in this research in a wider context. This includes the limitation in the traffic estimation model (TRAFOST) and the resulting generalisation of the research findings based on the results validation. The variability assessment further removes any bias on the conclusions drawn from the data. The results interpretation and application to traffic management contribute to understanding policy implications that include privacy and safety of the road users, and environmental pollution. The chapter closes with recommendations for future research.

Chapter 2. Critical Review of Literature on Bluetooth Traffic Monitoring and Applications in ITS

2.1 Introduction

This chapter presents a critical review of literature on Bluetooth technology as a novel traffic monitoring sensor for ITS (Intelligent Transport System) applications. Traffic monitoring is the process of collecting data that describes the use and performance of the road network (FHWA, 2013). The traffic data collected are used in a variety of ways to support traffic operations such as design, planning, analysis, and performance evaluation. However, a major drawback to some of the current data collection solutions, such as the inductive loops, is the requirement for significant capital investment, government commitment at several levels, as well as the support and backing of the public (Srinivasan, 2011). Bluetooth is a low-cost technology with the potential to address the current limitations by way of complimentary solutions and high value for money to address the problems of congestion. For example, data collected from across the roads using Bluetooth could be used to increase network intelligence, and to derive strategies for traffic management. However, such data need to be timely and reliable. A review of the literature identified research gaps regarding the reliability of Bluetooth data in traffic management, and this problem highlights the current research challenges. Therefore, this chapter covers the description of known methods for collection of traffic data, and a critique of the new method (Bluetooth approach).

Section 2.2 describes existing road traffic sensors, which include the data requirements. Section 2.3 presents a critical review of Bluetooth technology in contrast with other wireless technologies such as ZigBee and WiFi. Section 2.4 discusses estimation methods for analysing traffic sensor data. Section 2.5 presents the work done worldwide using Bluetooth for traffic sensing to define further specific research gaps before drawing conclusions in Section 2.6.

2.2 Sensors for Traffic Data Collection

2.2.1 Setting the data requirements for traffic management

The development of ITS requires high quality traffic information in real-time (Leduc, 2008). The real-time information collected by the road sensors are used in adaptive traffic management systems such as SCOOT for the management of road networks. Traditionally, three key measurements are used to monitor traffic operations on freeways (FHWA, 2013). They are volume, speed, and occupancy (the percentage of time a road section is occupied by a vehicle, and can be a surrogate for density) (FHWA, 2013). Other useful parameters for traffic management are; flow, travel times, O-D matrix, location, queue length, etc. Therefore, state-of-the-art traffic-sensing solutions should be able to provide archived information such as commute times and congestion patterns to help urban planners and traffic engineers make informed decisions in vital areas such as: where to improve road capacity, where and when to encourage car-pooling and where to enhance and increase the use of public transportation (Srinivasan, 2011). In this research, the key data requirements for traffic management considered are; flow, travel time, speed and O-D matrix. Yatskiv *et al.* (2013) highlighted the importance of these metrics in model construction, validation and calibration. As described in Table 2.1, other important criteria considered to ensure a holistic evaluation include sustainability (both in terms of acquisition and maintenance costs), sample size, and reliability. These assessment criteria provide a platform to compare the estimate of traffic metrics from Bluetooth with the existing methods to understand its strengths and limitations. The subsequent sections describe the methods, while more detailed information such as the operational principles is contained in the Traffic Monitoring Guide (FHWA, 2008).

Evaluation criteria	Description
Required traffic metrics (Flow, travel time, speed, and O-D matrix)	Flow: This is the rate at which vehicles pass a given point on the roadway and is stated as vehicles per hour. Flow is termed as traffic volume for specified time periods other than an hour, e.g. 15 minutes
	Travel time: This is the average of the total time including control delay spent by vehicles traversing a road segment measured in seconds or minutes
	Speed: The average speed of a traffic stream obtained from the length of a road segment divided by the average travel time is measured in kilometers (or miles) per hour (km/h)
	O-D matrix: This is achieved by applying the concept of flow estimation to an area-wide network
Sustainability (acquisition and maintenance costs)	Acquisition cost: This refers to the direct cost of acquiring a system or traffic sensor
	Maintenance cost: This refers to the costs incurred to keep an item in good and working condition
Transferability	This refers to how far traffic sensors can be conveyed or transferred to other contexts or settings
Availability	This is the ability to provide the required function and performance within a specified range
Accuracy and Reliability	Accuracy means how well a measured value agrees with the true value
	Reliability refers to the degree of consistency or repeatability of a measure
Sample size	This refers to the proportion of the detected vehicles compared to the actual population
Coverage	This refers to the maximum distance at which the approaching target or vehicle can be detected
Privacy issue	This relates to determining whether the technology can impinge on people's rights or not
Safety issue	This refers to the understanding of how well the technology can improve or affect road safety

Table 2.1: Description of the data requirements and the evaluation criteria for traffic management

Table 2.2 and Table 2.3 present a summary of the traffic sensors aligned with the evaluation criteria. While some of the current technologies are highly accurate in providing traffic information, they are not sustainable especially from a cost perspective, as they are either too expensive to acquire or maintain. However, emerging technology such as Bluetooth could be used to overcome the problem of cost without compromising accuracy. The subsequent sections describe the relevant sensors.

Traffic sensors	Required data for traffic management			
	Flow	Travel time	Speed	O-D matrix
Inductive loop detectors	Yes	No (estimation by algorithm)	Yes (with two consecutive loops)	No
Pneumatics tubes	Yes	No (not accurate)	Yes (with two detectors but not accurate)	No (not accurate)
Radar	Yes	No (except derived from local speed using specific algorithm)	Yes	No (except with special algorithm, and requiring high number of sensors)
Video detection	Yes	No (estimation by algorithm)	Yes	Not used
ANPR	Yes	Yes (by tracking number plates)	Yes	Yes
GNSS-based FCD	Yes	Yes	Yes	Yes
GSM-based FCD	Yes	Yes	Yes	Yes
Signpost system	Yes (if enough vehicles are equipped)	Yes	Yes	Yes (entry-exit)

Table 2.2: Comparison of relevant traffic sensors based on data requirements

Relevant sources: (Schmidt *et al.*, 2005; BITRE, 2014)

Traffic sensors	Other evaluation criteria									
	Capital cost	Operation and maintenance cost	Transferrability	Availability	Accuracy	Reliability	Sample size	Range of detection/ coverage	Privacy issue	Safety issue
Inductive loop detectors	Expensive	Expensive	No	Few	High	High	High	Short range and unidirectional	No	Installation and maintenance require lane closure
Pneumatics tubes	Moderate cost	Low cost	Yes	Few	High	High	High	Short range and multiple lanes	No	Relatively safe
Radar	Expensive	Expensive	Yes	Few	High	High	High	Short range and multiple lanes	No	Safe (if non-intrusive method)
Video detection	Low -high cost	Low cost	Yes	Few	Medium - high	High	High	Short range and multiple lanes	Low	Lane closure when camera is mounted over roadway
ANPR	Expensive	Expensive	Yes	Few	High	High	High	Short range and unidirectional	High	Safe
GNSS-based FCD	Expensive	Moderate	Yes	Ubiquitous	High	High	Low	Long range and unidirectional	High	Safe
GSM-based FCD	Low cost	Low cost	No	Moderate	Low	Moderate - high	Low	Medium range and unidirectional	High	Safe
Signpost system	Expensive	Expensive	Yes	Few	High	High	Low	Short range and multiple lanes	No	Safe

Table 2.3: Comparison of traffic sensors based on other relevant requirements

2.2.2 Inductive loop detectors

An inductive loop detector (ILD) is an electromagnetic communication or detection system of insulated wire embedded in the road surface, and consists of three main parts (a loop, loop extension cable, and a detector) (FHWA, 2013; Windmill, 2016). The loop utilises the principle that an electrical current is induced when a magnetic field is introduced near an electrical conductor (Windmill, 2016). For traffic monitoring, the vehicle acts as the magnetic field and the ILD as the electrical conductor, while a device at the roadside records the signals generated (Windmill, 2016). An increase in the oscillator frequency due to a change in the inductance of the loop makes vehicle detection possible (FHWA, 2013). During installation, the smallest detail matters to ensure accurate vehicle detection. Inductive loop detectors can accurately classify vehicles by type and detect speeds, but they also have significant drawbacks such as the cost of procurement (Leduc, 2008; Srinivasan, 2011). However, reducing traffic congestion and its attendant costs is one of the main goals of transport policy makers (Wang *et al.*, 2009). Besides being expensive, maintenance and installation work on the road often leads to traffic disruption (Srinivasan, 2011). Furthermore, since the speed of vehicles is calculated from the time taken to traverse the loops and congestion determined by the speed below a certain threshold, this means that there is a possibility of error in the estimation and inference (Chen and Miles, 1999; Morris, 2014). For example, vehicles close together may be interpreted as one long vehicle. Another limitation of these sensors is the inability for vehicle re-identification or the determination of O-D movements. Nevertheless, the 99% detection rate obtained from ILD shows that it is highly accurate for traffic data collection (Klein, 1997).

2.2.3 Pneumatic tubes

Pneumatic tubes placed on road lanes produce changes in pressure when vehicles pass over them (Leduc, 2008). One end of the data logger connects to the rubber tube(s) stretched across the road (Windmill, 2016). The air pressure in the tube activates the data logger as wheels pass over the tube and it records the time of event (Windmill, 2016). Pneumatic tubes can be stretched across several lanes of traffic. The data logger determines the direction of vehicles through the identification of the first crossing of the tubes (Alam, 2014). Consequently, simultaneous crossing may lead to erroneous estimation. Also, two close cars can be misinterpreted as one multi-axle vehicle (McGowen and Sanderson, 2011; Windmill, 2016). However, marketers claim an accuracy level of 99% but research based on 15-minute counts suggest approximately 10% absolute error (McGowen and Sanderson, 2011; Windmill, 2016). Typical traffic data captured by pneumatic tubes are vehicle speed, count and classification. It is relatively inexpensive and easy to install, and is useful for short-term traffic surveys of one or two weeks. This technology is easily damaged and unable to provide important traffic information such as travel time and O-D matrix.

2.2.4 Radar

A microwave radar system makes use of radar technology to detect moving vehicles. The detected transmitted energy scattered by the vehicle rear is converted to traffic information by the sensor, or in conjunction with the roadside controller (Klein *et al.*, 2006). Radar detectors emit frequencies ranging from 100MHz to 100GHz (FHWA, 2013). Vehicle speeds are calculated based on the Doppler principle with a decreasing frequency when the vehicle is moving away from the radar and an increasing frequency when the vehicle is approaching (Klein *et al.*, 2006). This technology can provide measurements of lane occupancy, vehicle count, speed, and vehicle classification (Klein *et al.*, 2006). It is limited in the provision of travel time and O-D information. The intrusive method of this technology can replace the loop detector with improved accuracy of 7.1% and 4.8% in length and speed respectively (Kim *et al.*, 2001). The non-intrusive method can achieve 8% accuracy over ILD both in length and speed (Kim *et al.*, 2001).

2.2.5 Video detection

Video detection makes use of video technology and systems that automatically analyse the video pictures as vehicles are passing through the detection zone (Windmill, 2016). The system consists of one or more cameras, a microprocessor-based computer for digitising and analysing imagery, and software for interpreting the images and converting them to traffic data (Klein, 1997; Klein *et al.*, 2006). A single camera can cover different directions of multiple lanes at once. Also, real-time modifications can be made to the detection zones from the control centre to accommodate the prevailing traffic conditions (Windmill, 2016). This vehicle counting technology has several advantages such as low procurement and maintenance costs and it can cover both directions and turning movements at once compared to loop detector and ANPR methods (Klein, 1997; Klein *et al.*, 2006; Windmill, 2016). Real-time data uploading and verification is simplified, with a detection accuracy similar to that of manual counting (Windmill, 2016). Video technology is important for ramp and lane management to enable informed decisions regarding any changes in traffic conditions to be made (Klein *et al.*, 2006; FHWA, 2013). This technology can replace inductive loops, and can classify vehicles by length, report vehicle presence, volume, lane occupancy, and speed for each vehicle class or lane (Klein *et al.*, 2006; FHWA, 2013). However, this technology is limited in the provision of O-D information as is the case with technologies such as radar, as vehicle re-identification across the network is not possible.

2.2.6 Automatic Number Plate Recognition (ANPR) camera

ANPR is a method used to detect and automatically read number plates using instruments such as the optical character recognition method (OCR) (Blythe, 2006; National Policing Improvement Agency, 2012). The OCR software can take repeated snapshots once a vehicle is near the camera, thus increasing the confidence level of detection (Blythe, 2006; Augustin and Poppe, 2012). ANPR is one of the methods most commonly used to calculate travel time and detect incidents on roads (Augustin and Poppe, 2012). Without any human intervention, ANPR systems can process video images of number plates taken by a roadside camera and convert this into the appropriate alphabetic/numeric characters (Blythe, 2006). This capability makes ANPR suitable for real-time application such as in crime detection and congestion management. While ANPR can provide O-D information, and has found applications in road-user charging, improved road safety, etc. the drawback is that ANPR cameras as with inductive loop detectors are expensive both in terms of procurement of the image processing software and installation (Biora *et al.*, 2012). Blythe (2006) also noted that while there is an improvement in the camera technology to provide clear images under certain conditions, some unresolved issues remain. These include, differences in shape and size of the letters, similarities in letters, blurring, poor lighting, masking of the number plate due to snow/fog/dirt and unrecognised number plate types such as number plates from foreign countries (Blythe, 2006; Augustin and Poppe, 2012).

2.2.7 Global Navigation Satellite System (GNSS)

Global Navigation Satellite Systems (GNSS) which include GPS and Galileo have varying applications in ITS (Misra and Enge, 2006). The operational principle comprises the interaction between space, ground, and user segments to provide accurate positions anywhere in the world using satellites as reference points (Trimble, 2007). The current technology-based systems which include GNSS provide a more dynamic and comprehensive solution than is possible using traditional systems (Hounsell *et al.*, 2009). For example, the emergence of satellite navigation systems has brought a fundamental change. Real-time tracking, route guidance, telematics, and location-based services are now carried out using GNSS solutions (Hounsell *et al.*, 2009). Booth (2005) highlighted the advantages of the GNSS technology to include route guidance in cars and buses, and warnings when approaching speed cameras. The technology has found applications in the estimation of travel time and speed (Quiroga, 2000; Mintsis *et al.*, 2004; Sadoun and Al-Bayari, 2007). However, this solution is sometimes limited in urban settings where positioning solutions are highly dependent on the availability and geometric distribution of satellites that in turn are sometimes constrained by tall buildings (Boneberg *et al.*, 2011). However, despite the ubiquitous nature of the GNSS technology, the probe vehicle method does not give the actual traffic count, but a proportion of the total traffic. Although, this is also the case with other technological-based options which include the GSM-based FCD (floating car data). Also, the presence of multi-path errors due to tall buildings in urban areas can degrade the quality of service of a GPS-based probe vehicle (Trimble, 2007).

2.2.8 Global System for Mobile Communications (GSM)

The Global System for Mobile Communications (GSM) or cellular-based FCD makes use of the radio modem (AVL, 2004). Mobile positioning is the technology used by telecommunication companies to approximate the location of a mobile phone and/or its user (Bar-Gera, 2007; Pourabdollah *et al.*, 2010). Advanced services with high mast station distribution such as in urban areas can attain about 50m accuracy and less in areas with masts widely spaced (AVL, 2004). GSM-based FCD is cost-effective but has a lower accuracy compared to the GPS-based and traditional systems. However, the sample size of 4% - 5% probe vehicles was estimated to be a reasonable range to estimate reliable travel times in metropolitan areas (Cheu *et al.*, 2002; Li and McDonald, 2007; Leduc, 2008). This technology relies on the positioning of the vehicles incorporating mobile phones to act as sensors over the network to capture traffic data (Leduc, 2008). This causes inaccuracy in the estimation of the O-D data while cost may be an issue in the implementation of the accurate GPS-based solution (Biora *et al.*, 2012). Nevertheless, the system has also shown potential in providing traffic data for system augmentation. For example, the integration of the system with other tracking and location-aware systems, such as GPS, offers a considerable advantage. This was demonstrated for the effective management of ambulance services (Derekenaris *et al.*, 2001). Bar-Gera (2007) used the technology to derive traffic speed, and travel time on a 14km freeway and found that it compared well with dual magnetic loop detectors, thereby showing promise for different practical applications. FHWA (2013) presents a detailed summary of the probe vehicle systems.

2.2.9 The signpost system

This is the technology used to track and locate vehicles along fixed routes. It utilises the proximity technique through RFID (Radio Frequency Identification) to determine location and allow for vehicle progress monitoring. Vehicle positions are determined as they pass through the sensor locations. The determination of travel time is through the information collected at two consecutive stations. While the most prevalent AVL (Automatic Vehicle Location) system for bus transit is GPS-based, several systems that provide real-time arrival/departure information are signpost-based including King County Metro in Seattle and Transport for London Buses (DoT, 2007). These systems are viable alternatives inside tunnels or other conveyances where there are blockages by terrain to GPS signals (AVL, 2004). Systems using RFID technology with appropriate algorithms and databases have found application in multi-vehicle, multi-lane, and multi-road junction areas to provide an efficient time management scheme (Al-Khateeb *et al.*, 2008). However, in terms of accuracy, the GPS-based system is better. The technology is capital intensive both in terms of investment and staff resources to develop, implement, and operate (FHWA, 2006a). While the technology can provide travel time and other information related to the vehicle and passengers, it is limited in coverage and non-representative given that its operation is mainly in buses (FHWA, 2006a; FHWA, 2013).

2.3 Bluetooth Technology

2.3.1 Description of Bluetooth

Bluetooth is a short-range, low-power wireless technology used for data communication and monitoring applications in the ITS domain (Andersson and Karlsson, 2000; Friesen and McLeod, 2014). Bluetooth operates in the globally unlicensed Industrial, Scientific and Medical (ISM) 2.4 GHz short-range radio frequency band (Information Age, 2001; Tabona, 2005). Bluetooth is named after the Danish King Harald Blåtand I (Kardach, 2008; BBC, 2011; Bluetooth, 2011). It was developed by the Bluetooth Special Interest Group (SIG), formed in May 1998 with five founding members: Ericsson, Nokia, Intel, IBM and Toshiba (Vainio, 2000). Bluetooth SIG has member companies in the areas of telecommunication, computing, networking, and consumer electronics (Tabona, 2005). In 2016, the SIG membership had reached 30,000 (Bluetooth SIG, 2016). The SIG oversees the development of the specification, manages the qualification program, and protects the trademarks (Information Age, 2001). Bluetooth technology has found application in several sectors including the automotive industry. Currently this technology is one of the emerging technologies with the potential to provide relevant traffic data. Within the road transport network, Bluetooth-enabled devices such as mobile phones, headsets, SatNavs and portable electronic devices are found onboard vehicles or carried by cyclists/pedestrians. The development in this sector is attributed to features such as hands-free calling, and security remote controls for locking and unlocking vehicles (Persistent Market Research, 2017). Bluetooth technology is considered the only proven wireless choice for both developers and consumers worldwide (Business Wire, 2010). Therefore, its potential to estimate traffic metrics for traffic management applications is considered in this research.

2.3.2 Bluetooth functionality

The installation of a Bluetooth sensor is usually on lamp posts at a height of about 3m above the ground (McDonald, 2013). The basic information collected by a typical Bluetooth sensor (Appendix 1) includes the date and time stamp of the occurrence of a Bluetooth device and the identification code referred to as MAC (Media Access Control) address. The MAC address is a combination of unique hexadecimal alphanumeric characters. The first six characters are allocated to the manufacturers (e.g. Nokia) and the device type (e.g. phone); while the last six characters relate to the wireless device as defined by the service provider (Barceló *et al.*, 2010). Appendix 2 presents example data. Bluetooth detected addresses are time-stamped with the possibility of re-identification at different locations. This principle is used to estimate travel time by computing the differences in the time stamps between different locations.

Figure 2.1 presents the concept of Bluetooth traffic sensing and metrics estimation.

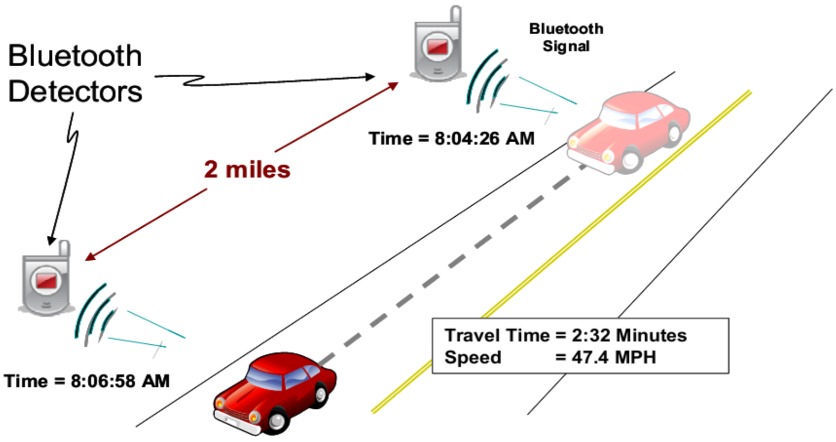


Figure 2.1: The concept of Bluetooth traffic sensing and metrics estimation (Source: UMCATT, 2008)

The dwell time in a location is computed from the timestamps at entry (first detection) and exit (last detection) of a device. For a device to be detected, Bluetooth security provides authorisation before pairing through device scanning or inquiry. However, Bluetooth traffic sensors do not require authorisation as they only detect and register the MAC addresses and the time stamps of the detected devices. A device can be detected up to 99% possibility at 5s inquiry (discovery) time (Kasten and Langheinrich, 2001; Peterson *et al.*, 2006). However, an inquiry time of 10.24s is recommended for the maximum detection of devices (Chakraborty *et al.*, 2010). Due to the inquiry time, not all of the devices are detected before leaving the zone. Experiments showed a capture rate of 80% (Gurczik *et al.*, 2012). While the detection rate is 2-50% of all vehicles depending on the study location and the type of antennae used (Young *et al.*, 2013). The detection rate is the ratio of the matched-pairs of Bluetooth detectable vehicles captured at two consecutive sensor locations compared to the actual link flow (Young *et al.*, 2013). However, the obtainable accuracy is dependent upon the installation environment as the formation of the RF (radio frequency) field of the antenna can be affected by trees, buildings, guard railings, and lighting columns (McDonald, 2013). Therefore, optimum performance of the system requires field inspection to select a site.

2.3.3 Relevant technical details of a Bluetooth system

Bluetooth operates on radio frequency (RF) technology (Bluetooth, 2011). The Bluetooth standard is IEEE802.15. The transmission of the Bluetooth signal to and from the cell phone consumes just 1 milliwatt of power which makes the battery of the phone virtually unaffected (Howstuffworks, 2011). Bluetooth has a typical range of 1- 100m depending on the class (Bluetooth SIG, 2001). Essentially, there are three categories of Bluetooth. Class 1 is used primarily in industry with a range of 100m; Class 2 is commonly found in mobile devices with a range of 10m; and Class 3 have a range of up to 1m and are mostly used in computer mouse and keyboard technologies (Bhaskar and Chung, 2013; Bluetooth SIG, 2015). For traffic sensing, there are two classes of Bluetooth antennae (the omni-directional and uni-directional) (TDC, 2011; Bhaskar and Chung, 2013). The omni-directional antennae can detect devices within the range of detection in every direction, while the uni-directional antennae can detect devices in one direction only, but with capability to detect devices travelling in opposing directions. The TDC uni-directional antenna used in this research has a detection range of 93m. This range also defines the maximum spatial error (positional error) that can be introduced to the data because the exact time of detection of a device and the location within the detection zone is unknown (Bhaskar and Chung, 2013). Therefore, the error in time can be up to the 10.24s standard inquiry time. The direction of travel of a device is determined by performing MAC address matching to determine the location of the first detection. That is, a device is said to be travelling in the direction ‘A to B’ if the time of detection at point A is before that of point B and *vice-versa*. This principle is used to carry out directional distribution of traffic.

2.3.4 Bluetooth capabilities and challenges

The growth of Bluetooth shows that billions of devices are expected to be enabled in the future (Gomez *et al.*, 2012). In the ITS domain, it is pertinent to note that Bluetooth potential for traffic monitoring started around 2010 (Friesen and McLeod, 2014). Currently, Bluetooth is potentially considered a viable technology in understanding traffic characteristics in both urban roads and motorways (Barceló *et al.*, 2010; Muhammed and Egemalm, 2012). For example, Bluetooth has shown the potential for O-D estimation to address the current challenges using existing technologies (Abrahamsson, 1998). If the opportunities offered by this technology are well-harnessed, Bluetooth systems may take over some of the functionalities of the traditional and more expensive monitoring systems.

Using Bluetooth data, a wide variety of error sources could impact greatly on the accuracy of the estimated traffic metrics that include travel time and O-D matrix if not properly handled (Araghi *et al.*, 2015; Bhaskar and Chung, 2013; Cragg, 2013). These error sources include; duplicates (more than one valid record for a device) in the data, especially during periods of congestion, error in MAC addresses leading to unrealistic speed estimation, a pedestrian or vehicle with multiple devices, road junctions with traffic lights and pedestrian crossings, business locations and car park areas near a Bluetooth station. This shows that Bluetooth traffic estimation in congested urban networks is more problematic than on the free flow motorway, and corroborates the research of Moghaddam and Hellinga (2013). For example, multiple devices may be counted as many vehicles during congestion leading to overestimation of the traffic volume. Bhaskar and Chung (2013) illustrated the effect of the entry and exit times of devices at the detection zones on the estimated metrics. The errors introduced are more pronounced on short links compared to the long links of motorways due to the aforementioned factors.

Generally, there are uncertainties regarding the carriers of Bluetooth enabled devices upon which assumptions are made for optimum results. For example, a device identified as a mobile phone may be carried in a vehicle or by a pedestrian. Also during congestion, it may be difficult to differentiate between the modes of transportation. With an increase in Bluetooth usage and new automobiles incorporating Bluetooth devices, periodic calibration of the detection rate will be required to obtain the actual flow estimation. However, the challenge is not only in the calibration but also in determining the frequency of the calibration for a continuous accurate estimation.

2.3.5 Bluetooth growth rate and market penetration in different sectors

Bluetooth, primarily designed for wireless connection of devices has found application in automotive, computing, networking and electronic devices such as speakers. Since the early 2000s, there has been an increasing penetration in the market for Bluetooth products, largely in mobile phones (Gray, 2007). In the automotive market, Bluetooth penetration started with vehicles beginning from 2003 models through the availability of features such as hands-free calling (In-Stat/MDR, 2002). In 2012, the Bluetooth SIG adopted the GNSS Profile version 1.0 to enable the sharing of positional data through a Bluetooth connection (Handheld, 2012). This adoption means that more Bluetooth devices can be detected thereby increasing the sample size and reliability of the data. The recent development in connected cars is also increasing the market penetration with a projection of connectivity in every car by 2025 (SBD, 2012). In 2014, Bluetooth had reached 90% penetration in all mobile phones (Bluetooth SIG, 2016). Currently, the Global Connected Car Market (GCCM) is poised to have CAGR (Compound Annual Growth Rate) of around 11.7% over the next decade with revenue of approximately $81.7 Billion by 2025 (PRNewswire, 2016). Table 2.4 presents the summary of the Bluetooth growth rate and market penetration in the relevant key sectors.

Bluetooth Market	Shipment/Market size	Projection	CAGR (%) Period	Current Sector
Bluetooth Beacon	Shipment was 80,000 in 2015	88.29 million by 2020	307.2 2015 - 2020	Retail, indoor navigation, telematics
Bluetooth Speakers	Shipment was 88.2 million in 2015	$7 Billion in revenue by 2019	38.73 2014 - 2019	Automotive, consummer electronics
Bluetooth Smart and Smart Ready	Greater than 2.5 billion shipment in 2013 with market size of $3.27 Billion	$5.57- 8.4 Billion in revenue by 2020	6.24 - 29 2014 - 2020	Automotive, consummer electronics, wearable electronics, retail, IoT, security, proximity sensing

Table 2.4: Bluetooth growth rate and market penetration in different sectors

2.3.6 Bluetooth vis-a-vis ZigBee and WiFi technologies

Wireless Sensor Networks (WSN), in particular the fusion of fixed and mobile networks, have been identified as having a significant role in delivering future intelligence to the transport sector for a safer, sustainable and robust future transport system, based on its ability to collect, process, disseminate and use data in a fully connected environment (Selvarajah *et al.*, 2012). Table 2.5 presents a comparison of the main features of ZigBee, Bluetooth and WiFi. WiFi is a technology based on the IEEE 802.11 standards, while ZigBee is an IEEE 802.15.4-based specification designed for small scale projects that require wireless connection (ZigBee, 2014). While these technologies offer a comparative advantage in terms of network range, Bluetooth is limited in bandwidth compared to WiFi (12Mbps against 54Mbps), but much better than ZigBee (250kbps). However, Bluetooth has a major advantage in the area of power consumption over WiFi (medium against high). WiFi is mostly used for internet connection with the advantage that it can connect many devices compared to Bluetooth; however, WiFi may become slow when many devices are connected (Bluetooth SIG, 2015). Like the cellular phone-based, WiFi is used for wider area networking but has lower accuracy compared to Bluetooth (Friesen and McLeod, 2014). Although ZigBee is designed to address the unique needs of low-cost, low-power wireless sensor, it has been used mainly for the interconnection of vehicles and infrastructure (Selvarajah *et al.*, 2008; ZigBee, 2014). Bluetooth remains the most widely-used wireless technology for in-vehicle communication due to its proven features (Kinney, 2003; Selvarajah *et al.*, 2008). For example, speakers and radio systems of new vehicles now incorporate Bluetooth. The development in the automotive market, has shown that despite some limitations these technologies have the potential to help deliver an integrated transport system; this includes application in connected vehicles. The major benefit of easy synchronisation and device connectivity gives Bluetooth an edge in the choice of wireless technology for traffic monitoring purposes. This further justifies the adoption of the Bluetooth method in this research.

Standard	**ZigBee 802.15.4**	**Bluetooth 802.15.1**	**WiFi 802.11g**
Automotive application	Inter-vehicle and vehicle to infrastructure communication	In-vehicle communication and device connectivity	Inter-vehicle and vehicle to infrastructure communication
Network range	Up to 100m	Up to 100m	Up to 100m
Bandwidth	250Kbps	12Mbps	54Mbps
Frequency	2.4GHz	2.4GHz	2.4GHz
Advantages	Low power; many devices; low overhead	Dominating PAN (Personal Area Network); easy synchronisation	Dominating PAN; widely available
Disadvantages	Low bandwidth	Consumes medium power. (Power output ranges between 1mW to 100mW)	Consumes high power

Table 2.5: Comparison of the relevant features of ZigBee, Bluetooth and WiFi (Modified from Selvarajah *et al*., 2008)

2.3.7 Bluetooth and Near Field Communications (NFC) technology

NFC is one of the more recent market entries with emphasis on low power and personal communication (Friesen and McLeod, 2014). NFC has its roots in radio-frequency identification (RFID) and is primarily used for devices of close proximity (4cm) without the need to set up a connection (Triggs, 2013; Faulkner, 2015; NFC, 2016). Table 2.6 presents the summary of NFC contrasted with Bluetooth. Although, NFC is much more power-efficient with faster connectivity, it is limited in range (less than 20cm), and transfer rate 424Kbps (APC, 2011; Triggs, 2013). Currently, the use of NFC is more business-focused. From a transport perspective, NFC has found application in seat adjustment and unlocking of cars; parking aid, ticketing, and for obtaining information on schedules and delays (NFC, 2016). However, given a 10cm range, NFC is not considered feasible for traffic management. This is also the case with Third Generation (3G) and Fourth Generation (4G) technologies which include Long Term Evolution (LTE) – the only true 4G (Rouse, 2014). However, they could be used to enhance traffic data collection. A recent application is the reporting of car data using LTE (Salvo *et al.*, 2016). The next section considers the estimation methods of analysing traffic sensor data.

NFC	Bluetooth
Much lower power consumption	Higher power consumption
Shorter range of about 10cm	Longer range up to 10m or more
Slower in data transmission (424kbits/s)	Faster in data transmission (2.1Mbits/s or 1Mbit/s for BLE
Faster connectivity (less than one-tenth of a second)	Slower connectivity but BLE can match the speed of NFC

Table 2.6: Summary of NFC/Bluetooth comparison

2.4 Estimation Methods of Analysing Data from Traffic Sensors

2.4.1 Current estimation methods

Traffic estimation refers to the calculation of metrics such as travel times based on known quantities up to the current point in time; while prediction forecasts traffic metrics up to a defined time in the future (van Lint *et al.*, 2005). Previous literature demonstrates that different estimation methods have been used in the past to analyse traffic data. The state of the art measurement for traffic estimation uses sensors such as loop detectors or traffic cameras, while a manually conducted survey is the state of the art methodology for recording origin-destination trips (FHWA, 2006b; Aslam *et al.*, 2012). There is also the use of the factoring method, identified as probably the simplest estimation method and most used worldwide (Leduc, 2008). This method consists of permanent traffic sites that are classified based on similarities in seasonal variability and traffic characteristics. Critical issues with this method include obtaining the optimal number of groups and assigning short counts to the seasonal factor groups (Leduc, 2008). This method has a low accuracy, and the short-term survey may not be representative. While traffic surveys and video surveillance methods can provide traffic information such as flow and speed, they have numerous drawbacks that include high cost of data collection and image processing (Abedi *et al.*, 2014).

The moving observer, floating car or probe vehicles, and historical data (cumulative curve) are three categories of estimation techniques identified by Maerivoet and Moor (2008). The moving observer technique involves a vehicle driven in both directions of a traffic flow, each time recording important information such as the number of oncoming vehicles, vehicles overtaken, and the time taken to complete the two trips (Krishnamoorthy, 2008; Maerivoet and Moor, 2008). Flow rate is calculated for the known average speed of the moving vehicle, road length and trip time (Mulligan and Nicholson, 2002). This method is economical according to the required accuracy. Beside the measurement of speed, travel time and flow, vehicle classification as well as other information such as location and causes of delay can be obtained. The disadvantages are that the method requires many moving observer runs to obtain accurate flow estimates (Mulligan and Nicholson, 2002; Krishnamoorthy, 2008). It is also sensitive to interconnecting traffic from side streets, and is limited in gathering O-D information (Mulligan and Nicholson, 2002). The floating cars or probe vehicles are comparable to the moving observer method with the difference of being equipped with GPS and GSM/GPRS devices for position determination and transmission of information. Probe vehicle data can provide accurate measurements of current traffic speeds and travel times (Bachmann *et al.*, 2013). These methods were described previously in Sections 2.2.7 and 2.2.8. The GPS-based method has been combined with GIS (geographic information systems) for urban traffic flow analysis (Rewadkar and Dixit, 2013). Using the historical data method, travel time is measured as the distance along the horizontal time axis. Based on this, the travel times over a period of many weeks, months, or even years can be analysed (Maerivoet and Moor, 2008). However, the evolution in ITS demands more timely information and a combination of the availability of modern, low-cost computing and communications technology. The availability of real-time traffic data will enhance rapid response to any anomalies by a way of re-routing to reduce congestion and the associated impacts (FHWA, 2013). Bluetooth as a direct method can be used in this regard to provide traffic information anytime and anywhere within the road network.

2.4.2 Emerging estimation methods

One of the emerging estimation methods includes the use of satellites and unmanned aerial vehicles (UAVs) (Fricker and Kumapley, 2002). The satellites and UAVs approach is primarily used to understand both temporal and spatial variability in traffic flow at any instant. However, the high cost of acquiring high resolution images and the processing software is a major disadvantage (Fricker *et al.*, 2002). Other limiting factors include weather, flight height, danger to aircraft, and privacy issues. While motion detection algorithms can detect each distinct moving vehicle, the algorithms are difficult to solve (Lee and Bovik, 2009). The optical flow estimation algorithms from traffic videos are considered as a better alternative, although they pose the problem of efficiency and computational complexity (Lee and Bovik, 2009). Other advances such as LIDAR (Light Detection and Ranging) approach and drone cameras are also emerging for the estimation of flow. Generally, there is a problem of incomplete datasets, and mostly the inability to estimate O-D matrix. The problem of incomplete datasets is usually addressed using estimation (predictive) and analytic methods such as log-linear model, linear regression, and neural network (Leduc, 2008).

2.4.3 Predictive and analytic methods

Incomplete datasets resulting from a number of factors such as equipment failure or scarce resources often lead to the requirement for data prediction. Csikos *et al.* (2015) classified prediction methods into two classes (the classical prediction methods and data driven methods). The classical prediction methods (model-based estimation) utilised statistical methods such as Bayesian network models, historical average, ARIMA (autoregressive integrated moving average), regressions, and Kalman filter theory. Forecast is based on analysis of historical time series data. Typical application includes the analysis of traffic flow using particle filtering (Polson and Sokolov, 2015). Particle filtering allows for posterior estimation of the most recent state with low computational complexity and the possibility for frequent updating compared to Kalman filtering. Generally, the classical approach is limited in an urban environment where the traffic conditions change rapidly (Csikos *et al.*, 2015). The data driven methods (machine learning) offer self-learning pattern recognition methods such as ANN (artificial neural network), fuzzy-rule based logics, k-mean clustering, and expectation maximisation based algorithms. This approach has the advantage of estimating and capturing the linkage of very complex traffic flows even under rapidly changing conditions. In particular, ANN algorithm was used to predict traffic speed in urban traffic networks (Csikos *et al.*, 2015). van Lint *et al.*, (2005) noted that ANN for travel time estimation is only suitable for freeway or urban arterial networks. Generally, the data driven methods are sensitive to the quality of the training data. However, this can be partly addressed by principal component analysis (PCA) to handle the missing input data. Another way to improve the accuracy is the combination of fuzzy logic and ANN as applied by (Gastaldi *et al.*, 2014). PCA has been used to analyse flow data, and is another method to overcome reliance on the knowledge of data distribution. PCA was used to measure variability in urban traffic flow to address the issue of both temporal and spatial correlation in time series data (Tsekeris and Stathopoulos, 2006). With the emergence of Bluetooth in traffic sensing, little reliance can now be placed on historical datasets and prediction. The Bluetooth approach provides platforms for the estimation of essential traffic data such as the area-wide O-D matrix in a cost-effective way to overcome the challenges posed by using traditional methods. Table 2.7 presents the summary of relevant predictive and analytics methods with “yes” signifying metrics where they are commonly applied.

Predictive / Analytical methods	Traffic metrics: Flow	Travel time	Speed	O-D matrix	Advantages	Disadvantages
Log-linear model	Yes	Yes		Yes	Suitable for count data; flexible; can readily estimate odd ratios	Not suitable for serial correlation in time series data; applicable to datasets with positive observations
Linear and multiple regression	Yes			Yes	Flexible; allows for interaction between variables	Not applicable to non-linear models; assumption of normality of errors
ARIMA	Yes	Yes	Yes		Can detect anomaly in data; good for short-term prediction	Assumption of normality of errors; memory intensive
Kalman filtering	Yes	Yes	Yes	Yes	Can estimate variables of diverse nature	Memory intensive; not applicable to non-linear models
Particle filtering	Yes		Yes	Yes	Low computational complexity; frequent updating is possible	Limited under rapidly changing traffic conditions
Historical average	Yes		Yes	Yes	Offers direct and quick solution	Reliance on historical data; data formats may require standardisation
Bayesian network models	Yes	Yes	Yes	Yes	Can improve linear regression accuracy	Requires independence between input characteristics; memory intensive
Generalised Least Squares	Yes			Yes	Allows the combination of traffic survey and count data and can be updated in short time; no assumption of distribution	Sensitive to non-negativity in datasets
Principal Component Analysis (PCA)	Yes			Yes	Can handle missing data; no reliance on data distribution; can account for temporal and spatial correlation	Reliance on orthogonal transformation of the original variables; it is not scale-invariant; the variables must be correlated
k-Nearest Neighbour (kNN)	Yes	Yes	Yes	Yes	Suitable for varied parameters such as delay and dwell times	Uses distances between attributes; memory intensive; consumes power
Artificial Neural Network	Yes	Yes	Yes	Yes	Can handle non-linearity in data; good accuracy with short-term prediction	Slow convergence; sensitive to the quality of the training data
Fuzzy logic	Yes			Yes	Can handle missing data	Sensitive to the quality of the training data
Spatial interaction model	Yes			Yes	Can handle missing data	Sensitive to the quality of the training data; few scholarly guides

Table 2.7: Summary of relevant methods of predicting and analysing traffic data

2.5 The Use of Bluetooth in Traffic Sensing

2.5.1 Bluetooth traffic sensing

This section presents a review of literature on the use of Bluetooth for traffic sensing. Consideration has been given to the areas adjudged to be the most relevant to this research to explore the gaps in knowledge. The increased awareness of the negative impacts of traffic congestion and the need for better transport through technology has led to a significant rise in the management of road traffic in recent years (White, 1989; Nellore and Hancke, 2016). Bluetooth is one of the emerging technologies for traffic sensing and ITS applications, which has been explored by authors such as Barceló *et al.* (2010), UMCATT (2008), Bhaskar and Chung (2013), and Araghi *et al.* (2015). This technology could also form an important part in the concept of “Big data”. Big data are gathered from different sources and formats that include mobile devices and the web (Troester, 2012). However, based on the available information gathered in this research, the current published studies on Bluetooth traffic sensing were carried out outside the UK. From the accessible publications, a significant gap identified was the absence of a comprehensive investigation of Bluetooth data for various traffic management applications, and the added benefits both in the short and long-term. Therefore, exploring the gaps in Bluetooth traffic sensing research to gain a better knowledge of the traffic metrics estimation capability is considered essential to support the delivery of a better optimised road network than is currently obtainable.

Bluetooth traffic sensing on rural freeways has shown great potential (Click and Lloyd, 2012). In urban freeway and arterial roads, Bluetooth has been studied for different purposes, such as the estimation of travel times (Wason *et al.*, 2008). Bluetooth traffic sensing has also found application in travel time prediction over congested periods in signalised urban arterial roads, as well as to understand delays in travel time in highway work zones (Haseman *et al.*, 2010; Quayle *et al.*, 2010); Khoei *et al.* (2013). More recent applications of Bluetooth include monitoring and tracking purposes (Stange *et al.*, 2011).

Researchers have also studied different antennae types (directional and omni-directional antennae) to understand performance and found that omni-directional antennae have a larger detection zone than the directional types (Malinovskiy *et al.*, 2010). On the other hand, Vo *et al.* (2012) and Click and Lloyd (2012) recommended using more than one sensor in a site to increase the data quality. Route choice analysis is another promising application for Bluetooth traffic sensing (Hainen *et al.*, 2011). However, results from these studies showed that noise in the data can cause signficant variance in the estimated metrics. Environmental factors such as weather could also impact upon the results. For example, Martchouk *et al.* (2011) showed that Bluetooth traffic sensing on a freeway segment under varying weather conditions (normal and abnormal) can present a significant difference in the computed mean and standard deviation of travel times. Therefore, these factors must be properly handled to obtain accurate and reliable estimations. Using Bluetooth, sample sizes of 5% - 7% of all vehicles are achievable with high levels of accuracy at a much lower cost (Tarnoff *et al.*, 2009). The subsequent sections present Bluetooth traffic sensing in relation to the four key data requirements considered in this research, with reference to other related applications.

2.5.2 Bluetooth for the estimation of link-flow

Flow is one of the key traffic data requirements considered in this research, being one of the most important raw traffic datasets for modelling and calibration in planning and congestion management applications. Bluetooth traffic sensing presents the opportunity to derive real time traffic flows to optimise the road networks. However, given that Bluetooth presents a sample of the actual traffic, it is important to understand this fraction (detection rate) in relation to the actual traffic. Table 2.8 to Table 2.13 present the summary of the review to understand this metric. This metric is classified into six different groups based on distinctly identifiable parameters that vary across the study locations. Table 2.8 presents the detection rates relating to people count versus the number of discoverable Bluetooth devices. Besides the limitation in scope, in terms of scale and period, the information obtained is rarely useful to infer the general traffic conditions given that the main cause of traffic congestion are vehicles and not people.

Author (Year)	Study Location	Detection Rate (%)	Method	Results
O'Neill *et al* . (2006)	Bath, UK	7	Discoverable devices were scanned whilst taking gate counts of people passing at four (4) locations for a short period of about 30 minutes. The counting of people was automated using the phone method	Linear correlation was observed between the number of people and discoverable Bluetooth devices. A detection rate of 7% was obtained.
Nicolai and Kenn (2007)	Bremen and San Francisco	2 and 6 respectively	It measured the percentage of people with discoverable Bluetooth devices whilst number of discoverable devices was plotted against the total number of people (gate count) passing through the gate.	The results obtained showed a positive linear correlation between the number of people and the discoverable devices. The difference in the detection rate over Bremen and San Francisco is attributed to population and variation in Bluetooth usage

Table 2.8: Table showing the detection rate of people with discoverable Bluetooth devices in Bath, Bremen and San Francisco

Table 2.9 presents the studies carried out on arterials of different urban areas across Europe. From this table, the minimum detection rate (15%) was computed by Roggendorf (2012) based on Bluetooth/manual count comparison. Despite being the lowest detection rate, there is a concern that the different Bluetooth pairs detected include those carried by cyclists and pedestrians. Consequently, the vehicular traffic proportion was not represented which explains the reason for the relatively high estimation of detection rate. A similar concern can be expressed over the high detection rates obtained in the other studies (Barceló *et al.*, 2010; Beca, 2011; Augustin and Poppe, 2012). Since a node-based method (comparison based on detected Bluetooth devices over the individual Bluetooth stations) was used in the computation of the detection rate as against the link-based method, two sources of errors can be identified as follows: i) contributions from vehicles from the opposing link; and ii) contribution from non-vehicular sources with Bluetooth devices such as pedestrians. Taking this into account is essential for a reliable estimation of traffic flow. Other limitations include the period of observations and limited information on the type of comparison made (Beca, 2011; Augustin and Poppe, 2012). However, these studies have provided vital information regarding Bluetooth traffic sensing over different geographical areas across Europe; thereby serving as *a priori* knowledge of the expectation in the UK. That is, the variation observed in the computed detection rates is indicative of levels of usage of Bluetooth-enabled devices in the study locations. If any of these locations share similar traffic characteristics and populations with a UK city, then one may assume that a detection rate consistent with such location(s) is representative in such a UK city. The above assumption informs an important research gap requiring the understanding of the detection rates over the chosen study area in the UK to enable a reliable estimate of traffic flow using Bluetooth. Hence, this research will build on the knowledge gained from previous studies to determine the detection rates. For example, detection rates will be computed based on directional link-flows on a long-term basis covering all the hours of the day, weekday, month and season; this is to minimise the errors in the estimation, and to fully explore the variations that may affect any inference made. Therefore, consideration will be given to these important research gaps to ensure a fundamental understanding of estimation of flow using Bluetooth.

Author (Year)	Study Location	Detection Rate (%)	Method	Results
Barceló *et al*. (2010)	Barcelona, Spain	30	Simulation and pilot study was conducted using well-calibrated inductive loops. Simulation was performed based on the detection rate and the available information in the area	The travel times predicted from the study show a high level of reliability on the use of Bluetooth to determine journey time
Beca (2011)	New Zealand	32.1 – 34.4	Bluetooth count was compared with traffic count from SCATS loops. Floating vehicle using GPS data logger was used to calibrate travel time and monitor speed along the route	The Bluetooth study using the BlipTrack system suggests a possibility. Detection rates of 32.1 - 34.4 were obtained over the study locations
Roggendorf (2012)	Aachen, Germany	15	Bluetooth compared with manual count of vehicles sampled between 8am to 5pm. Blids sensors were used at intersection to determine traffic flow	5250 different Bluetooth pairs were detected over 24 hours giving a detection rate of 15% against the manual count
Augustin and Poppe (2012)	Austria	38	Blids sensors were used in this study. Data used for the evaluation was not explicitly mentioned	Detection rate of 38% was obtained from the study

Table 2.9: Detection rates obtained from different urban arterials across Europe

Table 2.10 presents the results of the detection rates obtained from the studies conducted by BlipTrack (2012) and Araghi *et al.* (2012b). These studies conducted in a heavy traffic area of Aalborg both utilised the BlipTrack sensors, and as with the previous studies identified in Table 2.9, the estimation of detection rates was node-based. This means that it could not account for the uncongested area as well as the temporal variability that may be present in the data. There is also a concern with the method of installation of the sensors used in both studies. Keeping the sensors used on the ground means that they could easily be affected by many factors arising from displacement and damage that may consequently affect the configuration of the orientation and inclination, and the overall results. On the contrary, the Hi-Trac Blue sensors utilised in this research are installed on lamp posts. This installation method takes care of the risk of displacement and any possible accidental damage that could easily affect the sensors if they were ground-based, as used in the previous research. A major observation from the two studies is a difference of 7% in the detection rates (20 and 27%) computed at different sites in the same year in Aalborg. This significant variation shows that great attention must be paid to the network of varying characteristics, in order to obtain reliable results. Therefore, network configuration is considered an important factor and shall be explored in this research for a better understanding.

Author (Year)	Study Location	Detection Rate (%)	Method	Results
BlipTrack (2012)	E45 Aalborg, Denmark	27	Made use of historical flow record and also carried survey with some car dealers for Bluetooth information on car. Study conducted in the most heavily trafficked route in the region	The survey revealed that some cars have permanent discoverable Bluetooth hands free with a detection rate of 27%
Araghi *et al* . (2012b)	Denmark	20	Bluetooth was compared with general traffic volume. The sensors used were placed on the ground	The proportion of Bluetooth detection in the study area of Denmark gave 20% of the actual traffic

Table 2.10: Detection rates obtained in Denmark using the BlipTrack sensors

Table 2.11 presents a different cluster of the minute count ratio of Bluetooth to ANPR carried out on a motorway in Denmark by Muhammed and Egemalm (2012). The trial conducted on the motorway, E45 over 4 – 6 April 2012 distinguishes it from other studies piloted in arterials. Although this study attempted to capture all the varying periods, such as holidays, that could affect the estimation of travel time, there is concern about the choice of 5-minute interval count adopted. At this level of resolution, Bluetooth count is expected to yield a significant zero detection particularly during the off-peak periods thereby leading to unrealistic and unreliable estimation. Therefore, there is a concern that the result obtained contains a significant level of outliers arising from non-vehicular devices. Consequently, the minute count ratio of about 54 - 61% of the total time of the investigation between ANPR and Bluetooth is not regarded as the actual detection rate. However, the result obtained from travel time analysis was said to be reliable up to 95% matching.

Author (Year)	Study Location	Detection Rate (%)	Method	Results
Muhammed and Egemalm (2012)	Denmark	About 61% (Based on minute count ratio)	ANPR and Bluetooth comparative study with field test. Trial between 4-6 April 2012 on motorway, E45. Minute by minute count ratio of Bluetooth to ANPR was determined as against the standard method of determing detection rate.	The result obtained from travel time was accurate to 95%. The minute count ratio between ANPR and Bluetooth was up to 61%.

Table 2.11: Minute count ratio of Bluetooth to ANPR on Motorway

Table 2.12 presents the result of the detection rate computed over long distances (27.8 – 310km) in the Netherlands (Biora *et al.*, 2012). While this study has also provided useful knowledge on the potential of Bluetooth data, there is a concern regarding the estimation of detection rate over such long distances, particularly over 310km. At such range, not many vehicles are expected to travel that far except vehicles on tour. The computed 25 - 40% was based on the total devices captured, and this is rarely helpful for traffic planning and management purposes because it can lead to an exaggeration of the traffic volume. In this research, sections of roads of relatively short distances are considered within the urban arterials in the study locations as opposed to motorways. Also, metrics estimation is focused on vehicular traffic while the preliminary stage will investigate the general traffic.

Author (Year)	Study Location	Detection Rate (%)	Method	Results
Biora *et al* . (2012)	Netherlands	25-40	Made use of i-Travel systems to determine detection rates on four different sections ranging from 27.8 – 310 km long. Bluetooth was compared with total traffic volume. Sections of the road used are: A6, A7, A32 & A31	Varying results were obtained in the sections of the road investigated ranging from 25% to 40%

Table 2.12: The detection rates obtained over long distances in the Netherlands

In a study carried out over nine days in Scotland, Cragg (2013) compared Bluetooth station counts with data from ATC and obtained 20% and 33% for weekends and weekdays respectively as shown in Table 2.13. As observed from the literature, the node-based detection rate is rarely useful for traffic management purposes due to the influence of non-vehicular devices on the estimation and thereby resulting in incorrect and inflated rates. In general, the current challenges in the derivation of the detection that include variations over different geographical locations and network configurations, justifies the need for continued investigation into exploring the potential of Bluetooth technology for traffic flow estimation.

Author (Year)	Study Location	Detection Rate (%)	Method	Results
Cragg (2013)	Scotland	20 and 33 for weekend and weekdays respectively	Bluetooth station counts compared with ATC	The proportion of Bluetooth station count was consistent over different comparisons conducted between ATC and ANPR data

Table 2.13: The Bluetooth detection rate based on station counts against ANPR

2.5.3 Bluetooth for the estimation of travel times

The tendency to use Bluetooth technology for travel time estimation is rising for many reasons such as, an increase in Bluetooth-enabled devices among road users, anonymity of Bluetooth detections, flexibility of deployment and maintenance of Bluetooth sensors (Araghi *et al.*, 2012a). The first group believed to use this new approach to determine travel time for traffic purposes are a team of engineers from Indiana Department of Transportation and Purdue University (work reported in June 2008) (UMCATT, 2008; Haghani and Hamedi, 2013). Following this development, other researchers have conducted studies on the use of Bluetooth for the determination of travel times (TMCnet, 2011; Muhammed and Egemalm, 2012; Bhaskar and Chung, 2013; Araghi *et al.*, 2015). Araghi *et al.* (2012a) showed that travel times measured by Bluetooth compared well to those by tag readers (the use of radio frequency identification and detection – RFID). Consequently, Bluetooth shows promise for travel time estimation. Bluetooth travel time data is similar to that of ANPR with a considerable advantage of continuity (Biora *et al.*, 2012). Continuity defines the ability of a system to function over a given period without interruption (Langley, 2011). Bluetooth data is also not degraded in the case of poor visibility, nighttime, rainy, snowy and foggy conditions (Biora *et al.*, 2012). Bluetooth travel time estimation on motorways and on arterial roads has been shown to have comparable accuracy to video cameras (Wang *et al.*, 2011; Mei *et al.*, 2012). Webster *et al.*'s (2014) study also indicated the potential for travel time estimation on sections of motorways. Erkan and Hastemoglu (2016) examined the applicability of Bluetooth for travel time estimation in heterogeneous traffic in Istanbul, Turkey. A detection rate of 5 % of all vehicles was obtained from this study. The study utilised weighted linear regression methods to estimate travel time, with a conclusion that Bluetooth can be used to estimate travel time in heterogeneous traffic conditions. In addition, Bluetooth has been applied for real-time travel time prediction to improve the road network management (Qiao *et al.*, 2013).

UMCATT (2008) showed that by sampling a portion of the travelling vehicles' actual times from the traffic stream, Bluetooth traffic monitoring provided the opportunity to collect high-quality travel time data. UMCATT (2008) provided the knowledge of the basic concept of Bluetooth traffic monitoring; however, it is limited in scope both in terms of duration of the study and the area covered. Not only that, the few systems deployed were over a long distance (2-4 miles apart) while the data collection was over 48 – 96 hours. These limitations leave an uncertainty regarding the evaluation of the behaviour of the data collected over a short distance and a long period. These limitations need to be accounted for by investigating Bluetooth data over relatively short distances such as 500m links in different urban areas, and on a long-term basis spanning a year, to capture any seasonal variations in travel time estimated by Bluetooth. This research will explore this gap to increase the level of confidence of Bluetooth travel time estimation. Table 2.14 and Table 2.15 present the summary of the key research on Bluetooth for travel time estimation.

Authors	Bluetooth research areas	Methods	Conclusions	Remarks
Puckett (2010)	Travel time monitoring for border censoring	Anonymous Wireless Address Matching (AWAM) system.	Confirmed that Bluetooth can be used to measure border crossing times. Based on the AWAM proof-of-concept on urban arterials, Bluetooth device penetration is sufficient to collect high-quality travel time data	The report presented on behalf of TranStar presents no evidence/analysis or a detailed discussion of the method of validation used
Quayle *et al.* (2010)	Arterial travel time	Compared Bluetooth and GPS data	Concluded that Bluetooth has the ability to accurately measure travel time over long spans of time	Study conducted in Portland, Oregon
Malinovskiy *et al.* (2011)	Travel time estimation	Used three types of antenna with three different sensor arrangements on a short corridor (0.98 mile) of a varying configurations. Compared Bluetooth travel time to LPR	Larger detection zone is desirable while shorter corridor will have greater travel time errors. A pair of sensors mounted at opposing sides at each end of the corridor will result in significantly less error. Omnidirectional antennae have larger detection zone than unidirectional antennae but are subject to more temporal and spatial errors	The study was conducted on a short corridor of 0.98 mile
Porter *et al.* (2011)	Calibration of sensor and travel time estimation	Explored the suitability of five different types of Bluetooth antennae	Antenna type has an impact on the quality of the data collected	This may not require further study
Abbott-Jard *et al.* (2013)	Bluetooth and WiFi Scanning for travel time estimation	Used exist-exist method, and Excel and Matlab for data filtering. Used two types of antenna	The study conducted in Brisbane showed that Bluetooth has a higher match rate than WiFi - approximately 1:8; percent of usable data suggested 81 percent for Bluetooth and 19 percent for WiFi	WMS are not widely used, and their usage is still being explored. One day trial. No quantitative analysis of the travel time data
Bhaskar and Chung (2013)	Bluetooth as complementary data source	Explored the effects of detection zone on the accuracy of travel time estimation using Bluetooth	Proposed three mode of estimation for travel according to the modelled section of the signalised urban environment	Explored accuracy and reliability of travel time
Moghaddam and Hellinga (2013)	Travel time error evaluation	Evaluation of algorithms to detect outliers in travel time	Mean travel time error is always close to zero in all traffic conditions	The evaluation was based on simulation study constrained to the upstream and downstream of the traffic. This might not capture the errors arising from vehicles using other connecting routes
Platt (2013)	Travel time estimation	Bluetooth experimental set-up in South Wales was explored	The outcome of the experiment is positive as the information from Bluetooth is being fed to the management system for a display on VMS to aid commuters	No result was presented in this discussion
Qiao *et al.* (2013)	Real-time travel time prediction	The study implemented historical average, auto-regressive integrated moving average (ARIMA), Kalman filter, and K-nearest neighbours (KNN) models	Results showed that using the non-parametric approach, the prediction accuracy can improve by more than 10% for all day period and 20% for peak-hour periods over the other methods considered based the computed mean absolute percentage error (MAPE)	This study proposed a new model called KNN-T to improve travel time prediction accuracy, and also provided the knowledge of the suitable models to apply in travel time prediction

Table 2.14: Bluetooth for travel time estimation and traffic management – 2010 to 2013 studies

Authors	Bluetooth research areas	Methods	Conclusions	Remarks
Wu and Rilett (2014)	Reliability of real-time travel time estimation	Studied the prediction of travel time at a 15-minute level under different traffic conditions	The model comparison between the link-based and corridor-based prediction of travel times yielded comparable results	Established a correlation between the reliability of link-based and corridor-based short-term travel time prediction
Araghi *et al.* (2015)	Reliability of travel time estimation	Bluetooth and GPS consisting of 1000 trips were used as the controlled experiment. The GPS formed the ground-truth used to calibrate the Bluetooth detection rate.	Found that Bluetooth can be detected up to 80% of the time at a sensor location	The concern here is the use of only one vehicle for the experiment. This may introduce bias due to driving behaviour. The fact that it was also conducted on a link may not be representative enough
Stevanovic *et al.* (2015)	Accuracy and reliability testing of arterial travel times	Application of MAC readers to measure travel time in arterial roads. Used sensor developed by Florida Atlantic University (FAU) team. Four months field test of two test-bed networks around FAU. Used two type of antennae (omni and uni-directional), and compared results with GPS floating car technique. Also considered varying speed and antennae	Regression analysis between Bluetooth and GPS yielded R-Square equal to 0.65. Placement of Bluetooth in vehicle is significant (dashboard location is preferable)	Test statistics not presented but it was concluded that there is no significant difference in the travel time of Bluetooth and GPS at 95% level of confidence
Yu *et al.* (2015)	Travel times and volume for incident detection on arterial roads	The study used an incident detection algorithm based on moving average	Moving average was used to address the limitations resulting from sparse travel time sample data to obtain	Propose an incident detection algorithm that utilises travel time and traffic volume to establish a good balance between the actual detection rate and false-alarm rate
Araghi *et al.* (2016)	Mode-specific travel time estimation	Clustering techniques was used to explore the feasibility of Bluetooth to estimate mode-specific travel time	Clustering techniques can be used to carry out satisfactory classifiaction with an accuracy comparable to that of ANPR	The use of class of device for classification may not in all cases be feasible due to data encryption for private reasons
Park *et al.* (2016)	Performance of travel time at intersection	Utilised omnidirectional antennas for intersection-intersection analysis of travel times to estimate control delay at intersection. The data used spanned 6 - 19 December 2011. Received signal strength was used to transform the travel time while the estimate of flow was compared with data from loop detectors	Obtained detection rates between 5.8 - 84% over the different sections of the road. The estimated controlled delay was found to vary proportionally with the actual travel time. That is the control delay increases with an increase in travel time	This study did not consisder statistical analysis of the results

Table 2.15: Bluetooth for travel time estimation and traffic management – 2014 to 2016 studies

2.5.4 Bluetooth for the estimation of vehicle speed

Vehicle speed measurement, particularly where precise timing is important, is mainly carried out by a technological-based method for performance evaluation of road network and queue analysis (White, 1989). The GPS-based method has been used to gather precise travel times and speed information about the road traffic for real-time application. Bluetooth is now considered as a viable option in this regard. Table 2.16 presents the key studies. The effect of vehicular speed and multipath fading was considered by Pasolini and Verdone (2002), while Houston TranStar (2010) considered travel time speed estimation with a focus on cost-comparison with other sensors. Average speed and time are fundamental measurements of the traffic performance (May, 1990). Although, they are inverse measures, they are used differently in traffic engineering (Roess *et al.*, 1998). Further, the profile analysis of both travel times and vehicle speeds can be used to understand other traffic characteristics such as congestion, while flow and speed can be used to derive density – defined as the number of vehicles per unit length of the roadway (Roess *et al.,* 1998). Bachmann *et al.* (2013) compared data from Bluetooth and loop detectors with GPS data on a stretch of Highway 401 in Toronto, Canada. The analysis showed that the accuracy of traffic speed estimates obtained from loop detectors can be improved through Bluetooth data fusion. Also, the comparison of speeds based on GPS and Bluetooth data, and the simultaneous use of both datasets to improve estimation accuracy has been studied (Borresen *et al.*, 2016). However, Bluetooth traffic sensing for vehicle speeds estimation is currently under-investigated, and shall be explored in this research to contribute to knowledge.

Authors	Bluetooth research areas	Methods	Conclusions	Remarks
Pasolini and Verdone (2002)	Suitability of Bluetooth in ITS for provision of services for guidance support, and effect of vehicular speed and multipath fading.	Analytical and experimental text-bed. Examined the maximum distance between devices to communicate.	Bluetooth communication is sturdy but the presence of many vehicles can cause performance degradation due to the polling technique used by Bluetooth. Link performance is not limited by vehicle speed but by the amount of signal-to-noise ratio, and the transmitted power	An indoor experiment. Examined connection set-up delays and transmission reliability in a dynamic scenario. Found that file transfer delay is not affected at distances less than 60m
Houston TranStar (2010)	Speed, travel times, and cost comparison	Toll tag and Bluetooth data were used. 3,271 toll tag speed compared to 7,492 Bluetooth speed data sample	The two sets of data were virtually the same after filtering to remove outliers. Accuracy rate of Bluetooth as high as that of AVI system. AVI cost per unit - $75,000. LPR - $25,000 per four-lane installation; and Bluetooth - $2,000 is low-cost	Focus mainly on speed data. Low-rate not accounted for i.e. how to know the real traffic volume, variability not discussed, and detection rate is unknown. The study was conducted on a 2.2 miles road for 24 hours (1 day). Speed less than 5 mph on a freeway were removed. No result was presented in the report
Bachmann *et al.* (2013)	Freeway traffic speed estimation	Combined Bluetooth with loop detector data for improved speed estimation	Bluetooth and probe data such as GPS can improve estimation	The study is carried out on a freeway and not in urban roads that has different characteristics

Table 2.16: Bluetooth for vehicle speed estimation and traffic management

2.5.5 Bluetooth for the estimation of origin-destination matrix

Origin-destination matrices are estimated using the observed link-flow information (Aslam *et al.*, 2012). Traditional methods such as roadside surveys for the collection of O-D information often require additional resources in terms of time and cost, and may not provide up-to-date data (Srinivasan, 2011; Wang *et al.*, 2013). However, sustainable mobility requires a better management of the available infrastructure resources (Fernández-Lozano *et al.*, 2015). Presently, Bluetooth is one the technologies used to overcome these challenges within an urban network as have been previously demonstrated (Barceló *et al.*, 2012; Ayodele *et al.*, 2013; Barceló *et al.*, 2013; Bhaskar *et al.*, 2014). Bluetooth has been identified as a potential candidate for O-D estimation with the ability to provide real-time information as opposed to reliance on historical data (Barceló *et al.*, 2010; Bhaskar *et al.*, 2014). However, the literature shows that continued research is required to maximise the potential of this technology in traffic management. In fact, Nantes *et al.* (2014) noted that the issues of accuracies in Bluetooth data are yet to be adequately resolved.

Filgueiras *et al.* (2014) presented proof-of-concept deployment of Bluetooth technology to detect traffic flow conditions. The results showed that different information such as O-D matrices and travel times can be obtained using Bluetooth. The significance of Bluetooth traffic monitoring as a reliable source for O-D matrix was demonstrated in the study conducted in an urban area of Brisbane using seventy-nine Bluetooth sensors. This study compared Bluetooth results with loop detector data for assessment (Laharotte *et al.*, 2014; Laharotte *et al.*, 2015). O-D matrix estimation based on Kalman filtering has also shown promise for real-time estimation as previously demonstrated (Barceló *et al.*, 2013; Zhong and Lee, 2014). This feasibility was also affirmed by Fernández-Lozano *et al.*,(2015). Table 2.17 presents a summary of the key research on Bluetooth O-D estimation. This research will build on the available knowledge of the use of Bluetooth data to explore both the spatial and temporal variations in the estimated O-D matrix within GMN to reveal relevant underlying information about Bluetooth O-D estimation.

Authors	Bluetooth research areas	Methods	Conclusions	Remarks
Pasolini and Verdone (2002)	Suitability of Bluetooth in ITS for provision of services for guidance support, and effect of vehicular speed and multipath fading	Analytical and experimental text-bed. Examined the maximum distance required for devices to communicate	Bluetooth communication is sturdy but the presence of many vehicles can cause performance degradation due to the polling technique used by Bluetooth. Link performance is not limited by vehicle speed but by the amount of signal-to-noise ratio, and the transmitted power	An indoor experiment. Examined connection set-up delays and transmission reliability in a dynamic scenario. Found that file transfer delay is not affected at distances less than 60m
Blogg *et al.* (2010)	Travel time and O-D estimation	Utilised station count from Bluetooth sensors installed in Brisbane for O-D analysis	Reported that the results of the O-D estimation compared well with ANPR and Video data.	The study location is more or less a linear network. This may not be representative of the scenario for a complex O-D network. This study also utilised the MAC detection to estimate detection rate, and was subsequently compared with the actual volume. This does not reflect the true estimation level from Bluetooth
Barcelo *et al.* (2012)	Travel time and O-D estimation in freeway	Study conducted on a 40-km long section of road in Barcelona Spain	A caution on the use of Bluetooth for O-D matrix estimation	The data collection was over 2 months period in 2009. This study appears to make use of both Bluetooth and WiFi in the estimation of the O-D matrices. Therefore, the conclusion drawn cannot be generalised for Bluetooth
Barceló *et al.* (2013)	Estimation of O-D matrices	Kalman filtering approach	The numerical results shows Bluetooth possibility	The use of Kalman filter is memory intensive
Wang *et al.* (2013)	Dynamic O-D estimation and feasibility study	Used cell phone location tracking algorithms for data collection and estimation	Detection of 17.6 percentage of the daily traffic. The tracking algorithm is preferable for long distance or inter-city trips. It requires longer observations to increase the sample size	Six weeks observation in Kansas Metro Corridor
Bhaskar, *et al.* 2014	Estimation of traffic state	Integrated Bluetooth and loop data to estimate travel time and density	Bluetooth provides a good estimate of travel time but there is variability in sample size captured	The issue of variability in the sample collected is not discussed. Also the validation of the estimated density was through simulation. This is a common practice in anyway

Table 2.17: Bluetooth applications to origin-destination analysis

2.5.6 Other relevant use of Bluetooth traffic sensing

Table 2.18 presents related studies to Bluetooth traffic sensing with a focus on applications such as density estimation, sensors positioning and distributions, stand-alone traffic monitoring, and traffic light management (Nantes *et al*., 2014; Collotta and Pau, 2015; Park and Haghani, 2015; Salem *et al*., 2015). Also, Ayodele *et al*. (2014) presented the autonomic concept of Bluetooth to estimate vehicle emissions, while Bluetooth was used to detect passenger trips on public transport buses in Funchal, Portugal. Bhaskar *et al*. (2014) have demonstrated the potential of Bluetooth for density estimation to understand traffic characteristics. In the past, this quantity has been difficult and expensive to acquire but the problem can now be overcome using Bluetooth data (Bhaskar *et al.,* 2014). Figure 2.2 presents the generalised relationship among speed, density, and flow rate showing the three fundamental parts of a typical speed-flow curve viz: upper part – free flow, low part – congestion, and the projected part – busy (Purdue University, 2016). This and other applications such as modes classification (Araghi *et al*., 2012a), automatic vehicle identification, toll collection, and distress alert etc. can be explored to improve traffic management. Other applications include congestion study through the analysis of travel time index (TTI) – the ratio of the actual peak period to free-flow travel times (Lomax, 2010). A working definition of congestion is 'travel time or delay in excess of that normally incurred under light or free-flow travel conditions' (Gifford, 2003, page 181). HCM (2000) defined traffic delay as the delay component resulting from reduction of speed below the free-flow speed due to interaction of vehicles. When delivering a decision support system, objectives are set out and performance measures designed against the most appropriate option to be selected (Ayodele *et al*., 2014). LOS (level of service) measures the performance level of the network at various operating conditions (Mathew, 2014). In the future, Bluetooth might be used in this regard to deliver an efficient decision support system. For example, information gathered using Bluetooth may be sent to drivers based on the driving condition to optimise the speed and where possible to always arrive at junctions on a green light. Cooperative and integrated deployment of Bluetooth technology is another potential application. The European Commission defined cooperative systems in road traffic as: cooperation between road operators, infrastructure, vehicles, their drivers and other road users to deliver the most efficient, safe, secure and comfortable journey beyond what stand-alone systems can achieve (European Commission, 2004). Cooperative mobility on the other hand is defined as the sharing of information due to the interconnection of vehicles and infrastructure leading to better cooperation amongst drivers, vehicles and roadside systems (Boethius, 2011). The intelligent use of Bluetooth data in this way could help deliver a safe, sustainable and robust future transport system. In particular, the fusion of fixed and mobile networks (Edwards *et al*., 2012). Combining different data sources in connected and inter-connected environments begins to deliver the essential metrics that underpin a cooperative system. The added advantage of Bluetooth is that it is not affected by weather conditions such as snow or fog, unlike ANPR or video recording. This makes Bluetooth robust and complementary to the existing ITS sensors to deliver the cooperative objectives. Bluetooth could also serve as a 'big data' source to meet transport demands. Big data refers to enormity in five dimensions namely volume, variety, velocity, variability, and complexity, and are from different sources and formats that include mobile devices and the web (Troester, 2012).

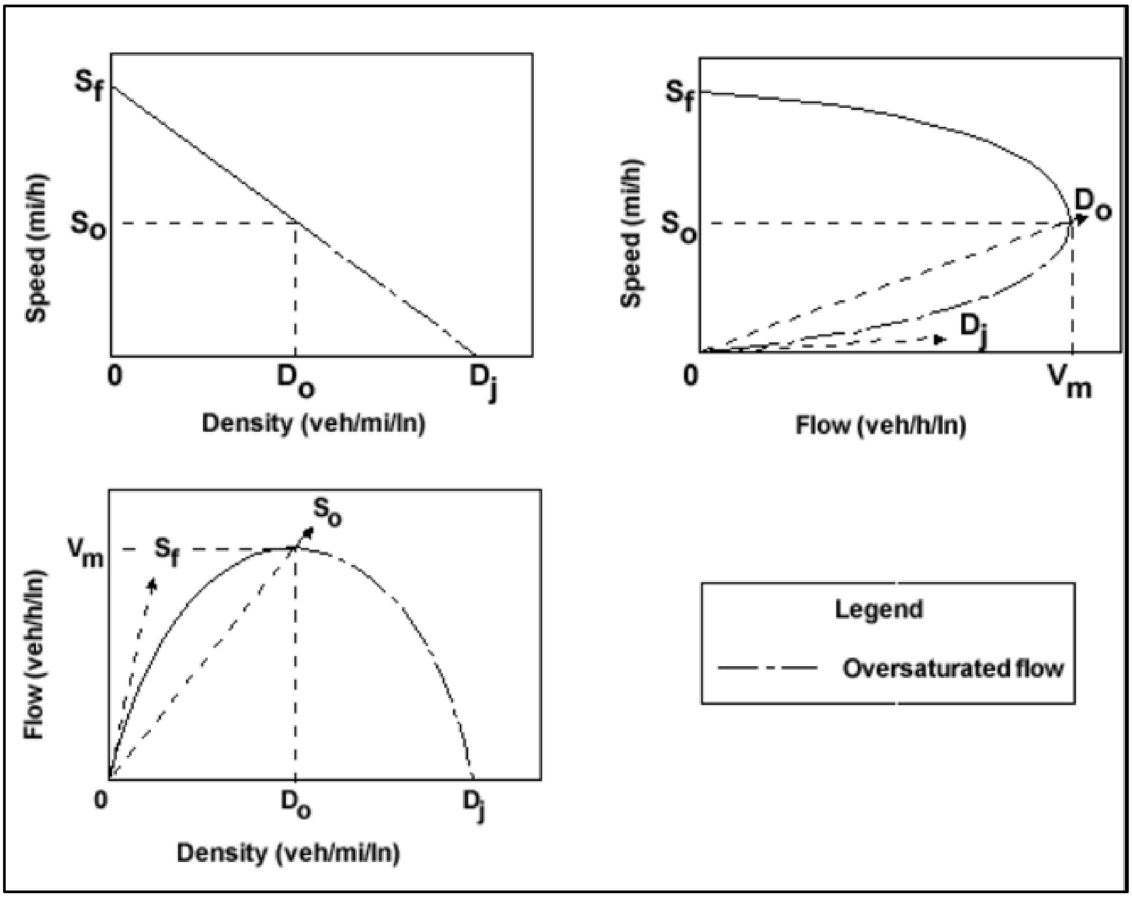


Figure 2.2: Generalised relationships among speed, density, and flow rate on uninterrupted-flow facilities

Source: Modified from Purdue University (2016)

Authors	Bluetooth research areas	Methods	Conclusions	Remarks
Hallberg *et al.* (2003)	Positioning	Bluetooth-based RSSI	Developed positioning system based on Bluetooth.	Not effective with personal Bluetooth devices
Castano *et al.* (2004)	Local positioning	Used RSSI (received signal Strength Indicator) distance estimation based on Kalman filter	The application can be used to track patients in the hospital	Only the transmitted power and the RSSI provided reliable information for distance estimation
Solon *et al.* (2006)	Bluetooth vulnerabilities	Explored the vulnerability levels of Bluetooth from three different manufacturers over 5 days	Detected over 340 Bluetooth-enabled devices of which Nokia presented the highest vulnerability rate (60%), Motorola (10%), and Sonny Erickson (30%)	The outcome is more of a concern for a Bluetooth manufacturer than for a traffic engineer
Browning and Kessler (2009)	Bluetooth Hacking	Explored different phones at varying operational conditions to test for attack	Concluded that there is possibility of Bluetooth hacking	The issue of Bluetooth vulnerability is limited in traffic management
Tarnoff *et al.* (2009)	Performance evaluation of freeway and arterials	Made use of Class 1 and Class 2 radios for vehicle detection on freeway and arterials	Introduced capabilities for Class 1 and Class 2 radios. Sample size of 5 - 7% with high level of accuracy	No results shown on the accuracy level
Martchouk *et al.* (2011)	Variation in different weather conditions (normal and abnormal)	Anonymous Bluetooth sampling on freeway using the hazard-based model	Significant difference in mean and standard deviation of travel time in different weather conditions	the two weeks data may not give the knowledge of any seasonal variability
Porter *et al.* (2011)	Calibration of sensor and travel time estimation	Explored the suitability of five different types of Bluetooth antennae	Antenna type has impact on the quality of the data collected	This may not require much further study
Abbas *et al.* (2013)	Microscopic modelling of control delay	Used Bluetooth and GPS probe vehicle data	The combination of Bluetooth and GPS data gives an added advantage	This area needs further investigation as results are based on simulation
Abedi *et al.* (2013)	Crowd data collection and monitoring	WiFi and Bluetooth data collection methods were contrasted. Investigated different antenna types	WiFi has shorter discovery time, and is preferable for crowd data	Benefits, challenges and enhancements were considered
Bhaskar and Chung (2013)	Bluetooth as complementary data source	Explored the effects of detection zone on the accuracy of travel time estimation using Bluetooth	Proposed three modes of estimation for travel according to the modelled section of the signalised urban environment	Explored accuracy and reliability of travel time
Hainen *et al.* (2013)	Quantitative evaluation of the operations of airport security check point	Exploratory analysis of the data was performed	Demonstrated that crowd source data obtained from mobile devices can be used to develop multi-modal transportation performance measures	Made use of 12 days (30 August -13 September 2010) data collected at George Bush International Airport. Exploratory analysis was performed and not quantitative analysis
Allström *et al.* (2014)	Calibration of traffic state	Calibration framework based on velocity based cell transmission model and ensemble Kalman filter	The results showed that for travel time estimation when calibrating the parameters on two-stage process is possible and even more important for travel time prediction	The scale needs to be extended for further generalisation
Bhaskar, *et al.* (2014)	Travel time and density estimation	Integrated Bluetooth and loop data to estimate travel time and density	Bluetooth provides a good estimate of travel time but there is variability in sample size captured	The issue of variability in the sample collected is not discussed. Also the validation of the estimated density was through simulation. This is a common practice anyway

Table 2.18: Other relevant applications of Bluetooth traffic sensing

2.6 Knowledge Gap

Important research gaps are identified from the literature review conducted. For example, little research has been carried out to understand the variability and errors in Bluetooth-derived metrics that usually cause uncertainty about conclusions drawn from the data (Turochy and Smith, 2002; Moghaddam and Hellinga, 2013). Also, evidence of the accuracy levels of the estimated traffic metrics as well as their statistical significance to ensure reliable reconstruction of traffic patterns and trends is hitherto under-investigated. This is because previous studies are limited in scale (period of data analysed) and also in terms of test and validation of field data. Simulation studies are often carried out instead. Besides, there is little knowledge of the variability and the spatial relationships in the estimated traffic metrics. In addition, there is little knowledge on the proportion (detection rate) of the actual traffic to understand the representativeness and reliability of traffic metrics estimation using Bluetooth sensors. For example, the current practice and research have estimated detection rates in different ways, namely: i) estimation based on the total devices captured at a station (Camacho *et al.*, 2010; Beca, 2011; Srinivasan, 2011; Cragg, 2013); and ii) estimation based on the combined (total) directional flow (O'Neill *et al.*, 2006; Cragg, 2013) (Section 2.3.4). The major limitation in the current practice of Bluetooth traffic flow and detection rate estimation is that such information is inadequate to plan and manage a complex transport network effectively. Three key reasons are identified for this limitation. Firstly, the aggregate representation of the traffic flow using the total devices captured does not represent the actual vehicular flow. For example, pedestrians carrying Bluetooth-enabled devices do not contribute significantly to traffic congestion or pollution. Secondly, the estimation of traffic flow using the total directional flow (summation of flows on the opposing links) does not present the level of service (LOS) each way in the network. Thirdly, the potential application and limitations of the Bluetooth approach to traffic management needs to be understood. In fact, Blogg *et al.* (2010) highlighted these problems as areas requiring improvement in knowledge. Therefore, a critical assessment of these limitations will enable a better understanding of the data to inform usability. Accordingly, clear distinctions between the different types of flow estimation are made to underscore the importance of the specific flows. The above challenges are considered very significant research gaps to inform usability; benefits derived, statistical confidence and sound inference on the subject.

2.7 Conclusions

This chapter presents a critical review of the relevant literature on the use of Bluetooth technology as an ITS sensor for traffic monitoring and metrics estimation (link-flow, travel times, speed and O-D matrix). The review focused on vehicular traffic while examining the issues of data requirements, accuracy and reliability. Currently, there is little work done in the area of ITS applications, particularly the applicability and viability of Bluetooth technology. The early studies showed that the availability of discoverable devices within the network is essential to the reliability of the results. Studies on the detection rates (2- 40%) of Bluetooth have been conducted on people and vehicles using different methods and over different geographical locations. There is a gap in knowledge regarding link-based estimations, accuracy, and the variability that may affect the results, and these are therefore taken into account in this study. Consideration was given to suitable analysis techniques such as exploratory and quantitative methods as the basis for results validation. With time, research into Bluetooth may form a key research area in the concept of Big data in solving transport problems. That is, Bluetooth may constitute an important part of the wide variety of data sources for transportation applications. Using a technological-based option such as Bluetooth to collect traffic data is considered a viable proposition. Therefore, Bluetooth could form an arm of traffic management functionalities to deliver performance measures such as travel time and speed to enhance traffic operations. However, the validity of these performance measures needs to be explored with respect to the established methods. The performance of Bluetooth at different temporal dimensions is considered an important research gap given that Bluetooth traffic monitoring is still a novel area. This review provides the motivation for continued research on the use of Bluetooth in ITS to support the realisation of better transport. Also highlighted are future directions and other potential applications.

This PhD research will explore, through validation, the reliability of Bluetooth for traffic sensing and metrics estimation in urban roads in the UK. Focus will be on four key traffic metrics (flow, travel time, speed, and O-D matrix). The next chapter presents the research methodology based on the Bluetooth approach to traffic sensing and metrics estimation.

Chapter 3. Research Methodology

3.1 Introduction

This chapter presents the methodology adopted in this research on Bluetooth traffic detection and metrics estimation, based on Bluetooth-enabled devices from vehicular traffic This chapter builds on the available knowledge such as that presented by UMCATT (2008) and Bhaskar and Chung (2013), to design and develop a Bluetooth-based data collection and processing model (TRAFOST). The model was used to derive and analyse traffic metrics (link-flow, travel time, speed and O-D matrix) at the chosen study sites in fulfilment of Research Objective number ii. While progress has been made in the area of travel time analysis, significant improvements are still required in order to understand the systematic procedure to derive useful traffic metrics. Therefore, this chapter presents a detailed discussion of the fundamental requirements to realise reliable estimates of traffic metrics using Bluetooth data. The discussion in this chapter encompasses research design through to results validation.

This chapter is structured as follows: Section 3.2 presents the research design detailing the research objectives and the corresponding methods of accomplishment, the data required, and the expected outcomes. The methods of Bluetooth data cleansing are presented in Section 3.3. This section considers the reliability and consistency of Bluetooth measurements of traffic data, representativeness of the measurements, multiple detection, and outliers to conclude the discussion. Section 3.4 presents the estimation methods of traffic metrics using Bluetooth data with a focus on travel time, flow, speed, O-D matrix, and detection rate. The validation methods for the results from Bluetooth data are presented in Section 3.5. This section deals with the strategies to validate Bluetooth results, before conclusions are drawn in Section 3.6.

3.2 Research Design

3.2.1 Justification of the research method

This section establishes the research problem and justifies the Bluetooth approach to traffic sensing and metrics estimation. As highlighted in research literature, conventional methods of traffic data collection and estimation are either very expensive to acquire and maintain, or difficult to implement. While technologies such as the FCD, ZigBee and WiFi present valid alternatives in terms of data requirements and cost, they have a lower penetration and growth rate compared to Bluetooth. For example, the penetration of Bluetooth in vehicles, mobile phones, and electronic devices gives Bluetooth an edge over other valid alternatives. Since the first report published in June 2008 on the use of Bluetooth for travel time estimation, there has been continuous evolution in this regard. However, the literature review clearly shows that continuous research is required to fully exploit the benefits of this technology in traffic management. Accordingly, this research considers the reliability of Bluetooth traffic sensing and metrics estimation, with a focus on the issues of accuracy and variability. To accomplish this, the Bluetooth results will be validated using already established methods to enable valid conclusions to be drawn on the applicability of the technology to enhance road traffic monitoring and management to reduce congestion. Central to this problem is the need to design, and develop a Bluetooth-based data collection, processing, and analysis procedure to derive useful traffic metrics. Currently the processing software are commercial-based, and are not available to the public. R programming language is adopted in this reasearch because R is free and open source unlike for example, Matlab that requires a licence. Figure 3.1 presents the diagrammatic flow of the research method showing the three main stages.

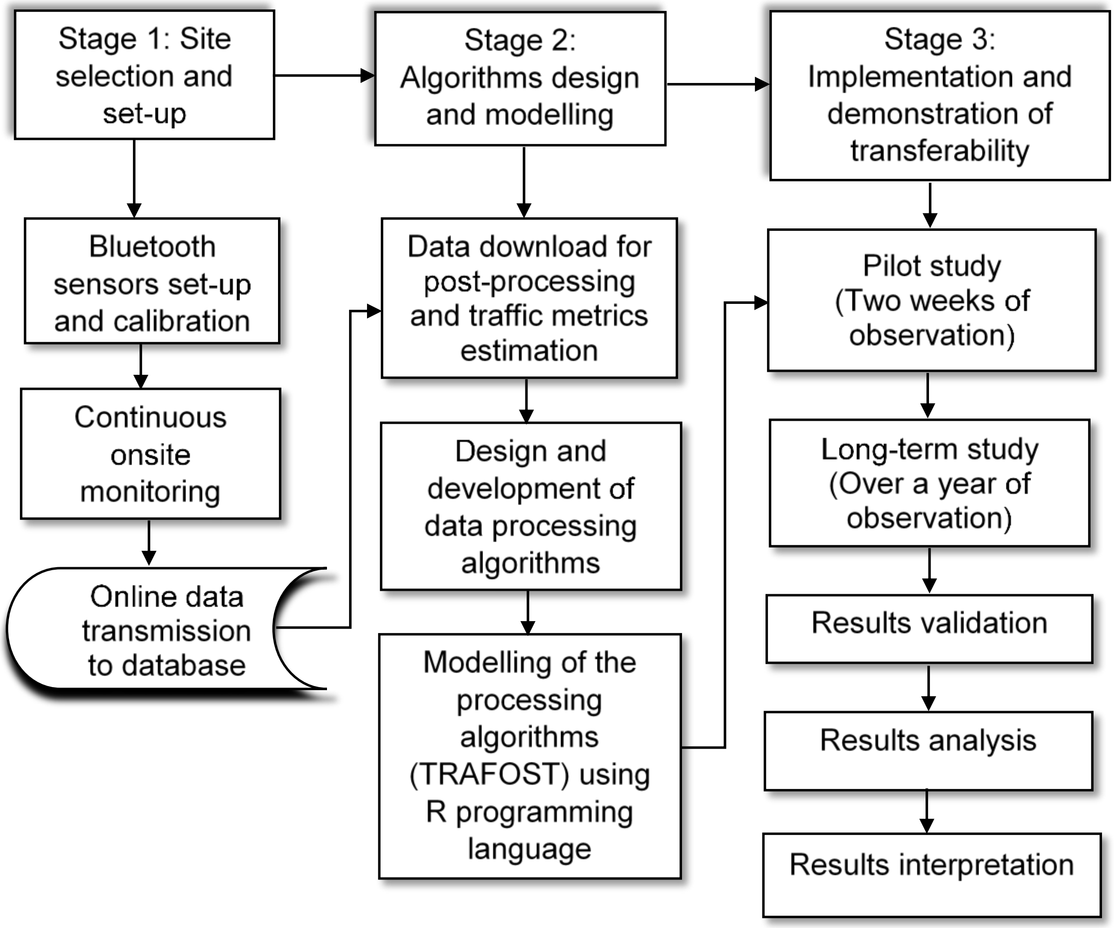


Figure 3.1: Research method diagrammatic flow

3.2.2 Data requirements and description

The research design consists of two sets of data: i) Bluetooth – the main data under investigation; and ii) the validation data sets – obtained from ATC and SCOOT loop detectors, ANPR cameras and GPS Traffic Master (TM) in the same-location as the Bluetooth sensors. The data from the ATC and SCOOT loop detectors are subsequently referred to as ATC and SCOOT flows. Table 3.1 presents a summary of the data required to accomplish this research. The summary includes the data type, period of collection, location, purpose and the number of stations and links used. Essentially, second-by-second Bluetooth encrypted raw data captured over the period 2011 to 2014 were used. Moreover, 15-minute Bluetooth counts obtained from C2-web – the software used by Transport for Greater Manchester (TfGM) also formed part of the data used for the validation. The short-term study consisted of two weeks' worth of data, while the long-term study made use of the data collected over the year 2013. The SCOOT data (15-minute link flows) complemented the ATC data (15-minute lane-by-lane flows) to ensure a sound and robust investigation. The limited ANPR data (vehicle-by-vehicle record over one day) were complimented with GPS Traffic Master data comprising six months (April – September 2013) of hourly averages to validate the estimated travel time and speed, in order to ensure statistical significance. The validation is essential to adequately establish the reliability of the Bluetooth approach to traffic metrics estimation.

Task	Data Source and Type	Period	Purpose	Location	Number of Stations/ Links
Model development and traffic metrics estimation	Bluetooth (Encrypted raw data and summary data from C2-Web)	Over 3 years (data from 2011 to 2014 inclusive)	Model building and estimation of traffic metrics; data quality, variability and transferability assessment	Liverpool, Birtley, and Manchester	55
Results validation	ATC (15-minute lane-by-lane flow)	1 year over 2013	Validation of flow	Manchester (Wigan – 1; Stockport – 2; Trafford – 2)	5
Results validation	SCOOT (15-minute link flow)	1 year over 2013	Validation of flow	Stockport, Manchester	2
Results validation	ANPR (Vehicle-by-vehicle record)	1 day in March 2014	Validation of journey times and speed	A6, Stockport Road	2
Results validation	Traffic Master (Hourly average of journey time and speed)	Six months from April - September 2013	Validation of journey time and speed	Manchester	4

Table 3.1: The description of the data requirements for the Bluetooth research

3.2.3 Bluetooth sensors set-up and data acquisition

TDC-Systems Ltd (TDC) developed and tested the Bluetooth sensors used in the acquisition of data in this research. TDC, in conjunction with the relevant Local Authorities, performed the set-up of the sensors for continuous onsite monitoring and transmission of data to the online database. The site selection was based on a careful consideration of factors that could affect the quality of data captured by the Bluetooth sensors (McDonald, 2013). That is, the sensor locations were selected for optimum performance. The stations were chosen and coordinated within the vicinity of existing traffic monitoring sensors, such as SCOOT and ATC loop detectors, within the road networks. The Bluetooth sensors were installed on an existing infrastructure at a height of 3m from the ground at the chosen stations. Data was captured through an automatic technique throughout the period of observations. Devices with their Bluetooth switched on and enabled were detected as they passed through the detectors' locations. This identification principle underpins the traffic data collection technique using Bluetooth technology. It is to be noted that the detected MAC addresses were encrypted before transmission to the online database for further analysis, either through real-time or post-processing. The encryption of the data complies with the Data Protection Acts to ensure that the privacy right of the device's owner is not compromised (TDC, 2011). The data used for post-processing and analysis of traffic metrics was downloaded from the online database through the access codes provided by TfGM.

Data availability and the reputation of TDC in producing traffic management systems are the reasons for making use of the data from TDC sensors. The sensors are 'Class 1' type designed to operate through continuous detection of Bluetooth discoverable devices carried by different traffic modes. The Hi-Trac Blue sensors utilised were developed in line with the core specifications of Bluetooth SIG (Special Interest Group) for automatic data capture (TDC, 2011). The sensors can cover up to six lanes at speeds up to 70mph, and they are fully compatible with all Bluetooth specifications (TDC, 2011). The sensors were designed to detect Bluetooth-enabled devices within the detection zone (range of 93m) seamlessly as opposed to the customary Bluetooth which is designed to connect with discoverable devices through password authentication. These Bluetooth sensors do not require code generation to initiate connection, and the process of detection is unnoticed by the device carriers (TDC, 2011). The data collection in this research was over five contrasting urban areas across three study sites (Liverpool, Birtley and Greater Manchester). The contrasting sites enable the understanding of the variability in results as well as transferability of the method. The study sites were chosen primarily due to data availability and the suitability to test the objectives of this research. However, it should be noted that changes to the research design due to limitations in the required sets of data from the Liverpool and Birtley studies brought about the additional sites in Manchester. The period of collection of Bluetooth and the validation data sets spanned 2011 through to 2014.

3.2.4 Description of the methods

This section describes the research objectives, the methods and data used as well as the expected outcomes. Table 3.2 and Table 3.3 present the research design classified as preliminary and evaluation stages respectively. This research design constitutes the plan to actualise the current problem, and to arrive at logical conclusions. The preliminary stages consist of objective numbers i to iii, while the evaluation stages consist of objective numbers iv to vi. In the design, a thorough review of the literature is first considered to establish gaps in knowledge and to contextualise the research. The second objective focuses on the development of a Bluetooth-based traffic data collection and processing procedure. Data collection and the pilot study were examined in objective number iii to round up the preliminary stages. At the evaluation stage, Bluetooth results were compared with the ground truth data to understand consistency, accuracy and variability in the data to enable critical analysis and interpretation in order to arrive at logical conclusions.

Objective number	Description of research objectives	Methods to achieve the objectives	Required data	Expected outcomes
i	This is to gain a comprehensive knowledge on Bluetooth application in traffic management and related applications to establish research gaps in current literature to extend the body of knowledge in this field of study	Online databases such as Web of Knowledge, Scopus, etc will be used to search for relevant articles and journals in this field	Data collection is not required at this stage. However, relevant information on the research topic will be acquired	Thorough knowledge of the field of study, and identification of gaps in the literature as well as contextualisation of this PhD research
ii	This involves the design and development of a model based on Bluetooth to derive traffic metrics	Acquisition of the relevant skills such as algorithm development, programming, data management and processing, etc. Liaising with the relevant stakeholders such as TfGM and TDC	Bluetooth data (few) to understand the physical properties such as structure and formats	The processing algorithms and a prototype Bluetooth-based model for traffic metrics estimation
iii	This objective involves data collection and the application of the model in targeted pilot studies in Liverpool, Birtley and Manchester for an overview of the potential of Bluetooth data for traffic management	Data collection shall be mainly through online download from TfGM database. Site visitation for verification where necessary, and model application for an overview study	Bluetooth, SCOOT, ATC, ANPR, and Traffic Master datasets shall be collected but only the Bluetooth data shall be utilised at this stage	Availability of the relevant data, and general understanding of the research based on the pilot studies

Table 3.2: Research design – preliminary stages

Objective number	**Description of research objectives**	**Methods to achieve the objectives**	**Required data**	**Expected outcomes**
iv	The performance of the model (TRAFOST) developed in Objective number ii will be examined for consistency, and fit for purpose, while the Bluetooth-derived traffic metrics will be tested for accuracy and reliability	Results from the model will be compared with the results from independent computation such as an independent software used by TfGM and Excel model. The use of repeated measurements where validation data sets are not available. Results will be validated against diverse independent measures of traffic to establish correlation. Relevant exploratory and quantitative analysis such as histogram, boxplot, and QQ plots will be explored. RMSE, MAD, and MAPE will be used as accuracy metrics to understand the degree of closeness of the estimated metrics to the actual or "true" values. ARIMA models shall be employed in the modelling of the estimated traffic metrics while the 80-20 rule of data splitting will be used to separate the training and test data sets. KL-D will be used to match Bluetooth data with the ground-truth to reach valid conclusions	Bluetooth, SCOOT, ATC, ANPR, and Traffic Master data	Calibrated and validated model and results. Establishment of the accuracy and reliability levels of the traffic metrics derived from Bluetooth. Statistical significance level of accuracy of Bluetooth-derived traffic metrics
v	Objective number v deals with the analysis of the variability in Bluetooth-derived traffic metrics to enable concrete deductions and sound inference	Exploratory analysis to understand some underlying properties; Principal Component Analysis (PCA) for data reduction; and 1-way ANOVA and Tukey's test to determine possible homogeneous subset. Variability statistics such as standard deviation and coefficient of variation (CV) shall be used. The representativeness of the sample shall be established using the package "samplesize4surveys" in R while CV will help to remove spatial differences such as scale in the data	A year (2013) Bluetooth data from the Greater Manchester Network (GMN) will be used	Understanding of the variability in Bluetooth-derived traffic metrics. Availability of releveant information to make informed decisions. Establishment of Bluetooth detection rates. Concrete conclusions to justify credibility
vi	To interpret the results and make deductions from the research findings in a wider context of applicability and viability and make recommendations for traffic management	Relevant skills such as critical interpretation and academic writing will be employed	The results obtained from the long-term study shall be interpreted for this purpose	Provision of relevant information to enhance traffic management using Bluetooth data. Contribution to the body of knowledge on the use of Bluetooth in ITS and traffic management

Table 3.3: Research design – assessment and interpretation stages

3.2.5 Development of TRAFOST for data processing

Given that the currently available software are commercial-based, and are not available to the public, a novel Bluetooth-based model (TRAFOST - Figure 3.2) was developed in this research to optimise Bluetooth data processing in order to derive useful insights. Appendix 3A presents the basic description of the four stages of the model, while the relevant codes are presented in Appendix 3B. The components relating to data cleaning and metrics estimation are presented in the subsequent sections. While effective data processing and cleansing requires the use of an existing or a novel algorithm, Heer (2014) stressed the importance of adequate data preparation before sending an algorithm over raw data to derive useful insights. Accordingly, the first major requirement will be the ability to manage and process the data to derive new insights. The processing of these huge data sets is usually carried out using machine learning, Hadoop (a free Java-based programming framework), programming languages such as R, cloud computing, and predictive analytics (Cook, 2014). Cleaning up data to the point where it becomes meaningful and useful is very demanding, and reconciling diverse data sources over which one has no control can take 80% of the total time (Smith, 2014). Therefore, in the design and application of TRAFOST, basic assumptions were made and tested in line with the research problem in order to obtain meaningful results.

1. All sources of errors (natural, instrumental, and personal) are assumed to be minimised at the time of installation of the Bluetooth sensors. Consequently, the results and any deductions made are not affected in this regard.
2. The Bluetooth traffic volume is expected to be higher during the congested period than at free flow, and similarly it is expected to be higher on weekdays than on weekends according to changes in the traffic situation and vice versa.
3. The Bluetooth sample of the traffic is expected to be consistently lower than the actual traffic, with a linear relationship corresponding to an increase or decrease in traffic level.

4. In a network of similar characteristics and under normal conditions, the detection rate was assumed to be constant over hours, days and across the network. This constancy is expected particularly if assumption number 3 is valid. Otherwise, Bluetooth might be difficult to apply to traffic management.
5. Following the *a priori* knowledge of the road network, devices travelling below 6km/h and above 120km/h are classified as outliers according to the boundary filter which was designed based on the average walking speed and the maximum speed the Bluetooth sensor can capture. These outliers include pedestrians, and high-speed vehicles (such as an ambulance), and are therefore not part of the traffic estimation. Consequently, advanced data filtering is required to cleanse other outlying values and noise remaining in the data, and these are taken into consideration in order to obtain valid results.
6. Irrespective of the filtering algorithm employed, noise arising from difficulty in differentiating devices during congestion, and unknown exact detection time of a device due to the inquiry time will be present in the estimation. Therefore, the design of the algorithm is subject to this limitation. However, estimation errors are expected to be minimised in a well-refined algorithm to obtain a valid result.
7. Another assumption made is that following appropriate data filtering by removing all sources of errors, Bluetooth results should present profiles and distributions similar to the actual traffic. Otherwise, the estimation algorithm will be considered to be in error and thus require modification; and where there is a marked difference not due to algorithm error, the data will be considered unusable.
8. If research assumption number 7 is valid, and the results are consistent with precision and accuracy, then the estimated metrics are considered reliable. Thus the reliability of Bluetooth for traffic metrics estimation to support traffic management and ITS applications such as in decision support systems and data augmentation will have been established.

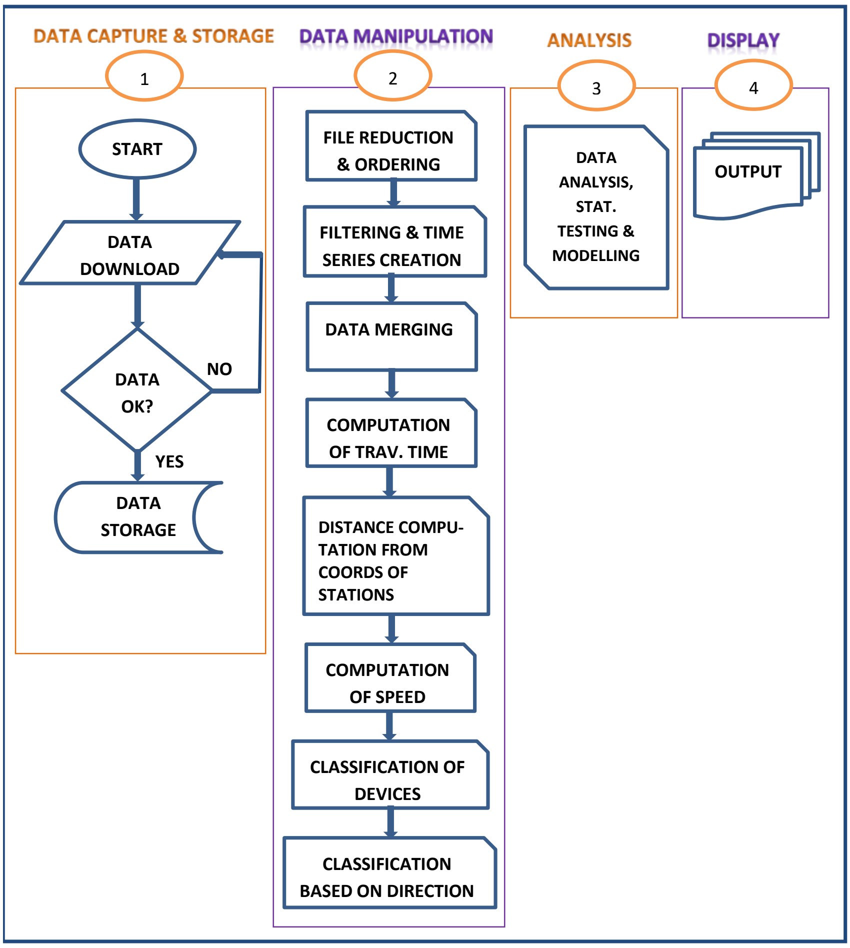


Figure 3.2: Bluetooth data processing algorithm design

3.3 Methods for Bluetooth Data Cleansing

3.3.1 The Rationale

Innovative traffic data collected from diverse sources such as Bluetooth, Twitter and a wide variety of other sources are often under-utilised (Ngoduy, 2013; Cook, 2014). This under-utilisation is believed to be primarily due to the problems inherent in the processing of these data to derive useful information. It is often challenging to analyse these data sets due to their enormity in volume and nature, leading to the frequent arrival of incoherent data in the database. As earlier stated in Sections 2.3.1 and 2.3.4, the data captured by Bluetooth consisted of MAC addresses from mobile phones, headsets and SatNavs carried by pedestrians, cyclists and onboard vehicles. This means that not all the Bluetooth devices detected by the sensors are from vehicular traffic. Also, a vehicle may have more than one Bluetooth enabled device on-board, and their location in the vehicle also influences their detection. For example, devices on the dashboard are 3.5 times more likely to be detected than when they are in a pocket or an obscured place (Stevanovic *et al.*, 2015). Therefore, the challenges in the Bluetooth data cleansing relate to the issues of reliability and consistency, representativeness, multiple detection, and outliers in the measurements. These important factors are considered in the next sections to address Bluetooth data cleansing.

3.3.2 Reliability and consistency of measurements

The reliability of MAC readers refers to successful detection of Bluetooth devices by the MAC readers (Stevanovic *et al.*, 2015). Reliability is also defined as the reciprocal of standard deviation (Bhaskar and Chung, 2013). This means that reliability can inform the knowledge of dispersion in the acquired data, and is in a way related to consistency that is determined by the precision (closeness) of one observation to the other in a group. Figure 3.3 presents an example plot of standard deviation of flows in both directions for weekdays' observation to underscore the importance of data cleaning before the final analysis. High reliability of a measure is determined through the ability to produce similar results under consistent conditions (Chen *et al.*, 2003). Therefore, for the estimated metrics to be reliable, the standard deviations computed under the same conditions must be similar (showing precision). This is in line with Shinya and Dragana (1999) that emphasised the need for the consistency of traffic volume data on different links of a network to ensure reliability. Accordingly, the estimated flows were filtered to remove the outlying values such as the spikes in the data using the Mahalanobis distance method. As in variability, the absence of consistency in data can influence measurements, analysis and in general the conclusion drawn (Lastdrager and Pras, 2009). It is to be noted that while reliable observations are consistent, the opposite cannot be said of consistent observations. Reliability is a function of a variety of factors such as location of the sensors, type of sensor, range and quality of the antenna used as well as the internal software settings such as the inquiry time.

Also, speed of the approaching vehicles, location of the Bluetooth devices within the car, and sensors' hardware and software can affect the reliability (Stevanovic *et al.*, 2015). Vehicles travelling at lower speeds are detected more reliably and omnidirectional antennae are detected more successfully (Stevanovic *et al.*, 2015). Successful detection in this context does not connote accuracy but the tendency to capture more Bluetooth devices that include non-vehicular modes. In particular, a longer range of detection zone is required to reduce random errors (Malinovskiy *et al.*, 2011). The closeness of vehicles to the sensor location also increases the rate of detection (Stevanovic *et al.*, 2015). In this research, the Bluetooth sensor used for the data collection has been configured to account for the range of detection and vehicle speed to reduce random errors, and to ensure reliability and consistency in Bluetooth detection. In other words, the manufacturer's settings of the sensors that include the inquiry time, second-by-second detection basis, and 93m detection range remain unchanged because this research has no control over the settings.

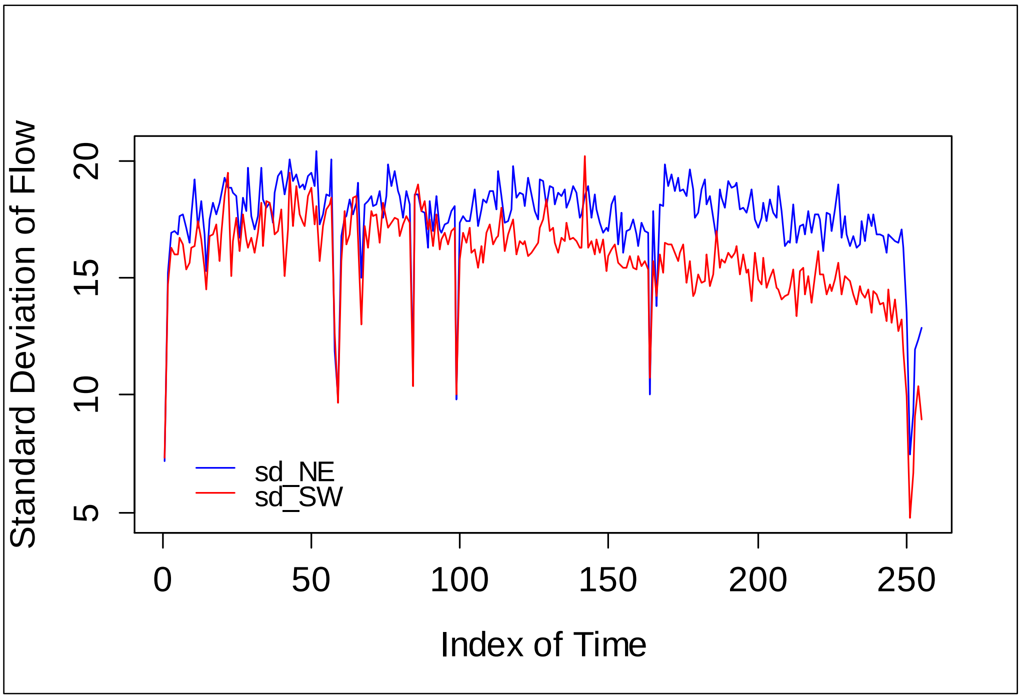


Figure 3.3: Standard deviation of flows in NE and SW directions for weekdays on Link0506

3.3.3 Representativeness of the measurements

Duplicate or multiple detection, and the location of the sensors can impact greatly on the representativeness of Bluetooth measurements by a way of introducing systematic or random errors. For instance, non-vehicular measurements captured by Bluetooth may be compared with traffic counts, which will give a false representation of the measurements. For example, if half of the detected devices are from non-vehicular sources, and are not removed from the data, the resulting estimate will be twice the actual vehicular detection. Blogg *et al.* (2010) referred to such representation as the capture rate and this is non-representative of the actual vehicular traffic. Other factors that could affect the representativeness of the estimation include missed detection – not all the devices can be detected while in the detection zone; and loss of information outside the detection zone, unlike the GPS method that could provide continous information throughout the journey. While these factors cannot be influenced after set-up, in this research, an appropriate data filtering that includes the removal of all error sources such as multiple detection, unrelialistic estimation, and outliers, is applied to ensure correct representativeness of the Bluetooth measurements. The filtered Bluetooth consisting of only the vehicular traffic is compared with the actual traffic count to obtain the detection rate. Literature shows that the current detection rate is greater than 2% of all vehicles, and it is considered a relatively modest sample size that is sufficiently large to provide a statistically robust performance evaluation (Hainen *et al.*, 2011; Hainen *et al.*, 2013). In this research, the validity of the Bluetooth representativeness shall be established using the package "samplesize4surveys" developed in R by Gutiérrez (2016).

3.3.4 Multiple detection

MAC noise arises from stationary and non-vehicular sources (Blogg *et al.*, 2010). However, appropriate data cleaning, extraction and aggregation are used to reveal the important information in a data set (Chang, 2014; Cook, 2014). This information includes travel time and speed to identify patterns and trends, and improving efficiency and safety within the road transport network. Araghi *et al.* (2015) investigated the effect of multiple detection using a three-antenna configuration. For a single MAC address, RSSI was used as the criterion to determine multiple detection. The study used the dwell time approach to classify duplicate records given that different antenna configurations were used. The time difference between the time of entry and exit of device at a detection zone gives the dwell time of the device. Based on this principle, duplicate records were removed from the data. The time of detection such as entry and exit times has also been used (Quayle *et al.*, 2010; Bhaskar and Chung, 2013). This research utilised the exit-to-exit and the dwell time approaches to deal with multiple detection. However, different antenna configurations as carried out by Araghi *et al.* (2015) could not be performed to test different scenarios due to the fact that the objectives of the Local Authorities that supplied the data used are independent of this research. Nevertheless, the dwell time approach is a valid method to identify multiple detection and invalid records. That is, any device with dwell time less than the average travel time of a link will be regarded as a duplicate record.

In the exit-to-exit approach, the time of last detection was used. While the dwell time approach utilised the *a priori* knowledge of the network to set travel time limits based on two conditions: (i) on a short link, say a length of 0.154km (minimum within the network) and at 48km/h speed limit, which corresponds to a travel time of 11.55 seconds, if the dwell time is less than 4 seconds (which allows a margin of error for possible delay in the actual detection) it is a duplicate; (ii) on a long link, say a length of 7.463km (maximum within the network) and at 48km/h speed limit, which corresponds to a travel time of 559.73 seconds, if the time difference between successive unique vehicle records was less than 300 seconds (also to allow a margin of error given that the data will be filtered), it is a duplicate because it is not expected that a vehicle would have made a return journey at less than such a travel time. The assumption here was that such a tracked device was either from a parked vehicle or vehicle with a stop over. Based on this principle, duplicate records were removed from the data.

3.3.5 Outliers

A detailed review of outlier detection methods can be found in Stavig and Gibbons (1977) and Seo (2002). However, an outlier is a value that deviates markedly from other observations in the same sample (Hodge and Austin, 2004). It is to be noted that all outlier detection methods have their strengths and weaknesses. Different outlier detection methods have been applied for Bluetooth data cleansing. For example, the Box-and-Whisker method has been used by Tsubota *et al.* (2011) while Kieu *et al.* (2012) combined Box-and-Whisker and MAD (Median Absolute Deviation) methods for outlier filtering. This research considers the Tukey's method (Box-and-Whisker) in conjunction with the Mahalanobis distance (MD). The Tukey's method defines outliers as values greater than $Q3 + 1.5 * IQR$ and values less than $Q1 - 1.5 * IQR$, where $Q1$, $Q3$, and IQR are the lower quartile, upper quartile, and inter-quartile range respectively (Crawley, 2005). This method is resistant to extreme values and is robust in handling large normal data, but is problematic with small data samples (Seo, 2002). The MD method as demonstrated by Warren *et al.* (2011) is robust to failures of assumption, flexible and incorporates both numerical and graphical outputs. The MD method implemented in this research utilises the chemometrics package in R (Filzmoser and Varmuza, 2013). The choice of the combination of the Tukey and MD methods is based on the recommendation of Warren *et al.* (2011) that any serious analysis of traffic or other pattern should utilise more than one technique. The robustness check is also necessary to avoid possible spurious outliers driving the model results as highlighted by Sebri (2016). Warren *et al.* (2011) have also shown that MD is very useful in analysing traffic volume data irrespective of the underlying assumptions. The traditional limitation of the MD is that it cannot be calculated if the number of variables exceeds the sample size due to the inverse of the weight matrix as shown in equation 3.1 (Brereton, 2015). However, this limitation is not obtainable in this research because the sample sizes obtained are far greater than the number of variables. Accordingly, the MD was used in this research to check for multivariate outliers and to account for differences in scale and variance of each of the variables in the data in line with Mahalanobis (1936) and Starkweather (2013). Mahalanobis distance is defined as (Brereton, 2015):

$$d_m = \sqrt{(x-\mu)S^{-1}(x-\mu)} \qquad (3.1)$$

Where $(x-\mu)$ is a matrix of distance from the mean, and S^{-1} is the inverse of the covariance matrix.

For the illustration of the MD method of outlier detection, Figure 3.4 presents the plot of Bluetooth flows against the Mahalanobis distances. The dotted line signifies the cut-off point (2.457) for determining outlying values. In the implementation, the R code based on the Moutlier function in R package Chemometrics was cross-checked in Minitab to ensure the results are free from systematic errors and blunders. Figure 3.5 and Figure 3.6 present the density plot and square of the MDs against Chi-square values. The concept is that the square of MDs has a Chi-square distribution with p degree of freedom, and when the sample is large, the MDs have approximately Chi-square distribution (Penn State Eberly College of Science, 2016). The expectation is that for a multivariate normal distribution, the plot of MDs against Chi-square distribution should follow a straight line while the density plot should be approximately normal. Also, outliers are classified as points with significant difference between the MDs and the Chi-square, and are shown at the upper right corner (Penn State Eberly College of Science, 2016). Figure 3.7 presents another application of the MD method in detecting outliers in a two-dimensional plot (scatter plot) using Bluetooth/ANPR journey times for illustration.

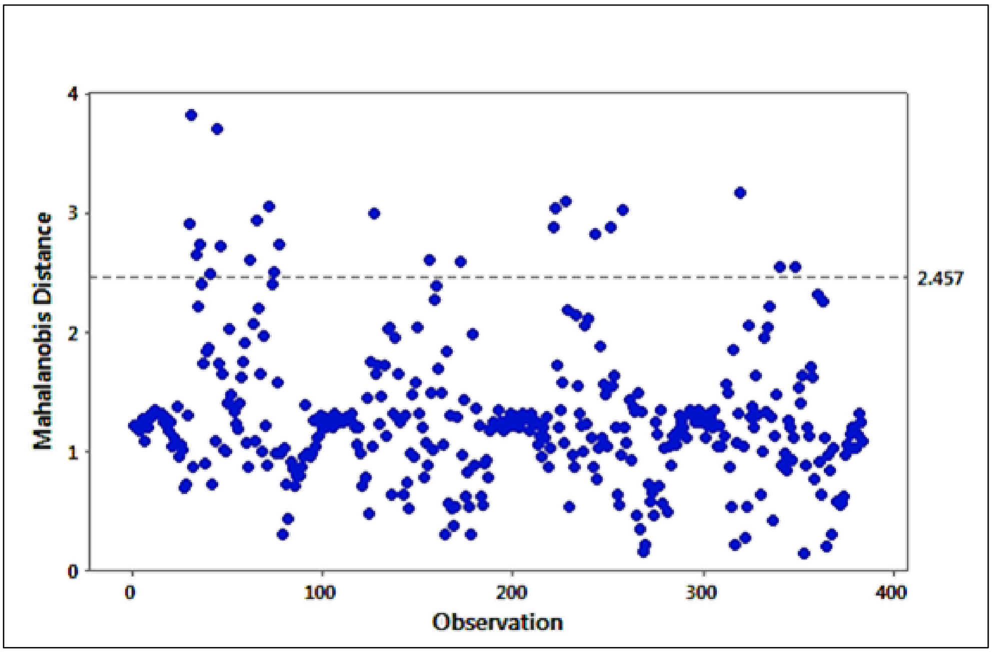


Figure 3.4: Plot of flow against Mahalanobis distance showing outlying points

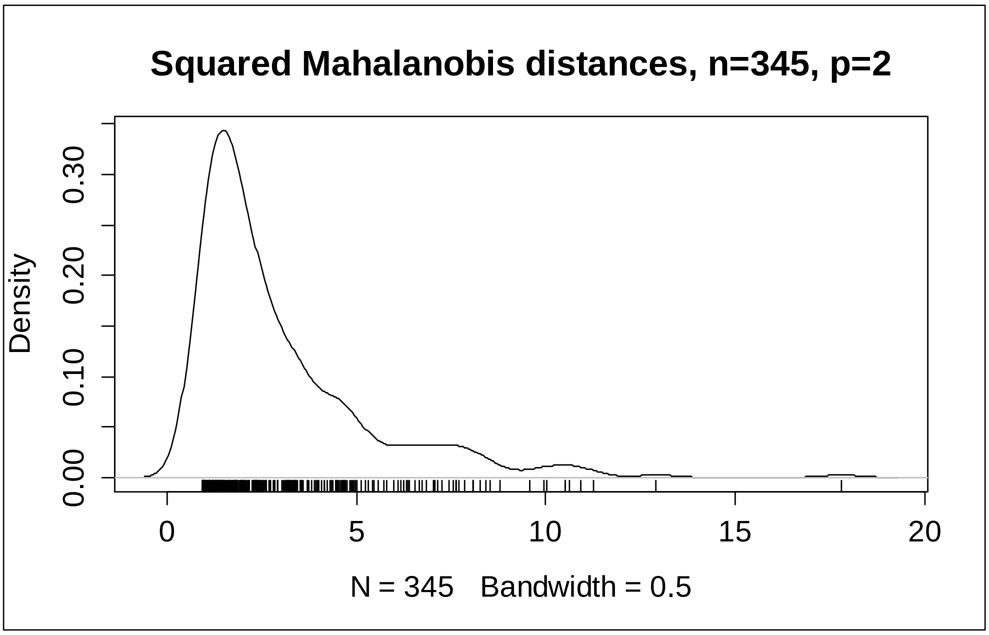


Figure 3.5: Density plot of Mahalanobis distances of 2-degree of freedom

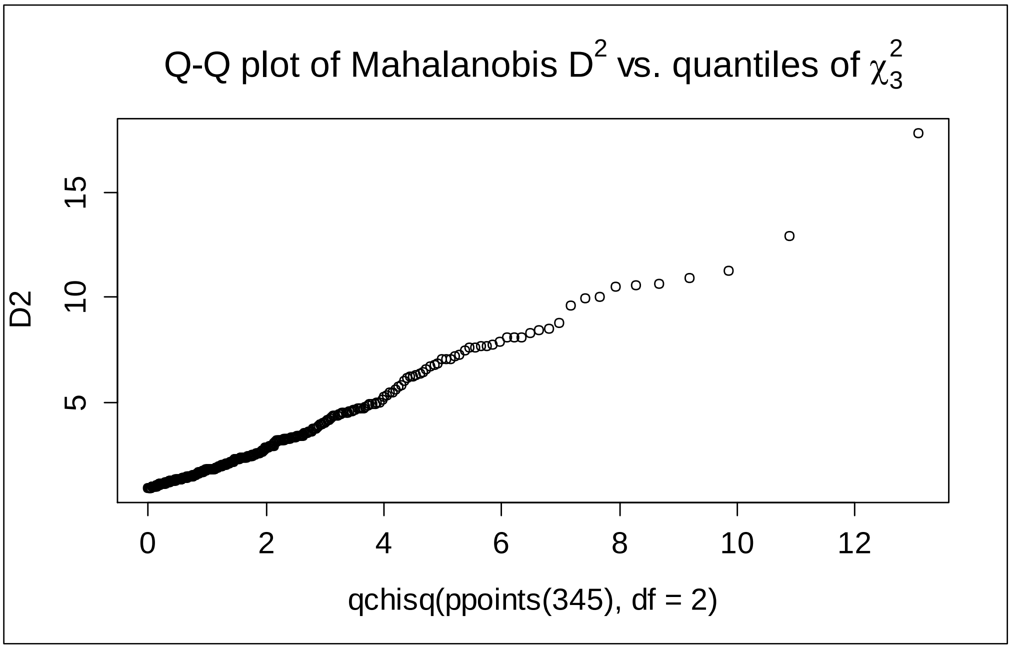


Figure 3.6: Q-Q plot of square of Mahalanobis distances against Chi-square of 2-degree of freedom

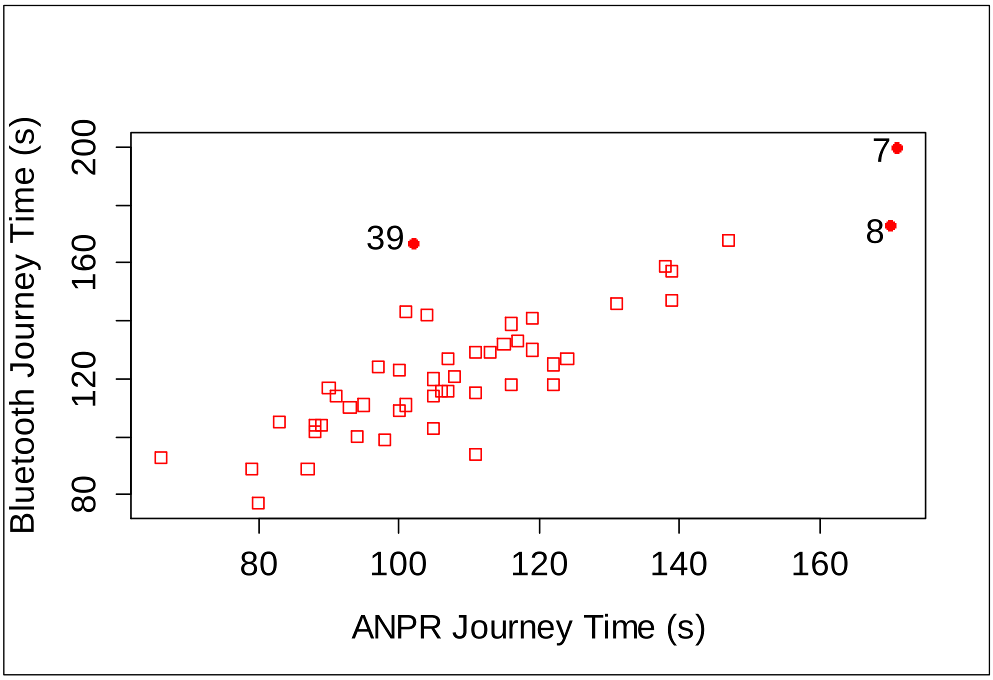


Figure 3.7: Plot showing outlying points in Bluetooth and ANPR journey time based on Mahalanobis distance method

3.4 Estimation Methods of the Traffic Metrics

3.4.1 Estimation of flow

The fundamental approach to determining travel time and other variables useful for inferring traffic conditions, such as flow and speed, using the Bluetooth approach has been previously discussed (Abbas *et al.*, 2013; Barceló *et al.*, 2013; Bhaskar and Chung, 2013). This concept is based on the re-identification principle from which travel time and other parameters can be estimated (UMCATT, 2008; Young *et al.*, 2013). The same concept is utilised in this research to estimate traffic metrics using the information collected by the Bluetooth sensors in the chosen study sites. Therefore, the term 'estimate' in this context refers to the use of Bluetooth information to determine or calculate an approximation for traffic metrics. The relevant codes for the estimation of the traffic metrics are contained in Appendix 3B (R-codes for Bluetooth data processing). Using Bluetooth, a single detector can detect and record the information on vehicles travelling in one or both directions in a road network. However, the challenge is that the data captured have no unique variables or parameters which identify the direction of travel of the detected devices. Therefore, the direction of travel of the devices cannot be differentiated using a single Bluetooth detector. However, when combined with data from another detector for example, the link-flow can be obtained through the separation and classification of the devices into their directions of travel. Table 3.4 presents the two (device and direction) main categories of classification performed by TRAFOST through data filtering. The first category is the device classification. As previously mentioned under the assumptions made in the algorithm design (Section 3.2.5), pre-defined minimum and maximum boundary limits (6km/h and 120km/h) were set based on the *a priori* knowledge of the road network and walking speed to remove outliers. Therefore, devices travelling at a speed greater than the upper limit are said to be an emergency vehicle, traffic violator, or error in the data such as an encryption error in the MAC addresses. Similarly, devices travelling at speeds less than the lower limit are said to be pedestrians, parked vehicles (vehicle stop over), vehicles making use of a bypass or alternative route, and vehicles possibly not detected on time, particularly on a short link. Devices in these categories were excluded from the metrics estimation as an initial step in the outlier removal process following the exclusion of duplicate records to obtain reliable results.

The second classification involves the direction of travel given that at each station, vehicles travelling both ways were detected and recorded together. With the stations' data merged, devices were classified into the direction of travel according to whether the computed travel time was positive or negative. Devices with a positive time difference were those travelling from origin-to-destination, and devices with a negative time difference were those travelling from destination-to-origin. This basic principle was used to group detected devices into directional clusters within the networks. Accordingly, the individual link-flows of the detected Bluetooth devices as they passed the detectors were estimated to provide the time series records of flow at different temporal dimensions; 5, 10, 15 and 60-minute averages as well as daily, weekly, weekday, and monthly averages. In the future, handling this problem may become more simplified with further technological advancement to improve efficiency in automation and computation.

Classification Types	**Classes**	
Device	Vehicles	Non-vehicles
Direction of Travel	Origin-to-destination	Destination-to-origin

Table 3.4: Summary of device and directional classifications

3.4.2 Estimation of travel time

The Bluetooth traffic monitoring approach makes use of the principle of identification and re-identification of vehicles at different stations within the road network to calculate travel times by matching MAC addresses at successive stations (Biora *et al.*, 2012; Young *et al.*, 2013). The time difference of the matched MAC address provides a measure of travel time and space mean speed determined by the link length between the successive stations within the detection zones (Haghani and Hamedi, 2013). For a reliable travel time estimation, Quayle *et al.* (2010) suggested the use of “ex-ex” (exit-to-exit) and “en-en” (entry-to-entry) detection time. Bhaskar and Chung (2013), on the other hand, recommended ex-ex travel time due to the delay observed at the upstream intersection. This recommendation is considered in this research in the estimation of travel time. The travel time between two stations, A and B, is given as (Bhaskar and Chung, 2013):

$$T_{AB} \quad = \quad TT_{Ex2Ex} = TT'_{Ex2Ex} + (\varepsilon_{D,{}^{d}/_{s}} - \varepsilon_{D,{}^{u}/_{s}}) \qquad (3.2)$$

Where TT_{Ex2Ex} and TT'_{Ex2Ex} denote the actual and the estimated travel time respectively, and $\varepsilon_{D,{}^{d}/_{s}}$ and $\varepsilon_{D,{}^{u}/_{s}}$ are the error terms at the two stations. The error terms are from the possible delay in the detection of a device due to the inquiry time. However, if the magnitudes of the errors are the same, then the estimated and actual travel time are the same.

3.4.3 Estimation of vehicle speed

The basic principle in time and speed calculations is that a vehicle with a unique MAC address detected at two different sensor stations (say A and B) separated by distance S_{AB}, metres will have travel time T_{AB} (T_B – T_A), seconds defined as equation 3.2 and speed V_{AB}, m/s between A and B expressed mathematically as follows: $V_{AB} = \frac{S_{AB}}{T_{AB}}$ (3.3)

Where V_{AB} is the average speed of a device from point A to B

S_{AB} is the network-based distance between stations A and B, and

T_{AB} is the time difference of the detection of the device at B and A. Where the network-based lengths are not available, they are measured on Google Earth using the path measuring tool. The measured link lengths were preferred to the computed lengths due to the curvature error in the computation, particularly where the road nature is irregular.

3.4.4 Estimation of O-D matrix

As stated earlier, the Bluetooth approach offers a direct method of sample estimation. However, using Bluetooth data collected from one station in isolation from another replicates the traditional link detector capability. The limitation in a single detector can be addressed by matching the MAC addresses between all the Bluetooth detectors across a network to create the origin and destination (O-D) information. In this research, an indication of O-D patterns within the areas of study was obtained by identifying and matching the same MAC addresses at different locations over the network. That is, the concept of the flow estimation described earlier, was applied to an area-wide network of Bluetooth array to estimate the network O-D matrix. Two types of O-D matrices classified as ‘one-many’ and ‘many-many’ according to the road network design and purpose, were estimated. In a one-many estimation, a reference station was chosen from where the origin-to-destination information is computed. On the other hand, the many-many estimation encompasses the computation of O-D information in both directions (origin-to-destination and destination-to-origin) across all the stations to obtain complete information about the network. Using the ‘igraph’ network analysis package in R, a typical O-D matrix was represented to show directional flow information. In the representation of the estimated O-D matrix, a one-headed arrow indicates one-way flow while a double-headed arrow indicates flows in both directions.

3.4.5 Estimation of detection rate

Bluetooth detection rate refers to the proportion of traffic captured by Bluetooth sensors compared to the actual traffic (Biora *et al.*, 2012; BlipTrack, 2012). Recall that Bluetooth does not immediately give the actual estimation of the traffic but a proportion of the traffic. Therefore, the detection rate is required to scale-up the Bluetooth sample of the traffic flow to the actual vehicular traffic obtained from the ground truth data. This metric is computed as the ratio between the estimate of Bluetooth flows and the corresponding SCOOT and ATC flows collected over the same period and location. If a regression analysis is performed between the two sets of data (estimated and actual), detection rate is obtained as the slope, β_i of the regression equation ($y_i = \beta_0 + \beta_i x_i + \varepsilon_i$). The use of the flows collected from the SCOOT and ATC links to determine the detection rates provide the opportunity to understand variability arising from relative location of the Bluetooth sensors to the ground truth sources. The computed ratio over different temporal dimensions were analysed to obtain the most probable value (mpv). The theoretical implication of mpv is that the estimation presents the best approximating values and not the actual value. That is, the actual value of the total traffic remains unknown. The hypothesis testing for variance in the detection rates was based on Bonett's test and Levene's test in Minitab to understand directional differences. These tests are used given that they give a type I error that is close to the specified significance level (α). They also allowed for a balance in sample size and skewness in the distribution.

3.5 Validation Methods

3.5.1 Model (TRAFOST) validation

This considers the steps taken in the validation of TRAFOST before considering the results generated using the model. Three steps are followed to accomplish this. The first step consists of results comparison between the model and the manual computation; while the second step involves the use of the output of C2-Web software. The last stage consists of cross-validation using the outputs of the model. Following these steps, the Bluetooth estimated metrics are validated using the ground truth data sets.

3.5.2 Results validation using diverse independent data sources

The availability of diverse sources of independent measures of traffic enabled both rigorous and sound validation of the model outputs. The estimated metrics computed by TRAFOST will be validated as follows:

i. The Bluetooth-derived traffic flows are validated using real-life traffic data obtained from the simultaneous observation of ATC and SCOOT flows at the same location. The use of independent data sets for the validation was performed over 2013 at different locations within the three networks in Manchester to demonstrate transferability in the Bluetooth approach.
ii. The estimated travel times and speed computed using TRAFOST will be validated using data from Traffic Master, consisting of six months' (April - September 2013) hourly averages covering four links in Stockport and Trafford. The Traffic Master data was complimented with ANPR data of 1-day in Stockport for further validation. However, while the 1-day ANPR data may be considered insufficient, it should otherwise be noted as an added advantage because its absence will not have had any effect on the conclusion of the results. The estimated O-D matrix, on the other hand, will be validated through repeatability using six months' worth of data over the three locations in Greater Manchester. The exercise was conducted primarily to test for consistency and variability in the estimated matrices as well as to evaluate the robustness of the model in handling large volumes of data.

The integration of the other sets of data with Bluetooth data for the validation exercise is essential as Bluetooth data presents only a sample that is lower than the actual traffic flow. However, the lower sample is expected because not everybody and all modes within the network have Bluetooth-enabled devices; and when they are switched on, the Bluetooth may not be enabled. Table 3.5 presents the summary of the methods of results validation. The comprehensive results validation and testing are presented in Chapter 5. The appraisal of the situation started with scatter plots to explore correlation. Edwards and Hamson (2001) advised that an alternative model formula must be considered if the linear fit is poor with low R^2 and noncolinearity suggesting an invalid proposition. Finally, considerations were given to the sites where simultaneous measurements of the Bluetooth and ground truth data are possible.

Bluetooth-derived metrics	Method of Validation	Period of the Data used
Link Flows	ATC and SCOOT flows	One year
Journey Time	ANPR and Traffic Master (TM)	Six months for TM and 1 day for ANPR
Journey Speed	ANPR and Traffic Master (TM)	Six months for TM and 1 day for ANPR
O-D Matrix	Repeated Measurements of O-D	Six months

Table 3.5: Table showing the methods of results validation

3.5.3 Statistical modelling of the Bluetooth estimated metrics

Characterising a time series data not only includes the estimation of mean and standard deviation but also the correlation between observations separated in time (Statgraphics, 2015). Time series models come in useful when dealing with serially correlated data. The serially correlated errors can be written as (Fox and Weisberg, 2010):

$$C(\varepsilon_t, \varepsilon_{t+s}) = C(\varepsilon_t, \varepsilon_{t-s}) = \sigma^2 \rho s \qquad (3.4)$$

Where ρs is the error autocorrelation at lag s. This research utilised Autoregressive Integrated Moving Average ($ARIMA$) models, being one of the two most widely used approaches for time series forecasting, and the models describe the autocorrelations in the data (Hyndman and Athanasopoulos, 2013). $ARIMA$ models can be estimated following the Box-Jenkins approach (Quddus, 2008), while the non-seasonal $ARIMA$ models are generally denoted as $ARIMA(p, d, q)$. The parameters p, d, and q represent the order of the autoregressive part, the degree of differencing, and the order of the moving-average model, and are non-negative integers (Cowpertwait and Metcalfe, 2009; Fox and Weisberg, 2010). The special cases of $ARIMA$ models such as autoregression – $ARIMA(p, 0,0)$, moving average $ARIMA(0,0, q)$ are presented in Hyndman and Athanasopoulos (2013). Combining Differencing, d with Autoregressive, $AR(p)$ and a Moving Average, $MA(q)$ model gives the following full $ARIMA(p, d, q)$ model:

$$y'_t = c + \varphi_1 y'_{t-1} + \cdots + \varphi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \qquad (3.5)$$

where y'_t is the differenced series; φ_i = parameters of the autoregressive part; θ_i= parameters of the moving average part; e_t = error terms; c = expectation of the model; and $i = 1\ to\ p\ and\ q$ respectively.

Using the backshift notation, equation (3.5) can be written as:

$$\left(1 - \varphi_1 B - \cdots - \varphi_p B^p\right)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)e_t \qquad (3.6)$$

AR(p) Differencing (d) MA(q)

As a first step in the modelling, the data used were prepared for conformity, and separated into two for calibration and validation through a data splitting technique. The separation of the calibration and validation datasets utilised the "80/20" rule using the Caret package in R. The 80/20 rule means that 80% of the data is used for calibration while the remaining 20% is used for validation (Brownlee, 2014). However, to build a model, the issue of stationarity of the series is essential to avoid any predictable patterns in the long-term (Fox and Weisberg, 2010). Therefore, the next step explores the data for stationarity through time series plots. The non-stationary series were stabilised through transformation, detrending, and differencing as highlighted by Hyndman and Athanasopoulos (2013). In addition to exploring the time plot, the ACF (autocorrelation function) and PACF (partial autocorrelation function) plots are also used to determine the parameters of the models. The 'auto.arima' function in the forecast package in R was also used to determine these possible values, while the adequate model selection utilised Akaike's Information Criterion (AIC) and personal judgement. AIC proposed by Akaike is an extension of the classical likelihood principle, and it is based on Kullback-Leibler information or distance as a fundamental basis for model selection (Burnham and Anderson, 2002). Using the AIC, the $K-L$ information computed for each model in the set helps in determining the most probable predictive model $(MPPM)$ given as (Burnham and Anderson, 2002):

$$AIC = -2\log\left(L(\hat{\theta} \mid y)\right) + 2K \qquad (3.7)$$

Where K = number of estimable parameters, and the expression $\log\left(L(\hat{\theta} \mid y)\right)$ is the numerical value of the log-likelihood at its maximum point. AIC provides a simple, effective, and objective means of model selection for both data analysis and inference. That is, AIC takes care of model parsimony (the principle of the simpler the better), and is therefore considered. As a final step, a portmanteau test was performed to understand whether differences in the group of autocorrelations are different from zero with a return of large p-value signifying white noise residuals according to Fox and Weisberg (2010). In summary, the following steps outlined by Srivastava (2015) are followed in the modelling. The steps are: i) visualise the series; ii) make the series stationary; iii) plot the ACF/PACF and find optimal parameters; iv) build the ARIMA model; and v) make predictions.

3.5.4 Exploratory and quantitative data analyses

In data analysis, the understanding of the distribution of the data is crucial to avoid invalid inference (Dixon and Massey, 1983). The normal distribution is considered in this analysis given its importance in statistics. Not only that, the hypothesis tested in this research is dependent upon the validity of normality and randomness of the residual errors. Another usefulness of the normal distribution to this research is in understanding the sampling distribution given that the Bluetooth-estimated traffic flow is a sample of the actual flow (population). Therefore, the first phase of the analysis explored the understanding of the distribution of the data. The analysis utilises quantile plots in conjunction with histogram plots. Examples of such plots are presented in Figure 3.8 (quantile plot) which suggests non-normality in the data distribution, while Figure 3.9 showing the histogram plot of journey time suggests a normality of distribution of the journey time data. This normality in the distribution informs the use of parametric methods. Based on the literature, a test based on the mean provided the best power for symmetric distributions with moderate tails (Brown and Forsythe, 1974). That is, the power of a test is the probability of not committing a type II error (error due to failure to reject the null hypothesis when it is false) (Minitab, 2014). However, the non-parametric approach was preferred given that the presence of an outlier in the data may invalidate the test result (Tukey, 1980). Again, the non-parametric technique was adhered to for the purpose of consistency. Dobson and Barnett (2008) highlighted the importance of giving consideration to separate analysis, which includes the understanding of the measurement scale, the shape of the distribution and the association within variables. Burnham and Anderson (2002) noted that deep thinking and exploratory data analysis (EDA) will result in good scientific questions and confirmatory data analysis; Tukey (1980) concludes that to properly implement the confirmatory hypothesis there is a need for extensive exploratory work such as histogram, box-and-whisker based on four features (location, dispersion, skewness, and potential outliers) and quantile plots to explore distribution and normality assumptions (Open Learn, 2015). To check for quality, each data was analysed separately as posited by Dobson and Barnett (2008) and Burnham and Anderson (2002). This premise forms the basis for employing both quantitative and exploratory statistical techniques in this research to properly implement the confirmatory hypothesis. For clarity, a ‘standard normal’ is given by (Acevedo, 2013, page 69) as:

$N_{0,1}(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$. While a normal $N_{\mu,\sigma}(x)$ is standardised to $N_{0,1}(z)$ by subtracting the mean and dividing by the standard deviation and is given as:

$Z = \frac{x - \mu_x}{\sigma_x}$. The standardisation could as well adopt methods such as specifying a range for the minimum or maximum (Minitab, 2014).

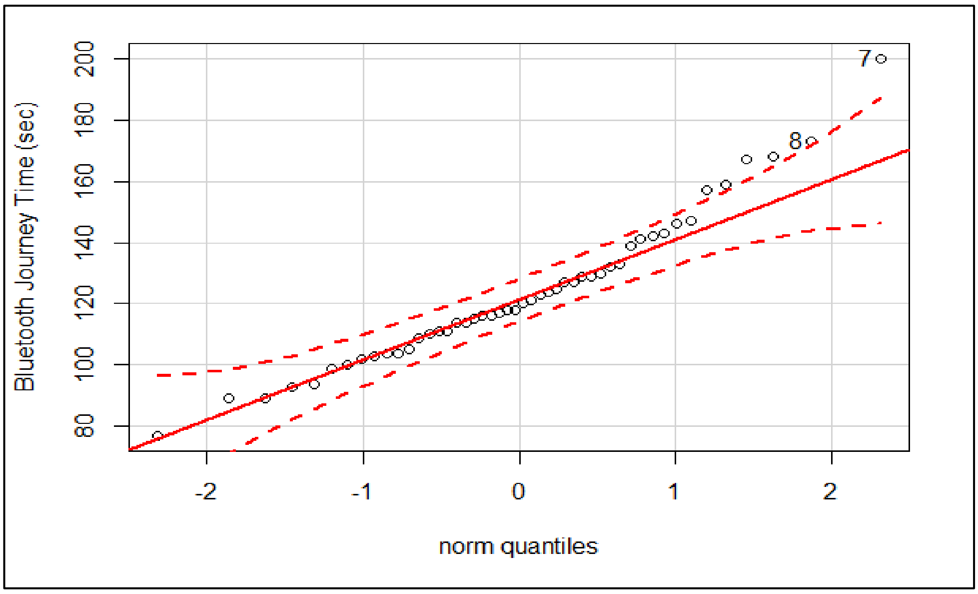


Figure 3.8: Quantile plot showing non-normality in distribution for Bluetooth journey times on Link7170 in Stockport on 3rd April 2014

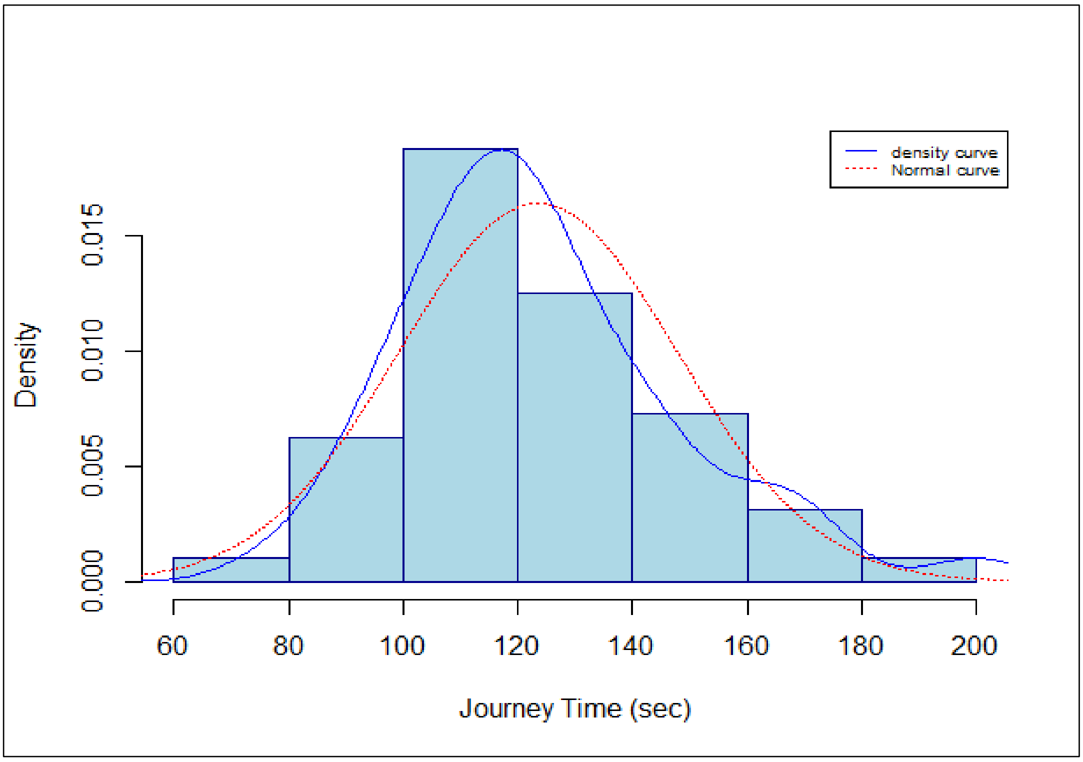


Figure 3.9: Histogram plot of Bluetooth journey time overlaid with normal and density curves on Link7170 in Stockport on 3rd April 2014

3.5.5 Relevant measures of variability in the data

Variability is a measure of spread in data and is used to understand similarities or differences in a data set (Dixon and Massey, 1983; Adedayo, 2006). Wedagama *et al.* (2007) highlighted the effect of variations in traffic flows and speeds over congested and uncongested periods in an urban area such as the GMN. Variability causes uncertainty and unpredictability and may affect the conclusions drawn. In transportation, this uncertainty is a major concern for both operators and commuters (Martchouk *et al.*, 2011). Tsekeris and Stathopoulos (2006) noted that the measurement of the spatial and temporal variation in traffic flow is a major issue in tackling the analysis of network congestion problems. Normally, a good way to start is to use the “range” of the distribution (Adedayo, 2006). However, since the range is subjected to extreme values and does not account for every value in the distribution, alternative statistics are considered and compared together to obtain a more reliable result. In this case, the variance and standard deviation are considered. These were chosen in lieu of mean absolute deviation (MAD) given that further statistical analysis can be performed on them, unlike the MAD that is based strictly on absolute values (Adedayo, 2006). The MAD considers the spread in the data, but it is affected by extreme values and is similar to standard deviation. Therefore, the best approximating value of any measurement is the mpv. It is the value that minimises the sum of the squares of the residuals and it is defined as the arithmetic mean given as equation (3.6) (Whyte and Paul, 1997).

$$mpv = \frac{\sum x}{n} \tag{3.8}$$

The precision (standard deviation) is given as: $\sigma = \sqrt{\frac{\sum v^2}{n}}$ (3.9)

The standard error of mean as: $\sigma_m = \frac{\sigma}{\sqrt{n}}$ (3.10)

According to Cooper (1974), the precision of a measurement is quoted as: $x = mpv \mp \sigma_m$. Another useful statistic considered in the evaluation of variability is the coefficient of variation (cv) defined as (Adedayo, 2006):

$$cv = \frac{\sigma}{\bar{x}} * 100\% \tag{3.11}$$

The coefficient of variation describes dispersion without dependence on the unit of measurement of the variable (Adedayo, 2006). This statistic was used to compare the spread in the detection rate distribution across GMN. The cv helped to account for geographical variations or changes in units over space since it was measured in percentages. Its application in comparing two sets of data is that the one with the smaller cv is the better of the two (Adedayo, 2006).

Since PCA is a variable reduction procedure (Minitab, 2014), it was used to reduce the measured daily flows to develop smaller numbers to account for most of the variance in the observed daily flows. This method not only provides the numerical values, but also graphical outputs for visualisation to enhance interpretation. The knowledge of data reduction is needed for optimisation and efficiency in traffic flow modelling to avoid redundancy. Analysis of variance (1-way ANOVA) was employed for post-analysis to further explore any significant variations among groups (the speed metrics). ANOVA is considered given that their distributions generally obey the parametric assumptions. The hypothesis testing for the post analysis utilised the Tukey test ($\alpha = 0.05$). The importance of accurate classification of the metrics can be found in model optimisation for improved efficiency. That is, it can help to determine when it becomes significant to change the traffic management plan such as in the timing of traffic lights.

3.5.6 Relevant measures of accuracy

This section discusses the key accuracy statistics and the quality measures utilised in this research that include both the absolute and relative metrics such as the mean absolute error (MAE), and root mean square error $(RMSE)$ (useful for the adjustment of unusual large errors) (Wood, 2012). Accuracy is defined as the closeness of the value of a measurement to the 'true' or theoretically correct value (Cooper, 1974). This was determined principally using quantitative analysis as well as time series plots to compare trends in the profiles. High levels of temporal similarity in trends and good performance metrics will inform reliability in the data. Using the accuracy statistics, small values close to zero were of good fit, while observations with a small standard error were of higher accuracy than observations with a big standard error. Different combinations of these metrics have been used in the past. For example, Tang *et al.* (2016) used the combination of MAE, $RMSE$ and $MARE$ (mean absolute relative error). According to Hyndman and Athanasopoulos (2013), the MAD (mean absolute deviation) is just another name for the MAE. The MAD was used in this case to compare similar models. The relative metrics include the mean absolute percentage error $(MAPE)$, mean percentage error (MPE), and mean squared percentage error $(MSPE)$ (Balcilar, 2007). In line with Sebri (2016), the $MAPE$ was used in this research because it is scale-invariant in order to account for the different locations and periods. Also according to Hyndman (2006), the mean absolute scaled error $(MASE)$, which is equally scale free, was used to avoid the problem of infinity (due to division by zero) or large value (due to presence of small numbers) in $MAPE$. Correlational analysis was employed to measure the linear association between the data sets, and a correlation coefficient $(r \geq 0.80)$ was considered to be a good relationship. Some of the accuracy metrics used in this research are hereby mathematically defined:

$$MAE = mean(|e_i|) \tag{3.12}$$

$$MAPE = mean(|p_i|) \tag{3.13}$$

Where $e_i = y_i - \hat{y}_i$ is the error term defined by the difference between the observed and the adjusted values; and $p_i = 100e_i/y_i$ is the percentage error.

For the statistician, the difficulty in discriminating between two populations with the best test determines their differences (Kullback and Leibler, 1951). The Kullback-Leibler distance (KL-D) measures the distance between two probability distributions to address the problem of scales in two random variables (Allison, 2016). It is a measure of the difference between two probability distributions, or a measure of dissimilarity or departure between two distributions (Wu, 2016). If the distributions are similar, the KL-D should be small, and it should be large if the distributions are far away from each other (Wu, 2016). That is, KL-D can be used to measure the quality of an estimation, and was used in this way. Note that the KL-D is generally not symmetric (Allison, 2016). Therefore, it is called a divergence instead of distance. KL-D is expressed mathematically as:

$$KL(p||q) = -\sum p(x) log \frac{q(x)}{p(x)} \tag{3.14}$$

Where $KL(p||q)$ is the KL-D relative to p; p represents the “true” distribution of the observation while q represents an approximation p. In this case, p and q correspond to the ground truth and Bluetooth data.

3.6 Conclusions

A description of the research methodology based on the Bluetooth approach to traffic metrics estimation was presented in this chapter. Bluetooth data captured from a road network consists not only of the devices from vehicular traffic but also from other sources such as pedestrians and cyclists. The raw data captured contain errors due to these different sources, and the mode of measurements such as multiple detection and inquiry time. Therefore, the methods of Bluetooth data cleansing to obtain a noise free data necessary for reliable traffic metrics estimation was discussed. This stage leads to the next step of estimation of traffic metrics that include flow and travel time. This chapter covered the relevant stages required for a reliable traffic metrics estimation for traffic management applications. While the methodology described was based on a post-processing approach, it could be adapted for a real-time application. A functional model termed TRAFOST was developed in R to automate and optimise the Bluetooth data processing and analysis. The use of diverse independently measured traffic data (ground truth data) for results validation was to ensure robustness in the analysis, and to establish the validity of the Bluetooth results. The validation methods for the estimated flows, travel times and speed were based on the ground truth data, and statistical modelling. The validation of the network O-D matrix was based on repeated measurements of Bluetooth data to understand consistency given that the ground truth data are not available on every link. In the research design, five different study sites of varying attributes over different geographical locations in the UK were considered due to different challenges encountered in the data acquisition. Basing the research design on more than one study site ensured the knowledge of transferability to inform results generalisation. It is noted that as with every model, the concept developed in this research is limited with its range of validity. Therefore, consideration should be given to the replication of this concept at a new study site to obtain reliable results.

Chapter 4. Data Collection

4.1 Introduction

4.1.1 Background to the data collection

Chapter 4 describes the data collection and short-term analysis performed to understand the Bluetooth potential in traffic metrics estimation. The data collection and short-term analysis provide the basis for the fulfilment of research objective iii, and to test the methodology described in Chapter 3 on Bluetooth traffic sensing and metrics estimation. Accordingly, this dicussion covers the challenges, limitations and the specific methods used in the collection of Bluetooth data in this research. The data collection covers three select study sites (Liverpool, Birtley, and Manchester) of different network attributes. This study covers three different study sites primarily due to the shortfalls encountered in the provision of the required data in the Liverpool and Birtley study sites. However, the different study sites have contributed in different ways that include the understanding of the spatial variability in Bluetooth usage in the UK. The Liverpool and Birtley studies cover a short-term data collection period over two weeks. The former was used for preliminary data quality assessment and the latter for flow and trip pattern analysis. The Manchester study site, which provided the long-term data collection of more than a year consists of three separate studies in Wigan, Stockport, and Trafford, and consolidated the Birtley study to demonstrate transferability. The preliminary analysis conducted in this chapter presents the initial understanding of the Bluetooth approach in different road networks such as urban arterials and linear networks. The data collection and the preliminary analysis also form the basis for the long-term study presented in Chapters 5-7 of this thesis to enable valid conclusions.

This chapter is structured as follows: Section 4.2 focuses primarily on preliminary data quality assessment (Liverpool pilot study) to understand the condition of the variables contained in the Bluetooth data as a first step towards understanding its relevance. Section 4.3 presents the evaluation platform for the estimation of flow and analysis of trip patterns using the Birtley pilot study.

Section 4.4 presents a description of the Manchester study sites by building on the methodology tested in the Birtley study. Section 4.5 presents trip patterns and speed distribution based on the Wigan study, while Section 4.6 (Stockport) considers the discrepancies in results based on the applied methods. Section 4.7 (Trafford) deals primarily with monthly variation, before conclusions are drawn in Section 4.8.

4.1.2 The study sites

This investigation was carried out in urban areas in the UK comprising Birtley in Tyne and Wear, Liverpool and Greater Manchester (consisting Wigan,Stockport and Trafford). This makes a total of five different networks of varying characteristics. These study sites were selected mainly based on data availability to meet the data requirements described in Section 3.2.1. That is, data were collected in these sites to meet the requirements of the research design to achieve the overall aim and objectives of this research. The distinguishing features of these study sites are primarily in their network configuration, the land use type and the area covered. The total length of the Birtley network is approximately 2km with seven Bluetooth stations while the Liverpool network covers 2.2km with eight Bluetooth stations. The three networks in Manchester are in residential and commercial areas and over larger areas (approximately 50km x 40km with forty Bluetooth stations) compared to Birtley and Liverpool study sites. The different geographical areas of dissimilar attributes considered in this research are important to understand variability in performance and transferability of Bluetooth approach of traffic monitoring. This will in turn provide the knowledge of the scalability of the technology over the study areas, and the UK in general. As described in Section 4.1.3, several factors brought about the study sites used in this research. Table 4.1 presents the summary of the description of the study sites, while Figure 4.1 shows their respective locations colour-coded on a UK map.

Location	Number of Case Studies	Case Study	Features	Number of Bluetooth Stations
Liverpool, UK	1	Liverpool	Short and relatively linear. Located close to high activity areas such as Docks and Shopping Centres in AQMA	8
Birtley, UK	1	Birtley	Short and relatively linear. Located in a less congested urban area in the vicinity of Banks and Commercial Centres	7
Greater Manchester, UK	3	Wigan	Non-linear and over a large area within a built-up area having access to M6 and train station	18
		Stockport	Linear on the A6 Buxton Road with high flows of commercial vehicles gaining access to local motorways	11
		Trafford	A longer linear network mainly embracing the A56 trunk road. Close to Old Trafford and having access to M60 and M602 among others	11

Table 4.1: The summary of the study sites description

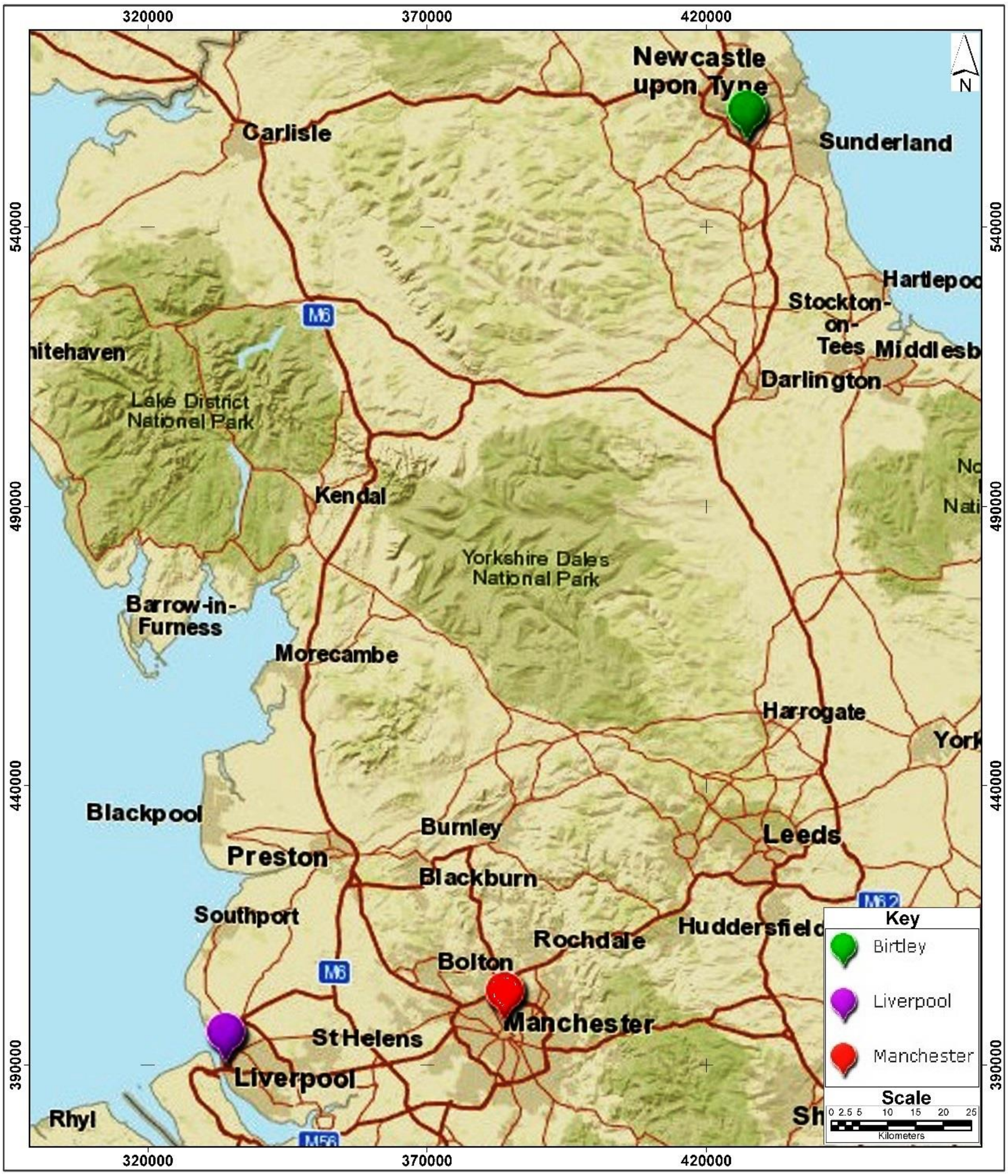


Figure 4.1: Map showing the study locations in the UK

4.1.3 Description of the Bluetooth traffic data collection

This section presents the specific methods used in the collection of Bluetooth data used for traffic estimation in this research. In all the locations, the data providers (Mouchel/2020Liverpool, Gateshead City Council, and TfGM in conjunction with TDC and SkyHigh) performed the set up for the on-site

Bluetooth traffic monitoring. As stated in Section 3.2.3, the study utilised the data collected by the Bluetooth sensors (Hi-Trac Blue) developed and tested by TDC Systems Ltd. These sensors were used to continuously detect the Bluetooth-enabled devices passing through the sensor locations within the networks and store their MAC addresses. While the focus was on vehicular traffic detection, Bluetooth-enabled devices carried by other network users such as pedestrians and cyclists in the traffic were also captured. The essential records stored by the sensors are the MAC addresses, timestamp and the details of sensor locations such as the coordinates.

Essentially, the detected MAC addresses in raw form were encrypted for the purpose of privacy and security before they were transmitted to an online database (C2-Web) through the traffic control network using routers and direct cables from site to server. Specifically, in this research, all the data sets used were either downloaded through the internet using access codes or received as attachments either through e-mail or on an external hard disk. However, there were site visits to the Liverpool and Birtley study locations primarily for better understanding of the locations of the sensors and the traffic stream.

4.1.4 Challenges and limitations in the data collection

At the onset, the experimental design for this research was based on the use of SCOOT measured flows and GPS tracking data, and where possible with ANPR data to validate the Bluetooth results. However, different challenges were met at different stages that put the completion of this research at a risk. For example, the different councils that provided the data used have their specific objectives that are independent of this research. Consequently, this research had no control over when to deploy or remove the sensors. Originally, Liverpool was considered as the only study site with the expectation to meet the data requirements for this research through a collaboration with Mouchel/2020Liverpool. However, the trial conducted over a short period of two weeks from 15th to 27th June 2011 was discontinued, thereby resulting in insufficient Bluetooth data with a limitation in scope and coverage as well as in the provision of the validation datasets. Due to these limitations, the research design was adjusted to further this investigation. Following the modification to the research design, the Birtley study site was chosen to provide solutions to the challenges encountered in the Liverpool study site. However, similar challenges as in the Liverpool study site were also encountered. In order to achieve the research goal, further adjustment was made to the research design to address the limitations in scope and coverage as well as in the provision of the validation datasets, and in this case the Greater Manchester study site was chosen. In Manchester, the installation of the Bluetooth sensors is on a permanent basis across the three networks. Some of the sensors are installed near SCOOT and ATC loop detectors for independent measurement of the traffic, and for validation. However, prior to the final acquisition of the data sets, there were further challenges in the process. Prior to the acquisition of the data from the Manchester study site, data from ATC and ANPR were proposed as the new datasets for results validation while the study area included Scotland due to the availability of data for results validation. However, the inclusion of the Scotland study site was discarded due to positive results from the Greater Manchester area, leading to the provision of the required validation data sets that include ATC and SCOOT flows captured over the same period as Bluetooth.

4.2 Liverpool: Preliminary Study on Data Quality Assessment

4.2.1 Background to Liverpool study

Figure 4.2 presents the map of the study area showing the distribution of the Bluetooth sensors, while Table 4.2 presents the description of the locations of the eight Bluetooth sensors strategically chosen within an Air Quality Management Area (AQMA) in Liverpool. As discussed above, Liverpool was originally considered as the study site to meet the data requirements following a mutual understanding between Newcastle University and

Mouchel/2020Liverpool. However, the Bluetooth sensors installed in the Liverpool study site were disengaged after two weeks of data acquisition. This means that the objective of this research could not be realised based on just two weeks worth of data, and thus required a modification to the research design to further the research. However, the data collected over the two weeks were used for preliminary data quality assessment to understand the structure and condition of the variables in the data to aid further analysis.

Figure 4.2: Location of Bluetooth sensors in the Liverpool study site

Station Number	Coordinates		
	Latitude	Longitude	Station description
1	53.4114	-2.99908	Bath Street
2	53.4116	-2.99686	King Edward Street
3	53.4081	-2.99236	Chapel Street
4	53.4076	-2.98977	Dale Street
5	53.4052	-2.99070	James Street
6	53.4009	-2.98687	Liver Street
7	53.3970	-2.98582	Wapping
8	53.3941	-2.98301	Chaloner Street

Table 4.2: Description of the Bluetooth sensors locations in the Liverpool study site

4.2.2 Data quality assessment

The aspects of data quality assessment include accuracy, reliability, presentation and consistency. This section considers the preliminary aspects of the Bluetooth data quality to understand its relevance to this research. The aspects under consideration in this section are data presentation, completeness, update status (timeliness) and consistency. An example data is presented in Appendix 4A, which shows that Bluetooth data are well-presented with clear headers (variables) describing the data. The data also come as comma separated values (csv) file format, which gives the data a defined structure in terms of presentation. In terms of completeness, the Bluetooth sensors are capable of continuous data recording throughout the day, which makes the data acquired complete and adequate for studies such as temporal status monitoring. Also, the time stamp recorded by the sensors is on a second-by-second basis that shows the timeliness of the data. This one-second level of precision attribute of the data is very significant in the classification of devices during data filtering. For example, speed can be calculated on a second-by-second basis, while travel time for devices detected on short links can be differentiated. For instance, a short link of length 500m with a speed limit of 48km/h will require a travel time of 37.5 seconds to traverse the link. Clearly, this value is greater than the 1-second resolution level measured by Bluetooth and thus confirms the sufficiency of the precision level of the measurement. Furthermore, the quality of the Bluetooth data was analysed for consistency to serve as the benchmark for the subsequent analysis. Therefore, the preliminary data quality investigation was conducted under the following assumptions.

1. Bluetooth data are expected to present profiles similar to real life traffic situations by capturing the variations in traffic flows over the day; and
2. The captured data are expected to present continuous profiles with minimal gaps (missing data) over the period of observation with weekday profiles contrasted to weekend profiles.

4.2.3 Results presentation and analysis

For a better understanding of the daily flows, Table 4.3 and Figure 4.3 present respectively the summary of traffic flow over fourteen days and the equivalent profile over seven days at Station 7. The profile of Station 7 having the highest count (Figure 4.4) was presented as an example of the individual station analysis because of the configuration of its position in the network to assess consistency. Results from the other stations are presented in Appendix 4B. The profile over seven days is presented to show the similarities and consistency observed in the traffic count within this period. From Table 4.3, the total count for the period of observation (Wednesday 15th June – Tuesday 28th June 2011) is 103,520. The highest daily count (8515) was observed on Friday 17th June while the lowest count (4513) was observed on Sunday 26th June 2011. From the profile, there are two prominent peak periods in the weekdays' observations, the morning and evening peak periods with an average count of 700 devices over the hours of 8 am and 5 pm. These are related to the period of trips to and from work as is the case with real life traffic data, and thus confirming the first assumption of representing real-life traffic. In fact, the dual peaks observed in the data were also observed in previous studies (Beca, 2011; Augustin and Poppe, 2012; Cragg, 2013). Similarly, the second assumption is confirmed through the continuity observed in the profiles and the similarities and high positive correlation observed in the weekdays/weekend data. In the next chapter, the validity of the assumptions will be verified against real life traffic data to ensure data quality assurance.

Count of NumbPlat	Days														
Hours	15-Jun	16-Jun	17-Jun	18-Jun	19-Jun	20-Jun	21-Jun	22-Jun	23-Jun	24-Jun	25-Jun	26-Jun	27-Jun	28-Jun	Grand Total
00	51	74	51	160	180	66	57	34	72	74	150	218	73	46	1306
01	34	22	48	113	161	49	32	29	32	37	113	127	41	41	879
02	22	22	24	84	108	26	22	20	21	20	95	140	21	18	643
03	16	25	27	70	92	23	25	18	13	18	81	118	27	21	574
04	37	18	32	53	63	35	30	26	32	27	45	79	43	39	559
05	83	71	69	69	80	72	78	64	77	70	52	70	81	86	1022
06	151	171	145	79	82	134	178	149	179	159	57	56	146	165	1851
07	488	461	397	130	98	488	465	457	444	431	127	63	446	476	4971
08	675	648	612	182	91	644	723	691	682	635	182	78	648	671	7162
09	641	614	543	248	123	612	588	607	579	557	288	145	604	620	6769
10	526	459	507	332	281	513	449	476	486	498	354	199	474	514	6068
11	495	546	484	409	359	535	506	488	470	460	418	222	506	446	6344
12	521	576	532	460	401	437	511	486	475	588	456	254	501	532	6730
13	495	499	539	520	492	543	518	542	600	611	477	303	526	520	7185
14	545	484	579	533	459	518	550	496	508	574	461	419	478	592	7196
15	585	644	622	578	426	496	561	523	576	675	448	363	489	622	7608
16	714	700	660	459	442	627	630	676	668	662	458	359	600	713	8368
17	692	725	675	462	428	693	720	635	670	620	457	342	588	726	8433
18	487	501	509	446	301	434	406	462	449	450	432	245	395	430	5947
19	302	334	428	417	263	287	271	301	306	340	414	230	268	284	4445
20	189	200	309	309	214	165	202	202	225	278	261	168	159	208	3089
21	175	164	340	206	166	164	146	159	171	198	229	116	152	158	2544
22	122	134	200	175	137	103	126	122	131	164	222	115	96	133	1980
23	88	105	183	202	106	78	99	125	84	175	342	84	82	94	1847
Grand Total	8134	8197	8515	6696	5553	7742	7893	7788	7950	8321	6619	4513	7444	8155	103520

Table 4.3: Count of detected Bluetooth-enabled devices at Station 7 in June 2011

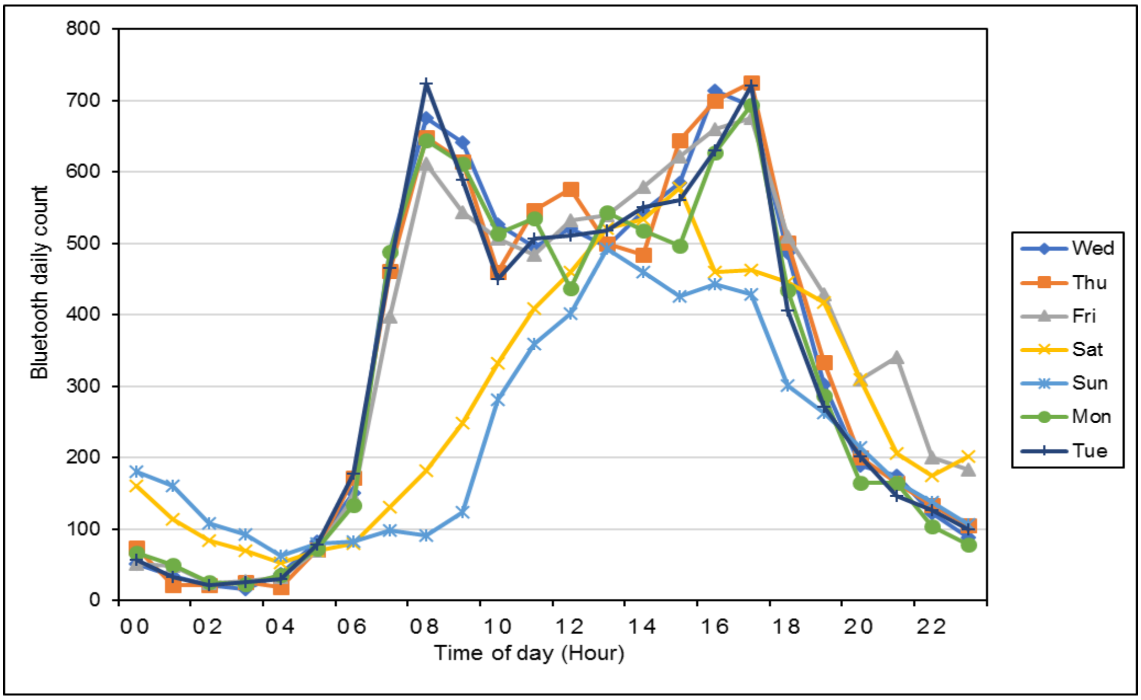


Figure 4.3: Daily profiles of counts of detected devices at Station 7 over seven days

Figure 4.4 presents the summary of the daily flows over eight days in Liverpool. The lowest and highest flows were observed respectively at Stations 1 and 7 over the days. Stations (2, 7, and 8) located close to the Docks and on A5036 connecting A562 in the South and A565 in the North have the highest flows as expected compared to Stations (1, 3, 4, 5, and 6) located along minor roads. Stations 5 and 6 on the other hand exhibited a different trend over the two Saturdays (18/06/2011 and 25/06/2011) with a higher flow compared to the weekdays. This change in trend at Stations 5 and 6 is attributed to the activities around St John's Shopping Centre and Liverpool John Moore's University. However, there is consistency in the data over days and stations. The consistency in the result obtained at this level is very interesting because Bluetooth data shows a strong indication to model the real world traffic and, in that case, a candidate to provide transport data. The data were further analysed as contained in Table 4.4 and Table 4.5. However, the reliability of the results will be tested through validation in Section 5.3.1.

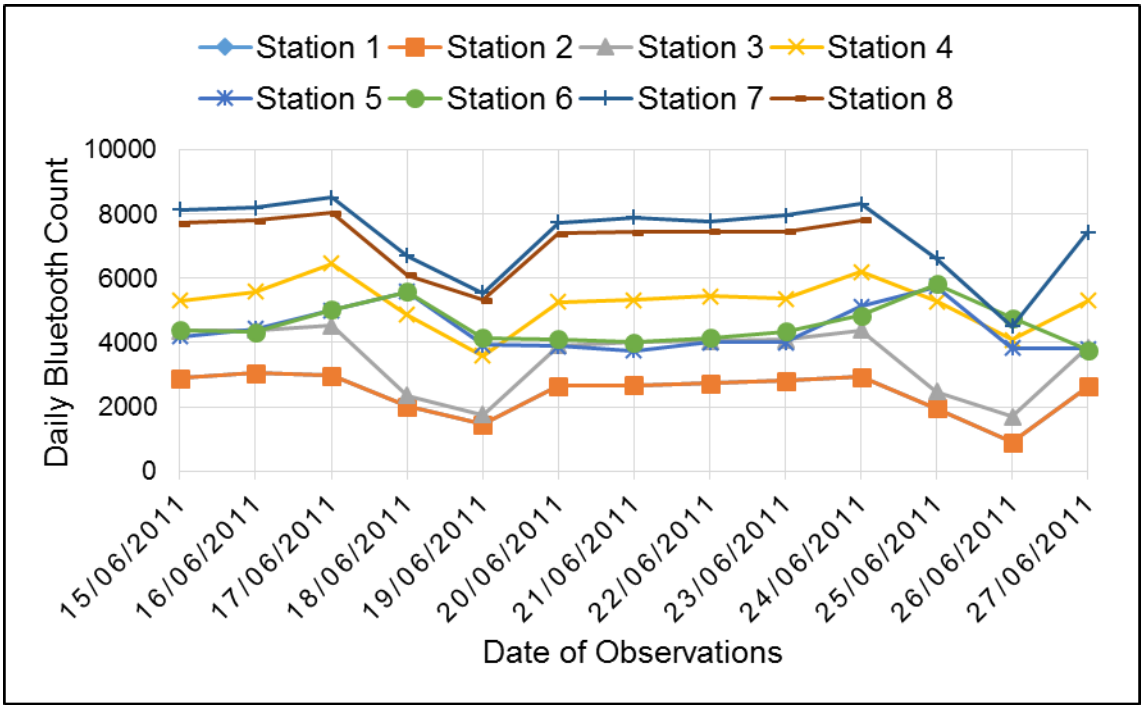


Figure 4.4: Summary of the variations in daily flows over eight stations in Liverpool

From the second assumption, since the traffic count profile of the weekdays differs significantly from that of the weekend, the two sets of data were analysed separately. Table 4.4 shows the correlation analysis of the weekdays' count while Table 4.5 presents the descriptive statistics. Analysis performed on the weekdays (Monday to Friday) showed a very strong positive relationship (>0.09) with a corresponding high level of similarities in the data between the weekdays as shown from the descriptive statistics, and thus indicates a level of quality in the data. For example, the values of the kurtosis, a measure of the peakedness of the distribution relative to the normal distribution as defined by Adedayo (2006), showed that they all exhibit similar distribution and peakedness. However, on Fridays (17th and 24th June), the results exhibit negative skewness (-0.25 and -0.08 respectively) as would be expected due to a translation from weekdays to weekends. The change in the skewness of the data on a Friday is attributed to a change from weekdays to the weekend normally associated with weekend travel and activities. Figure 4.5 shows the scatter plot of Saturday and Sunday hourly flows overlaid with regression line showing a very strong correlation ($R^2 = 0.896$).

	15-Jun	*16-Jun*	*17-Jun*	*20-Jun*	*21-Jun*	*22-Jun*	*23-Jun*	*24-Jun*	*27-Jun*	*28-Jun*
15-Jun	1									
16-Jun	0.992	1								
17-Jun	0.967	0.969	1							
20-Jun	0.991	0.980	0.954	1						
21-Jun	0.991	0.987	0.958	0.989	1					
22-Jun	0.995	0.988	0.967	0.991	0.991	1				
23-Jun	0.991	0.986	0.969	0.989	0.991	0.995	1			
24-Jun	0.977	0.979	0.981	0.963	0.975	0.979	0.984	1		
27-Jun	0.991	0.983	0.951	0.993	0.990	0.994	0.990	0.972	1	
28-Jun	0.995	0.987	0.966	0.985	0.993	0.989	0.991	0.982	0.985	1

Table 4.4: Correlation analysis between weekdays (Monday – Friday)

Descriptors	15-Jun-11	16-Jun-11	17-Jun-11	18-Jun-11	19-Jun-11	20-Jun-11	21-Jun-11	22-Jun-11	23-Jun-11	24-Jun-11
No of Observations	24	24	24	24	24	24	24	24	24	24
Mean	338.9	341.5	354.8	322.6	328.9	324.5	331.3	346.7	310.2	339.8
Standard Error	51.55	51.50	47.47	49.03	49.97	49.13	49.21	49.11	46.36	51.79
Median	395	397	413	361	339	379	375	386	332	357
Standard Deviation	252.54	252.30	232.57	240.18	244.83	240.68	241.09	240.59	227.13	253.72
Kurtosis	-1.71	-1.66	-1.55	-1.75	-1.57	-1.65	-1.64	-1.66	-1.79	-1.65
Skewness	0.03	0.03	-0.25	0.05	0.11	0.02	0.02	-0.08	0.03	0.09
Range	698.00	707.00	651.00	670.00	701.00	673.00	669.00	657.00	627.00	708.00
Minimum	16	18	24	23	22	18	13	18	21	18
Maximum	714	725	675	693	723	691	682	675	648	726

Table 4.5: Descriptive statistics for the weekdays count from 15th – 28th June 2011

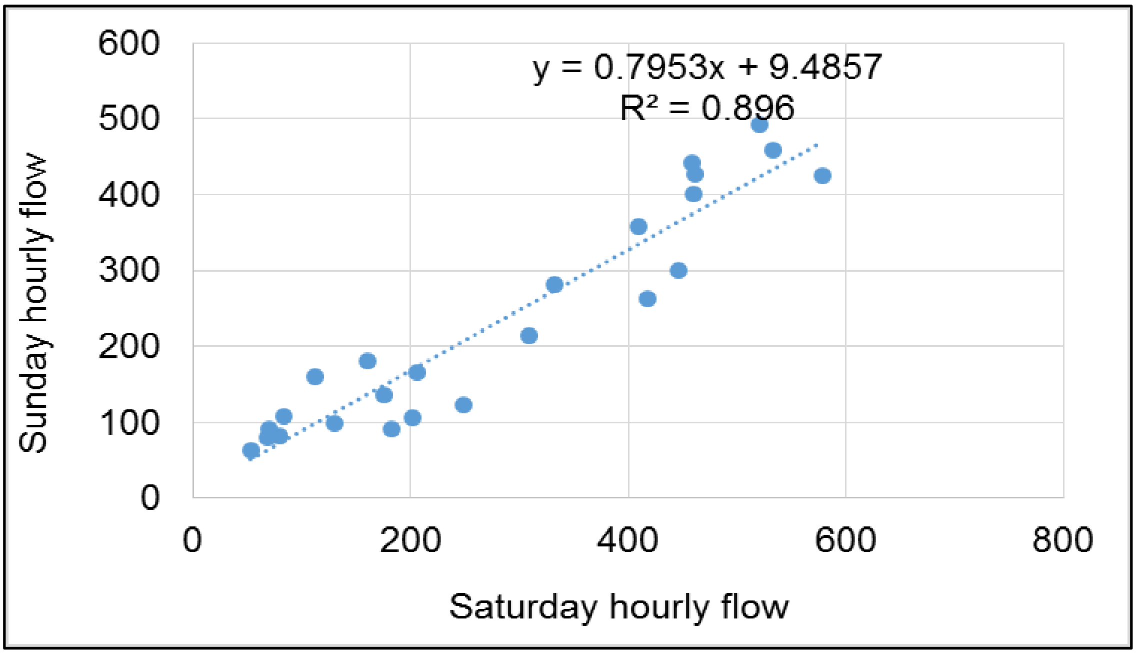


Figure 4.5: Scatter plot of weekend flows overlaid with regression line

4.2.4 Conclusion from the Liverpool study

The data collection conducted in Liverpool for quality assessment showed that Bluetooth data is of high-resolution (one-second), consistent, and with a well-structured presentation. Two peak periods consistent with real life traffic data were observed: the morning peak hours (7-9am) and the evening peak hours (4-6pm). Correlation analyses performed showed a very strong positive correlation between weekdays and between weekend observations as would be expected of real life traffic. The descriptive statistics also showed a high level of consistency.

However, the reliability of the data will be established in Chapter 6. Summarily, the preliminary data quality assessment conducted justifies the need for continued research on Bluetooth data to establish its relevance and maximise its potential to support the delivery of an enhanced traffic management.

4.3 Birtley: An Evaluation Platform for Bluetooth Traffic Metrics Estimation

4.3.1 Background to the Birtley study

The Birtley study area is located north of County Durham and South-West of Gateshead. The study consisted of seven Bluetooth monitoring stations located mainly along the A167, Durham Road as shown in Figure 4.6. Table 4.6 presents the description of the location of the sensors. The data were collected over two weeks from 5th March to 16th March 2012. The aim of the study was to create an evaluation platform for Bluetooth data to enhance traffic management by employing a post-processing data analysis technique developed in this research. The major assumption made under this section builds on the Liverpool study to further the preliminary assessment of Bluetooth data. The assumption is that under normal conditions, the proportion of the Bluetooth-enabled devices captured will vary in time and space (geographical location) with variations in traffic patterns. The results of this test are presented in the next section.

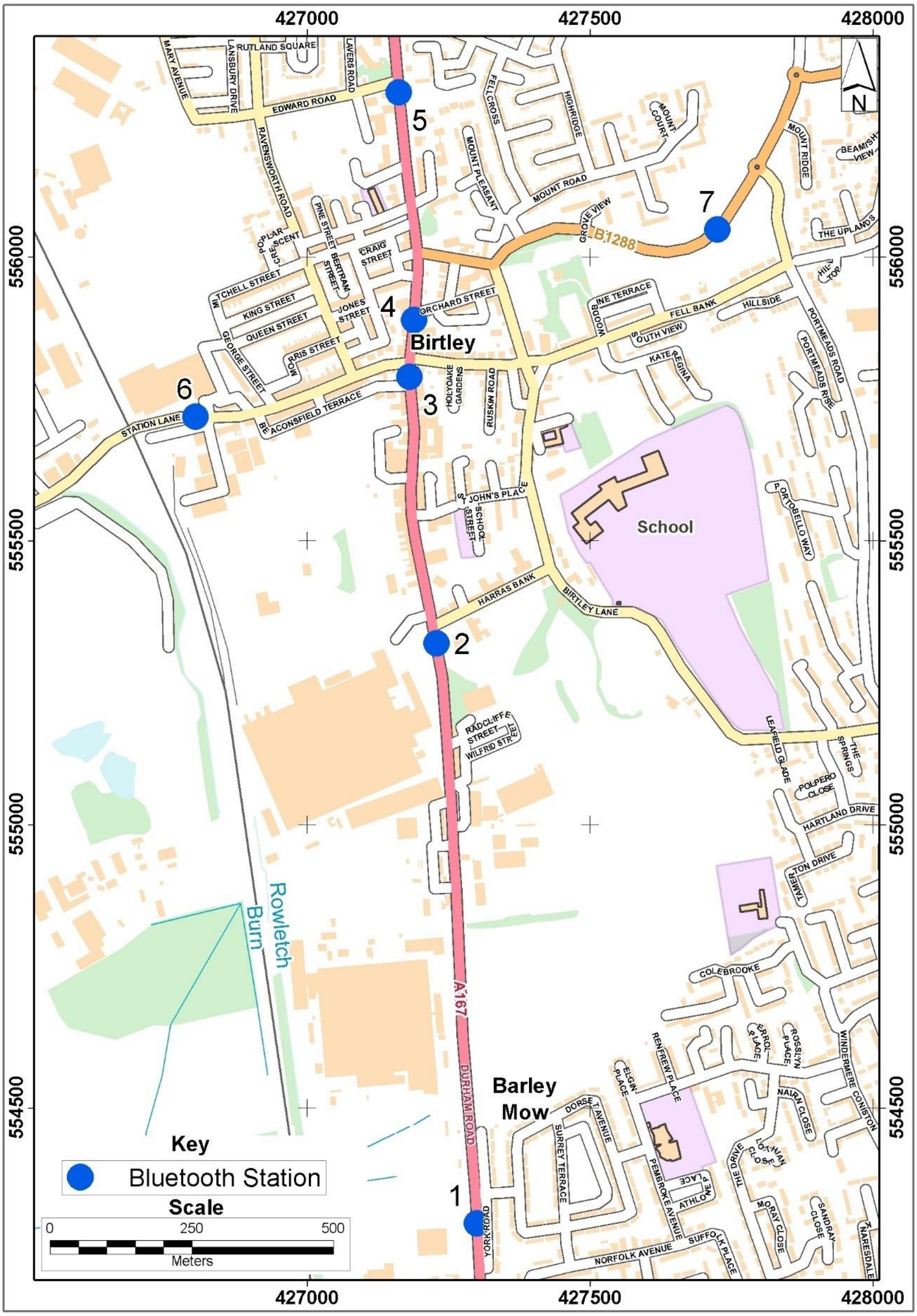


Figure 4.6: Location of Bluetooth sensors in the Birtley study site

Station Number	Coordinates		Station description
	Latitude	Longitude	
1	54.88269	-1.57599	A167 Durham Road, Birtley (South of Dorset Avenue)
2	54.89187	-1.57700	A167 Durham Road, Birtley (South of Harras Bank)
3	54.89610	-1.57770	A167 Durham Road, Birtley (South of Station Lane)
4	54.89700	-1.57757	A167 Durham Road, Birtley (South of Orchard Street)
5	54.90060	-1.57795	A167 Durham Road, Birtley (South of Edward Road)
6	54.89549	-1.58360	Station Lane, Birtley (West of Factory Access)
7	54.89841	-1.56920	Mount Pleasant Road, Birtley (South of Portmeads Road)

Table 4.6: Location description for the Bluetooth sensors in the Birtley study site

4.3.2 Results and Analysis

Figure 4.7 shows the count of MAC addresses captured daily throughout the 11-day survey period at Stations 6 and 7 and for the 6-day period for Stations 1 to 5. The non-uniformity observed in the daily count particularly at Stations 1 to 6 is due to the difference in the start and end time of the period of data acquisition according to the time of installation and removal of the sensors. For example, observations started at 1 pm on the first day and ended at around 4 pm on the 10^{th} day. The counts of the detected MAC addresses represent the proportion of the total traffic (all modes) passing the detectors. The proportion of the actual flow is assumed to depend on the level of the road usage and the consistency of detecting Bluetooth-enabled devices from day-to-day. However, it is clear from Figure 4.7 that the Bluetooth counts from each station over 24 hours are similar from day to day. The spatial variations in the Bluetooth count represent the level of Bluetooth usage across the stations. For example, the lowest number of devices was recorded at Station 7 (60 devices) over the weekdays. The highest number of devices was recorded at Station 3 with an average of 210 devices over the weekdays.

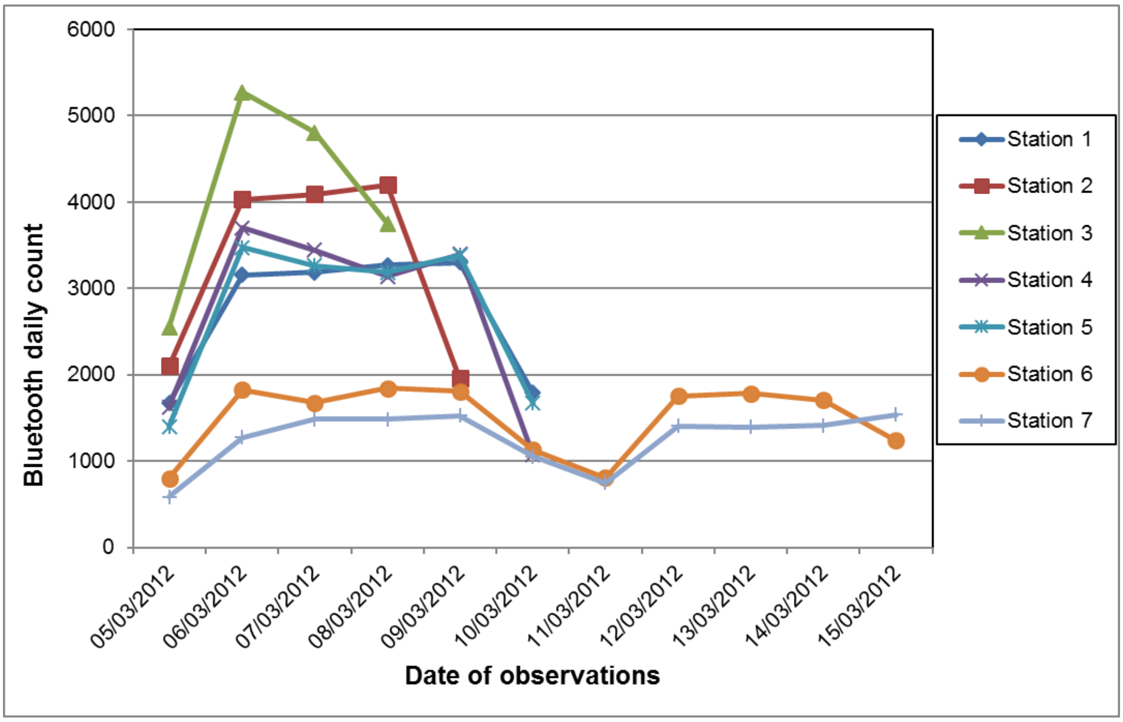


Figure 4.7: The profiles of Bluetooth hourly count at the seven stations

Daily trips were explored by matching MAC addresses at different stations and calculating travel times. From the analysis of travel time and the direction of travel, two main trip patterns emerged, based on their characteristic road usage when passing along a link between two consecutive stations designated as “single trip” and “round trip” commuters. Trips made without a return on the same day were classified as a single trip, while any trip with a return trip on the same day was classified as a round trip. Given that only a sample of the actual traffic was captured by the Bluetooth sensors, other types of trip were classified under the above two broad classifications.

- The “undetected” who were assumed to have either made the return trip but no longer with the Bluetooth switched on (perhaps due to weak battery or the device was switched off) or they left the network without passing a Bluetooth detector and were not detected on the return trip, or possibly made a return trip but were not detected by the sensors.

- Those who stayed within the detection zone (either employed in the area, were visiting or lived there) for a period and were detected later in the day is another possibility.

Table 4.7 presents the scaled Bluetooth counts of devices through the network in one (single trip) or both (round trip) directions. Scaling was done to ensure uniformity in the presentation of the data. The links breakdown shows more single-trip commuters than round trip commuters as would be expected. Analysis shows that the smallest mean ratio (2.5) of single trip to round trip was observed on Link36 while the highest (5.2) was observed on Link47. The ratio gives a level of understanding of the usage of the routes. Link12 exhibits the most similar characteristics based on the precision (range of 0.1) of the ratio observed on the link. The widest departure (ratio 4.9 - 5.6) was observed on Link47 with a range of 0.7. The spatial variation observed in the data confirmed the assumption made on the data with detection dependent upon the location of the sensors. Although not validated, there is an obvious reflection of the movements of commuters (O-D patterns) across the network.

Analysis of the link speeds for the journeys made each way along selected links between links (12), (23), (36) and (47) showed that the typical speed is in the range of 10km/h and 65km/h with a higher percentage of the vehicles travelling within 40km/h. This is considered reasonable given the average speed limit for Gateshead (20.7mph) and Tyne and Wear 23.4mph (Thorgil, 2007; Tyne and Wear, 2010; DfT, 2011; Tyne and Wear, 2011). This result also showed that Bluetooth can provide estimates of speed for individual vehicles along stretches of road as well as the proportion of vehicles moving at a particular speed, as parameter to measure or understand delay (Ayodele et al., 2013). At this preliminary stage of the analysis and without access to independent measures of traffic flows, it was assumed that the Bluetooth estimates are representative of the actual traffic.

Link	Link length (m)	Date	Single trip	Round trip	Trip ratio	Mean Ratio
		06/03/2012	1112	426	2.6	
		07/03/2012	1155	430	2.7	
L12	1,023.17	08/03/2012	1214	464	2.6	2.6
		06/03/2012	1271	472	2.7	
		07/03/2012	1241	430	2.9	
L23	479.84	08/03/2012	1039	331	3.1	2.9
		06/03/2012	572	255	2.2	
		07/03/2012	495	220	2.3	
L36	396.66	08/03/2012	467	160	2.9	2.5
		06/03/2012	457	94	4.9	
		07/03/2012	529	106	5.0	
L47	685.14	08/03/2012	507	90	5.6	5.2

Table 4.7: Summary of trip patterns on four prominent links in the Birtley study site

4.3.3 Conclusion from the Birtley study

A preliminary study on the exploration of Bluetooth data to estimate traffic metrics to enhance traffic management has been carried out. Analysis of the trip patterns showed that a single trip was more prominent than a round trip over the select road sections in the Birtley urban area. The preliminary results and analysis indicate that Bluetooth could be used to understand the trip patterns in a network. The ability to identify trip patterns (origins and destinations) offers the potential to considerably enhance decision making with respect to managing traffic demand and providing information to users of the network across modes (Bell *et al.*, 2012). Although at this stage, only a preliminary analysis of the pilot survey is available but some interesting applications emerge. The counts from day to day were consistent suggesting that the origins and destinations in the area could be monitored successfully over time of the day. Such information is useful to model traffic conditions, and to provide better congestion management systems. On a link basis, this will enable a realistic evaluation of network performance.

4.4 Manchester: Exploring Transferability

4.4.1 Background to the Manchester study

Table 4.8 presents the description of the Bluetooth stations in the three study sites of Wigan, Stockport and Trafford located in Greater Manchester. Due to the small scale of the Manchester study site, the zoomed in (detailed) map of each of the study sites is presented in subsequent sections for clarity. This pilot study in Manchester builds on earlier work carried out in Birtley (Section 4.3), which demonstrated the potential of Bluetooth data to classify network users (such as, round-trip or single-trip commuters). It is also used to identify the patterns of movement through a simple network to show the capability for enhanced traffic management. This study in comparison to the earlier work in Birtley was carried out on a larger scale (utilising 23 stations compared to 7 stations in Birtley) to demonstrate the transferability of the research method. The data collection consists of three study sites – Wigan, Stockport and Trafford, which have "non-linear, linear, and longer-linear" network layouts respectively. In this case, the non-linear network is defined as the array of sensors over urban roads with interconnecting routes forming area-wide O-Ds. The linear network is defined as the array of sensors mainly in a linear form over a road segment not exceeding 4km. On the other hand, the longer-linear network is the array of sensors primarily in a linear form over a road segment up to 4km or greater. The three case studies were chosen to investigate whether there are any differences in traffic patterns over the entire network. Bluetooth data captured from Wigan were analysed for trip patterns and speed distribution, while data from Stockport were analysed for transferability checking for possible differences in the results and interpretation. Data from Trafford were analysed mainly to explore monthly variation. The results for each demonstration are presented in turn with conclusions drawn and next steps articulated.

Station Name	Coordinates		Station description
	Latitude	Longitude	
MAC1012WG	53.51902	-2.65240	Warrington Road
MAC1013WG	53.52892	-2.65498	Warrington Road/Smithy Brook Road
MAC1014WG	53.54142	-2.64781	Wallgate Saddle Gyratory
MAC1015WG	53.54323	-2.63559	Wallgate/Caroline Street
MAC1016WG	53.52564	-2.64757	Poolstock Lane/St Pauls
MAC1017WG	53.52975	-2.64372	Poolstock Lane/Rushdene
MAC1018WG	53.54121	-2.63077	Chapel Lane
MAC1021WG	53.56371	-2.63169	Wigan Lane/Brock Mill Lane
MAC1022WG	53.55925	-2.62833	Wigan Lane/Royal Albert Edward Hospital
MAC1023WG	53.54873	-2.62713	Central Park Way
MAC1024WG	53.55758	-2.66141	Woodhouse Drive/Scot Lane
MAC1025WG	53.55275	-2.66532	Scot Lane/Challenge Way
MAC1026WG	53.53649	-2.68501	Orrell Road/Fleet Street
MAC1027WG	53.53570	-2.67097	Ormskirk Road/Sherwood Drive
MAC1028WG	53.53768	-2.65721	Ormskirk Road/Alker Street
MAC1029WG	53.51690	-2.68382	Pemberton Road VMS
MAC1030WG	53.53230	-2.66539	Billinge Road/Little Lane
MAC1031WG	53.52168	-2.66892	Holmes House Avenue
MAC1033ST	53.39596	-2.14980	A6 Buxton Rd/Nangreave Rd
MAC1034ST	53.39295	-2.14634	A6 Buxton Rd/Kennerley Rd
MAC1035ST	53.38990	-2.14066	A6 Buxton Rd south of Woodsmoor Rd
MAC1036ST	53.38672	-2.13178	A6 Buxton Rd north of Dialstone Ln
MAC1037ST	53.38432	-2.12700	A6 London Rd/Newmoor Rd
MAC1038ST	53.38317	-2.12574	A6 London Rd south of Vernon St
MAC1039ST	53.38057	-2.12262	A6 London Rd se of Hope St
MAC1040ST	53.37903	-2.11899	A6 London Rd south of Grundey St
MAC1041ST	53.37547	-2.11382	A6 London Rd/Buxton Rd
MAC1001TR	53.39044	-2.35031	Junction of Woodlands Road A56 / Church Street
MAC1002TR	53.39516	-2.35224	Junction Manchester Road A56 / Barrington Road
MAC1003TR	53.39766	-2.35218	Junction Manchester Road A56 / Navigation Road
MAC1004TR	53.40614	-2.34743	Junction Manchester Road A56 / Park Road
MAC1005TR	53.41149	-2.34117	Junction Washway Road A56 / Eastway
MAC1006TR	53.41964	-2.33187	Junction Washway Road A56 / Marsland Road
MAC1007TR	53.42565	-2.32525	Junction Washway Road A56 / Ashton Lane
MAC1008TR	53.43115	-2.31897	Junction Cross Street A56 / Dane Road
MAC1009TR	53.39103	-2.34762	Junction Woodlands Road A560 / Barrington Road
MAC1010TR	53.39123	-2.34157	Junction Woodlands Road A560 / Stockport Road
MAC1011TR	53.39270	-2.31750	Junction Shaftsbury Avenue A560 / Thorley Lane
MAC1070MR	53.44900	-2.19217	Stockport Road/Matthew's Lane
MAC1071MR	53.44429	-2.19162	Stockport Road/Albert Road

Table 4.8: Location description for the Bluetooth sensors in Greater Manchester

4.5 Study Site 1: Wigan

4.5.1 The Wigan network

Figure 4.8 presents the Wigan network (Study site 1), near Central Park Way - a busy urban area, with the reference station (MAC1014WG – highlighted with a red circle near Poolstock Brook) presenting a strategic advantage for comprehensive data capture due to its central position within the network. The Wigan network presents a good comparison with the linear networks of Sites 2 and 3 due to the area-wide positioning of the Bluetooth sensors within the road network.

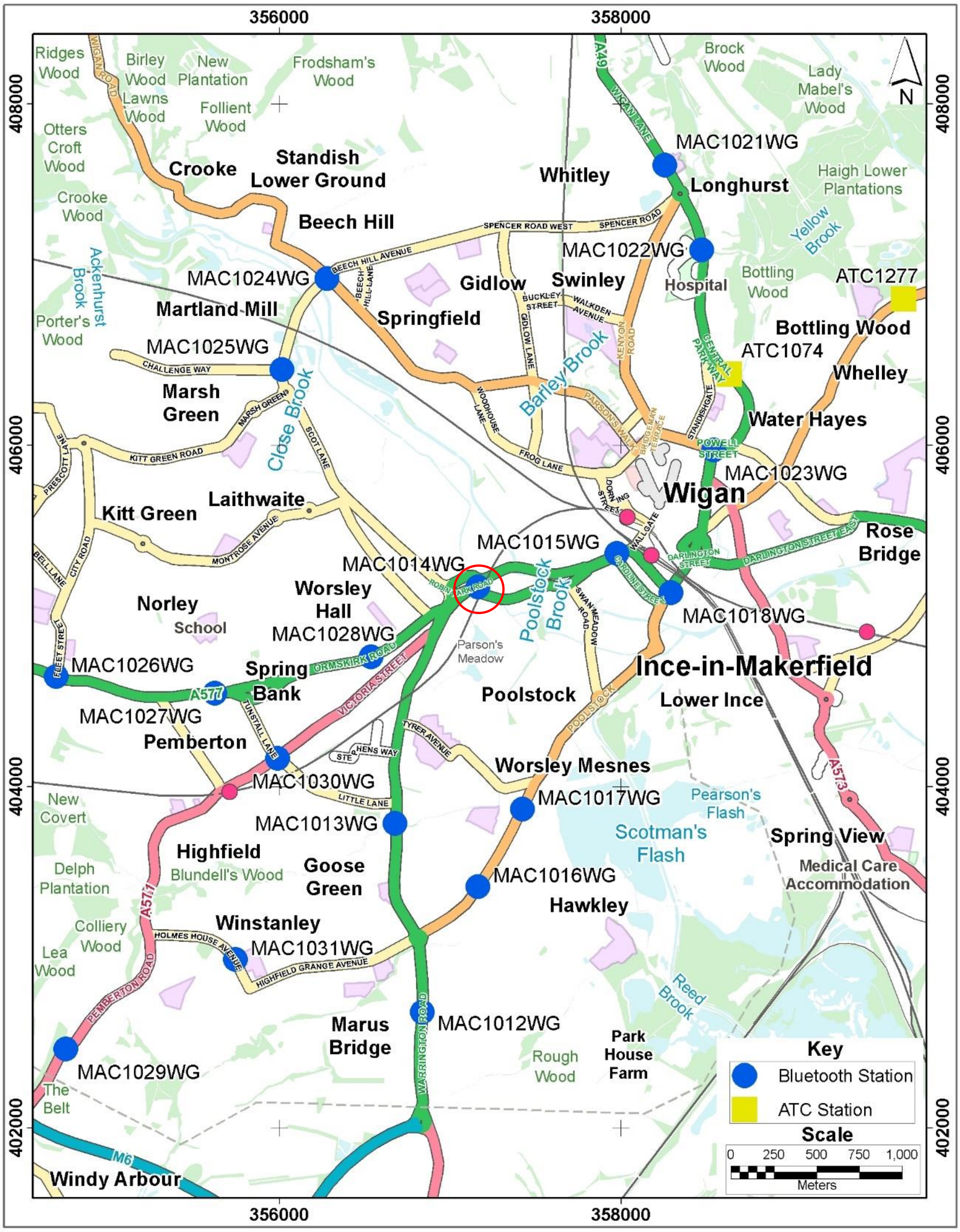


Figure 4.8: Map of the Wigan network showing Bluetooth and ATC stations

4.5.2 Estimation of traffic counts

Figure 4.9 presents the daily Bluetooth count for the period 3rd -10th September 2011 (inclusive) for all stations. A key assumption made in this study is that there is daily consistency in the percentage of the detected Bluetooth-enabled devices with variations across the different stations. Lower counts are also expected during the weekend. An interesting observation is that traffic counts at Stations 12, 14 and 18 on 10th September were systematically lower than the 3rd September at all stations despite both days being a Monday. This shows the potential of Bluetooth to respond to changes in the network by capturing the temporal changes in the traffic levels. The assumption is that any difference in the observation represents the actual changes in traffic levels on the street. Such changes were noticed on Friday (drop in flow below Station14) and Saturday (rise in flow above Station12) at Station 21.

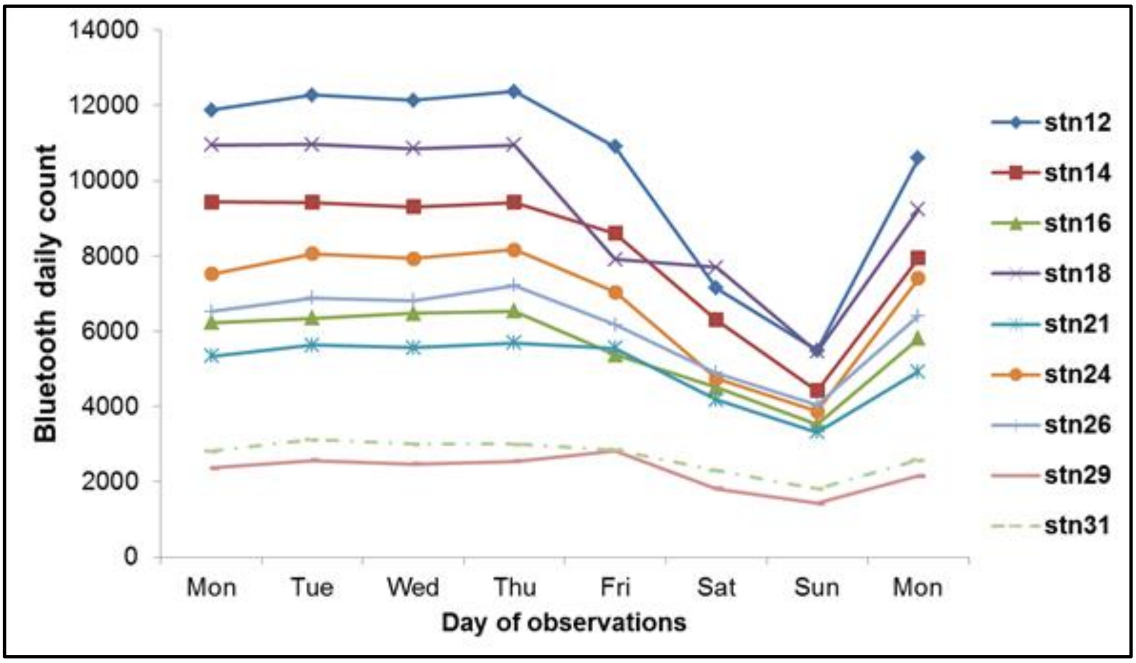


Figure 4.9: Bluetooth daily count of devices at nine stations (stn) in Wigan

4.5.3 Travel time parameters

Araghi *et al.* (2012) proposed four statistical parameters to evaluate the accuracy of travel time estimation using Bluetooth. The research showed that the minimum and median travel times provide more robust estimates of typical route travel time compared to the maximum and the mean (Araghi *et al*, 2012). However, since the minimum, maximum and mean travel times are all functions of extreme values, this PhD research considered three additional parameters namely: Twentieth Percentile, First Quartile and Third Quartile in order to establish a richer understanding of travel times to enhance interpretation, and to overcome the effect of extreme values. Overall, the results (Appendix 4C) indicate that the maxima give a clear indication of the longest delay on the road segment, while consistent with Araghi *et al.* (2012) and Araghi *et al.* (2013), the median is considered the most robust and stable measure of travel times and thus reflects the prevailing traffic conditions on the road. In the long-term study, the mean and median travel time will be explored further for statistical significance of the results.

4.5.4 Estimation of vehicle speeds

The speed of the captured devices was computed based on the methodology described in Section 3.4.3. Figure 4.10 illustrates the average over eight days of the distribution of speeds for the three major links within the network overlaid on the study site area map to indicate location. The profiles are presented as line graphs rather than as bar charts to allow for easy comparison with the distribution of speeds for different links. Dual peak, which reflects the proportion of traffic during the eight days at the particular location in the network was observed. The first mode of the bimodal distribution reflects congestion with speeds typically 10km/h and the mode at the higher level (35 - 50km/h) reflects free-flow on the road. The highest flow level (25%) was observed on Link1412 in both directions, which can be attributed to the effect of the high levels of cross-flows at the junction along the route. The least congested link was Link1418 with a substantial number of vehicles on this link travelling at speeds between 35 and 65km/h. The modal speed for Link1426 in both directions was determined to be 45km/h, and is considered reasonable given the stated speed limit (48km/h). With this information, appropriate control measures can be implemented to optimise the flow of traffic in the network.

On Links1214, 1814, and 2614, similar interpretation as described on the opposing links above is given to the results. Overall, the similarities in the profiles of the opposing links particularly on Link1412 and Link1214 means that the same plan or strategy can be implemented to control the traffic on the links.

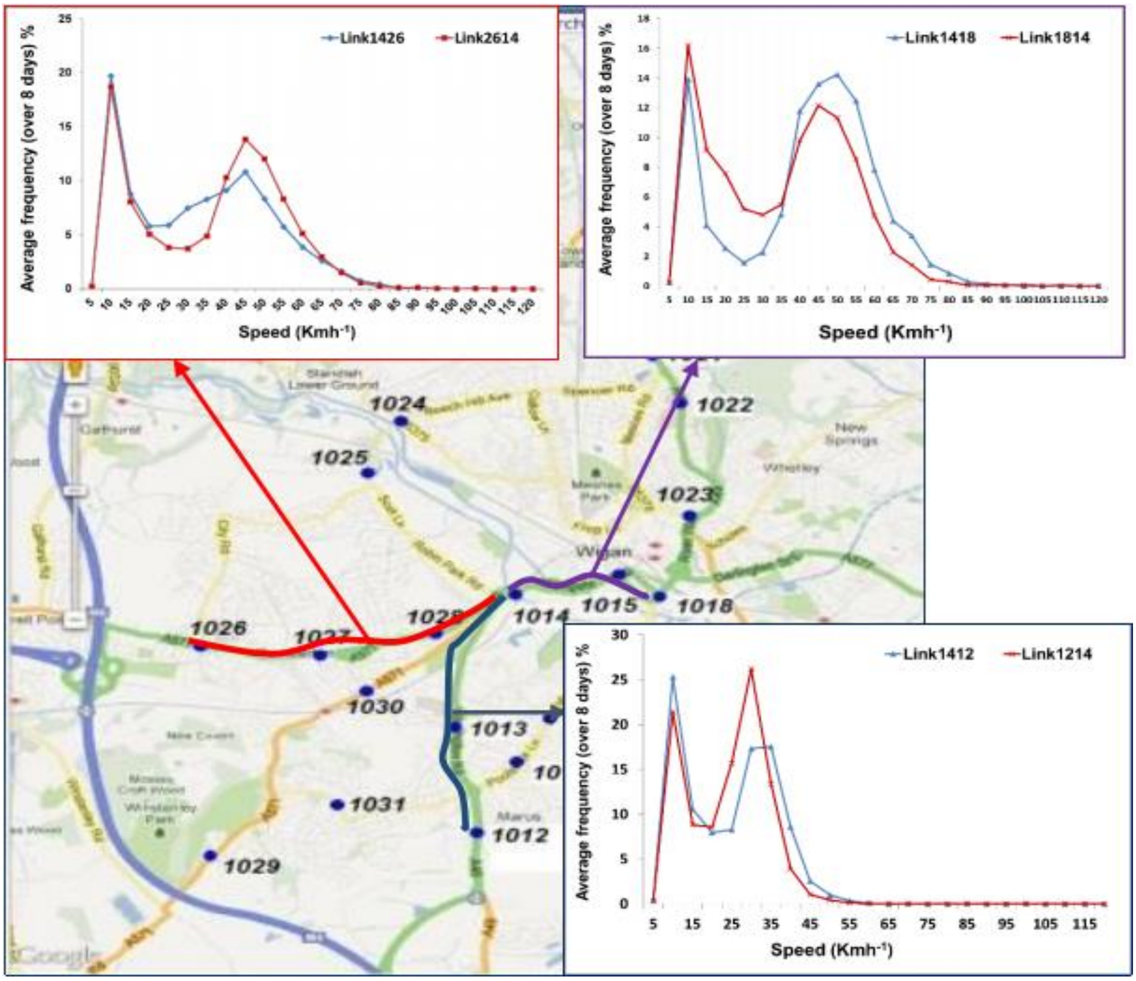


Figure 4.10: Map of Wigan showing the distribution of speed across three links (1412, 1418 & 1426) for each direction

4.5.5 Origin and destination analysis

Origin-destination analysis was carried out with Station 14 chosen as the critical reference node due to its strategic location. One mode is considered at this stage to test the research methods before extending the concept in the further study. This station was considered strategic due to its central position forming a nodal point for all the major routes. A “one-to-many” (defined as the estimation of link-flows from a reference station to all other stations in the O-D array). The O-D matrix presented in Appendix 4D showing the flow levels alongside travel times was generated based on the research method. From the results, the highest number (38%) of all vehicles tracked at Station 14 was found at Station 18 while the least (1.9%) was tracked at Station 31. The highest and lowest percentages of the vehicular total flow were also observed on these links but in the opposing direction with 40% and 1.7% respectively. This is expected as Station 18 leads to a commercial area while Station 31 is on a minor road not directly linked either upstream or downstream to Station 14. With this type of information, Bluetooth may be used in a variety of transport applications such as planning and management.

4.5.6 Defining journey types using Bluetooth data

Table 4.9 shows the summary of the trip types classified based on the trips made across the three major links, namely 1412, 1418 and 1426 identified within the network. These links are considered very important because they connect the network of the area to the M6 and Wigan North-Western Train Station which as such are expected to be busier than the other links in the network. “Out_unique” and “In_unique” as used in this context correspond to the number of unique vehicles identified leaving for example, point A to B and from point B to A respectively. The journey types are classified as either a single trip or a round trip as earlier defined under the Birtley study (Section 4.3.2). The first column under “count validation” gives the sum of out_unique and in_unique while the second column gives the sum of single trip and 2 times the round trip. The round trip is multiplied by 2 in this context to reflect the contributions from the two opposing links. The small discrepancies observed on some days with a maximum difference of six Bluetooth devices on 4th September on Link1426 is attributed to the problem of non-uniqueness of MAC address or encription error (See row 1 of an example data – Appendix 2B). Although these results are not verified by any other method, they show Bluetooth potential for journey type classification.

Table 4.9 presents the analysis for the 8-day observations as presented in columns 7 and 8. Link1426 showed the highest consistency in the count of devices with a range of 0.5 between single trip to round trip ratio. Link1412 had the highest range and a mean ratio corresponding to 1.6 and 3.9 respectively. However, the least mean ratio was observed on Link1418, signifying the highest amount of return journeys, thus indicating that this link probably has the highest demand for local access in the area due to its proximity to a commercial area and the train station (Ayodele *et al*., 2013). Consequently, the link was further analysed to investigate the hourly count profiles for consistency over the weekdays as presented in Figure 4.11.

Link	Link length (km)	Date	Out_Unique (Count/day)	In_Unique (Count/day)	Single trip (Count/day)	Round trip (Count/day)	Trips-Ratio	Mean-Ratio	Count Validation	
		03/09/2012	745	812	931	312	3.0		1557	1555
		04/09/2012	800	779	995	292	3.4		1579	1579
		05/09/2012	840	801	1069	286	3.7		1641	1641
		06/09/2012	817	741	984	286	3.4		1558	1556
		07/09/2012	595	649	860	192	4.5		1244	1244
		08/09/2012	352	374	492	117	4.2		726	726
		09/09/2012	250	270	358	81	4.4		520	520
Link1412	2.712	10/09/2012	643	628	883	194	4.6	3.9	1271	1271
		03/09/2012	2271	2431	2056	1323	1.6		4702	4702
		04/09/2012	2324	2448	2064	1354	1.5		4772	4772
		05/09/2012	2309	2459	2128	1320	1.6		4768	4768
		06/09/2012	2297	2385	2038	1322	1.5		4682	4682
		07/09/2012	1747	1582	1755	787	2.2		3329	3329
		08/09/2012	1467	1571	1330	854	1.6		3038	3038
		09/09/2012	1040	1294	1146	594	1.9		2334	2334
Link1418	1.284	10/09/2012	1905	1858	1757	1003	1.8	1.7	3763	3763
		03/09/2012	1277	1303	1290	645	2.0		2580	2580
		04/09/2012	645	688	665	337	2.0		1333	1339
		05/09/2012	1215	1247	1277	591	2.2		2462	2459
		06/09/2012	1154	1211	1149	608	1.9		2365	2365
		07/09/2012	1085	998	1137	473	2.4		2083	2083
		08/09/2012	830	819	785	432	1.8		1649	1649
		09/09/2012	577	569	580	283	2.0		1146	1146
Link1426	2.700	10/09/2012	1096	1168	1166	549	2.1	2.1	2264	2264

Table 4.9: Summary of journey types on the top three busiest routes

The percentage hourly count of the profile of the Bluetooth devices presented in Figure 4.11 and Figure 4.12 showed a high level of consistency between the weekdays on both opposing links 1418 and 1814. The highest percentage flow (about 12%) occurring at about 8-10am (morning peak) on Link1418 and 10% between 2 – 4 pm on Link1814. The graphs showed the variation in the traffic flow over the day that provides knowledge of when the section of the road may be congested. With this knowledge, traffic engineers and planners may begin to put strategies in place to mitigate any impact arising from the traffic level at those periods. From the analysis, the weekend distributions presented a clear departure from the other weekdays as expected and, as a result, were analysed separately. The information gathered was found to reveal patterns and characteristics of the traffic such as high and low flows with a high level of consistency even over the eight days of study both in terms of flow and speed. This result thus demonstrates the value of Bluetooth useful traffic metrics for traffic modelling performance evaluation for each link across the area.

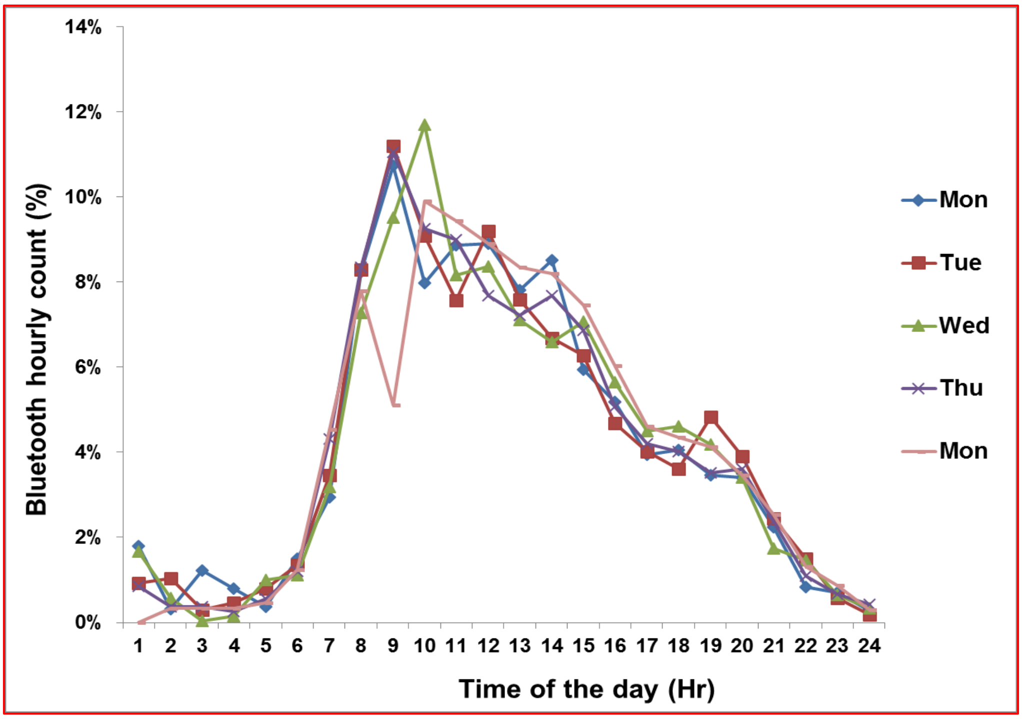


Figure 4.11: Bluetooth hourly count profile over the day for Link1418

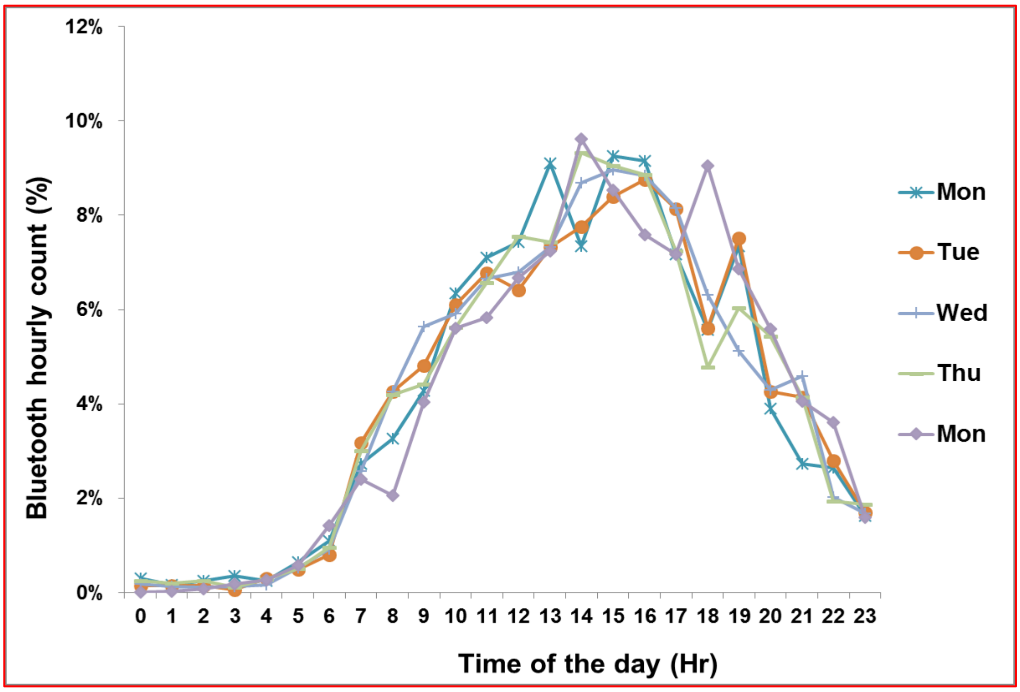


Figure 4.12: Bluetooth hourly count profile over the day for Link1814

4.6 Study Area 2: Stockport

The Bluetooth data for this study covered eight days. A similar analysis to the Wigan study was performed by utilising the research method to demonstrate reproducibility and transferability. This step provides the opportunity for a preliminary validation of the results through repeatability. Figure 4.13 presents the map of the Stockport network showing the nine Bluetooth stations (MAC1033ST – MAC1041ST) and ATC (ATC1500 and ATC1013) locations. Stockport (Study site 2) is a linear network on the A6 Buxton/London Road. The characteristics of this study site contrast with the non-linear network-based ones of Study site 1. Station MAC1033ST, which is located at the junction of Nangreave/Aquinas College Road and Buxton Road leading to London Road, was chosen as the reference point for Study area 2 in order to understand whether the Bluetooth stations that are far apart have any influence on the results. A key observation worthy of note in this study is that the two stations furthest apart (MAC1033ST and MAC1041ST) have the least match records as would be expected due to the possibility of vehicles making a detour between O-D pairs. The results from this study site are presented in Appendix 4E, and were not reported further given that they are similar to the Wigan analysis and added nothing additional.

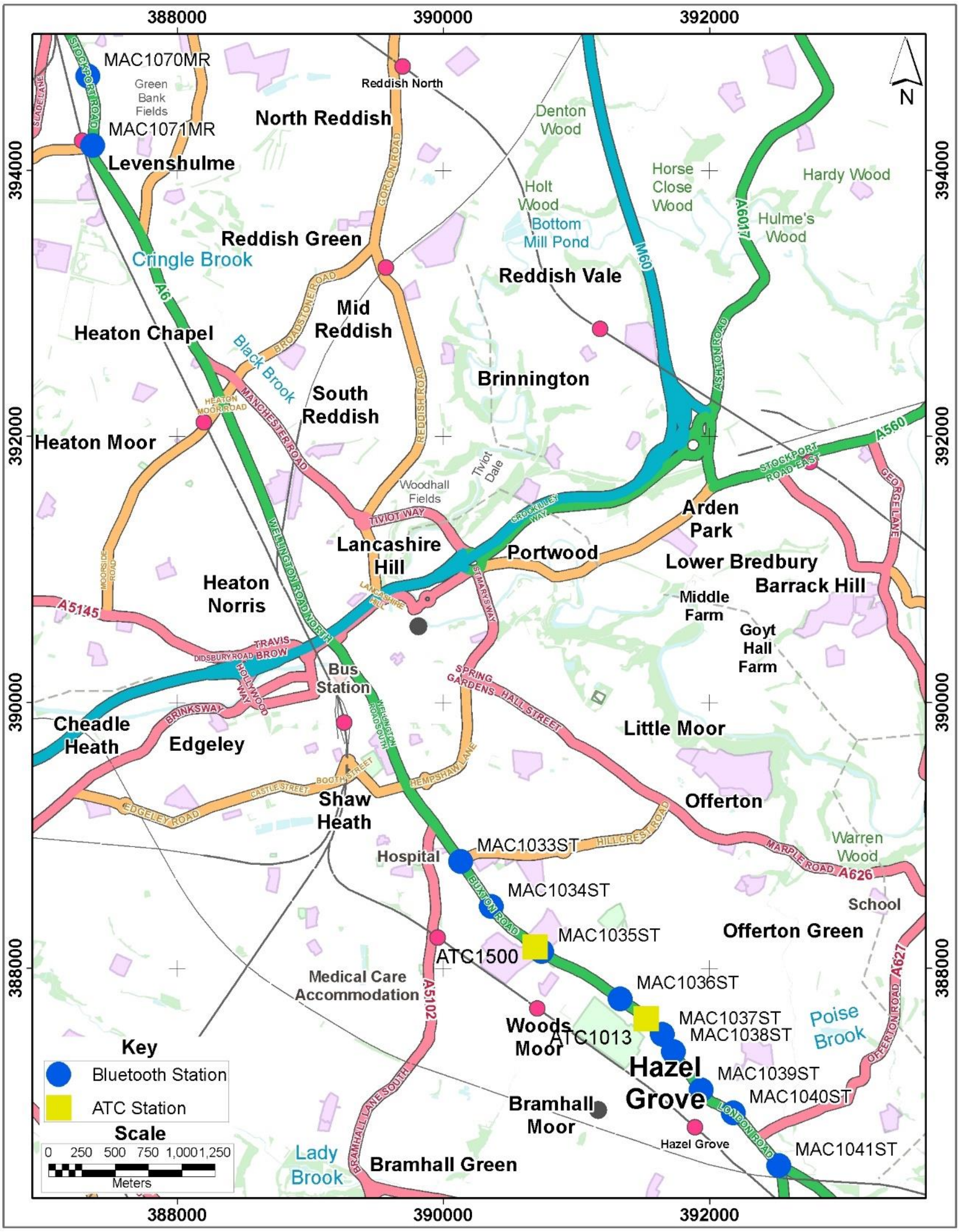


Figure 4.13: Location of Bluetooth sensors and ATC in the Stockport study site

4.7 Study Area 3: Trafford

4.7.1 The Trafford network

Figure 4.14 presents the distribution of Bluetooth sensors and ATCs over Study site 3, a longer linear network mainly embracing the A56 trunk road. Five Bluetooth stations comprising MAC1001TR – MAC1005TR (where access to data were first granted) were analysed to explore monthly variations over six months for the period 1st October 2011 to 31st March 2012, as well as exploring speed/flow relationships.

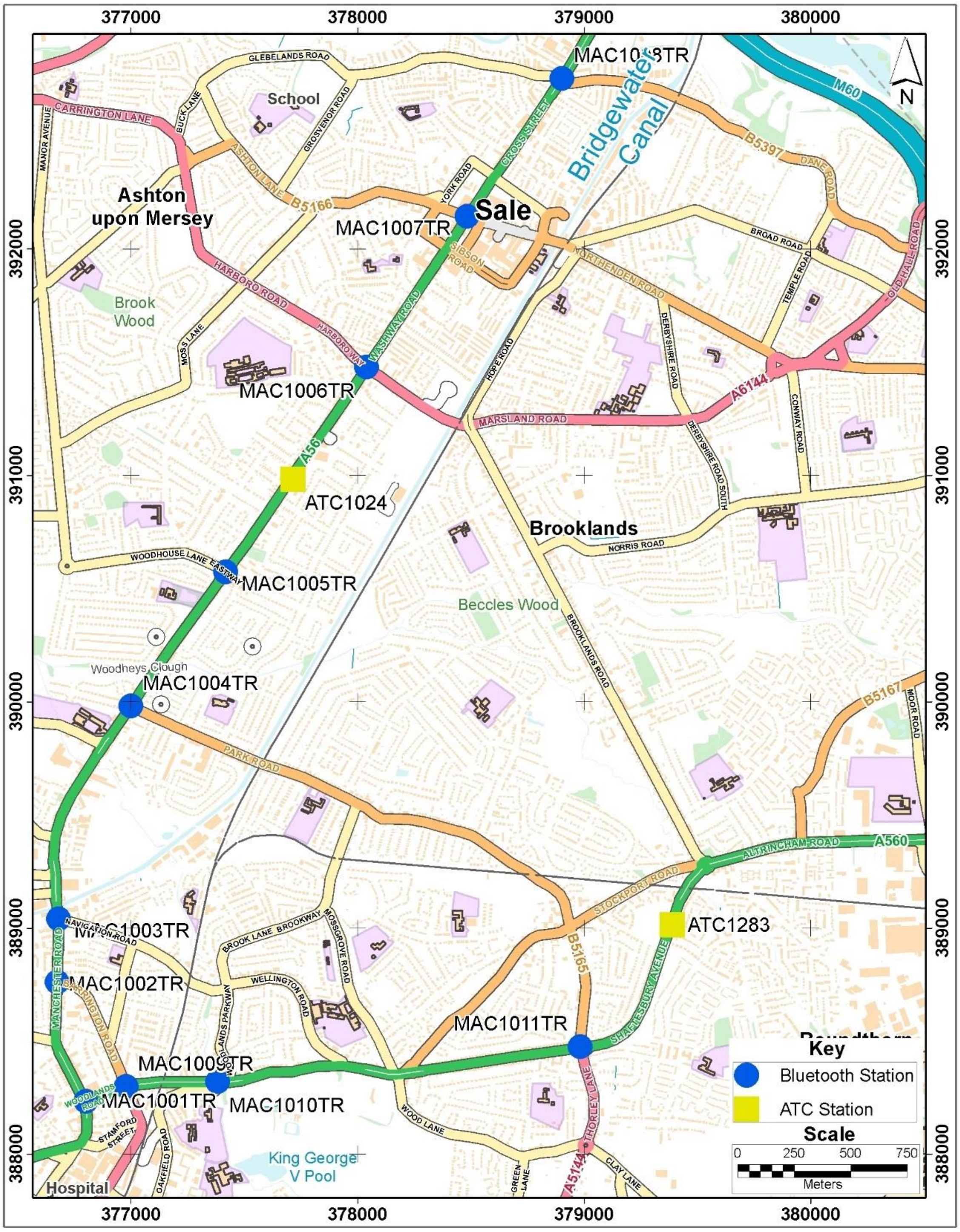


Figure 4.14: Location of Bluetooth sensors and ATC in the Trafford study site

4.7.2 Understanding monthly flow levels

Having gained an initial understanding of the daily flow levels analysed in study site 2, six months of 15-minute Bluetooth average flow collected were analysed to understand consistency and temporal variation. Table 4.10 presents the monthly correlation analysis, while Figure 4.15 presents the profiles of the flow showing a clear consistency over the period. However, temporal variations were observed particularly at peak periods as expected. Flows in the months of October, November and March are slightly above the average while the flows in January, February, and December were slightly below the average flow. The correlation analysis performed presents a better understanding of the monthly flow. The highest correlation (correlation coefficient - 0.987) was reported between the months October and November. The least correlation (correlation coefficient - 0.971) between December and March is attributed to holiday in the period. However, the range of the correlation coefficients (0.015) showed that the difference is not significant. Therefore, the average flow over the period (six months) may well be representative of a typical monthly flow level. The consistency observed in the data from day to day and over months with a strong positive correlation $(r \geq 0.97)$ is indicative of a level of reliability in the data. This consistency in the data is highlighted in the work of Biora *et al.* (2012). This type of consistency is necessary for efficient traffic models to characterise the network.

	Jan	*Feb*	*Mar*	*Oct*	*Nov*	*Dec*
Jan	1					
Feb	0.982	1				
Mar	0.976	0.974	1			
Oct	0.979	0.986	0.981	1		
Nov	0.984	0.986	0.980	0.987	1	
Dec	0.985	0.978	0.971	0.976	0.979	1

Table 4.10: Correlation analysis for six months average flow in Trafford

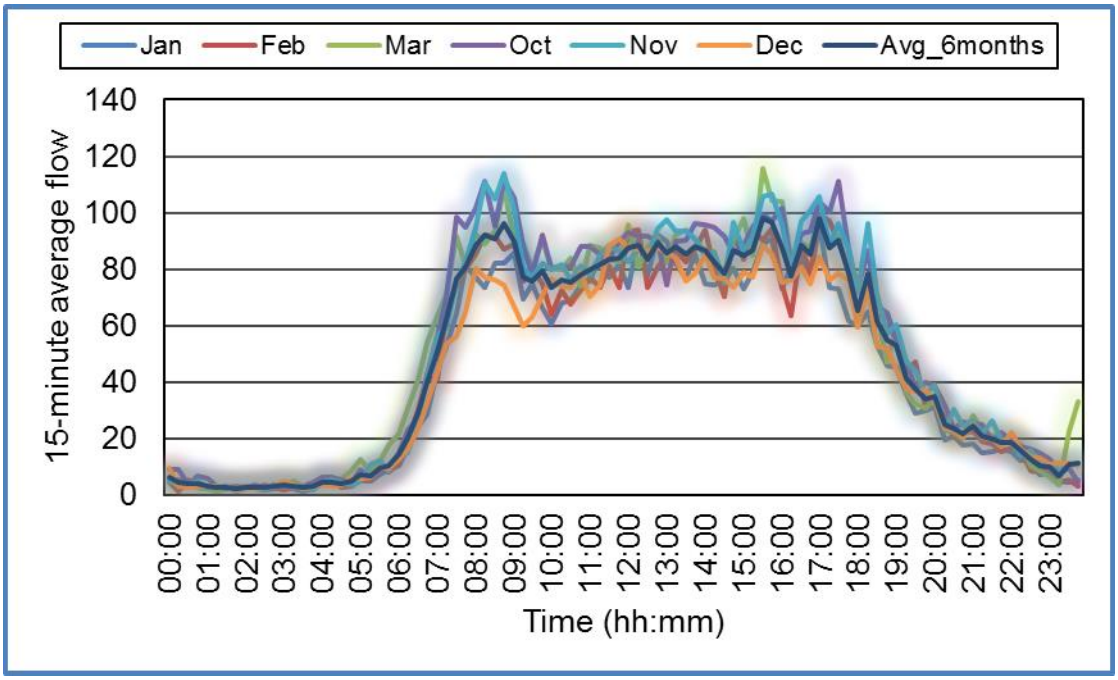


Figure 4.15: Average flow for six months (Oct 2011 – Mar 2012) at MAC1001TR at Trafford

4.7.3 Estimation of the link volume

Table 4.11 presents the summary of the analysis of the detected devices from Station MAC1001TR through Station MAC1005TR designated as Stations 1- 5 in the subsequent text. The second column presents the unfiltered MAC devices detected at a station; while the third column shows the number of duplicates present at each station. The column of the matched records presents the number of the MAC devices detected at two consecutive stations. The filtered column shows the number of unique MAC devices captured at a station over the day following the application of the boundary filtering condition, and the exclusion of the duplicate records. The column of the link volume presents the number of vehicles in each direction following a directional classification as described in the methodology. The summation of the directional flows equals the number of the filtered records in both directions. The results showed that the traffic volume is greater in the opposite direction for all the links, which points to the area of higher activities. The difference in flows was examined as shown in Figure 4.16, which shows a typical flow on the link. Other matches carried out between Station 1 to Station 5 showed that the two detectors furthest apart have the lowest match rate as seen on Link15. The reasons for this can be largely due to drivers making use of the bypass and rat running in the network.

Station No	*No of unfiltered records*	*No of duplicate records*	*Matched records*	*Filtered records ($6 \leq V \leq 120$) (V in Km/h)*	*Link distance (m)*	*Link volume (No of Vehicles/day)*	*Link*
1	4092	929					
2	3628	773	2937	1,257	540.65	403; 854	12
3	4142	881	4,279	1,875	278.05	669; 1,206	23
4	6546	1786	3,777	1,508	994.87	478; 1,030	34
5	2996	495	3,767	1,858	726.09	626; 1232	45
1	4092	929	1,172	586	2,539.66	172; 414	15

Table 4.11: Summary of the link volume analysis over the Trafford network

Figure 4.16: MAC1001 located at the junction of Church Street, A56 Trafford

4.7.4 Understanding speed and travel time patterns

For the time-of-day speed distribution, links through Stations 3 to 5 are designated 30mph (approximately 48km/h roads). It was observed that very few vehicles violated the speed limit especially at midnight and between 12 noon - 2 pm on Link34 (Figure 4.17). Despite the violations observed, this result shows a high level of speed compliance in the area. On Link12, which is a dual carriageway, a higher level of speed was observed (Appendix 4F) compared to Link34. These types of results show that Bluetooth data can be used to infer speed patterns within the network to aid policy formulation such as emission, safety, and economic policies. Further statistical analysis of the travel time from Station 1 to Station 2 shows that it is positively skewed with a value of 4.41, with its mean and standard deviation as 63.67 and 42.94 respectively. The traffic profile of Link12 shows the most populated cluster of vehicles at about 11-12 noon on the day signifying the most congested period along the stretch of the road. Since congestion patterns are expected to be more pronounced during the peak periods than in the off-peak, the pattern observed on this day may be due to an incidence occurrence. Therefore, incident monitoring is another possible application of Bluetooth.

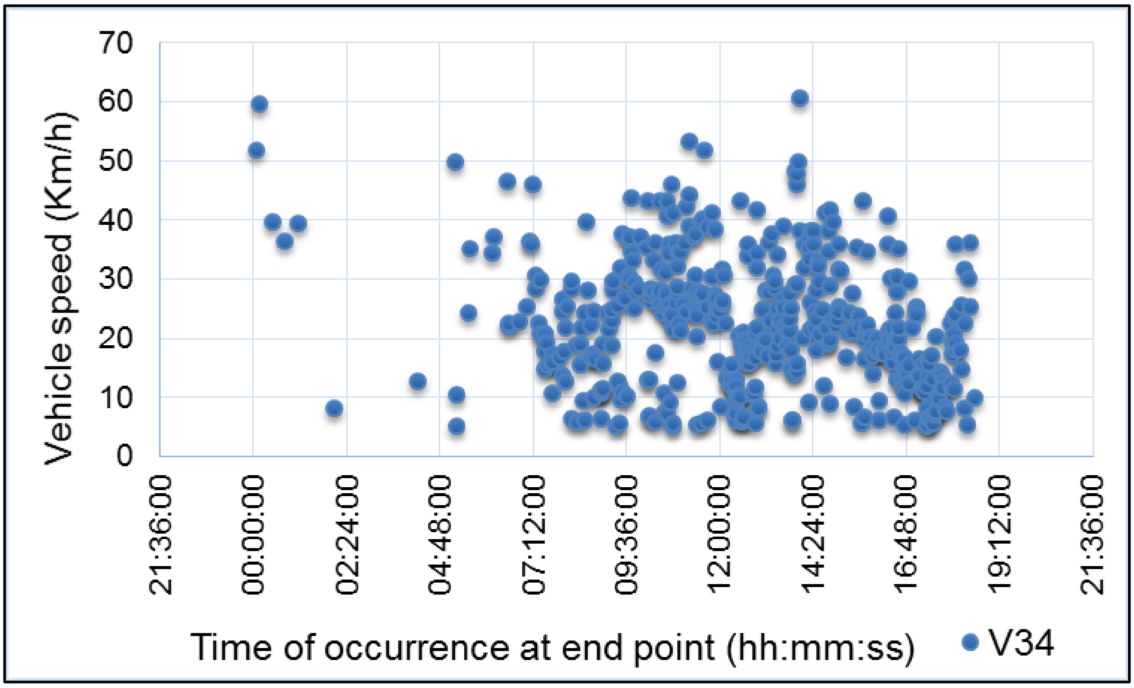


Figure 4.17: Speed distribution over hours of the day from Station 3 to Station 4 in Trafford

V34 is the speed from station 3 to 4.

4.8 Conclusions

A description of the data collection for Bluetooth traffic metrics estimation was presented. The description consisted of three pilot studies: Liverpool, Birtley, and Greater Manchester all in the UK. The study conducted in Liverpool showed that the quality of Bluetooth data is sufficient to estimate traffic metrics. The Birtley study showed that Bluetooth has the potential to identify traffic patterns through the analysis of trips of commuters. The Manchester study built on the results from the Birtley study in an area-wide context to demonstrate transferability. More Bluetooth matches were detected between closer stations than stations farther apart as expected within an urban network that may at times experience rat running or the use of side roads for other activities. Generally, the Manchester study revealed higher traffic volumes in Stockport and Trafford (Sites 2 and 3 respectively) compared to Wigan (Site 1). The preliminary results obtained showed that Bluetooth could provide a viable means of acquiring origin-destination information that has been difficult and expensive to acquire in the past. The results also showed a high level of consistency typified by strong positive correlation coefficient $(r \geq 0.80)$. The characteristics' peak and off-peak nature of normal traffic were equally observed in the data. This suggests the ability of Bluetooth data to represent the actual traffic. The possibility of this application means that Bluetooth provides the platform to acquire traffic data in a cost-effective way, thereby contributing to the delivery of sustainable transport systems. At this stage, Bluetooth data is believed to possess the potential for traffic management applications. The next chapter discusses the validation of the results for large-scale applications based on the concept of these pilot studies.

Chapter 5. Validation of Results

5.1 Introduction

In Chapter 4, the preliminary analysis of Bluetooth data was performed on a short-term scale to gain an initial understanding of the data quality and its potential application in traffic metrics estimation. Chapter 5 builds on the pilot studies presented in Chapter 4 with a specific objective of validating the results obtained to establish the level of accuracy of the data. This is in fulfilment of the Research Objective iv as stated in Section 1.4. The validation of the Bluetooth results in this chapter utilises diverse independently measured traffic data obtained from ATC, SCOOT, ANPR and Traffic Master (TM) In addition, different validation techniques were used to assess the results from the long-term study to ensure sound and robust judgement and maintenance of fit for purpose concept. This is because there is a limited knowledge on the accuracy and reliability of Bluetooth data conducted based on field tests. This validation is also necessary because the available bespoke commercial software for Bluetooth traffic metrics estimation is presently not accessible to the public. Therefore, this chapter examines the question of whether Bluetooth data is accurate enough to provide essential traffic metrics that include travel time and speed. Hence, the following specific objectives are considered: i) calibration of the traffic metrics estimation model (TRAFOST) developed in this research; ii) validation of results using diverse independently measured traffic data; and iii) modelling of the results using ARIMA models to understand the predictive capability of Bluetooth data. The subsequent sections present the discussion of the calibration and validation.

Chapter 5 has the following structure: Section 5.2 presents the calibration of TRAFOST before the validation of the estimated metrics using independently measured traffic data sets. This calibration is to ensure the validity of the model outputs before any comparison of its results with other data sources. Three steps contribute to the calibration namely i) the use of independent computation; ii) the use of C2-Web outputs; and iii) cross validation using the model outputs. The validation of results using diverse measures of traffic data is presented in Section 5.3. The analysis focused on the links where simultaneous capture of Bluetooth and independent measurement of traffic data were possible. In all, four specific links were investigated on a directional basis to ensure better understanding and clarity of purpose on the use of Bluetooth data. The estimated speed was further verified where possible using live traffic information (example in Appendix 5A). Overall, the assessement is essential to establish the validity of Bluetooth estimation by establishing its relatonship with the "true" value. Section 5.4 presents the results of Bluetooth estimation based on ARIMA models to conclude the validation process before conclusions are drawn in Section 5.5.

5.2 Calibration of TRAFOST

5.2.1 Calibration of the model outputs against independent computation

This section describes the calibration of TRAFOST against an independent computation utilising the Excel model (manual computation). The independent checks introduced in the calibration is to detect and correct for any likely difference or error in the TRAFOST-generated results. That is, the ability to reproduce the independently generated results is a way of building proof into the model. However, where necessary, consultations were made to TfGM and TDC for clarifications of results. Table 5.1 presents the summary of such comparisons. From the table, all the metrics from the two models present a high level of precision with standard errors (0.298, 0.226, 0.095) for flow, journey time and speed respectively. The maximum difference being: flow ($5veh/hour$), which occurred during the peak period; journey time ($4s$); and speed ($1km/h$). An important observation is that TRAFOST-derived metrics is consistently higher throughout the day. This difference is attributed to approximations and iterations in TRAFOST, and not the presence of systematic errors. TRAFOST is adjudged to be correct due to the reproducibility of the previous results and the day-to-day precision between the two methods of estimation.

	Manual Estimation			TRAFOST Estimation			Difference in Estimation		
Period (Hour)	Volume (Veh/h)	Journey Time (s)	Speed (km/h)	Volume (Veh/h)	Journey Time (s)	Speed (km/h)	Volume (Veh/h)	Journey Time (s)	Speed (km/h)
0	15	41	48	15	41	49	0	0	-1
1	5	43	44	5	43	44	0	0	0
2	3	35	53	3	35	54	0	0	-1
3	4	54	40	4	54	41	0	0	-1
4	10	51	42	10	51	43	0	0	-1
5	13	33	56	13	34	57	0	-1	-1
6	44	38	50	45	42	50	-1	-4	0
7	119	45	44	121	47	45	-2	-2	-1
8	136	52	38	137	53	39	-1	-1	-1
9	132	46	43	135	48	43	-3	-2	0
10	184	50	40	184	51	41	0	-1	-1
11	182	49	41	186	51	41	-4	-2	0
12	167	43	45	167	43	46	0	0	-1
13	158	46	42	158	47	43	0	-1	-1
14	192	51	40	197	54	40	-5	-3	0
15	168	56	36	170	57	37	-2	-1	-1
16	170	47	41	170	48	42	0	-1	-1
17	134	45	43	134	46	44	0	-1	-1
18	136	44	44	139	47	44	-3	-3	0
19	104	45	45	105	47	45	-1	-2	0
20	62	45	45	63	48	45	-1	-3	0
21	56	45	44	56	46	45	0	-1	-1
22	46	40	49	46	40	50	0	0	-1
23	12	38	50	12	38	50	0	0	0

Table 5.1: Results of the model calibration against independent computation

5.2.2 Calibration of the model against C2-Web outputs

Another assessment of the validity of the model developed in this research considers a comparison of the model estimation of traffic counts with those obtained from C2-Web. C2-Web is commercial software developed by Drakewell/TDC used by TfGM for Bluetooth traffic analysis. A month of data (July 2013) as available from the Wigan study area was used to carry out this exercise. Wigan was used in this case primarily due to the configuration of the road network connecting the Bluetooth stations (1022 and 1023) relative to the validation station (ATC1074) as presented in Figure 4.8 (Section 4.5.1). Figure 5.1 presents the scatter plots and the adjusted R-squared of the weekday's traffic counts over the month. The results of the calibration showed that there is a strong positive relationship ($r > 0.8$) between the C2-Web software and TRAFOST. An examination of the reason for the difference in estimation showed that the C2-Web estimation was without exclusion of any Bluetooth-enabled device. Notwithstanding, both software (C2-Web and TRAFOST) showed a perfect agreement when compared on the basis of total devices captured. Also, despite the observation in the C2-Web results, an independent check has been provided for TRAFOST at the traffic count level.

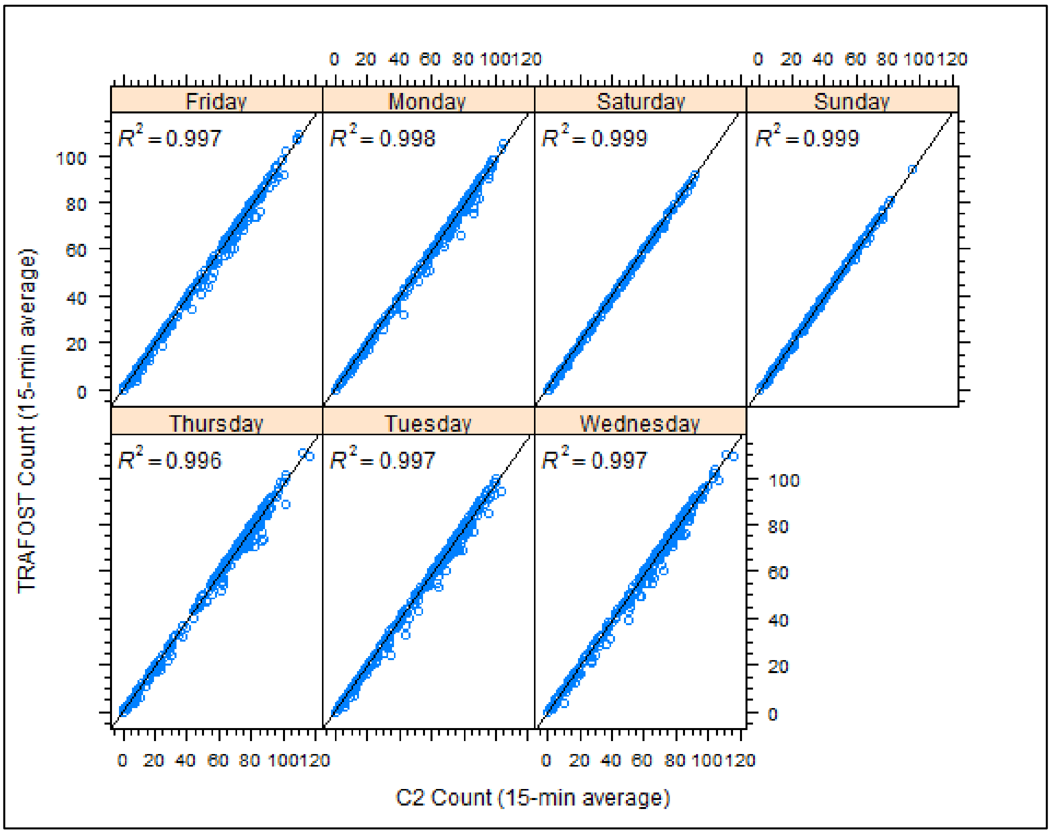


Figure 5.1: Calibration of TRAFOST against C2-Web count on Link2223 in Wigan over July 2013

5.2.3 Cross-validation using journey time and speed results

In order to ensure a high level of reliability in the estimation, cross-validation was incorporated into the verification exercise to reveal the presence of any systematic errors in the estimation. This process serves as an external check and by so doing building further proof into the process. In this case, journey times and speed results were used to provide the proof given that journey time and speed curves are expected to produce a close reflection of each other. The validity of this proof provides confidence in the estimated metrics. Also, cross-validation is considered very useful as any mistake and/or systematic errors in either of the metrics can be discovered thereby making the process robust. This concept was extended to the examination of speed-time plot (Appendix 5B). The speed-time graph is expected to produce a hyperbolic curve whose area under the curve defines the distance travelled. That is, the distance travelled by the individual vehicles or the average over time is expected to be approximately equal to the actual link distance. The hyperbolic curve produced by the plot conforms to the expectation, thereby building another level of confidence in the estimation model. These theoretical concepts were all considered in the design and verification exercise to further assess the accuracy and reliability of the model and the derived metrics.

Figure 5.2 and Figure 5.3 show the hourly distributions of journey times and speed over the month of July on Link3435 in Stockport. From the two graphs, it is evident that they both produce a mirror reflection of each other as postulated. Both plots respectively captured the morning and evening peak periods with a relatively uniform average journey time and speed over the weekend. The highest journey time (52s) was observed on Monday over the morning peak period corresponding to the lowest speed (38km/h). The graphs also showed that the least travel time corresponding to the highest speed for the month was observed over the early and late hours of the day as well as on the weekend. The validity of the model outputs is further justified by the computed relative absolute error of distance $(absolute\ error/measurement) \approx\ 0.03\%$. This shows that the chances of the measurements being in error is less than 1%.

Irrespective of the time taken, all vehicles are expected to travel a distance very close to the link distance (0.511km). Figure 5.4 presents the profile of the distance travelled averaged over hours of the day for the month of July 2013. The 95% confidence limit for the distance is 0.514 to 0.519. Based on the

0.511km actual link distance, the result obtained over the month is accurate to 1cm level of accuracy both on an hourly and daily basis. The high level of accuracy and precision obtained gives another level of confidence and reliability to the model and the estimated metrics. The next step considers the use of independently measured traffic data for results validation.

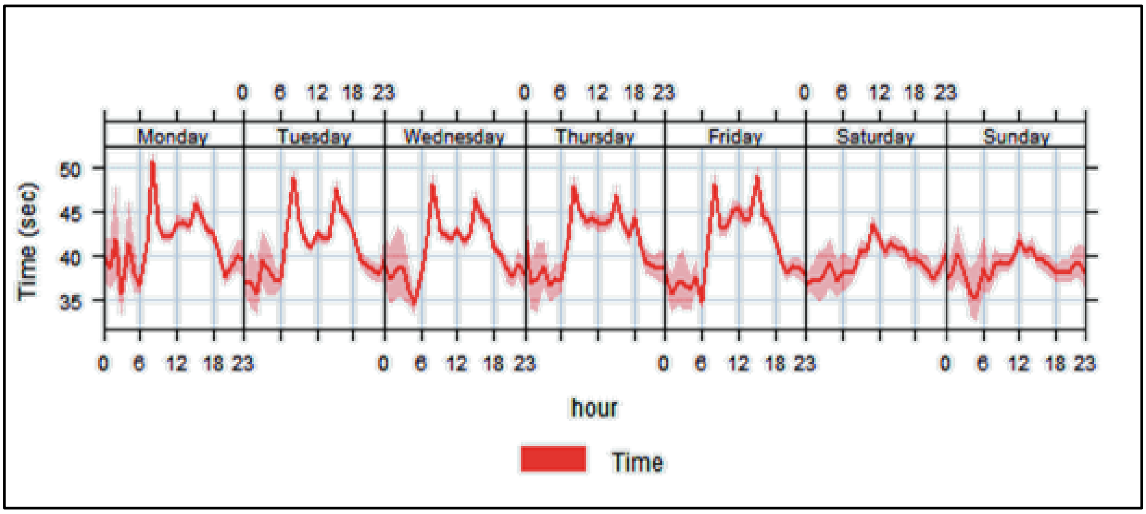


Figure 5.2: Profile of Bluetooth average journey time overlaid with 95% confidence limit over July 2013 in Stockport

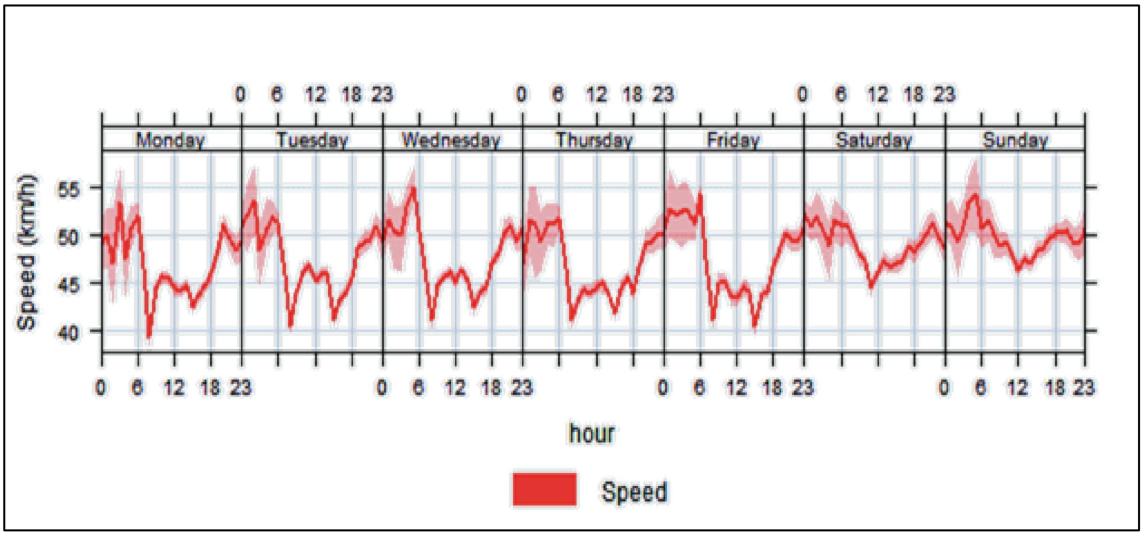


Figure 5.3: Profile of Bluetooth average journey speed overlaid with 95% confidence limit over July 2013 in Stockport

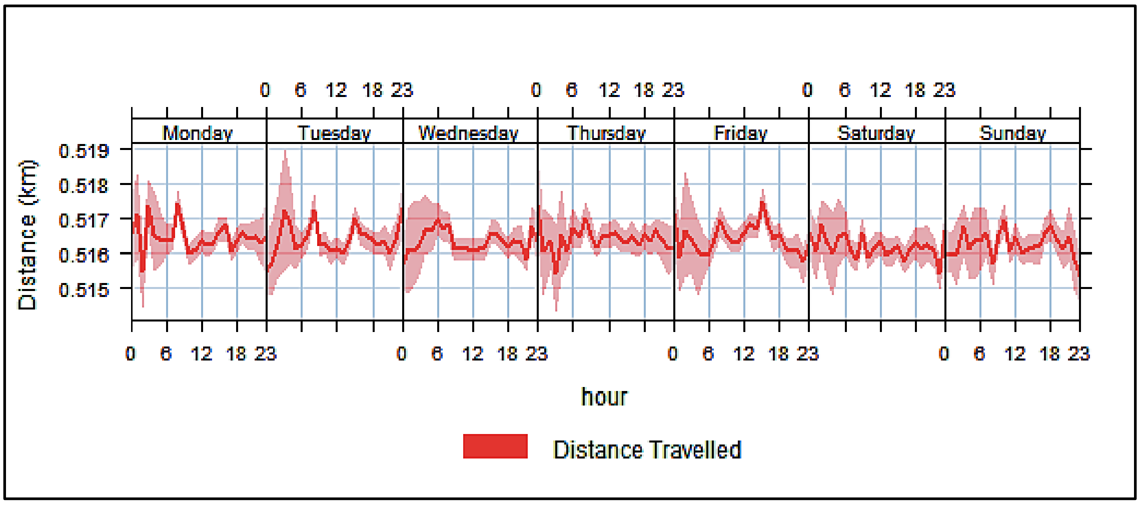


Figure 5.4: Profile of distance travelled overlaid with 95% confidence limit over July 2013 in Stockport

5.3 Validation of Results against Independent Measures of Traffic Data

5.3.1 Validation of flow

This section presents the results of Bluetooth estimated flow validated against the flows measured by three other independent data collection systems (ATC, SCOOT and ANPR) to understand their relationships. The question here is whether Bluetooth can be used to reliably reconstruct the traffic patterns and trends observed in the established systems. As a start, scatterplots and other descriptive statistics were carried out to assess both the direction and strength of the relationships between the traffic flow data collected by Bluetooth, ATC and SCOOT over the weekdays. Table 5.2 presents the coefficients of the correlation analysis performed on the weekday flows for the three variables in both directions, for Stockport and Wigan validation stations. Generally, the analysis of the link flows comparison showed that a strong positive correlation $(r \geq 0.80)$ exists between SCOOT/ATC/Bluetooth flows from day-to-day. This means that where there is no actual flow, Bluetooth data could be used as a proxy measure or to augment the historical data to avoid network failure. Given ATC, the strength comparison over Link3435 and Link3637 showed that higher correlation was observed on Link3637 compared to Link3435 in both directions. Also, Bluetooth/SCOOT presented a stronger relationship due to the values of correlation coefficient compared to Bluetooth/ATC. This result suggests a better performance with SCOOT compared to ATC. The difference is attributed to the spatial location of the SCOOT and ATC detectors relative to the Bluetooth stations. However, the focus of this analysis is not on SCOOT/ATC comparison, The SCOOT links are positioned upstream and downstream of the link close to the Bluetooth locations while the ATC detectors are positioned in-between the two ends of a link. In Wigan, the results obtained ($R^2 = 0.77 - 0.82$) are very similar and are comparable in both directions meaning the same level of confidence can be placed on the observations.

	Adjusted R-Square Based on Location and Variables							
	Stockport, BT3435T/ATC1500		Stockport, BT3435T/SCOOT3435T		Stockport, BT3637T/ATC1013		Wigan, BT2223T/ATC1074	
Weekday	NW	SE	NW	SE	N	S	SE	NW
Mon	0.73	0.77	0.91	0.83	0.81	0.79	0.81	0.79
Tue	0.74	0.76	0.92	0.84	0.81	0.76	0.79	0.81
Wed	0.65	0.67	0.91	0.82	0.79	0.73	0.79	0.79
Thu	0.66	0.71	0.9	0.87	0.8	0.78	0.79	0.81
Fri	0.78	0.78	0.91	0.86	0.78	0.78	0.79	0.82
Sat	0.78	0.76	0.83	0.82	0.78	0.78	0.82	0.79
Sun	0.78	0.78	0.88	0.86	0.80	0.80	0.77	0.80

Table 5.2: The adjusted R-square showing the strength of relationship over weekdays in Wigan and Stockport validation stations

Table 5.3 presents the adjusted R-square values between Bluetooth and ATC over weekdays in Trafford on Link0506 in both directions. The results of the validation showed a strong positive relationship over the days with the adjusted R^2 values ranging from 0.713 – 0.914 for weekdays. The highest value was observed on Saturday (0.914) and the lowest on Tuesday (0.874), giving the knowledge of the level of variability in the weekday flow. The degree of the variability in the data will be explored in the next chapter (Section 6.2.3). The combined directional flows presented higher correlation coefficients, thereby suggesting a better result compared to directional-based analysis and may be preferable. However, total directional flows present less information regarding the level of service (LOS) each way compared to directional flow estimation. Overall, the coefficient of correlation, which explains the amount of variation in the data, coupled with the scatter plots, showed that both data sets are strongly positively correlated. Further analysis of the estimated flows showed that observations taken in two directions can be used to reduce systematic errors as noted by Cooper (1974). The significant increase in the correlation coefficients as observed from Table 5.3 confirmed the validity of this principle in reducing systematic errors. Thus, it is argued that estimation based on total directional flow is preferrable if errors in the estimated metrics are to be minimised. This means that in a network of similar characteristics, directional estimation may not be the preferred option because it may not give any added advantage and could be a waste of resources. Overall, the strong positive relationship between Bluetooth and ATC flows over the Trafford network is consistent with the Wigan and Stockport networks, which is indicative of consistency and the possibility of reliable traffic measurement.

		Adjusted R-Square		
Variables	Weekday	Southbound	Northbound	Combined Direction
BT/ATC	Mon	0.724	0.766	0.889
BT/ATC	Tue	0.713	0.763	0.874
BT/ATC	Wed	0.718	0.758	0.881
BT/ATC	Thu	0.724	0.759	0.883
BT/ATC	Fri	0.753	0.778	0.891
BT/ATC	Sat	0.805	0.811	0.914
BT/ATC	Sun	0.768	0.763	0.894

Table 5.3: The adjusted R-square values between Bluetooth (BT) and ATC at the Trafford validation station

In order to reach a valid conclusion, the flow was further analysed. Initially, a month's worth of data was analysed over the Greater Manchester Network (GMN) for this purpose. This was later extended to twelve months to examine monthly consistency and any seasonal variation. To explore these data sets, the function "*timeVariation*" in the R package "*openair*" (Carslaw and Ropkins, 2012) was adapted to produce four different plots, showing the normalised traffic metrics over four different dimensions to examine temporal consistency.

A normalised time series is presesnted to enable comparison between the two data sets. After standardisation, the aggregation of the data is as follows: i) hourly-weekday (top), ii) hourly (bottom-left), iii) monthly (bottom-middle) and iv) weekday (bottom-right). Figure 5.5 presents the combined plot of Bluetooth and ATC flows to understand their relationships. Interestingly, the day-to-day consistency in the patterns observed in ATC was also evident in the Bluetooth estimation. This consistency includes the capturing of the peak and off-peak periods as well as the weekdays/weekend variations. The absence of coincidence in the results and the consistency in replicating the actual traffic characteristics further highlight the credibility of Bluetooth data.

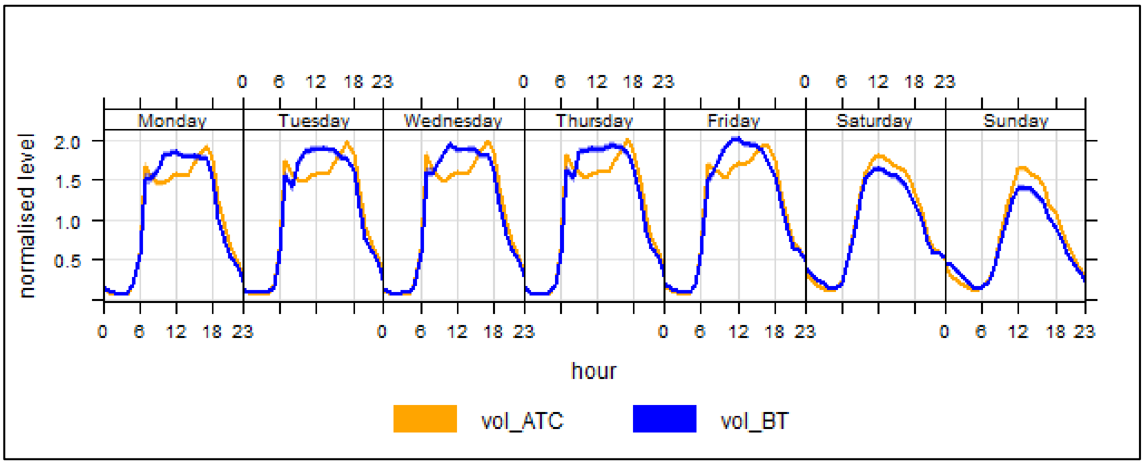


Figure 5.5: Hourly-weekday time series plot of Bluetooth and ATC flows over a year on Link0506 in Trafford (N = 33,646)

Given the similarity in the weekdays' plots and the fact that variabilities are observed over the peak period, a typical weekday's (Monday) average is further analysed and presented below (Figure 5.6) for a better understanding of the relationships between the two data sets. Figure 5.6 presents the normalised profiles of the Bluetooth/ATC flows, showing a high level of precision between the measured flows over the off-peak periods of the early and late hours of the days. However, between the hours of 7am to 6pm, variability is evident from day-to-day and over the months. Further analysis of the results showed that the proportion of Bluetooth to ATC on average is 14%. The histogram and normal plots (Appendix 5C) showed that the distributions are not normally distributed.

Therefore, the Mann-Whitney test (Wilcoxon test), the equivalent of a t-test was employed. The test result showed that the Bluetooth estimated flow is not statistically significantly different from the ATC measured flow at (alpha =0.05) with a p-value of 0.7807 for the test, $\eta 1 = \eta 2$ vs $\eta 1 \neq \eta 2$ and CI (-0.462, 0.419) for $\eta 1 - \eta 2$ for a point estimate of 0.041. As a final step, the Kullback-Leibler divergence (KL-D) was computed for the whole data over the year using the package "*entropy*" in R to compare the closeness or separateness of the distributions. The KL-D value (0.0272) alludes to the closeness of the distributions of the two data sets. Similarly, Figure 5.7 presents the SCOOT flow equivalent showing the normalised hourly flows over the weekday in the NW-direction. For holistic assessment, the combined plot of the directional flows from Bluetooth, SCOOT and ATC is presented in Figure 5.8. Additional results such as the opposing directional flow profiles and scatter plots for Bluetooth/ATC/SCOOT are presented in Appendix 5D. In a nutshell, combining the results from ATC, SCOOT and Bluetooth has led to increased understanding and conviction on Bluetooth-derived flows.

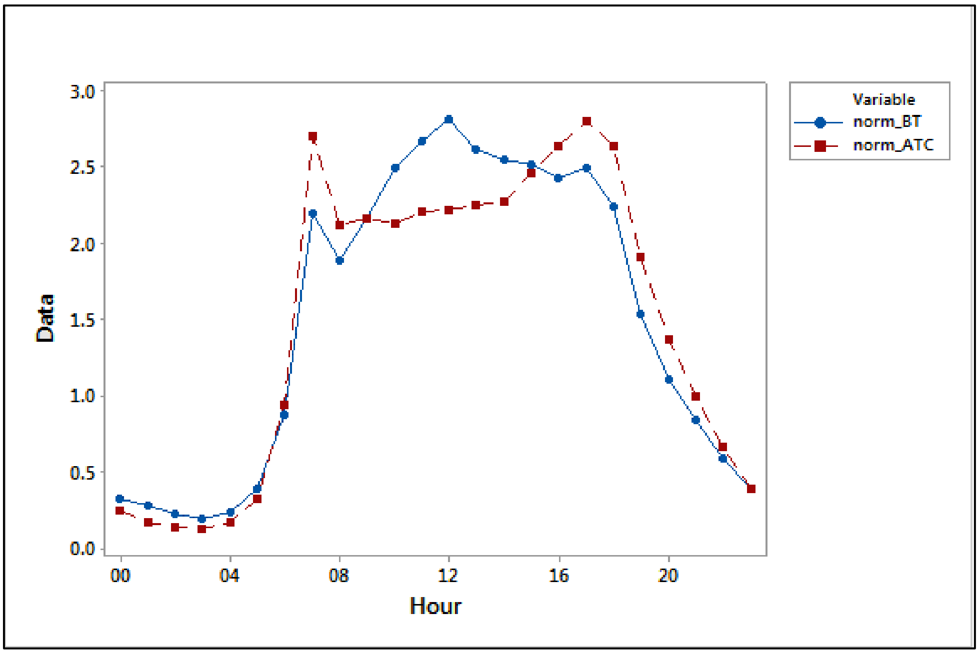


Figure 5.6: Normalised profiles of Bluetooth and ATC hourly flows (all Mondays) in November 2013 on Link0506 in Trafford (N=24)

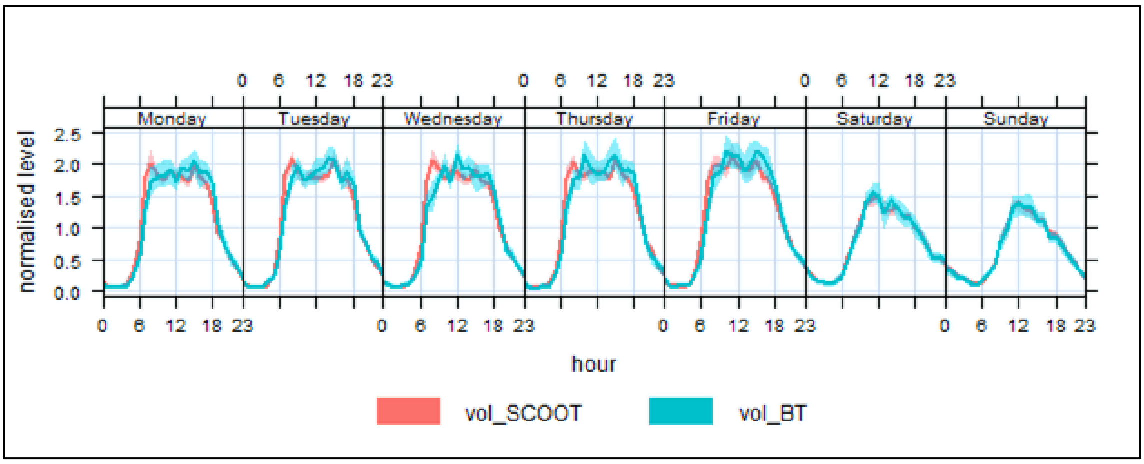


Figure 5.7: NW-directional flow time series profiles of SCOOT and Bluetooth in Stockport (N=2976)

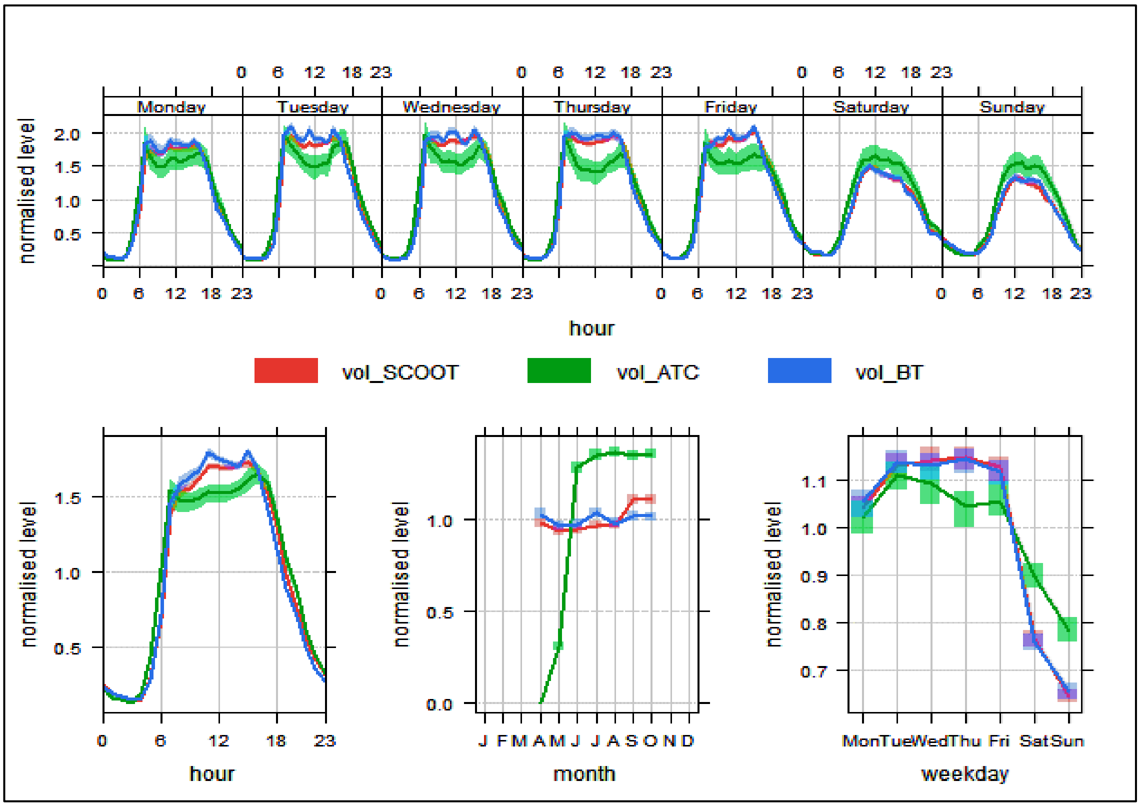


Figure 5.8: Combined normalised NW flow between Bluetooth, ATC and SCOOT on Link3435 over 2013 in Stockport (N=18761)

Figure 5.9 presents the time plot of Bluetooth and ANPR flows to assess the relationship between the two variables. Appendix 5E presents the descriptive statistics for flow, journey times and speed for both ANPR and Bluetooth. Unlike the journey time and speed results presented in the subsequent sections, the flow comparison showed a poor correlation $(R^2 = 0.23)$ between Bluetooth and ANPR. This is primarily due to the data sample – one day of observations, and the temporal dimension used. However, the resultant difference in the trend particularly over the morning hours of about 7 am – 10 am may be due to other factors given that the corresponding estimated journey times and speed are strongly correlated with the ANPR measurements. However, the detection rate from the two flows (12%), falls in the range of the detection rates obtained from both ATC and SCOOT comparison. Detailed discussions on detection rate are presented in Chapter 6 (Section 6.5). While there is a poor correlation between the Bluetooth and ANPR flow data, at this level of the analysis, a conclusion cannot be drawn given that only one-day data was available for the analysis. However, the consistency of the detection rate with SCOOT and ATC-derived rates suggests that with a large sample, there is a possibility of Bluetooth/ANPR augmentation. Table 5.4 presents the summary of the quantitative assessment of Bluetooth flow.

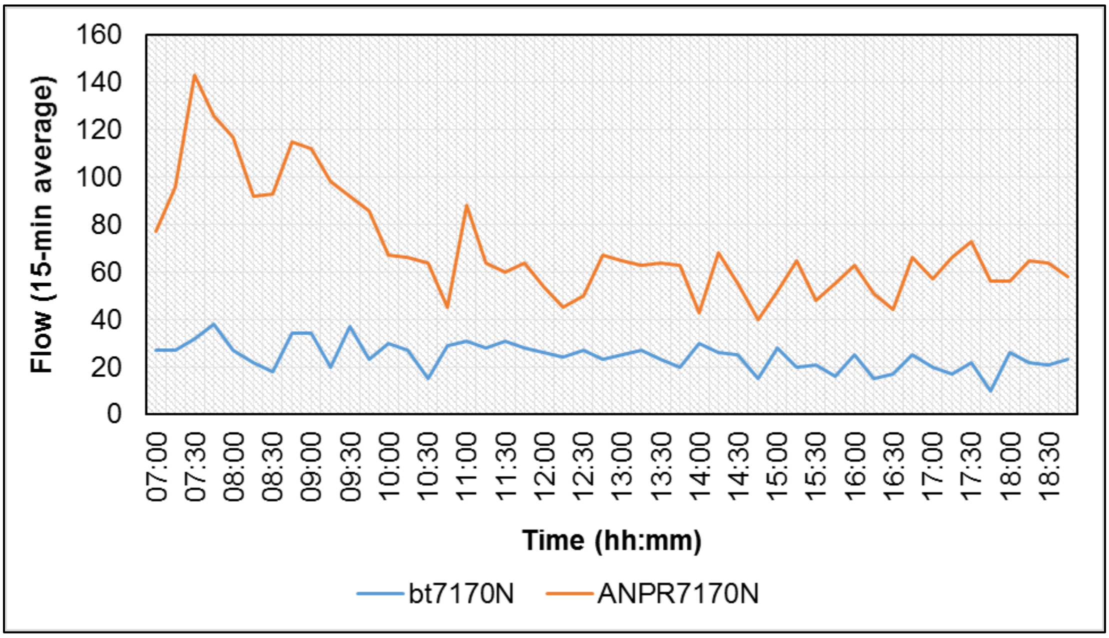


Figure 5.9: Time plot of Bluetooth and ANPR flows of 3rd April 2014 on Link7170 in Stockport (N=48)

Metrics	Point estimate	CI	P-value	KL-D	N	Link
Normalised Bluetooth/ATC Flows	0.04	(-0.462,0.419)	0.781	0.027	24	0506
Bluetooth/ATC Flows	-192.00	(-206.00,-175.00)	0.000	0.028	384	3534
Bluetooth/ATC Flows	-197.00	(-210.00,-166.00)	0.000	0.022	384	3435
Normalised Bluetooth/SCOOT Flows	-0.06	(-0.4708,0.1977)	0.452	0.025	24	3435
Bluetooth/SCOOT Flows	-129.00	(-148.99,-102.00)	0.000	0.027	384	3534
Bluetooth/SCOOT Flows	-112.00	(-132.00,-84.00)	0.000	0.044	384	3435
Bluetooth/ANPR Flows	-40.00	(-44.00,-36.00)	0.000	0.043	48	7170

Table 5.4: The summary of flow validation using IMTD

5.3.2 Validation of journey times

Two sets of IMTD (TM and ANPR) are considered in this section. For a quick exploration, Appendix 5F presents the boxplots of both TM and Bluetooth-derived journey times on four routes where data were available for validation in Stockport (A6) and Trafford (A56) in both directions. From the exploration on the A56, the journey times are comparable for both technologies (Bluetooth and TM) showing that Point 33 is an outlying point. On the A6, the SE-bound journey times presented more outlying points as observed in both Bluetooth and TM than in NW-bound, which has in both cases Point 57 as an outlier. Less time is spent along the SE (40s – 50s) compared to NW (40s – 75s). On the other hand, correspondingly similar travel times in the range of 70s – 140s for TM and 78s – 112s for Bluetooth were spent on route A56 in both directions and were both higher than the A6, as will be expected given that it is about twice the length of the A6.

Figure 5.10 presents the scatter plots of Bluetooth against TM journey times on four routes over GMN. The scatter plots present a quick appreciation of both the direction and strength of the two variables to understand the relationship between them. A visual inspection of the graph indicates that all the routes are positively correlated with a stronger relationship on the A6 compared to the A56. Table 5.5 shows the values of the adjusted R^2 for both journey times and speed on the four routes. A further analysis of the routes on weekdays/weekend basis as observed from the table showed that weekdays performed better than weekends in terms of correlation. This observation is connected with the low sample rates on weekends given that both technologies (Bluetooth and Traffic Master) presented samples of the total traffic thereby leading to a low count rate. In both the A6 and A56, the NW/SW-bound analysis presented a better match compared to the SE/NE-bound equivalent. This shows that the observations from the NW/SW flow are more reliable than the SE/NE flows. Following this exploration, the next step considers the time plots of the data to understand spatial relationships.

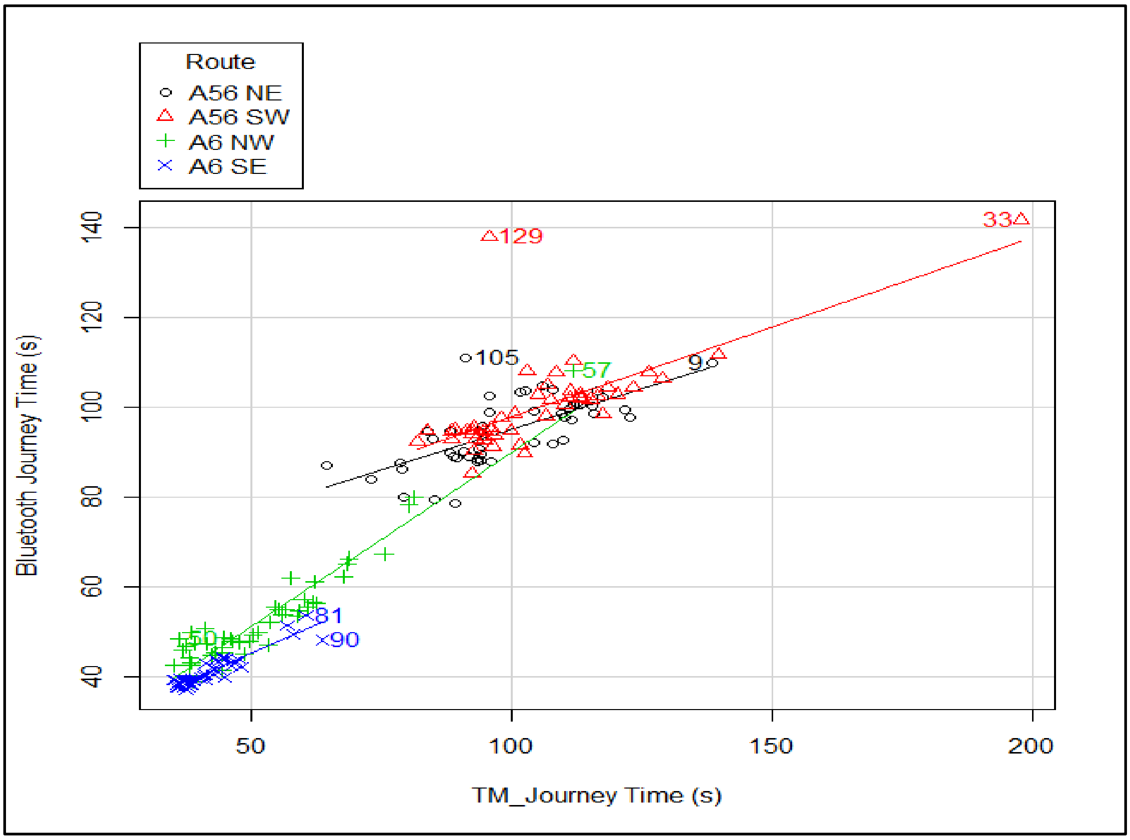


Figure 5.10: Scatter plots of Bluetooth against TM journey times on four routes

Route	Direction	Location	Adjusted R-Square (Journey Times)		Adjusted R-Square (Speed)	
			Weekdays	Weekend	Weekdays	Weekend
A56	NE	Trafford	0.7826	0.4202	0.8267	0.4925
A56	SW	Trafford	0.9231	0.6039	0.7779	0.4875
A6	NW	Stockport	0.9376	0.8339	0.9228	0.8043
A6	SE	Stockport	0.8805	0.6788	0.8933	0.6480

Table 5.5: The adjusted R-square values between Bluetooth and Traffic Master validation for journey times and speed comparison

Figure 5.11 presents the time series plot of journey times for Bluetooth and TM across the A6 route in Stockport. The results of the Trafford network are presented in Appendix 5F. The first section of the plot shows the journey times on weekdays in the NW direction while the second section of the NW series presents the weekend journey times. Obviously and as expected, the weekdays' travel times are higher and with higher variability due to a higher volume of traffic than on the weekend. Similarly, in the SE sections of the profiles, travel times are higher and with higher variability over the weekdays (first part) than the weekends (second part – last section). One key observation is the similarity in trend between the two sensors as observed by Quayle *et al.* (2010) and Haghani *et al.* (2010). However, dissimilarity in trend can be observed at some points in the series, which may be due to a limitation in Bluetooth. Therefore, a quantitative analysis technique was employed to reach a logical conclusion. Table 5.6 presents the summary of the quantitative analysis showing that there is no statistically significantly difference between the two distributions of Bluetooth and TM journey times. The next discussion is focused on the validation of Bluetooth journey times using ANPR measurements.

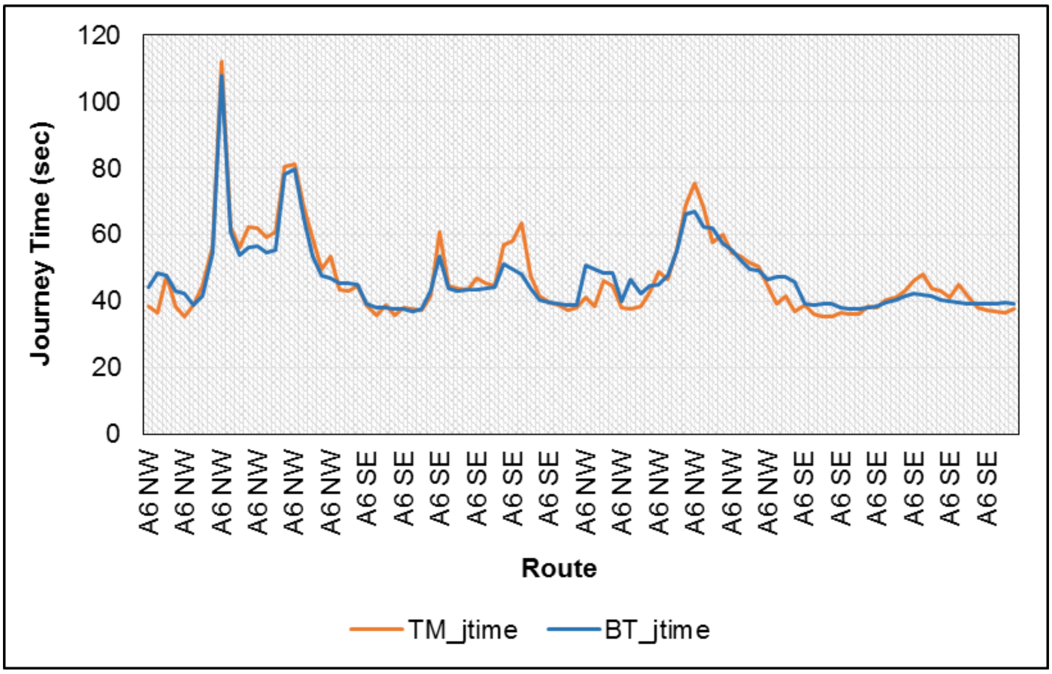


Figure 5.11: Profiles of Bluetooth and TM journey times over six months by Routes in Stockport (N=96)

Figure 5.12 presents the boxplots of both ANPR and Bluetooth journey times on Link7170 in Stockport. The exploratory analysis shows that Bluetooth-derived journey times compared well with the ANPR in many respects such as in skewness (positive – mean greater than the median journey times) of the data and interquartile range (35s – 37s). The similarity in the results as observed by Stevanovic *et al.* (2015) is very interesting giving another level of credence to Bluetooth application in traffic management. Further appraisal of the similarity in the results through scatter plots (Appendix 5G), showed that Bluetooth and ANPR are positively strongly correlated for journey times ($R^2 = 0.71$). Figure 5.13 presents the time plot of the two data sets. The observation started at 7am and ended at 7pm. The journey times for the observations fluctuate between 50s and 200s. From the plot, although there is similarity in trend, variability is much more pronounced than ANPR, over the hours of 3pm – 5pm with intermitent over/under-estimation of travel time. To conclude the analyis, a Mann-Whitney test was performed to understand if there is any significant diference between the two distributions. The test results (point estimate 14.0 and CI (6.0,22.99) - overlap) showed that the two groups are not statistically significantly different from each other at $\alpha = 0.05$. The $KL - D$ (0.006) also showed that the two distributions are similar and are closely related (see Table 5.6 for the summary of journey times validation). In conclusion, it is evident from all the tests conducted that Bluetooth is accurate enough to be used to estimate travel time.

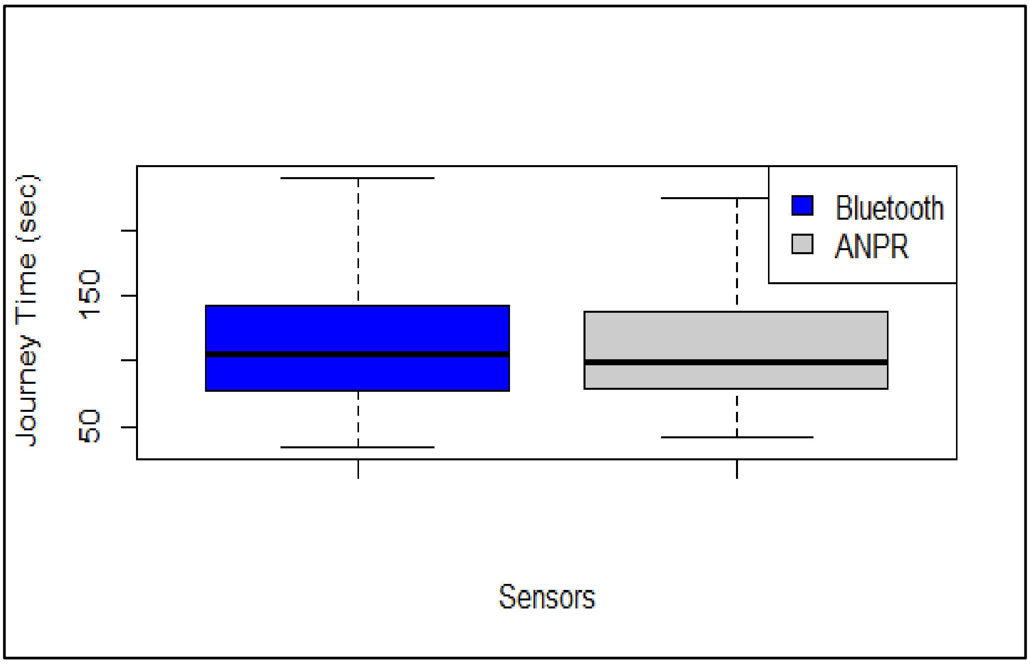


Figure 5.12: Boxplot of Bluetooth and ANPR journey time of 3rd April 2014 on Link7170 in Stockport

Metrics	Point estimate	CI	P-value	KL-D	N	Link
Bluetooth/ANPR Journey Times	14.00	(6.00,22.99)	0.001	0.006	48	7170
Bluetooth/TM Journey Times	0.94	(-1.001,2.751)	0.261	0.004	96	A 6
Bluetooth/TM Journey Times	-4.00	(-7.55,-0.73)	0.015	0.006	96	A 56

Table 5.6: Summary of journey times validation based on IMTD

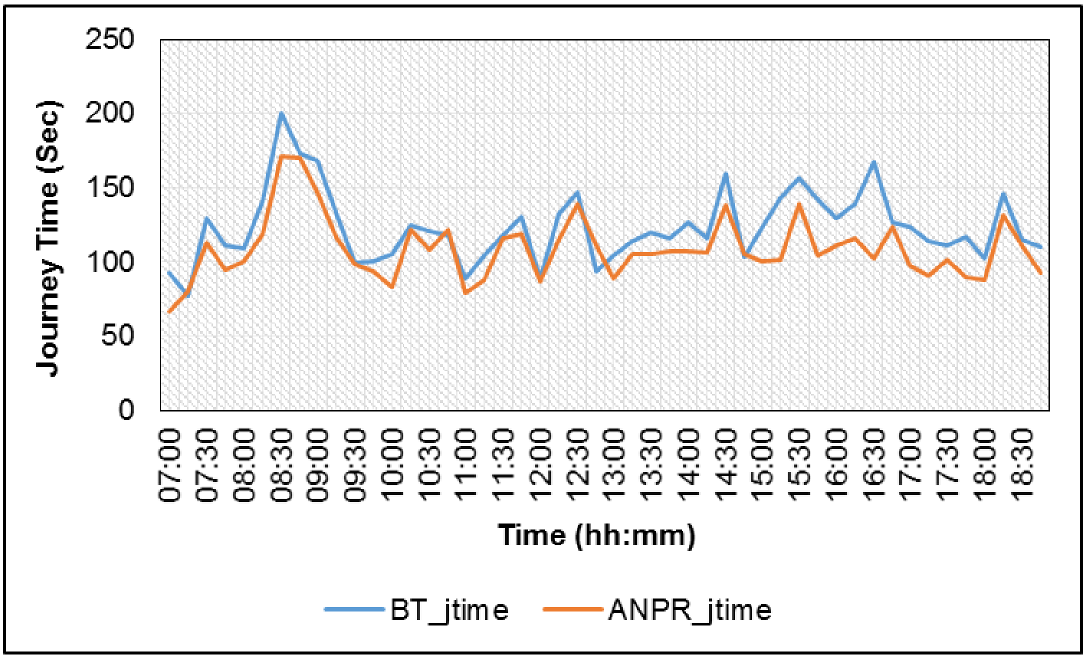


Figure 5.13: Time plot of Bluetooth and ANPR journey times of 3rd April 2014 on Link7170 in Stockport (N=48)

5.3.3 Validation of speed

Figure 5.14 presents the time plot of Bluetooth and TM speeds in both directions (NW and SE) in Stockport. The first section of NW and SE represents the weekdays speed while the second section represents the weekend speed. Across the groups, the speeds fluctuate between 15km/h and 55km/h typifying periods of free flow and congestion. Also, the speed distribution is higher on the A6 with lesser variability compared to the opposing link speed. The weekend speeds are higher in both directions as would be expected. Table 5.7 presents the summary of the quantitative analysis showing that there is no significant difference between the two distributions of Bluetooth and TM journey speeds. The next discussion is focused on the validation of Bluetooth journey times using ANPR measurements.

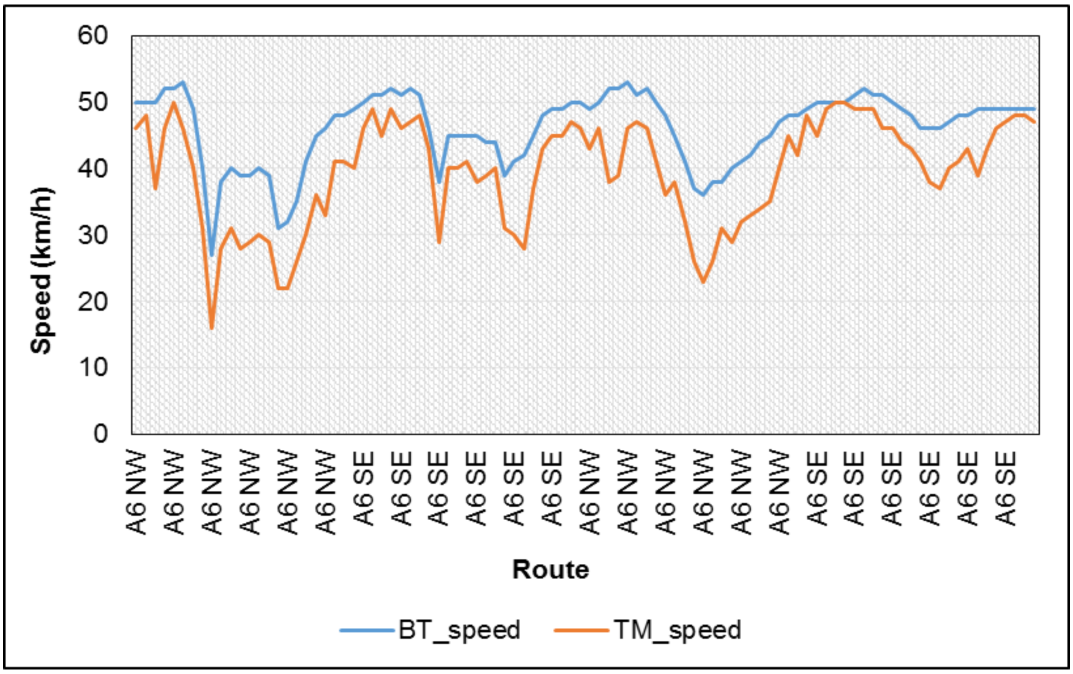


Figure 5.14: Profiles of Bluetooth and TM speed over six months by Routes in Stockport (N=96)

Figure 5.15 presents the time plot of Bluetooth and ANPR speeds for observations starting from 7am – 7pm on 3rd April 2014. The highest variability between the two series occured between 3pm – 6pm. The journey speed for the observations fluctuates between 10km/h and 35km/h. From the plot, although there is similarity in trend as well as evidence of strong correlation ($R^2 = 0.71$), variability is much more pronounced than ANPR, over the hours of 7am – 10am with occasional over/under-estimation of journey speed. To conclude the analyis, a Mann-Whitney test was performed to understand if there is any significant diference between the two distributions. The test results (point estimate (-2.0), CI (-3.0,001) - overlap) showed that the two groups are not statistically significantly different from each other at $\alpha = 0.05$. The $KL - D$ (0.006) also showed that the two distributions are similar and are closely related. Table 5.7 presents the summary of the test statistics. Summarily, the test results showed that Bluetooth is sufficiently accurate to be used for the estimation of speed.

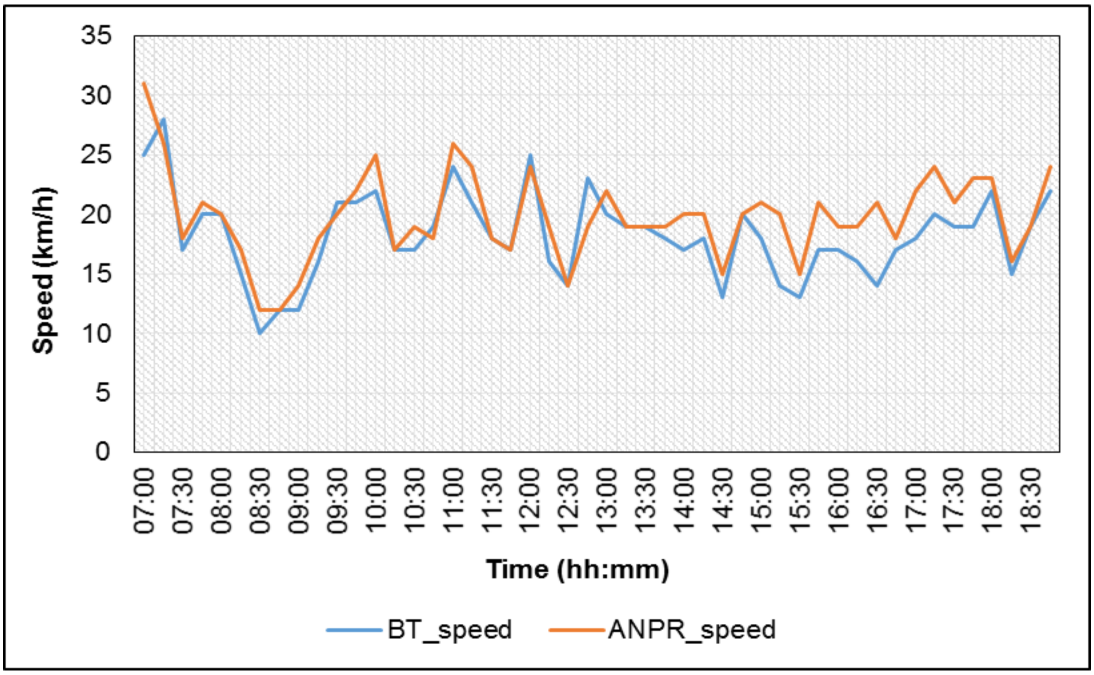


Figure 5.15: Time series plot of Bluetooth and ANPR speeds of 3rd April 2014 on Link7170 in Stockport (N=48)

Metrics	Point estimate	CI	P-value	KL-D	N	Link
Bluetooth/ANPR Journey Speed	-2.00	(-3.000,-0.001)	0.028	0.006	48	7170
Bluetooth/TM Journey Speed	6.00	(4.000,8.000)	0.000	0.006	96	A 6
Bluetooth/TM Journey Speed	7.00	(6.000,8.000)	0.000	0.005	96	A 56

Table 5.7: Summary of journey speed validation using IMTD

5.3.4 Validation of O-D matrix

For the O-D matrix, six months of Bluetooth data (April – September 2013) were analysed over the three networks in Greater Manchester for day-to-day consistency. Across the networks, over 6,000 O-D matrices generated using TRAFOST were analysed. The day-to-day correlation analysis between the matrices showed a high level of positive relationship between the days over the six months. This shows the potential of Bluetooth to support the delivery of O-D matrices using low-cost sensors as demonstrated by Blogg *et al.* (2010) and Barceló *et al.* (2012). Table 5.8 presents an example of such correlation analysis. One interesting thing from this result is the high value of the correlation coefficients compared to those obtained from the link flows. However, this is expected given that the O-D matrix correlations were computed from larger samples compared to the link flows. The next step considers the predictive capability of Bluetooth traffic estimation using the ARIMA models to finalise the validation.

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Sun	1						
Mon	0.96	1					
Tue	0.93	0.98	1				
Wed	0.96	0.96	0.96	1			
Thu	0.87	0.95	0.97	0.91	1		
Fri	0.96	0.98	0.96	0.98	0.94	1	
Sat	0.89	0.96	0.97	0.94	0.96	0.95	1

Table 5.8: Correlation analysis over weekdays in the Wigan network

5.4 ARIMA Modelling of Bluetooth Traffic Metrics

5.4.1 Modelling of flow data

After data splitting, the training and testing samples for flow consist of 26,188 and 6546 data points respectively. Figure 5.16 presents the time series plot of the training sample based on a daily average (for clear visualisation) on Link0506 in Trafford. The same approach was adopted for the processing of the journey times and speed data. The exploration of the flow plotted in Figure 5.16 shows that the mean and variance are not constant (changing with time) due to some sparks, and there is a visible cut off between the first day and 100th day. The exploration also shows that the data exhibit trend and seasonal effect. The presence of sparks and the lack of decay in the plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) (Figure 5.17) portray trend and seasonality. Therefore, a first order regular difference was performed to make the data stationary, and a logarithm transformation to improve the performance of the prediction. The expectation at this level is a model of form $(p, 1, q)$. Figure 5.18 presents the residuals plot of flow after first difference and logarithm transformation showing that the residuals are distributed about the mean, zero although with few sparks.

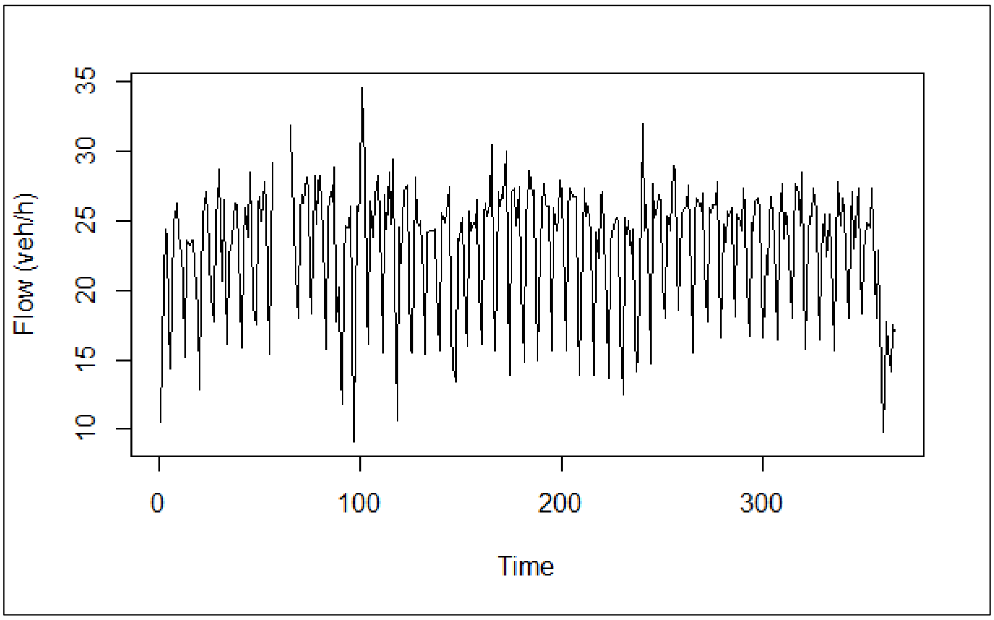


Figure 5.16: Time series plot of Bluetooth flow on Link0506 in Trafford

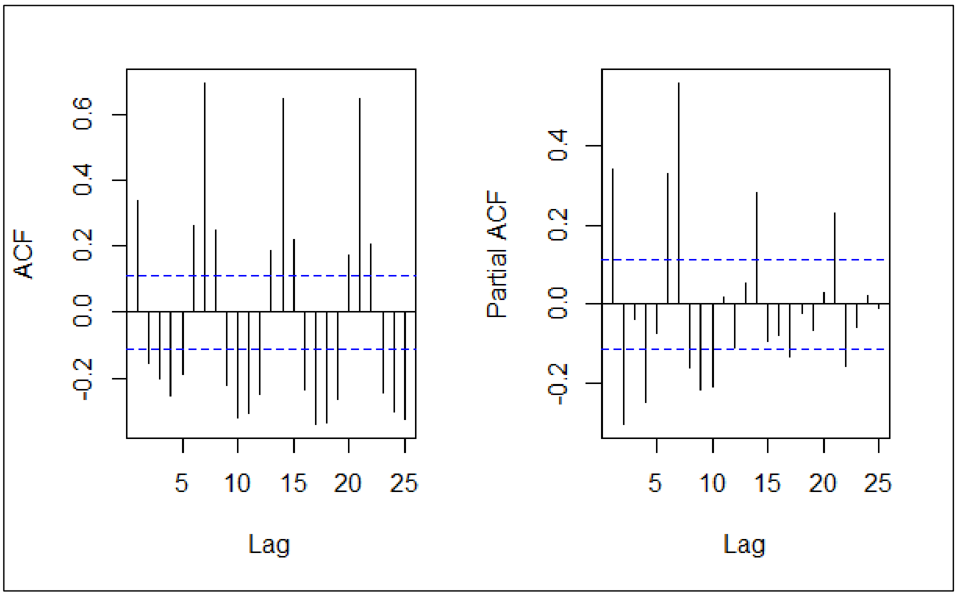


Figure 5.17: Plots of ACF and PACF from Bluetooth flow on Link0506

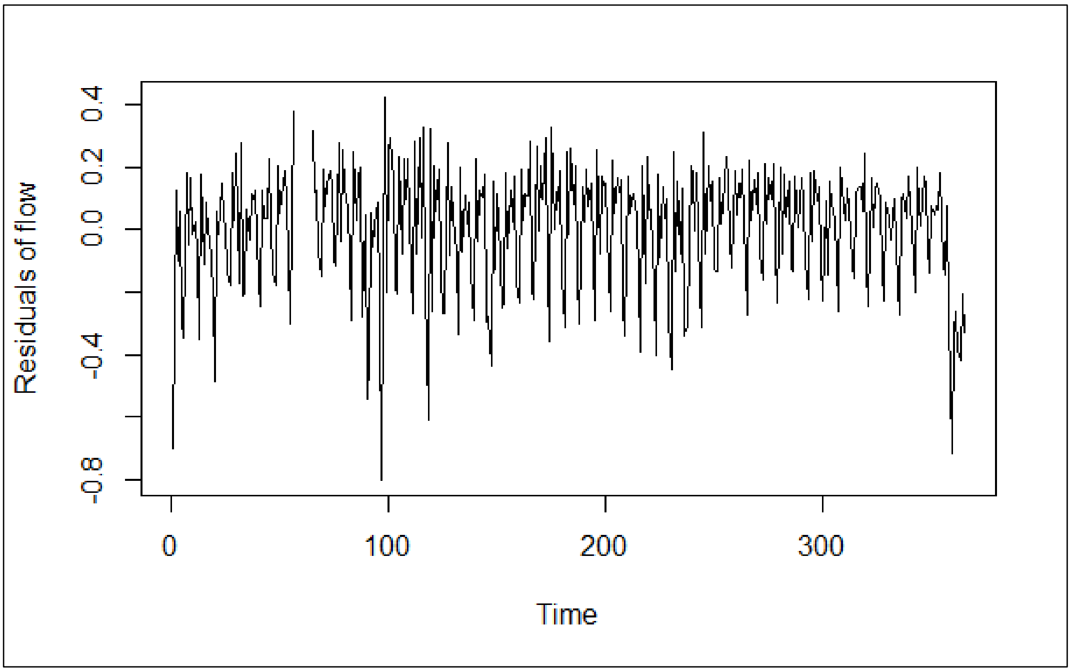


Figure 5.18: Time series plot of residuals of flow after log and first difference transformation

To determine the optimum parameters for the model, the ACF and PACF plots were used as guides. Exploring the plots, the cut off after the first lag in the ACF plot suggests that the AR parameter, p should be zero (0) while the MA parameter, q should be greater than or equal to 1. As a start, a model of the form $ARIMA(0,1,1)(0,1,1)$ of period 12 was postulated due to the presence of seasonal variation. Other combinations were explored, including the use of the auto function in R to determine the best model (i.e the most probable predictive model –$MPPM$). Given the least AIC, a model of the form $(0,1,2)$ with a seasonal component is considered the most parsimonious and adequate model. This model not only presents the least MAE (0.147), MAPE (4.917) and MASE (0.790), but also an RMSE (0.195) comparable to the least value (0.191) among the groups. The MAPE value shows that normally, the forecast will capture 95% of the trend (i.e. 95% accuracy level), and will possibly be off by approximately 5%. Given that the MASE is less than 1 also shows a good performance. However, a MASE of 1.3 was proposed in a competition as a cut-off point (Hyndman, 2006). A portmanteau test to check for the randomness or autocorrelation of the residuals returned a p-value (0.824) which suggests that the residuals are white noise. Consequently, this model was used to make predictions. Figure 5.19 presents the visualisation of the training data with the prediction. Table 5.9 presents the postulated models with their corresponding accuracy statistics.

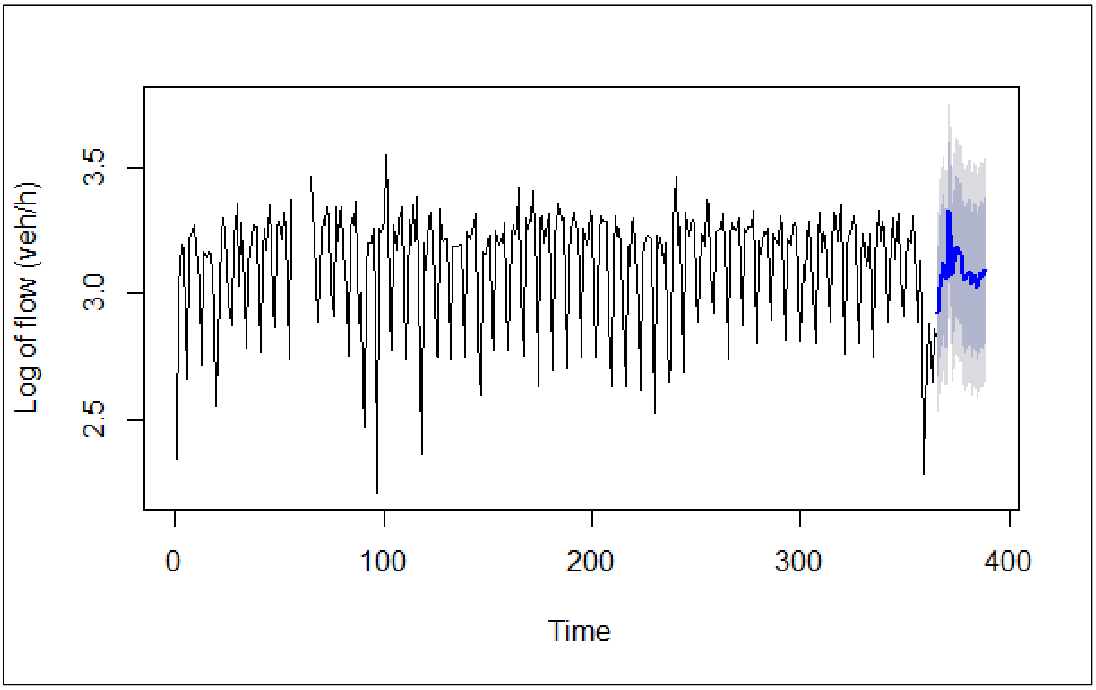


Figure 5.19: The log of flow and the prediction overlaid with 80% and 95% confidence limits

Forecast Series	AIC	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA(1,0,1)(1,0,1)	-152.26	-0.002	0.191	0.150	-0.521	5.043	0.811
AUTO.ARIMA(1,0,3)	-126.77	0.000	0.199	0.162	-0.455	5.426	0.876
ARIMA(0,0,1)	-123.18	0.000	0.202	0.165	-0.445	5.507	0.890
ARIMA(1,1,1)	-95.90	0.123	0.208	0.171	-0.051	5.694	0.923
ARIMA(1,0,2)	-127.78	0.000	0.199	0.164	-0.455	5.463	0.883
ARIMA(0,1,2)(0,1,2)	-70.00	-0.024	0.195	0.147	-1.152	4.917	0.790

Table 5.9: Forecast series and accuracy statistics for flow

5.4.2 Modelling of journey time data

The training and testing samples used consist of 537,226 and 134,304 data points respectively after splitting. Figure 5.20 presents the time series plot of the training sample (daily average) on Link0506 in Trafford. The exploration of the journey time data shows the evidence of trend and seasonal effect. The presence of sparks and slow decay in the plots of ACF and PACF (Appendix 5H and Figure 5.21) portray trend and seasonality. Therefore, a first order regular difference was performed to make the data stationary, and a logarithm transformation to improve the performance of the prediction. The expectation at this level is a model of form $(p, 1, q)$. Figure 5.22 presents the residuals plot of journey times after first difference and logarithm transformation showing that the residuals are distributed about the mean although with few sparks.

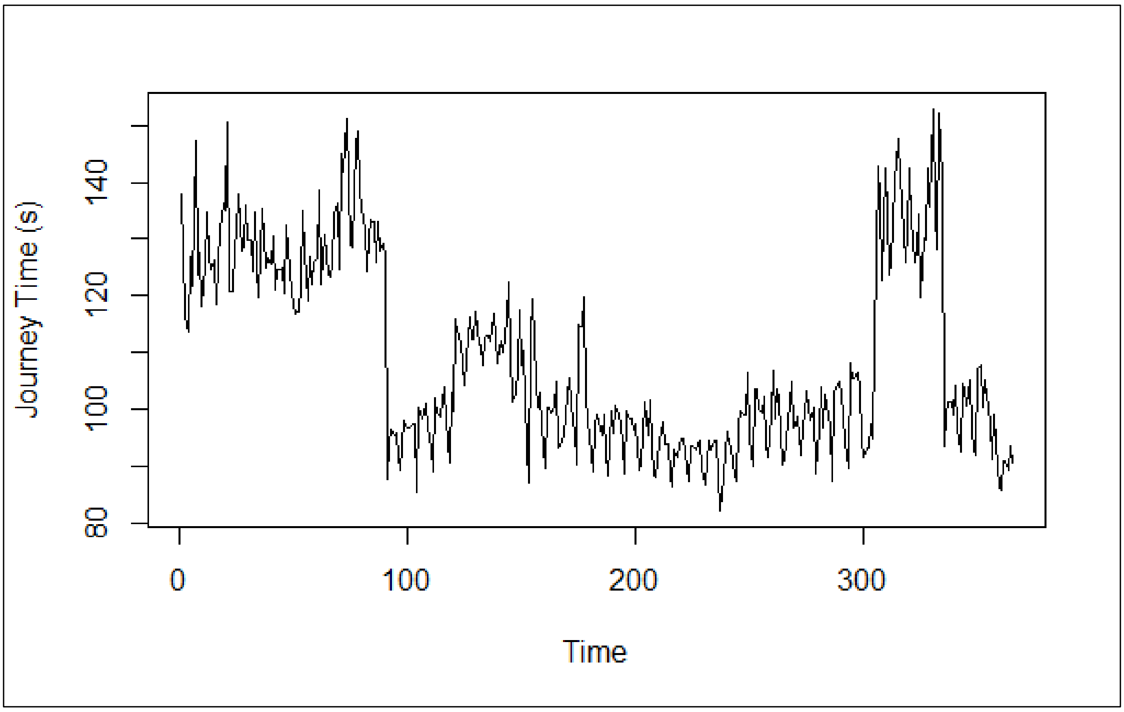


Figure 5.20: Plot of Bluetooth journey time on Link0506 in Trafford (N=365)

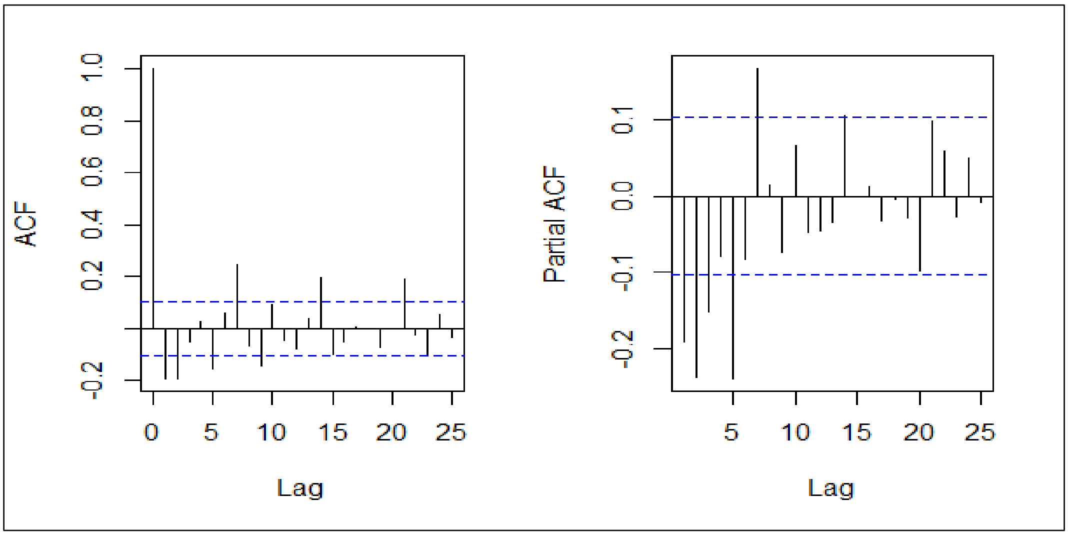


Figure 5.21: Plots of ACF and PACF of Bluetooth journey times on Link0506 after first difference and log transformation

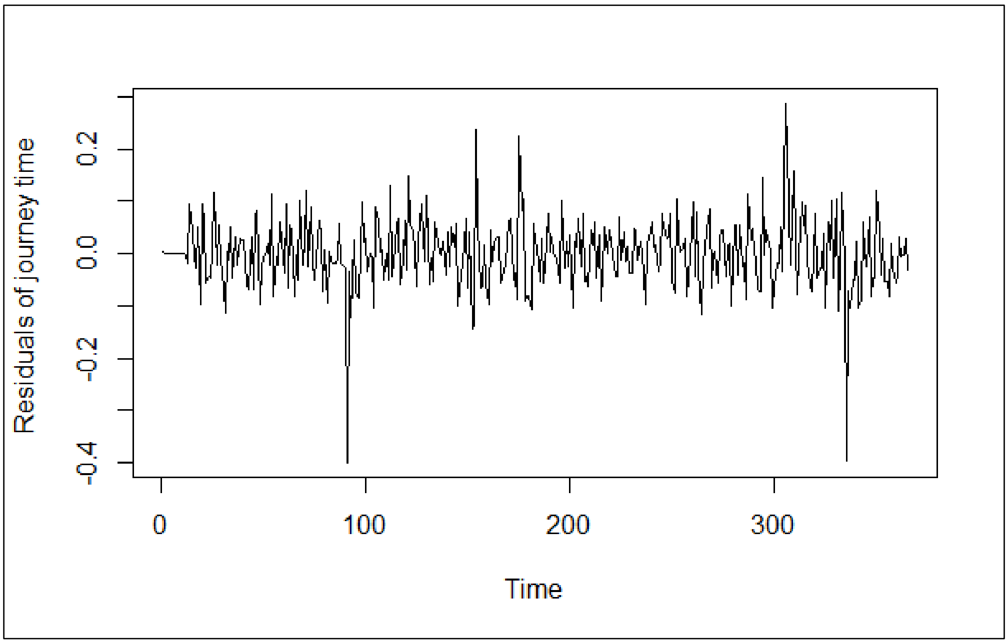


Figure 5.22: Residuals of journey times after log and first difference transformation

As with flow, a model of the form $ARIMA(0,1,1)(0,1,1)$ of period 12 was postulated due to the presence of seasonal variation and the behaviour of the ACF and PACF plots. Other combinations were also explored to determine an optimum model based on the least AIC. A model of the form $(0,1,1)$ with a seasonal component presents the least AIC (-818.23). However, the outcome of the portmanteau test p-value (0.768) suggests the adoption of the model of form $(0,1,2)$ with a seasonal component having an AIC value (-842.45). This model also has a better MAE (0.050), MAPE (1.073) and MASE (0.946), but also an RMSE (0.069) compared to the model with the least AIC. The MAPE value shows that less than 2% of the forecast will possibly be in error. Given that the MASE is less than 1 also shows a good performance. Summarily, all the computed accuracy statistics suggest the validity of the model. Consequently, this model was used to make a prediction. Figure 5.23 presents the visualisation of the training data with the predicted journey times. Table 5.10 presents the postulated models with their corresponding accuracy statistics.

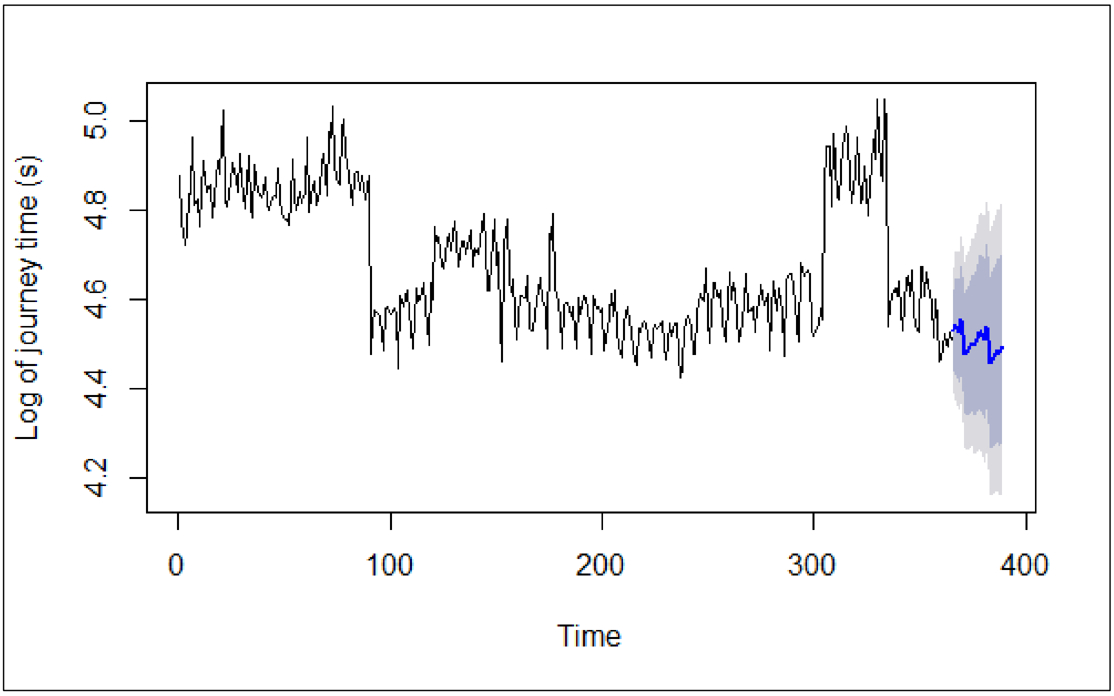


Figure 5.23: Plot showing the log of journey times and prediction overlaid with 80% and 95% confidence limits

Forecast Series	AIC	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA(0,1,1)(0,1,1)	-818.23	0.000	0.071	0.053	-0.154	1.130	0.996
AUTO.ARIMA(2,1,2)	-889.80	-0.002	0.068	0.049	-0.064	1.048	0.926
ARIMA(0,1,1)	-879.50	-0.002	0.071	0.052	-0.055	1.109	0.980
ARIMA(1,1,1)	-894.91	-0.003	0.069	0.050	-0.071	1.065	0.941
ARIMA(0,1,2)	-897.82	-0.002	0.069	0.050	-0.069	1.060	0.936
ARIMA(0,1,2)(0,1,2)	-842.45	-0.001	0.069	0.050	-0.025	1.073	0.946

Table 5.10: Forecast series and accuracy statistics for journey times

5.4.3 Modelling of speed data

The training and testing samples used for the modelling of the estimated speed consist of 537,226 and 134,304 data points respectively after splitting. The same procedure described in the modelling of the journey time was followed. The exploration of the speed data also revealed the presence of trend and seasonality as would be expected, and as is the case with the estimated journey time. Figure 5.24 presents the residuals plot of speed after first difference and logarithm transformation showing that the residuals are distributed about the mean although with few sparks. This observation from the residuals plot points to the practicality of modelling the estimated speed. Also, as with journey times, a model of the form $ARIMA(0,1,2)(0,1,2)$ of period 12 was adopted following a series of combinations to determine the optimum model. This model presents the second least AIC (-1186.73), (the least being -1125.43) from the model of form $ARIMA(0,1,1)(0,1,1)_{12}$. Despite the similarities in the accuracy statistics between the two models, the preference was due to the outcome of the portmanteau test with a p-value (0.668), which suggests the randomness of the residuals and the adoption of the model. A key observation is that the selection criterion or the use of auto.arima to determine the best model may also require personal judgement to determine the optimum model. From the selected model, the MAPE value (0.822) shows that less than 1% of the forecast will possibly be in error. Also, given that the MASE is less than 1 this suggests a good performance. Summarily, all the computed accuracy statistics are small (close to zero) which points to good performance of the model. Table 5.11 presents the postulated models and their corresponding accuracy statistics, while Figure 5.25 presents the visualisation of the training data with the prediction made using the adopted model. Other results including the modelling of the data on a monthly basis are presented in Appendix 5I. Overall, the high level of accuracy obtainable using Bluetooth estimated speed is a significant benefit given that Bluetooth is a low-cost sensor. Therefore, using Bluetooth in this way can contribute to achieving better transport through technology.

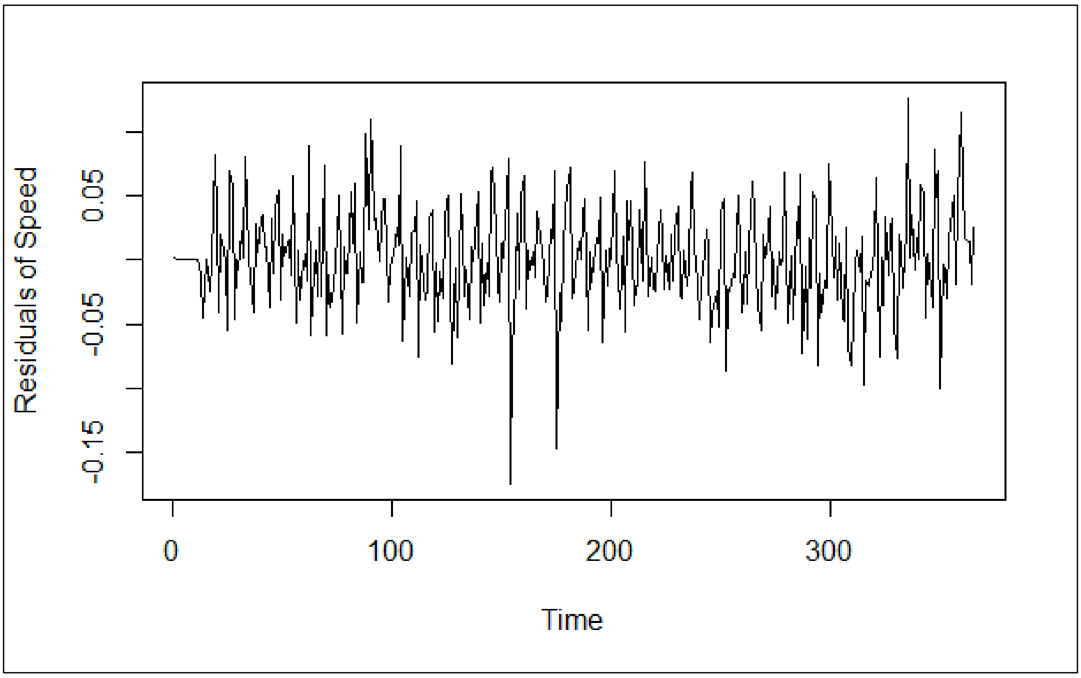


Figure 5.24: Plot of residuals of speed after logarithm and first difference transformation

Forecast Series	AIC	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA(0,1,1)(0,1,1)	-1125.43	0.002	0.045	0.035	0.049	0.913	0.955
AUTO.ARIMA(2,1,2)	-1288.77	0.001	0.041	0.031	0.009	0.833	0.869
ARIMA(0,1,1)	-1215.73	0.001	0.045	0.036	0.008	0.941	0.982
ARIMA(1,1,1)	-1261.71	0.001	0.042	0.033	0.011	0.883	0.921
ARIMA(0,1,2)	-1274.44	0.001	0.042	0.033	0.012	0.868	0.906
ARIMA(0,1,2)(0,1,2)	-1186.73	0.002	0.041	0.031	0.044	0.822	0.858

Table 5.11: Forecast series and accuracy statistics for speed

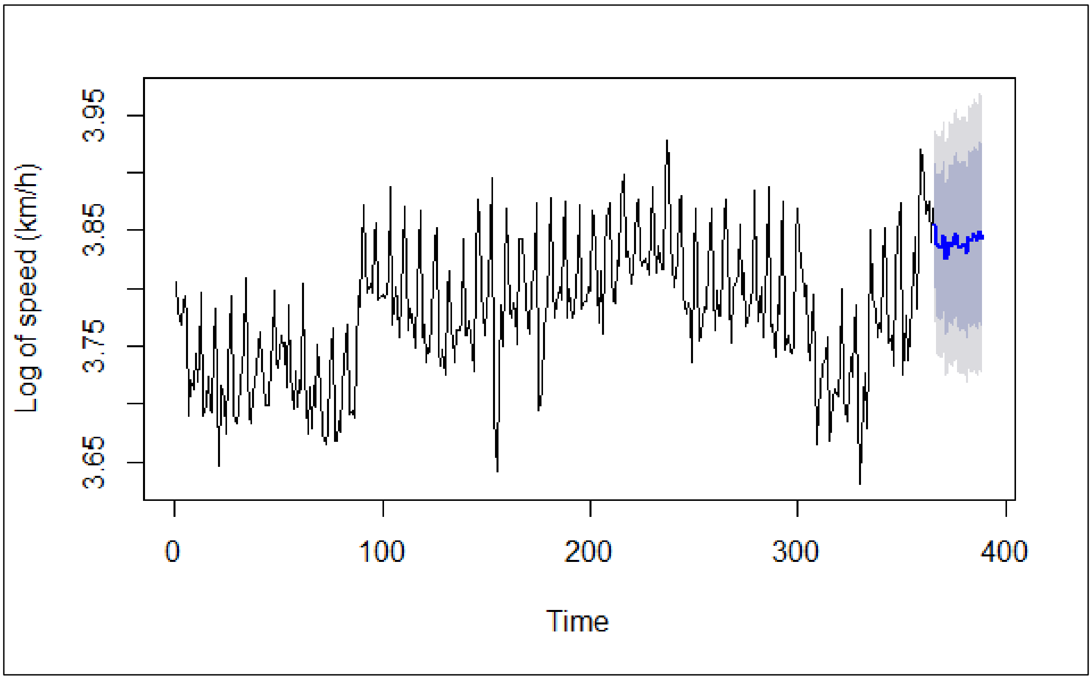


Figure 5.25: Plot showing the log of flow and the prediction overlaid with 80% and 95% confidence limits

5.4.4 Model validation of flow

This section presents the validation results using the test data based on flow estimation to conclude the assessment. Figure 5.26 to Figure 5.28 present the time plot, density plot, and the normal distribution plot of the validation results. Although the forecast seems to be under-estimating with a lower density and wider spread, the quantitative analysis showed that the difference is not significant. The correlation analysis between the forecast using the training data set and the validation using the test data set gives 0.824 with a p-value = 0.000 showing the significance of the result. In addition, the Mann-Whitney test and 95% confidence interval give a point estimate of -1.280 and CI (-2.073, -0.625), and the test statistic is significant at 0.0007. The results show that the two distributions are not statistically significantly different at $\propto = 0.05$ confidence level. The value of KL-D (0.0015) further buttressed the results. Overall, the test data produced the following accuracy statistics: RSME =0.193077; MAE= 0.145761 MAPE= 4.89 with a p-value = 0.5858 for the portmanteau test which signifies the validity of the estimation.

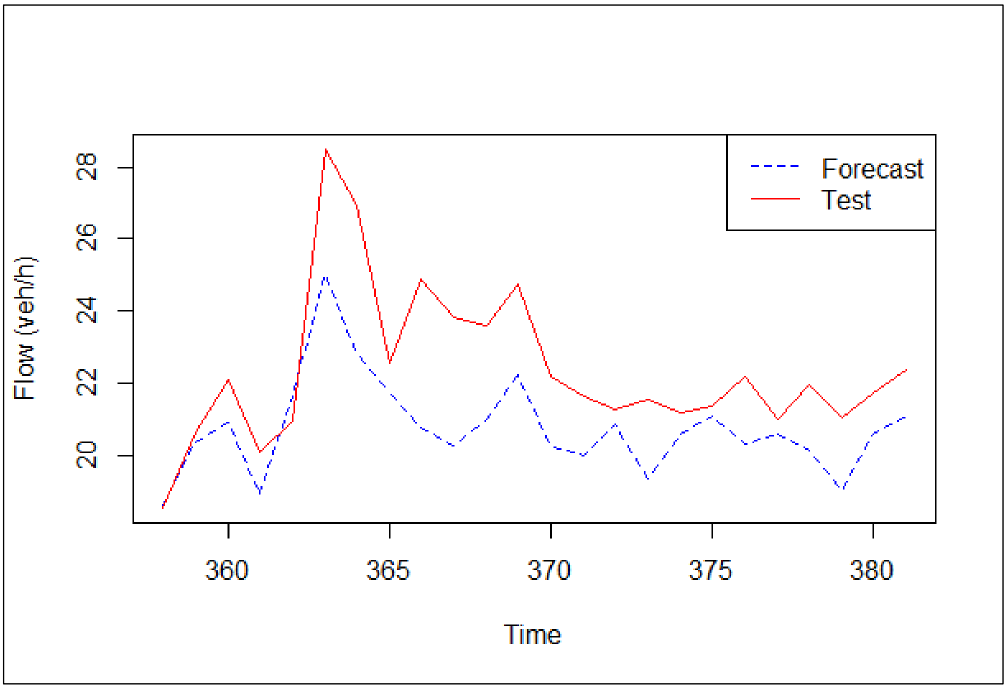


Figure 5.26: Plot of forecast and validation (test) data

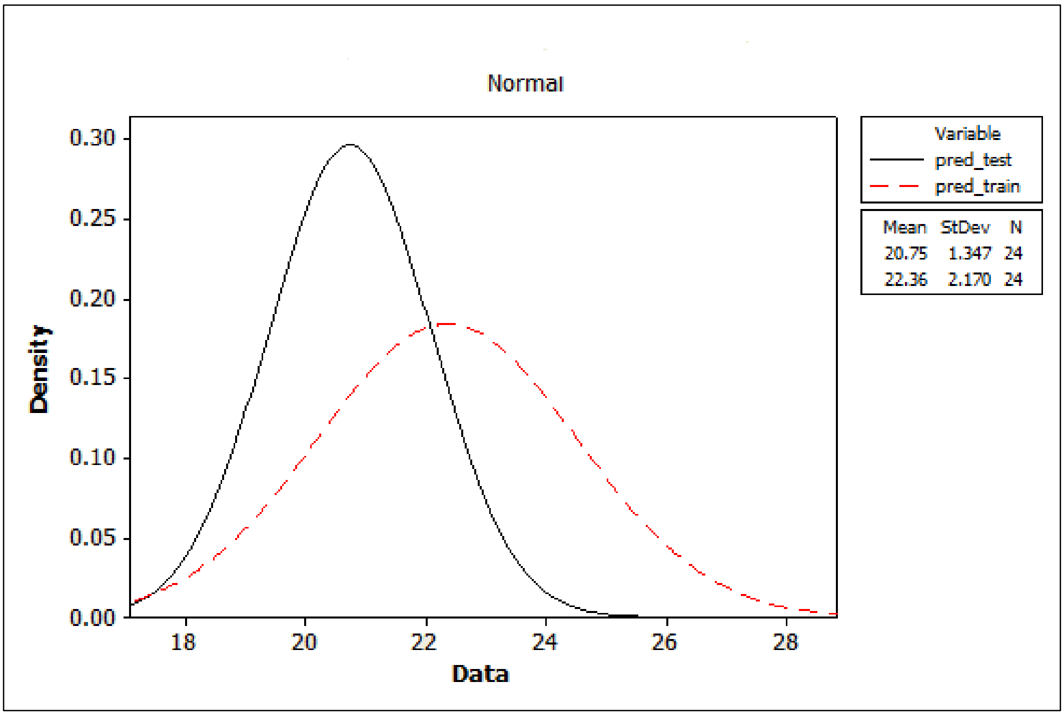


Figure 5.27: Density plot of forecast (red) and validation (black)

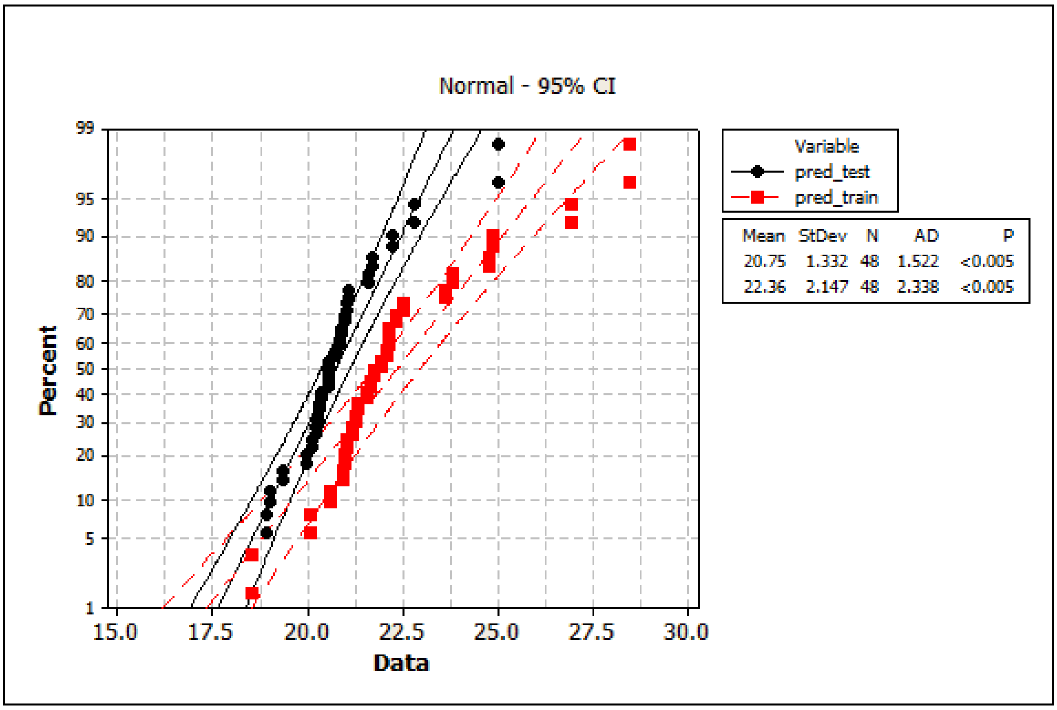


Figure 5.28: Normal probability and 95% confidence Interval plot of forecast (red) and validation (black)

5.5 Conclusions

This chapter presents the results of the validation exercise carried out in this research. The assessment started with the calibration of the estimation model (TRAFOST) developed in this research to maintain the concept of fit for purpose. This step was followed by the validation of the estimated metrics against the independent measures of traffic. The validation concludes with ARIMA modelling and forecasting to understand the predictability and validity of the estimation. The exploratory and quantitative analysis techniques employed ensured that a robust validation was performed. The outcome of the Mann-Whitney-Wilcoxon test, Kullback-Leibler divergence as well as the forecast accuracy statistics for flow, journey times and speed showed a high level of precision and accuracy given a 95% confidence level. The overall result implies the validity and practicality of the estimation – that is the possibility to derive performance measures such as journey times and vehicle speeds, to enhance traffic management using Bluetooth. Not only that, the forecast accuracy suggests a possibility of predicting the future traffic state as well as data augmentation to realise enhanced traffic planning and management. It is noted that only the range of conditions covered limits the resulting generalisation in this validation. It is to be noted that the validity of the O-D matrix validation will require further analysis to reach a logical conclusion. Therefore, validation and testing need to be conducted to investigate whether the same generalisation holds for data in other locations and for other related metrics such as the O-D matrix and density. Interestingly, the results obtained agree with the findings from the previous research. The next chapter considers the variability in the estimated metrics to enhance the knowledge of the data usage and to avoid invalid judgement and conclusion.

Chapter 6. Exploring Variability in Bluetooth-Derived Traffic Metrics

6.1 Introduction

Chapter 6 builds on the validation presented in Chapter 5 by investigating the variability in the estimated metrics to ensure a valid statistical underpinning (Research Objective number v). The understanding of this important factor in Bluetooth is considered essential to ensure reliability, given that a number of error sources can influence the estimated metrics, in particular, the variability relating to the long-term variation in order to understand practicality. Consequently, this chapter considers the following specific objectives: i) investigation of possible reason(s) for over/under-estimation (that is, the issue of over/under-sampling which may be due to outliers); ii) understanding of consistency and the modelling capability of the data; iii) examining daily/weekday temporal changes to understand the reliability of the metrics; and iv) understanding of any long-term variation. Therefore, the Bluetooth data collected over the Trafford network on Link0506 were analysed for this purpose using a combination of exploratory and quantitative analysis techniques. Accordingly, the variability in the Bluetooth derived metrics and its significance to ITS applications in road traffic management was explored.

This chapter is structured as follows: Sections 6.2, 6.3 and 6.4 consider the variability in the Bluetooth estimated metrics (flow, journey time and speed, respectively) with a focus on over/under-sampling, the issue of consistency and the modelling capability of Bluetooth and the day-to-day and long-term dynamics in the estimated metrics. The spatio-temporal assessment of the variability in Bluetooth detection rates is presented in Section 6.5; the problem considered in this section focuses on the changes in the detection rates over GMN, and whether the result holds, irrespective of the data source and location, before conclusions are drawn in Section 6.6.

6.2 Understanding Variability in Flow

6.2.1 Exploration of estimated flows

The estimated traffic flows were explored over different temporal dimensions (with a focus on hourly, weekday and monthly averages), direction of travel and over different periods of observation to understand variations. This investigation was necessary to describe the estimated flows accurately to understand possible limitations. Boxplots and other exploratory techniques were used to rapidly characterise the flows. The results of the exploratory analysis are presented in Appendix 6A. Following the exploration, Table 6.1 presents the summary of NE and SW-directional flows based on the application of Mahalanobis distance (MD) filtering. The mean and median values corresponding to 21veh/h, 19veh/h, 18veh/h and 16veh/h for NE and SE flows, respectively. On an average, the flows on the opposing links are similar. This could mean that the two opposing links' flows can be averaged to manage the network using the same strategy, thereby reducing the amount of planning and improving efficiency in performance.

	Directional Flow (veh/h)		
	NE	SW	MD
Min.	0	0	0.02
1st Qu.	6	4	0.90
Median	19	16	1.19
Mean	21	18	1.18
3rd Qu.	36	31	1.35
Max.	63	56	2.45

Table 6.1: Summary of NE and SW-directional flows based on MD filtering

Figure 6.1 presents the time series plot of flows in both directions aggregated on four temporal dimensions. The results showed that the monthly average has the highest variability. Appendix 6B presents further results on the analysis of the flow data such as the table of adjusted R^2 to understand the goodness of fit. The result obtained gives a level of reliability to the data, and the possibility for data reduction to improve computational and operational efficiency. That is, the data could be averaged to reduce the number of the variables to be modelled. In addition, the monthly analysis of the flow data is consistent with Johnson (1989) and DfT (2014), which stated that the neutral months of April/May and September/October are supposed to have minimum variability of flows. The combined flows over these months averaged 42veh/h. Other analysis performed also showed that Bluetooth flows aggregated at high resolutions, (such as a 5-minute average), present many dispersions between weekdays. Higher aggregate levels on the other hand showed better precision (less dispersion), which signifies a better level of estimation for traffic prediction. Generally, there exists a high level of temporal consistency with the maximum variability being about 3veh/h for all. This temporal consistency was analysed further through their mean and standard deviation plots. However, given the day-to-day similarities in the flows from the opposing links, the subsequent discussion is focused on the NE-directional flows while the SW-equivalent flows are presented in the Appendix 6C.

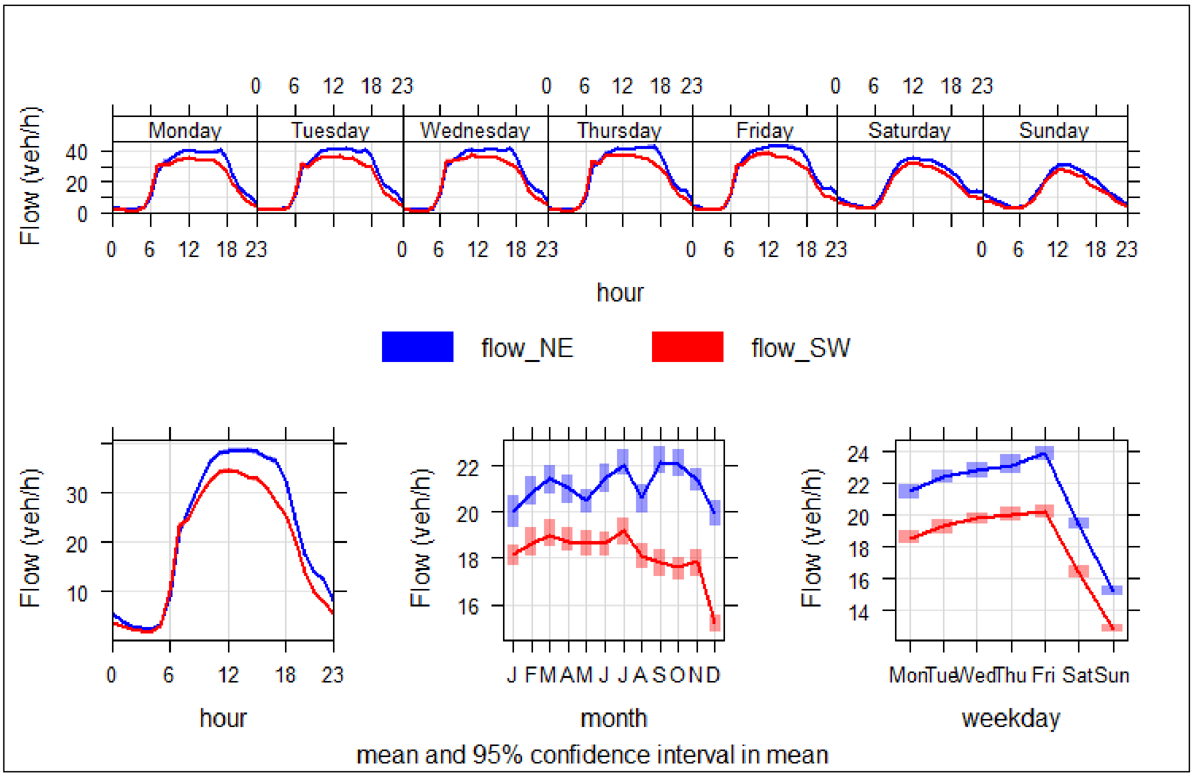


Figure 6.1: Time series plots of directional flows on Link0506 (N=31937)

Figure 6.2 presents the time series plot of the NE-directional flows averaged on a daily basis. The flow average over the year is between 13-25veh/h. From day-to-day, there is an evidence of seasonality caused by the daily and weekly effect. A few sparks are also noticed on the 100th and 208th day. The gap between the 52nd and 54th day may be due to equipment failure or corrupt data as it is not expected that no vehicles were recorded over these periods. There is also a significant drop in the flow at the end of the year, which relates to the festivity during this period. The trend in the data will be explored further in later discussions to understand long-term variation. The next step considers the consistency of the data. Consistency in this context as earlier defined is when the Bluetooth estimation corresponds to the actual traffic pattern given any temporal dimension, such as hourly or daily average, and is measured in terms of the precision of the mean and standard deviation of the data on a given average. In this case, standard deviation shall be used to measure consistency.

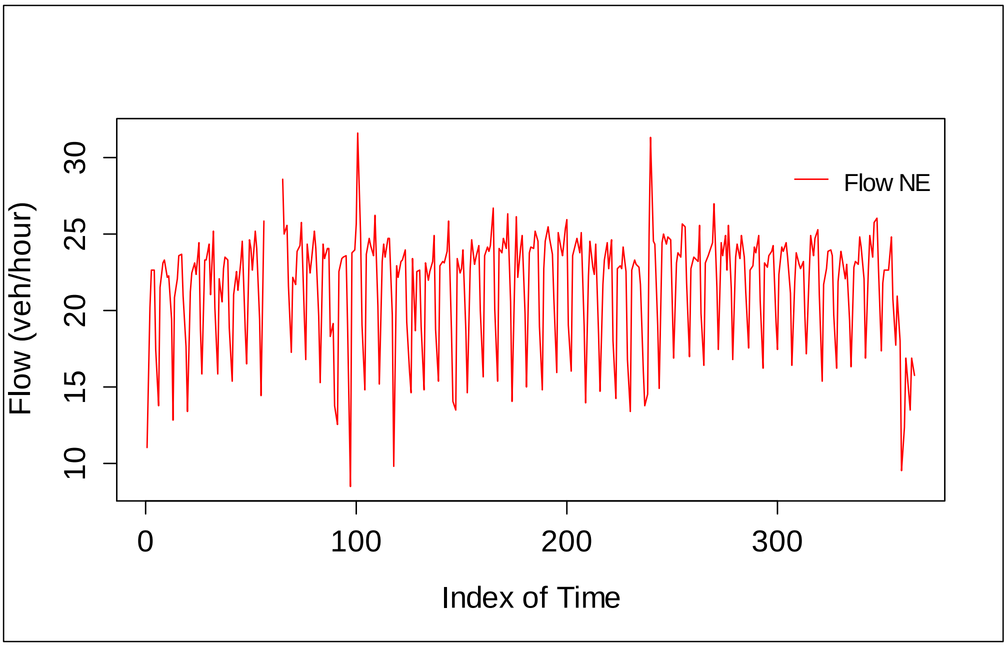


Figure 6.2: Time series plot of NE-directional daily average flow

6.2.2 Understanding consistency and reliability in flow

This section explores the use of standard deviation to understand the precision of the estimated flows to establish reliability. This investigation is expanded further in the next section by exploring the degree of variability in flow to provide answers to the specific objectives ii and iii in this chapter. Figure 6.3 presents the standard deviations of flow in the NE direction to understand dispersion and consistency in the data. The result shows that standard deviations of flows are clustered mainly between 12veh/h and 18veh/h with a few fluctuations at some points, such as on days 99 and 100. Generally, the standard deviation of the NE flow is consistent and is considered to not change with time. Although the result portrays the daily-weekly seasonal effect, the reproducibility of these measurements confirms the reliability of the Bluetooth estimated flow data on this temporal dimension as a useful traffic metric.

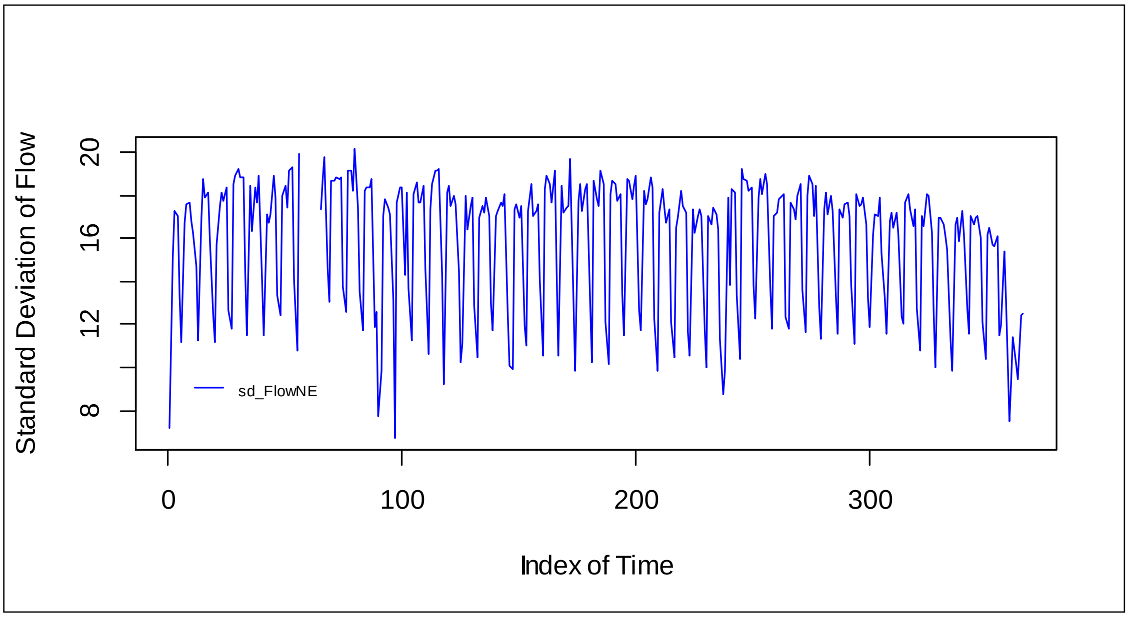


Figure 6.3: Standard deviation of flows in both directions after filtering

6.2.3 Understanding the degree of variability in flow

As a further step, the data was analysed to understand the reliability and modelling capability using Principal Components Analysis (PCA) and seasonal decomposition. PCA was used for the weekdays' flows given that it is a useful tool in understanding complexity in large urban networks (Tsekeris and Stathopoulos, 2006). On the other hand, the seasonal decomposition helps in understanding both the seasonal effect and the trend in the data to aid modelling. Using PCA, the starting point is to explore the data to understand the correlation between the variables. Box 1 shows that there is a correlation between the weekday's flows signifying the presence of redundancy in the observations and a strong indication to use PCA. For example, between the weekdays (Monday to Friday), the correlation is very high (>0.95). The p-value of 0.000 suggests that the results are highly significant. Therefore, the assumption here is that two distinct groups are possible, consisting of weekdays and weekends as would be expected. This assumption is investigated further in the analysis of the eigenvalues (the variances in the traffic flows).

The analysis of the eigenvalues presented in Box 2 shows that the first principal component has a variance of 6.2549 and accounts for 89.4% of the total variance. The second principal component has a variance of 0.5272 and accounts for 7.5%. The first two components together account for more than 96% of the total variance and are deemed sufficient to explain the variability in the data. This was confirmed in the scree plot presented in Figure 6.4, which shows a sharp drop from the first principal component to the second principal component while the rest of the principal components are very close to zero, and are considered not significant. From the first two components, equation (6.1) and equation (6.2) were formed. From equation (6.1), the coefficients of all the variables are positive but with higher values over the weekdays than over the weekend. Furthermore, given two decimal places, the coefficients of the weekdays are equal (0.39), showing a high degree of agreement indicative of redundancy in the observations. On the other hand, the weekend coefficients are also similar (0.37 and 0.33) for Saturday and Sunday, respectively. From equation (6.2) the transition or change in the algebraic sign of the coefficients from negative to positive, from weekdays to weekends, further implies the possibility of a reduction of the data into two smaller components to represent the whole in the future analysis. The differences noted in the coefficients of the variables typify the daily changes in flows between the weekdays. The assumption was further assessed using a loading plot, which confirms the validity of this assumption, for visual examination and interpretation. The implication of this result is that the use of PCA to analyse traffic flow can help in capturing temporal dynamics in a complex urban network, such as the GMN.

	Mon	Tue	Wed	Thu	Fri	Sat
Tue	0.966					
	0.000					
Wed	0.972	0.966				
	0.000	0.000				
Thu	0.969	0.969	0.979			
	0.000	0.000	0.000			
Fri	0.958	0.957	0.970	0.968		
	0.000	0.000	0.000	0.000		
Sat	0.811	0.839	0.824	0.849	0.855	
	0.000	0.000	0.000	0.000	0.000	
Sun	0.706	0.725	0.703	0.733	0.718	0.885
	0.000	0.000	0.000	0.000	0.000	0.000

Cell Contents: Pearson correlation
P-Value

Box 1: Box showing the correlation matrix and p-values of weekday flows

Eigenanalysis of the Correlation Matrix

Eigenvalue	6.2549	0.5272	0.0984	0.0413	0.0303	0.0291	0.0188
Proportion	0.894	0.075	0.014	0.006	0.004	0.004	0.003
Cumulative	0.894	0.969	0.983	0.989	0.993	0.997	1.000

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Mon	0.387	-0.249	0.273	0.119	-0.727	-0.408	-0.069
Tue	0.390	-0.191	0.097	0.735	0.495	-0.124	0.047
Wed	0.389	-0.250	0.070	-0.253	-0.014	0.394	0.749
Thu	0.392	-0.182	0.038	-0.097	-0.025	0.632	-0.634
Fri	0.390	-0.184	-0.228	-0.566	0.395	-0.511	-0.158
Sat	0.365	0.452	-0.751	0.186	-0.240	0.038	0.070
Sun	0.327	0.753	0.542	-0.132	0.115	-0.030	0.004

Box 2: Box showing the eigenvalues of the correlation matrix of weekday flows

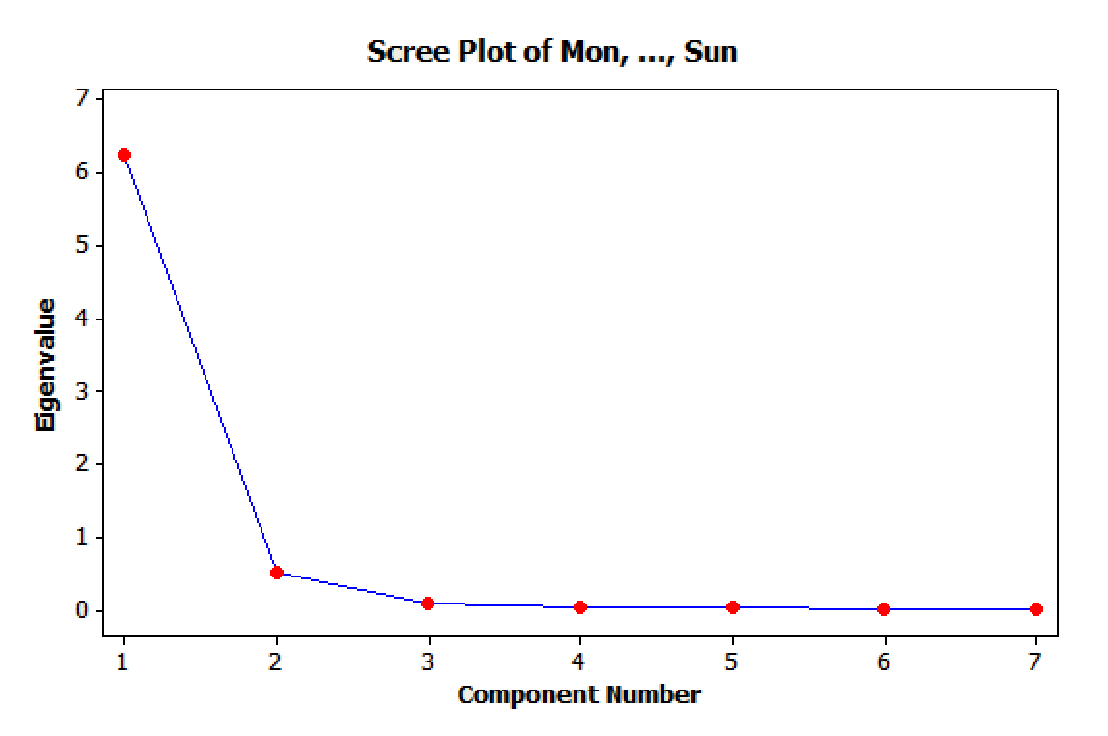


Figure 6.4: Scree plot to judge the relative magnitude of eigenvalues

$$Z1 = 0.387\mathrm{Mon} + 0.390\mathrm{Tue} + 0.389\mathrm{Wed} + 0.392\mathrm{Thu} + 0.390\mathrm{Fri} + 0.365\ \mathrm{Sat} + 0.327\mathrm{Sun} \tag{6.1}$$

$$Z2 = -0.249\mathrm{Mon} - 0.191\mathrm{Tue} - 0.25\mathrm{Wed} - 0.182\mathrm{Thu} - 0.184\mathrm{Fri} + 0.452\ \mathrm{Sat} + 0.753\mathrm{Sun} \tag{6.2}$$

Figure 6.5 presents the plot of loadings for the second component (y-axis) versus the loadings for the first component (x-axis) with a line drawn from each loading to the (0, 0) point based on Minitab (2014). The analysis of the loading plot showed that the groups (weekdays and weekend flows) started off at the same point and diverged with an increase in the first component particularly with Sunday flows showing higher loading. The clustering of the weekdays' loadings signifies closeness in observations (presence of redundancy), and therefore, higher precision compared to weekend flows. Irrespective of the separation observed in the weekend flows, they are considered as another cluster as revealed in the earlier analysis above; but in this case with a better understanding of the separation in the weekend flows.

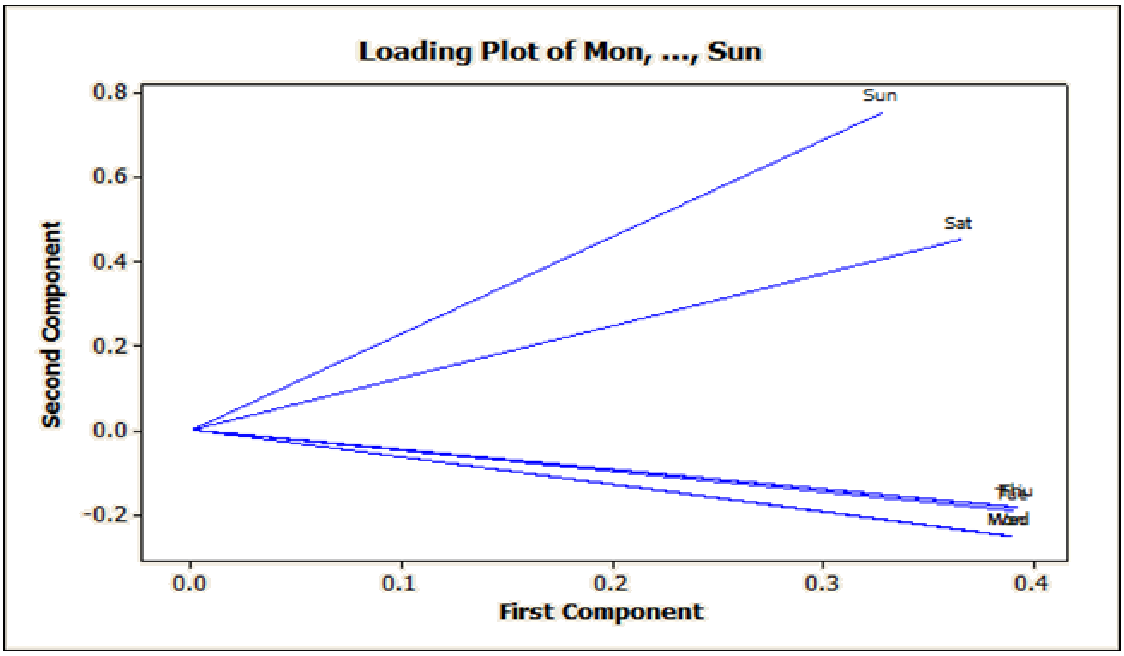


Figure 6.5: Loading plot of weekday flows showing two different groups in flow

Figure 6.6 presents the seasonal decomposition of the flow data showing four components. The first component (top) is data, which comprises all the other three components while the second component presents the seasonality. The third component is the trend in the data while the fourth (bottom) is the remainder after the removal of the seasonal and trend components from the data. The results show that the seasonal component does not change with time while trend presents the entire movement in the series with a flexible pattern. The start and end of the year have a low flow that corresponds to negative values in the remainder component. The bars at the end of the plots represent the relative scales and the amount of variation of the components (Hyndman and Athanasopoulos, 2013). For example, the long bar in the seasonal component means smaller variation compared to the data and remainder components with short bars. The modelling capability of the flow data is further confirmed in Figure 6.7 that shows the autocorrelation and cross-autocorrelation between the two-directional flows. The ACF plots show that there remains some serial correlation in the data; nevertheless, there is a strong indication that the data can be modelled as shown in Chapter 5.

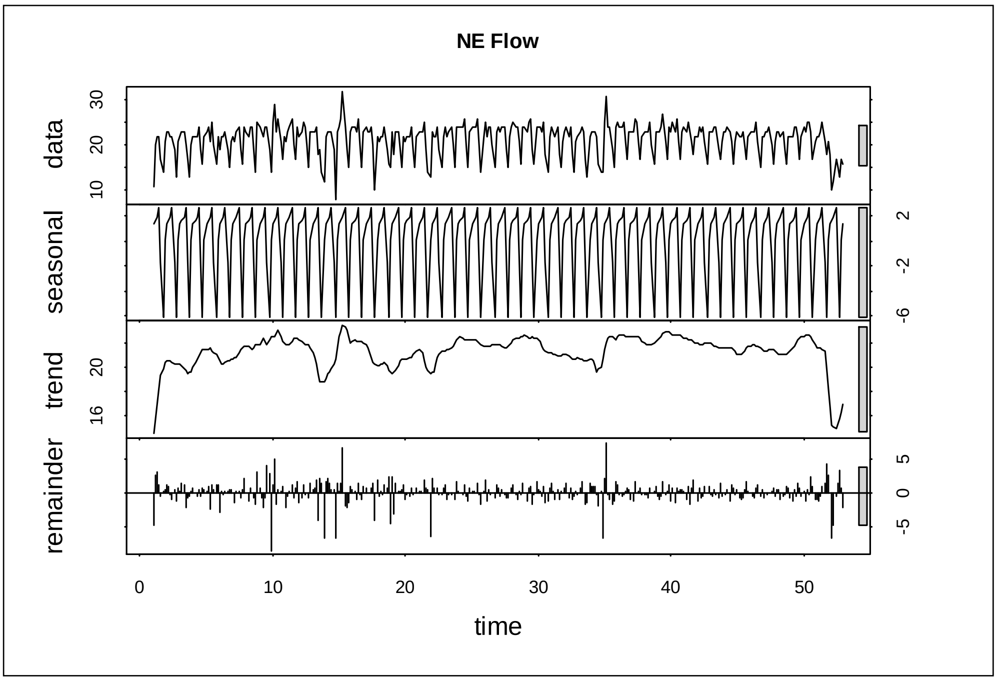


Figure 6.6: Time series decomposition of NE-directional flow

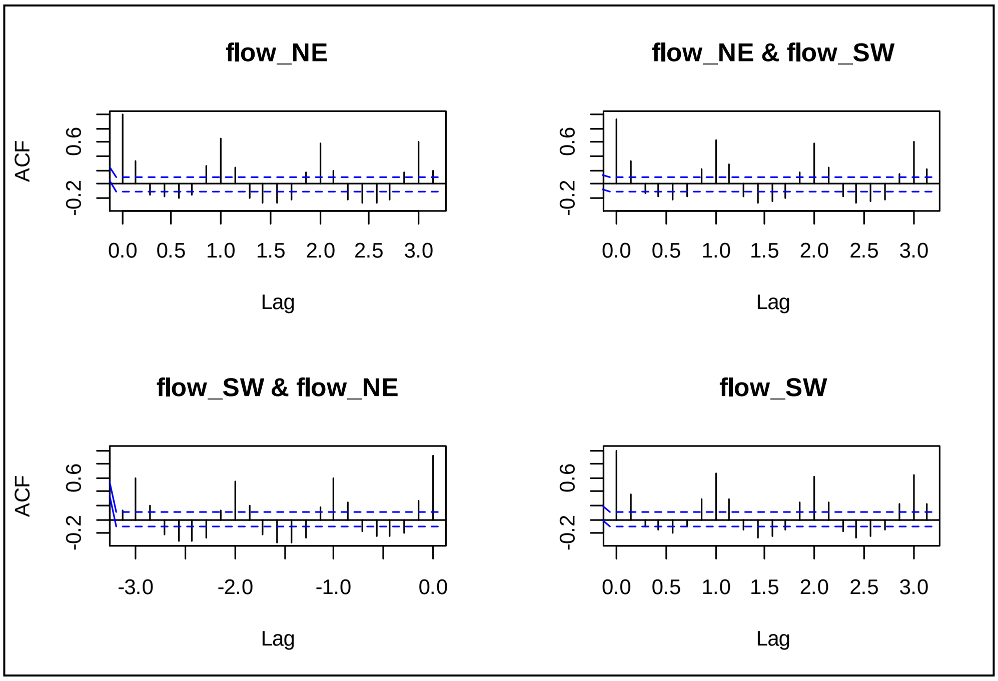


Figure 6.7: Autocorrelation and cross-autocorrelation of directional flows

6.2.4 Post-analysis of flows to understand temporal changes

This section concludes the investigation and the focus is to understand temporal variations within homogenous groups. Zhang *et al.* (2013) highlighted the necessity to understand the evolution of traffic states in both time and space as a critical step to improving freeway modelling and operations. The assessment is based on five groups of periodic flows over 24 hours in a month. Table 6.2 and Table 6.3 present the output of the post-analysis based on $(\alpha = 0.05)$ using Statistical Package for the Social Sciences (SPSS), where alpha defines the cut-off point upon which a rejection or acceptance of the hypothesis test is determined. The tables show the test results of multiple comparison and homogenous subsets of the grouped flows. Analysis of Table 6.2 suggests that the groups designated 07-10hrs and 16-20hrs are the most variable groups compared to other groups. This information points to the period of less precision in the estimated flows as evident in the computed standard error (column 3 of Table 6.2). The significance of this result can be found in weight assignment in modelling to ensure accurate and precise prediction. That is, the knowledge of the period of high or low level of reliability in flow can be determined based on this information. Table 6.3 presents a clearer picture of the significant differences among the grouped flows. From this result, four homogeneous subsets were identified among the five groups. The p-value (0.257) computed for the subset 3 clearly shows that the means of the two most variable groups are statistically not significantly different from each other. If the same condition is applicable in all situations, a typical traffic plan or strategy can be implemented to manage the two periods. This thereby reduces the amount of planning and operational activities and consequently increases efficiency in production and optimisation of input.

Multiple Comparisons

Measure: MEASURE_1

	(I) period	(J) period	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	01-07hrs	07-10hrs	-22.36*	1.053	.000	-25.29	-19.43
		10-16hrs	-28.66*	.849	.000	-31.02	-26.29
		16-20hrs	-20.26*	.956	.000	-22.92	-17.60
		20-23hrs	-4.88*	.956	.000	-7.54	-2.22
	07-10hrs	01-07hrs	22.36*	1.053	.000	19.43	25.29
		10-16hrs	-6.29*	1.079	.000	-9.29	-3.29
		16-20hrs	2.10	1.165	.378	-1.14	5.34
		20-23hrs	17.49*	1.165	.000	14.24	20.73
	10-16hrs	01-07hrs	28.66*	.849	.000	26.29	31.02
		07-10hrs	6.29*	1.079	.000	3.29	9.29
		16-20hrs	8.39*	.985	.000	5.65	11.13
		20-23hrs	23.78*	.985	.000	21.04	26.52
	16-20hrs	01-07hrs	20.26*	.956	.000	17.60	22.92
		07-10hrs	-2.10	1.165	.378	-5.34	1.14
		10-16hrs	-8.39*	.985	.000	-11.13	-5.65
		20-23hrs	15.38*	1.079	.000	12.38	18.39
	20-23hrs	01-07hrs	4.88*	.956	.000	2.22	7.54
		07-10hrs	-17.49*	1.165	.000	-20.73	-14.24
		10-16hrs	-23.78*	.985	.000	-26.52	-21.04
		16-20hrs	-15.38*	1.079	.000	-18.39	-12.38

Table 6.2: Table of multiple comparison tests between the grouped flows

MEASURE_1

	period	N	Subset 1	2	3	4
Tukey HSD[a,b]	01-07hrs	28	4.33			
	20-23hrs	16		9.21		
	16-20hrs	16			24.59	
	07-10hrs	12			26.69	
	10-16hrs	24				32.98
	Sig.		1.000	1.000	.257	1.000

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 9.313.

a. Uses Harmonic Mean Sample Size = 17.500.

b. Alpha = .05.

Table 6.3: Table showing the homogeneous subset of the grouped flows

6.3 Understanding Variability in Journey Time

6.3.1 Understanding temporal variability in journey times

Figure 6.8 presents the combined plots of mean and median journey times over four temporal dimensions – hourly-weekday (top), hour (bottom-right), month (bottom-middle) and weekday (bottom-right). Over the four averages, the mean journey times are consistently higher, with the highest variability in the monthly average (88s-140s). Variability is also higher over the peak periods than in free flows, so also on weekdays as compared to weekends. However, the two estimators present similar trends over the year with the exception of the trends over the weekday average. The median travel time presents relatively the same time over Monday to Wednesday but in the case of the mean, the travel time on Tuesday is a little higher. The obvious difference is the reversal in trend from Wednesday to Friday. While the mean travel time increases, the median equivalent decreases. However, they both show a decrease in travel time from Friday to Saturday as expected. Overall, a conclusion is reached that the day-to-day variability captured by the mean estimator better represents the real-life situation. However, an assessment of any significant differences between the mean and median estimators presented in Table 6.4 based on the Mann-Whitney test shows that the two journey time metrics are good estimators for traffic management. Other relevant results are presented in Appendix 6D.

To establish the day-to-day reliability in travel time estimation, the daily average is analysed. Yildirimoglu *et al.* (2015) emphasised the significance of the day-to-day travel time variability and reliability. Figure 6.9 presents the time series plot of journey times over a year on a daily average to understand the day-to-day variability. Unlike the flow data, the journey time over the year fluctuates with irregular patterns between 80s and 140s. Over the year, three clusters of journey times can be identified with the highest variability over the first three months and in November. The most consistent period (day 150-300) corresponds to June-October, signifying the most reliable period of the year as captured by the established systems. Given this information, an unreliable travel time that has been identified as one of the problems of congestion can be addressed accordingly using Bluetooth. This includes the provision of real-time traffic information obtainable from Bluetooth instead of reliance on archived data for better prediction of journey times and by extension leading to improved service delivery and more confidence in route planning. Accordingly, real-time traffic information obtained from Bluetooth can be disseminated based on the changes in the network traffic.

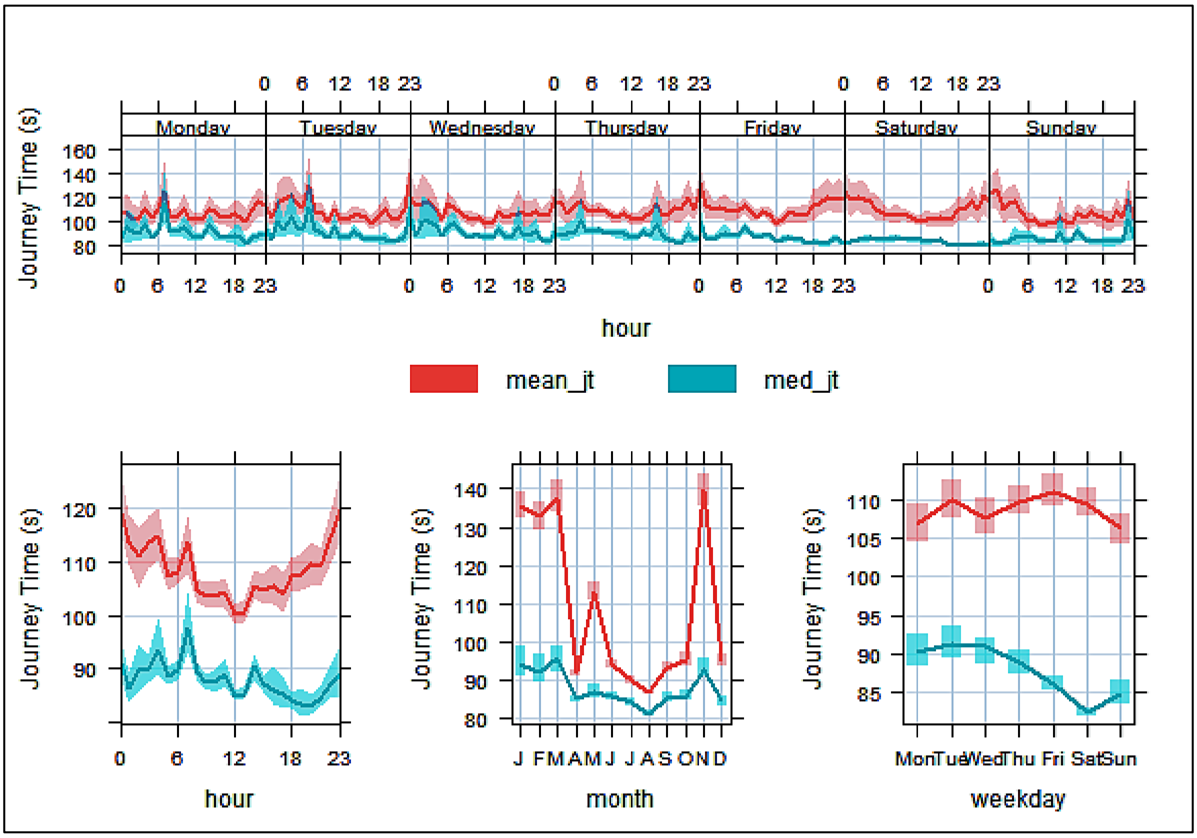


Figure 6.8: Mean and median (med_jt) journey times on Link0506

Mann-Whitney Test				
Parameter	Point Estimate	95% Confidence Interval	P-Value	N
Hourly	-13.00	(-14.000,-13.000)	0.0000	168
Hourly-Weekday	-17.00	(-18.000,-16.000)	0.0000	8205
Daily	-11.25	(-12.991,-9.880)	0.0000	365
Week	-10.42	(-20.943,-7.920)	0.0000	53
Month	-9.25	(-33.500,-5.600)	0.0001	12
Weekday	-17.00	(-19.000,-14.000)	0.0022	7

Table 6.4: Table showing the summary of Mann-Whitney Test over different temporal dimensions

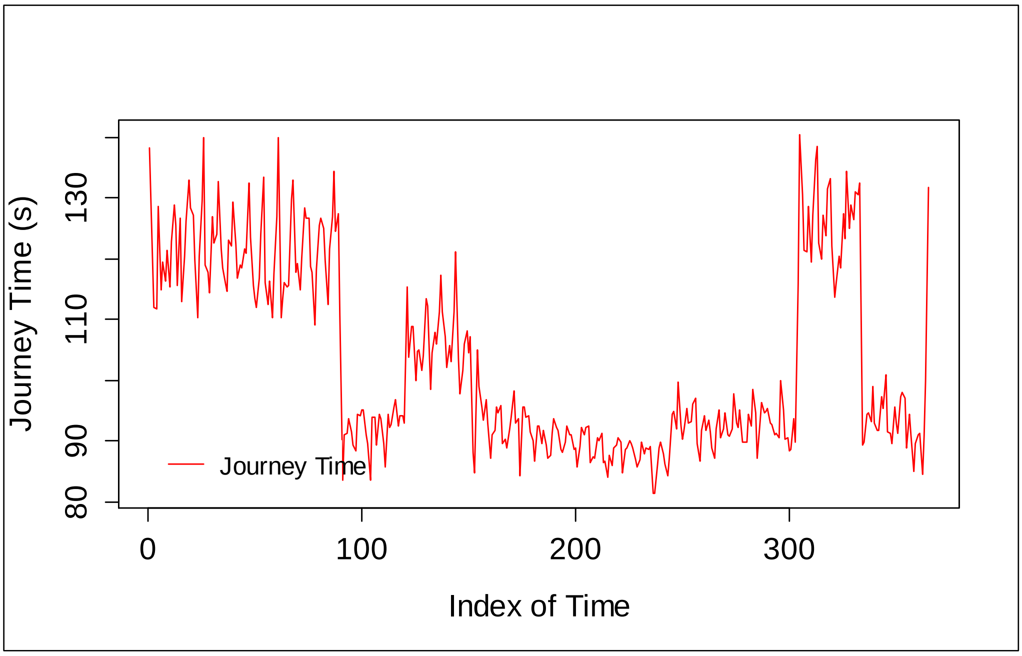


Figure 6.9: Time series plot of daily journey time on Link0506 in Trafford

6.3.2 Understanding consistency and reliability in journey time

Figure 6.10 presents the standard deviation plot of journey time over the year showing non-uniformity in the daily pattern, which is usually what happens in real life. The analysis shows that standard deviations of journey times range between 5s and 33s symbolising the degree of variability over the year. Approximately, drivers may experience up to a 30s time difference compared to normal in traversing this link. Variability is highest over the first three months and least between the months of June and October. A decomposition of the journey times (Figure 6.11 – plotted on a weekly time scale) shows a better understanding of the trend in the data over the year. Clearly, from the plot, the weeks over the period 23-44 present the minimum variability. The results also show that there is a constant seasonal effect, with the trend showing three to four distinct patterns. The bars on the plots show that variability is least in the seasonal component compared to others, while the positive and negative values in the remainder component signify the points of rise and fall in trend, as well as, the description of the amount of variability in the trend. For example, the rise in journey time over the end of the year corresponds to a positive rise in the remainder that may be due to the end of year rush or other events. This result is consistent with Martchouk *et al.* (2011) who noted that factors, such as weather conditions and driver behaviour, may significantly influence variability in travel times over a particular period.

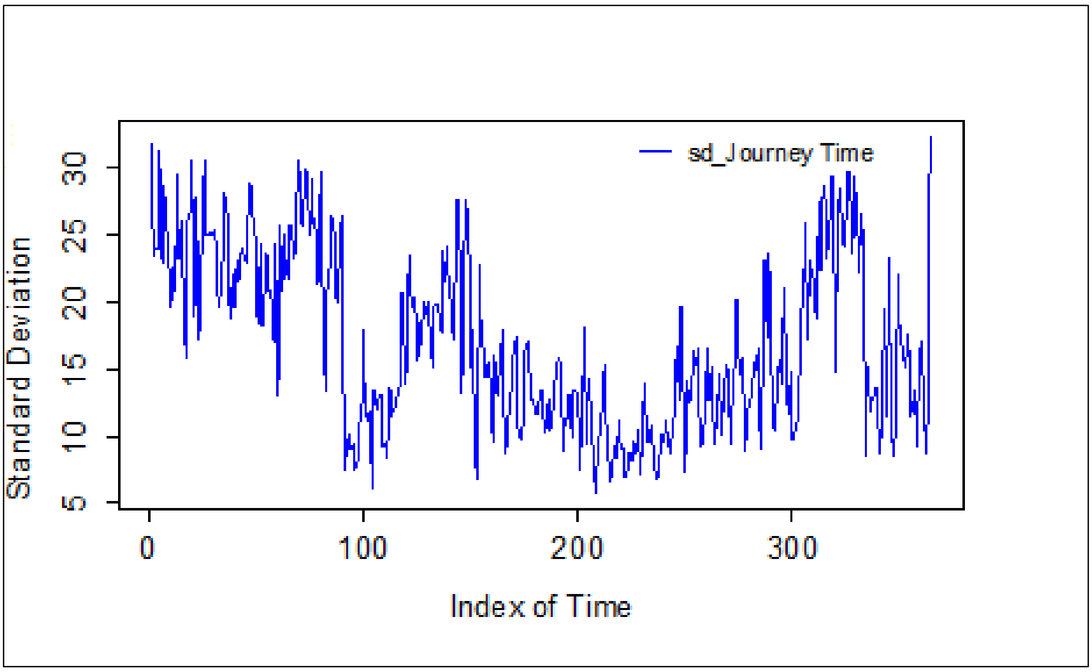


Figure 6.10: Standard deviation of daily journey time on Link0506 in Trafford

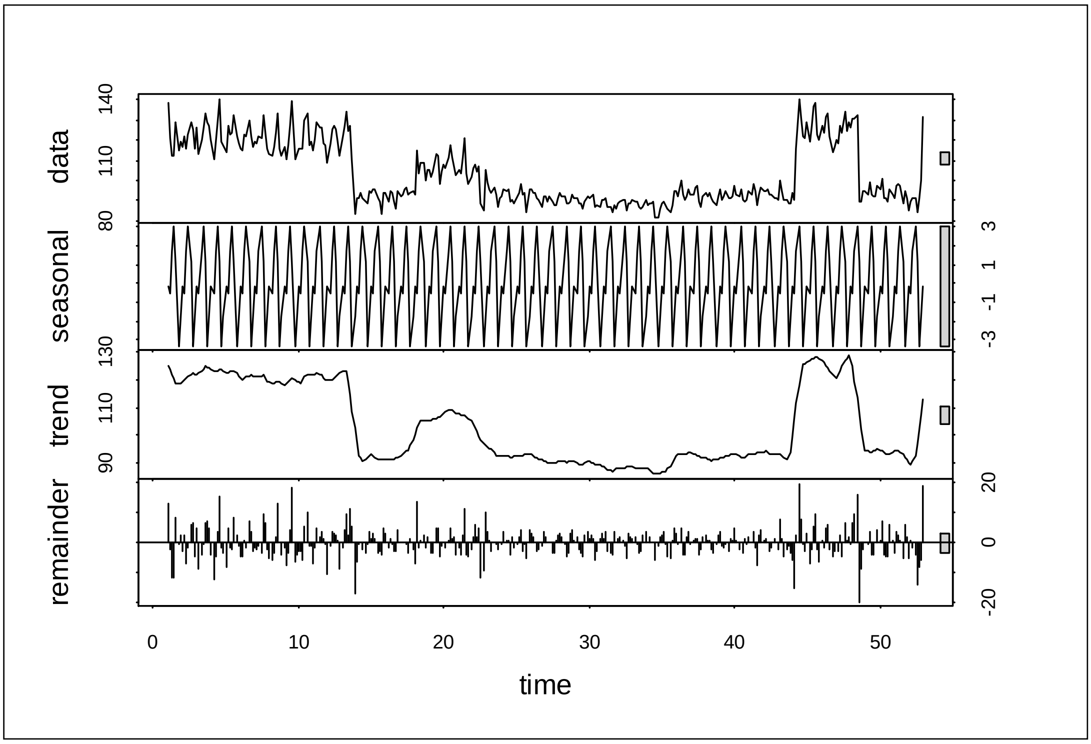


Figure 6.11: Seasonal decomposition of daily journey time over a year on Link0506 in Trafford

6.4 Understanding Variability in Speed

6.4.1 Understanding temporal variability in journey speed

Figure 6.12 presents the combined plots of mean and median journey speeds averaged over four different temporal dimensions to illustrate both short and long-term variations. The hourly variation presents almost a perfect agreement between the two estimators and has a minimum variability compared to the other averages. Consistent with the journey time estimators, the monthly average presents the highest variability over the year with a range of 5km/h. This indicates the degree of variability that could be experienced in speed over the year on the link. Further discussion of the implication of the results will be presented in the next chapter. However, to conclude the analysis, the Wilcoxon test was performed to test for significant differences between the two estimators. The test results showed that the two estimators are not significantly different at a 95% confidence limit with a p-value less than 0.01. Figure 6.13 presents the time series plot of vehicle speeds over a year on a daily average to understand day-to-day variability, while Appendix 6E presents other relevant results. Consistent with the journey times, there is a fluctuation in speed over the year between 42km/h and 53km/h. On an average, two dominant clusters of speed can be identified with the period of high-speed corresponding to a short journey time. The next section examines the long-term variability on a daily basis by exploring the standard deviation and decomposition of speed over the year.

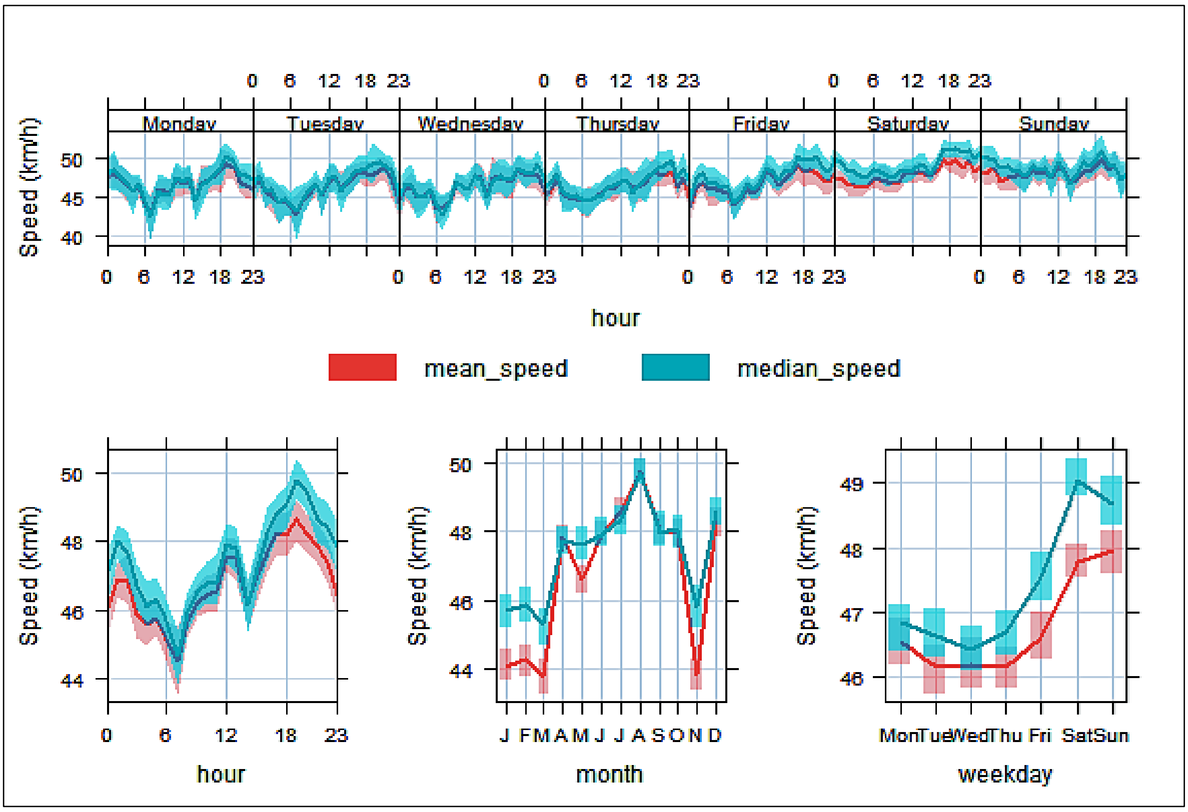


Figure 6.12: Mean and median vehicle speeds on four temporal dimensions

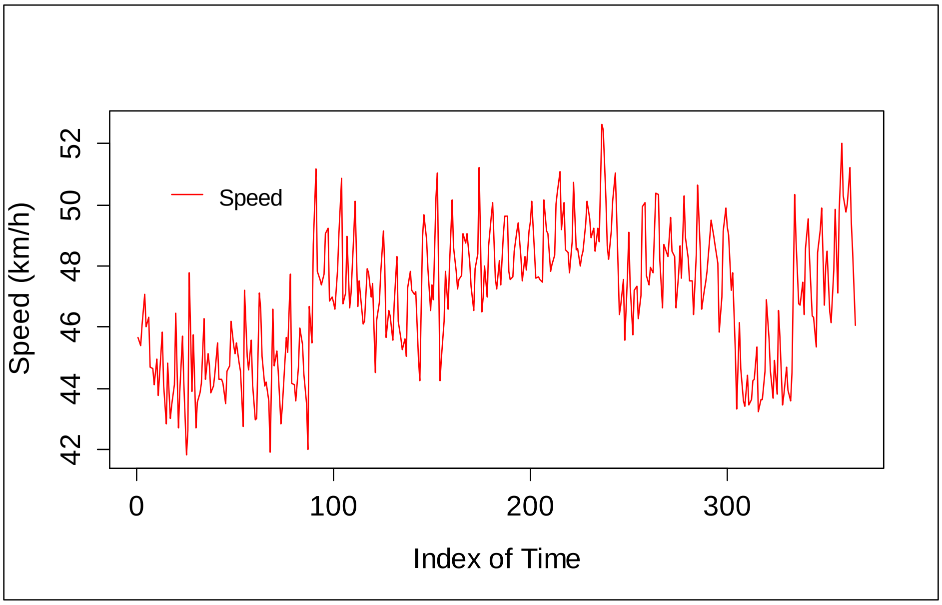


Figure 6.13: Profile of daily vehicle speeds on Link0506 in Trafford

6.4.2 Understanding consistency and reliability in journey speed

Figure 6.14 presents the standard deviation of speed over the year. Consistent with the journey times' data, there is non-uniformity in the daily pattern. The analysis also shows that the degree of variability of the daily vehicle speeds over the year is depicted by the range of the standard deviations (between 2km/h-8km/h). In terms of reliability of vehicle speed over the year, the period that is more consistent is more reliable. A decomposition of the vehicle speed (Figure 6.15 – plotted on a weekly time scale) presents a clearer understanding of the changes over time. From the results, while the seasonal component is not changing with time, the trend component is (as in journey time). The remainder plot presents the magnitude of the variability in the trend. The bars on the plots show that the variability in the trend and remainder components is about half the variability in the data and about three times the variability in the seasonal component. The fall in speed at the end of the year also corresponds to the equivalent rise in journey time over the year as previously observed. Clearly, there is evidence of changes over time as would be expected due to different traffic regimes, such as free flow and congestion. Ability of Bluetooth speed to reconstruct the actual traffic situation captured by the established systems gives credence to its applicability in this regard.

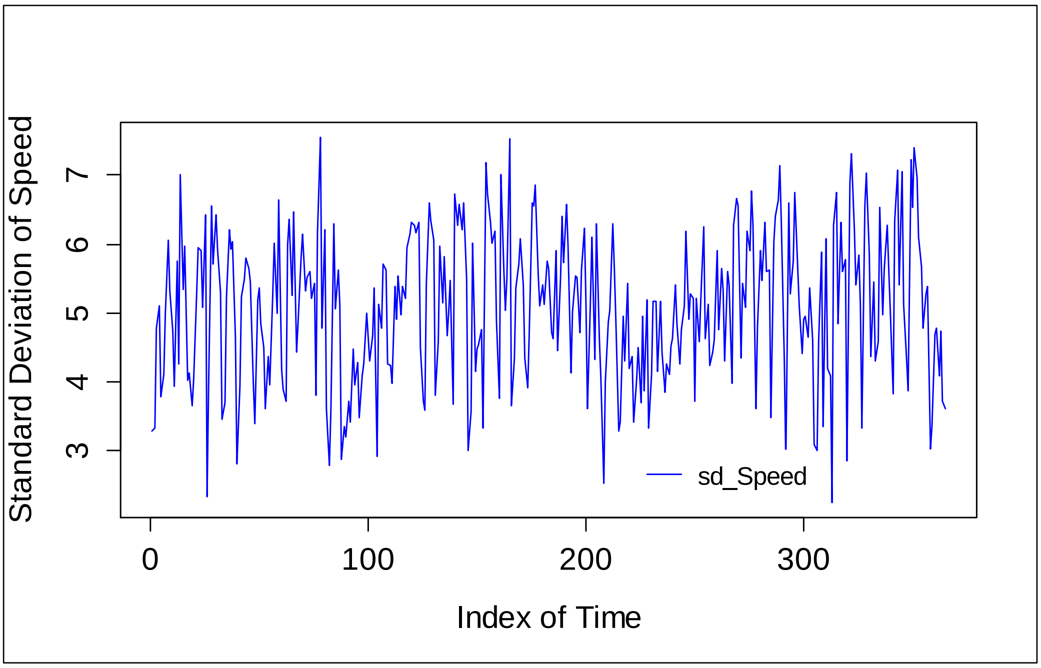


Figure 6.14: Standard deviation of vehicle speeds

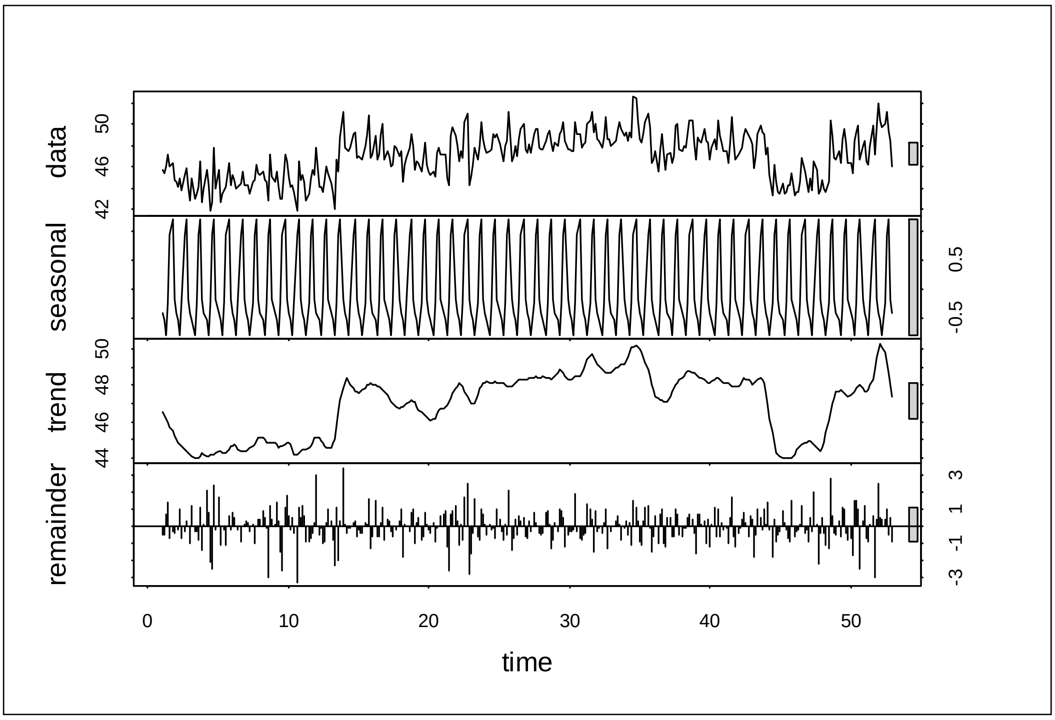


Figure 6.15: Time series decomposition of vehicle speeds

6.4.3 *Post analysis of journey speed to understand temporal changes*

In this section, the focus is on the Bluetooth-estimated speed, given that it is a derivative of journey time and not a direct measurement. This investigation helps in determining whether Bluetooth derivatives, such as the journey speed, are as reliable as the direct measurements, such as journey times. To investigate this, variability in journey speed was explored using analysis of variance $(1 - way\ ANOVA, \alpha = 0.05)$. ANOVA was considered because speed data is normally distributed. The alpha level determines the rejection or acceptance cut-off point of the test statistic (IBM Corporation, 2012). In this case, directional speed data based on weekdays and months were analysed. The null hypothesis (H_0) testing assumes equality of means in the speed distributions across the groups. In this case, acceptance of H_0 means that there is no evidence of significant change across the groups and hence no periodic or temporal variations. Otherwise, the rejection of H_0 (the acceptance of the alternative hypothesis) means that there are temporal variations. Post analysis based on the Tukey test ($\alpha = 0.05$) using R (R Core Team, 2013) was used to identify differences in the group means as shown in Table 6.5 and Table 6.6 – Link0506; and Table 6.7 and Table 6.8 – Link0605. Tukey tests helped to classify the periods that are statistically significantly different from one another. From the tables, the first column (Groups) shows different classes in the treatments (weekday and months) indicated by the letters assigned to the groups. The test results indicate temporal variations across weekdays and over months in the speed distributions. If this is consistent over time on the link, it means that different strategies will be required to manage and control traffic over the different groups.

From Table 6.5 to Table 6.8, means with the same letter are not significantly different from each other. For example, with an Honestly Significant Difference (HSD) of 0.389, (Table 6.5) suggests that the means of speed on Thursday and Monday (groups “a” and “b”, respectively) are significantly different from the means of the other weekdays at alpha level 0.05. Interestingly on the reverse link (Table 6.7), the same feature was identified, but in this case with a lower speed averaging 45km/h (HSD: 0.318) compared to the opposing link averaging 47km/h. Another interesting feature revealed by this analysis is the fact that all the weekdays' means are within the speed limit of the road (48km/h) and are reasonably close to one another. These results suggest a level of speed limit compliance and the possibility of using Bluetooth for this application.

Groups	**Treatments**	**Means**
a	Thurs	47.29
ab	Sat	47.21
ab	Fri	47.05
ab	Tues	46.95
ab	Wed	46.95
ab	Sun	46.94
b	Mon	46.83

Table 6.5: HSD test for weekday means of speed (km/h) over Link0506

Table 6.6 and Table 6.8 show the HSD tests of the monthly speed variations on both sides of the road. Unlike the weekday summary that presented similar output, in this case, it was not so. However, this is expected to be a possibility given the effects of seasonal variations. As earlier stated in Section 6.2.1, the neutral months of April/May and September/October are supposed to have the minimum variability of flows (DfT, 2014). This, in turn, is expected to influence the computed speed over these periods. From Table 6.6 with HSD (0.563), clearly, the mean speed of the months of April and May (groups “c” and “d”, respectively) are significantly different from the means of the other months. September and October are in the same cohort of “bc” together with the months of June and July. August is in a separate cohort “a” while December is in “b” and the other four months (January, February, March and November) are in the cohort “e”.

Groups	Treatments	Means
a	Aug	48.59
b	Dec	47.83
bc	Oct	47.69
bc	Jul	47.60
bc	Sep	47.48
bc	Jun	47.41
c	Apr	47.24
d	May	46.65
e	Mar	45.65
e	Jan	45.63
e	Feb	45.56
e	Nov	45.51

Table 6.6: HSD test for monthly means of speed (km/h) over Link0506

In Table 6.8 (HSD: 0.466), seven significantly different groups are identified compared to the six groups identified on the opposing link. In this case, May and August are in the same group “a”. While April and June are in different groups of “ab” and “bc”, serving as the connection or transition between group “a” and the next group “c” consisting of January, February, March and July. September and December are in different groups of “d” and “de”, respectively while October and November constitute the last group, “e”. An interesting outcome of this analysis is that while the weekdays’ speed showed the same subsets over opposing links, the same cannot be said of the monthly speed. For example, the month of August classified in a different group on Link0506 is grouped with May on the opposing Link0605. Therefore, estimating speed using Bluetooth data may be better considered on a weekday basis than on a monthly basis, particularly when considering the LOS (level of service) each way due to the significant variations in the monthly average speed. However, the monthly variation in the speed data is consistent with the flow; thereby indicating a level of reliability in Bluetooth derivatives as indirect measurements. The next step considers the Bluetooth detection rates to round up the investigation.

Groups	Treatments	Means
a	Thurs	45.36
ab	Mon	45.13
b	Sat	45.04
b	Tues	45.03
b	Wed	45.03
b	Fri	45.00
b	Sun	44.97

Table 6.7: HSD test for weekday means of speed (km/h) over Link0605

Groups	Treatments	Means
a	May	46.00
a	Aug	45.92
ab	Apr	45.86
bc	Jun	45.45
c	Mar	45.39
c	Feb	45.38
c	Jul	45.28
c	Jan	45.12
d	Sep	43.98
de	Dec	43.58
e	Oct	43.41
e	Nov	43.21

Table 6.8: HSD test for monthly means of speed (km/h) over Link0605

6.5 Understanding Variability in Bluetooth Detection Rates

6.5.1 Background to the detection rate

Bluetooth presents a sample of the actual traffic. Therefore, it becomes imperative to understand the spatio-temporal variability in the estimated proportion of the actual vehicular flow captured by Bluetooth to inform usability. The early studies on Bluetooth have suggested that approximately 5% of all vehicles contain a form of Bluetooth-detectable device (UMCATT, 2008). However, with an increase in Bluetooth usage, as well as differences over different geographical locations (Beca, 2011; Biora *et al.*, 2012; Roggendorf, 2012), a proper understanding of this factor is considered essential. The knowledge of applicability is necessary to avoid over-generalisation, particularly given the contrasting nature of GMN. To this end, temporal and spatial variations in detection rates were investigated using the Bluetooth data collected over 2013 in Manchester, U.K. Overall, the problem considered here is to determine the changes in the detection rates over GMN, and whether the result holds irrespective of the data source and the location.

Clear distinctions were made between the rates obtainable based on different types of estimation and the ground-truth data used to inform usability. This distinction consisted of all devices, directional and total directional-based estimation to account for spatial relation and transferability. Dissanayake *et al.* (2012) noted the importance of the consideration for spatial transferability of a model. As a result, the consistency of the detection rate over the GMN was also examined to understand the differences and similarities spatially. Of course, differences are expected to be seen because the three networks are of a differing nature but how significant they are remains unknown. Appendix 6F presents other relevant results such as the detection rate variability plots over the hours of the day to understand temporal changes. Further, the hourly variation, weekday, monthly and seasonal variations were all examined for any significant differences over the different temporal dimensions. This information is also useful in determining the temporal transferability of a model to ensure efficiency in management.

6.5.2 Detection rate: all detected devices

In this section, the detection rate derived from the estimation of traffic flow based on the total devices detected is presented to understand the proportion of the total traffic equipped with Bluetooth-enabled devices. The Bluetooth data captured at Station 1011 co-located at ATC1283 location (the second validation station in Trafford), was filtered to remove duplicates and processed into 15-

minute flows. The total flow on both sides of the road measured by ATC was compared against the total flow at the Bluetooth station over the weekdays. A very strong relationship exists between Bluetooth and ATC flows with adjusted R^2 of 0.87 and 0.89 for the weekdays and weekends, respectively. The analysis of the data based on total modes gave the detection rate of 30% based on the flow ratio. A similar value (33%) was obtained over the weekday in the study conducted in Scotland (Cragg, 2013). However, it should be noted that this figure is not a representative of the actual vehicular proportion detected, and hence, rarely useful for congestion control. Consequently, directional estimation is considered in the next sections.

6.5.3 Detection rate: Wigan study site

In Wigan, Bluetooth Stations 1022 and 1023 are co-located with the validation Station (ATC1074) and are therefore considered for the analysis in this section. From the configuration, the location of ATC1074 is closer to the Bluetooth Station 1023 than 1022 (≈585m to 790m). The network configuration also suggests a possibility of a reduction in the devices detected at Station 1023 before reaching Station 1022. This is due to the possibility of vehicles taking an alternate route from Station 1023. Following the recap of the location description, the analysis of the detection rates (Table 6.9) showed a constant rate of 10% over the weekdays (Mon-Sun) in NW-bound. The rate of 14-16% was observed in SE-bound with the lowest rate observed on a Sunday. The difference between the directional rates is attributed to the relative position of the Bluetooth stations to the validation station as shown in the network configuration. The 5% difference on average between the NW and SE-detection rates showed that station calibration might not be sufficient to scale up an entire network especially in a network of varying characteristics. That is, the use of either of the computed rates (NW or SE) to predict the traffic flows on both links will result in over or under-estimation.

ATC1074										
Direction		Mon	Tue	Wed	Thu	Fri	Sat	Sun		MAC1022
NW		10	10	10	10	10	10	10		MAC1023
SE	Weekday	16	15	15	15	16	16	14	Adj. R^2	0.77 - 0.82

Table 6.9: Detection rates (%) derived from ATC, Wigan

6.5.4 Detection rate: Stockport study site

Two different independent measures of traffic flows (ATC and SCOOT) were used to derive detection rates in Stockport. The use of the data from the two ground truth sources presents the opportunity to explore variability in the rates arising from the two systems given they are positioned differently on the road. Each of the independent systems comprises two sets of validation stations. ATC1500 and SCOOT links 1034 and 1035 are co-located with Bluetooth Stations 1034 and 1035. Station 1035 is located close to ATC1500 (≈ 63m) while Station1034 is further apart (≈ 450m) with two main cross routes contributing to the traffic towards 1035. The SCOOT links, on the other hand, are located upstream and downstream from the link. Similarly, ATC1013 is a little closer to Station 1037 than 1036 (173m and 248m, respectively).

From the analysis, the detection rate of Bluetooth to SCOOT (13-16% as presented in Table 6.10) is higher than that of Bluetooth to ATC (7-12%). SCOOT rates were observed to be more consistent and precise than the ATC derived rates. This result showed that the location or positioning of the validation source relative to Bluetooth stations is significant in determining detection rates. This is due to the contributions from the connecting routes by way of vehicles leaving or joining the traffic. Therefore, an important practical implication of this result is that combining the SCOOT and ATC flows over a complex urban network to derive the Bluetooth detection rate may not yield the best result. This further means that the use of either of the two may be the preferred option for the purpose of consistency instead of the combined rates. Irrespective of the differences, an important observation from the results is that estimation of the detection rate is affected by both temporal and spatial variations. Therefore, the choice of sensor location is considered to play a critical role in the reliability of the resulting generalisation. Consequently, it must be taken as an important element for consideration during the installation of Bluetooth sensors.

ATC1500/SCOOT											
Direction	Sensor	Weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun		MAC1034
NW	ATC		12	12	11	12	12	10	8		MAC1035
SE	ATC		9	9	9	10	10	7	7	Adj. R^2	0.65 - 0.78
NW	SCOOT		16	16	16	16	16	16	16		
SE	SCOOT		13	13	13	14	14	13	13	Adj. R^2	0.82 - 0.92
ATC1013											
		Weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun		MAC1036
N			13	14	13	13	13	9	8		MAC1037
S			8	7	7	8	8	5	5	Adj. R^2	0.73 - 0.81

Table 6.10: Detection rates (%) derived from ATC and SCOOT, Stockport

6.5.5 Detection rate: Trafford study site

As with the previous cases, the validation station (ATC1024) is not centrally located between the two Bluetooth detectors. ATC1024 is closer to Station 1005 (508m) than 1006 (589m). The NE-bound and SW-bound monthly rates are 13-14% and 10-13%, respectively (Table 6.11). The January to September rate remains constant in a NE-direction while it remains constant from January to August for SW-bound traffic. December has the lowest rate (10%) while the September to November rate is 11% in a SW- bound direction. The 2% difference between August and September in the SW direction, as well as, the 2-3% difference over September to December between the directional rates, gives a strong indication of periodic variation. This means that periodic calibration will be required. This was established further through the seasonal differences observed in the data. For the seasonal-weekday studies, the summer period has the highest rate (15%) while the lowest rate was observed during autumn and winter (13%). However, for the “seasonal-weekend”, a constant rate of 12% was observed over the four seasons in the NE-bound direction and remains consistent (10-12%) in the SW-bound direction. The better precision observed over the weekend was as expected, given that variability is less over the free-flow than in congested period. Spatial consistency was observed where the networks exhibited similar characteristics and configurations. Overall, there is a presence of temporal and spatial variability in the estimated rates in the network. Therefore, Bluetooth traffic estimation requires both periodic and spatial calibration to obtain up-to-date and reliable predictions.

Direction	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
NE	14	14	14	13	14	14	14	14	14	13	13	13
SW	13	13	13	13	13	13	13	13	11	11	11	10
			Weekday				Weekend				Type	Adj. R^2
			Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Month	0.72 - 0.82
NE			14	15	13	13	12	12	12	12	Season	0.85 - 0.92
SW	Season		13	13	11	12	12	11	10	11	Weekday	0.76 - 0.81
			Mon	Tue	Wed	Thu	Fri	Sat	Sun			
NE			14	14	14	14	14	12	11			
Sw	Weekday		12	12	12	12	12	11	10		MAC1005, MAC1006	

Table 6.11: Detection rates (%) derived from ATC1024, Trafford

The monthly directional detection rates were analysed further to understand the statistical significance of the results. The monthly detection rates at the ATC1024 location in Trafford were investigated to understand monthly variability and the possibility of obtaining the most representative value. As a first step, the possibility to compute the mean of the two monthly directional rates were established through statistical testing. Figure 6.16 presents the summary of the analysis showing that the NE detection rates have $\sigma = 0.452; (CI = 0.252, 0.971)$ whilst the SW detection rates have $\sigma = 1.138; (CI = 0.706, 2.194)$. The overlap in the CI signifies that the two groups are not different and can be averaged. Further statistics showed that the ratio of σ = 0.397 and ratio of σ^2 = 0.158. Bonett's test (p-value = 0.039) suggests a significant difference between the two groups, and therefore, a rejection of the null hypothesis that the ratio is equal to 1. However, given that the sample size of 12 is less than 20, the Levene's test (p-value = 0.171), which is the greater of the two tests coupled with the overlap in the 95% CI for standard deviations, suggests the acceptance of the null hypothesis. This thus signifies that there is no significant difference in the ratio. Consequently, 13% was computed as the most probable value, (mpv) for the monthly detection rates over the Trafford based site on the mean of the two-directional monthly detection rates. The evidence from this research suggests that 13% is the best approximating value of the detection rate of Bluetooth vehicular detection over Trafford. The theoretical implication of this figure is that it only represents the best approximating value and not the true value. Therefore, it may be subject to over or under-estimation in some cases and thus require further investigation. The next step considers the comparison of detection rates in both directions on a day-to-day basis to understand long-term variability and the reliability of the detection rates.

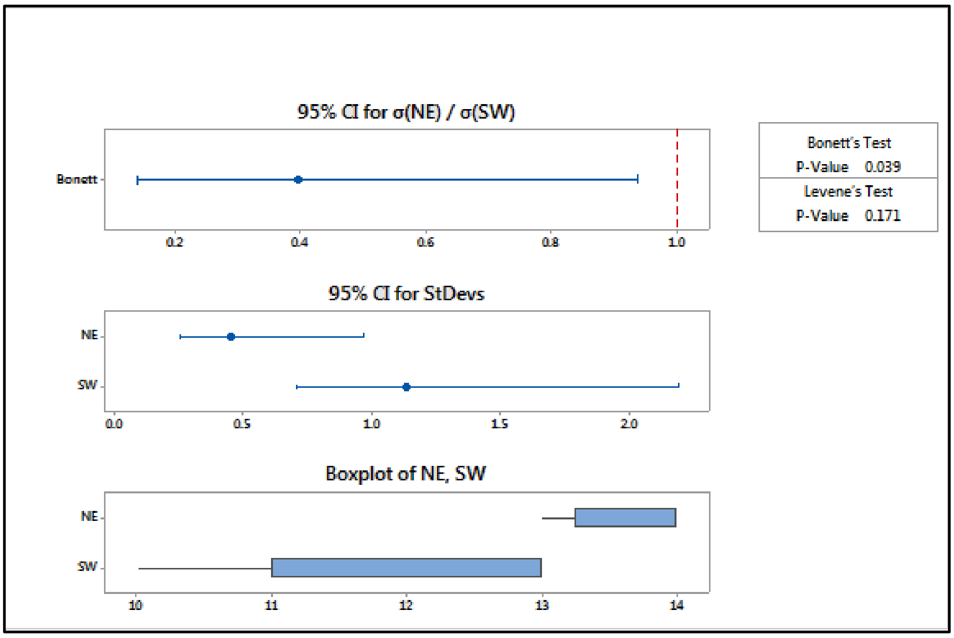


Figure 6.16: Plot of the F-test and CI for variances of NE and SW detection rates over Trafford

6.5.6 Understanding consistency and reliability in detection rates

Figure 6.17 and Figure 6.18 present the mean of the day-to-day NE-detection rates and Total-detection rates (both directions) over a year. The summary of the descriptive statistics for all the directional flows is presented in Table 6.12 while Appendix 6G presents additional results and the SW-equivalent time series plot. The quantitative analysis of the opposing directional rates showed that they could be averaged, as they are not significantly different. The hypothesis test of their distributions gave a point estimate of 0.02188 and 95% CI (0.02031, 0.02346) and the result is significant at 0.0000. In addition, the p-values for F-test (0.107) and Levene's test (0.118) at $\propto = 0.05$ confidence level for ratio $\sigma = 0.918, (CI = 0.822, 1.023)$ and $\sigma^2 = 0.843, (CI = 0.676, 1.046)$ showed that the ratios are not significantly different from 1. As a result, the plots of NE and Total-detection rates are presented in the discussion with reference to the SW-detection rates. Generally, the results show that detection rates fluctuate between 10% and 17% of all vehicles according to the time of the day and the day of the week. However, this result is in accordance with literature, such as Blogg *et al.* (2010), which obtained 17% on an average with a range of 2%-30% depending on the time of day. The combined detection rates present better stability and appears to be more reliable than the individual directional flow ratio, given the coefficient of variation 5.02 compared to 6.74 and 8.63 for the NE and SW ratios, respectively. On directional basis, variability is higher in the SW-direction compared to the NE-direction. The histogram plot presented in Appendix 6H as well as the coincidence of the mean and median show normality and symmetry in the detection rates. Overall, the results showed a negligible error and little variability over time. This signifies a high level of consistency and reliability in the estimation. For a better appreciation of the spread, the standard deviations are analysed further.

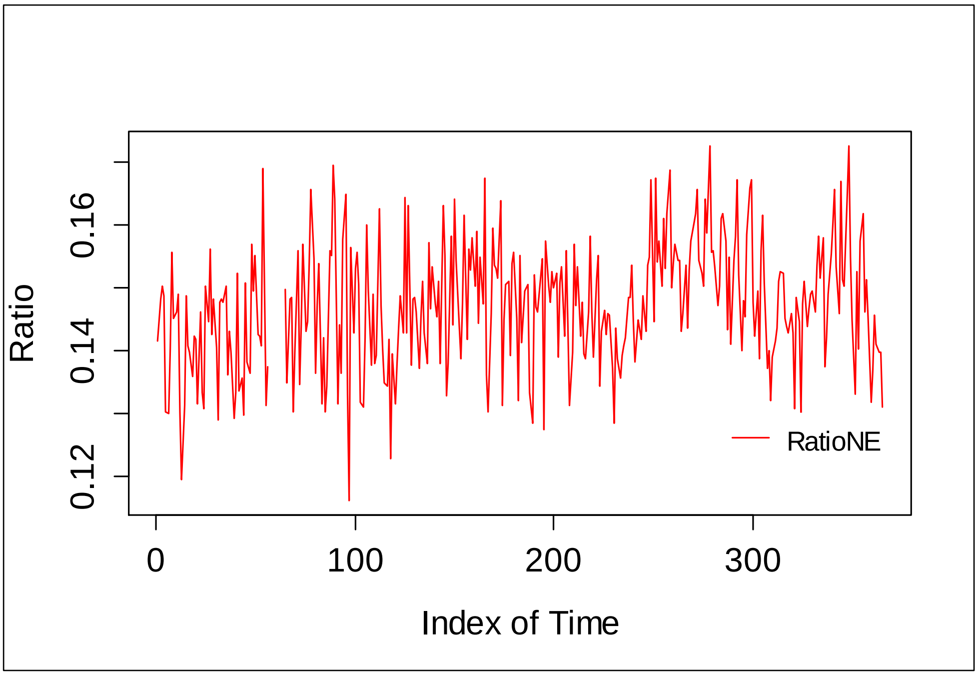


Figure 6.17: Mean plots of NE-directional flow ratio

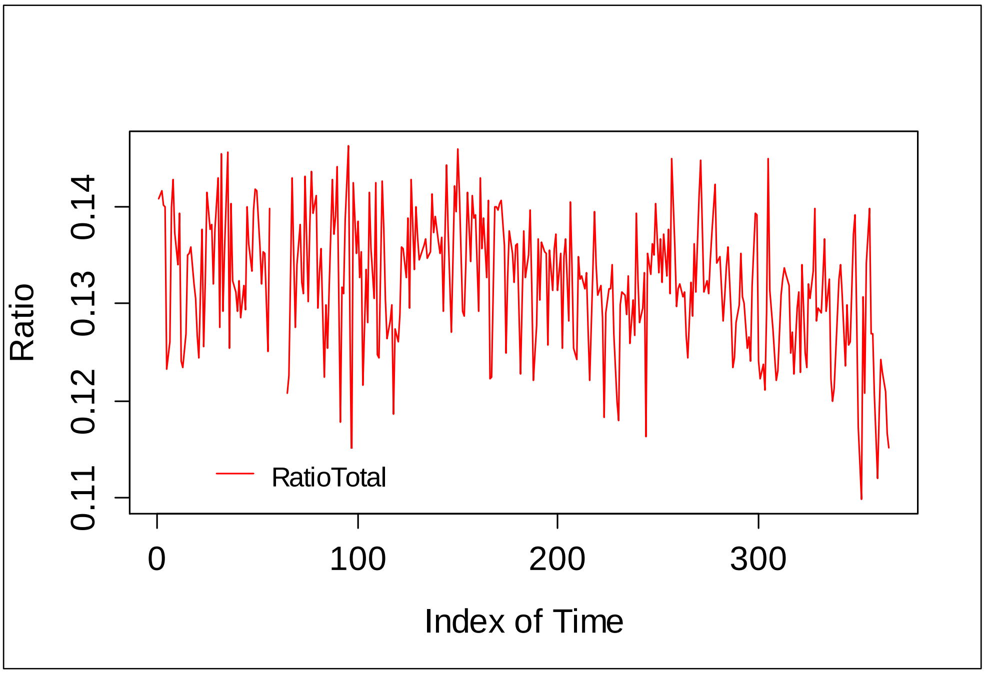


Figure 6.18: Time series plot of mean total directional flow ratio

Descriptive Statistics: Ratio_NE, Ratio_SW, Ratio_Total			
Variable	Ratio_NE	Ratio_SW	Ratio_Total
Total Count	357	357	357
Mean	0.15	0.13	0.13
SE Mean	0.00	0.00	0.00
StDev	0.01	0.01	0.01
CoefVar	6.74	8.63	5.02
Minimum	0.12	0.10	0.11
Q1	0.14	0.12	0.13
Median	0.15	0.13	0.13
Q3	0.15	0.13	0.14
Maximum	0.17	0.16	0.15

Table 6.12: Descriptive statistics of directional flow ratios

Figure 6.19 presents the time series plot of the standard deviations of day-to-day NE-detection rates over a year, while Figure 6.20 presents the standard deviation of the combined detection rates. From the results, the highest variability was observed in August for the NE detection rates and in November for the total detection rates. As shown earlier, the result of the estimated detection rates for the total flow (sum of the flows on the two opposing links) presents a better level of reliability given that it is more consistent than the individual link detection rates. However, any resulting generalisation must take into account the nature of the network. For example, two opposing links of differing attributes will present a different scenario. However, for all, the standard deviations of the detection rates clearly show a high level of precision – NE (0.03 – 0.10) and Total (0.03 – 0.06) – which signifies a high level of reliability. The representativeness of the Bluetooth sample of the actual traffic flow is established in Figure 6.21. The result showed that estimated sample sizes of 2331 and 8275 are required to obtain a maximum coefficient of variation of 5% and a maximum relative margin of error of 5%, respectively. The estimation corresponds to approximately 3% and 10% of the actual traffic, respectively, which is less than the average sample size obtainable over GMN. Interestingly, this result is also greater than the 2% sample size required to provide a statistically robust description of system performance as posited by Hainen *et al.* (2013).

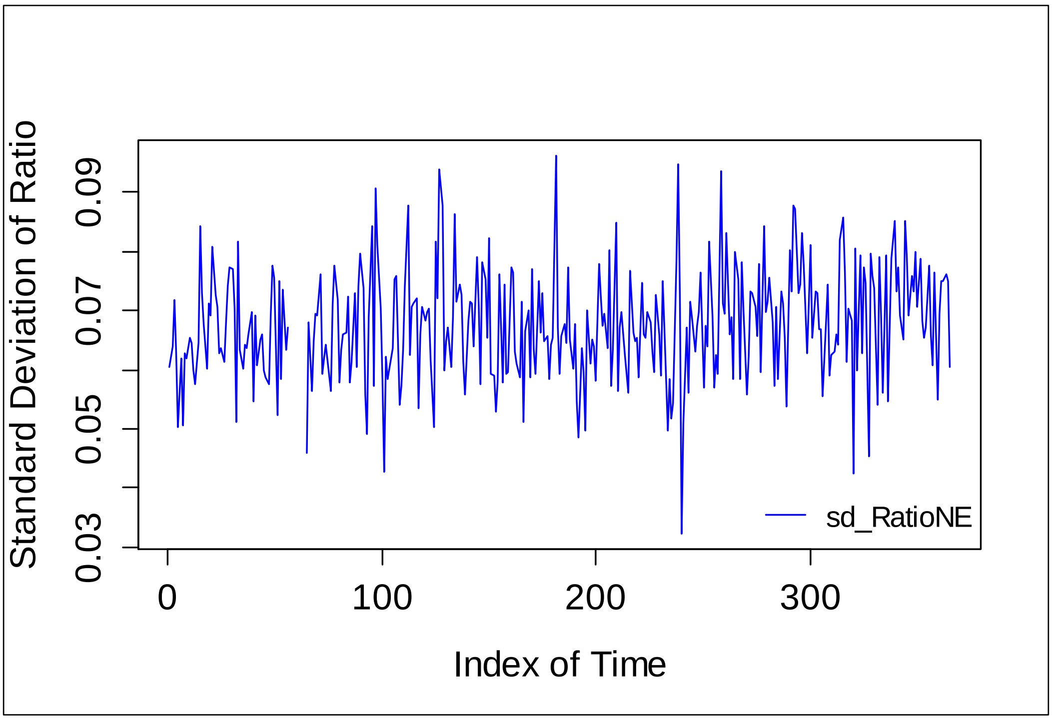


Figure 6.19: Time series plot of standard deviation of NE flow ratio

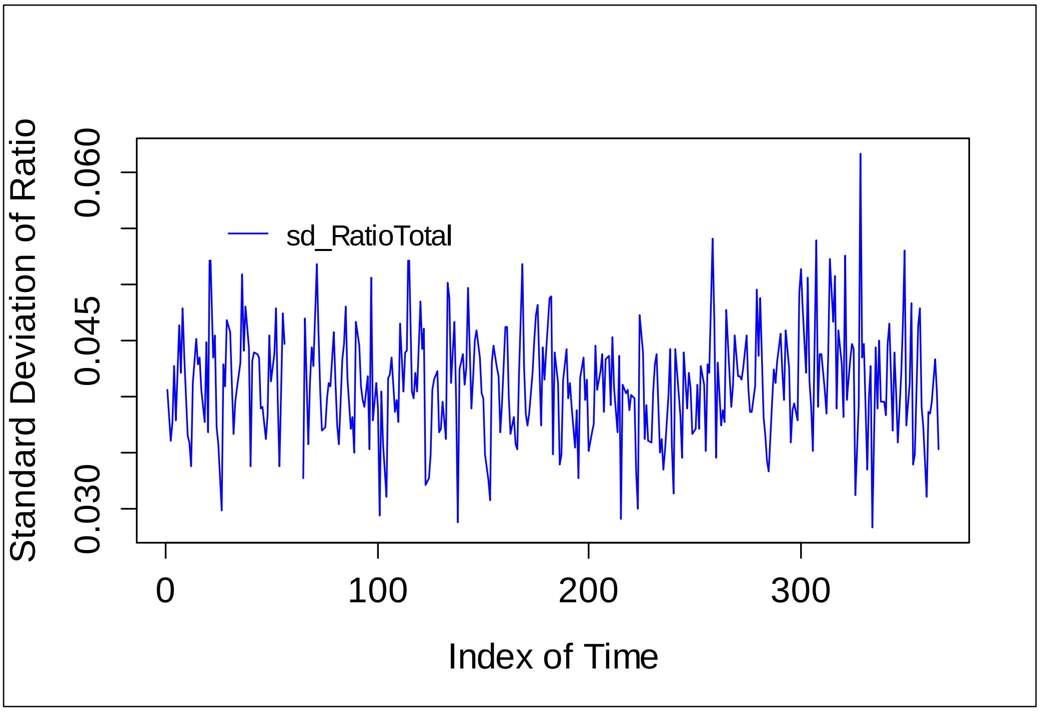


Figure 6.20: Time series plot of standard deviation of total directional flow ratio

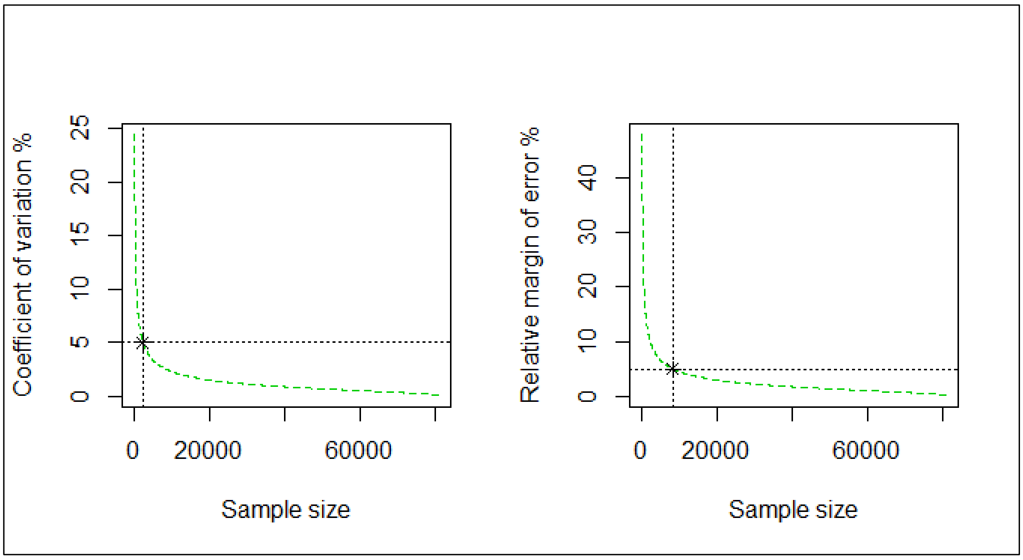


Figure 6.21: Plots showing the sample size in relation to coefficient of variation and relative error margin in percentage

6.6 Conclusions

This chapter presented the investigation conducted on the variability in the estimated Bluetooth traffic metrics to understand consistency and reliability and how the validity of the results might be affected by temporal variations. Exploratory analysis was used to understand the underlying properties of the estimated metrics, while post-analysis using the Tukey test confirmed the presence of significant temporal variations. The metrics showed contiguous homogeneous subsets over the am and pm peak and off-peak periods as would be expected in a real traffic situation. The test performed provided a concrete answer to the question “can Bluetooth capture temporal variations in traffic?” Analysis showed that the weekday average is the most consistent compared to other averages. Spatially, the highest variability was observed in Stockport, while in a network of similar attributes, total directional estimation is preferable on the grounds of accuracy and cost compared to the individual opposing links. The detection rates required to calibrate the Bluetooth estimate of the actual vehicular traffic computed over GMN using ATC and SCOOT flows yielded variable results, with an $, mpv$ (most probable value) of 13% in Trafford. This means that in GMN, a unique detection rate is not representative as a scaling factor for practical applications to avoid over/under-estimation. Therefore, caution must be taken in generalising the results over the entire network, and by extension over other geographical locations in the U.K. The relative position of the validation stations to the Bluetooth stations is also significant, and must be considered to obtain optimal results.

Finally, the results have so far shown that estimation of traffic metrics using Bluetooth can yield highly consistent and reliable results both in the short and long-term, and at the same time capturing the expected temporal variability. The results have also shown factors that must be considered, such as the level of aggregation of the data and the placement of the Bluetooth sensors relative to the validation stations. It is argued that harnessing this information might form an essential building block for more advanced theory on the use of Bluetooth data in ITS for traffic monitoring and management. The next chapter presents a discussion of the results interpretation and potential applications.

Chapter 7. Results and Interpretation of the Estimated Metrics

7.1 Introduction

In the previous two chapters, the Bluetooth estimated traffic metrics (flow, travel time and speed) were validated using diverse independently measured traffic data to establish consistency and the level of accuracy of Bluetooth traffic measurements. The estimated journey times, vehicle speeds and link-flows following the validation exercise all portend a high level of temporal and spatial consistency and a high level of accuracy. The results were also assessed for variability to avoid any biased conclusions. Following these steps, this chapter presents primarily the discussions of the Bluetooth results and their potential applications in traffic monitoring and management as well as the added benefits derivable from using Bluetooth in traffic sensing. This is in partial fulfilment of Research Objective iv to be complemented by Chapter 8, which considers the applicability and viability of the estimated traffic metrics in a wider context. This discussion focuses on the results obtained from the long-term study in the Greater Manchester Network (GMN) following the validation to avoid any bias in the interpretation.

Chapter 7 is structured as follows: the estimated traffic flow and the interpretation to traffic management application is described in Section 7.2. Three different types of traffic flow estimation using Bluetooth are presented for a better understanding of the applications namely i) all devices; ii) directional estimation; and iii) total directional estimation. Section 7.3 considers the interpretation of the results from the Bluetooth journey time by building on the previous studies such as UMCATT (2008) and Araghi *et al.* (2013) which showed that by sampling a portion of the travelling vehicles' actual times, reliable journey times data can be provided. In Section 7.4, are the results from the vehicle speeds and the interpretation to congestion management and traffic pattern analysis through a reconstruction of the actual traffic state at the time of observation. Section 7.5 discusses the results from the estimated O-D matrix and its usefulness in transportation planning and optimisation. The key focus of this section utilised a well-structured spatio-temporal analysis of origin-destination data from Bluetooth to provide answers to many relevant questions that may arise in the course of decision-making such as: i) Given a set of Bluetooth data, which part of the network is free of congestion? ii) Where within the network are road users likely to be exposed to pollution? iii) What time of the day/year is the congestion level highest/lowest? iv) Which link is most/least used? These important questions are considered before conclusions are drawn in Section 7.6.

7.2 Estimated Traffic Flows using Bluetooth Data

7.2.1 Estimation of total flow based on all Bluetooth detected devices

Figure 7.1 presents the profiles of the total (unfiltered generic traffic) and estimated (filtered vehicular traffic) flows derived from the Bluetooth data on Link3435 in Stockport. Link3435 is considered a good example for this illustration because it is a relatively short link (approximately 511m) with a speed limit (30mph – ≈ 48km/h). Further, the Bluetooth stations on the link are co-located with ATC and SCOOT detectors for validation. From the flow profiles that are deemed to be representative of the reconstruction of the real traffic at the time of detection on Link45 in Stockport, the total flow measured up to 700 vehicles an hour on average. However, there is, for example, the presence of other road users in the measurements that shows the profile does not reflect the true status of the vehicular traffic. From Figure 7.1, the presence of other modes accounts for more than 50% of the devices detected (from ≈ 300 to 700 vehicles at peak periods). Recall that Bluetooth sensors capture a range of enabled devices such as mobile phones and laptops carried by different road users moving in both directions once they are within the detection range. Consequently, any traffic parameter or metric derived from such measurement as performance measures will contain other road users that may not necessarily contribute to the traffic. Accordingly, such estimation is rarely useful for traffic management. However, data filtering as described in Section 3.2 helped in removing outliers, which reduced the flows to 200 – 300 vehicles on average over the 7 am – 6 pm period. The filtered flow (the lower profile of Figure 7.1) presents the actual reconstruction of the traffic state as shown in the validation presented in Figure 7.6. The filtered flow also presents typical traffic flow regimes compared to the total estimation that contains some amount of noise that usually causes unpredictability. However, the data filtering applied enabled the realisation of the ideal state of the vehicular traffic. This result shows the necessity for adequately handling outliers in Bluetooth data to obtain a realistic estimation. Otherwise, the results may be misleading.

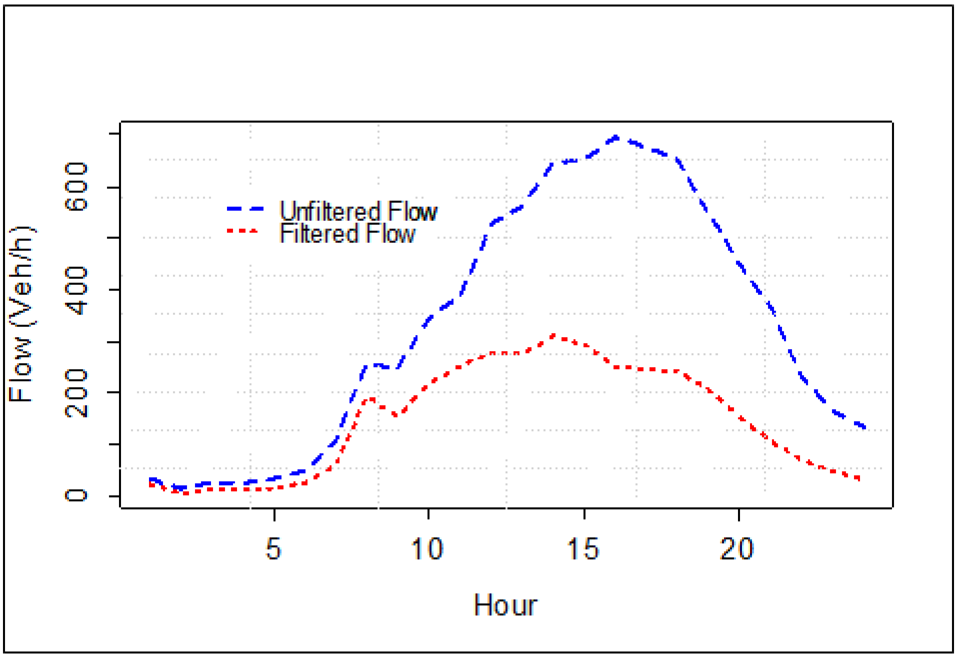


Figure 7.1: Flow profiles of unfiltered and filtered devices on Link3435

For a better understanding of the errors that might arise from the use of total flows, the speed distributions were plotted using histograms as shown in

Figure 7.2. The histograms present the opportunity to understand the distribution of speed variations from the detected devices. The knowledge of these speed distributions can be utilised in congestion and vehicle emissions studies. Generally, speed distribution is usually normal or approximately normal. The plot of the unfiltered speed does not follow this behaviour while the filtered speed is best represented with a normal distribution. The histogram plot of the filtered devices presents a clearer picture of the speed distributions over the

links. The variation in speed as would be expected is due to temporal changes in traffic volume on the link over the day. The first bin of the histogram of the unfiltered devices presents the highest frequency (n=15,500) associated with the group of devices travelling at less than 10km/h. This group was classified as non-motorised modes, and other extreme cases such as vehicle stop-over were excluded from devices classified as vehicles in the analysis. Accordingly, devices that were too slow or too fast were rejected according to the boundary and outliers filters described in Section 3.2. The filtering of the data leads to a 13% detection rate compared to 30% of the total estimation. Further analysis of the link speed on hourly average showed that vehicle speeds range between 35 – 51km/h. This signifies a high level of speed limit compliance as would be expected in UK urban areas. This is expected in an urban road given the UK policy on traffic violation that includes strict penalties.

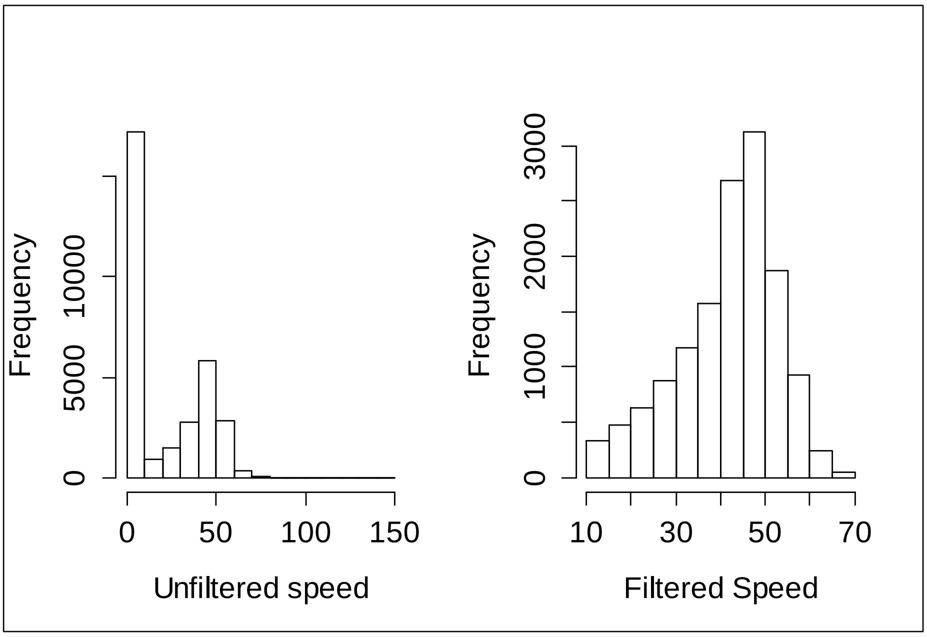


Figure 7.2: Histogram plots of the speed of all and filtered devices on Link3435

In addition, note that the station data captured by Bluetooth contains no information to indicate the direction of travel and is therefore limited in application without combining it with data from another station. Combining or merging Bluetooth data captured at different locations across the network leads to the realisation of the direction of travel, O-D matrix and the overall traffic metrics estimation. For example, the merging and analysis of the tracked devices at Stations 34 and 35 (a total of 33,651 devices - Figure 7.3 showed that about 50% of the total devices on the link were travelling within the boundary of the first filtering condition (speed ≥ 6km/h and speed ≤120km/h based on average walking speed of 5km/h). The implication of this is that if all the captured devices contribute to traffic and are considered as vehicles, about 701m road length per hour on average will be required to accommodate all the vehicles. This is assuming an average vehicle length of 2m in a 4-lane road (both directions) and with no gap between the cars and without scaling up the estimation. That is, the product of number of devices (33651) and the vehicle length (2) divided by the product of number of lanes (4) and number of hours (24). However, this situation is practically impossible considering the road configuration given above. This type of unrealistic scenario is presented when analysing Bluetooth data based on total devices captured. For a clearer picture, the speeds of all the 33,651 merged devices was analysed using the Mahalanobis distance method while boxplot was used to understand the properties of the distribution. The results clearly showed an unrealistic skewness in the data (except in extreme and rare occasions such as heavy congestion, which in this case, is not). Figure 7.4 shows the skewness in the data and the Mahalanobis cut-off point (2.448) for the outlying values.

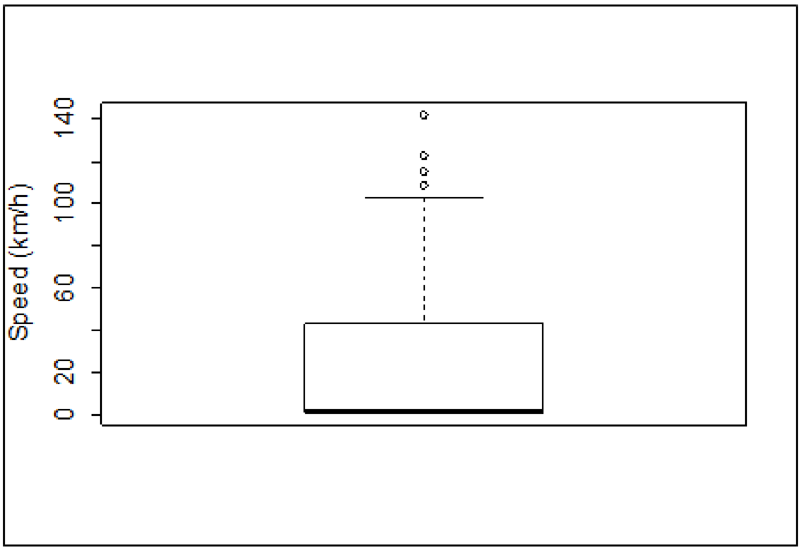


Figure 7.3: Boxplot showing the speed distribution of all devices on Link3435

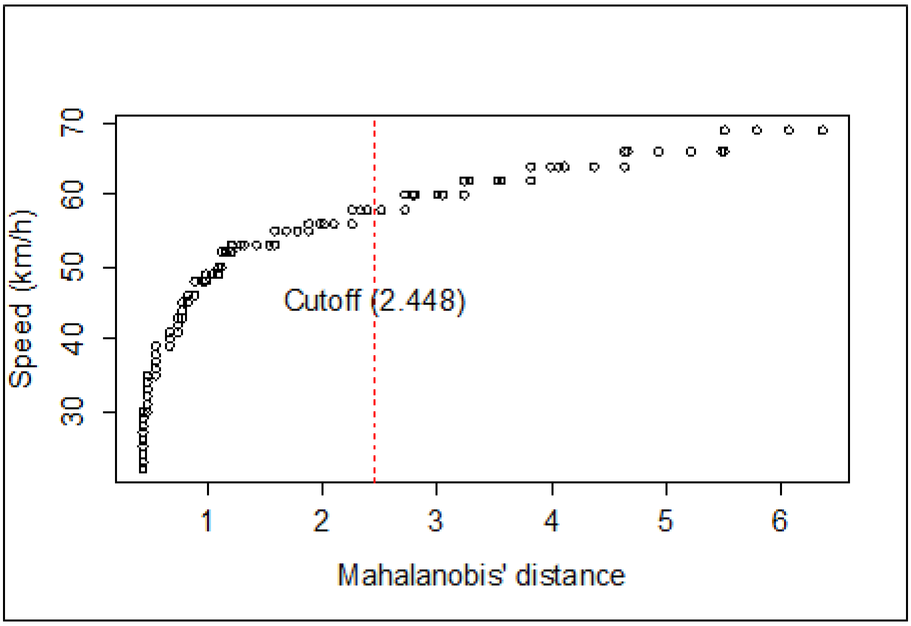


Figure 7.4: Plot of speed against the Mahalanobis' distances with the cut-off point

Figure 7.5 presents the boxplot of the vehicle speeds after filtering. The boxplot representation of the filtered devices showed that speeds <15km/h are outliers. Based on evidence from the SCOOT comparison, the boxplot of the filtered devices shows a better and cleaner representation of the road conditions compared to Figure 7.3. As shown in Figure 7.5, about 50% of the vehicles travelled between 35- 50km/h, which is more realistic and sensible, based on the road configuration. However, it should be noted that for a short distance, estimation errors might increase due to locational uncertainty arising from the detection zone. That is, the actual position of the detected device within the detection zone is unknown. If the device was detected at the exit and entry points at two consecutive stations, this will lead to an underestimation of travel time, and may consequently be interpreted as over speeding. Similarly, a detection of a device at the entry and exit points of the detection zone will lead to an over-estimation of travel time, leading to lower speed than reality. This is by extension affecting the vehicle count and any subsequent analysis such as pollution level monitoring. This is because valid vehicle records may be regarded as outliers and filtered out. Given this knowledge, the application of Bluetooth for speed compliant monitoring on a short link may not be desirable, particularly in an urban setting, as in this case.

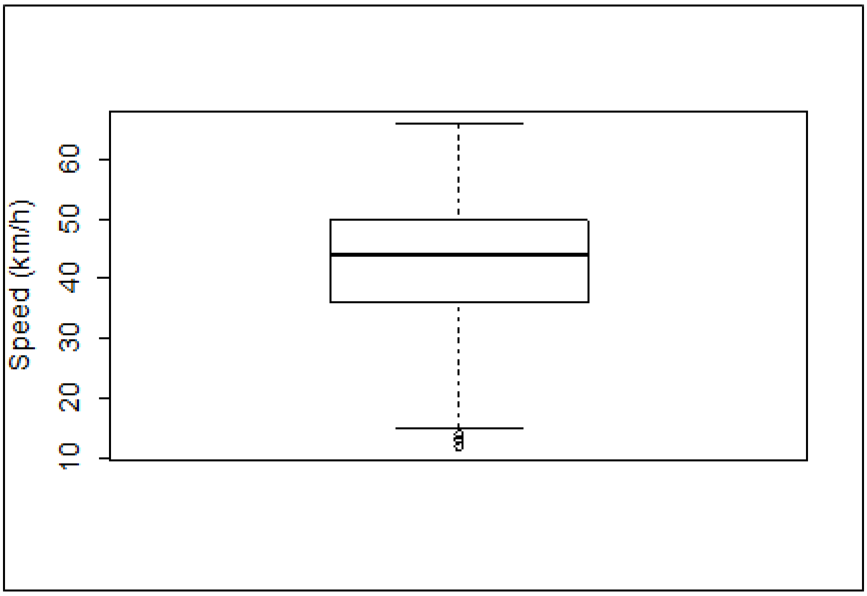


Figure 7.5: Boxplot showing the filtered speed on Link3435 in Stockport

7.2.2 Estimation of directional flows

In Section 5.3.1, the combined normalised flow profiles of Bluetooth, ATC and SCOOT in the NW-direction was presented showing a high level of consistency and reliability. This section presents the SE-equivalent of the results with emphasis on the hourly-weekday temporal variation over the months of April to October, to build on the discussion of flow estimation using Bluetooth. Consistent with the NW-directional flow profiles, Bluetooth and SCOOT flows present a better correlation compared to Bluetooth and ATC comparison over all the temporal dimensions considered. Figure 7.6 presents the hourly-weekday profiles of the flows over the period. Overall, variability is much pronounced over the peak period particularly with ATC flows. However, at low flow, the measured flows by Bluetooth showed a very strong relationship with the flows measured by SCOOT and ATC detectors. The evidence from this research shows that despite the variability, there exist the possibility of data reduction to minimise redundancy and consequently increase efficiency in data processing and information dissemination. This is evident from the day-to-day consistency in weekdays' (Monday – Friday) average and over the weekend

(Saturday and Sunday) as would be expected of real life traffic.

Overall, the consistency observed in the data over time following the validation presented in Section 5.3.2 signifies reliability, which indicates that the Bluetooth estimated flow can be used to build-up historical data for traffic management in the event of network failure. That is, the typical flow level obtained from Bluetooth for a particular day may be used as a substitute to avoid disruption in operation. In addition, the temporal correlation of Bluetooth with the ground-truth data implies the validity of the estimated flows. This is evidenced from the reproducibility of the actual pattern observed from the SCOOT measured flows. Therefore, Bluetooth has potential to understand temporal variability in a traffic network. In turn, this knowledge will serve as an aid to traffic signal timing and adjustment to ensure efficiency in the network.

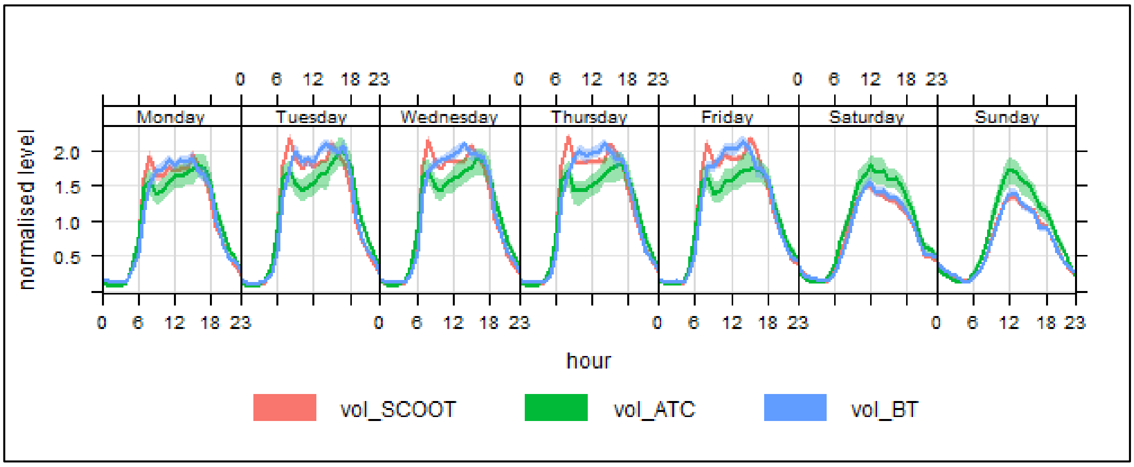


Figure 7.6: SE-directional flow profiles on link3435 in Stockport (18761)

7.2.3 Estimation of total directional flows

Figure 7.7 presents the profiles of the total directional flow on four different temporal dimensions on Link0506 in 2013 in Trafford. Figure 7.8 on the other hand, presents the superposition of the directional flows for a better understanding of the differences in the level of service each way on the link. The interpretation of the total directional flow profiles is consistent with the directional flows but in this case, the summation of the flows on the opposing links is presented. However, the slightly different peaks noticed (as expected) in the directional flows in the morning and evening for the SW and NE-bound flows respectively are evened out in the total flow. From the evidence presented in Section 6.5.5, the total flow thus shows less variability compared to the directional flow. One interesting thing about this result is that the temporal variation arising from work/school time and the close of work was captured in the data as reflected in the opposing links. The variability that is more pronounced in the NE-bound monthly flow has also been smoothed with the precision (less dispersion in the trend of the data) observed in the NE-bound monthly flow. Similarly, all the profiles at the varied resolutions present less variability in flows compared to the directional flow. The NE flows were higher than the SW flows on the weekdays and months in the year. However, similar trends such as variation between the peak and off-peak periods were observed on the opposing links. Through this analysis, one could infer the period of the day (giving the knowledge of "*when*") different strategies may be required on the opposing links because of differences in the level of service. For example, different strategies may be required between the hours of 12 noon and 6 pm as observed from the hourly-weekday profiles of Figure 7.8. Based on the evidence provided in this research, the results obtained showed a possibility of Bluetooth application to traffic congestion monitoring using Bluetooth measured flows. The next step examines the potential application of Bluetooth estimated flow.

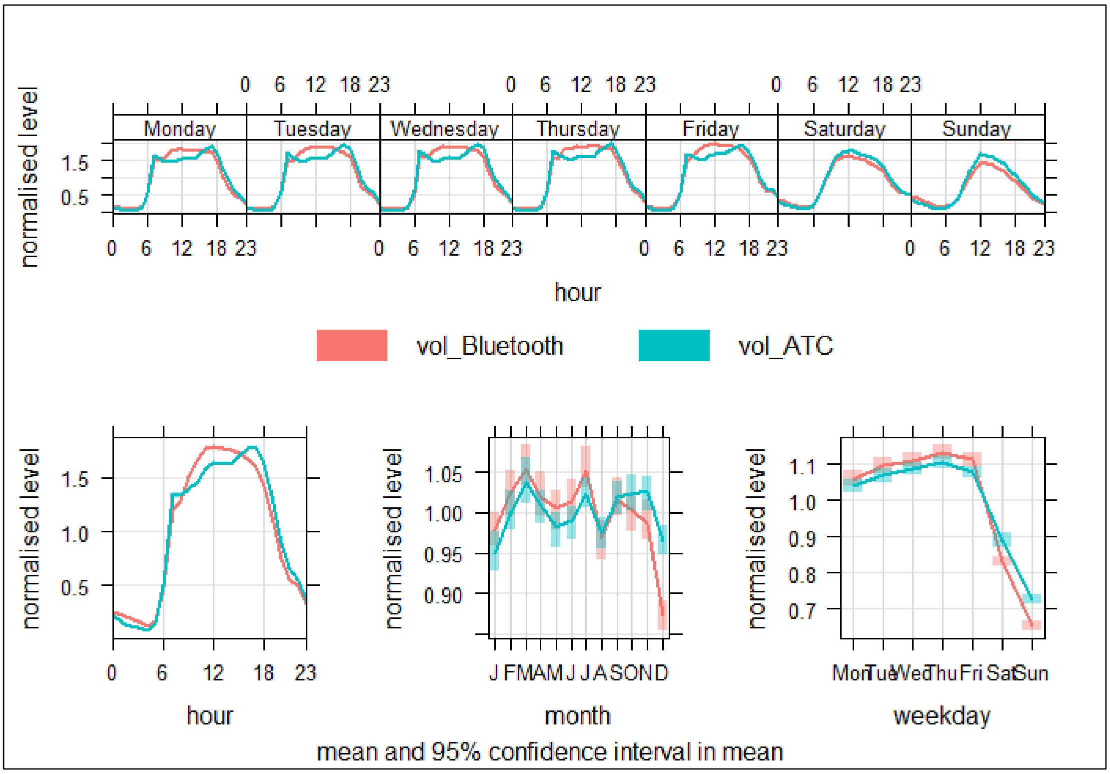


Figure 7.7: Profiles showing the total directional flow at different resolutions on Link0506 over 2013 in Trafford (N=31306)

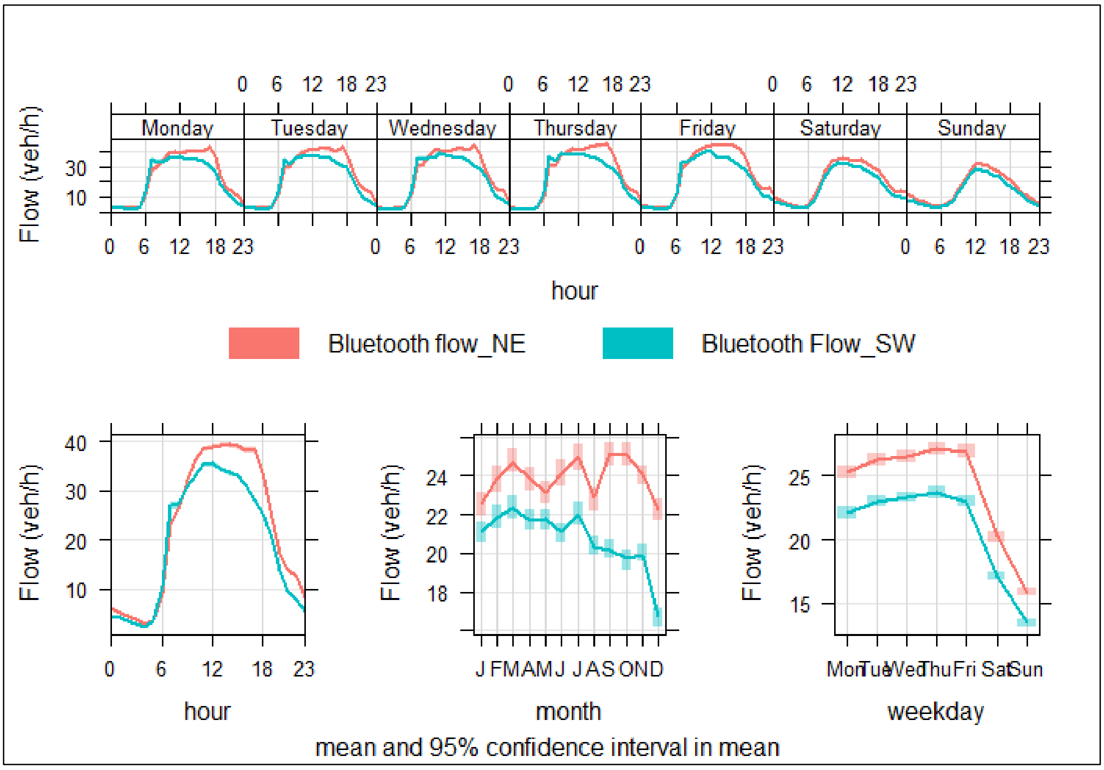


Figure 7.8: Profiles showing the superposition of the directional flows at different temporal dimensions on Link0506 over 2013 in Trafford (N=31306)

7.2.4 Understand temporal and spatial variations in flow

This section explores the capability of Bluetooth to understand both temporal and spatial variations in a network. Three non-consecutive links with independent measurement of flows in the Stockport network on the A6, Buxton Road were analysed to carry out this investigation. The first two links (3534 and 3736) are separated by 694m while the third link (4039) is 815m away from the second link. Normally, little variation is expected in the flows across the links due to the connecting routes and given the fact that the total link length is 2km. (See the location map - Figure 4.13 for the road configuration). However, if there should be any significant variation across the road, the first two links are expected to be more closely related given the evidence from the ATC validation. Therefore, the applicability of Bluetooth to capture both temporal and spatial variations in the measured flows is explored in this way as presented in the following figure (Figure 7.9) using a month's data over July 2013. Interestingly, all the links are closely related with a fine precision with the SCOOT measured flows as evidenced from the overlapping profiles. The correlation coefficients between the Link3534 to Link3736, Link4039 and the SCOOT measured flows are 0.999, 0.954, and 0.958 respectively. The results showed a very strong relationship between Bluetooth and the SCOOT measured flows with consistently similar patterns of traffic regimes over the hours and weekdays. Therefore, what readily comes to mind is that traffic from the connecting routes in this case has had no effect on the volume of the road network. That is, the number of vehicles joining and leaving the road section seems to cancel each other out. Given this situation, the same strategy might be sufficient to manage the network. The uniformilty in the traffic volume also means that the same detection rate may be used to scale up the estimated flow across the links whilst achieving the same level of accuracy in the estimation. However, special cases involving a network of different attributes must be carefully considered when computing detection rates to be used as a scaling factor for other links where they have not been directly computed, to avoid estimation error. Where the difference in volume reflects the actual change in flow levels spatially, for instance, this may signify a higher activity on that link than the other links. Therefore, to keep the traffic flowing to prevent congestion and blocking back on the link with the higher volume, more time will be required when the traffic light is on green. Temporal changes characterised by higher flows over the weekdays than on the weekend are evident from the results with evidence from the SCOOT modelled flow pattern. Therefore, the capability of Bluetooth to provide both spatial and temporal status information if utilised will inevitably contribute to efficient network management. For instance, real-time provision of traffic data to inform both temporal and spatial changes will enhance the management of traffic such as in traffic signal control for an optimised road. The knowledge of the spatial changes in traffic level will also facilitate a timely solution to avoid the building up of traffic. This will in turn, help the road users in the choice of optimum route during congestion to save time and fuel used in traffic. Using Bluetooth in this way offers a considerable advantage over the more expensive conventional data collection systems particularly in terms of cost.

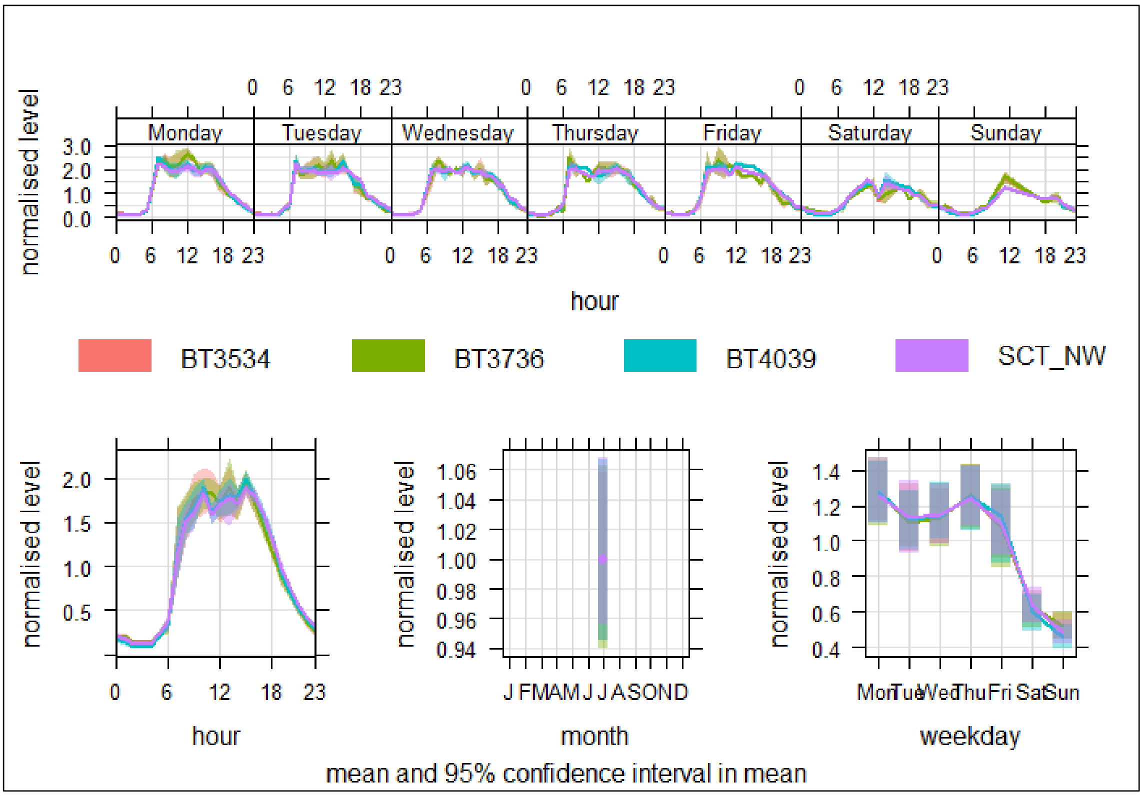


Figure 7.9: Bluetooth (BT) flow profiles on three routes over the month of July overlaid with SCOOT (SCT) flows northwards on London/Buxton Road, A6 (N=2976)

Figure 7.10 presents the normalised time series plot of Bluetooth and SCOOT flows captured during the month of July 2013 in Stockport over Link3435 SE-bound. The normalised flow profiles showed that Bluetooth is consistent with the SCOOT measured flow and representative of the actual flow captured by the SCOOT links over all the averages. This research has also shown that Bluetooth can detect temporal changes not only in the long-term but also on a day-to-day basis. An interesting thing in the use of Bluetooth as seen from these results and as evidenced in Chapter 5 is that despite being a low-cost sensor measuring a lower flow, quality is not compromised. This is a clear advantage offered by Bluetooth technology in terms of sustainable options over the conventional methods. Therefore, the application of Bluetooth for temporal status monitoring is considered a possibility. A significant advantage of Bluetooth technology in this respect over the conventional methods such as the inductive loop detector (ILD) is that Bluetooth can be installed in large numbers in a network, thereby leading to a more comprehensive monitoring of the network traffic than would be possible using ILD.

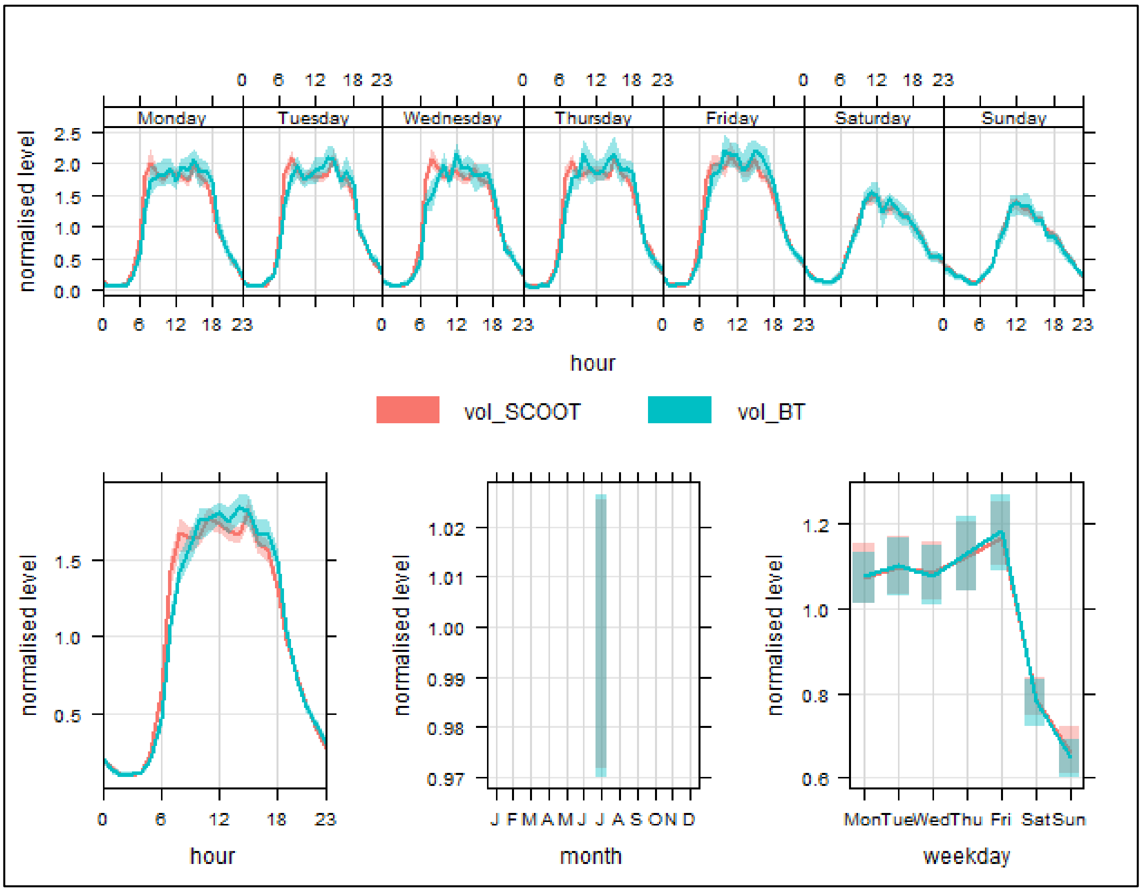


Figure 7.10: SE-directional time series flow profiles of Bluetooth and SCOOT on Link3435 in Stockport (N=2976)

7.2.5 Using Bluetooth estimated flow for data augmentation

One of the merits of Bluetooth technology is highlighted in its possibility in an integrated system through data fusion and augmentation to ensure continuous undisrupted network management (Bhaskar *et al.*, 2014). The possibility of this application is accentuated in the scatter plots presented in Figure 7.11 showing positive correlation in the monthly flows of Bluetooth and SCOOT. The scatter plots also showed hourly correlation with dispersion more pronounced over the peak periods. The evidence following the validation presented in Section 5.3 shows that Bluetooth data could be utilised to augment the existing systems as previously demonstrated by Bhaskar and Chung (2013). In some cases, such as understanding of the generic stream (total traffic), the technology may serve as a stand-alone sensor. Also, changes in temporal relationships of Bluetooth data with the IMT data inform the knowledge of usage and performance such as in peak and off-peak periods.

The concept of data augmentation becomes even more significant when there is a system failure arising from the traditional method of data collection. However, it may be argued that system failure is not frequent and, as a result, may not be a cause for concern. While this argument is valid, the application of Bluetooth helps removes reliance on archive data by the traditional system and is thus a significant added advantage to the existing system. Application of Bluetooth for data augmentation will include node-to-node data adjustment and fine-tuning of erroneous data points, thereby leading to avoidance of disruption in service provision. However, it should be noted that for a complex urban road network monitoring, Bluetooth may be insufficient because error from the detection rate will result in poor accuracy in the estimation. Therefore, its application should take into account the limiting factors highlighted in Section 2.3.3. Despite some limitations such as low count rate, the analysis of the results of Bluetooth data as seen in the scatter plots and based on the evidence from the validation exercise, suggests the possibility of data augmentation. Therefore, harnessing the potential of Bluetooth in data fusion and augmentation to extract value will be essential to capitalise on investment and to benefit from the resulting opportunities as noted by Harris (2014).

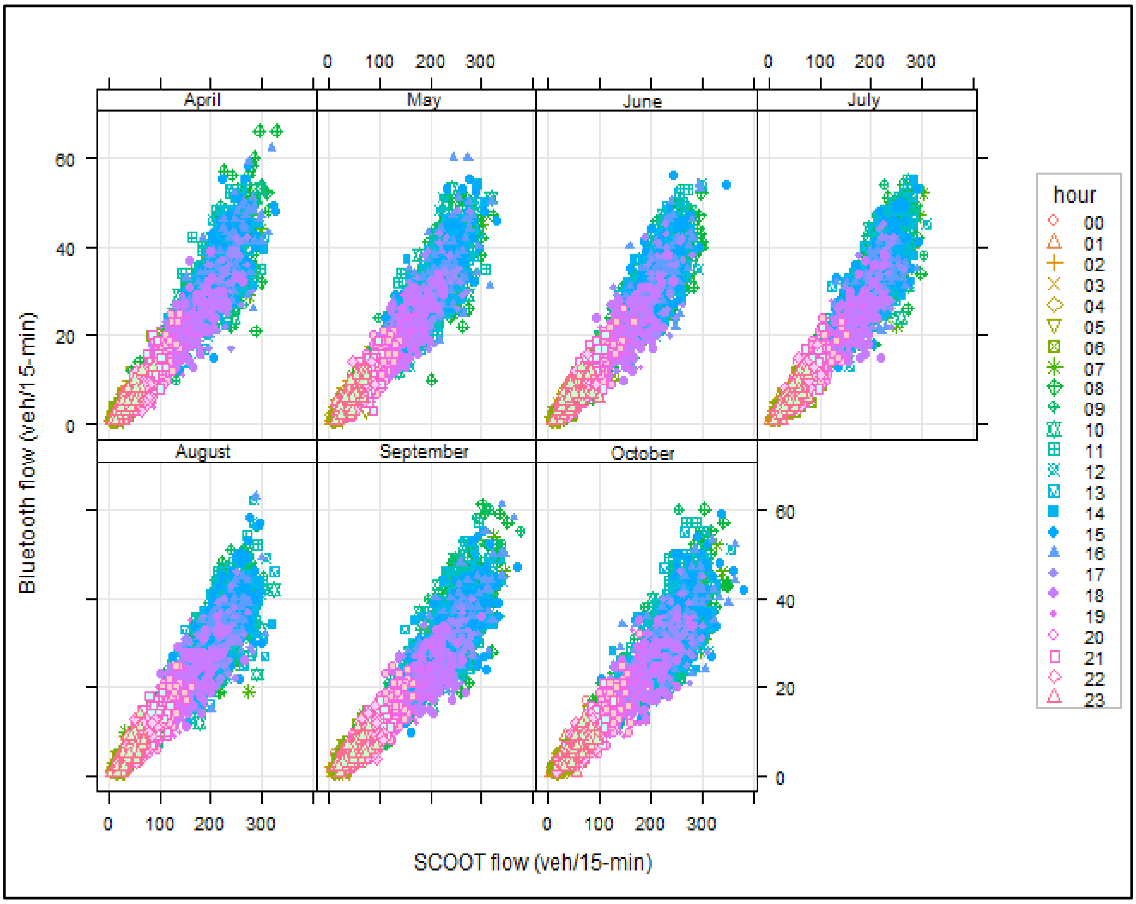


Figure 7.11: Monthly scatter plots of Bluetooth against SCOOT measured flows on Link3435 NW-bound, Stockport

7.3 Using Bluetooth Journey Time for Traffic Management

7.3.1 Journey time management using mean and median travel times

From the pilot study (Section 4.5.4), it was shown through the preliminary analysis that the median is the best estimator of journey times. In Section 6.3.1, a discussion of the analysis of the two estimators on a larger scale (using a year data) was presented. The analysis conducted a test for any significant difference between the two estimators to support the findings of the pilot study. That is, to find out if the median journey time is a better estimator than the mean for journey time management. On exploration of the time series plots presented in Figure 6.8, the median journey times can initially be argued as a better estimator. However, the confirmatory test performed between the two estimators at 95% confidence level stated otherwise. The result showed that the probability of a difference between the two estimators (irrespective of the temporal dimension) is less than 0.01. This is evidenced from the overlap of the point estimate with the 95% confidence interval (CI), with the p-values showing the significance level of the results as presented in Table 6.4. Therefore, the application of Bluetooth mean journey time is comparable to the median, and is thus considered a good estimator for journey time management.

7.3.2 Journey times for network planning

This section presents the results of journey times to show both temporal and spatial variation in travel time across the Wigan network. Figure 7.12 presents the O-D analysis of the network journey times distribution on an hourly basis across the Wigan Network. The results showed the amount of time it takes to traverse the network from Station 12 located on the A49 southwards to the respective stations under consideration as shown in this figure. The x-axis presents the different routes under consideration. From the weekly and daily analyses of the results, it was observed that it takes a longer time to move from Station 12 to Station 21 (a distance of 5.83km) connected with a major road than to move to Station 24 (a distance of 4.87km). The shortest distance 0.89km (Link1216) has the shortest journey times (150 seconds on average). However, this is expected as seen in the network configuration between the two stations – the location map is presented in Figure 4.8 – Section 4.5. The capability of Bluetooth to capture the spatial variations implies the possibility to support network planning for the delivery of enhanced services. While these variations could be captured by other methods of data collection, these other methods cannot be deployed in large numbers, unlike Bluetooth.

The analysis of the network in terms of the journey times showed clearly that the length between two stations might not necessarily correlate with their journey times. For example, Link1231 has a shorter distance (1.6km) with an average journey time of 294s compared to Link1218 (3.02km) with a journey time of 423s on average. This shows that the journey times within an urban network are not solely dependent on the link length, but also on other important variables, including but not limited to, the road types linking the stations together and the land use of the area. With the capability of Bluetooth data to be transmitted and analysed in real time, this type of information presents network engineers with the opportunity to optimally manage the network for efficient flow. This can be seen in the areas of traffic signal timing control, suggestion of alternative route(s) and parking guidance through a personal alert system or VMS. Summarily, the consistency observed in the data as in Bhaskar and Chung (2013), and as noted by Beca (2011), gives a level of reliability to Bluetooth journey time estimation to support decision-making for network optimisation. From the above, answering questions such as “which is the optimum route in the network?”; “what is the time it takes from one origin to another destination?”; “what time of the day is the journey time longer?” or “when or where can congestion be experienced in the network?” becomes realistic. Any change or sharp departure from the normal trend might be indicative of an incidence occurrence that needs to be investigated and/or given attention. This type of information is also useful in understanding travel time index (TTI).

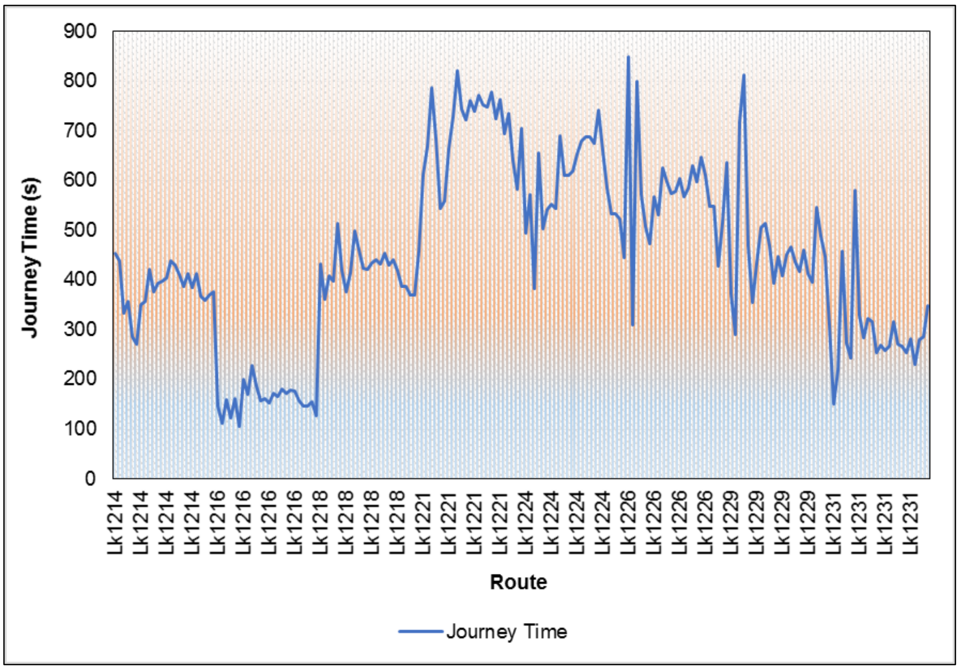


Figure 7.12: Weekly distribution of journey times across the Wigan Network

7.3.3 Using Bluetooth for the study of travel time index

This section presents the knowledge of how Bluetooth might be used to evaluate travel time index. A working definition of congestion is 'travel time or delay in excess of that normally incurred under light or free-flow travel conditions' (Gifford, 2003, page 181). HCM (2000) defined traffic delay as the delay component resulting from reduction of speed below the free-flow speed due to the interaction of vehicles. Travel time index (TTI) found significance in calculating and understanding of the reliability of performance measures through the day-to-day variation in travel time (Lomax, 2010). In Section 6.3.3, the knowledge of the day-to-day variability in journey times was explored. According to Lomax (2010), this variation describes the amount of time that road users should allow for in an important trip. Furthermore, reliability measures are particularly useful for identifying the effect of system management strategies designed for efficient traffic operations (Lomax, 2010). TTI is defined as (Lomax, 2010, page 6):

$$Travel\ Time\ Index\ (TTI) = \frac{Delay\ Time + Free\ flow\ Time}{Free\ flow\ Time} \tag{7.1}$$

Simplifying Equation (7.1) becomes:

$$TTI = \frac{Time\ at\ Congestion}{Time\ at\ Free\ flow} \tag{7.2}$$

The mean journey times over free-flow and congested periods on the A56, NE-bound, Trafford were analysed.

Figure 7.13 presents the hourly journey times over the month of November showing temporal variations. Consistent with other results on journey times which are characterised by temporal changes, the analysis showed that more time was spent in traffic during the congested period (250 s) compared to the free-flow period (100 s). Over Link0506, a TTI of 2.5 was computed based on equation 7.2. This factor (TTI) is useful in determining the amount of extra time spent in traffic. For example, a journey of 100 seconds at free-flow will translate to 114 seconds during the congested period for a TTI of 1.14. This showed that an extra time of 14 seconds was spent in traffic given a TTI of 1.14. With a higher value of TTI, the amount of time spent in traffic will increase accordingly.

This information provides not only the knowledge of the additional time spent in traffic but also the idea of the changes that might occur in the network. Such information can also be evaluated in terms of the amount of fuel consumed to evaluate the economic impact of the additional time spent in traffic. The application of Bluetooth to understand this phenomenon can help route planners to put in place an appropriate management strategy to reduce unpredictability in travel time that may in turn affect driving behaviour. However, while other methods of traffic data collection can be used in this regard, Bluetooth offers the advantage of cost. Summarily, the accuracy of travel time estimation using Bluetooth data suggests the possibility of TTI application.

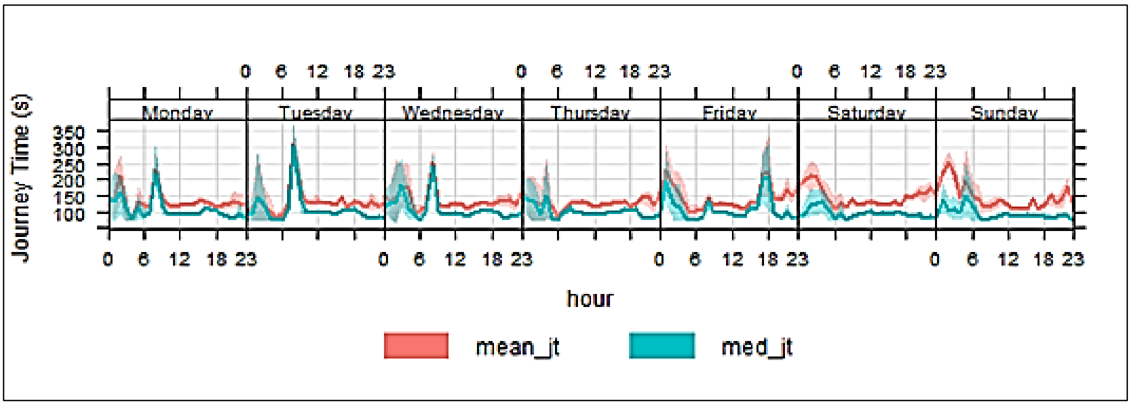


Figure 7.13: Hourly travel time over the month of November on Link0506 (N= 2880)

7.4 Using Bluetooth Journey Speed for Traffic Management

7.4.1 Using the mean and median speeds for congestion management

This section explores the use of the Bluetooth estimated mean and median speeds to reconstruct traffic state to understand congestion patterns such as free flow and congestion. Journey speed in kilometers per hour (km/h) is the average speed of a traffic stream obtained from the length of a road segment divided by the average travel time (HCM, 2000). Figure 7.14 presents the reconstruction of journey speed over each hour of the day with consistency observed in the profiles from day-to-day. Unlike the journey time, the mean and median speeds overlap each other showing clearly that there is no significant difference between the two estimators. This shows that either of the two estimators can be used to understand congestion patterns to achieve the same results. Figure 7.14 shows that the speeds of vehicles using both estimators are higher over the early hours and late in the night than during the working hours of the day. The regular dip observed over weekdays at about 8 am indicates that this is the most congested period of the day, thereby giving an indication to when pollution may be highest in the day. The speed over the weekend is higher with less variability compared to the weekdays as expected. From the available evidence and the validation of the vehicle speed presented in Section 5.3.3, which showed a high level of accuracy, it is concluded that Bluetooth can be used in congestion management to minimise pollution arising from vehicle emissions. Potential applications include congestion level monitoring and density estimation. Already, density estimation has been demonstrated using Bluetooth through data fusion (Bhaskar *et al.*, 2014).

In addition, the traffic regimes depicted in the Figure 7.14 showing variations in speed level are analogous to the reconstruction of real traffic at the time of occurrence. As expected, speeds of vehicles are higher over the weekend (47km/h) compared to 45km/h over the weekdays. The closeness of the speed distribution over the weekday is attributed to the speed regulation, and thus suggests the possibility of Bluetooth to contribute to monitoring the speed compliance level in a given road network. Based on the available evidence, using Bluetooth in this way will assist traffic managers to understand what time of the day or day of the week speeds are usually low such as days on which football matches are played, or an hour before or after football matches. If this trend is monitored efficiently over time using Bluetooth, appropriate control measures could be put in place based on the information provided by Bluetooth to minimise traffic congestion, and thus its attendant environmental pollution. Control measures may include re-routing by relaying the information gathered from Bluetooth to road users through VMS. For example, the displayed information may include a restriction to private cars on key corridors to promote the use of public buses, and thus a reduction in the number of vehicles on the road and the amount of emissions generated. Alternatively, there could be the implementation of road-user charging communicated using Bluetooth to control traffic in such instances. In that case, private car users will have to pay to use the roads over these specific periods. A significant advantage of using Bluetooth in this regard is that additional infrastructure such as radio frequency identification (RFID) tag may not be required.

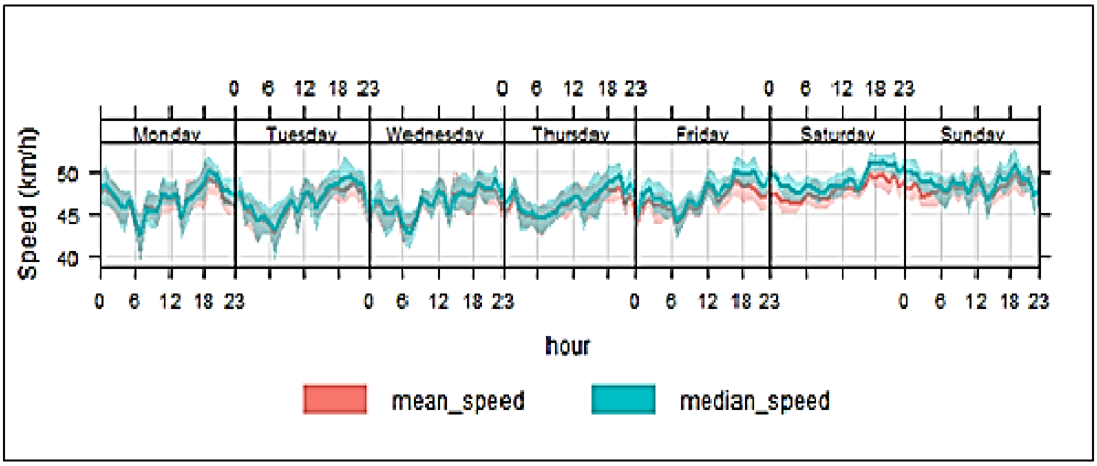


Figure 7.14: Non-normalised mean and median journey speed on Link0506

7.4.2 *Application of Bluetooth for speed limit compliance monitoring*

A speed limit is defined as the maximum, legally permissible driving speed along a specific good road section and under good travel conditions (RTA, 2011). Speed limits are usually imposed on roads to control traffic and are primarily for two things: i) To reduce risks imposed by drivers' speed choices leading to potential vehicle conflicts; and ii) To provide the basis for punishment for road offenders who endanger the life of others (DoT, 2015). In this regard, Bluetooth was analysed to understand whether it could be used to monitor the compliance level of motorists to the speed limit and to understand the safety level of the road users. From the investigation, analysis showed that on a road with 48km/h speed limit, from day-to-day, the average speed over the link is between 30km/h and 65km/h. The first observation from this result is that a certain percentage of the detected vehicles travelled at speeds above and below the speed limit. The compliance level analysis also suggests that about

20% of the total vehicles plying the route travelled above the speed limit while about 80% are speed limit compliant. Given that 20% of the vehicles travelled above the speed limit over the year, the conclusion here is that this percentage cannot be attributed to only high-speed vehicles such as ambulances but also some road offenders who break the speed limit. With this type of information, there are different possibilities to address this issue to reduce the risk posed by the offenders which include: i) deployment of security personnel to arrest offenders; ii) the use of VMS to warn road users of over speeding; iii) introduction of traffic calming where necessary, and iv) implementation of a policy to register the MAC address of a vehicle which, in this case, will be synonymous to the registration of vehicles' numbers. The possibility of this type of application will be of significant benefit in terms of both cost and safety. However, it may raise security and privacy issues. That is, motorists may attribute such application as a breach of privacy right and that they are covertly monitored. To clear any doubts will require policy review and public sensitisation to educate the road users. If this is a welcome idea by the public, then the implementation of a real-time warning system to reduce the risk on the road also becomes a possibility. However, it is to be noted that a certain percentage of the defaulters may not be captured given that Bluetooth only captures the sample of the total traffic. Nevertheless, Bluetooth can be harnessed in this regard to complement existing technology such as speed cameras to derive safety benefits and an enhanced operation. The next section considers the possible application areas of O-D matrix using Bluetooth data.

7.5 Using Bluetooth O-D Matrix for Traffic Management

7.5.1 Origin-destination matrix for network planning

In this research, 6,159 hourly O-D matrices were analysed over six months across the three networks in Greater Manchester to understand temporal and spatial variations in the network traffic. Figure 7.15 conceptually shows the origins and destinations for the Wigan network and, for each hour, a matrix was produced. The stations used were selected within the network at strategic locations to cover the spread of the Bluetooth sensors while improving the computational efficiency. A one-headed arrow indicates one-directional flow while a double-headed arrow indicates bi-directional flows. For computational efficiency and ease of understanding, the data were transformed to vector form. That is, each O-D matrix was transformed from a 2-dimensional object to a 1-dimensional object. Table 7.1 gives the proportion of the vehicles tracked across the network. O-D pair 12 and 16 has the highest proportion (12%) of the hourly network flow while O-D pairs 21-29 and 26-29 both have the lowest proportion (0.1%).

The vectors for each hour from day-to-day were compared against each other using the function ‘rcorr’ to compute the correlation coefficients and the p-values as given in Table 7.2. The function rcorr is in R statistical package (R Core Team, 2013). The p-values help in understanding the significance of the results. A very strong positive correlation was observed from day-to-day, with a high significance level at 95% confidence. Unlike the link-based analysis, the O-D matrices comparison showed a strong correlation between weekdays and weekends. The improvement in the correlation coefficients is expected given the volume of the data used in the O-D matrix compared to the link flow estimation. The strong relationship in weekday data thus suggests a possibility for data reduction to improve computational efficiency. The day-to-day consistency in the measurements means a level of reliability in the data. This demonstrates that the day-to-day monitoring of the O-D can provide the data needed to compute and plan traffic management interventions in response to, for example, air pollution events and incidents. In addition, routine assessment of the impact of the intervention is made possible. More important is the monitoring of any significant changes in O-D that may occur because of roadworks and accidents.

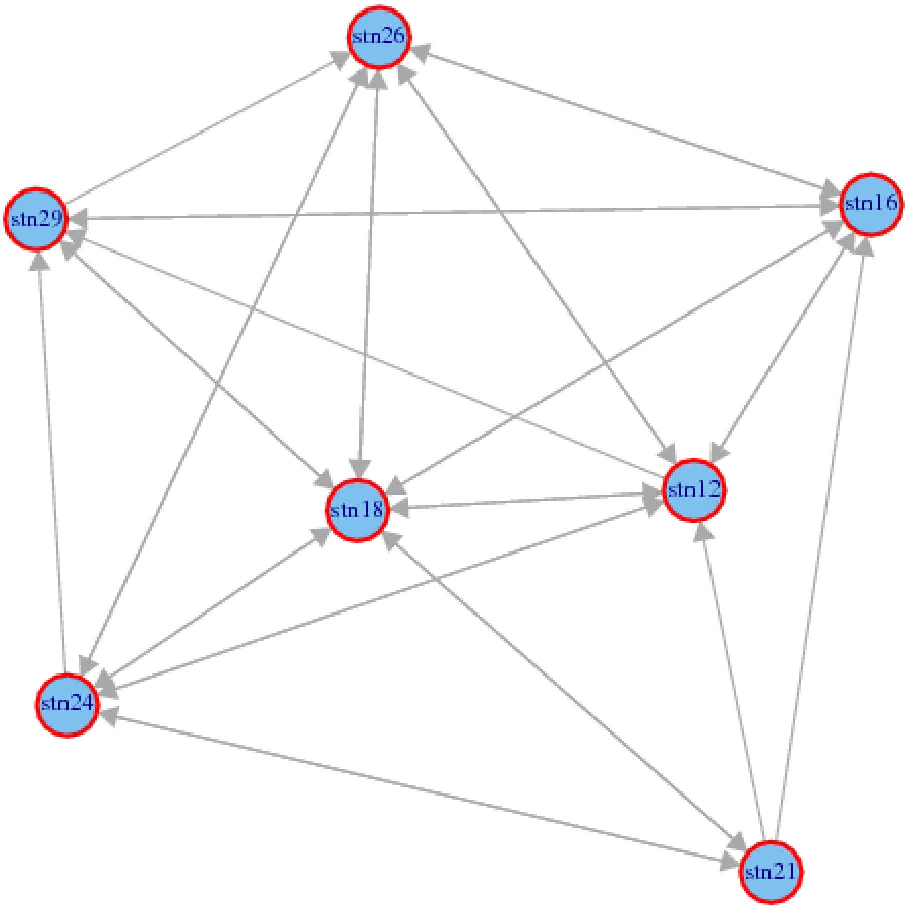


Figure 7.15: A typical plot of an O-D matrix in the Wigan network

	Stn12	**Stn16**	**Stn18**	**Stn21**	**Stn24**	**Stn26**	**Stn29**
Stn12	0	0.085	0.072	0.003	0.019	0.01	0.011
Stn16	0.12	0	0.096	0.002	0.004	0.003	0.017
Stn18	0.071	0.107	0	0.024	0.017	0.022	0.014
Stn21	0.017	0.018	0.044	0	0.032	0.005	0.001
Stn24	0.022	0.003	0.018	0.014	0	0.017	0.003
Stn26	0.014	0.005	0.029	0.004	0.021	0	0.001
Stn29	0.017	0.005	0.007	0	0.004	0.007	0

Table 7.1: Proportion (%) of traffic flow across Wigan network

	Sun	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**
Sun		0.0001	0.0003	0	0.0041	0	0.0022
Mon	0.0001		0	0	0.0003	0	0
Tue	0.0003	0		0.0004	0	0	0
Wed	0	0	0.0004		0.0054	0	0.0013
Thu	0.0041	0.0003	0	0.0054		0.0014	0.0002
Fri	0	0	0	0	0.0014		0.0003
Sat	0.0022	0	0	0.0013	0.0002	0.0003	

Table 7.2: P-values of hourly O-Ds in Wigan for seven days

7.5.2 Hourly origin-destination matrix for network optimisation

In the previous section, the day-to-day consistency in Bluetooth O-D matrix estimation was demonstrated through correlation analysis to understand relationship and strength. The results from Table 7.3 demonstrate the ability of Bluetooth as a technology to provide hour by hour O-D matrices as demonstrated by Barceló *et al.* (2012). Such information can be used in traffic models to explore solutions for tactical intervention plans to optimise specific (or a combination of) performance measures for the smooth running of the network. Finally, Table 7.4 presents the correlation coefficients (0.96 – 0.98) of the weekday O-D matrices. The values of the correlation coefficients signify a strong positive relationship between the weekday O-D matrices. The consistency of the hourly O-D from day-to-day signifies the possibility of building up historical data, for example, in the event of data failure. Based on the evidence provided in this research and from literature, Bluetooth is considered a viable option to enhance traffic management. This enhancement can be seen in different traffic management applications using the O-D matrix information for planning and implementation purposes.

Flow (Veh/h) JT (S) Speed (Km/h)	Stn1012	Stn1016	Stn1018	Stn1021	Stn1024	Stn1026	Stn1029
Stn1012		78 131 26	38 390 29	3 1038 21	11 1546 16	3 996 21	3 396 29
Stn1016	114 121 28		56 247 32	3 609 30	1 1097 16	3 2193 16	11 330 25
Stn1018	66 362 31	119 237 33		25 574 23	18 788 19	8 1534 11	16 836 29
Stn1021	6 1380 21	7 1225 22	23 388 30		28 239 37	1 1147 19	1 878 34
Stn1024	16 1337 18	3 1836 12	17 891 18	35 419 30		10 517 27	3 1695 15
Stn1026	2 1266 14	5 1082 26	22 466 32	3 1456 21	14 1262 14		4 729 48
Stn1029	6 525 28	11 342 34	7 565 34	1 908 32	1 540 41	6 1187 24	

Table 7.3: O-D matrix showing flow, journey times (JT) and speed in the Wigan network

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Sun	1	0.99	0.98	0.97	0.97	0.98	0.99
Mon	0.99	1	0.99	0.99	0.99	0.98	0.98
Tue	0.98	0.99	1	0.99	0.99	0.99	0.98
Wed	0.97	0.99	0.99	1	0.99	0.99	0.96
Thu	0.97	0.99	0.99	0.99	1	0.98	0.97
Fri	0.98	0.98	0.99	0.99	0.98	1	0.97
Sat	0.99	0.98	0.98	0.96	0.97	0.97	1

Table 7.4: Correlation analysis between the weekday O-D matrices in Wigan

7.5.3 Using origin-destination matrix to understand the impact of traffic

Table 7.5 presents the analysis of the traffic impacts across GMN to understand the location that is most affected as part of the potential application of Bluetooth. Across the three locations, the amount of time spent in traffic based on the Bluetooth information was analysed per unit kilometre to normalise the data. The Stockport link is the shortest (3.37 km) while Trafford and Wigan are 5.24 km and 5.83 km respectively. From the sample analysed, Stockport had the highest number of vehicles in both directions (1147 and 1189 vehicles), followed by Trafford (209 and 258 vehicles) and Wigan (144 and 198 vehicles) from which average speed and time were computed over the period under consideration. Given that Stockport links are the shortest and have the highest sample (number of vehicles) over the same period suggests that Stockport possesses the highest number of vehicles incorporating Bluetooth devices. However, this is not necessarily so given the fact that over short links, contributions arising from connecting routes, whether by a way of reducing or increasing the volume on the main link, is minimised compared to long links. For example, the arterial network of Wigan is expected to have the lowest match rate over a long distance compared to the other two networks in Stockport and Trafford. Besides, one of the interesting features captured by the analysis is that more time per kilometre (4.942 and 5.575 for inbound and outbound flows respectively) corresponding to the lowest speed/km is spent in the conurbation (Wigan) than on the routes within Trafford and Stockport. The Stockport links had the least time spent per kilometre (2.746 and 3.268 for inbound and outbound flows respectively). Consequently, it may be inferred that within GMN, given the same factors such as vehicle composition, weather and period, spatially, more fuel will be burnt in Wigan which is therefore more susceptible to pollution, while in Stockport, less fuel will be used thereby saving cost with less pollution. Using Bluetooth to enhance this understanding and other useful applications in traffic management is considered viable with the obvious advantage of low-cost compared to the traditional methods of traffic data collection. The next discussion presents the conclusions drawn.

Location	Link	Distance (km)	Daily Average Speed (km/h)	Daily Average Time (min)	Normalised Speed (Speed/km)	Normalised Time (Time/km)	Sample
Trafford	1001-1008	5.24	21	22.24	4.005	4.241	209
Stockport	1033-1041	3.37	25	11.01	7.421	3.268	1147
Wigan	1012-1021	5.83	16	32.50	2.744	5.575	144
Trafford	1008-1001	5.24	23	20.14	4.386	3.841	258
Stockport	1041-1033	3.37	27	9.25	8.04	2.746	1189
Wigan	1021-1012	5.83	18	28.81	3.087	4.942	198

Table 7.5: Analysis of traffic impacts across GMN for a typical weekday

7.6 Conclusions

In Chapter 7, the discussion and interpretation of Bluetooth estimated traffic metrics comprising flow, journey time, speed and O-D matrix was presented. The results and interpretations covered the relevance of Bluetooth technology and how it provided the knowledge of potential traffic management applications. Potential applications include the use of journey time and speed metrics to reconstruct typical traffic regimes to understand temporal variations arising from peak and off-peak periods, and analysis of traffic impact. For example, over the GMN, Stockport links are the most efficient with the least time spent per kilometre (2.746 and 3.268 minutes for inbound and outbound flows respectively) while Wigan is least efficient due to the effect of conurbation. Spatially and temporally, the consistency observed in the data provides the opportunity to build historical data, and thus the possibility for data augmentation. The realisation of O-D information using Bluetooth is justified by the high level of temporal consistency, which signifies reliability. Using Bluetooth in this way presents an added advantage in terms of both cost and time of data acquisition as well as safety benefits. The significantly low-cost of acquisition, installation and maintenance of Bluetooth sensors compared to the traditional systems of data collection presents another added advantage to densify the road networks for an area-wide coverage. This will of course bring about timely response to incident management as an incident can be localised to the exact scene with precision.

However, the limitations observed in Bluetooth data such as the low count rate, and challenges in accurately differentiating between carriers of a Bluetooth-enabled device during congestion means that it cannot be used as a stand-alone system in all applications. Similarly, the estimation of the actual traffic flow is dependent upon calibration against an independent measure of traffic to determine the scaling factor, which is obtained from the detection rate. In a network of similar attributes, estimation of flow based on combined directional flows is preferable to the link-based (directional) estimation on the grounds of accuracy and cost. However, the link-based estimation presents a better reconstruction of the traffic states and level of service in each direction. Data filtering is required to obtain the proportion of the vehicle captured by the Bluetooth sensors, and estimation based on all the detected devices does not provide the actual traffic state. Based on the available evidence, in particular, from a typical network within GMN, Bluetooth has a number of viable applications in traffic management. The next chapter concludes the discussion on the potential applicability and viability of Bluetooth in a wider context to round up Research Objective iv.

Chapter 8. Bluetooth Traffic Monitoring in the Context of Applicability

8.1 Introduction

Chapter 8 considers the key policy and technological implications emanating from the research conducted using Bluetooth for traffic sensing and metrics estimation for traffic management applications. This discussion concludes Research Objective iv, and sets the platform for possible recommendations from the research. This chapter also considers issues relating to public acceptance, and the economic benefits offered by the technology over other possible alternatives. That is, it considers key issues relating to the reliability of Bluetooth in traffic sensing as well as the applications and benefits it could deliver both in the present and future. Exploring the applications in this way will help traffic engineers and ITS managers as well as policy makers understand how the technology could potentially improve traffic management. That is, at a glance, how technological improvements through the use of Bluetooth can lead to an enhanced solution in traffic management and can be understood. The potential of this technology in traffic management includes an optimised road network and improved safety, reduction in pollution and fuel consumption through reduced traffic congestion. Using Bluetooth, twelve potential applications are presented with their benefits to inform usability. The evaluation criteria considered in this research include a consideration for cost, accuracy and precision, and temporal and spatial consistency of the data. Exploring Bluetooth in this way is in agreement with the recommendation to use pricing and technological measures as solutions to traffic congestion (Mitchell *et al.*, 2011). Therefore, knowledge of the Bluetooth approach in a wider context of traffic management might form the foundation for viable alternatives and essential policy formulation.

Chapter 8 comprises the following key sections; Section 8.2 completes the discussion on the applicability of Bluetooth technology in traffic management while the transferability of the Bluetooth approach is presented in Section 8.3.

Section 8.4 examines the theoretical implications while Section 8.5 considers the policy implication. The concept of 'Bluetooth Economic 4-Way Test' is presented in Section 8.6 to underpin the cost-accuracy benefits of the technology, before conclusions are drawn in Section 8.7.

8.2 Applicability of Bluetooth in Traffic Management

Table 8.1 presents the summary of the potential applications of Bluetooth and their benefits in traffic management. UTMC, as we know it today, can be made to respond better to the management of road traffic if the opportunities offered by technologies such as Bluetooth are well-harnessed (Ayodele *et al.*, 2014). From the evidence in this research and literature, exploring this option will lead to significant potential benefits. The derivable benefits include low procurement and operational cost, potential to support a reduction in traffic delay and improved road safety. Derivation of traffic metrics such as journey time and speed through the detection of Bluetooth-enabled devices carried onboard vehicles, and of other modes of transportation is a possibility. Potentially, the efficiency of the signal control models such as SCOOT can be improved upon through the use of, for example, hour by hour O-D matrices provided by Bluetooth instead of reliance on the traditional fixed simulation periods of typical daily peak and off-peak (Ayodele *et al.*, 2014). Journey times and vehicle speeds from Bluetooth can contribute to performance measures required to determine the effectiveness of the road network.

Traffic metrics such as O-D matrix and density that have been difficult and expensive to acquire in the past can now be obtained in a fast and cost-effective manner compared to the traditional systems (Barceló *et al.*, 2013; Bhaskar *et al.*, 2014). The possibility of computing the penetration rate presents the opportunity for scalability and transferability of Bluetooth estimated flow over other links of similar traffic characteristics. In addition, the possibility of real-time communication will contribute to road safety such as in collision avoidance, particularly on sharp bends and at road junctions. It is anticipated that Bluetooth sensors may take over some of the functionalities of the current systems in both infrastructures and vehicles. Possible transportation applications besides traffic management will also emerge which include wireless tyre pressure monitoring, keyless entry, and the emergence of an ecosystem, where the head unit consists of Bluetooth sensors instead of a combination of different wireless technologies (Kuchinskas, 2013). Bluetooth also has more potential in electric vehicles (EVs) such as in reduced weight through a reduction in the wiring systems (Kuchinskas, 2013). The positioning applications and telemetry services can now be achieved with efficiency using Bluetooth (Gakstatter, 2014). Therefore, using Bluetooth technology to support traffic management applications is recommended.

S/N	Application	Traffic Metric	Benefit
1	Link-flow estimation for congestion control	Link-Flow	Cost benefit, improved traffic prediction, optimised road through congestion management
2	Data augmentation	Link-Flow/Journey time/Speed	Improved accuracy, avoidance of network failure and better reliability
3	Temporal and spatial status network monitoring	O-D matrix/Journey time/Speed	Enhanced traffic management leading to safety, cost and health benefits
4	Support for network optimisation	O-D matrix/Journey time/Speed	Enhanced traffic management leading to safety, cost and health benefits, optimised road network
5	Traffic impact analysis	O-D matrix	Health and cost benefits as well as social and psychological benefits
6	Incident detection	Journey time/speed	Enhanced traffic management through rapid response to emergency situations
7	Dwell time analysis	Journey time	Cost and safety benefits, enhanced fleet management and vehicle monitoring
8	Travel time index study	Journey time	Cost benefit, variability index and congestion management for an optimised road
9	Speed limit compliant level monitoring	Journey speed	Safety benefit
10	Level of service analysis	Flow/Speed	Enhanced traffic management
11	Density estimation	Flow/Speed	Enhanced traffic management
12	Decision support system	O-D matrix/Journey time/Speed	Enhanced traffic management

Table 8.1: Bluetooth potential traffic management applications and benefits

8.3 Transferability of Bluetooth Traffic Monitoring Method

One of the factors to consider in the choice of any system is transferability (Srinivasan, 2011). Although there are different vendors of Bluetooth sensors such as TDC Systems Ltd, BlipTrack, and Blids, the approach for traffic monitoring remains the same. Few exemptions can be seen in performance such as in data acquisition and transmission. Besides, the basic operational principle remains the same irrespective of vendor or geographical location. For example, the method described in the literature was built upon in this thesis, and was applied to the data collected from different urban areas in the UK. However, it should be noted that network configuration and attributes play an important role in the results obtained. For example, sensors installed at roundabouts will detect more vehicles going in different directions than those at T-junctions. Consequently, for vehicular traffic detection, care must be taken to distinguish between different road users. Otherwise, the method for the generic network traffic is not transferable for a vehicular detection. Care must also be taken when transferring the method over networks of different attributes such as in urban or rural networks to avoid over/under-estimation. Nevertheless, the investigation conducted over the different geographical locations and the evidence from the literature showed that Bluetooth technology application for ITS purposes is transferable on temporal and spatial dimensions. Another advantage is that it is not difficult to move a Bluetooth detector from one location to another (UMCATT, 2008). However, as with any equipment, there is the requirement to calibrate the sensor over the new location to determine the detection rate to be used as the scaling factor to obtain the actual traffic flow. Overall, the transferability of the Bluetooth approach presents a significant economic benefit to support transport sustainability.

8.4 Theoretical Implications of the Research

The scope of this study within the UK means that more case studies and, in particular, real-time application will be required for further assessment and generalisation on this subject. That is, the results obtained are considered valid based on the data used in this research. This means that a new set of data may produce different results particularly if there is a significant rise in Bluetooth usage in the next few years. This further means that the use of Bluetooth for traffic estimation will require periodic calibration to account for any changes in usage. However, the major challenge is in determining how often the calibration will be required to ensure continuously accurate and reliable estimations. Not only that, the transferability of the technology is another important factor to be critically examined. While the methodology and the processing techniques may be the same and transferable, there is a need to consider the differences in the road networks where the Bluetooth sensors are deployed. For example, the results obtained in a less urbanised area may not be transferable to a more urbanised city due to the increase in the traffic volume in the new location and vice versa. Similarly, traffic estimation within the city centres or in congested networks will require a more robust validation to account for uncertainties arising from the contributions from other road users such as pedestrians than in a free-flowing network such as the motorway. For example, video recording may be required to obtain the disaggregation of traffic to accurately classify different modes to remove uncertainty. As with any technology and a direct consequence of the methodology, this research encountered some limitations, which need to be considered. This includes:

- Low count rate
- Heterogeneous data sources leading to difficulty in differentiating traffic modes during congestion
- The requirement for high-speed processing platform to handle the timely processing of the high-resolution data.

Irrespective of the challenges, the outcomes of this research, which spanned quality assessment to a demonstration of transferability, and proof of concept showed that the Bluetooth approach to traffic solutions is a viable proposition. The accuracy and reliability of the results obtained suggests the possibility of using Bluetooth data to inform policies that will help to optimise road transport planning and management.

The current literature on studies conducted outside the UK suggests the practicality of the Bluetooth approach to traffic monitoring and management. This concept of Bluetooth-based traffic monitoring and metrics estimation was analysed further and proven viable at 95% confidence through the validation exercise. This research demonstrated the possibility of Bluetooth application in temporal and status monitoring through the use of flows, journey times, vehicle speeds, and O-D matrix to support an optimised road network. The design, development and implementation of a model termed TRAFOST aided the resulting contribution to knowledge on the Bluetooth concept. In the future, research on Bluetooth applications in ITS can benefit from the use of TRAFOST to improve the understanding of Bluetooth approach. Currently, Bluetooth data processing algorithms are custom-based and are not available to the public.

The application of a novel and a low-cost wireless sensor such as Bluetooth to enhance the management systems to address congestion problems within the road transport network constitutes ground-breaking and cutting-edge research. The obvious benefits in terms of optimised road network are improved safety and reduced travel time leading to a reduction in pollution and the amount of fuel consumed, thereby saving cost. The provision of timely and accurate data that have been difficult and expensive to acquire in the past addresses the problem of data availability in transport modelling. Bluetooth technology has the potential for real-time applications and can account for a network of varying characteristics to provide traffic data. The required number of sensors to be deployed depends on the nature of the network and the purpose of the data collection. For instance, an O-D survey will require more sensors that are well-distributed at strategic nodes than a link-based study. Similarly, a complex urban area will require more sensors than a free-flowing motorway. Bluetooth technology is an emerging solution in ITS and related transport applications. Currently the only available publications which have been identified are from studies conducted outside the UK. Therefore, an investigation into the reliability of the applicability of Bluetooth data to address road congestion at UK study site areas, constitutes a significant contribution. That is, the enhancement of the knowledge of the applicability and viability of Bluetooth data as a novel solution to traffic congestion. This contribution to knowledge also includes the understanding of the variability in Bluetooth-derived metrics to enable sound inference and avoidance of uncertainty and unreliability in journey time predictions. Currently, there is limited information on this.

The advances in in-vehicle technologies make this research more compelling. This PhD research conducted within the UK has generated a fundamental understanding of spatial and temporal variations within the GMN more than is possible using traditional systems. Not only that, this research has informed the knowledge of the quality, limitations and usability of Bluetooth data through exploratory and quantitative analysis techniques to realise efficient and smarter decision support systems. In addition, to demonstrating the modelling and forecasting capability of the data using seasonal ARIMA models, the knowledge of the detection rate required to obtain the actual traffic flow is also enhanced. This knowledge thus provides the platform and justification for further research on the use of Bluetooth for transport applications. Clearly, the outcome of this research will undoubtedly put the City of Greater Manchester and the UK in general at the forefront of utilising low-cost and innovative technologies to enhance the road network through a better management of the increasingly congested roads.

8.5 Policy Implications of the Research

One particular policy issue relates to how technological-based solutions can be embraced to establish a balance in the road network through smart management without compromising the privacy of the road users. Such policies can be seen in the objective of ITS-UK and the Foresight projects. However, public awareness of the benefits of Bluetooth technology will be essential in the process. This awareness will help to remove concern for covert monitoring from the public. The fact that carriers of Bluetooth-enabled devices have full control on the discoverability and connectivity is to be stressed. The use of encrypted data coupled with an unnoticed process of detection that constitutes no interference is another added advantage. The empirical findings based on vehicular traffic in this research showed that Bluetooth application is a possibility in ITS to realise smarter and efficient solutions to congestion management. Therefore, policy formulation such as road speed limit using Bluetooth as a complementary monitoring system will be beneficial. In this way, Bluetooth can contribute to meeting the ITS objective of a safe and efficient network. Moreover, the objective of the Climate Change Act can be achieved through a better-managed network leading to all-inclusive benefits. The next section considers the assessment of Bluetooth on a cost-accuracy scale to understand their implications.

8.6 The Economic 4-Way Test of Bluetooth Application

Addressing the problem of traffic congestion from a technology perspective requires exploring different alternatives. From a transport and sustainability perspective, one of the different alternatives is Bluetooth. This is justified in the 'Bluetooth Economic 4-Way Test' presented in this section. This concept stands on the principle of 'economy of accuracy', which simply means maintaining a balance between the standard of accuracy aimed at, and the needs of the particular task (Whyte and Paul, 1997). The general rule of thumb is that the higher the standards of accuracy required, the higher the cost in terms of both time and money (Whyte and Paul, 1997). Figure 8.1 shows the concept of the Bluetooth Economic 4-Way Test consisting of four quadrants segmented based on cost and accuracy. Evidence from this research and literature showed that Bluetooth falls in the upper left quadrant of low-cost and high-accuracy, considered as the 'green zone'. Also, despite some of the limitations of Bluetooth, it could be used in some cases to characterise the road network more than is possible using the traditional systems This highlights the smart benefits that could be derived using Bluetooth data for traffic monitoring and management as well as other related transport applications. Bluetooth is a low-cost smart solution and is cheaper than the traditional systems both in terms of cost of acquisition, installation and maintenance, and is recommended.

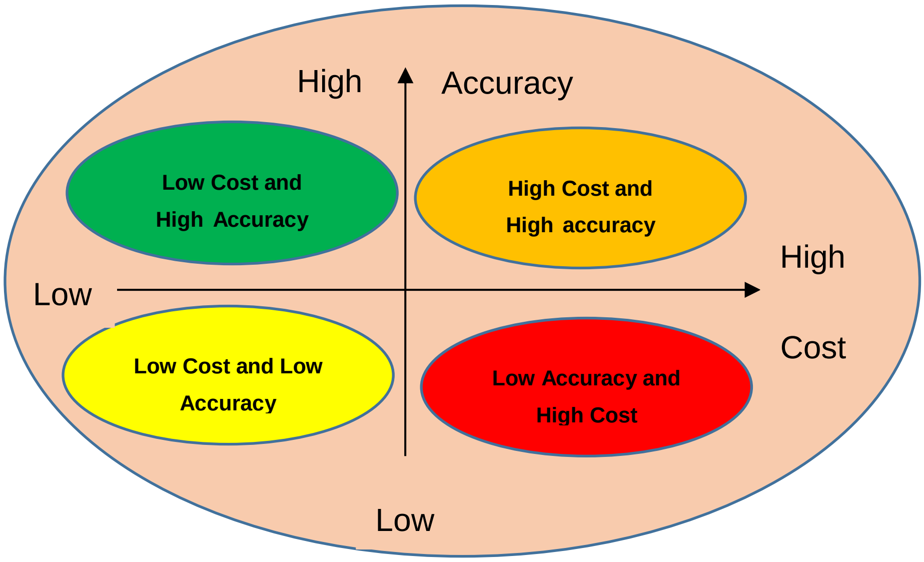


Figure 8.1: The concept of Bluetooth Economic 4-Way Test

8.7 Conclusions

In this chapter, key policy and technological issues relating to Bluetooth for traffic metrics estimation in a wider context of traffic monitoring and management are presented. Using the Bluetooth approach, twelve potential applications and their benefits were presented to inform usability. Generally, Bluetooth presents a smarter solution than is currently possible with the traditional systems both in terms of deployment and cost. The use of Bluetooth for traffic management will contribute to improved mobility, safety, efficiency, reliable journey time management and economic benefits. Bluetooth application will equally contribute to a reduction in waste and pollution through enhanced performance in traffic management systems. Therefore, the applicability of Bluetooth technology will support the establishment of a balanced network.

In a connected environment, Bluetooth could help improve the accuracy and the reliability of the monitoring sensors through data fusion and augmentation in a smart way. In fact, the dividends are all encompassing. Hence, this discussion is by no means exhaustive; meaning a need for future research on this subject. Over time, new applications such as automatic vehicle identification, toll collection, and distress alert, etc., will almost certainly continue to appear due to the novelty of the technology in the domain of ITS. The deductions made on the research findings in the wider context of applicability present a broad knowledge of the potential applications of Bluetooth technology in traffic management. The conclusion is that the Bluetooth approach, irrespective of any limitations, presents an innovative means that changes the way traffic information can be collected. The next chapter presents the conclusions and the thesis summary as well as recommendations for further research on the use of Bluetooth in ITS and related transport applications.

Chapter 9. Conclusions and Recommendations for Future Research

9.1 Introduction

This research has explored the concept of Bluetooth-based traffic monitoring and metrics estimation as an effective, smart and low-cost means to enhance traffic management systems to mitigate road congestion. The study found within the UK study sites, the nature, limitations and characteristics of Bluetooth-derived traffic metrics; the correlation with other independently measured traffic data (IMTD); the variability in the estimated metrics, and the usability of the traffic metrics in traffic management. This research has assessed the potential applications of the Bluetooth approach to traffic management in a wider context of traffic sensing and metrics estimation as well as whether the technology can enhance the traditional systems as a low-cost sensor. The need for low-cost consideration is to establish a balance in the road networks through innovative thinking – such as the use of novel and emerging technologies as viable alternatives or to complement the existing systems. This enormous potential makes research into the use of Bluetooth in ITS of high relevance, particularly with the UK being one of the leaders in ITS with a focus on 'better transport through technology'. This research sought an answer to the question: *is Bluetooth data reliable and of sufficient accuracy to estimate traffic metrics for traffic management applications to reduce congestion?* It found within the UK study sites that Bluetooth data is reliable, sufficiently accurate and low-cost for traffic management applications.

In the remaining three sections of this chapter, Section 9.2 presents the findings from the key chapters. Section 9.3 presents the recommendations for future research before the overall conclusion in Section 9.4.

9.2 Findings from the Key Chapters based on the Objectives of the Thesis

A number of research objectives were outlined at the outset of this research; this section summarises how each of these objectives were addressed.

Research Objective i: To carry out a comprehensive and critical review of the literature on the application of Bluetooth technology in traffic management, and to consider other technological options in road traffic monitoring. Chapter 2 critically addressed Research Objective i. The literature review showed that the knowledge of the reliability and validity of Bluetooth traffic sensing and metrics estimation for traffic management remains largely unknown due to the novelty of the technology in the area of ITS. The early research was conducted on journey time management and O-D estimation both on arterials and motorways. The positive outcome of the early research regarding the applicability of the technology provided the motivation for continued research towards the realisation of the ITS objective of a safe and efficient road network. This chapter of the thesis explored the gaps in methodology, usability and limitations in the Bluetooth approach to traffic sensing and metrics estimation with a focus on the reliability and validity of the solution for various road transportation applications. Exploring Bluetooth in this way will contribute to knowledge in realising the potential of the technology in ITS and related applications.

Research Objective ii: To design and develop a Bluetooth-based data processing procedure (a model) to derive origin-destination matrix, link-flow, journey time and speed in the chosen study areas. Chapter 3 addressed Research Objective ii and presented the description of the research design, methods of Bluetooth data cleansing, estimation and the validation methods of the traffic metrics. A Bluetooth-based traffic detection and estimation model termed TRAFOST (Traffic Flow Origin-destination Speed and Travel-time) was developed to accomplish the data processing. The model was developed based on R-programing language to estimate traffic metrics following the earlier Excel and Fortran models. Relevant assumptions were made such as in establishing the boundaries for the outlying values in the development of the model. The model’s significance is in the acceleration of the data processing and the reproducibility of the estimated traffic metrics (link-flow, origin-destination matrix, journey times and vehicle speeds). The final research design incorporating diverse independent measures of traffic for results validation ensured that a sound and robust investigation was carried out. This design involving the use of diverse IMTD provided an unbiased interpretation of the results and consequently increased reliability. The robust procedure thus removed any bias that might surround the analysis if it were only Bluetooth-validated (i.e., using the base data for validation). The processing and the analysis procedures, as well as the TRAFOST described in this thesis, contributed to knowledge of the methodology on the use of Bluetooth data. The procedure described here can be used in the future research on Bluetooth by other researchers with an interest in Bluetooth study. The validity of the model outputs given a 95% confidence level means the research assumptions are valid. However, the results should be applied within a limited range of validity given the prevailing conditions.

Research Objective iii: To apply the model in targeted pilot studies in selected study sites consisting of Liverpool, Birtley and Manchester, for an overview of the potential of Bluetooth-derived traffic metrics. Chapter 4 addressed Research Objective iii by exploring the potential of Bluetooth data to support the delivery of a smarter and more efficient transport network. The preliminary data quality assessment in the Liverpool study provided the motivation for continual investigation on the use of Bluetooth data to estimate traffic metrics. The Birtley study, on the other hand, served as an evaluation platform to test the research methodology for both limitations and strengths. The Manchester study implemented the research methods in an area-wide context to demonstrate the transferability of the methods, taking into account the limitations discovered in the earlier study. The demonstration of the credibility of Bluetooth data was in the form of consistency of the repeated measurements with the correlation coefficient $(r > 0.80)$ between weekdays' observations. The time series plots of the preliminary results showed similarity in the periodic trend signifying consistency over time. The time series plots also showed clear evidence of typical traffic patterns associated with temporal variations, having morning peak hours (7-9 am) and evening peak hours (4-6 pm). The outcome of the data collection over the study site provided the platform and the justification for the long-term study and validation of the estimated metrics using diverse methods. The key findings are:

- Each Bluetooth sensor provides the records of all the detected devices passing through the location (site) irrespective of direction. This shows the possibility of understanding the level of service each way at a given period without the need to install additional Bluetooth sensors to monitor the opposing link, thereby saving cost.
- Analysis of two weeks' worth of data collected in the study area of Liverpool and Birtley showed that Bluetooth has the potential to provide traffic flows and journey time, and can be used to understand journey patterns.
- Correlation analyses showed a very strong positive correlation $(r \geq 0.90)$ between weekdays and weekend observations. While the descriptive statistics also showed a high level of consistency in terms of both spread and distribution of the data which suggests reliability in the data.
- This reliability can also be observed in the form of spatial variability reflecting the volume of traffic across the networks.
- The high-resolution data (one-second) provided by Bluetooth presents the opportunity to estimate traffic metrics to support up-to-date traffic information without reliance on archive data.

Research Objective iv: To examine the performance of the model (TRAFOST) developed in Objective ii and the consistency of Bluetooth-derived traffic metrics for accuracy and reliability through validation against diverse independent measures of traffic and modelling. Chapter 5 addressed Research Objective iv using diverse IMTD (Independently Measured Traffic Data). The use of TRAFOST in this research facilitated the data processing and analysis by combining automation, repeatability and efficiency. delivery This advantage, in turn, culminated into an in-depth knowledge of the traffic flow patterns and spatio-temporal variations within the study sites. The development of TRAFOST proved to be of significant advantage in terms of both the credibility and reliability of the estimated metrics, as well as in speed and reproducibility. The comparisons of Bluetooth data against the IMTD as well as the modelling of the data showed a strong relationship and an accuracy level up to 95% given the MAPE values (0.822 – 4.917). In addition, Kullback-Leibler divergence analysis with values 0.004 – 0.044, showed a very good match between the data sets. Bluetooth/SCOOT presented a better correlation than Bluetooth/ATC. However, the difference cannot be attributed to technological differences but to their spatial positioning. SCOOT links are positioned upstream and downstream of the link while ATCs are in-between the link. The data from ANPR and TM are not co-located; therefore, were not compared against each other. Individually, the two data sets showed a strong relationship with Bluetooth-derived journey time and speed ($R^2 > 0.70$). Detection rates required to calibrate the estimated flows were computed (from the ratio of the flows or slope of the regression equation) between 7-15% for ATC, 13-16% for SCOOT and 12% for ANPR. Scaling up this rate over the network showed that estimations are best at the validation link and degrade further away with changes in the network characteristics, thus informing the knowledge of usability. That is, the range (8%) of the detection rate obtained in GMN means that spatial variation must be taken into consideration when generalising the results. Combining the results from ATC, SCOOT, TM and ANPR with Bluetooth has led to an increased understanding and conviction of the potential of Bluetooth data for traffic metrics estimation. Generally, the accuracy statistics from the ARIMA models all portend a high level of reliability and validity of the estimation.

Research Objective v: To analyse the variability in Bluetooth-derived traffic metrics to enable concrete deductions and sound inference based on the analysis of year 2013 data from the Greater Manchester Network (GMN). Chapter 6 addressed Research Objective v as a way of further validation. Overall, the results showed that Bluetooth can capture the temporal and spatial dynamics in the traffic network. The aggregation on a weekday basis presented the best consistency and accuracy while the monthly average presented the highest variability. This type of information is very significant to the practical applications of Bluetooth data to avoid unpredictability. The knowledge of the data distribution informed the statistical method applied. Generally, and consistent with SCOOT and ATC flows, the Bluetooth flows are not normally distributed while journey time and speed are best represented with a normal distribution. Higher variability was observed in the directional flows (coefficient of variation = 6.74 and 8.63 for NE and SW flows respectively) compared to the total directional flows (coefficient of variation = 5.02), signifying a better result and a higher reliance level in the total directional flows compared to the directional flows. Using Bluetooth, particularly in a network of similar traffic characteristics, total directional flow estimation may be preferred to directional-based estimation according to this information. Similarly, variability was more pronounced over the congested period than in free-flow, thus informing the knowledge of the period of better reliability. Higher variability was observed in ATC-derived detection rates than in the SCOOT-derived. An mpv (most probable value) of 13% for ATC-derived penetration rate was obtained in Trafford, based on monthly and daily directional flows. The day-to-day analysis of the detection rates on a long-term basis showed a high level of precision with a standard deviation of 0.01. This value is considered the representative proportion of the total vehicles detected by Bluetooth sensors in the Trafford network. Spatially, Stockport presented the highest variability with a cv (coefficient of variation) of 0.14 – 0.20. Post-analysis tests showed that hourly and periodic metrics can be grouped into different homogenous subsets to enhance traffic prediction. The variability study in general, provided the knowledge of essential factors that must be considered in the application of Bluetooth-derived metrics that include the averaging of the data and time of observation. It is noted that harnessing this information is critical to arriving at valid and sound conclusions from the results, and thus contributing to the reliability of the solution. The importance of variability can be seen in reliable journey time prediction and thus a removal of uncertainty in the mind of road users. That is, road users can effectively plan their routes and journeys without having to worry about unpredictability in journey time.

Research Objective vi: To interpret the results and make deductions on the research findings in a wider context of applicability and viability and make recommendations for traffic management. Chapter 7 partly addressed the interpretation element of Research Objective vi, while Chapter 8 summed up the applicability of the results. This chapter serves as the basis for informing the knowledge of the applicability and viability of the results. That is, the reliability of the Bluetooth approach to traffic monitoring as well as how the results obtained provided knowledge of the overarching research question. Generally, and consistent with the validation results, when the opposing links are of differing traffic characteristics, the link-based estimation presents a better reconstruction of the actual traffic compared to the total link-flow. However, if the level of service is similar, the total link-flow is preferable. The Bluetooth approach showed the possibility of answering questions relating to problem or incident identification in a network such as recurrent patterns and where a delay happens in a section of a road; this can be seen in the form of an unusual spike in a trend that calls for attention. The understanding of turning points and origin-destination (O-D) matrix of the network flows; and travel time and traffic regimes characterised by peak and off-peak periods are very important metrics to characterise and manage traffic for an optimised road network. From the results obtained, it is obvious that the requirement to provide accurate and reliable traffic information to support the delivery of enhanced traffic management can be met using Bluetooth data. The analysis of the Bluetooth-estimated metrics in this chapter enabled a deeper knowledge of the characteristics of Bluetooth data. This knowledge includes the various applicability and limitations of the metrics, the spatio-temporal variations, station and link-based estimation of traffic metrics, and reliability in different time averages (hourly, monthly, seasonal, etc.). In particular, it enabled the knowledge of a typical network within GMN (Greater Manchester Network) than is possible using any of the traditional systems.

Chapter 8 partly addressed the applicability and viability element of Research Objective vi to complement Chapter 7. This chapter presented the summary of the wider context of the applicability and viability of the results as well as the knowledge of what applications might emerge to sum up the aim of the research. Analysis of the Bluetooth economic 4-way test to understand Bluetooth scope in terms of cost and accuracy showed that the technology presents a means of collecting accurate traffic data at a low-cost. This is a major advantage considering the need for a sustainable transport network using low-cost technological options without investing heavily in a new infrastructure. The adoption of Bluetooth data for transportation applications means smart or innovative thinking (based on safety, economic and environmental benefits). The overriding benefits are accruable to both the road users, traffic engineers and other stakeholders. Overall, the Bluetooth approach presents an innovative means that changes the way traffic information can be collected. Bluetooth is considered a potential candidate in automated vehicle and the provision of big data for transport application. Overall, twelve different applications such as data fusion and augmentation, journey time management, and network planning and optimisation were presented with reference made to other possibilities. The conclusion drawn does not in any way assume a generalisation for the whole public, but a personal judgement based on the research outcomes and evidence provided.

9.3 Recommendations for Future Research

Bluetooth-based traffic monitoring is an emerging solution to congestion problems, and it is almost certain that different applications will continue to emerge. The scope of this research is within the study sites considered in the UK (Birtley, Liverpool and Manchester) using Bluetooth sensors developed by TDC Systems. To obtain a more generalisable result over different geographical locations and networks of differing traffic characteristics requires more study sites spread over the UK. This study has offered an assessment of the reliability and validity of Bluetooth traffic monitoring and metrics estimation as a contribution to a novel approach to traffic management. Overall, it is evident from this research that further investigation is needed to continue to exploit the potential of this near-ubiquitous technology. For example, the real-time application of the technology as a decision support system to enhance traffic management is a welcome idea and is strongly recommended.

The following are the recommendations for further research.

i. Real time/autonomic application of Bluetooth technology in an ITS environment such as the UTMC. This area was not investigated and tested in real-time. Therefore, future work should consider the real-time application of this technology to fully explore the opportunities offered by Bluetooth. This application, of course, will involve the knowledge of artificial intelligence, artificial neural network, data mining, Kalman filtering and particle filtering. The implementation of this application will undoubtedly require collaboration for research and development between the Research University and relevant stakeholders such as the Transport for Greater Manchester (TfGM) and TDC-Systems. This collaboration will provide a balance in resources and technical know-how.

ii. The model developed runs on a Windows platform that is limited in memory compared to platforms such as Linux. Consequently, a huge amount of data cannot be processed instantaneously. For a large scale and real-time deployment, parallel or cloud computing is recommended. Exploring the research in this way will enable an area-wide, timely and efficient solution. Also, exploring the current R applications such as “data.table” instead of “read.csv” and interactive online package such as “shiny” will be an advantage in the data processing and real-time analysis. In the future, these aspects of data import and analysis need to be explored to improve the efficiency of Bluetooth deployment in real-time applications.

iii. Using Bluetooth data to classify the network traffic based on the mode of transport is considered a research area for the future. This is another useful area to explore in the future to support a multi-modal transport system. In this way, better policy relating to the use of roads can be designed to accommodate all modes to ensure the safety of all road users.

iv. Vehicle and pedestrian tracking and monitoring: This area of research is recommended for future study given the need for security of lives and properties. Exploring Bluetooth in this context is considered paramount for the delivery of efficiency and safety in freight and allied services as well as pedestrian's safety.

9.4 Overall Conclusions on Bluetooth-Based Traffic Monitoring and Metrics Estimation

Increased levels of population and car use mean that the problem of traffic congestion will remain within the road networks. The negative impacts of traffic congestion cuts across both economic and health spheres. Different approaches to congestion management have been considered in the past. These include the use of technological solutions such as the traffic management systems, road-user charging and road expansion. However, capital investment on new infrastructures such as road construction and/or expansion as well as continued reliance on the traditional systems for traffic data collection and management are not sufficient to achieve smarter solutions. Bluetooth is a novel technology that can be integrated into ITS to achieve smarter solutions through the provision of accurate and real-time traffic data. Bluetooth is in a state of evolution in ITS. In this research, the understanding of the reliability and validity as well as the underlying factors that could affect the application of Bluetooth technology in traffic management is improved upon to demonstrate practicality. This research has demonstrated that Bluetooth traffic sensing and metrics estimation for the enhancement of traffic management systems to reduce road congestion is a viable proposition and is recommended.

References

Abbas, M., Rajasekhar, L., Gharat, A. and Dunning, J. P. (2013) 'Microscopic modeling of control delay at signalized intersections based on bluetooth data', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 17(2), pp. 110-122.

Abedi, N., Bhaskar, A. and Chung, E. (2014) 'Tracking spatio-temporal movement of human in terms of space utilization using Media-Access-Control address data', *Applied Geography*, 51, pp. 72-81.

Abrahamsson, T. (1998) *Estimation of Origin-Destination Matrices Using Traffic Counts - A Literature Survey* (IR-98-021/May). Laxenburg, Austria: International Institute for Applied Systems Analysis. [Online]. Available at: http://webarchive.iiasa.ac.at/Admin/PUB/Documents/IR-98-021.pdf.(Accessed: 18 September 2014).

Adedayo, O. A. (2006) *Understanding Statistics*. Akoka-Lagos: JAS Publishers.

Al-Khateeb, K. A. S., Johari, J. A. Y. and Al-Khateeb, W. F. (2008) 'Dynamic Traffic Light Sequence Algorithm Using RFID', *Journal of Computer Science, Science Publications* 4(7), pp. 517-524.

Alam, T. (2014) *Traffic volume studies*. Available at: http://www.slideshare.net/tanviralam31337/traffic-volume-studies (Accessed: 23 July 2016).

Allison, L. (2016) *Kullback Leibler Distance (KL)*. Available at: http://www.csse.monash.edu.au/~lloyd/tildeMML/KL/ (Accessed: 09 November 2016).

Andersson, F. and Karlsson, M. (2000) *Secure Jini Services in Ad Hoc Networks*. MSc thesis. Royal Institute of Technology (KTH) [Online]. Available at: http://www.e.kth.se/~e95_fan/exjobb/Thesis.pdf (Accessed: 27 April 2012).

Andrienko, N. and Andrienko, G. (2006) *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach* Germany: Springer.

APC (2011) *Inside NFC: how near field communication works*. Available at: http://apcmag.com/inside-nfc-how-near-field-communication-works.htm/ (Accessed: 25 July 2016).

Araghi, B. N. (2012) *Feasibility Study for application of Bluetooth Sensors for Traffic Measurement*. PhD Reseach (slides). cartogis.ugent.be.

Araghi, B. N., Christensen, L. T., Krishnan, R. and Lahrmann, H. (2012a) *Application of Bluetooth for Mode specific travel time estimation on arterial roads: potentials and challenges* [Computer program].

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

Araghi, B. N., Pedersen, K. S., Christensen, L. T., Krishnan, R. and Lahrmann, H. (2012b) 'Accuracy of travel time estimation using Bluetooth technology: Case study Limfjord Tunnel Aalborg', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

Aslam, J., Lim, S. and Pan, X. (2012) 'City-Scale Traffic Estimation from a Roving Sensor Network', *SenSys'12, November 6–9, 2012*. Toronto, ON, Canada.

Augustin, D. and Poppe, C. (2012) 'A Bluetooth-based traffic data collection system supporting international traffic management plans', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

AVL (2004) *Arizona Phase II Final Report: Statewide Radio Interoperability Needs Assessment* (31 October). Arizona. [Online]. Available at: http://en.wikipedia.org/wiki/Automatic_vehicle_location. (Accessed: 31 October 2011).

Ayodele, E. G., Bell, M. C., Galatioto, F. and Thorpe, N. (2013) 'A study of the potential use of Bluetooth sensor data in delivering sustainable transport networks', *45th Annual UTSG Conference*. Oxford, UK, 2-4 January, 2013.

Ayodele, E. G., Bell, M. C., Thorpe, N. and Galatioto, F. (2014) 'The Role of Bluetooth Technology in Autonomic Decision Support Systems', *46th Annual UTSG Conference*. Newcastle, UK, 6-8 January, 2014.

Bachmann, C., Roorda, M. J., Abdulhai, B. and Moshiri, B. (2013) 'Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 17(2), pp. 152-164.

Balcilar, M. (2007) *Forecast Accuracy*. Available at: http://www.emu.edu.tr/mbalcilar/teaching2008/econ604/notes/forecast_accuracy.pdf (Accessed: 10 November 2016).

Banister, D. (1995) *Transport and Urban Development*. London: Chapman & Hall.

Bar-Gera, H. (2007) 'Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel', *Transportation Research Part C: Emerging Technologies*, 15(6), pp. 380-391.

Barceló, J., Montero, L., Bullejos, M., Serch, O. and Carmona, C. (2011) 'A Kalman Filter Approach for the Estimation of Time Dependent OD matrices Exploiting Bluetooth Traffic Data Collection', *91st Transportation Research Board 2012 Annual Meeting*. Washington, DC, November 2011.

Barceló, J., Montero, L., Bullejos, M., Serch, O. and Carmona, C. (2012) 'Dynamic OD matrix Estimation Exploiting Bluetooth Data in Urban Networks', A: International Conference on Automatic Control, Modelling & Simulation, 'Proceedings of the International Conference'. Saint-Malo: 2012, p. 116-121.

Barceló, J., Montero, L., Bullejos, M., Serch, O. and Carmona, C. (2013) 'A kalman filter approach for exploiting bluetooth traffic data when estimating time-dependent od matrices', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 17(2), pp. 123-141.

Barceló, J., Montero, L., Marques, L. and Carmona, C. (2010) 'Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring', *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C.*, (No. 2175), pp. 19-27.

BBC (2011) *About Bluetooth*. Available at: http://www.bbc.co.uk/webwise/guides/about-bluetooth (Accessed: 24 October 2011).

Beca (2011) *Bluetooth Deployment - Puhoi to Warkworth Pilot Study*. New Zealand Transport Agency (NZTA). [Online]. Available at: http://www.bliptrack.com/uploads/media/NZI-4002843-Bluetooth_Deployment_-_Puhoi_to_Warkworth.pdf (Accessed: 17 January 2012).

Bell, M. C., Ayodele, E. G. and Galatioto, F. (2012) 'Creating an evaluation platform to deliver sustainable urban networks using Bluetooth technology', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

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

Bhaskar, A., Tsubota, T., Kieu, L. M. and Chung, E. (2014) 'Urban traffic state estimation: Fusing point and zone based data', *Transportation Research Part C: Emerging Technologies*, 48, pp. 120-142.

Biora, F., d'Aprile, F. and Marino, R. (2012) 'A large scale application for Bluetooth-based travel time measurement in the Netherlands', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

BITRE (2014) *New traffic data sources - An overview*. Sydney. [Online]. Available at: https://bitre.gov.au/events/2014/files/NewDataSources-BackgroundPaper-April%202014.pdf (Accessed: 9 February 2017).

BlipTrack (2012) 'Penetration Studies Version 1.0', *19th World ITS Congress* Vienna, Austria, 22-26 October 2012.

Blogg, M., Semler, C., Hingorani, M. and Troutbeck, R. (2010) 'Travel Time and Origin-Destination Data Collection using Bluetooth MAC Address Readers', *Australasian Transport Research Forum 2010 Proceedings 29 September – 1 October 2010* Canberra, Australia.

Bluetooth (2011) *Bluetooth*. Available at: http://www.bluetooth.com (Accessed: 24 October 2011).

Bluetooth (2012) *The latest news about Bluetooth products and technology*. Available at: http://www.bluetooth.com/Pages/News-Detail.aspx?ItemID=413. (Accessed: 24 April 2012).

Bluetooth SIG (2001) *Specification Volume 1. Specification of the Bluetooth System, Profiles. Version 1.1*.

Bluetooth SIG (2015) *Bluetooth technology and WiFi*. Available at: http://www.bluetooth.com/Pages/Wi-Fi.aspx (Accessed: 09 September 2015).

Bluetooth SIG (2016) *Our History*. Available at: https://www.bluetooth.com/about-us/our-history (Accessed: 23 February 2017).

Blythe, P. T. (2006) 'Experiments with Mobile Adhoc Networks and Smartdust Sensors to Deliver Pervasive Road User Charging, Wide-Scale Environmental Monitoring and other ITS Services', *ITS World Congress*. London, 13 October 2006.

Boethius, E. (2011) *European Activities towards deploying cooperative systems* Available at: http://www.drive-c2x.eu/tl_files/publications/DRIVE_C2X_ITS_Lyon_2011_01.pdf (Accessed: 01 September 2015).

Boneberg, L. K., Hancook, C. M. and Roberts, G. W. (2011) *The Journal of the Chartered Institution of Civil Enginering Surveyors*, (December/January).

Booth, S. (2005) 'Satnav to oust paper maps?', *Geomatics World*, 13(3).

Borresen, J. L., Jensen, C. S. and Torp, K. (2016) *Proceedings - IEEE International Conference on Mobile Data Management*. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84981748210&partnerID=40&md5=74e602e50c997af7a34e3f637c1234a8. (Accessed: 01 February 2017).

Brereton, R. G. (2015) 'The Mahalanobis distance and its relationship to principal component scores', *Journal of Chemometrics*, 29(3), pp. 143-145.

Brown, B. M. and Forsythe, B. A. (1974) 'Robust Tests for the Equality of Variance', *Journal of the American Statistical Association*, 69(346), pp. 364-367.

Brownlee, J. (2014) 'How To Estimate Model Accuracy in R Using The Caret Package', *R Machine Learning*. Available at: http://machinelearningmastery.com/how-to-estimate-model-accuracy-in-r-using-the-caret-package/. (Accessed: 30 July 2016).

Burnham, K. P. and Anderson, D. R. (2002) 'Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach'. New York: Springer, pp. 1-454.

Business Wire (2010) *Bluetooth Technology to Be Standard in More Than 90 Percent of Automobiles by 2016*. Available at: http://www.businesswire.com/news/home/20101118006520/en/Bluetooth-Technology-Standard-90-Percent-Automobiles-2016 (Accessed: 10 January 2017).

Camacho, T., Kostakos, V. and Mantero, C. (2010) *A Wireless Infrastructure for Delivering Contextual Services and Studying Transport Behavior*. Madeira Interactive Technologies Institute at the University of Madeira Portuguese Foundation for Science and Technology (FCT) (Accessed: 24 November 2011).

Camsys and Texas Transportation Institute (2004) *Traffic Congestion and Reliability: Linking Solutions to Problems* Cambridge, Massachusetts: Federal Highway Administration. [Online]. Available at:

http://www.ops.fhwa.dot.gov/congestion_report_04/congestion_report.pdf. (Accessed: 14 November 2016).

Chakraborty, G., Kshirasagar, N., Chakraborty, D., Shiratori, N. and Wei, D. (2010) 'Analysis of the Bluetooth device discovery protocol.', *Wireless Networks*, 16(2), pp. 421-436.

Chang, W. (2014) 'Interactive Graphics with ggvis', *R-bloggers.com*. Available at: http://www.r-bloggers.com/ (Accessed: 14 August 2014).

Chen, C., Skabardonis, A. and Varaiya, P. (2003) 'Travel-Time Reliability as a Measure of Service' *Transportation Research Record*. pp. 74-79. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-1942537014&partnerID=40&md5=9bfb461318649b9b2b9128c92235363c. (Accessed: 24 February 2015).

Chen, K. and Miles, J. C. (1999) *ITS Handbook 2000: Recommendations from the World Road Association (PIARC)*. Boston: Artech House.

Cheu, R. L., Xie, C. and Lee, D-H. (2002) 'Probe Vehicle Population and Sample Size for Arterial Speed Estimation', *Computer-Aided Civil and Infrastructure Engineering*, 17(1), pp. 53–60.

Chowdhury, M. A. and Sadek, A. (2003) *Fundamentals of Intelligent Transportation Systems Planning*. Boston: Artech House.

Click, S. and Lloyd, T. (2012) 'Applicability of bluetooth data collection methods for collecting traffic operations data on rural freeways', *Transportation Research Board 91st Annual Meeting*, Washington D.C, 22-26 January 2012.

Collotta, M. and Pau, G. (2015) 'New solutions based on wireless networks for dynamic traffic lights management: A comparison between IEEE 802.15.4 and bluetooth', *Transport and Telecommunication*, 16(3), pp. 224-236.

Conservation (2012) *Take simple steps to live green*. Available at: http://www.conservation.org/act/simplesteps.aspx (Accessed: 2 March 2012).

Cook, D. (2014) 'Not Drowning, Waving', *R-Blogger*. Available at: http://www.r-bloggers.com/ (Accessed: 14 August 2014).

Cooper, M. A. R. (1974) *Fundamentals of Survey Measurement and Analysis*. London: Granada.

Cowpertwait, P. P. S. and Metcalfe, A. V. (2009) *Introductory time series with R*. Springer Science + Business Media, LLC.

Cragg, S. (2013) *Bluetooth Detection - Cheap but challenging*. Available at: http://www.transportscotland.gov.uk/system/files/documents/tsc-basic-pages/Bluetooth_Report_-_S_Cragg.pdf (Accessed: 01 December 2014).

Crawley, M. J. (2005) *Statistics: An Introduction to R*. England: Wiley.

Csikos, A., Viharos, Z. J. and Kis, K. B. (2015) 'Traffic speed prediction method for urban network - An ANN aproach', *2015 Models and Technologies for Intelligent Transportation Systems (MIT-ITS)*. Budapest, Hungary, 3-5 June 2015. IEEE. Available at: http://www.traffic.bme.hu/data/uploads/Publication/MT-ITS_2015_IEEE_cikk_FINAL.pdf. (Accessed: 07 February 2017).

Dalgaard, P. (2002) *Statistics and Computing: Introductory Statistics with R*. United States of America: Springer.

Dalgleish, M. and Hoose, N. (2009) 'Intelligent Transport Systems', *TEC*

Darey, E. (2012) *Legislation and Compliance*. Available at: http://cutcarbon.info/legislation/?&gclid (Accessed: 02 March 2012).

Data Protection Commissioner (2003) *Data Protection Acts 1988 and 2003: A Guide to Your Rights*. Available at: http://www.dataprotection.ie/documents/guidance/A_Guide_to_Your_Rights_web_version.pdf (Accessed: 16 September 2014).

DataCamp (2014) 'Data analysis the data.table way: introducing DataCamp's newest course', Available at: *http://www.r-bloggers.com/* (Accessed: 8 October 2014).

Derekenaris, G., Garofalakis, J., Makris, C., Prentzas, J., Sioutas, S. and Tsakalidis, A. (2001) 'Integrating GIS, GPS and GSM technologies for the effective management of ambulances', *Computers, Environment and Urban Systems*, 25(3), pp. 267-278.

DfT (2011) *Transport Statistics Great Britain*. Available at: http://assets.dft.gov.uk/statistics/releases/transport-statistics-great-britain-2011/tsgb-2011-stats-release.pdf (Accessed: 31 May 2012).

DfT (2014) *Data Sources and Surveys*. London: Department for Transport. [Online]. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/427119/webtag-tag-unit-m1-2-data-sources-and-surveys.pdf. (Accessed: 01 February 2015).

Diebold, J. (1995) *Transportation Infostructures: The Development of Intelligent Transportation Systems*.

Dissanayake, D., Kurauchi, S., Morikawa, T. and Ohashi, S. (2012) 'Inter-regional and inter-temporal analysis of travel behaviour for Asian metropolitan cities: Case studies of Bangkok, Kuala Lumpur, Manila, and Nagoya', *Transport Policy*, 19(1), pp. 36-46.

Dixon, W. J. and Massey, F. J. (1983) *Introduction to Statistical Analysis*. Fourth edn. Auckland: McGraw-Hill Book Company.

Dobson, A. J. and Barnett, A. G. (2008) *An Introduction to Generalised Linear Models*. Third edn. Boca Raton: Taylor & Francis (Chapman & Hall).

DoT (2007) *Automatic Vehicle Location (AVL)/Rural Transit*. Available at: http://www.pcb.its.dot.gov/factsheets/avl/avlRur.pdf (Accessed: 31 October 2011).

DoT (2012) *Focus on Congestion Relief*. Available at: http://www.fhwa.dot.gov/congestion/describing_problem.htm (Accessed: 12 May 2012).

DoT (2015) *Why Speed Limits?* Available at: https://www.codot.gov/library/Brochures/Establishing_Realistic_Speed_Limits_Brochure.pdf (Accessed: 31 August 2015).

Eddington, S. R. (2006) *The Eddington transport study: the case for action: Sir Rod Eddington's advice to Government*. Available at: http://www.sortclearinghouse.info/research/343 (Accessed: 11 August 2012).

Edwards, D. and Hamson, M. (2001) *Guide to Mathematical Modelling*. New York: Palgrave.

Edwards, S. J. F., Evans, G., Blythe, P., Brennan, D. and Sevarajah, K. (2012) 'Wireless Technology Applications to Enhance Traveller Safety', *IET Intelligent Transport Systems.*

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

European Commission (2004) *3rd eSafety Forum*. [Online]. Available at: http://ec.europa.eu/information_society/activities/esafety/doc/esafety_communication/esafety_communication_vf_en.pdf (Accessed: 20 April 2012).

Faulkner, C. (2015) *What is NFC? Everything you need to know*. Available at: http://www.techradar.com/news/phone-and-communications/what-is-nfc-and-why-is-it-in-your-phone-948410 (Accessed: 02 May 2016).

Fernández-Lozano, J. J., Martín-Guzmán, M., Martín-Ávila, J. and García-Cerezo, A. (2015) 'A wireless sensor network for urban traffic characterization and trend monitoring', *Sensors (Switzerland)*, 15(10), pp. 26143-26169.

FHWA (2006a) 'CHAPTER 2. SENSOR TECHNOLOGY', in *Traffic Detector Handbook*. Third edn.

FHWA (2006b) *Freeway Management and Operations Handbook: Chapter 15 - Detection and Surveillance*. [Online]. Available at: http://ops.fhwa.dot.gov/freewaymgmt/publications/frwy_mgmt_handbook/chapter15_01.htm. (Accessed: 27 July 2016).

FHWA (2013) *Traffic Monitoring Guide*. U.S. Department of Transportation.

Filgueiras, J., Rossetti, R. J. F., Kokkinogenis, Z., Ferreira, M., Olaverri-Monreal, C., Paiva, M., Tavares, J. M. R. S. and Gabriel, J. (2014) 'Sensing bluetooth mobility data: Potentials and applications' *Advances in Intelligent Systems and Computing*. Springer Verlag, 262, pp. 419-432. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84921739840&partnerID=40&md5=06d6a61f907a366b3b64427b3ee1248f. (Accessed: 24 February 2015).

Filzmoser, P. and Varmuza, K. (2013) *Package Chemometrics*. Available at: http://cran.r-project.org/web/packages/chemometrics/index.html (Accessed: 24 May 2016).

Foresight (2006) *Intelligent Infrastructure Futures: Project Overview*. United Kingdom: Office of Science and Technology Crown.

Fox, J. (2005) 'The R Commander: A Basic Statistics Graphical User Interface to R', *Journal of Statistical Software*, 14(9), pp. 1--42.

Fox, J. and Weisberg, S. (2010) *Time-Series and Generalized Least Squares in R: An Appendix to An R Companion to Applied Regression,* Second Edition, Sage, Thousand Oaks, CA.

Fricker, J. D. and Kumapley, R. K. (2002) *UPDATING PROCEDURES TO ESTIMATE AND FORECAST VEHICLE-MILES TRAVELED* (FHWA/IN/JTRP-2002/10). West Lafayette, IN: Purdue University. [Online]. Available at: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1685&context=jtrp. (Accessed: 07 February 2017)

Friesen, M. R. and McLeod, R. D. (2014) 'Bluetooth in Intelligent Transportation Systems: A Survey', *International Journal of ITS Research*, 13(3), pp. 143–153.

Gakstatter, E. (2014) 'RTK on Your Smartphone or Tablet'. Available at: http://gpsworld.com/rtk-on-your-smartphone-or-tablet/ (Accessed: 15 July 2016)

Gastaldi, M., Gecchele, G. and Rossi, R. (2014) 'Estimation of Annual Average Daily Traffic from one week traffic counts. A combined ANN-Fuzzy approach', *Transportation Research Part C*, 47, pp. 86-99.

Gifford, J. L. (2003) *Flexible Urban Transportation*. Amsterdam: Elsevier.

Gomez, C., Oller, J. and Paradells, J. (2012) 'Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology', *Sensors*, 12, pp. 11734-11753.

Gray, T. (2007) *Bluetooth Market Continues Growth, But Rate is Slowing*. Available at: http://www.tmcnet.com/voip/ip-communications/articles/9173-bluetooth-market-continues-growth-but-rate-slowing.htm (Accessed: 03 August 2016).

Gurczik, G., Junghans, M. and Ruppe, S. (2012) 'Conceptual approach for determining penetration rates for dynamic indirect traffic detection based on Bluetooth', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

Gutiérrez, A. (2016) 'Computing Sample Size for Variance Estimation', *R-bloggers*.

Haghani, A. and Hamedi, M. (2013) 'Application of bluetooth technology in traffic detection, surveillance, and traffic management', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 17(2), pp. 107-109.

Haghani, A., Hamedi, M., Sadabadi, K. F., Young, S. and Tarnoff, P. (2010) 'Data collection of freeway travel time ground truth with Bluetooth sensors' *Transportation Research Record*. pp. 60-68. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-78651311179&partnerID=40&md5=e9aaab895cf4177022399ddfd80b3181. (Accessed: 24 February 2015).

Hainen, A., Wasson, J., Hubbard, S., Remias, S., Farnsworth, G. and Bullock, D. (2011) 'Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles', *Transportation Research Record: Journal of the Transportation Research Board*, 2256, pp. 43-50.

Hainen, A. M., Remias, S. M. and Bullock, D. M. (2013) 'Collection and analysis of multi-modal airport land side probe data from Bluetooth enabled mobile devices', *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. 6-9 October 2013. pp. 1304-1309.

Handheld (2012) *Bluetooth Group Adopts GNSS Standard*. Available at: http://www.gpsworld.com/consumer-oem/handheld/news/bluetooth-group-adopts-gnss-standard-12798?utm_source=GPS&utm_medium=email&utm_campaign=navigate_03_27_2012&utm_content=bluetooth-group-adopts-gnss-standard-12798 (Accessed: 28 March 2012).

Harris, R. (2014) 'Big Data: Challenges and opportunities', *ITS-UK Review*.

Haseman, R., Wasson, J. and Bullock, D. (2010) 'Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking', *Transportation Research Record: Journal of the Transportation Research Board*, 2169, pp. 40-53.

HCM (2000) *Highway Capacity Manual*. Transportation Research Board. [Online]. Available at: https://sjnavarro.files.wordpress.com/2008/08/highway_capacital_manual.pdf. (Accessed: 18 November 2016).

Heer, J. (2014) 'Data Cleaning is a critical part of the Data Science process', *New York Times,* 18 August 2014.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1985) *Exploring Data, Tables, Trends and Shapes*. USA: John Wiley & Sons.

Hodge, V. J. and Austin, J. (2004) 'A survey of outlier detection methodologies', *Artificial Intelligence Review*, 22(2), pp. 85-126.

Hounsell, N. B., Shrestha, B. P., Piao, J. and McDonald, M. (2009) 'Review of urban traffic management and the impacts of new vehicle technologies', *Institution of Engineering & Technology, Intelligent Transport Systems*, 3(4), pp. 419-428.

Houston TranStar (2010) 'Bluetooth speed and travel data collection shows cost savings', *ITS International*, July/August 2010.

Howstuffworks (2011) *Bluetooth*. Available at: http://electronics.howstuffworks.com/bluetooth1.htm (Accessed: 24 October 2011).

Hyndman, R. (2006) 'Another look at forecast-accuracy metrics for intermittent demand', *Foresight*, (4).

Hyndman, R.J. and Athanasopoulos, G. (2013) *Forecasting: principles and practice*. OTexts. Available at: http://otexts.org/fpp/ (Accessed: 28 November 2016).

IBM Corporation (2012) *GLM Repeated Measures*. Further info needed

Information Age (2001) *The Bluetooth Blues: Short-distance wireless communications technology is now reaching the consumer. Thousands of high-tech companies are banking on its success, but will they be disappointed?* Available at: http://www.information-age.com/article/2001/may/the_bluetooth_blues (Accessed: 24 October 2011).

Johnson, F.J. (1989) 'Traffic monitoring in Great Britain', in *Second International Conference on Road Traffic Monitoring*. London: IEE.

Jones, P. (2011) Bluetooth technology brief, 17 November 2011.

Jones, P. (2013) Bluetooth Data Capture and Transmission, 03 May 2013.

Kapur, J. N. (1988) *Mathematical Modelling*. New York: John Wiley & Sons.

Kardach, J. (2008) *Tech History: How Bluetooth got its name*. Available at: http://www.eetimes.com/electronics-news/4182202/Tech-History-How-Bluetooth-got-its-name (Accessed: 23 April 2012).

Kasten, O. and Langheinrich, M. (2001) 'First Experiences with Bluetooth in the Smart-Its Distributed Sensor Network'. Available at: http://www.vs.inf.ethz.ch/publ/papers/bt-experiences.pdf (Accessed: 05 March 2014)

Khoei, A. M., Bhaskar, A. and Chung, E. (2013) 'Travel time prediction on signalised urban arterials by applying SARIMA modelling on Bluetooth data', *36th Australasian Transport Research Forum (ATRF)*. Brisbane, Australia.

Kieu, L. M., Bhaskar, A. and Chung, E. (2012) 'Bus and car travel time on urban networks: Integrating Bluetooth and bus vehicle identification data', *25th ARRB Conference – Shaping the future: Linking policy, research and outcomes, 2012*. Perth, Australia. Available at: http://eprints.qut.edu.au/58733/1/LeMinhKieu_ARRB_2012.pdf (Accessed: 06 March 2017).

Kim, I. S., Jeong, K. and Jeong, J. K. (2001) 'Two Novel Radar Vehicle Detectors for the Replacement of a Conventional Loop Detector: Introduction to intrusive and non-intrusive radar vehicle detectors developed at 24 GHz', *Microwave Journal*, (3241).

Kindleysides, E. I. S. (2014) 'What are ITS?', *Intelligent Transport Systems Review*.

Kinney, P. (2003) 'ZigBee Technology: Wireless Control that simply works', *Communication Design Conference*. 2 October 2003.

Klein, L. A. (1997) *Vehicle Detector Technologies for Traffic Management Applications*. Available at: www.cotrip.org/.../Vehicle%20Detectors/Klein%20Part%202%20-%20Vehicle%20Detection (Accessed: 09 November 2016).

Klein, L. A., Mills, M. K. and Gibson, D. R. P. (2006) *Traffic Detector Handbook* (FHWA-HRT-06-139). Washington, DC: US Department of Transportation, Federal Highway Administration.

Kosta, M., Wiedersheimb, B., Dietzelc, S., Scaubb, F. and Bachmord, T. (2011) *WiSec 2011 Demo: PRECIOSA PeRA - Practical Enforcement of Privacy Policies in Intelligent Transport Systems*. Available at: www.uni-ulm.de/fileadmin/website.../2011-MC2R-perademo.pdf (Accessed: 10 December 2011).

Kostakos, V. (2008) 'Towards sustainable transport: wireless detection of passenger trips on public transport buses', *International Energy Outlook 2007*. Washington DC. United States Department of Energy. Available at: http:www.eia.doe.gov/oiaf/ieo/index.htmlk. (Accessed: 10 December 2011).

Krishnamoorthy, R. K. (2008) *Travel time estimation and forecasting on urban roads*. PhD thesis. Imperial College London [Online]. Available at: http://standard.cege.ucl.ac.uk/workshops/PDF_Files/KrishnanPhD.pdf. (Accessed: 08 November 2016).

Kuchinskas, S. (2013) *Bluetooth's smart future in telematics*. Available at: http://analysis.tu-auto.com/infotainment/bluetooths-smart-future-telematics (Accessed: 20 May 2015).

Kullback, S. and Leibler, R. A. (1951) 'On Information and Sufficiency', *The Annals of Mathematical Statistics*, 22(1), pp. 79-86.

Laharotte, P., Billot, R., Come, E., Oukhellou, L., Nantes, A. and El Faouzi, N. (2014) 'Spatiotemporal Analysis of Bluetooth Data: Application to a Large Urban Network', *IEEE Transactions on Intelligent Transportation Systems*.

Laharotte, P. A., Billot, R., Come, E., Oukhellou, L., Nantes, A. and El Faouzi, N.E. (2015) 'Spatiotemporal analysis of bluetooth data: Application to a large urban network', *IEEE Transactions on Intelligent Transportation Systems*, 16(3), pp. 1439-1448.

Langley, R. (2011) *Diggig GPS integrity*. Available at: http://www.gpsworld.com/gnss-system/innovation-diggig-gps-integrity-12254 (Accessed 10 December 2011).

Lastdrager, E. and Pras, A. (2009) 'Consistency of Network Traffic Repositories: An Overview', in Sadre, R. and Pras, A. (eds.) *AIMS 2009, LNCS 5637*. Springer, pp. 173 - 178.

Leduc, G. (2008) *Road Traffic Data: Collection Methods and Applications*. Spain: European Commission. [Online]. Available at: http://ftp.jrc.es/EURdoc/JRC47967.TN.pdf (Accessed: 10 September 2015).

Lee, J. and Bovik, A. C. (2009) *2009 16th IEEE International Conference on Image Processing (ICIP)*. 7-10 Nov. 2009.

Li, J-W., Lu, H-P., Geng, X-F. and Wang, H-W. (2011) 'Study on Dynamic Origin-Destination estimation of large-scale road Networks', *ICCTP 2011*.

Li, Y. and McDonald, M. (2007) 'Determining the sample size of probe vehicles', *Proceedings of the Institution of Civil Engineers - Transport*, 160(4), pp. 201-205.

Lomax, T. (2010) *Incorporating Sustainability Factors Into The Urban Mobility Report: A Draft Concept Paper*.

Lomax, T., Turner, S., Shunk, G., Levinson, H. S., Pratt, R. H., Bay, P. N. and Douglas, G. B. (1997) *Quantifying congestion. Volume 1: Final Report* (NCHRP Report 398). Washington, DC: Transportation Research Board. [Online]. Available at: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_398.pdf. (Accessed: 14 November 2016).

Maerivoet, S. and Moor, B. D. (2008) 'Traffic Flow Theory', [Online]. Available at: https://www.scribd.com/document/246430124/Traffic-Flow-Theory. (Accessed: 07 February 2017).

Mahalanobis, P. C. (1936) 'On the generalised distance in statistics', *Proceedings of the National Institute of Sciences of India*. pp. 49-55. Available at: http://www.unt.edu/rss/class/Jon/MiscDocs/1936_Mahalanobis.pdf (Accessed: 24 May 2016).

Malinovskiy, Y., Lee, U-K., Wu, Y-J. and Wang, Y. (2011) 'Investigation of Bluetooth-Based Travel Time Estimation Error on a Short Corridor', *Transportation Research Board 90th Annual Meeting,* Washington, D.C, 23-27 January 2011.

Malinovskiy, Y., Wu, Y-J., Wang, Y. and Lee, U. K. (2010) 'Field experiments on bluetooth-based travel time data collection', *Transportation Research Board 89th Annual Meeting, CD-ROM Paper*. Washington D.C, 10-14 January 2010.

Martchouk, M., Mannering, F. and Bullock, D. (2011) 'Analysis of Freeway Travel Time Variability Using Bluetooth Detection', *Journal of Transportation Engineering*, 137(10), pp. 697-704.

Mathew, T. V. (2014) 'Capacity and Level of Service (LOS)', in *Transportation Systems Engineering* Available at: www.nptel.ac.in/courses/ (Accessed: 30 October 2014).

McDonald, J. (2013) Bluetooth configuration and installation, 15 October 2013.

McGowen, P. and Sanderson, M. (2011) *Accuracy of Pneumatic Road Tube Counters*. Western District Annual Meeting: Institute of Transportation Engineers. [Online]. Available at: http://www.westernite.org/annualmeetings/alaska11/Compendium/Moderated%20Session%20Papers/8C-Patrick%20McGowen.pdf. (Accessed: 23 July 2016).

Mei, Z., Wang, D. and Chen, J. (2012) 'Investigation with Bluetooth Sensors of Bicycle Travel Time Estimation on a Short Corridor', *International Journal of Distributed Sensor Networks*, 2012, pp. 1-7.

Meng, X., Yang, L., Aponte, J., Hill, C., Moore, T. and Dodson, A. H. (2008) *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*. 12-15 October 2008.

Miles, J. C. and Chen, K. (2004) *The Intelligent Transport Systems Handbook: Recommendations from the World Road Association (PIARC)*. 2nd edn. India: Andrew Barriball.

Miller, B. A. and Bisdikian, C. (2001) *Bluetooth Revealed: The Insider's Guide to an Open Specification for Global Wireless Communications*. 2nd edn. Prentice Hall.

Minitab (2014) 'Minitab 17: Software for Quality Improvement'. Available at: http://www.minitab.com/en-us/. (Accessed: 02 September 2014).

Mintsis, G., Basbas, S., Papaioannou, P., Taxiltaris, C. and Tziavos, I. N. (2004) 'Applications of GPS technology in the land transportation system', *European Journal of Operational Research*, 152(2), pp. 399-409.

Misra, P. and Enge, P. (2006) *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press.

Mitchell, G., Hargreaves, A., Namdeo, A. and Echenique, M. (2011) 'Land use, transport, and carbon futures: the impact of spatial form strategies in three UK urban regions', *Environment and Planning A*, 43(9), pp. 2143-2163.

Moghaddam, S. and Hellinga, B. (2013) 'Quantifying measurement error in arterial travel times measured by bluetooth detectors' *Transportation Research Record*. pp. 111-122. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84897073297&partnerID=40&md5=5175c507212cf3eb2305ec9bb4f8578f. (Accessed: 24 February 2015).

Morris, T. (2014) Update on Data Collection, 04 September 2014.

Muhammed, R. and Egemalm, A. (2012) 'Travel time estimation based on Bluetooth sensors: Results from a comparative field test', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

Mulligan, A-M. and Nicholson, A. (2002) 'Uncertainty in Traffic Flow Estimation Using the Moving-Observer Method'. Transportation Group, NZ. Available at:

www.transportationgroup.nz/papers/2002/35_Mulligan_Nicholson.pdf. (Accessed: 07 February 2017).

MVP Programs (2014) *What are Box & Whisker Plots?* Available at: http://mvpprograms.com/help/mvpstats/graphics/WhatAreBoxAndWhiskerPlots (Accessed: 01 October 2014).

Nantes, A., Miska, M. P., Bhaskar, A. and Chung, E. (2014) 'Noisy bluetooth traffic data?', *Road and Transport Research*, 23(1), pp. 33-43.

National Policing Improvement Agency (2012) *Automatic Number Plate Recognition*. Available at: http://www.npia.police.uk/en/10505.htm (Accessed: 04 January 2012).

Nellore, K. and Hancke, G. P. (2016) 'A survey on urban traffic management system using wireless sensor networks', *Sensors (Switzerland)*, 16(2).

NFC (2016) *Near Field Communication Technology Standards*. Available at: http://nearfieldcommunication.org/technology.html (Accessed: 02 May 2016).

Ngoduy, D. (2013) *Urban traffic state estimation problems from bluetooth data*. Available at: http://ias.ust.hk/workshop/tmtm201312/doc/Dong_Ngoduy.pdf (Accessed: 25 May 2015).

O'Neill, E., Kostakos, V., Kindberg, T., gen. Schieck, A. F., Penn, A., Fraser, D. S. and Jones, T. (2006) 'Instrumenting the city: Developing methods for observing and understanding the digital cityscape', *8th International Conference on Ubiquitous Computing*. Springer, pp. 315-332.

Open Learn (2015) *Interpreting data: Boxplots and tables*. Available at: http://www.open.edu/openlearn/science-maths-technology/mathematics-and-statistics/statistics/interpreting-data-boxplots-and-tables/content-section-3.1.3 (Accessed: 17 June 2015).

Park, H. and Haghani, A. (2015) 'Optimal number and location of Bluetooth sensors considering stochastic travel time prediction', *Transportation Research Part C: Emerging Technologies*, 55, pp. 203-216.

Pasolini, G. and Verdone, R. (2002) 'Bluetooth for ITS', in The 5th International Symposium on Wireless Personal Multimedia Communications, 27-30 October 2002, Honolulu, Hawaii, 1, pp. 315-319.

Penn State Eberly College of Science (2016) *Applied Multivariate Statistical Analysis: Multivariate Normality and Outliers*. Available at: https://onlinecourses.science.psu.edu/stat505/node/59 (Accessed: 02 December 2016).

Persistent Market Research (2017) *Bluetooth in Automotive Market: Global industry analysis and forecast 2016 - 2026*. Available at: http://www.persistencemarketresearch.com/market-research/bluetooth-in-automotive-market.asp (Accessed: 10 January 2017).

Peterson, B. S., Baldwin, R. O. and Raines, R. A. (2006) 'Bluetooth Discovery Time with Multiple Inquirers', *Proceedings of the 39th Hawaii International Conference on System Sciences*, Kauai, Hawaii, 4-7 January 2006.

Polson, N. and Sokolov, V. (2015) 'Bayesian analysis of traffic flow on interstate I-55: The LWR Model', *The Annals of Applied Statistics*, 9(9), pp. 1864–1888.

Pourabdollah, A., Xiaolin, M. and Jackson, M. (2010) *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010*. 14-15 October 2010.

PRNewswire (2016) *Global Connected Car Market to Reach $81 Billion by 2025 - Cars Enabled with Smart-phones are Driving Market Growth*. Research and Markets. [Online]. Available at: http://www.fox34.com/story/33852902/global-connected-car-market-to-reach-81-billion-by-2025-cars-enabled-with-smart-phones-are-driving-market-growth-research-and-markets (Accessed: 10 January 2017).

Purdue University (2016) *Lesson 3 on Traffic Flow Models and Analysis: Continuous and Deterministic Flow* Available at: https://engineering.purdue.edu/ce361/LECTURE/Lcontin_flo.html (Accessed: 18 November 2016).

Qiao, W., Haghani, A. and Hamedi, M. (2013) 'A nonparametric model for short-term travel time prediction using bluetooth data', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 17(2), pp. 165-175.

Quayle, S. M., Koonce, P., Depencier, D. and Bullock, D. M. (2010) 'Arterial Performance Measures with Media Access Control Readers: Portland, Oregon, Pilot Study', *Transportation Research Record: Journal of the Transportation Research Board*, pp. 185-193.

Quddus, M. A. (2008) 'Time series count data models: an empirical application of traffic accidents', *Accident Analysis and Prevention*, 40(5), pp. 1732-1741.

Quiroga, C. A. (2000) 'Performance measures and data requirements for congestion management systems', *Transportation Research Part C: Emerging Technologies*, 8(1–6), pp. 287-306.

R Core Team (2013) *R: A language and environment for statistical computing.* Vienna, Austria: Computing, R.F.f.S. [Online]. Available at: http://www.R-project.org/. (Accessed: 7 October 2013).

Rewadkar, D. N. and Dixit, T. (2013) 'Review of Different Methods Used for Large -Scale Urban Road Networks Traffic State Estimation', *International Journal of Emerging Technology and Advanced Engineering*, 3(10), pp. 369-373.

Roess, R. P., McShane, W. R. and Prassas, E. S. (1998) *Traffic Engineering*. Second edn. New Jersey: Prentice Hall.

Roggendorf, S. (2012) 'Traffic counting at intersections by Bluetooth-technology', *19th ITS World Congress*. Vienna, Austria, 22-26 October 2012.

Rouse, M. (2014) *Long Term Evolution (LTE)*. Available at: http://searchmobilecomputing.techtarget.com/definition/Long-Term-Evolution-LTE (Accessed: 30 July 2016).

RTA (2011) *Speeding - Did you know? Why do we need speed limits?* Available at: http://www.rms.nsw.gov.au/saferroadsnsw/speedlimits-why.pdf (Accessed: 31 August 2015).

Sadoun, B. and Al-Bayari, O. (2007) 'Location based services using geographical information systems', *Computer Communications*, 30(16), pp. 3154-3160.

Salem, A., Nadeem, T., Cetin, M. and El-Tawab, S. (2015) *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84950286299&doi=10.1109%2fITSC.2015.123&partnerID=40&md5=8f4a9f497a042b1ac0f3044ddc38f599. (Accessed: 01 February 2017).

Salvo, P., Turcanu, I., Cuomo, F., Baiocchi, A. and Rubin, I. (2016) 'LTE Floating Car Data application off-loading via VANET driven clustering formation', *2016 12th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*. Available at: http://dl.ifip.org/db/conf/wons/wons2016/p192-salvo.pdf. (Accessed: 10 January 2017).

SBD (2012) *2025 Every Car Connected: Forecasting the Growth and Opportunity*. GSMA. [Online]. Available at: http://www.gsma.com/connectedliving/wp-content/uploads/2012/03/gsma2025everycarconnected.pdf (Accessed: 10 January 2017).

Schmidt, M., Giorgi, L., Chevreuil, M., Paulin, S., Turvey, S. and Hartmann, M. (2005) *GALILEO: Impacts on road transport* (EUR 21865 EN). Spain: European Commission.

Schrank, D., Eisele, B. and Lomax, T. (2012) *TTI's 2012 URBAN MOBILITY REPORT*. Powered by INRIX Traffic Data.

Science Daily (2014) *Traffic Engineering*. Available at: http://www.sciencedaily.com/articles/t/traffic_engineering_%28transportation%29.htm (Accessed: 14 November 2014).

Scofield, W. (2001) *Engineering Surveying: Theory and Examination Problems for Students*. Fifth edn. Amsterdam: Elsevier.

SCOOT-UTC (2011) *SCOOT- The World Leading Adaptive Traffic Control System*. Available at: http://www.scoot-utc.com/ (Accessed: 24 October 2011).

Scullion, P. (2011) *Billions wasted on congested roads*. Available at: http://politics.co.uk/news/2011/09/15/billions-wasted-on-congested-roads (Accessed: 24 May 2012).

Sebesta, R.W. (1999) *Concepts of Programming Languages*. Fourth edn. Reading: Addison-Wesley.

Sebri, M. (2016) 'Forecasting urban water demand: A meta-regression analysis', *Journal of Environmental Management*, 183, Part 3, pp. 777-785.

Selvarajah, K., Arief, B., Tully, A. and Blythe, P. (2012) *Deploying Wireless Sensor Devices in Intelligent Transportation System Applications*. Crotia: InTech.

Selvarajah, K., Tully, A. and Blythe, P.T. (2008) *ZigBee for intelligent transport system applications* (No. CS-TR-1097). Newcastle University.

Seo, S. (2002) *A review of comparison of methods for detecting outliers in univariate data sets*. MSc thesis. University of Pittsburgh.

Sethi, R. (1996) *Programming Languages: Concepts and Constructs*. Second edn. Reading: Addison-Wesley.

Shinya, K. and Dragana, M. (1999) 'Method to Preprocess Observed Traffic Data for Consistency: Application of Fuzzy Optimization Concept', *Transportation Research Record: Journal of the Transportation Research Board*, 1679, pp. 73-80.

Smith, D. (2014) 'Data Cleaning is a critical part of the Data Science process', *R-bloggers*. Available at: http://www.r-bloggers.com/ (Accessed: 18 August 2014).

Srinivasan, V. (2011) 'Turning to Bluetooth to Reduce Traffic Congestion. Strategic White Paper', [Online]. Available at: www.alcatel-lucent.com/wps/portal (Accessed: 20 December 2011).

Srivastava, T. (2015) 'A Complete Tutorial on Time Series Modeling in R', *Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/. (Accessed: 08 June 2016).

Stange, H., Liebig, T., Hecker, D., Andrienko, G. and Andrienko, N. (2011) 'Analytical workflow of monitoring human mobility in big event settings using Bluetooth. ', *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*. Chicago, Illinois. 01-04 November 2011.

Starkweather, J. (2013) *Multivariate outlier detection with Mahalanobis distance*. Available at: https://it.unt.edu/sites/default/files/moutlier_jds_july2013.pdf (Accessed: 24 May 2016).

Statgraphics (2015) *Time Series Analysis and Forecasting*. Available at: http://www.stagraphics.com/time-series_analysis.htm (Accessed: 12 January 2015).

Stavig, G. R. and Gibbons, J. D. (1977) 'Comparing the mean and the Median as measures of Centrality', *Internaltional Statistics Review*, 45, pp. 63-70.

Stevanovic, A., Olarte, C. L., Galletebeitia, Á., Galletebeitia, B. and Kaisar, E. I. (2015) 'Testing Accuracy and Reliability of MAC Readers to Measure Arterial Travel Times', *International Journal of Intelligent Transportation Systems Research*, 13(1), pp. 50-62.

Tabona, A. Z. (2005) *802.11b vs. Bluetooth*. Available at: http://www.windowsnetworking.com/articles-tutorials/wireless-networking/80211b-Bluetooth.html (Accessed: 02 August 2016).

Tang, J., Zou, Y., Ash, J., Zhang, S., Liu, F. and Wang, Y. (2016) 'Travel Time Estimation Using Freeway Point Detector Data Based on Evolving Fuzzy Neural Inference System', *PLoS ONE*, 11(2), p. e0147263.

Tarnoff, P. J., Bullock, D. M., Young, S. E., Wasson, J., Ganig, N. and Sturdevant, J. R. (2009) 'Continuing Evolution of Travel Time Data Information Collection and Processing', *Transportation Research Board 88th Annual Meeting*. Available at: http://en.wikipedia.org/wiki/Telematics (Accessed: 17 December 2011).

TDC (2011) *Hi-Trac Blue:Traffic impact analysis system*. Available at: http://www.tdcsystems.co.uk (Accessed: 23 December 2011).

Texas Transportation Institute (2011) *Detector Life‐Cycle Costs and Considerations: Mobility Measurement in Urban Transportation Pooled Fund Study*. [Online]. Available at: http://mobility.tamu.edu/resources (Accessed: 22 December 2011).

Thomson, J. M. (1978) *Great Cities and Their Traffic*. Great Britain: The Chancer Press.

Thorgil (2007) *Congestion Baseline Report*. [Online]. Available at: https://www.newcastle.gov.uk/sites/default/files/wwwfileroot/planning-and-buildings/planning/09.08_congestion_baseline_report_july_2007.pdf (Accessed: 02 December 2016).

Thorpe, N. (2005) *Public Attitudes to Road-User Charging: A Case Study of the Toll-Rings in Norway*. Newcastle University

TMCnet (2011) *A Cost-Effective, Bluetooth-Focused Approach to Traffic Monitoring*. Available at: http://blog.tmcnet.com/next-generation-communications/2011/10/a-cost-effective-bluetooth-focused-approach-to-traffic-monitoring.html (Accessed: 14 December).

Triggs, R. (2013) *What is NFC & how does it work?* Available at: http://www.androidauthority.com/what-is-nfc-270730/ (Accessed: 02 May 2016).

Trimble (2007) *GPS: The First Global Navigation Satellite System*. USA: Trimble.

Troester, M. (2012) *Big Data Meets Big Data Analytics*. SAS.

Tsekeris, T. and Stathopoulos, A. (2006) 'Measuring variability in urban traffic flow by use of principal component analysis', *Journal of Transportation and Statistics*, 9(1), pp. 49-62.

Tsubota, T., Bhaskar, A., Chung, E. and Billot, R. (2011) 'Arterial traffic congestion analysis using Bluetooth duration data', in Tisato, P., Oxlad, L. and Taylor, M. (eds.) *Australasian Transport Research Forum 2011, 28 - 30 September 2011*. Adelaide Hilton Hotel, Adelaid, SA.

Tukey, J. W. (1980) 'We Need Both Exploratory and Confirmatory', *The American Statistician*, 34(1), pp. 23-25.

Turochy, R. E. and Smith, B. L. (2002) 'Measuring variability in traffic conditions by using archived traffic data' *Transportation Research Record*. pp. 168-172. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0036998508&partnerID=40&md5=1dc50ce9304b88850d0e8f3e2bec24b3. (Accessed: 24 February 2015).

Tyne and Wear (2010) *Local Transportation Plan Tyne and Wear*. [Online]. Available at: http://www.tyneandwearltp.gov.uk/wp-content/uploads/2010/09/CRPpt2.pdf (Accessed: 02 December 2016).

Tyne and Wear (2011) *Keep Tyne and Wear Moving*. [Online]. Available at: http://www.tyneandwearltp.gov.uk/wp-content/uploads/2011/03/TW-LTP3-Delivery-Plan-Mar-2011-for-upload.pdf (Accessed: 02 December 2016).

UMCATT (2008) 'Bluetooth Traffc Monitoring Technology: a Concept of Operation & Deployment Guidelines', pp. 1-5 [Online]. Available at: www.i95coalition.org/ (Accessed: 14 December 2011).

Vainio, J. T. (2000) 'Bluetooth Security '. Available at: http://simson.net/ref/2004/bluesec.html (Accessed: 23 April 2012).

van Lint, J. W. C., Hoogendoorn, S. P. and van Zuylen, H. J. (2005) 'Accurate freeway travel time prediction with state-space neural networks under missing data', *Transportation Research Part C: Emerging Technologies*, 13(5–6), pp. 347-369.

Vo, T., Suh, W., Guensler, R., Guin, A., Hunter, M. P. and Rodgers, M. O. (2012) 'Assessment of multi-antenna array performance for detecting bluetooth enabled devices in a traffic stream', *Transportation Research Board 91st Annual Meeting,* 22-26 January 2012.

Wang, C., Quddus, M. A. and Ison, S. G. (2009) 'Impact of traffic congestion on road accidents: A spatial analysis of M25 motorway in England', *Accident Analysis & Prevention*, 41(1), pp. 798-808.

Wang, C., Quddus, M. A. and Ison, S. G. (2013) 'The effect of traffic and road characteristics on road safety: A review and future research direction', *Safety Science*, 57, pp. 264-275.

Wang, X. and Zhang, N. (2005) 'GLS Estimation of OD matrix with traffic counts and information from ATIS', *Proceedings of the Eastern Asia Society for Transportation Studies*. pp. 1188 – 1196.

Wang, Y., Malinovskiy, Y., Wu, Y-J. and Lee, U. K. (2011) *Error modelling and analysis for travel time data obtained from Bluetooth MAC address matching*. Seattle, Washington: University of Washington.

Warren, R., Smith, R. E. and Cybenko, A. K. (2011) *Use of Mahalanobis distance for detecting outliers and outlier clusters in markedly non-normal data; a vehicular traffic example* (AFRL-RH-WP-TR-2011-0070). Air Force Research Laboratory.

Wason, J. S., Sturdevant, J. R. and Bullock, D. M. (2008) 'Real-time travel estimates using MAC address matching', *Institiute of Transport Engineers Journal*, 78(6), pp. 20-23.

Watt, D. A. (1990) *Programming Language: Concepts and Paradigms* New York: Prentice Hall.

Webster, N., Darter, M. and Hranac, R. (2014) *21st World Congress on Intelligent Transport Systems, ITSWC 2014: Reinventing Transportation in Our Connected World*. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84929191839&partnerID=40&md5=d0383779649d218e825e6c34ecd5ada4. (Accessed: 01 February 2017).

Wedagama, D. M. P., Bird, R. N. and Dissanayake, D. (2007) 'The Influence of Urban Land Use on Pedestrian Accidents during Congested and Uncongested Periods ', *Journal of the Eastern Asia Society for Transportation Studies*, 7, pp. 2830-2843.

Weigelt, H. R., Gotz, R. E. and Weiss, H. H. (1973) *City Traffic- A System Digest*. Translated by Wengatz, G. F. New York: Van Nostrand Reinhold Company.

White, M. T. (1989) 'Traffic Monitoring- Local Highway Authority Requirements', in *Second International Conference on Road Traffic Monitoring*. London: IEE.

WHO (2005) *Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide*. Geneva, Switzerland: World Health Organization.

Whyte, W. S. and Paul, R. E. (1997) *Basic Surveying*. Fourth edn. Oxford: Laxtons.

Windmill (2016) *Vehicle Sensing: Ten Technologies to Measure Traffic*. Available at: http://www.windmill.co.uk/vehicle-sensing.htm (Accessed: 23 July 2016).

Wood, T. (2012) *Using Mean Absolute Error for Forecast Accuracy*. Available at: http://canworksmart.com/using-mean-absolute-error-forecast-accuracy/ (Accessed: 02 December 2016).

Wu, C. (2016) 'Variation inference Part 2'. North-Eastern University. [Online] Available at: https://www.youtube.com/watch?v=uKxtmkfeuxg. (Accessed: 02 December 2016).

Yildirimoglu, M., Limniati, Y. and Geroliminis, N. (2015) 'Investigating empirical implications of hysteresis in day-to-day travel time variability', *Transportation Research Part C: Emerging Technologies*, 55, pp. 340-350.

Young, S. E., Sharifi, E., Sadrsadat, H., Serulle, N. U. and Sadabadi, K. F. (2013) 'Detection Probability Models for Bluetooth Re-identification Technology', *ITSWC Conference*. America. Available at: https://itswc.confex.com (Accessed: 11 September 2015).

Zhang, S., Wang, H., Quan, W. and Liu, X. (2013) 'Measuring Variability in Freeway Traffic States Using Real-time Loop Data in Jilin', *Procedia - Social and Behavioral Sciences*, 96, pp. 2676-2683.

Zhong, Z. and Lee, J. (2014) *21st World Congress on Intelligent Transport Systems, ITSWC 2014: Reinventing Transportation in Our Connected World*. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84929207209&partnerID=40&md5=ad928468937874dcd66b95fac4c64acb. (Accessed: 01 February 2017).

ZigBee (2014) *About ZigBee Technology*. Available at: http://www.zigbee.org/About/AboutTechnology/ZigBeeTechnology.aspx (Accessed: 14 August 2014).

APPENDICES

Appendix 1

Appendix 1: A typical Bluetooth sensor mounted on a lamp post

Appendix 2

Site ID	Date	Lane	Lane Name	Direction	Direction Name	Class Scheme	Class	Class Name	Length (ft)	Headway (s)	Gap (s)	Speed (mph)	Weight (lb)	Vehicle Id	Flags	Flag Text	Num Axles
'MAC090001001	01/09/2012 00:00:13	1	MAC	2	South		0	Class 0						9.7093E+11	0		0
'MAC090001001	01/09/2012 00:00:22	1	MAC	2	South		0	Class 0						521E0E0051C7	0		0
'MAC090001001	01/09/2012 00:00:33	1	MAC	2	South		0	Class 0						DD22550541DC	0		0
'MAC090001001	01/09/2012 00:06:27	1	MAC	2	South		0	Class 0						180F720072EC	0		0
'MAC090001001	01/09/2012 00:09:47	1	MAC	2	South		0	Class 0						3473BE41D113	0		0
'MAC090001001	01/09/2012 00:10:49	1	MAC	2	South		0	Class 0						4E0A6E0029FC	0		0
'MAC090001001	01/09/2012 00:12:42	1	MAC	2	South		0	Class 0						D5E2408E4D69	0		0
'MAC090001001	01/09/2012 00:16:56	1	MAC	2	South		0	Class 0						87555FC55C87	0		0
'MAC090001001	01/09/2012 00:18:42	1	MAC	2	South		0	Class 0						56D5D102A073	0		0
'MAC090001001	01/09/2012 00:19:44	1	MAC	2	South		0	Class 0						1E36FA8E4B61	0		0
'MAC090001001	01/09/2012 00:19:53	1	MAC	2	South		0	Class 0						B12ECC001FB3	0		0
'MAC090001001	01/09/2012 00:21:20	1	MAC	2	South		0	Class 0						A3194E00A6C9	0		0
'MAC090001001	01/09/2012 00:21:22	1	MAC	2	South		0	Class 0						D4DD36073F26	0		0
'MAC090001001	01/09/2012 00:21:36	1	MAC	2	South		0	Class 0						D315D700D6CF	0		0
'MAC090001001	01/09/2012 00:21:36	1	MAC	2	South		0	Class 0						FE13F800AA4E	0		0
'MAC090001001	01/09/2012 00:21:57	1	MAC	2	South		0	Class 0						6B9448CE35BB	0		0
'MAC090001001	01/09/2012 00:22:07	1	MAC	2	South		0	Class 0						E52AF5005CF7	0		0
'MAC090001001	01/09/2012 00:25:42	1	MAC	2	South		0	Class 0						E08065007982	0		0
'MAC090001001	01/09/2012 00:25:51	1	MAC	2	South		0	Class 0						C123FD004FE9	0		0
'MAC090001001	01/09/2012 00:28:03	1	MAC	2	South		0	Class 0						3E77EC8C1759	0		0

Appendix 2: An example Bluetooth sensor data (encrypted) captured at station MAC1001 in Trafford

Appendix 3

Appendix 3A: Description of TRAFOST

The development of TRAFOST has very significant advantages such as in processing time, the volume of data processed, reproducibility and reliability. The TRAFOST and outputs enabled a well-structured analysis and presentation of the data unlike the output from the Excel/manual computation. In line with Sebesta (1999), TRAFOST is considered reliable since it is reproducible and time-saving as well as performing to specification. Another advantage derivable from the use of the TRAFOST is in organisation. Programming languages provide ways of organising computations (Sethi, 1996). However, its choice depends partly on the programming to be done, and partly on other external factors that include availability, support, and training (Sethi, 1996). Another factor that calls for consideration is the semantics of a programming language that concerns how programs behave when executed (Watt, 1990). Several factors such as cost, accessibility and speed of processing were given consideration before arriving at the choice of R. For example, Matlab programming language was considered suitable, but it does require the purchase of a licence, unlike R that is open source. That is, R is available for free download and works on multiple computing platforms (Dalgaard, 2002). Not only that, for many years, R is a leading software in terms of data and results visualisation (Chang, 2014). The basic four stages of the model (TRAFOST) developed in this research as well as the input sources and the formats of the data used are presented. The stages are: i) data capture and storage; ii) data manipulation; iii) analysis; and iv) display of results.

Stage 1: Data Capture and Storage

Data upload and storage

As discussed in Section 3.2.2, following the on-site data capture and online data storage, the encrypted data (for privacy reasons) were downloaded and assessed for physical quality such as in resolution, structure and format before the final storage in the processing environment. The storage format is comma separated variables (csv). The downloaded files were stored accordingly in a directory with unique names for easy manipulation and retrieval. At the same time, the original data remains unaltered for future use.

Input sources and data types

The input sources and the data types used in the implementation of the model are as previously described in Section 3.2.1. As a reminder, TRAFOST was implemented using Bluetooth, SCOOT, ATC, ANPR and TM data consisting of varying resolutions and formats.

Stage 2: Data Manipulation

According to Andrienko and Andrienko (2006), data manipulation is chiefly to derive new data from existing data for more convenient or comprehensive analysis. The TRAFOST deployed was used to massage the data into a useful form. The process of the data manipulation was automated and executed in turn, over different phases. These stages include recoding and renaming, sorting and merging datasets, aggregating, reshaping, sub-setting using some specified criteria through the use of arithmetic and logical operators as well as statistical functions. The operations include data merging, file reduction and ordering, data filtering and the creation of time series objects as well as merging data from different stations to create O-D patterns of the network.

It is a known fact that data size is a key factor requiring adequate consideration in any data processing for the purpose of software efficiency. Therefore, file reduction is paramount to conserve memory and gain computational speed. Each originally downloaded data file used in this study contains 15, and in some cases 20 variables. The initial set of the data collected contains 15 variables while the subsequent data collected contains 20 variables following the modification of the software of the sensors. Some of the variables include lane, lane name, direction and direction name. However, only three of the variables (Station Id, Timestamp, and Vehicle Id) are required in this research. Hence, each file was reduced to the required three variables with the original data file unaltered. Data ordering, on the other hand, was performed to organise the data appropriately to enable easy and efficient manipulation. The captured Bluetooth data contain MAC addresses of different devices such as mobile phones (on pedestrians or in vehicles) and vehicles (cars, buses, HGVs). However, since the research focuses on vehicle detection to estimate traffic metrics rather than tracking pedestrians or other road users, high-level data filtering is required to separate the devices reasonably. Hence, filtering is considered one of the intricate aspects of Bluetooth data processing. The filtering involved different phases to carry out the data mining process. Section 3.3 discusses the methods of Bluetooth data cleansing.

The timestamp of the Bluetooth data was used to create time series records of different resolutions such as 5-min, 10-min, 15-min etc. This is necessary to examine Bluetooth profiles at different temporal dimensions to come to a logical conclusion on the usability of the data. That is to understand at what levels of resolution the data could be of best use. It is also to determine whether fluctuations in hourly/daily/monthly traffic flows provide any evidence of some underlying change in traffic that must be taken into account. The understanding of such variations, as well as travel patterns and movement across a network, is fundamental to effective traffic flow modelling, and was considered in the algorithm design.

Link distance computation

Table A1 presents the summary of the sources of the road (link) length used, the formats and the input mode at the execution stage. TRAFOST takes distance information (input) from either an existing file or onscreen. Distances are also computed from station coordinates either in the form of a grid or geographic coordinates where possible according to the road configuration. Other sources of distance information include TfGM database and Google Earth/Maps.

Metric	Source	Input type	Input mode	Purpose
Length	Station coordinates, TfGM database and Google maps	Real number (Grid or geographic coordinates) and grid length	File import or onscreen	Distance computation

Table A1: Summary of the distance functional component of the TRAFOST

Traffic metrics estimation components

Table A2 presents the summary of the estimated metrics using TRAFOST. Source defines the primary variables such as MAC address, timestamp and distance used to derive the metrics. Input type defines the nature of the data used such as raw, summary, date/time and link length. The format gives the form of the data such as character/string, factor, hour, minute, second, metre or kilometre and the like. The output type defines the form of the processed data that include integer and real variables while extension presents an appendage to the primary function of the module. For example, the chief role of the matrix module is to compute matrices of traffic flow data but it can also compute O-Ds for journey times and speed.

Estimated Traffic Metrics	Source	Input type	Format	Output type	Extension
Traffic Count	MAC address	Encrypted raw data	Character/string	Integer number	
Flow	MAC address	Encrypted raw data	Character/string	Integer number	SCOOT and ATC link-by-link flows
Journey Time	Time stamp	Date and time	dd/mm/yyyy hh:mm:ss; or dd-mm-yyyy hh:mm:ss	Real number	
Journey Speed	Time stamp and link-distance	DateTime and real number	dd/mm/yyyy hh:mm:ss; m or km	Integer number	
O-D Matrix	MAC address, link-distance and time stamp	Encrypted raw data, DateTime, length	Character/string, m or km and dd/mm/yyyy hh:mm:ss	Real and integer number	Journey time and speed

Table A2: Table showing the traffic metrics estimation components

Data aggregation and integration

Table A3 presents the summary of the types of data aggregation and integration performed by TRAFOST. The aggregation types range from 5-min to monthly averages. "*Yes*" or "*No*" defines whether such averages were performed on a specified data or not. They also define whether the validation data sets were analysed against Bluetooth at the specified temporal dimension or not. The column of "Integrated metrics" on the other hand presents the types of metrics integrated with the IMTD for accuracy and validity assessments.

	Types of Aggregation and Integration							
Data class	**5-min**	**10-min**	**15-min**	**Hourly**	**Weekday**	**Daily**	**Monthly**	**Integrated metrics**
Bluetooth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Flow, journey times and speed
SCOOT	No	No	Yes	Yes	Yes	Yes	Yes	Flow
ATC	No	No	Yes	Yes	Yes	Yes	Yes	Flow
ANPR	No	No	Yes	Yes	No	No	No	Flow, journey times and speed
TM	No	No	No	Yes	No	No	No	Journey times and Speed

Table A3: Summary of the types of data aggregation and integration

Stage 3: Data Analysis

Data analysis helps in the understanding of the phenomena in data (Andrienko and Andrienko, 2006). TRAFOST was used in this research to characterise Bluetooth data through data analysis to understand its underlying behaviour. TRAFOST incorporates both exploratory and quantitative methods of data analysis to obtain a richer understanding of the Bluetooth data than could be obtained using any manual method. The implementation of the model is dependent upon R statistical packages.

Detection of outliers and data cleaning

Cleaning of the data to remove outliers to obtain an accurate estimate of the traffic stream is essential. The error sources include the possibility of redundant observations (occurring due to repeated measurements or multiple matches of a device at a location); conflicting MAC address (arising from WiFi devices or encryption error); unknown mode or carrier that may lead to inclusion of other modes during classification, particularly at peak periods; unknown exact time of detection of a device leading to error in the estimation; missed detection – not all the devices can be detected leading to small sample or low detection rate; and loss of information outside the detection zone, unlike the GPS method that could provide continous information throughout the journey. TRAFOST incorporates the Mahalanobis distance method of outliers' detection and data cleaning due to its versatility to handle multivariate normal data as well as the possibility to handle markedly non-normal traffic data as demonstrated by Warren *et al.* (2011). Boxplots were used to visualise the data for exploratory assessment.

Integration of diverse data sources for validation of results

The availability of diverse sources of independently measured traffic data enabled both rigorous and sound validation of the model outputs. The integration of the other sets of data with Bluetooth data for the validation exercise is essential as Bluetooth estimates present only a sample of the total population that is lower than the actual traffic flow. The model design accommodated validation, refinement and re-validation using these set of data for the purpose of establishing a generalisable relationship between them. The comprehensive results of the validation and testing are presented in Chapter 5. In accordance with Edwards and Hamson (2001) the model and the methodology developed in this research is not thought of as the only right and proper solution for Bluetooth traffic metrics estimation.

Stage 4: Display of Output

Good data visualisation provides for a balance between scepticism and discovery, which helps in the general understanding of the data (Cook, 2014). Therefore, offline and web data display techniques and technologies were used for presentation of results to discover and characterise salient features in Bluetooth data. The outputs of the data processing were primarily two-fold: quantitative and graphical outputs. The quantitative output comprises information such as the network summary, daily and hourly flow. These were stored as a csv file. The graphical outputs were either displayed on R graphical console or customised where possible to be viewed on Google maps or Google Earth. Results were also explored using statistical data graphics covering static data visualisations as well as interactive and dynamic graphics. Table A4 presents the summary of the capabilities of TRAFOST both in terms of display and output of results. The "*Yes*" or "*No*" in the table is according to whether the indicated functionality is available or not.

	Metrics					
Display option	**Map**	**Count**	**Flow**	**JT**	**Speed**	**O-D Matrix**
Static	Yes	Yes	Yes	Yes	Yes	Yes
Interactive	Yes	Yes	Yes	Yes	Yes	No
Google Earth	Yes	No	No	No	No	No
Line Graph	No	Yes	Yes	Yes	Yes	No
Bar Graph	No	Yes	Yes	Yes	Yes	No
Bubble Graph	No	Yes	Yes	Yes	Yes	No
Colour	Yes	Yes	Yes	Yes	Yes	No

Table A4: Summary of the TRAFOST display and output capabilities

Typical time taken for the processing of sample data

TRAFOST was implemented on Windows-based computing systems. Table A5 presents the summary of the typical time taken to process data using TRAFOST. For example, a data size of 1.01GB processed with the Laptop described above over four O-D nodes and for 7 days worth of data took 1 hr 33 mins from upload to subsetting of the data and to the final processing of the hourly O-D matrix. Using the Desktop, it took 2 hrs 10 mins to complete the same process. This shows a significant change in the time spent. Another trial based on an increase in the number of days and nodes also showed a significant increase in time spent using the Desktop. It took 5 hrs 22 mins to complete the processing of 30 days of five nodes of O-D extracted from 660MB of data. Similarly, a significant decrease in time was observed with a decrease in the volume of the uploaded data (35MB) and the number of days processed at a time (8 days) despite an increase in the number of O-D nodes (9). In this case, the Desktop processing time was 8 mins. This shows an improvement in productivity, therefore, informing the knowledge of the management of TRAFOST for speed and efficiency. However, with cloud computing and the recent developments in R packages such as the introduction of 'data.table' DataCamp (2014), greater speed and efficiency can be achieved in real-time application.

Computer Specification	**Type**	**Data size**	**No of stations**	**No of days processed**	**Hourly O-D processing time**
Intel ® Core ™ i5-3230M CPU @2.60 GHz 2.60, 6GB RAM, 64-bit	Laptop	1.01GB (13 months)	4	7	1hr 33mins
Intel ® Core ™ i5 CPU 650 @ 3.20 GHz 3.19, 4GB RAM, 64-bit	Desktop	1.01GB (13 months)	4	7	2hrs 10mins
Intel ® Core ™ i5 CPU 650 @ 3.20 GHz 3.19, 4GB RAM, 64-bit	Desktop	660MB (7 months)	5	30	5hrs 22mins
Intel ® Core ™ i5 CPU 650 @ 3.20 GHz 3.19, 4GB RAM, 64-bit	Desktop	35MB (8 days)	9	8	8mins

Table A5: Typical time taken to process Bluetooth data on a Windows platform based system and data configuration

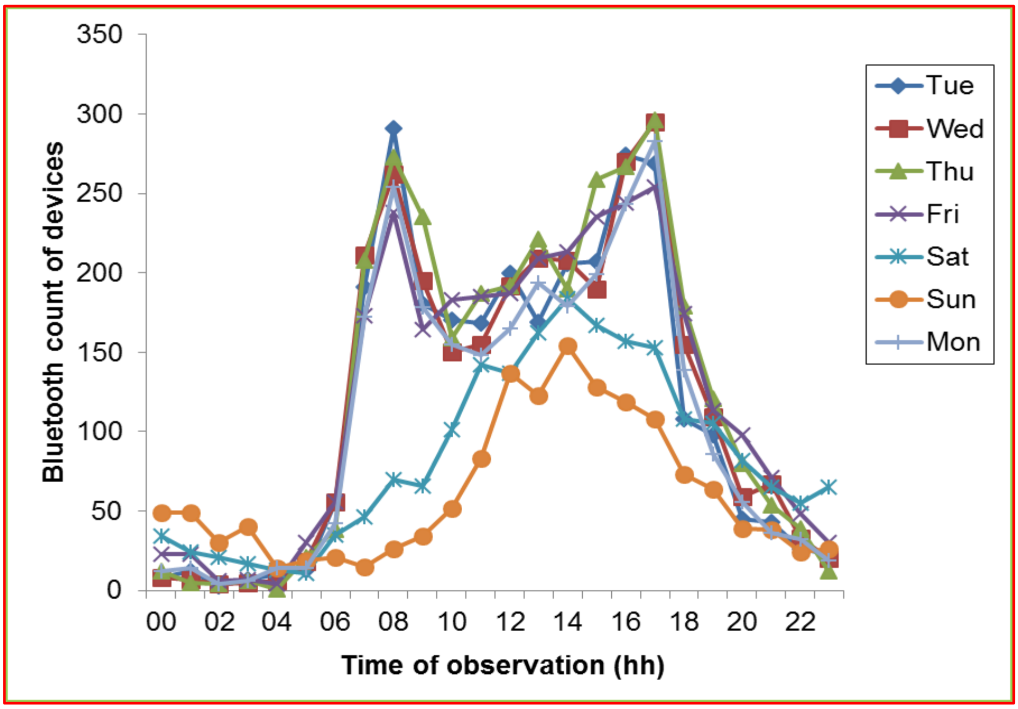
Appendix 3B: R-codes for Bluetooth processing

See codes at the end of the appendices

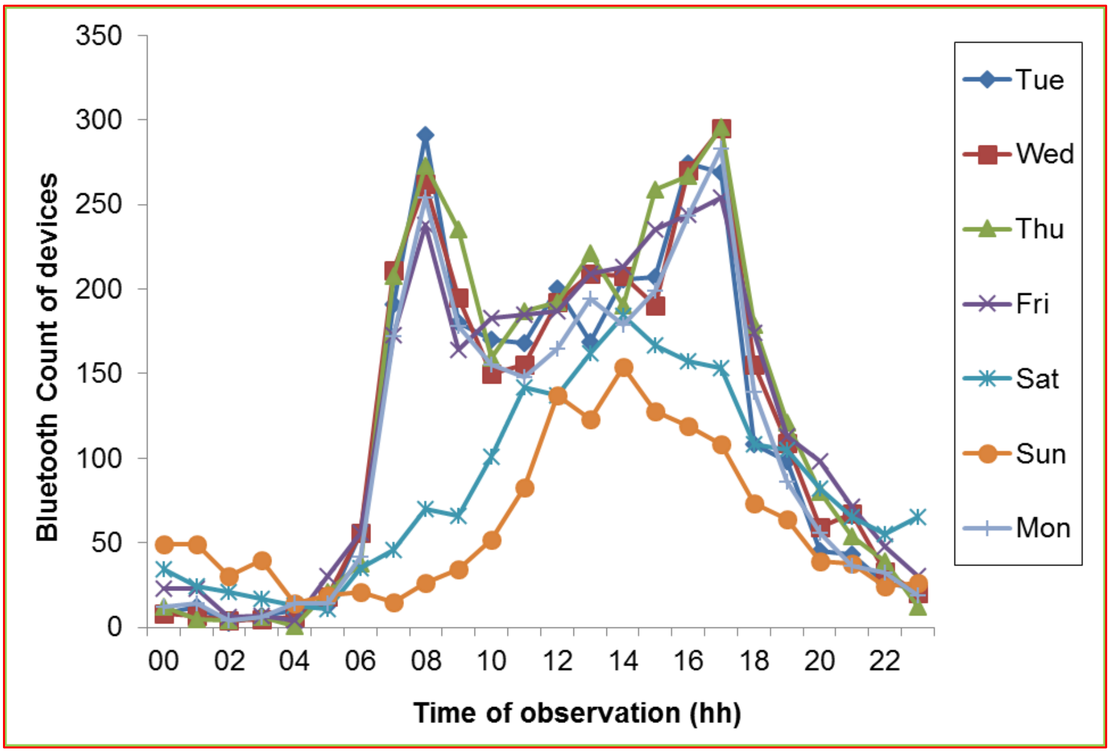
Appendix 4

SiteId	"MAC000000001"
SiteName	"1"
SiteDescription	"Bath Street"
SiteLatitude	53.41139
SiteLongitude	-2.99908
RecTime	VehicleId
13/06/2011 16:43:43	"8914E600163E"
13/06/2011 16:43:43	"8914E600163E"
13/06/2011 16:43:43	"83FD3507895A"
13/06/2011 16:43:43	"83FD3507895A"
13/06/2011 16:43:43	"8914E600163E"
13/06/2011 16:43:44	"8914E600163E"
13/06/2011 16:43:44	"3E19C600D3AA"
13/06/2011 16:43:45	"D31E740086C7"
13/06/2011 16:43:45	"D31E740086C7"
13/06/2011 16:43:45	"83FD3507895A"
13/06/2011 16:43:45	"D31E740086C7"
13/06/2011 16:43:45	"83FD3507895A"
13/06/2011 16:43:46	"83FD3507895A"

Appendix 4A: An example data (encrypted) for the Liverpool study area

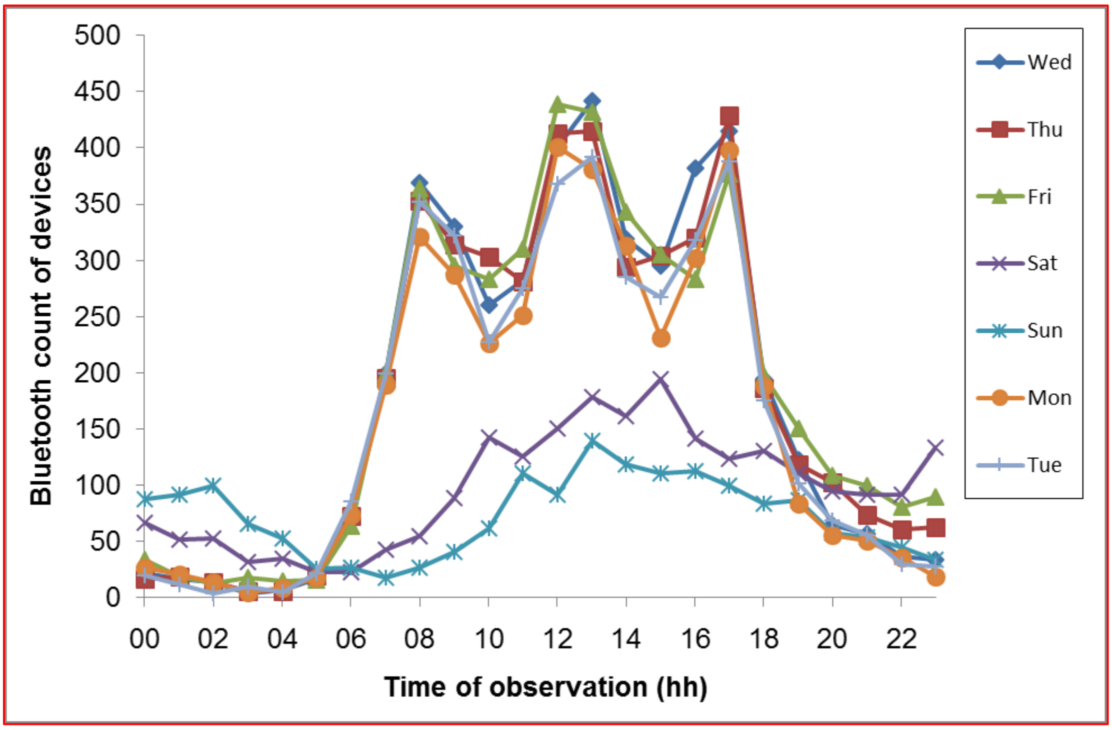


Appendix 4B-1: Profile of count of detected devices at Station 1 over weekdays

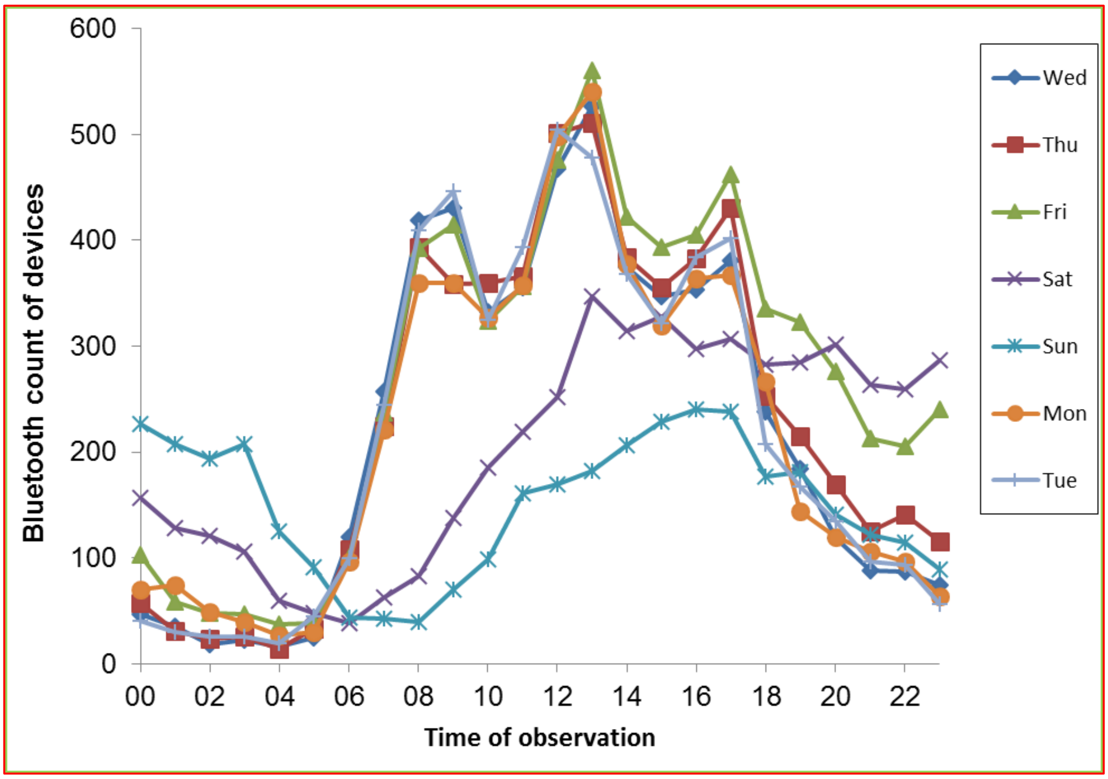


s

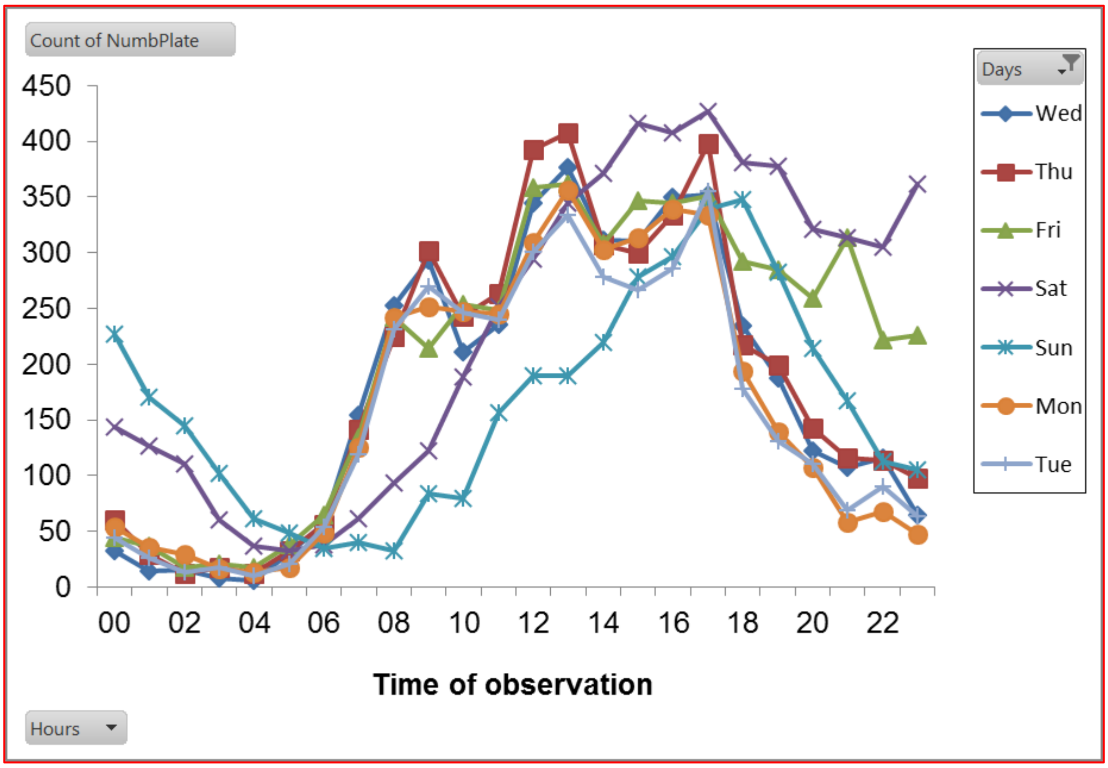
Appendix 4B-2: Profile of count of detected devices at Station 2 over weekdays



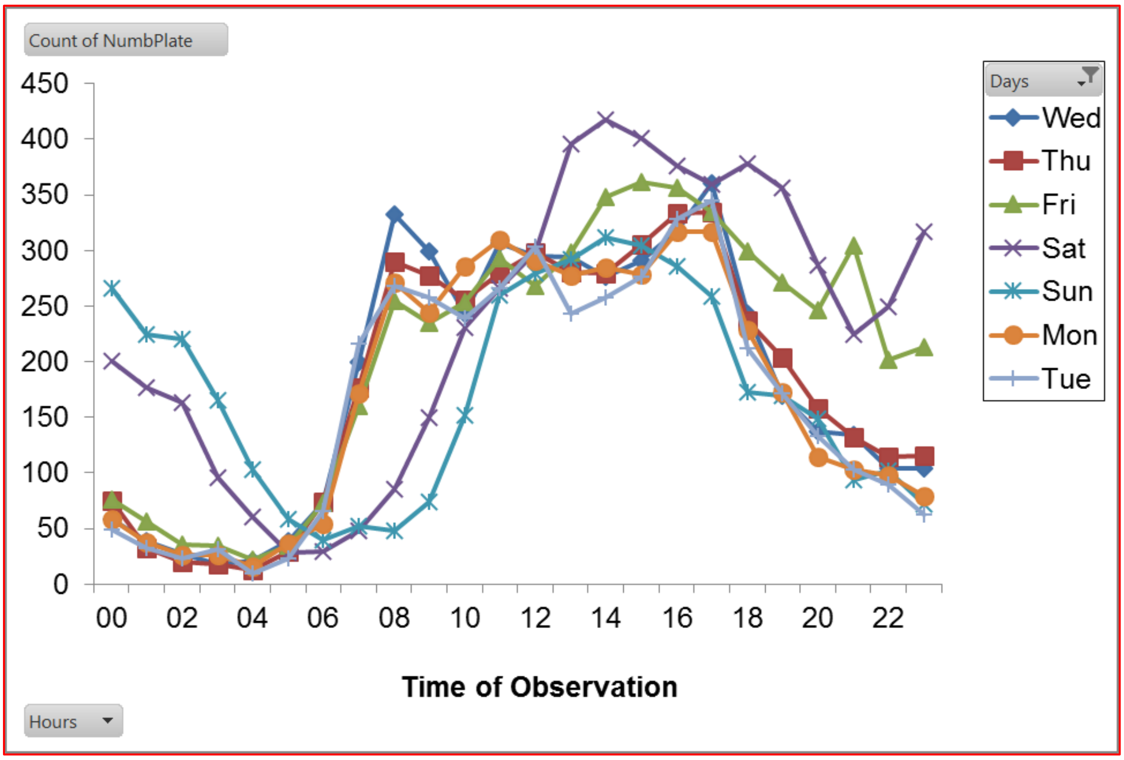
Appendix 4B-3: Profile of count of detected devices at Station 3 over weekdays



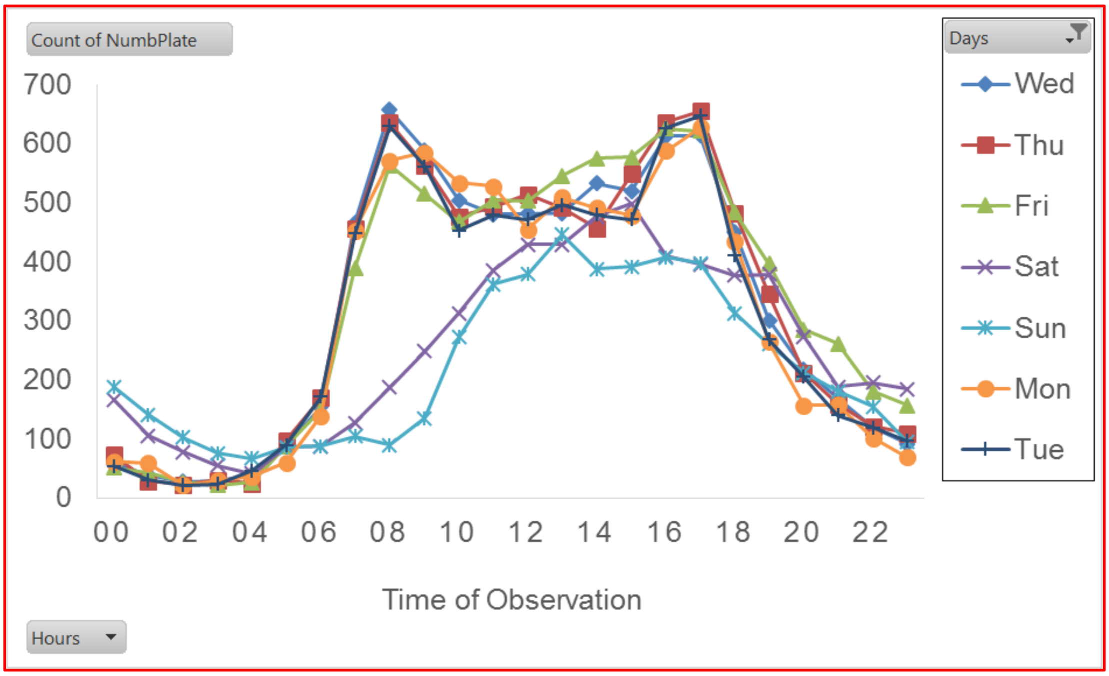
Appendix 4B-4: Profile of count of detected devices at Station 4 over weekdays



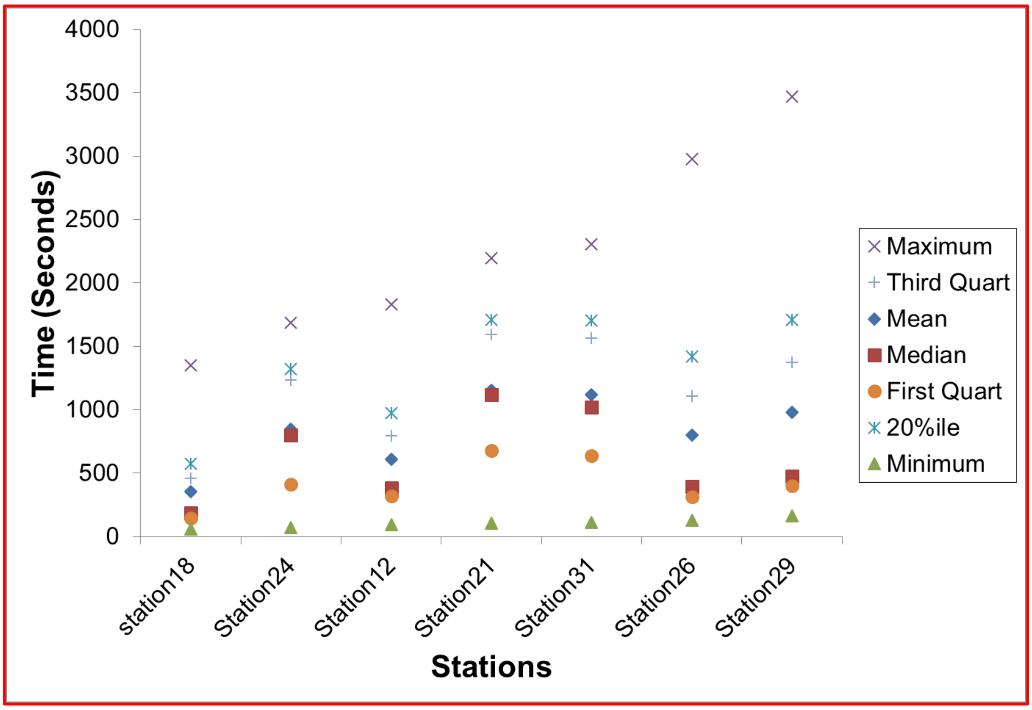
Appendix 4B-5: Profile of count of detected devices at Station 5 over weekdays



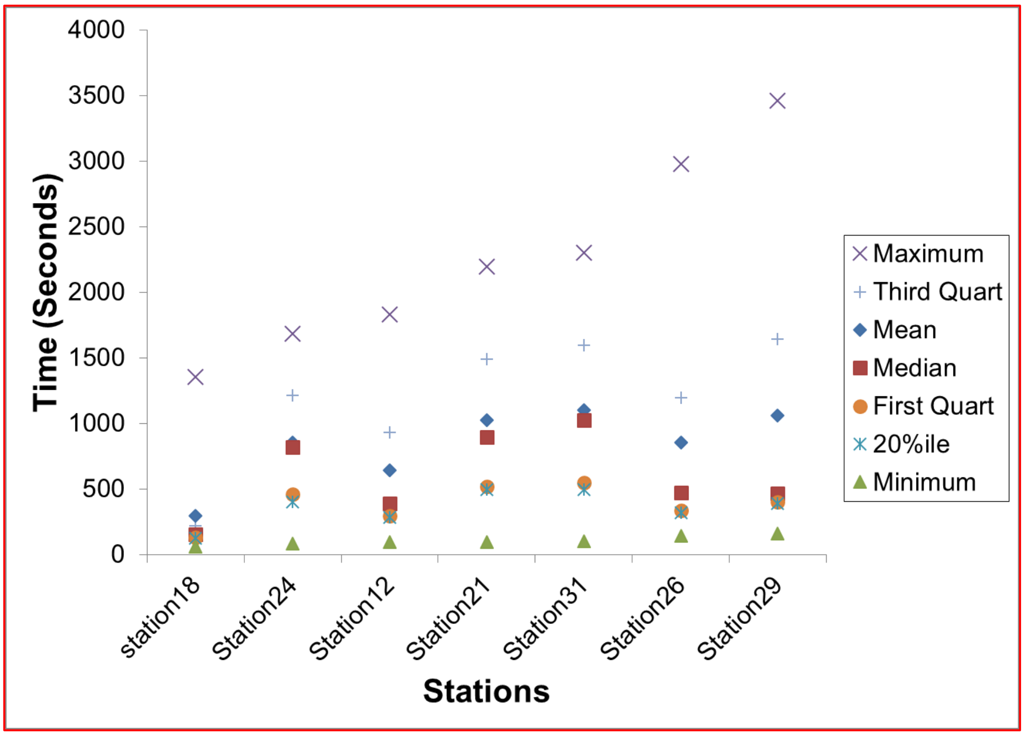
Appendix 4B-6: Profile of count of detected devices at Station 6 over weekdays



Appendix 4B-7: Profile of count of detected devices at Station 7 over weekdays



Appendix 4C-1: Plot of time estimate parameters to station 14



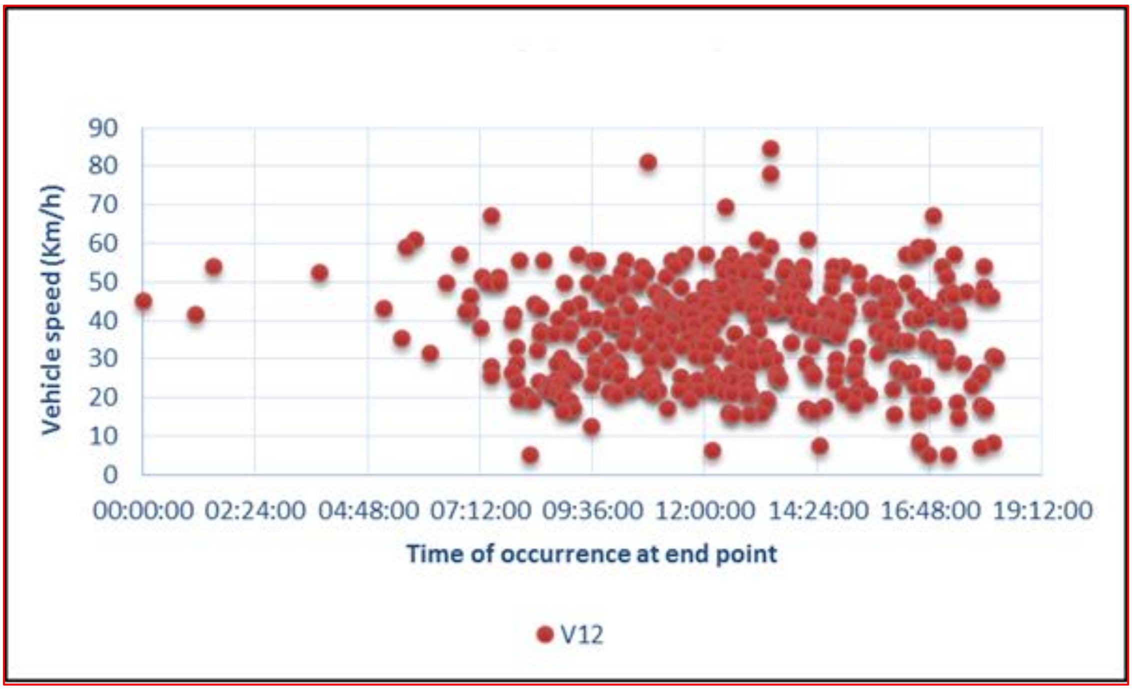
Appendix 4C-2: Plot of time estimate parameters from station 14

	STATION	12		16		18		21		24		26		29		31		
UNITS	(SEC, KM/H)	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TOTAL
	N	5216		1112		12501		1352		2037		7745		1920		616		32499
	MEAN	640.9	21.9	721.7	11.1	293.7	39.8	1025.3	14.6	855.7	13.7	851.4	30.8	1057.8	29.6	1103.0	15.2	
FROM	MEDIAN	387	23.7	728	8.7	156	43.4	894.5	12.3	817	10.3	474	31.5	465	37.4	1022.5	11.3	
STATION	MIN	95	5	84	5	57	5	93	5	82	5	144	5	162	5	102	5	
14	MAX	1832	96.4	1263	75.3	1353	118.7	2195	118.1	1683	102.8	2980	103.5	3462	107.2	2304	113.1	
	20%ile	281	8.3	424	6.1	124	18.6	492.2	6.8	402.2	6.4	316.8	10.08	386.8	9	493	6.6	
	1ST QUART	293	9.8	465.75	6.5	130	30.6	521	7.4	457	7	335	12.5	398	10.6	549.75	7.2	
	3RD QUART	931.25	31.3	972.25	13.6	221	52.1	1490	21.1	1212	18.4	1195	44.5	1644.5	43.6	1596.5	21	
	%TOTAL	16.05		3.42		38.47		4.16		6.27		23.83		5.91		1.90		100.00
	N	5138		1041		13721		1104		2155		8364		2102		568		34193
	MEAN	607.5	21.5	635.1	13.6	354.9	32.6	1154.6	12.4	847.7	14.2	798.9	33.4	978.4	30.3	1116.613	14.11954	
TO	MEDIAN	381	24	562	11.3	188	36	1119.5	9.8	803	10.5	393	37.9	477	36.4	1022.5	11.25	
STATION	MIN	94	5	85	5	58	5	106	5	71	5	128	5	161	5	113	5	
14	MAX	1832	97.5	1262	74.4	1353	116.7	2195	103.6	1685	118.7	2980	116.5	3472	107.9	2305	102.1	
	20%ile	977	9.4	985	6.4	577	11.7	1711	6.4	1322	6.4	1417.4	10.5	1709	10.2	1701.2	6.8	
	1ST QUART	317.0	11.5	354.0	6.8	145.0	14.7	678.8	6.9	414.0	6.8	314.0	13.5	399.0	12.7	641	7.4	
	3RD QUART	795.8	28.9	924.0	17.9	459.0	46.7	1594.3	16.2	1237.0	20.4	1107.3	47.5	1371.3	43.5	1566	18	
	%TOTAL	15.03		3.04		40.13		3.23		6.30		24.46		6.15		1.66		100.00

Appendix 4D: One-many origin-destination matrix in the Wigan network

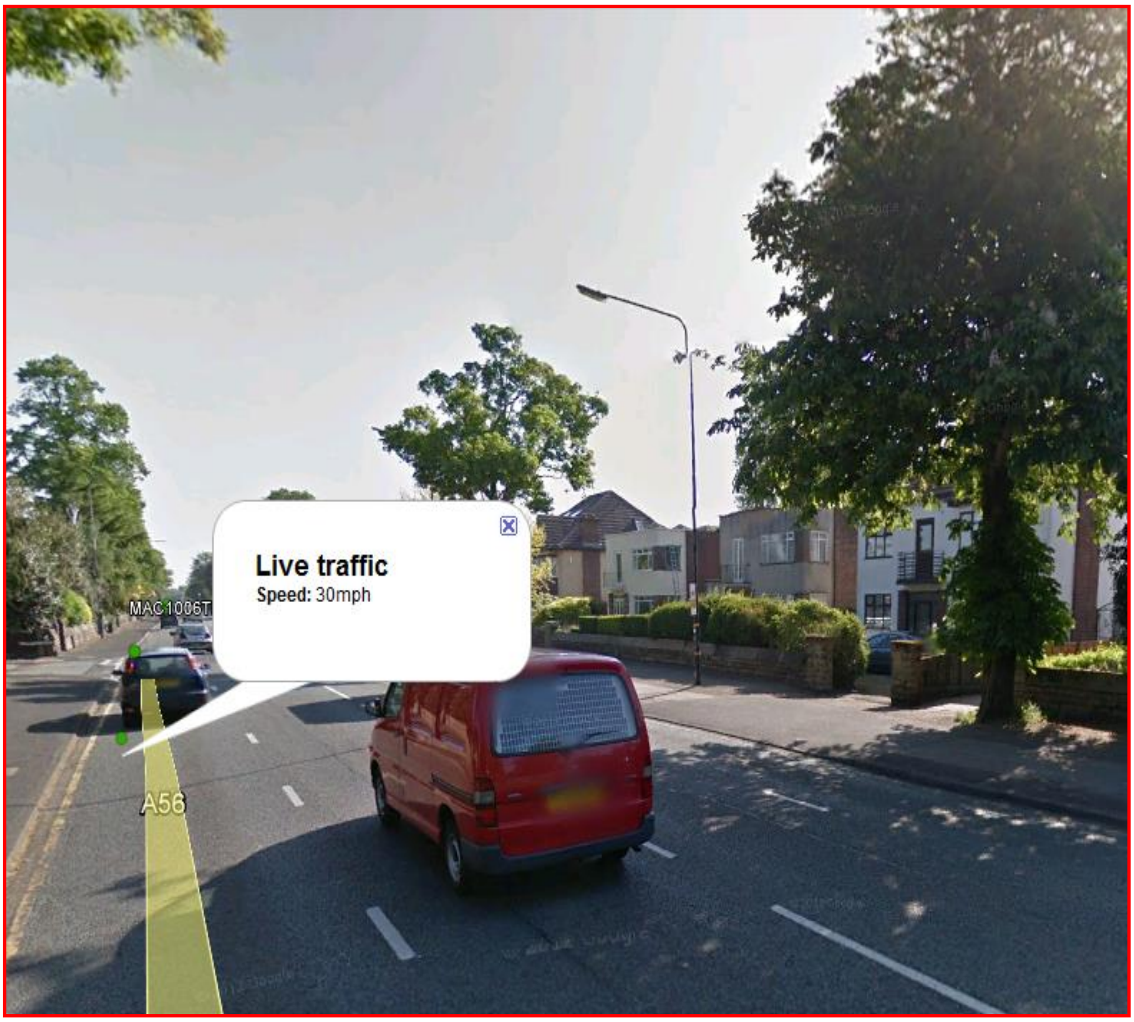
	STATION	34		35		36		37		38		39		40		41		
UNITS	(SEC, KM/H)	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TIME	SPEED	TOTAL
	N	29018		22010		50330		24786		21641		46433		20993		31000		246211
	MEAN	61.0	38.5	123.8	45.8	281.3	44.2	404.8	39.0	449.0	36.9	544.4	34.9	586.5	36.0	741.5	33.9	
FROM	MEDIAN	48	38.3	94	46.7	174	46.6	250	41	281	38.9	341	36.7	371	38	461	36	
STATION	MIN	16	5	38	5	68	5	89	5	102	5	105	5	119	5	141	5	
33	MAX	367	114.9	878	115.6	1622	119.3	2051	115.3	2183	107	2503	119.2	2816	118.4	3318	117.7	
	20%ile	35	23.9	74	34.1	137	28.9	194	24.1	220	22.9	265	20.2	292	22.4	365	20.4	
	1ST QUART	37	25.9	77	36.3	143	34.2	204	29	230	27.6	278	25.1	306	26.9	383	24.8	
	3RD QUART	71	49.7	121	57.1	237	56.7	354	50.3	395	47.5	498	45	523	46	670	43.3	
	%TOTAL	11.80		8.90		20.40		10.10		8.80		18.90		8.50		12.60		100.00
	N	31397		22189		54499		25486		22603		48861		21155		31792		257982
	MEAN	78.5	31.4	154.0	38.7	279.5	41.8	405.4	37.8	435.7	35.8	535.4	33.6	587.0	34.1	673.7758	35.56055	
TO	MEDIAN	63	29.2	113	38.9	188	43.2	257	39.9	292	37.4	361	34.7	398	35.4	450	36.9	
STATION	MIN	16	5	37	5	68	5	99	5	103	5	107	5	118	5	142	5	
33	MAX	367	114.9	878	118.8	1622	119.3	2051	103.6	2183	106	2503	117	2817	119.4	3318	116.8	
	20%ile	97	18.9	175	25.1	291	27.9	442	23.2	457	23.9	592	21.1	638	22.1	706.8	23.5	
	1ST QUART	44.0	20.9	88.0	28.4	153.0	32.7	214.0	29.2	245.0	29.0	297.0	25.1	332.0	26.3	376	27.8	
	3RD QUART	88.0	41.8	155.0	49.9	248.0	53.0	351.0	47.9	376.0	44.6	499.0	42.1	536.0	42.4	596	44.1	
	%TOTAL	12.17		8.60		21.13		9.88		8.76		18.94		8.20		12.32		100.00

Appendix 4E: Journey times, speeds and O-D matrix in Stockport network

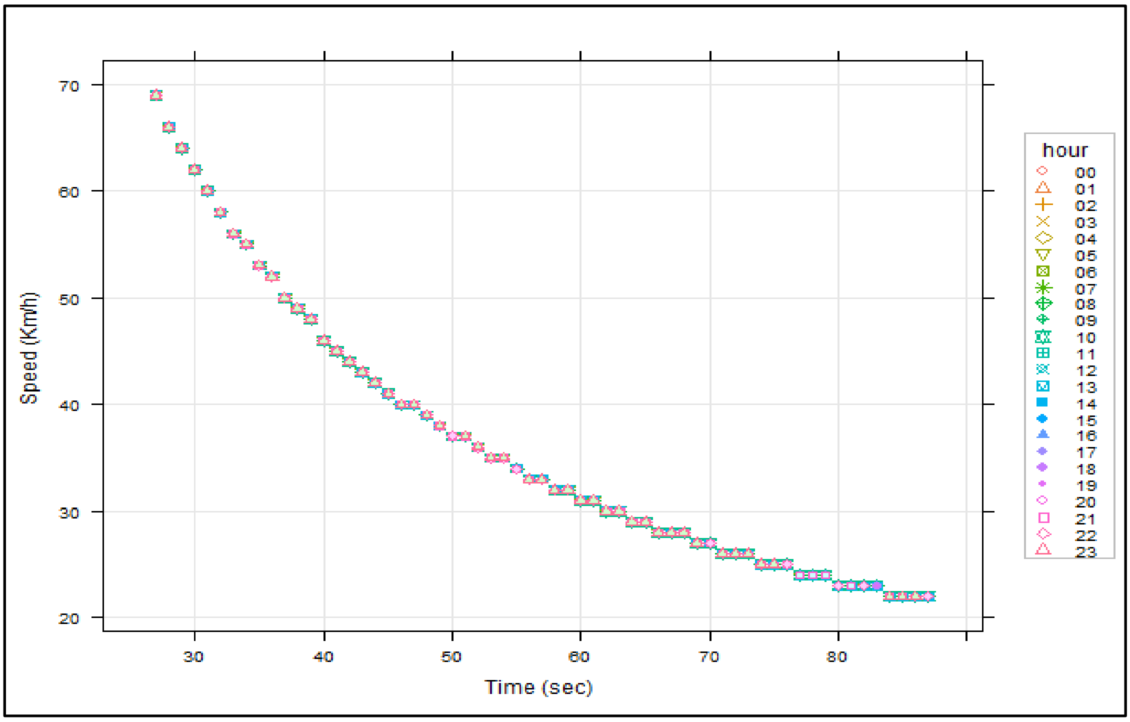


Appendix 4F: Speed distribution over hours of the day from Station 1 to Station 2 in Trafford

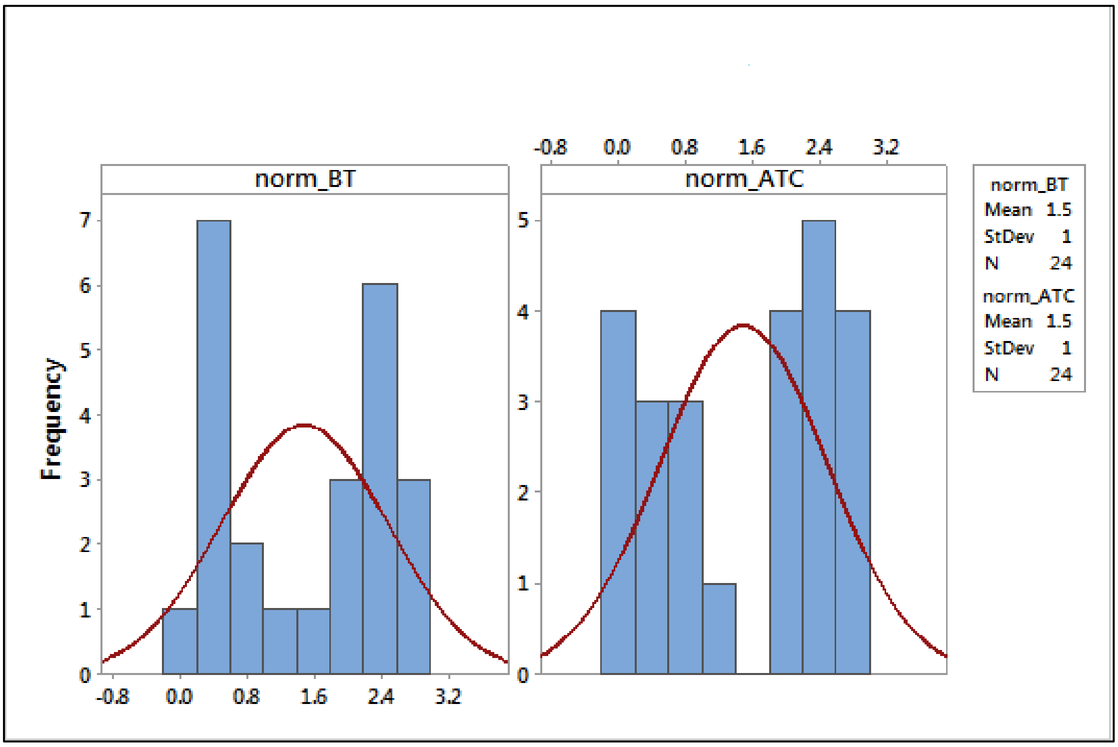
Appendix 5



Appendix 5A: Validation of journey speed with live traffic information on the A56 Washway Road, Trafford



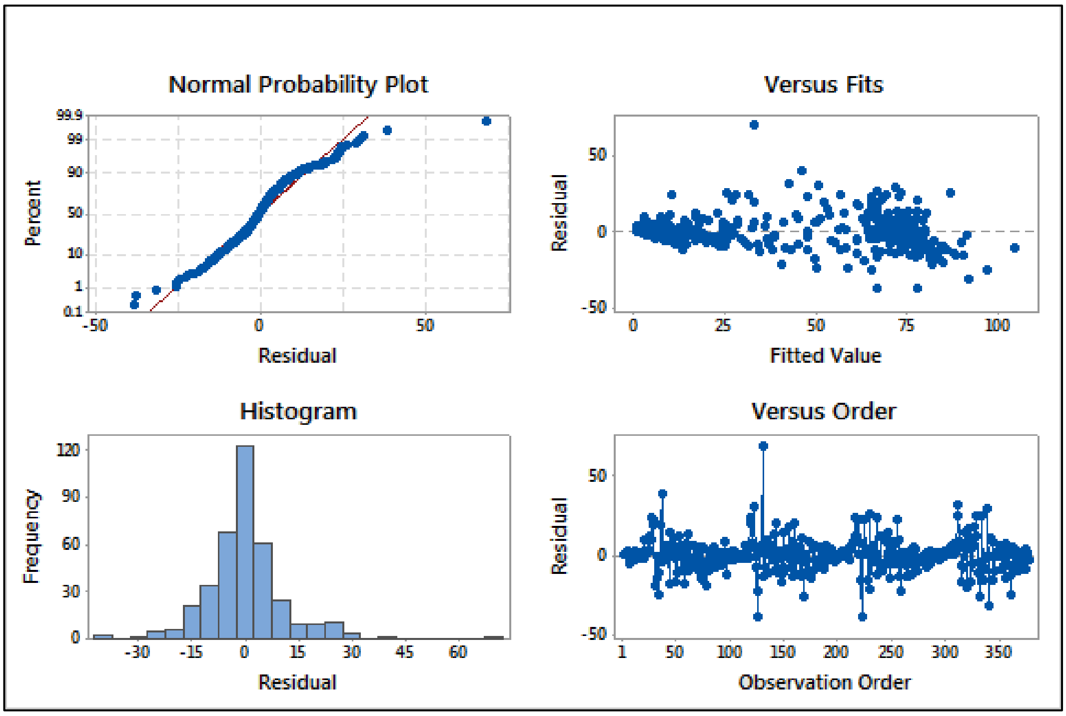
Appendix 5B: Scatter plot of speed against time grouped by hour on link3435 in Stockport



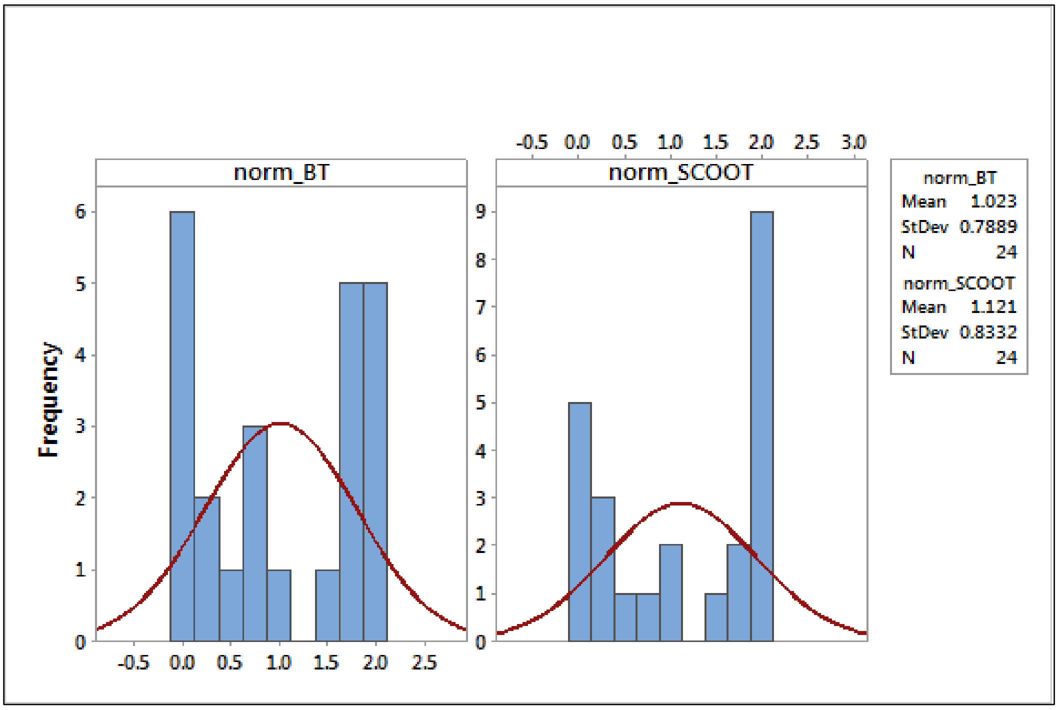
Appendix 5C-1: Histogram plots of normalised flows of ATC and Bluetooth overlaid with normal curve

Descriptive Statistics: norm_BT, norm_ATC										
Variable	Total Count	Mean	SE Mean	StDev	CoefVar	Minimum	Q1	Median	Q3	Maximum
norm_BT	24	1.5	0.204	1	66.67	0.19	0.389	1.709	2.496	2.811
norm_ATC	24	1.5	0.204	1	66.67	0.119	0.336	2.014	2.267	2.804

Appendix 5C-2: Descriptive statistics of normalised flows of ATC and Bluetooth



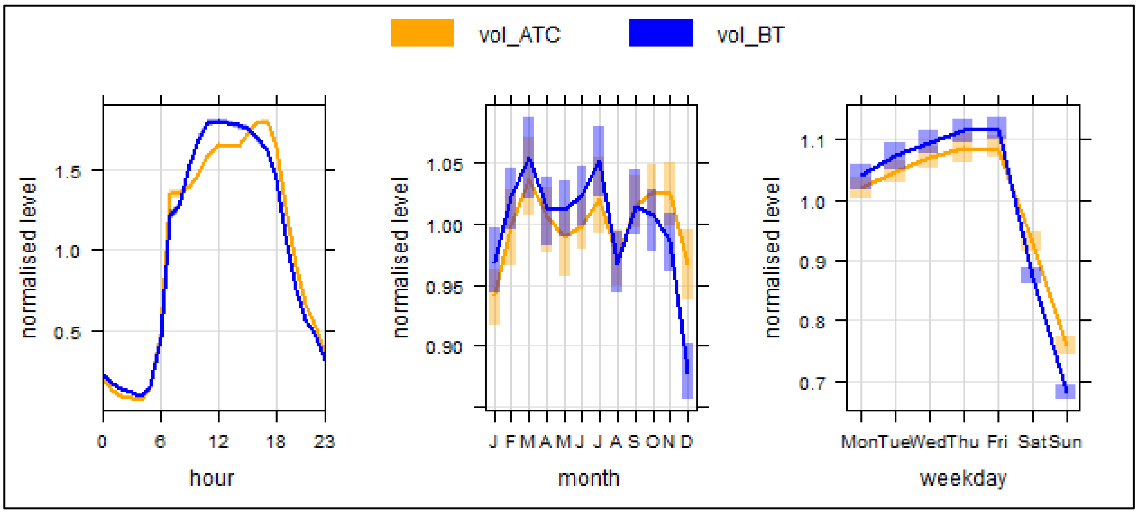
Appendix 5C-3: Diagnostics plots of Bluetooth flows for all Mondays in November (N=378)



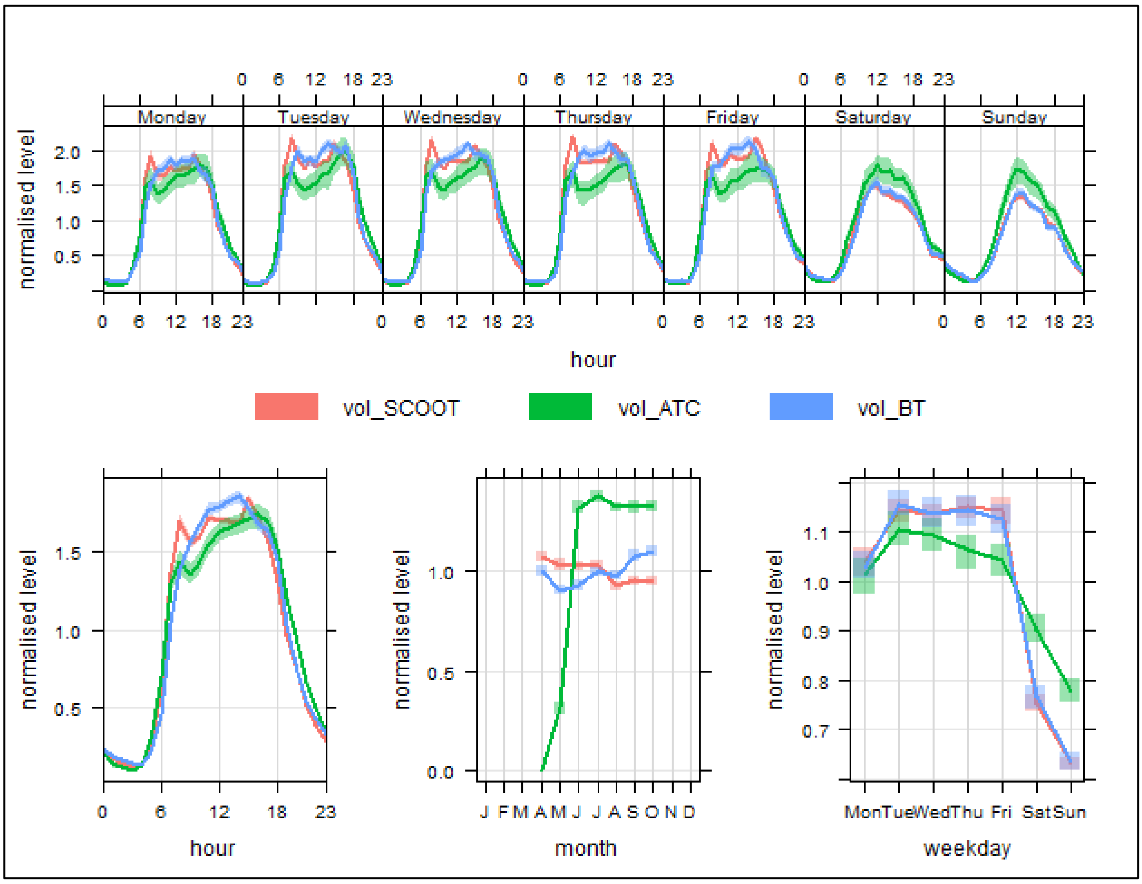
Appendix 5C-4: Histogram plots of normalised flows of SCOOT and Bluetooth overlaid with a density curve

Descriptive Statistics: norm_BT, norm_SCOOT										
Variable	Total Count	Mean	SE Mean	StDev	CoefVar	Minimum	Q1	Median	Q3	Maximum
norm_BT	24	1.023	0.161	0.789	77.1	0	0.147	0.947	1.783	2
norm_SCOOT	24	1.121	0.17	0.833	74.35	0	0.187	1.216	1.962	2

Appendix 5C-5: Descriptive statistics of normalised flows of SCOOT and Bluetooth



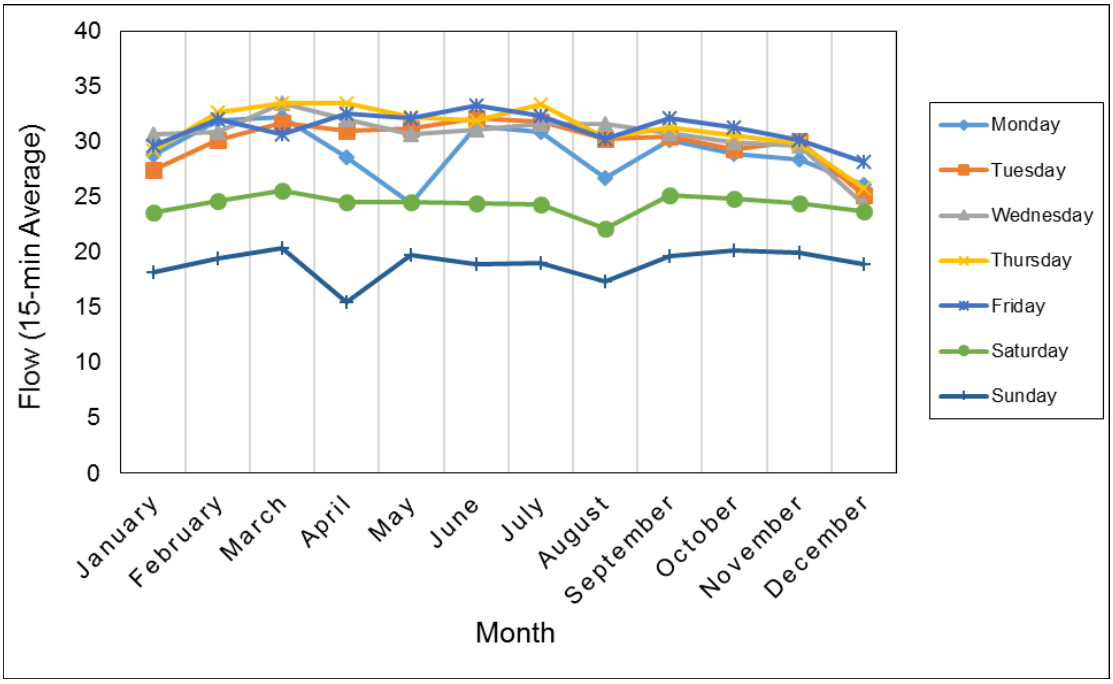
Appendix 5D-1: Flow profiles of Bluetooth and ATC on Link0506 in Trafford (N=33,646)



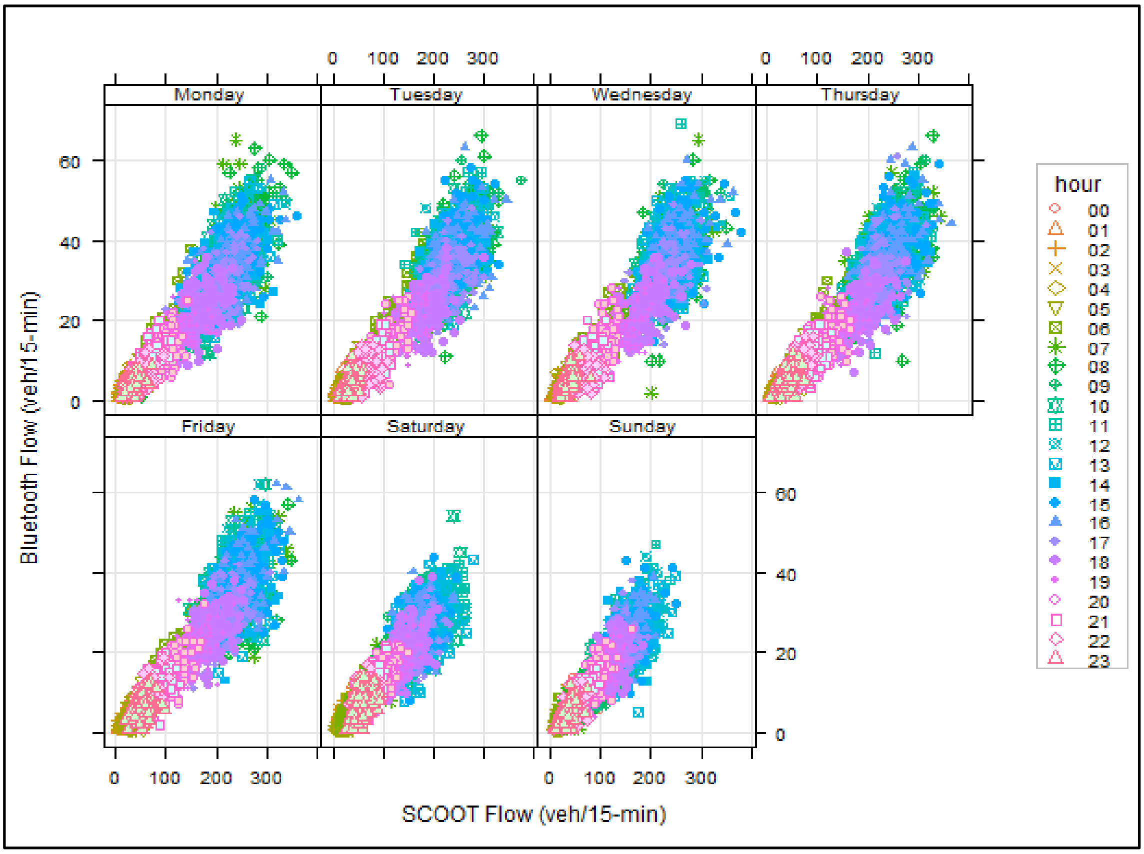
Appendix 5D-2: SE-directional flow profiles on link3435 in Stockport (N=18,761)

cor(julBAST2[,2:7])						
	BT_NW	ATC_NW	BT_SE	ATC_SE	SCT_NW	SCT_SE
BT_NW	1					
ATC_NW	0.84	1				
BT_SE	0.84	0.78	1			
ATC_SE	0.85	0.94	0.85	1		
SCT_NW	0.95	0.88	0.92	0.90	1	
SCT_SE	0.96	0.87	0.92	0.90	0.98	1

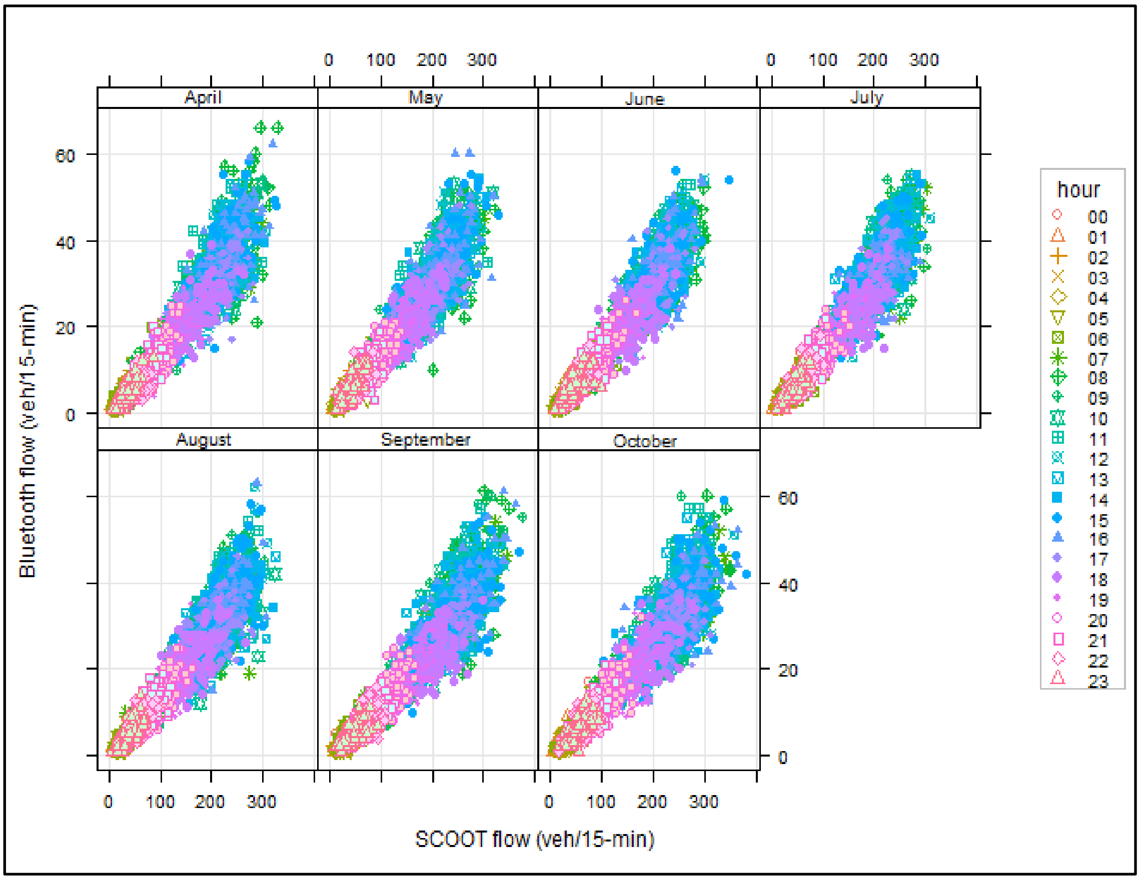
Appendix 5D-3: Table of correlation coefficients between the measured flows in both directions



Appendix 5D-4: Profiles of Bluetooth monthly-weekday flows on Link0506



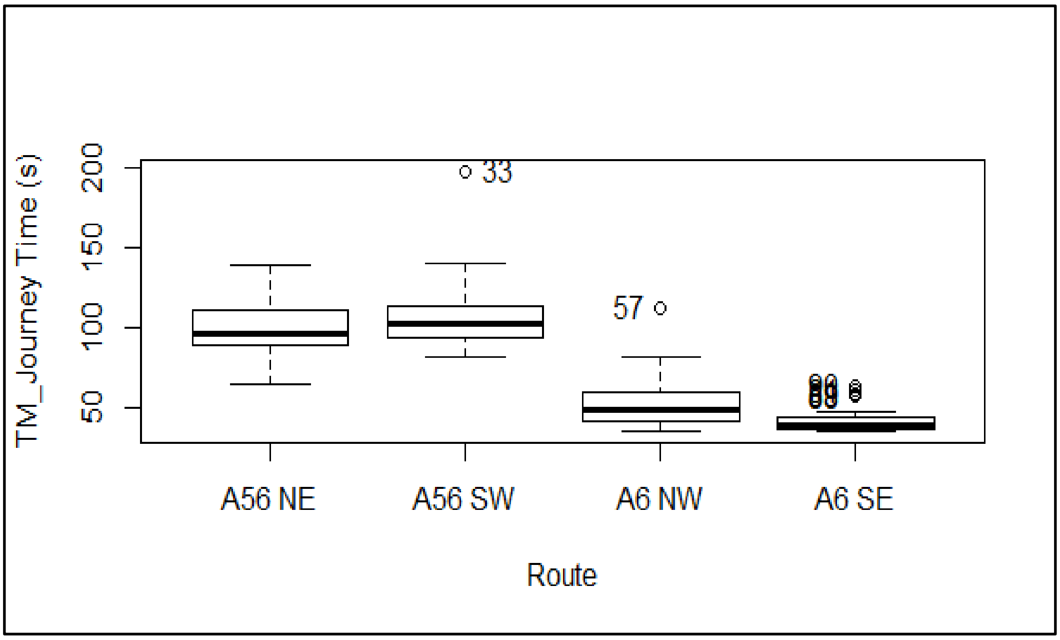
Appendix 5D-5: Weekday scatter plot of Bluetooth against SCOOT flow (NW-bound)



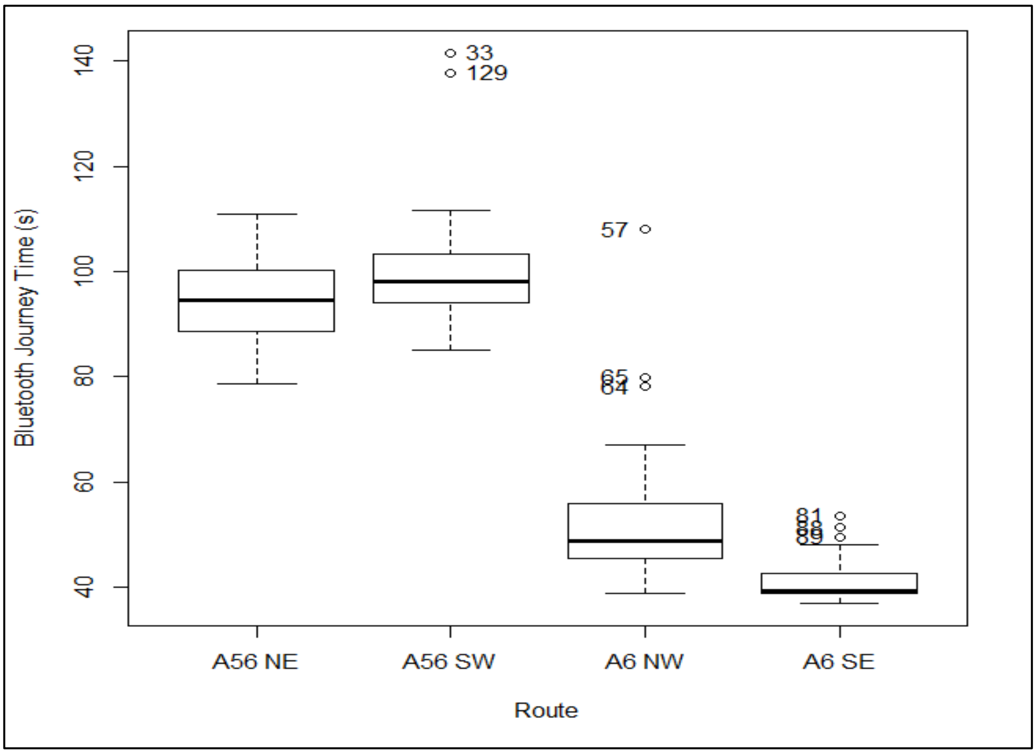
Appendix 5D-6: Monthly scatter plot of Bluetooth against SCOOT flow (NW-bound)

Descriptive Statistics: BT_Flow, ANPR_Flow, BT_jtime, ANPR_jtime, BT_speed, ANPR_speed										
Variable	Total Count	Mean	SE Mean	StDev	CoefVar	Minimum	Q1	Median	Q3	Maximum
BT_Flow	48	24.521	0.859	5.95	24.27	10	20.25	25	28	38
ANPR_Flow	48	70.44	3.34	23.14	32.84	40	55.25	64	83.75	143
BT_jtime	48	123.54	3.51	24.34	19.7	77	106	119	137.5	200
ANPR_jtime	48	108.75	3.07	21.24	19.53	66	94.25	105.5	118.5	171
BT_speed	48	18.208	0.528	3.661	20.11	10	16	18	20	28
ANPR_speed	48	19.813	0.532	3.682	18.59	12	18	19.5	22	31

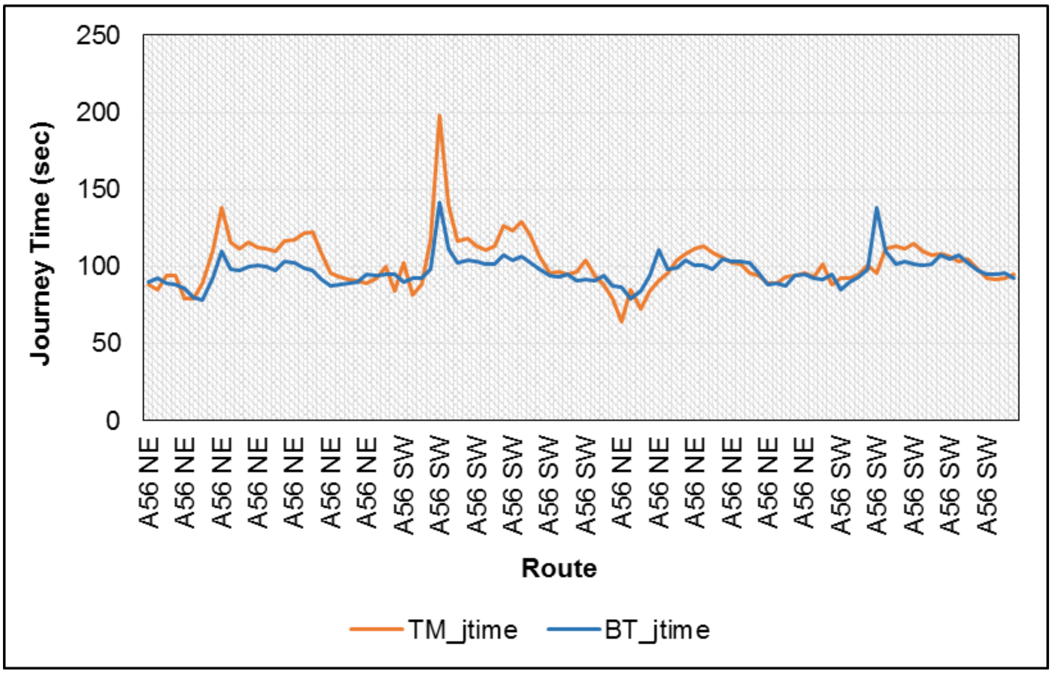
Appendix 5E: Table of descriptive statistics for flow, journey times and vehicle speeds



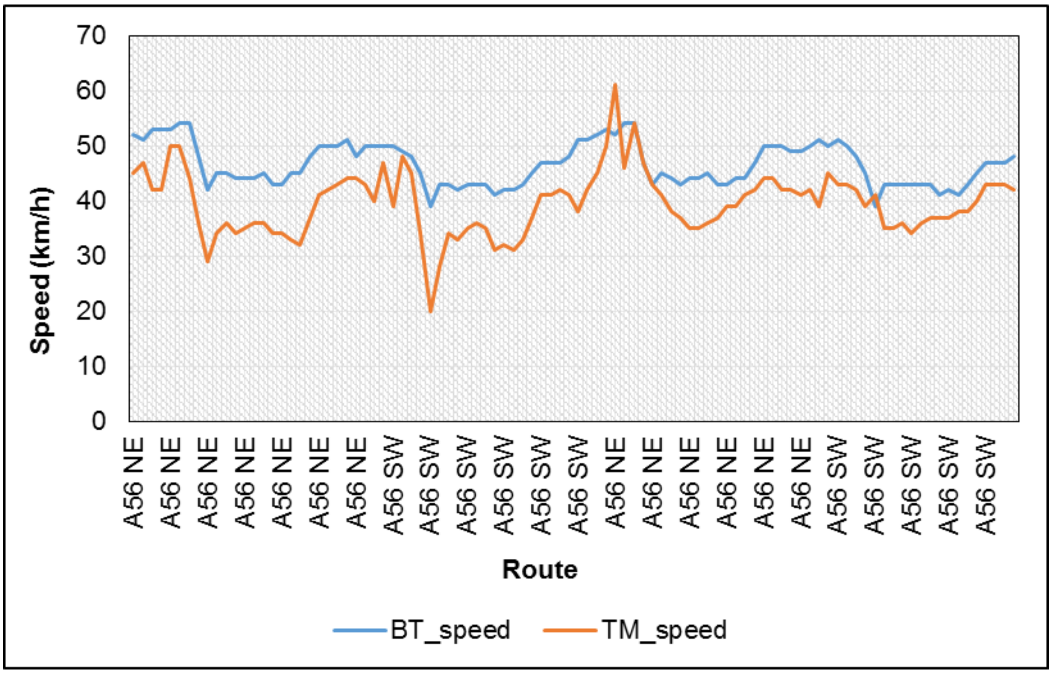
Appendix 5F-1: Boxplot of TM journey times over four routes in GMN



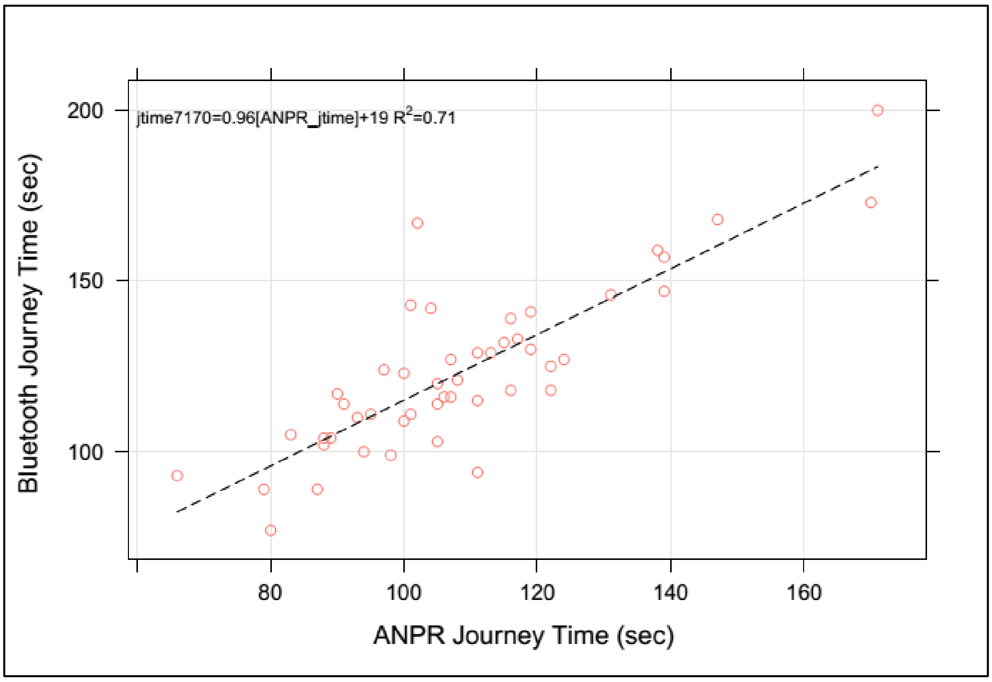
Appendix 5F-2: Boxplot of Bluetooth journey times over four routes in GMN



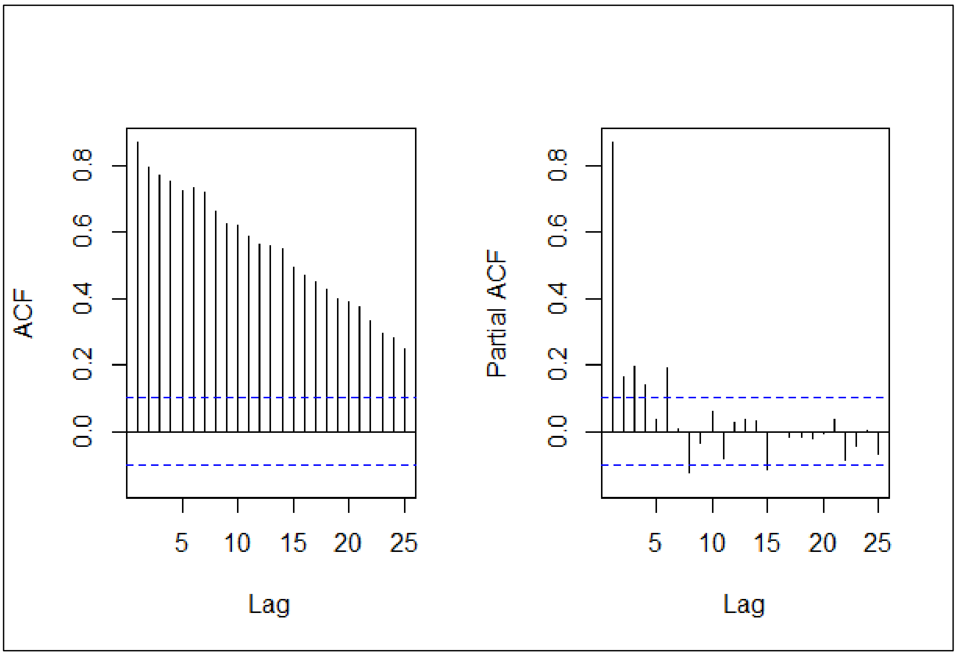
Appendix 5F-3: Profiles of Bluetooth and TM journey times over six months by Routes in Trafford (N=96)



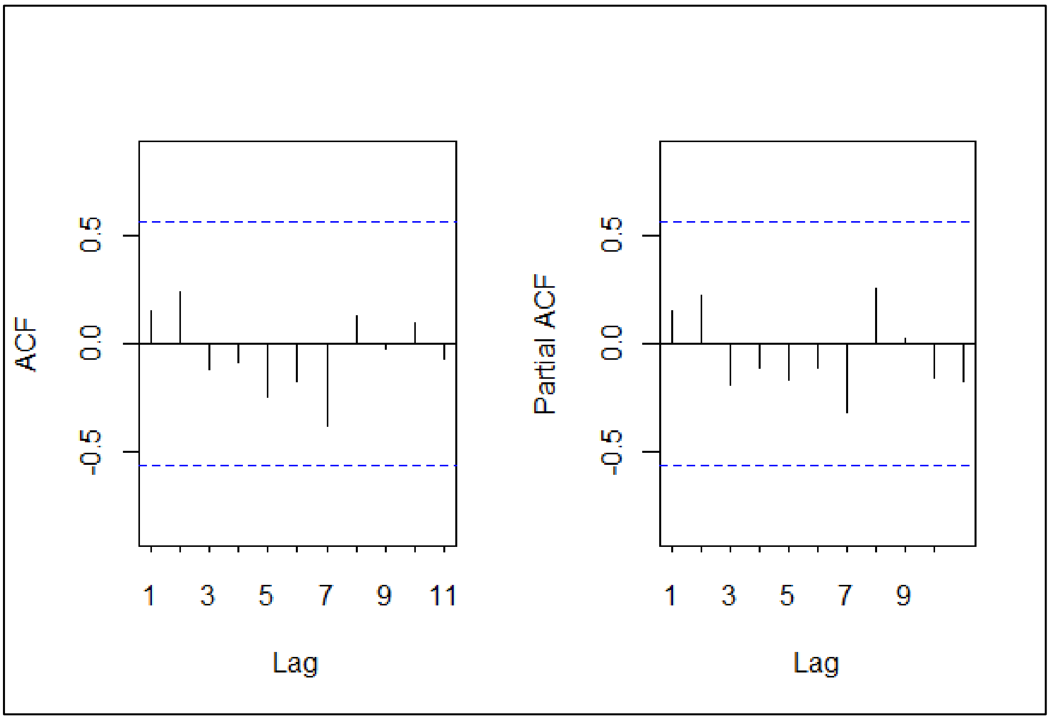
Appendix 5F-4: Profiles of Bluetooth and TM journey times over six months by Routes in Trafford (N=96)



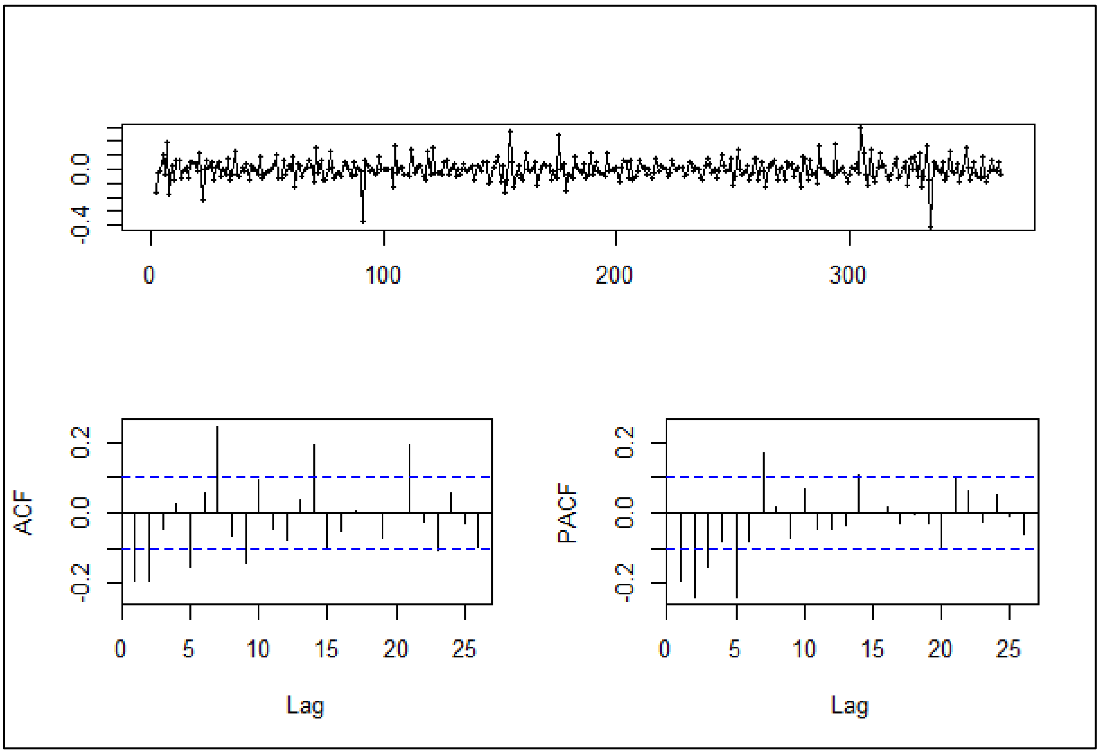
Appendix 5G: Scatter plot of Bluetooth against ANPR journey times (overlaid with regression line) of 3rd April 2014 on Link7170 in Stockport



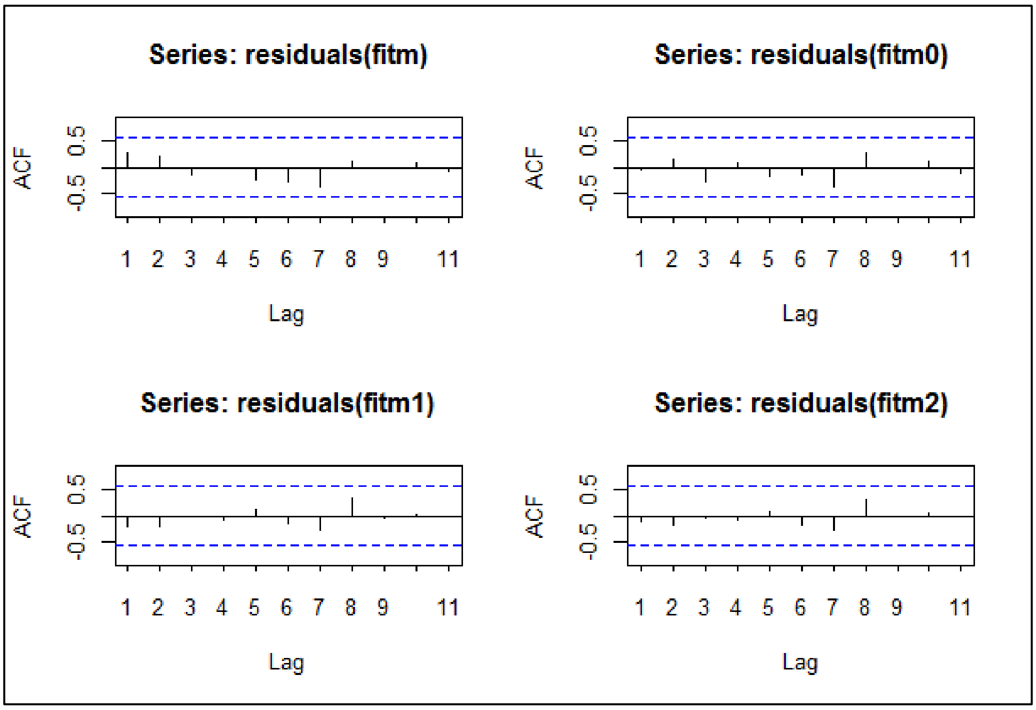
Appendix 5H: Plot of Autocorrelation and Partial Autocorrelation Function of journey times before transformation



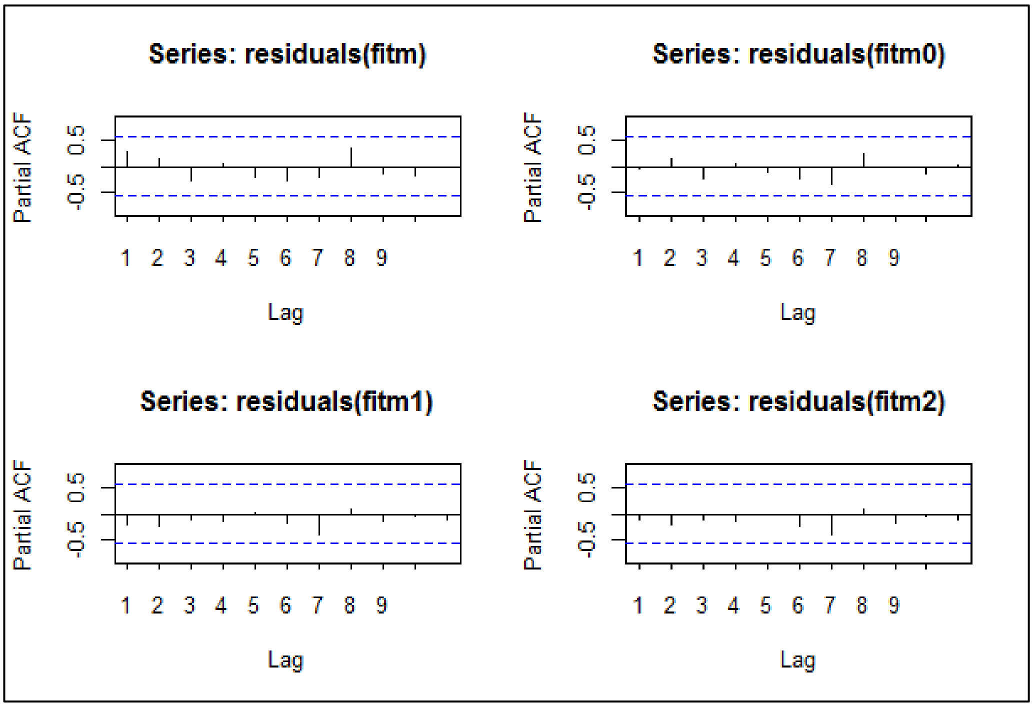
Appendix 5I-1: ACF and PACF from monthly journey times modelling



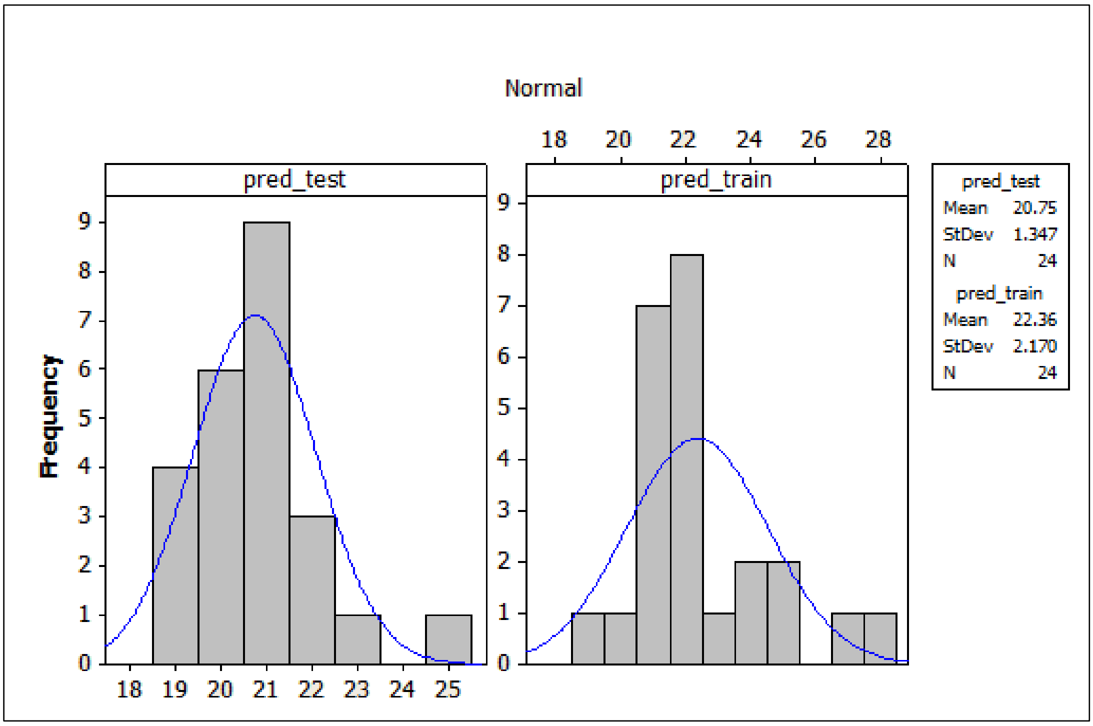
Appendix 5I-2: ACF, PACF, and Residuals plots after transformation of journey times



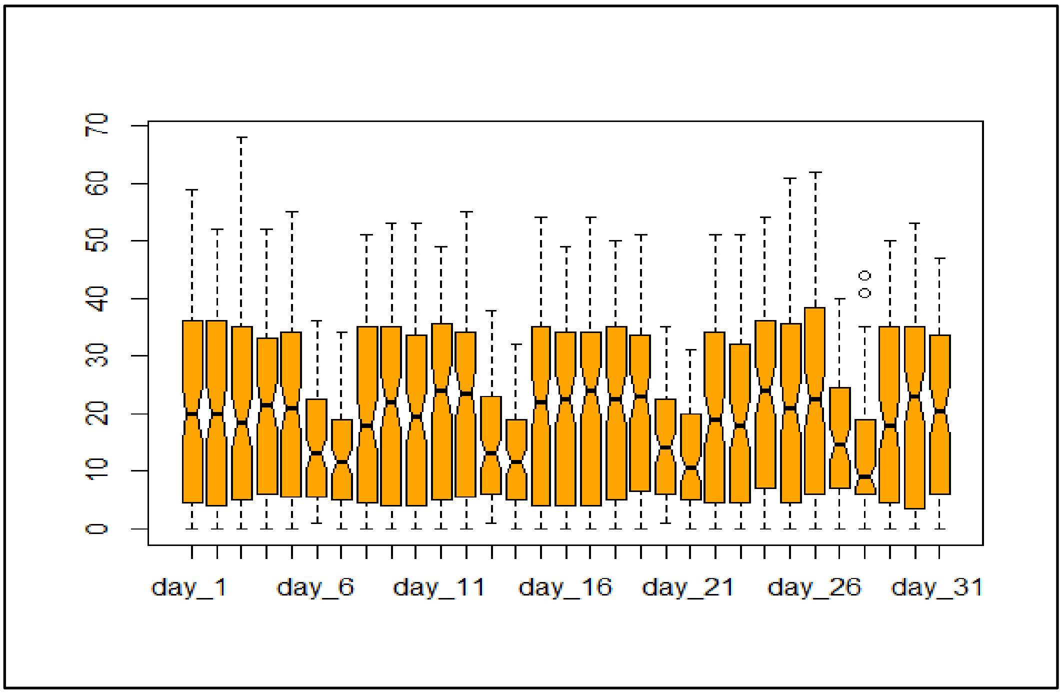
Appendix 5I-3: Plot of Autocorrelation Function of flow for different ARIMA models



Appendix 5I-4: Plot of Partial Autocorrelation Function of flow for different ARIMA models

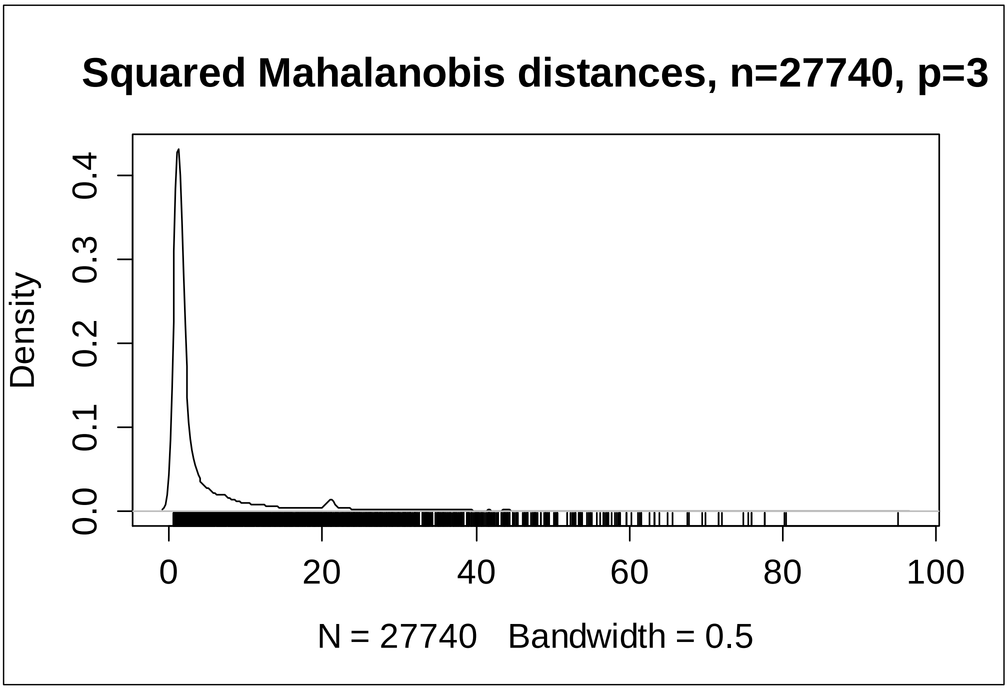


Appendix 5I-5: Histogram plots of the forecast and validation data over 24 days

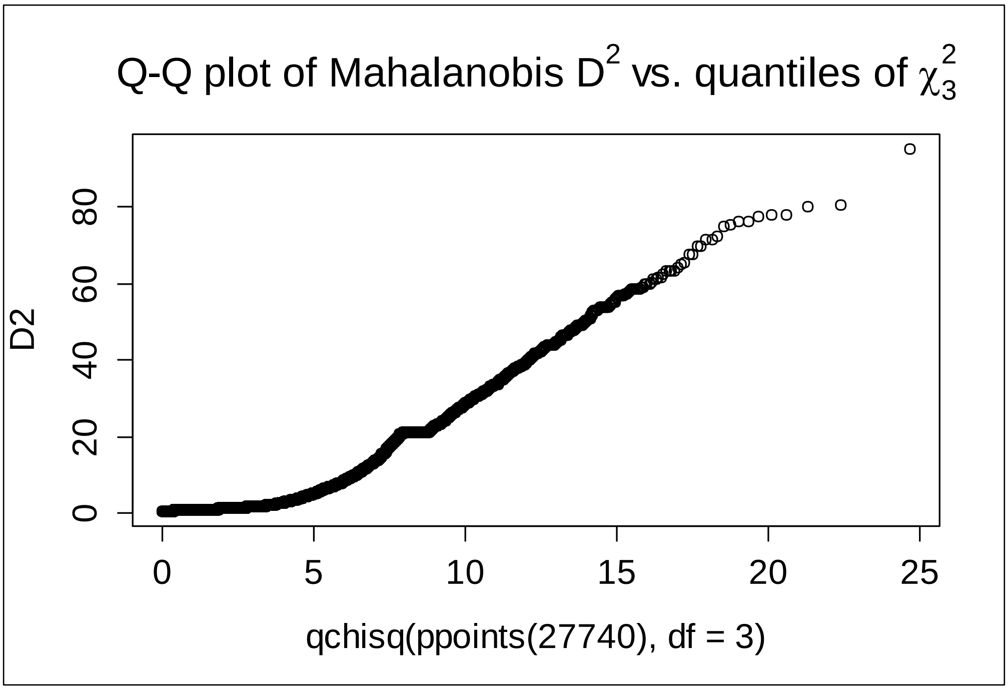


Appendix 5I-6: Box and Whisker plot of Northbound Bluetooth flow for July 2013 on Link3637, Buxton Road

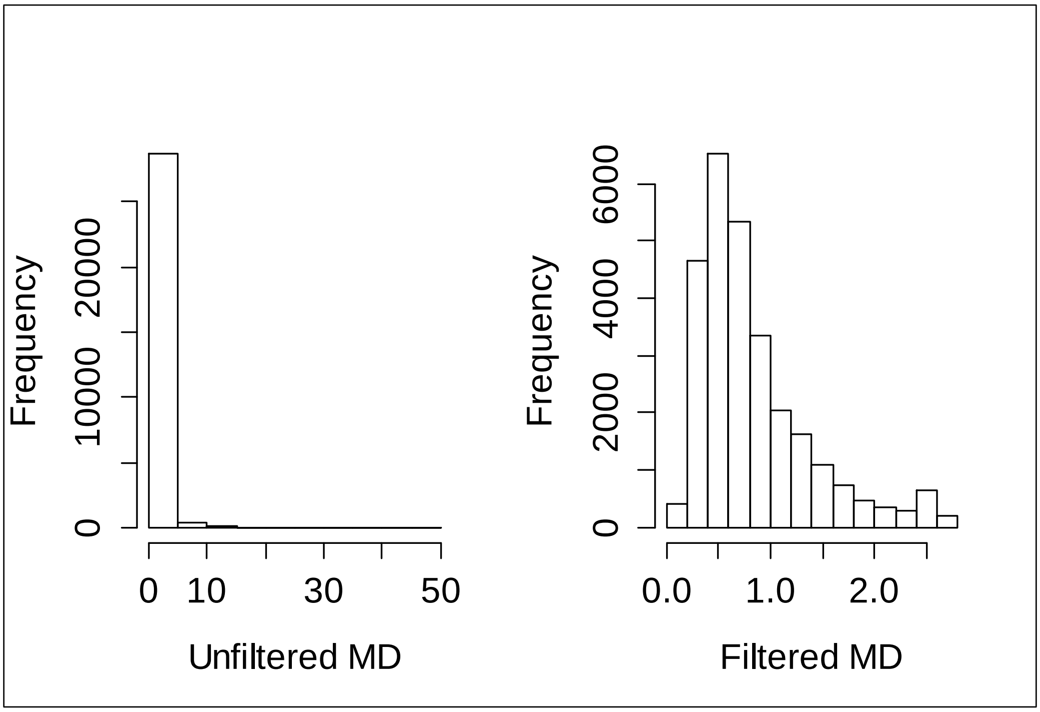
Appendix 6



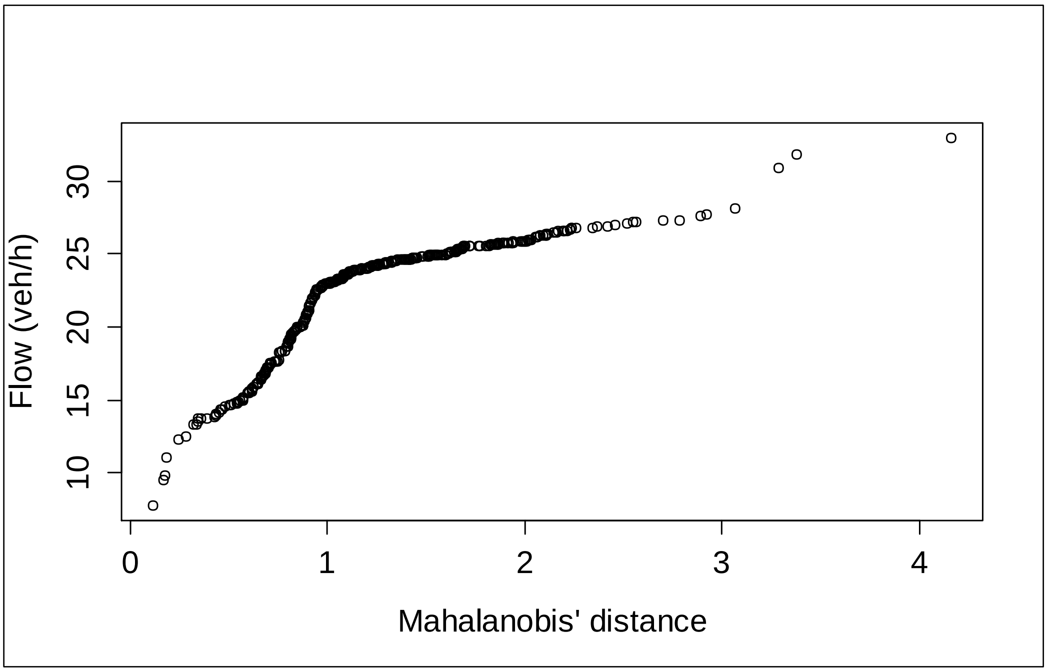
Appendix 6A-1: Density plot of squared of Mahalanobis distances



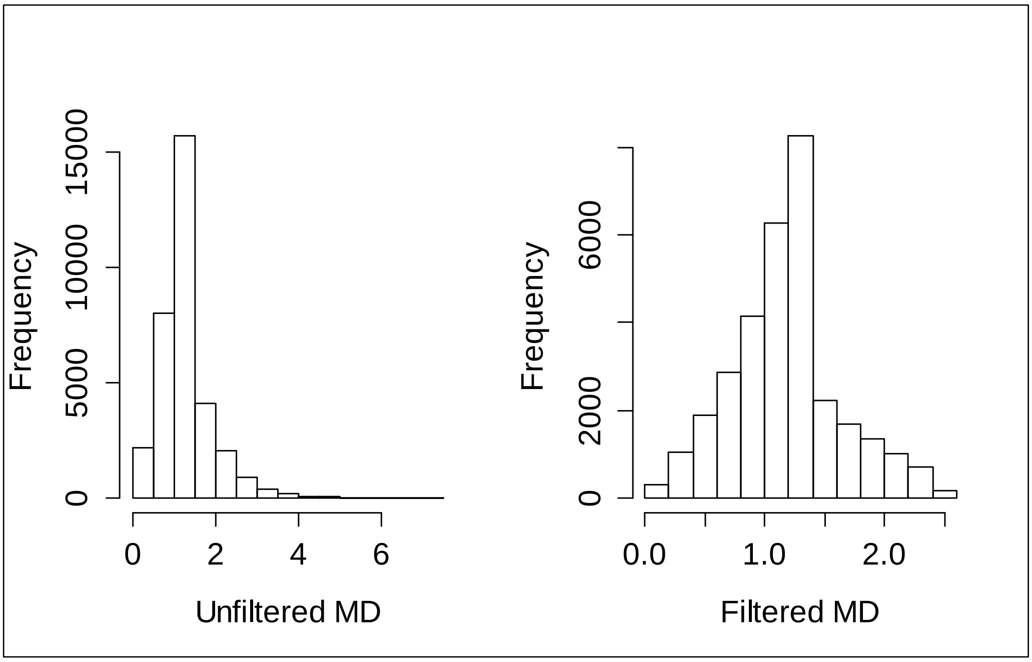
Appendix 6A-2: Q-Q plot of Squared of Mahalanobis distance against quantiles of Chi-square of degree of freedom 3



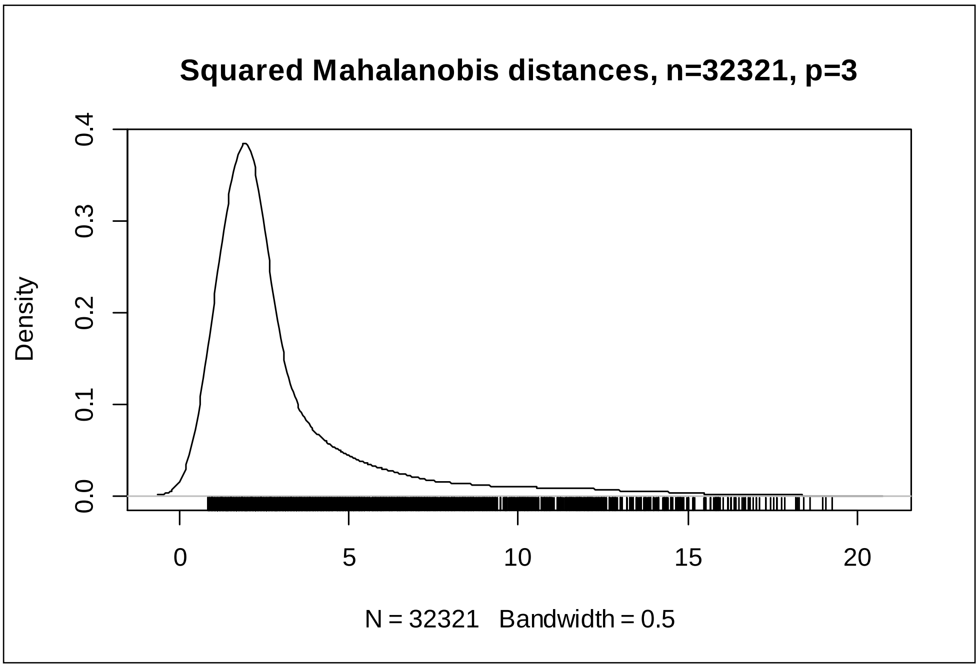
Appendix 6A-3: Histogram plots of Unfiltered (left) and Filtered (right) Mahalanobis distances



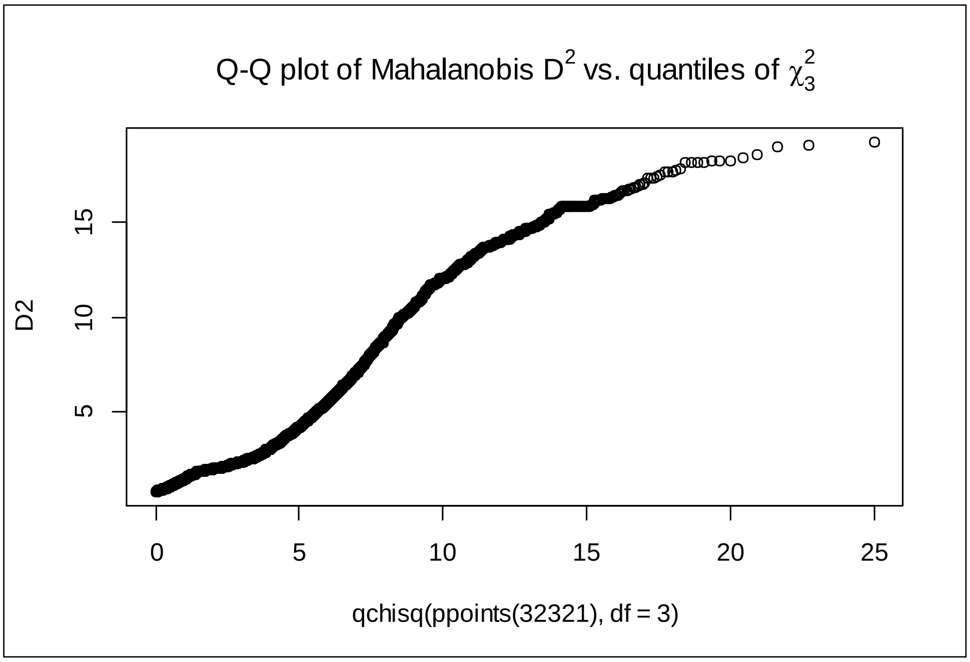
Appendix 6B-1: Plot of flow against Mahalanobis distances



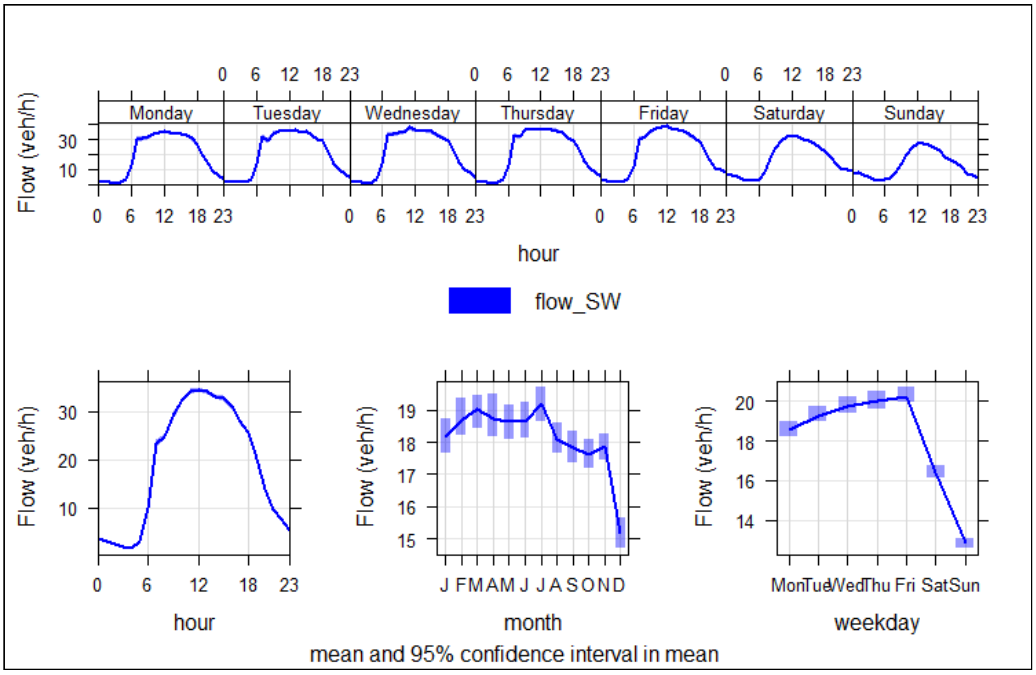
Appendix 6B-2: Histogram plots of Unfiltered (left) and Filtered (right) Mahalanobis distances



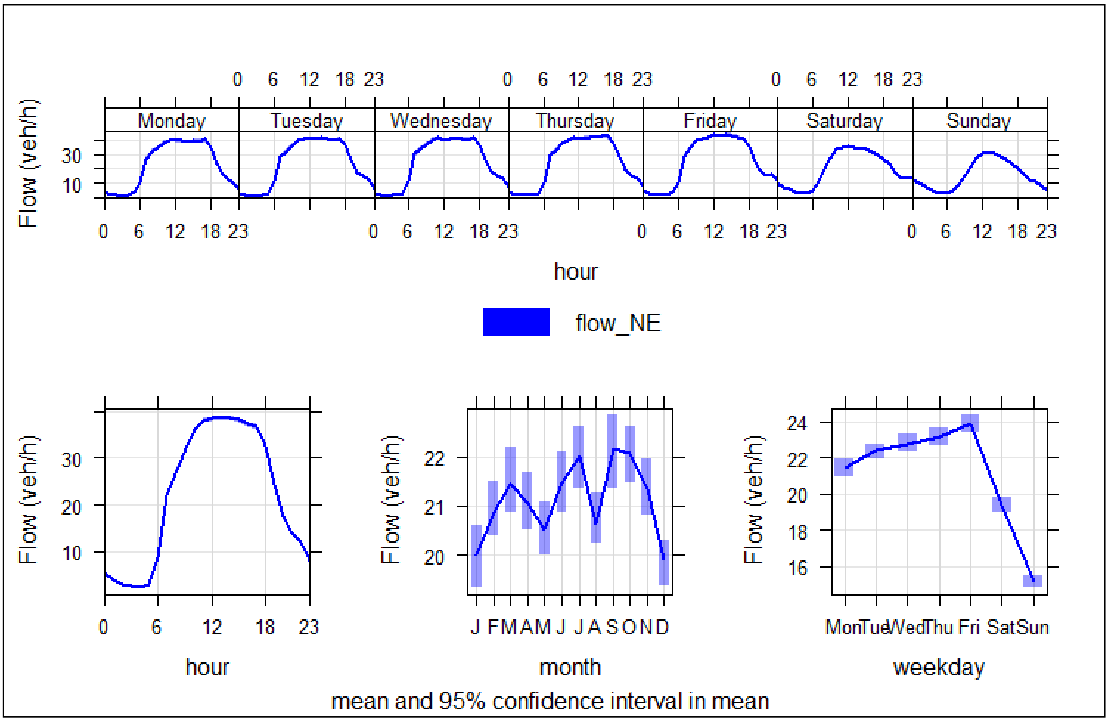
Appendix 6B-3: Density plot of squared of Mahalanobis distances



Appendix 6B-4: Q-Q plot of Squared of Mahalanobis distance against quantiles of Chi-square of degree of freedom 3



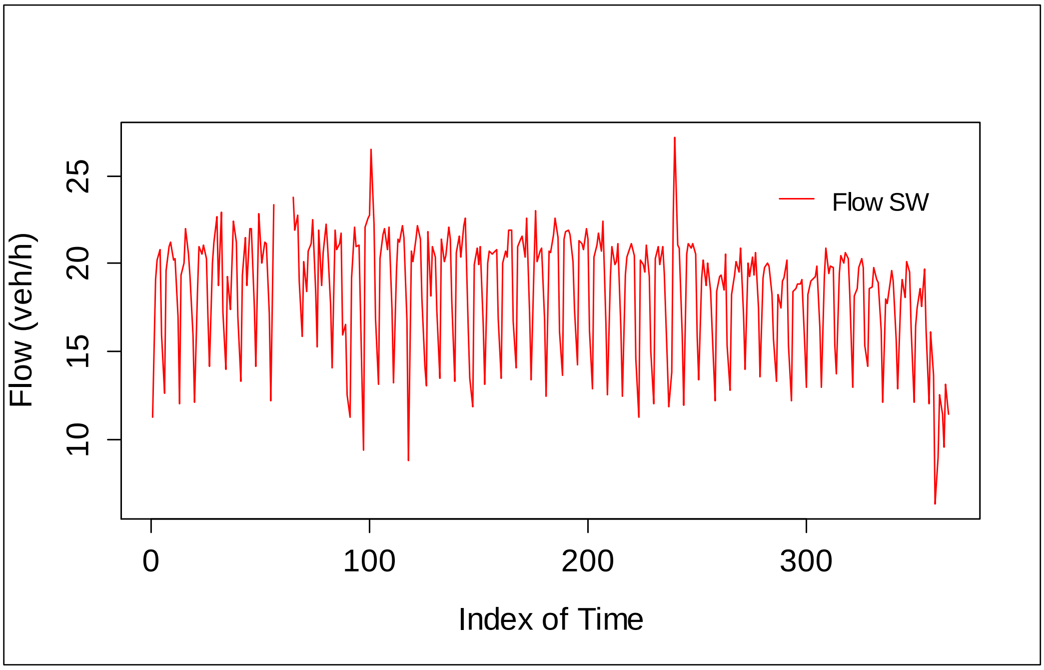
Appendix 6C-1: Time series plots of SW-flows on Link0506 in Trafford in 2013



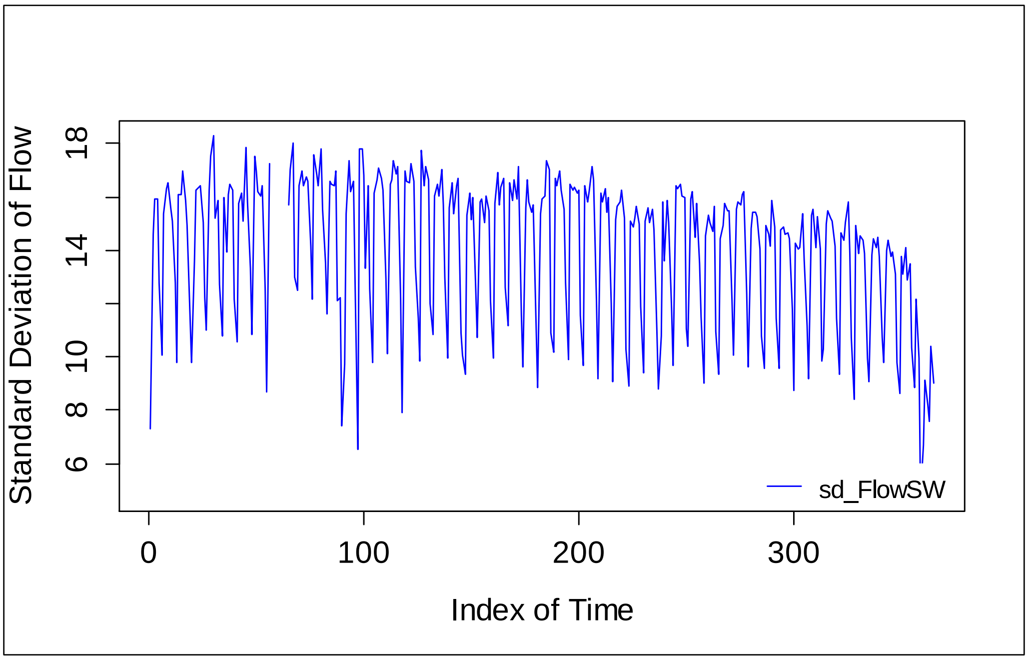
Appendix 6C-2: Time series plots of NE-flows on Link0506 in Trafford in 2013

Variables	Year Month	Adjusted R-Square		
		Southbound	Northbound	Combined Direction
BT/ATC	Jan	0.745	0.721	0.883
BT/ATC	Feb	0.737	0.761	0.884
BT/ATC	Mar	0.736	0.788	0.891
BT/ATC	Apr	0.769	0.757	0.879
BT/ATC	May	0.776	0.751	0.886
BT/ATC	Jun	0.765	0.796	0.887
BT/ATC	Jul	0.756	0.816	0.904
BT/ATC	Aug	0.775	0.817	0.901
BT/ATC	Sep	0.736	0.767	0.890
BT/ATC	Oct	0.724	0.767	0.879
BT/ATC	Nov	0.720	0.746	0.870
BT/ATC	Dec	0.733	0.768	0.887

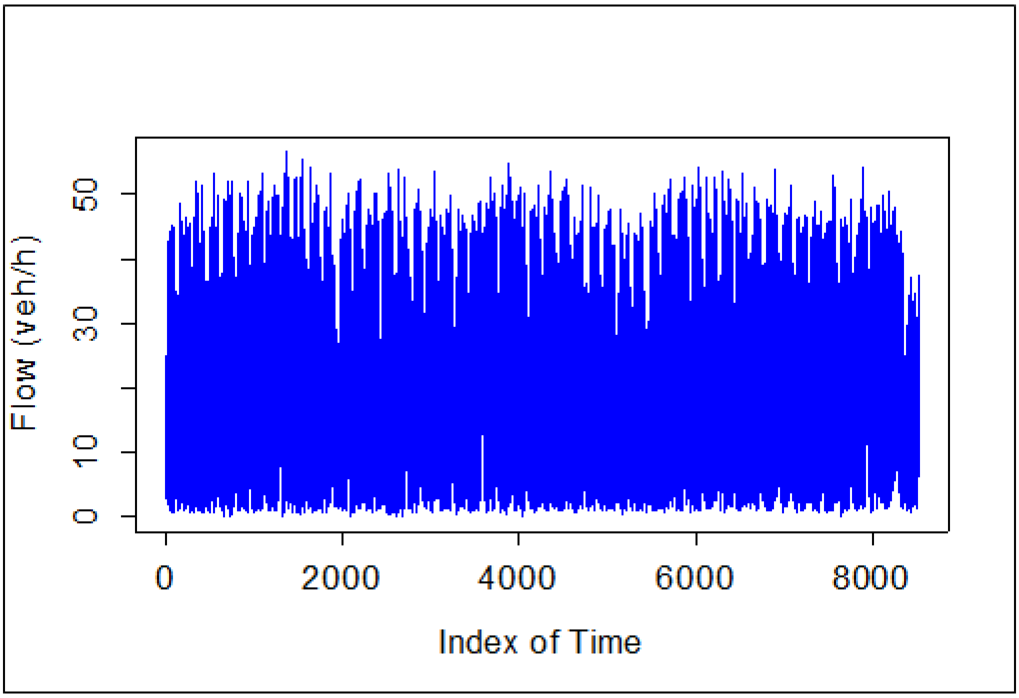
Appendix 6C-3: Table showing the monthly adjusted R-square for directional and combined flows on Link0506T in Trafford in 2013



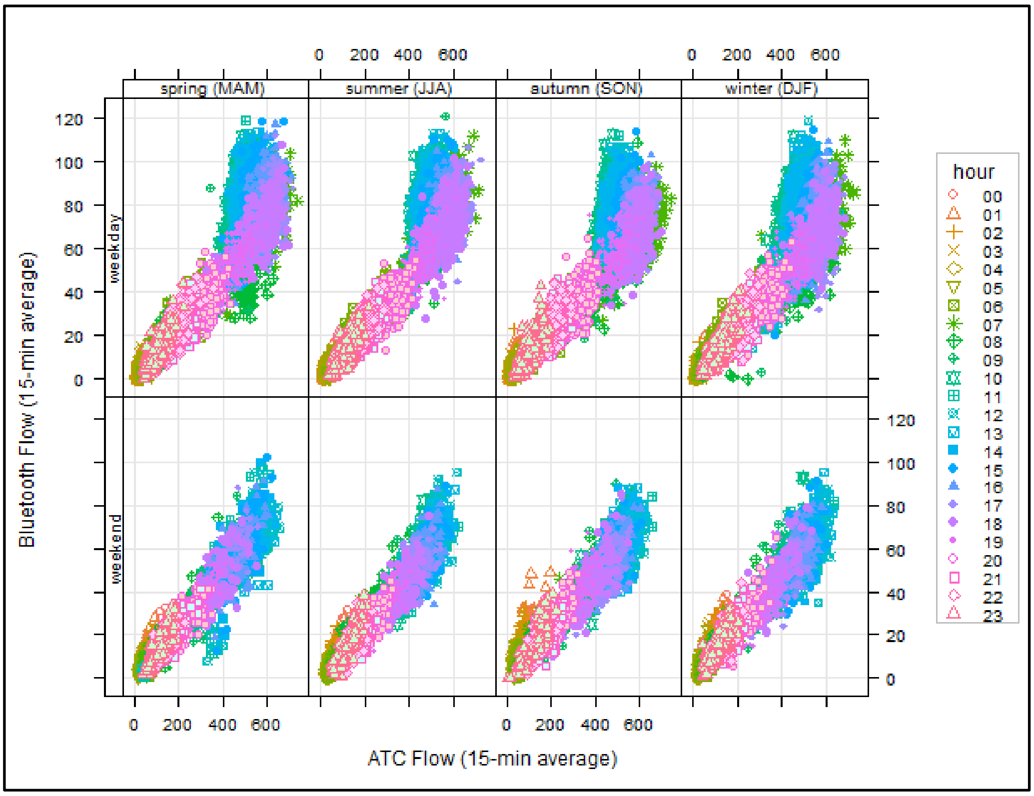
Appendix 6C-4: Day-by-day SW-directional flow over a year



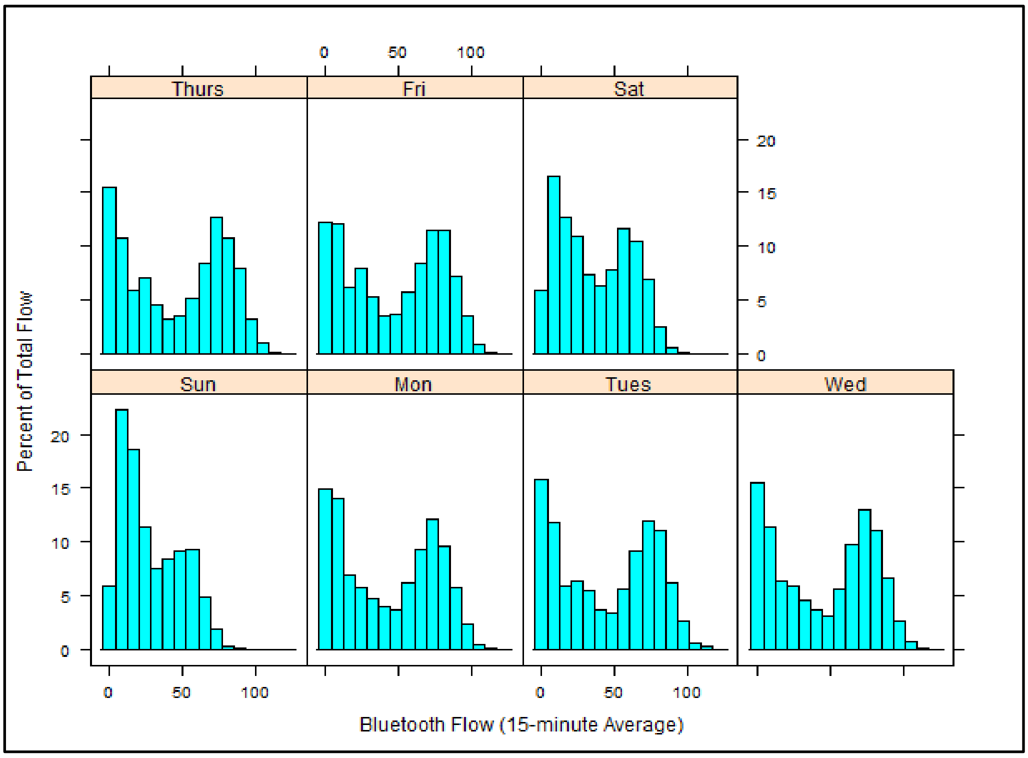
Appendix 6C-5: Standard deviation of SW-directional flow



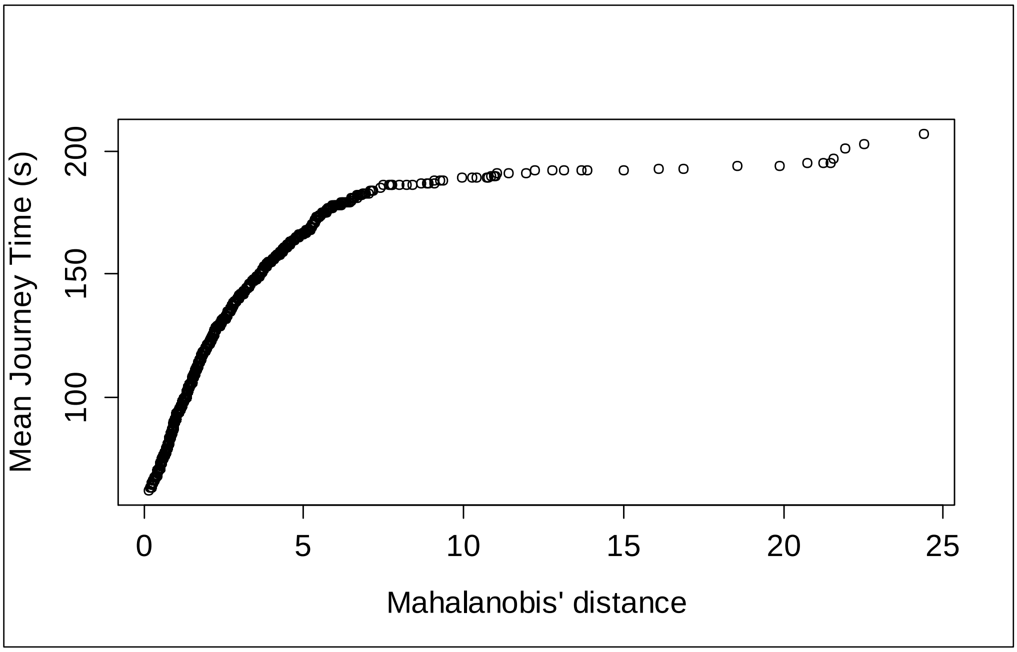
Appendix 6C-6: Hour-by-hour SW-directional flow over a year



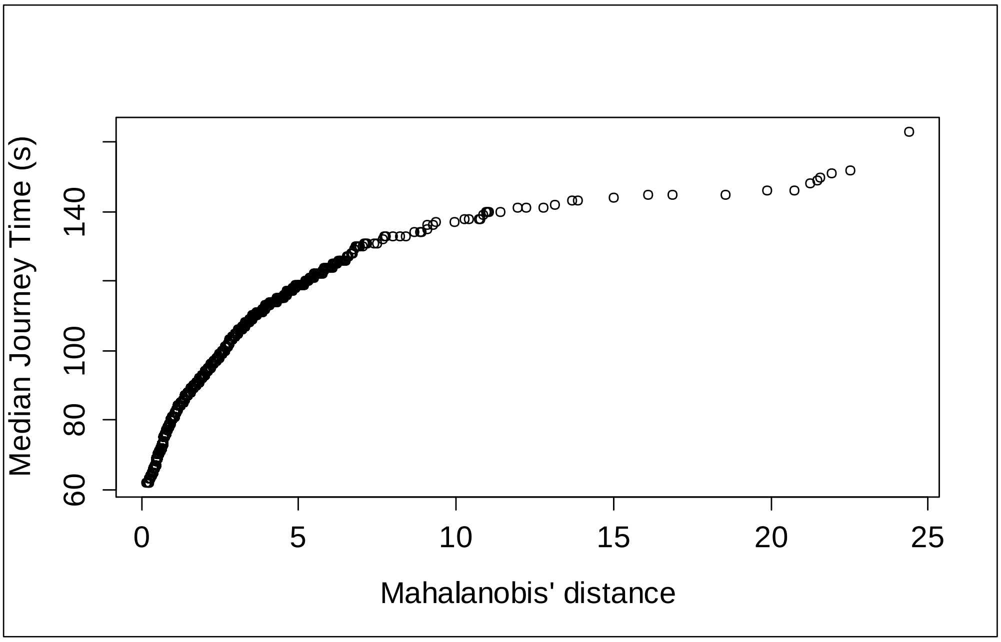
Appendix 6C-7: Scatter plots of Link0506 total flow between Bluetooth and ATC by season in Trafford



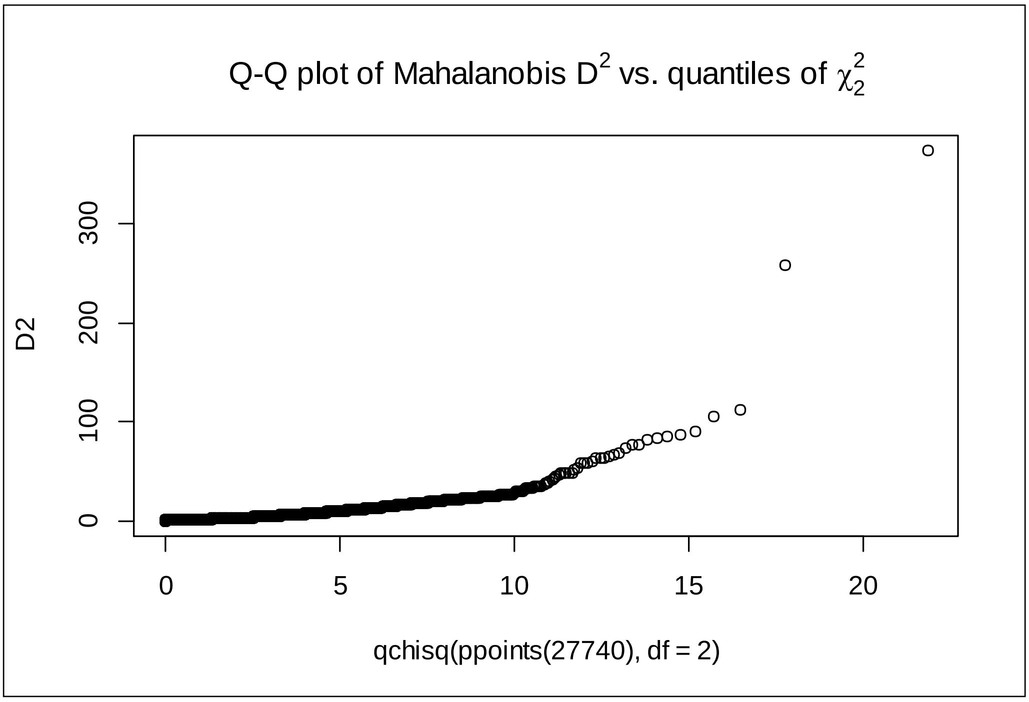
Appendix 6C-8: Typical daily flow profiles for each day in the week showing variation (%) in traffic proportions



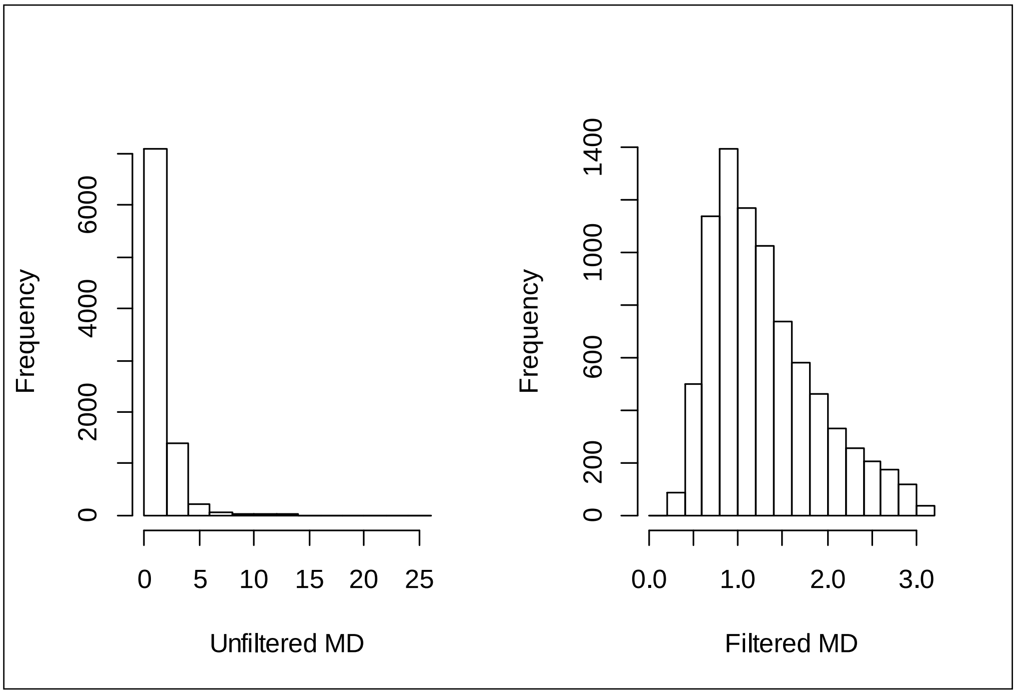
Appendix 6D-1: Plot of mean journey times against Mahalanobis distances



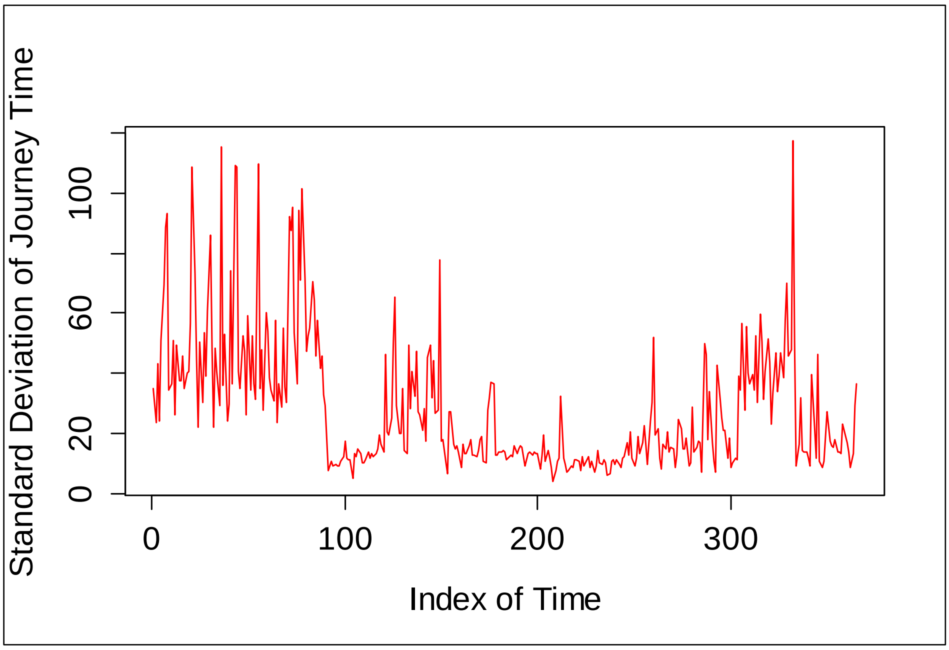
Appendix 6D-2: Plot of median journey times against Mahalanobis distances



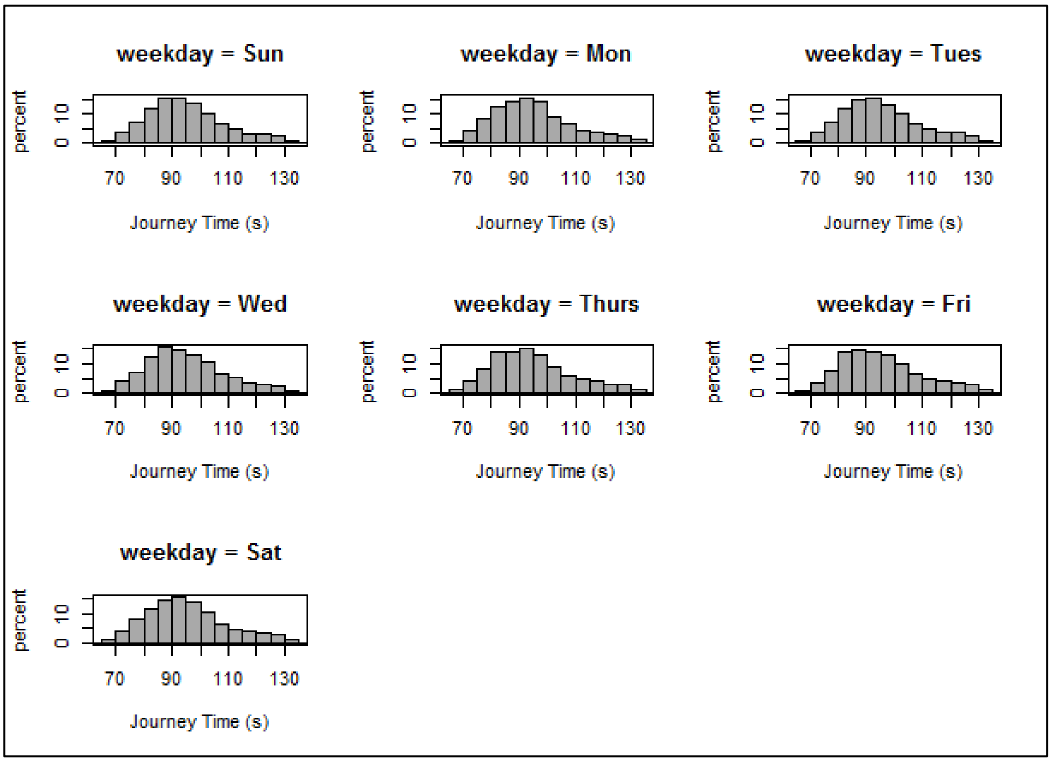
Appendix 6D-3: Plot of squared of Mahalanobis distances against Chi-square of degree of freedom 2



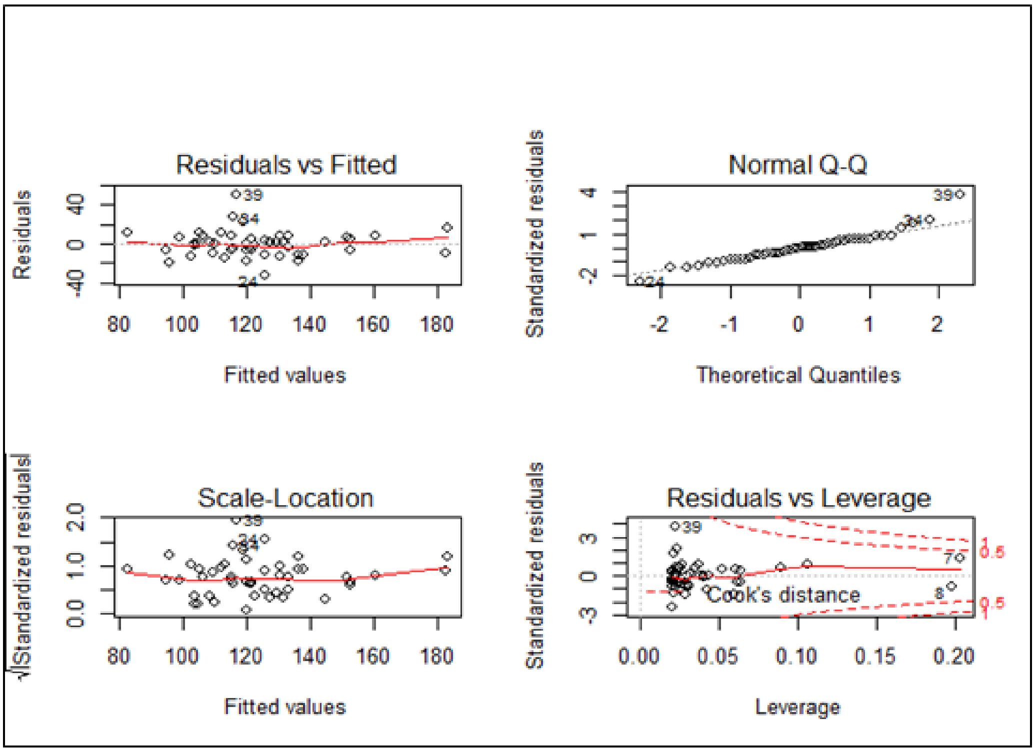
Appendix 6D-4: Histogram plots of Unfiltered (left) and Filtered (right) Mahalanobis distances



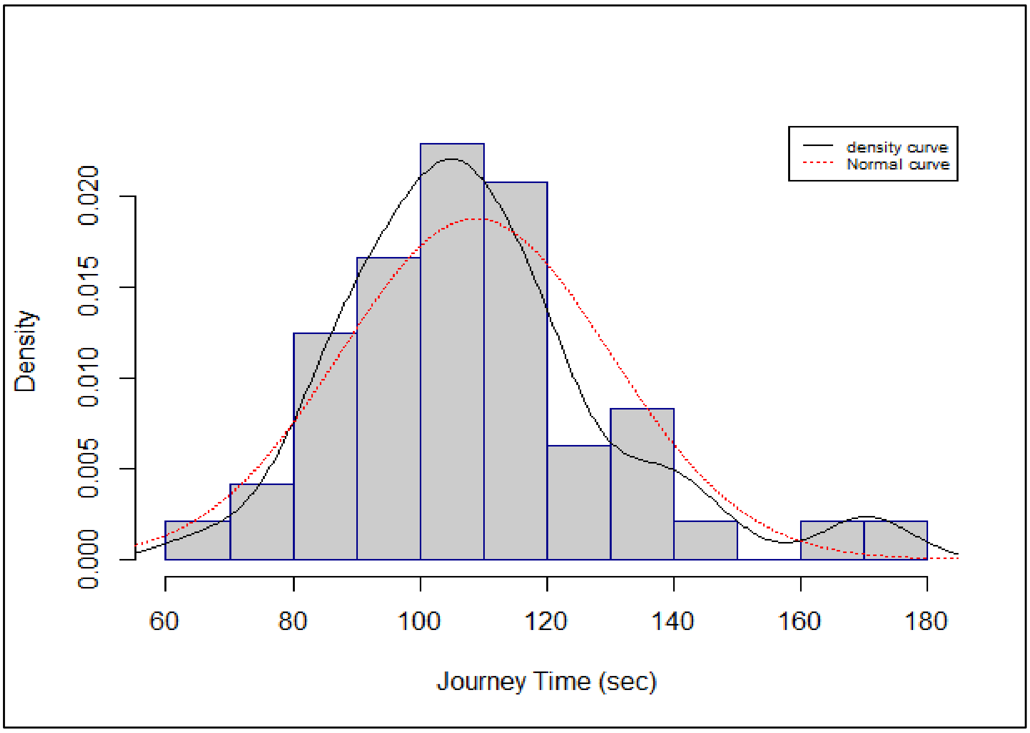
Appendix 6D-5: Standard deviation of journey time before cleansing



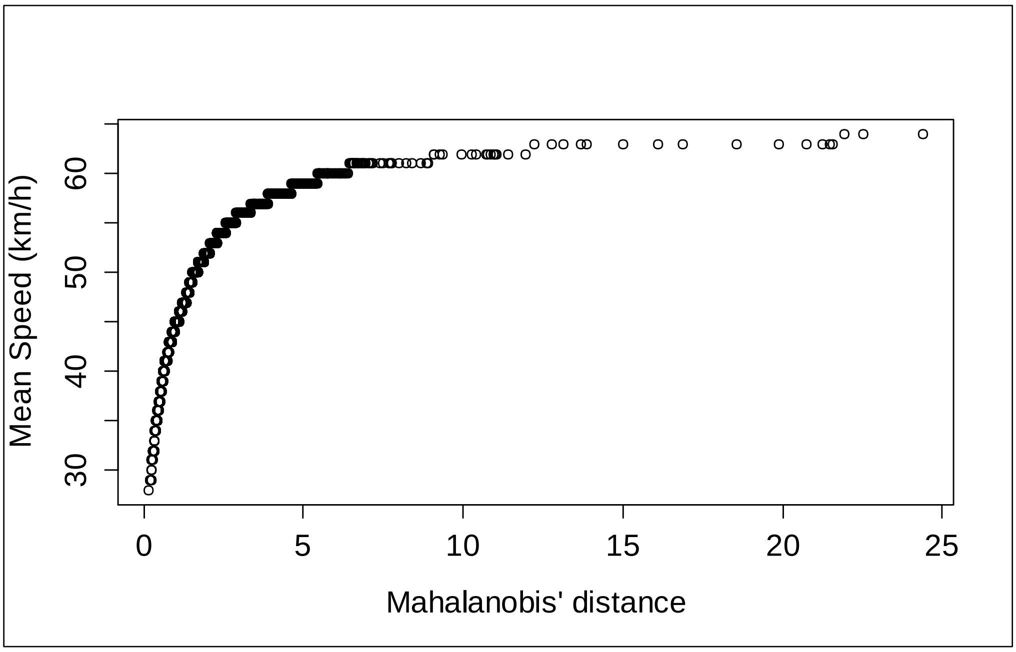
Appendix 6D-6: Histogram plots of weekday journey times on Link0605 (SW)



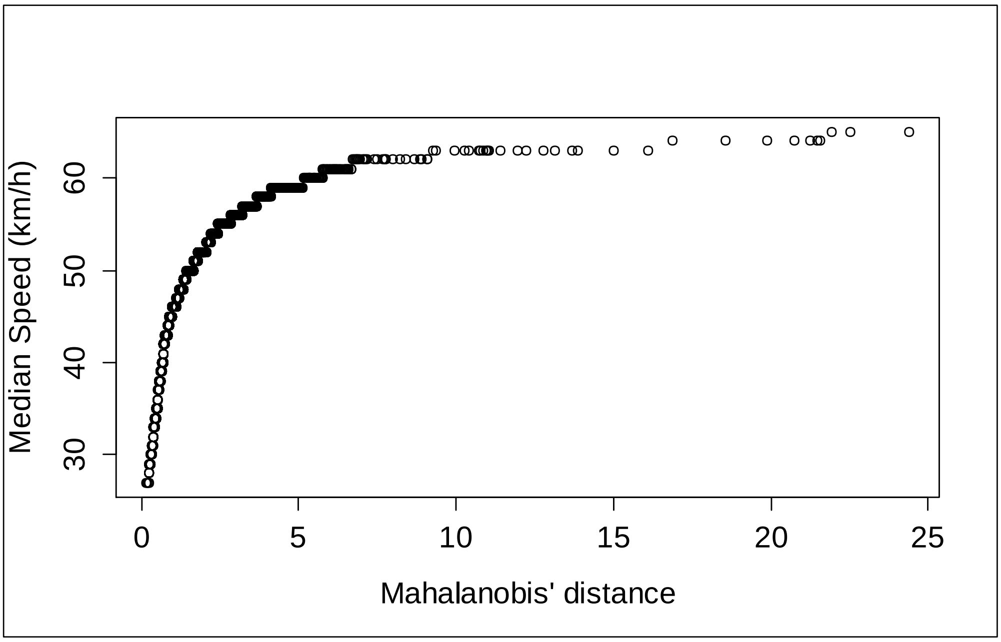
Appendix 6D-7: Diagnostic plots of linear modelling of ANPR and Bluetooth journey times



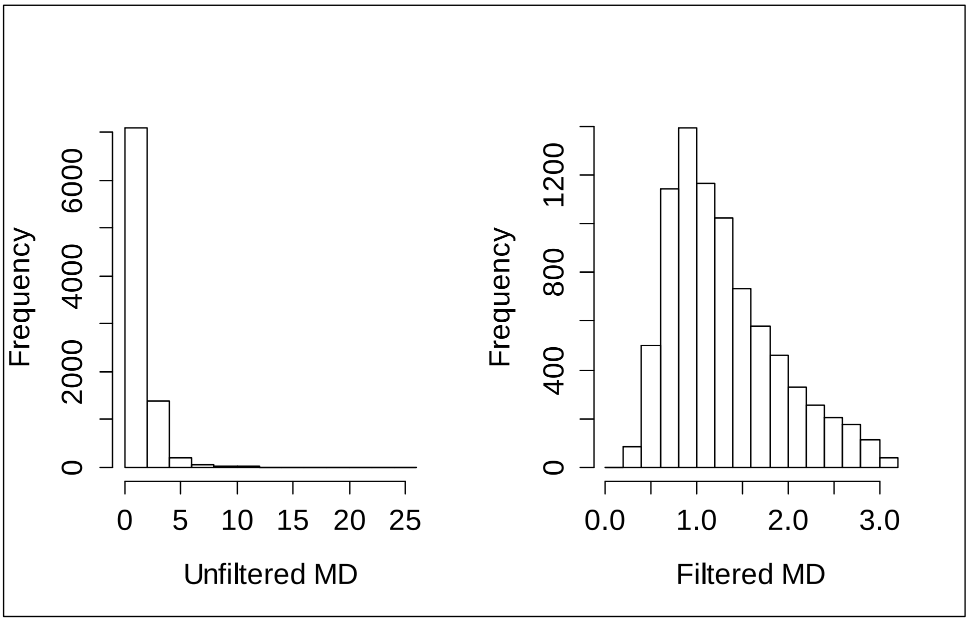
Appendix 6D-8: Histogram of ANPR journey times overlaid with Normal and Density Curves



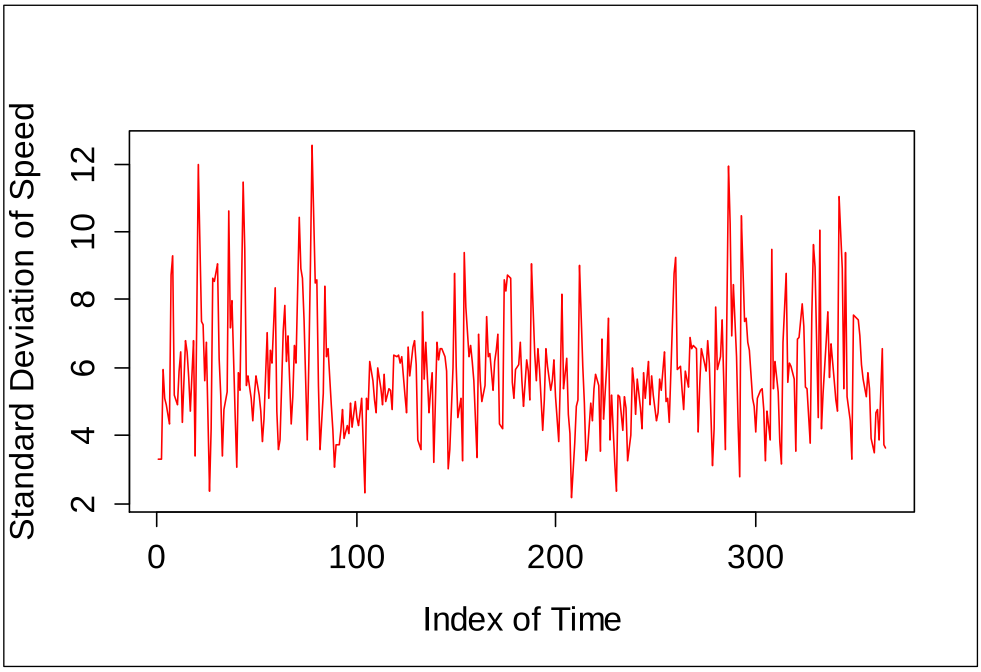
Appendix 6E-1: Plot of mean speed against the Mahalanobis distances



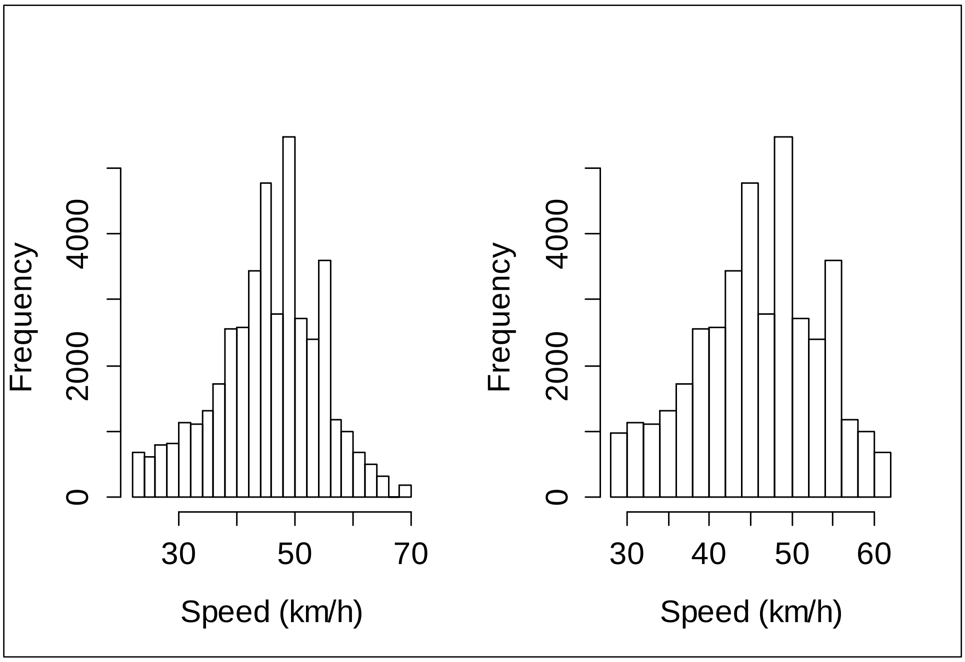
Appendix 6E-2: Plot of median speed against the Mahalanobis distances



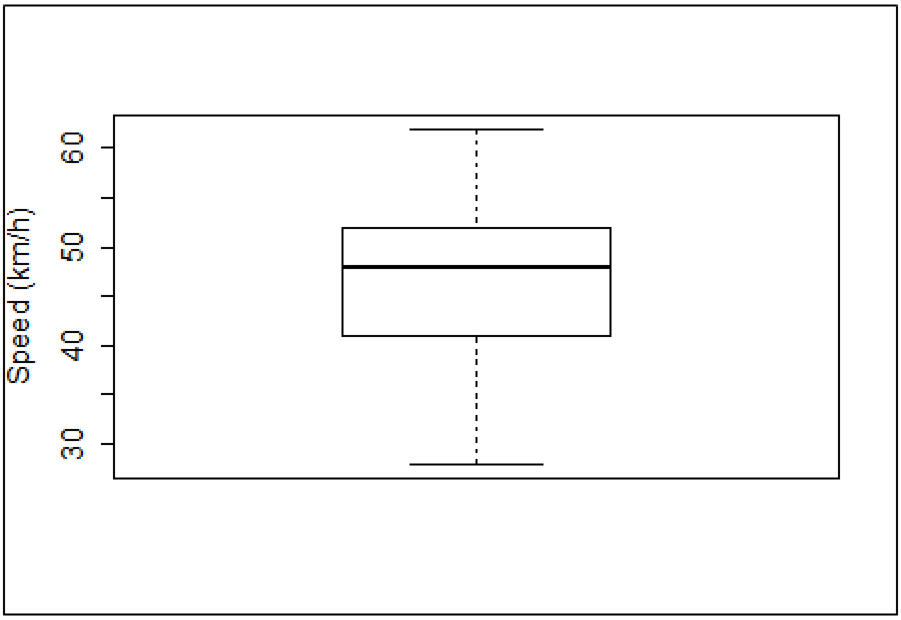
Appendix 6E-3: Histogram plots of Unfiltered (left) and Filtered (right) Mahalanobis distances



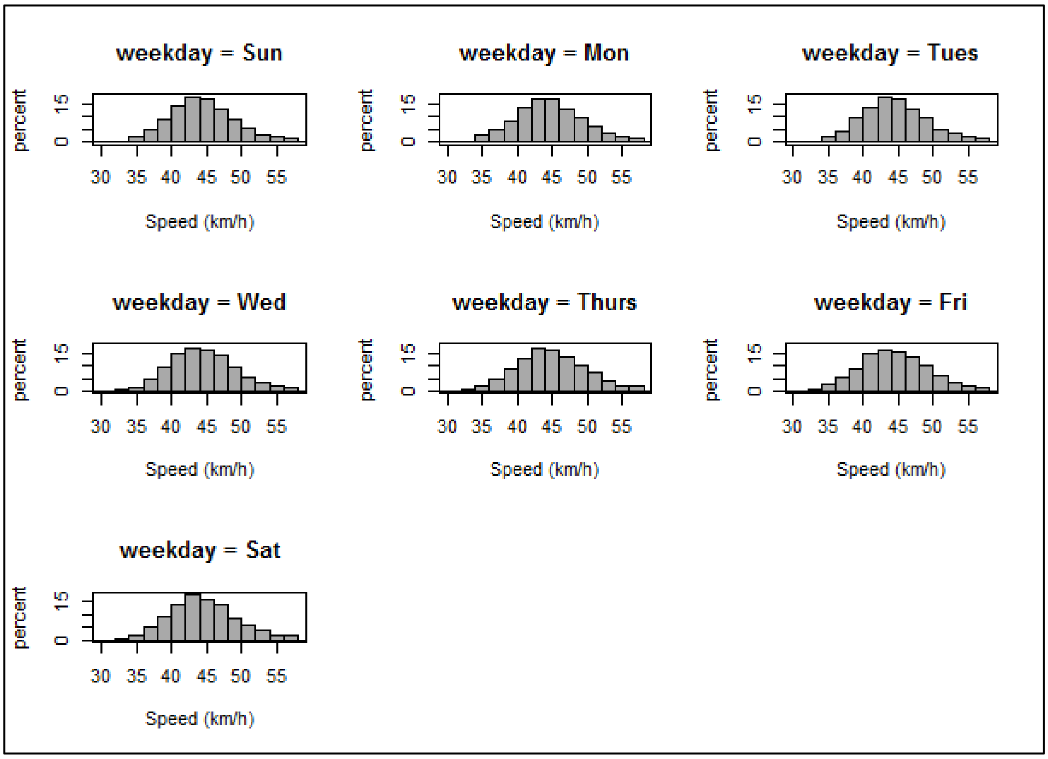
Appendix 6E-4: Standard Deviation of vehicle speeds before filtering



Appendix 6E-5: Histogram of vehicle speeds before (left) and after (right) filtering



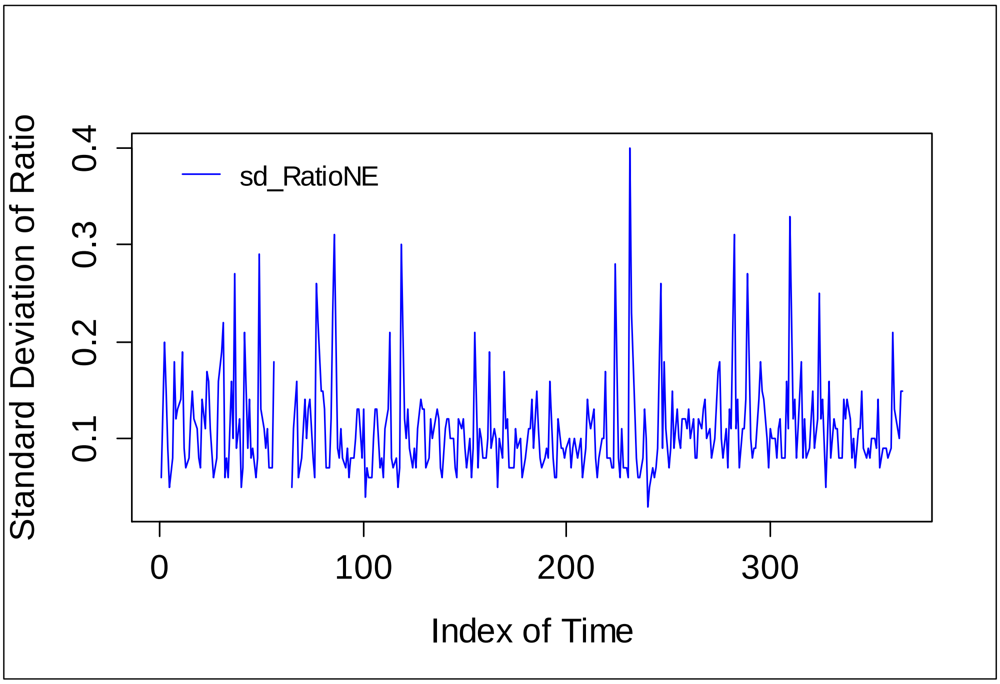
Appendix 6E-6: Boxplot of speed after filtering on Link3435



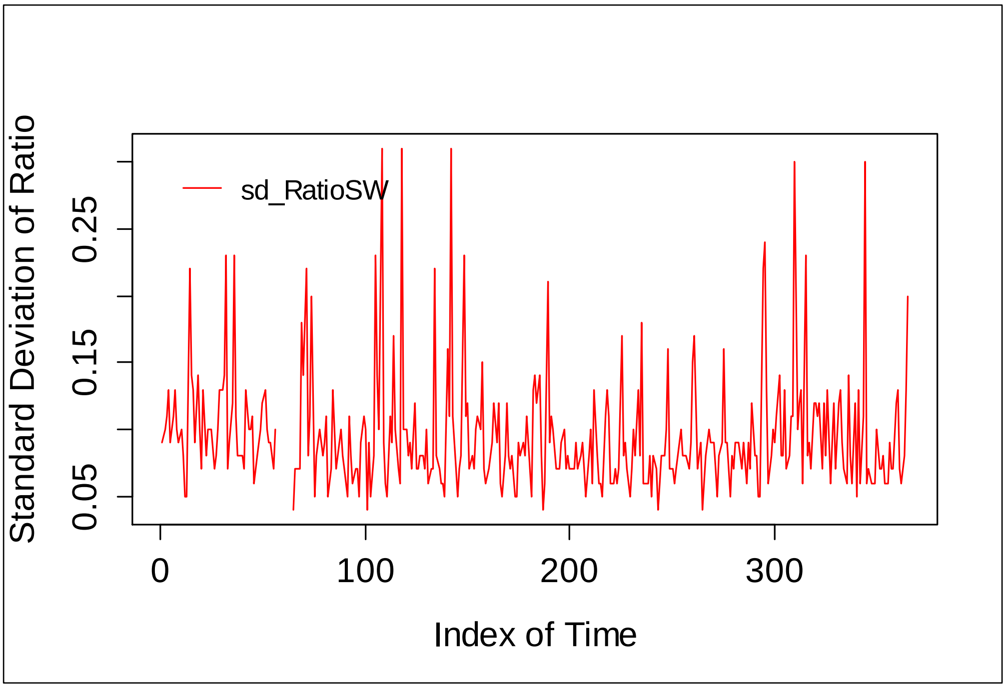
Appendix 6E-7: Histogram plots of weekday journey speeds on Link0605 (SW) in Trafford

	Speed (km/h)		Journey Time (s)	
	Median	Mean	Median	Mean
Min.	27	28	62	62
1st Qu.	44	43	78	85
Median	48	46	84	98
Mean	47.51	47.07	85.66	102.3
3rd Qu.	51	51	91	115
Max.	65	64	163	207
N	8760	8760	8760	8760

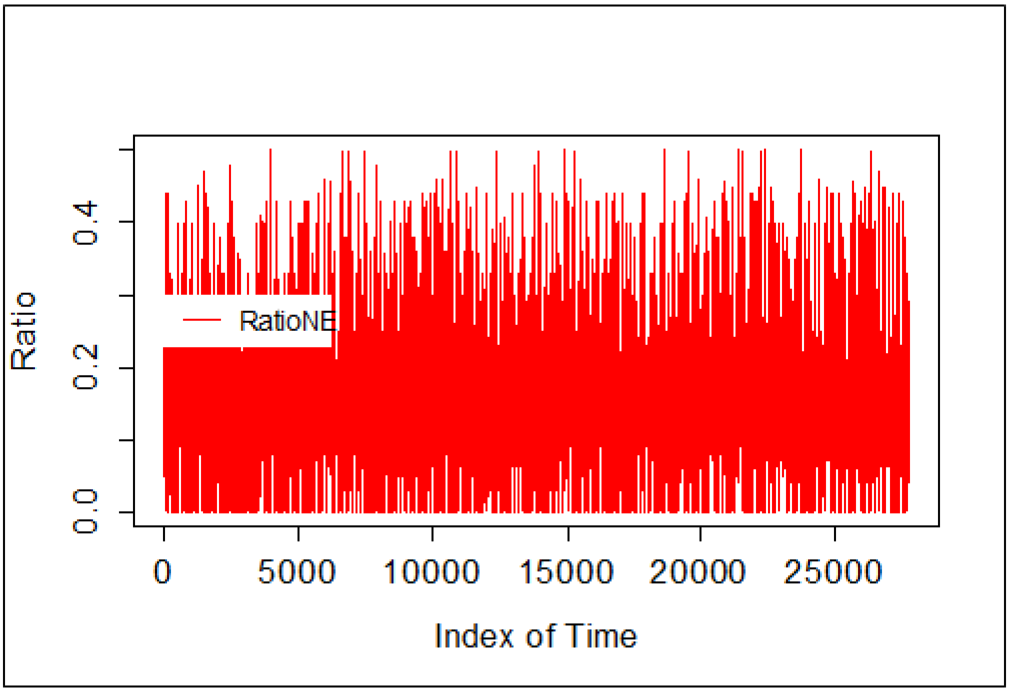
Appendix 6E-8: Summary of journey times and speed after the application MD filtering



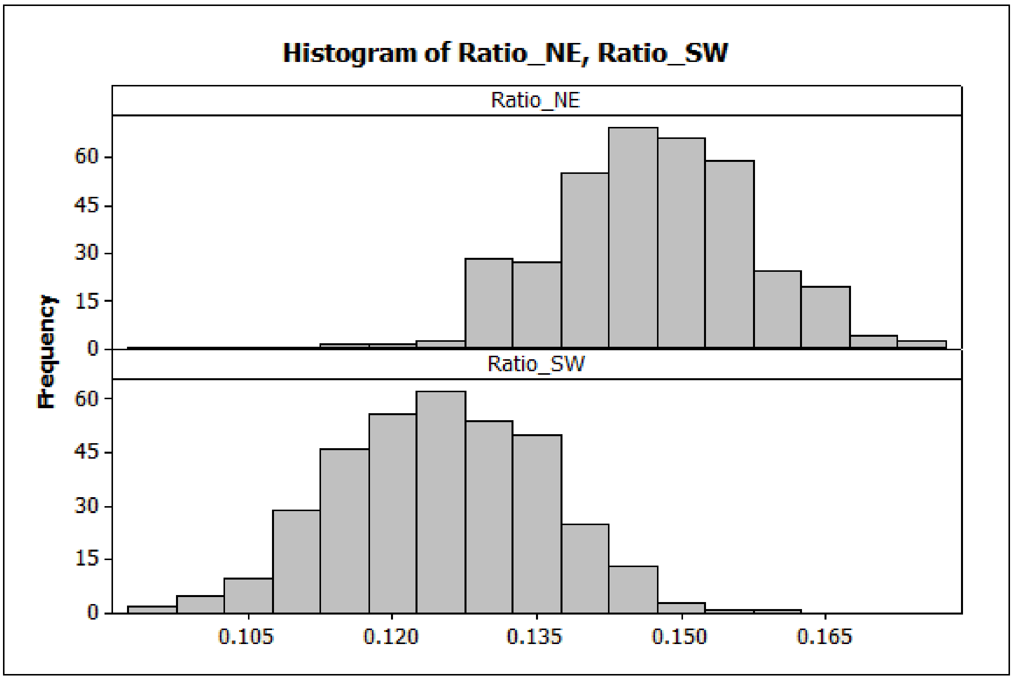
Appendix 6F-1: Standard deviation of NE ratio (detection rate) before cleansing on Link0506



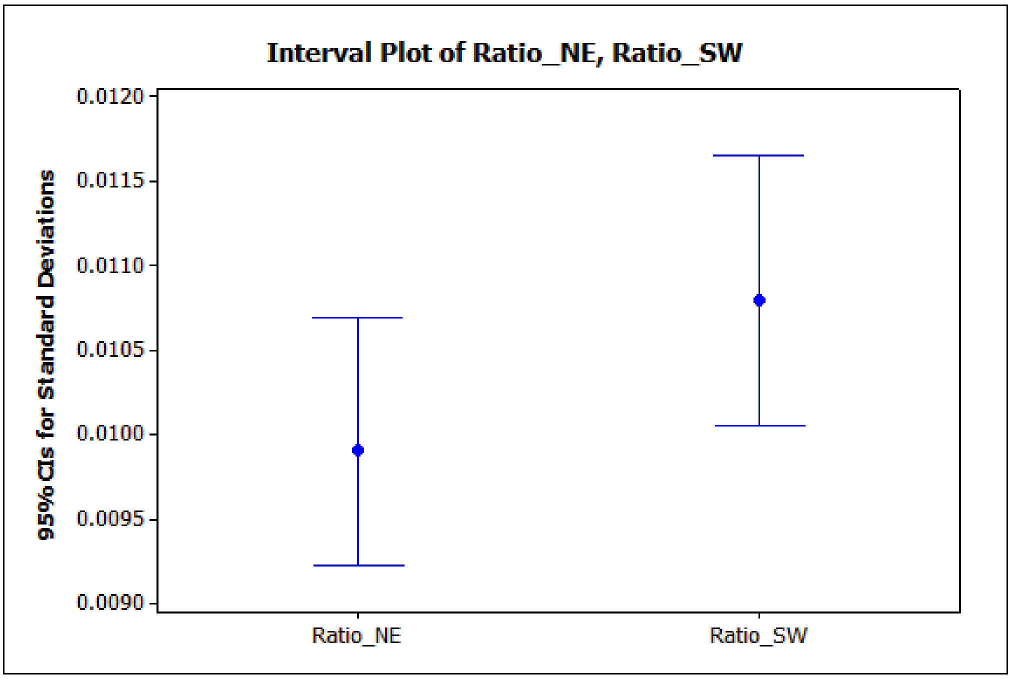
Appendix 6F-2: Standard deviation of ratio (SW) before cleansing on Link0506



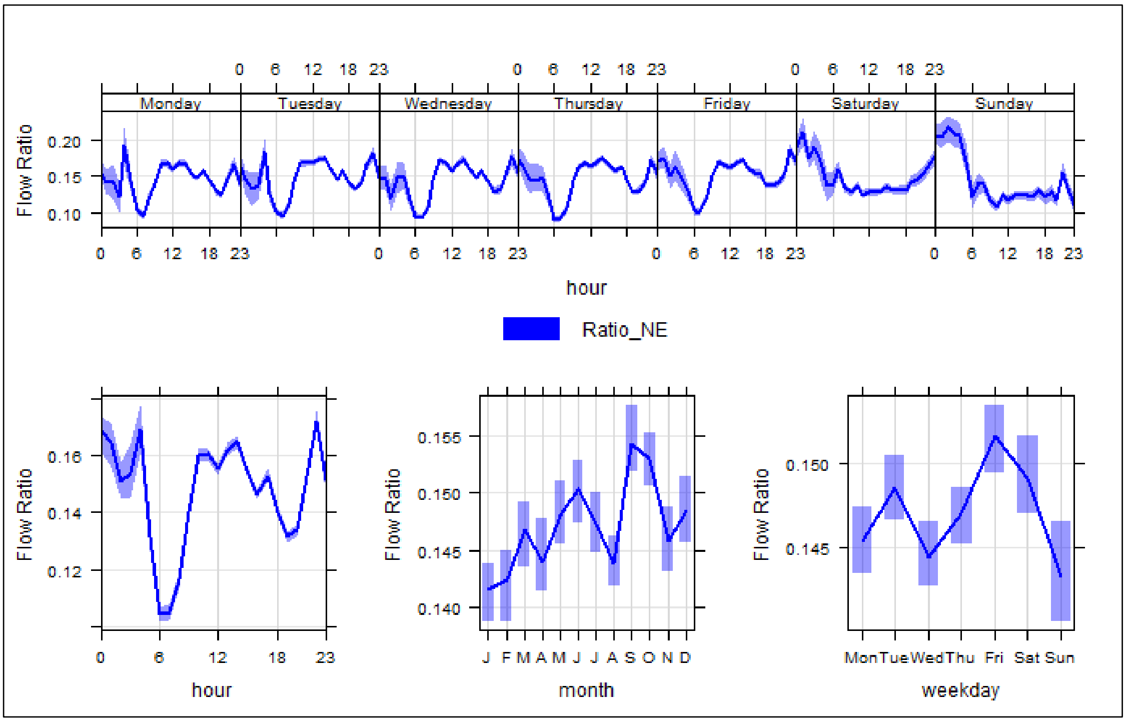
Appendix 6F-3: Mean of ratio (NE) before cleansing on Link0506



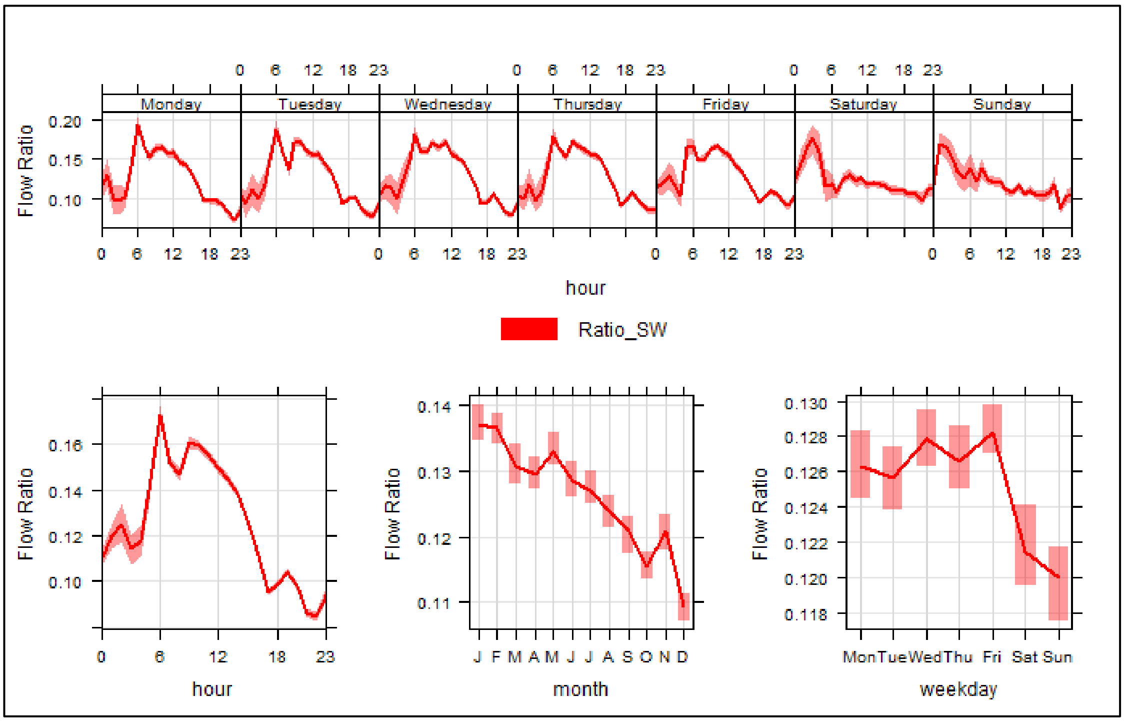
Appendix 6F-4: Histogram plots of ratio in both directions on Link0506



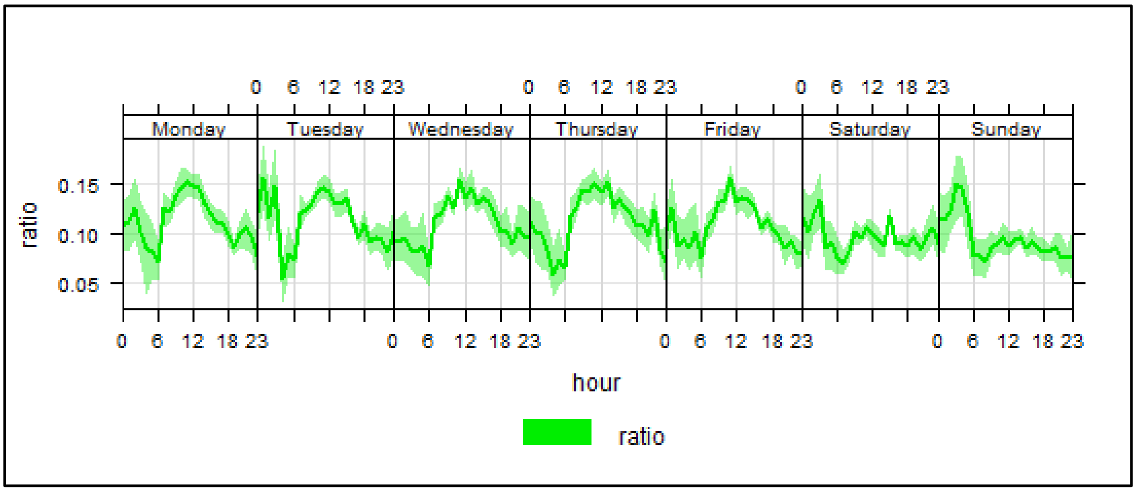
Appendix 6F-5: Interval plot of standard deviations of ratios at 95% confidence level



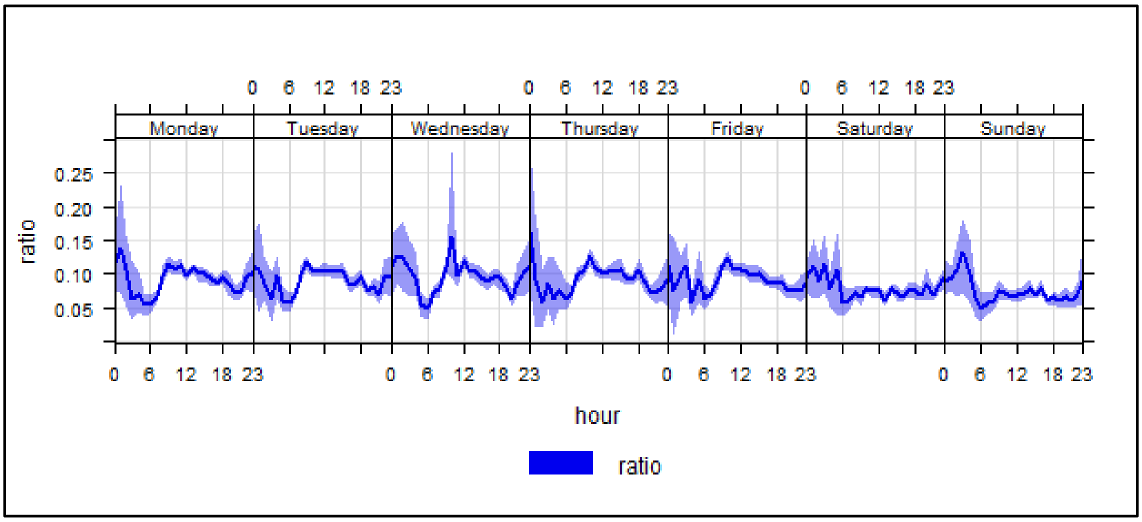
Appendix 6F-6: Plot of ratios on four temporal dimensions on Link0506 (NE)



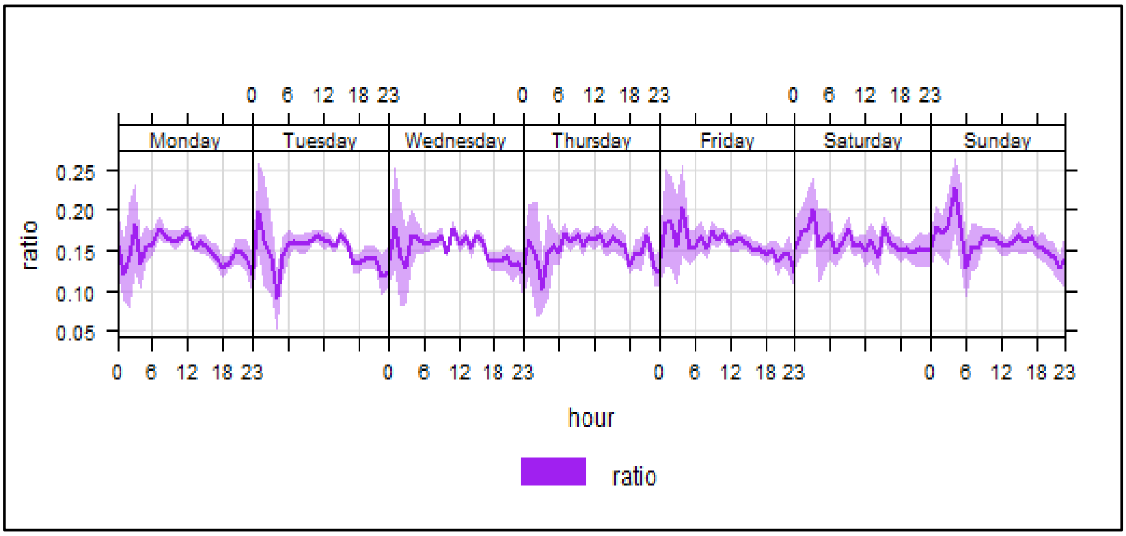
Appendix 6F-7: Plot of ratios on four temporal dimensions on Link0506 (SW)



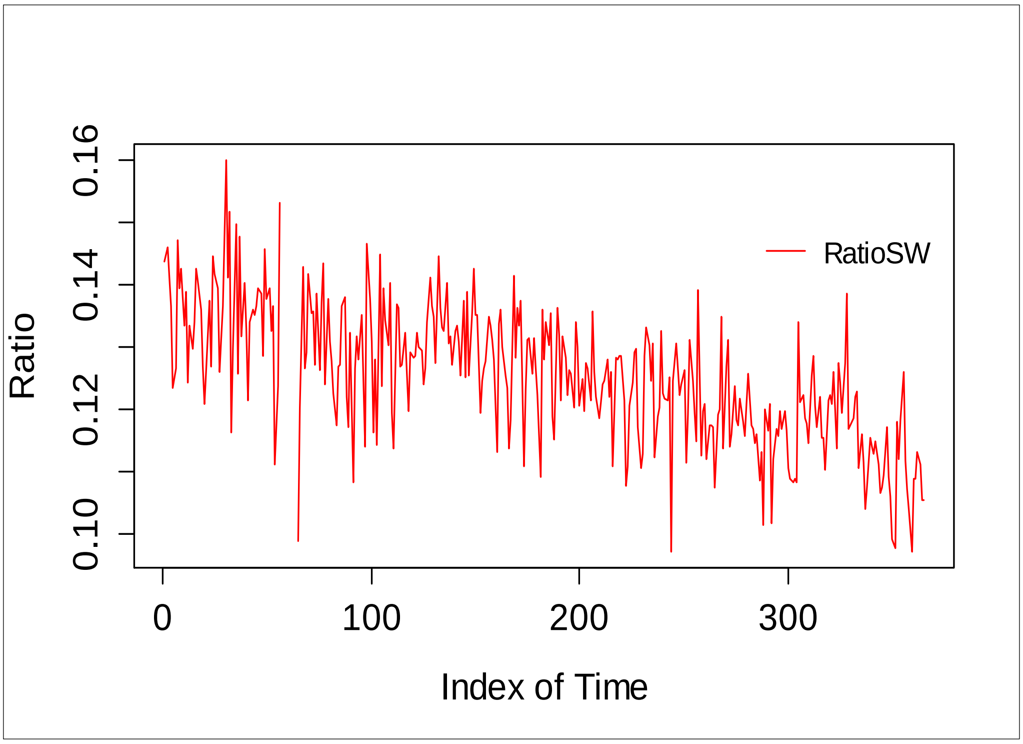
Appendix 6F-8: Bluetooth-ATC flow ratio profiles on Link3534 (Northbound)



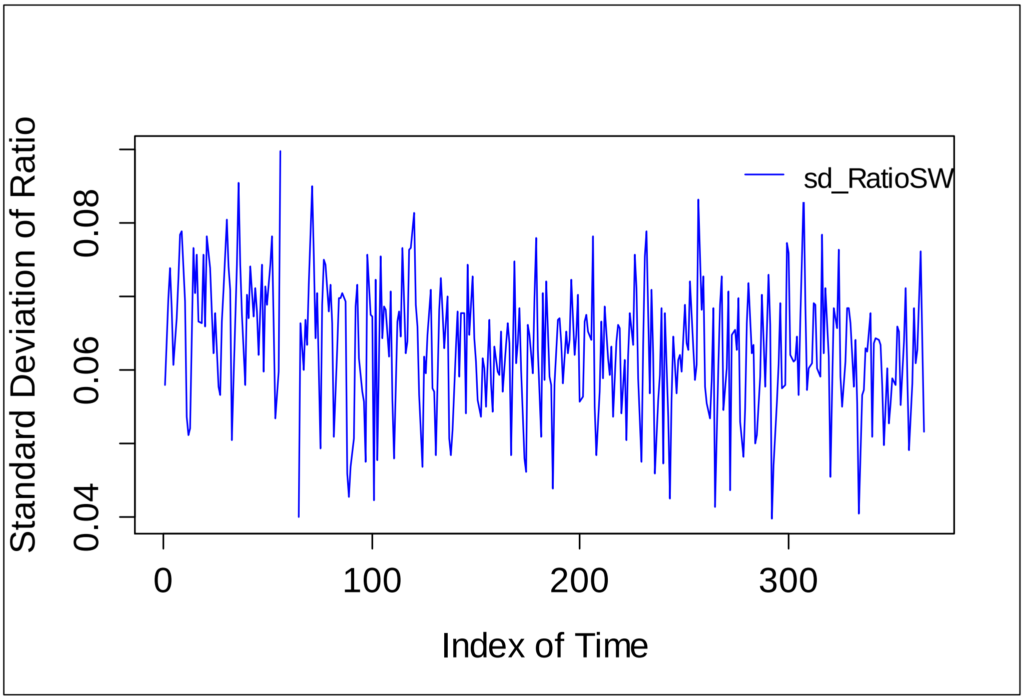
Appendix 6F-9: Bluetooth-ATC flow ratio profiles on Link3435 (Southbound)



Appendix 6F-10: Bluetooth-SCOOT flow ratio profiles on Link3534 (Northbound)



Appendix 6G-1: Day-to-day SW-bound ratio on Link0506



Appendix 6G-2: Day-to-day SW-bound standard deviations of ratio on Link0506

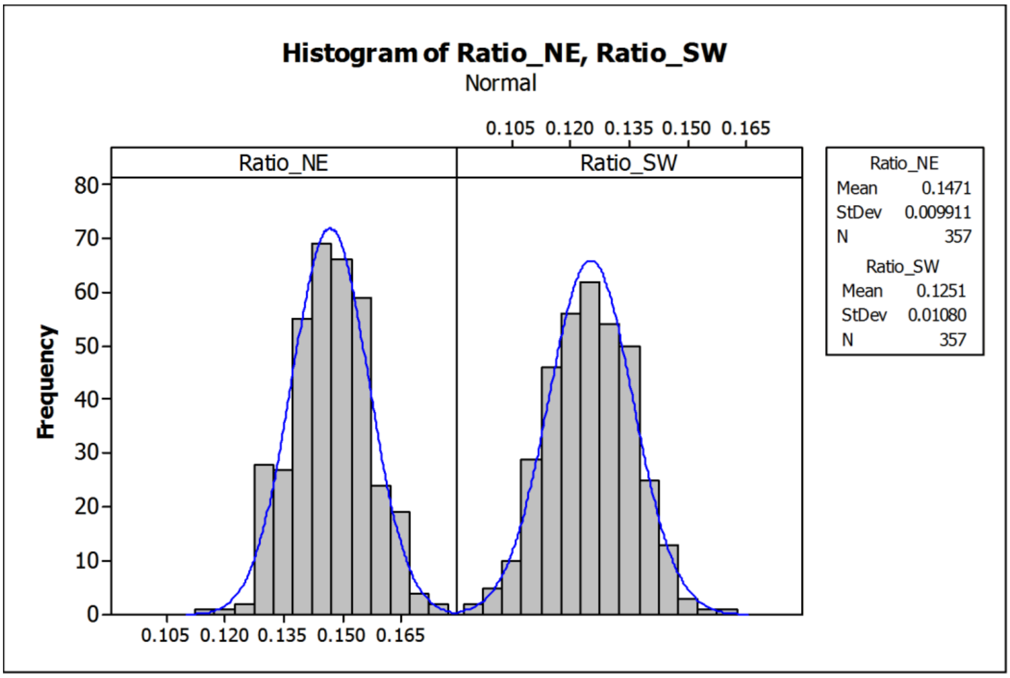
	Directional Ratio			MD
	NE	SW	Total Flow	
Min.	0.00	0.00	0.00	0.09
1st Qu.	0.11	0.09	0.11	0.45
Median	0.14	0.12	0.13	0.69
Mean	0.15	0.13	0.13	0.83
3rd Qu.	0.18	0.16	0.16	1.04
Max.	0.50	0.44	0.29	2.79
N	27740	27740	27740	27740

Appendix 6G-3: Summary of directional flow ratios

Ratio Statistics for Ratio_NE / Ratio_SW

Std. Deviation	Price Related Differential	Coefficient of Dispersion	Coefficient of Variation Median Centered
.137	1.008	.093	11.7%

Appendix 6G-4: Statistics of NE-ratio to SW-ratio



Appendix 6H: Histogram plots of day-to-day NE and SW detection rates

Appendix 3B: R Codes for Bluetooth Processing

```
## Bluetooth script
## Program lines starting with # are for comments
## This program is used to load, process, analyse and display
#Bluetooth data and the estimated traffic metrics. ##
## The model termed TRAFOST computes traffic metrics such as: Flow,
# O-D Matrix, vehicle speeds and travel times from Bluetooth data.
## Analysis performed using TRAFOST includes daily, hourly, 15-minutes
## etc. summary, geospatial analysis and data sets comparison for
## validation and computation of penetration or detection rates.
## The program also performs graphical presentation of results on the
## plot window. It also has a function with the capability to produce
## motion charts on Google map as well as plotting locations on Google
## Earth/Google map. TRAFOST makes use of the different functions of
# the computational model to accomplish the four stages of the
#analysis procedures developed in this research. TRAFOST is developed
#in R language and is dependent upon R packages for effective running.
# The Key packages used in plotting are ‘openair’ by Carslaw, 2006 and
# Rcommander (‘Rcmdr’) by Fox, 2005.
## Program written by E. G. Ayodele, PhD Civil Engineering and
#Geosciences
## Newcastle University, United Kingdom. 2013 Edition. Last modified
## December 2016.
## For further information, please contact:
#e.g.ayodele@newcastle.ac.uk
paste("Start date/time is", date()) # to write the start time and date
#of operation
## loading some pre-installed packages to be used
library(plyr)
library(lubridate)
library(reshape)
library(ggplot)
library(cluster)
library(latticeExtra)
library(grid)
## specify the files directory and load the data files
path.files <- "H:\\R\\stockport\\"
bt.data <- lapply(list.files(path = path.files, pattern = ".csv"),
function(.file) read.csv(paste(path.files, .file,
sep = ""), header = TRUE))
## Examine part of the data to access structure
head(bt.data[[1]])
######################################################################
## Function to reduce the data size to 3 columns. Columns 1, 2 and 15
## “Site ID”, “Date” and “Vehicle Id” are required.
btr.data <- lapply(bt.data, function(bt.data) bt.data[c(1:2,15)])
## Examine part of the data to ensure that the output is correct
head(btr.data[[1]])
######################################################################
## Function to order the data by vehicleId. All the columns are kept
bt.order<-lapply(btr.data, function(btr.data)
btr.data[order(btr.data[,3]),1:3])
######################################################################
## Apply time format to the list and remove duplicates from data using
## function fdup
btr.data <- bt.order # assign a new name to the ordered data list
fdup <- function(btr.data){
for(i in btr.data){
tm <- dmy_hms(btr.data$Date)
btr.data$day <- day(tm) # retrieve date value from the data
btr.data$hour <- hour(tm) # extract hour component and add to df
# Extract the minute component of dateTime and add to the df
btr.data$min <- minute(tm)
btr.data$second <- second(tm) #This extracts the seconds part of the
## data
# Compute time in seconds and add to station data
```

```
btr.data$tsec <- as.numeric(btr.data$hour*3600 + btr.data$min*60
+ btr.data$sec)
# Convert data to vectors to apply unique()
hour <- btr.data$hour
min <- btr.data$min
# compute the 15-minute interval summary
min15 <- floor(as.numeric(min)/15)
# multiply by 15 to obtain the 15-minute format
btr.data$min15 <- min15*15 +0 # addition of 0 makes the 1st 0-15mins 0
# compute the 10-minute interval summary
min10 <- floor(as.numeric(min)/10)
# multiply by 10 to obtain the 10-minute format
btr.data$min10 <- min10*10 +0 # add 0 to make the 1st 0-10mins 0
# compute the 5-minute interval summary
min5 <- floor(as.numeric(min)/5)
# multiply by 5 to give the minutes a proper format
btr.data$min5 <- min5*5 +0 # add 0 so that the first 0-5mins will be 0
## assign new variables to tsec and VehicleId
x <- btr.data[,8]
y <- btr.data[,3]
n <- length(x)
## Compute time difference in seconds between successive points
btr.data$secdif <- c(0,as.numeric(abs(diff(x))))
## Remove the duplicate records from the data to obtain a subset
y1<-y[2:n]
y2<-y[1:(n-1)]
yc<-as.character(y1)!=as.character(y2)
btr.data$yc <-c("TRUE",yc)#add "TRUE" to the 1st point for completion
ndup <-
btr.data[(btr.data$yc=="FALSE"&btr.data$secdif>=300)|(btr.data$yc=="TR
UE"),]
return(ndup)
 }
}
bt.rdup <- lapply(btr.data,fdup)
####################################################################
# Function to reduce the file size to 8 columns before merging
bt.rdup2 <- lapply(bt.rdup,function(bt.rdup) bt.rdup[c(1:5,9:11)])
###################################################################
# Duplicates are removed before files are merged to avoid creating
# unwanted large files
###################################################################
# Merging more than 2 data files (one-many mapping) or (many -many)
# Create 2 lists of the reduced data to enable the merging process
mdata <- bt.rdup2 # 1st list
mdata1 <- bt.rdup2 # 2nd list
# enter 0 or 1 for "mgopt" according to merging option (1-many or
# many-many)
mgopt <- 0
# data merging starts here
if(mgopt==0){
mgr <- function(mdata){
mdat <- mdata1[[1]]
mdat <- merge(mdat, mdata, by = "VehicleId", sort=T,all = FALSE)
return(mdat)
}
mg <- lapply(mdata,mgr) # The list of the merged files is created here
}else{
##############################################################
mgr <- function(mdata){
mdat<-list()
for(j in 1:length(mdata1)){
mdat1 <- mdata1[[j]]
# names(mdata) <- names(mdat)
mdat[[j]] <- merge(mdata, mdat1, by = "VehicleId", sort=T,all = FALSE)
}
return(mdat)
}
```

```
system.time(mg <- lapply(mdata,mgr))# list of the merged files is
#created
}
# many- many produces a list of lists
####################################################################
# Select the desired list file if many to many mapping
if(mgopt==1){
mga <- mg # save the large file with a different name for preservation
mg1 <- mga[[1]] # the first list n the bigger list (change the index
## accordingly)
mg1 <- mg1[-1] # drops the first file in the first list
mg <- mg1 # assign a new mg to the created list
} else {mg <- mg[-1] }# to remove the unwanted file
####################################################################
# The next step is to compute distance from station coordinates and
# subsequently the vehicle speed
coords=read.csv("H:\\R\\bt_st_details.csv",header=T)
## Function to compute distance,time and speed
fsvt=function(mg){
st1 <- substring(as.character(mg[1,2]),2,13)
st2 <- substring(as.character(mg[1,9]),2,13)
lat1 <- coords[as.character(coords[,1])==as.character(st1),c(8)]
lat2 <-coords[as.character(coords[,1])==as.character(st2),c(8)]
lon1 <-coords[as.character(coords[,1])==as.character(st1),c(7)]
lon2 <-coords[as.character(coords[,1])==as.character(st2),c(7)]
# computation of time differences between two data ponits. And
# addition of the computed differences to the merged dataframe
t1<-strptime(mg$Date.x,"%d/%m/%Y %H:%M:%S")
t2<-strptime(mg$Date.y,"%d/%m/%Y %H:%M:%S")
jtime<- difftime(t2,t1,units="secs")
# jtime<- difftime(t2,t1,units="auto")
## Distance computation using spherical coordinates. Distance in km
R=6378137 # WGS84 radius of the earth
sn=sin(lat1)*sin(lat2)
cs=cos(lat1)*cos(lat2)*cos(lon2-lon1)
dist=(acos(sn+cs)*pi/180)*R/1000
dst <- as.numeric(sprintf("%.2f",dist))
# or use "dst <- print(dst,digits=3)"
## computation of vehicle speed begins here
tme <- as.numeric(jtime/3600)
# thr <- as.numeric(sprintf("%.2f",tme))
tmin <- as.numeric(tme*60)
tmin <- as.numeric(sprintf("%.2f",tmin))
spd <- ceiling(as.numeric(abs(dst/tme)))
mg2 <- data.frame(mg, jtime,tmin,spd)
# remove point data with different days merged together
tf1 <- dmy_hms(mg2$Date.x)
tf2 <- dmy_hms(mg2$Date.y)
day1 <- day(tf1)
day2 <- day(tf2)
mg2<-subset(mg2,day1==day2) # Subset for same day merged records
# remove vehicles travelling at very low speed and at very high speed
mg2 <- subset(mg2,spd>5&spd<=120)
# mg2 <- subset(mg2,spd>=0&spd<=120)# all the tracked devices
#mg2 <- subset(mg2,spd>=0&spd<=5)# assumes to be pedestrians and
# cyclists
# mg2$wf <- cut(mg2$hour.y,5)
return(mg2)
}
svt <- lapply(mg,fsvt)
svt1<-svt[[1]] # to obtain the first element of the list
head(svt1) # to examine the data (the first link)
####################################################################
## Normality test using quantile plots
# spd.ntp <-
lapply(svt,function(svt){qqnorm(svt$spd);qqline(svt1$spd)})
## remove outliers to obtain 95% of the remaining data if normally
## distributed
```

```
#attach(svt1)
#ulim <-mean(spd) +1.96*sd(spd)
#llim <-mean(spd) -1.96*sd(spd)
#spdf95<-subset(mg2,spd>=llim) # to obtain all the values greater than
llim
#spdf95<-subset(spdf95,spd<=ulim) # to remove values greater than ulim
#spdf95<-subset(mg2,(spd>=llim)&(spd<=ulim))
#detach(svt)
######################################################################
## The following is to separate the merged files into two based on the
## travel direction
# par(mfrow=c(3,3))
drn_pos <- lapply(svt,function(svt) svt[svt$jtime>0,])
drn_pos1 <-drn_pos[[1]]
######################################################################
## Rule of thumb to remove outliers Crawley, 2005
attach(drn_pos1 )
lmtquant <-subset(drn_pos1 , spd<=upquant&spd>=lwquant)
outquan <- subset(drn_pos1,spd<lwquant|spd>upquant) # outliers
######################################################################
## Station Summary
# convert list to a dataframe using ldply function and do a summary of
# daily count per station
dfbt.rdup <- ldply(bt.rdup)
## reduce the file size to the desired variables
#(site.id,date,vehicleid,day,hour,min15,min10, min5, secdif)
dfbt.rdup_red <- dfbt.rdup[c(1,4:5,9:11)]
dfbt.rdup_red <- dfbt.rdup_red[dfbt.rdup_red$day==3,] #to subset day 3
## summarise the data for daily count per station
#stn <- t(stn)
stn_sum <- as.data.frame(ftable(dfbt.rdup_red, row.vars=c(1),
col.vars=c(2))) #long format of above
## plot the bar chart of the daily count data
barplot(stn_sum[,3], main="Sept 3 bar plot", ylab="Daily count",
col=c(1:8,"purple"), xlab= "Stations 33-41",las=2,
cex.main=1.0,cex.lab=0.8,cex.axis=0.8)
#legend("topleft","Stockport")
dfbt.rdup_red <- dfbt.rdup[c(1,4:5,9:11)]
stn <- ftable(dfbt.rdup_red, row.vars=c(2), col.vars=c(1))
ftable(dfbt.rdup_red, row.vars=c(2), col.vars=c(1))
## Bar plot stations. Plot the stations side-by-side
#barplot(stn[,2:9], main="Sept 3-10 bar plot", ylab="Daily count",
# col=c(1:9,"purple"), xlab= "Stations 33-41",las=2,beside=TRUE)
## Line plot station
plot(stn[,2], xlab=c("Weekdays from Mon-Mon"),ylab="Daily count",
ylim=c(3500,12000), main="Stockport Bluetooth daily profile at nine
stations",
xaxt="n",cex.main=1.0)
mtext(side=1,at=1:8,text=c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8",
"Sun9","Mon10"))
for(k in 1:9){
lines(stn[,k],col=k,lty=k)}
legend("topright",lty=c(1:9),col=c(1:9),legend=paste("Stn",33:41,sep="
"),cex=0.6)
######################################################################
print("Begin inbound processing and analysis from here")
# Section to analyse the vehicles travelling from origin to
# destination (pos direction)
######################################################################
## Data summary, statistical analysis and plotting. Each link is
# processed in turn and stacked over one another
drn_pos <- lapply(svt,function(svt) svt[svt$jtime>0,])
# reduce the file size. drn.posr is a list containing reduced data
drn.posr <- lapply(drn_pos,function(drn_pos)drn_pos[c(1:2,9:16,18)])
#drn.posr <- lapply(drn_pos,function(drn_pos)drn_pos[c(1,10:14,16)])
# summarise the data using package plyr
ld_drn.posr <- ldply(drn.posr) # convert list to a dataframe
# order the data file by datetime
```

```
ld_drn.posr.order <- function(ld_drn.posr){
posr.order <- ld_drn.posr[order(ld_drn.posr[,4]),1:11]
return(posr.order)
}
ldposr <- ld_drn.posr.order(ld_drn.posr)
# write the data to file
write.csv(ldposr, "H:\\R\\stockport2\\ldposr.csv")
#ldposri <-ldposr[1]
stn.num <- as.numeric(substring(as.character(ldposr[,3]),10,13))
link.num <- as.numeric(substring(as.character(ldposr[,3]),10,13))
# Extract a specific link by day based
link3334.3 <- ldposr[ldposr$day.y==3&stn.num==1034,]
write.csv(link3334.3, "H:\\R\\stockport2\\link3334.3.csv")
#day.num <- ldposr$day.y
## Extract a specific link by day and create a list for all the days
## Function to run the days in turn
# lnk <- function(ldposr){
#for(ldy in 3:10){
# link3334.5 <- subset(ldposr,day.num==ldy&link.num==1034)
# return(link3334.5)
# }
# write.csv(link3334.5, "H:\\R\\stockport2\\link3334.5.csv")
# }
#link <-lnk(ldposr)
# Get the count of VehicleIds
#id.count <- ddply(ldposr, .(Site.ID.y,day.y,VehicleId), "nrow")
#write.csv(id.count,file="idcount.csv")
# This takes some time to run
# create a summary of the data based on the specified variable
#avgtsec <- ddply(ldposr, .(Site.ID.y, day.y),
summarise,Vcount=length(VehicleId),
# min_jtime= min(jtime), max_jtime= max(jtime), mean_jtime=
mean(jtime))
#speed <- ldposr$spd
#Date.y <- ldposr$Date.y
# 5-minute speed flow summary based on repeated flow within an
interval
#sum_link3334.5.3 <- ddply(link3334.5.3, .(Site.ID.y,
day.y,hour.y,min5.y),
summarise,Vcount=rep(length(VehicleId), length(VehicleId)))
speed <- link3334.3$spd
Date.y <- link3334.3$Date.y
Site.ID.y <- link3334.3$Site.ID.y
vq_link3334.5.3 <- ddply(link3334.3, .(hour.y,min5.y),
summarise,Vcount=rep(length(VehicleId), length(VehicleId)))
#scatterplot3d(link3334.5plot)
summary(link3334.5plot)
#plot(link3334.5plot[,2],link3334.5plot[,3])
tsplot <- ts(link3334.5plot)
plot(tsplot[,3])
boxplot(tsplot[,3])
hist(tsplot[,3],col="light blue",border="dark blue", freq=FALSE,
######################################################################
## 10-minute summary
vq_link3334.10.3 <- ddply(link3334.3, .(hour.y,min10.y),
summarise,Vcount=rep(length(VehicleId), length(VehicleId)))
vq_link3334.10.3 <-
data.frame(Site.ID.y,Date.y,vq_link3334.10.3,speed)
write.csv(vq_link3334.10.3, "H:\\R\\stockport2\\vq_link3334.10.3.csv")
######################################################################
## 15-minute summary
vq_link3334.15.3 <- ddply(link3334.3, .(hour.y,min15.y),
summarise,Vcount=rep(length(VehicleId), length(VehicleId)))
vq_link3334.15.3 <-
data.frame(Site.ID.y,Date.y,vq_link3334.15.3,speed)
write.csv(vq_link3334.15.3, "H:\\R\\stockport2\\vq_link3334.15.3.csv")
######################################################################
## Summary per link for all days. Change the index in turn to compute
```

```
## the entire network on link basis
drn.posri <- drn.posr[[1]]
#spd.count <- ddply(drn.posri, .(day.y,min5.y,spd,todp.class), "nrow")
var.count5 <- ddply(drn.posri, .(day.y,hour.y,min5.y), "nrow")
write.csv(var.count5, "H:\\R\\stockport2\\var.count5.csv")
#attach(spd.count5)
#coplot(hour.y~nrow|day.y) # for conditioning plots
#coplot(spd~nrow|todp.class)
#detach(spd.count5)
# Summarise the daily traffic flow at different links by mean and
# median of journey time and speed
var.sum15 <- ddply(drn.posri, .(day.y,hour.y,min15.y),
summarise,Vcount=length(VehicleId),
med_spd = ceiling(median(spd)),mean_spd = ceiling(mean(spd)),
med_jt = ceiling(median(jtime)),mean_jt = ceiling(mean(jtime)))
write.csv(var.sum15, "H:\\R\\stockport2\\var.sum15.csv")
#var.sum15 <- var.sum15[order(var.sum15$todp.class),1:8]
var.sum15.d3 <- var.sum15[var.sum15$day.y==3,] # change this
#accordingly
## compute percentage count
var.sum15.d3$Vcount.pct <-
round((var.sum15.d3$Vcount/sum(var.sum15.d3$Vcount))*100,2)
palette <- c("red","yellow","blue","green","orange")
# map.class <- avgspd15.d3$todp.class
# plot of average speed grouped by time of the day
#plot(avgspd15.d3$mean_spd,avgspd15.d3$Vcount.pct,
# xlab="15-minute average speed (km/h)",ylab="Time of the day class",
# main="Daily Speed Classification",pch=21)
#################################################################
drn.posri<-drn.posr[[1]] # change the index in turn according to the
list #length
# Classify the Bluetooth count according to peak and off-peak
# periods
todp.class <- rep("0 - 07hrs", times=nrow(drn.posri))
todp.class[drn.posri$hour.y>=7&drn.posri$hour.y<10] <- "07 - 10hrs"
todp.class[drn.posri$hour.y>=10&drn.posri$hour.y<16] <- "10 - 16hrs"
todp.class[drn.posri$hour.y>=16&drn.posri$hour.y<20] <- "16 - 20hrs"
todp.class[drn.posri$hour.y>=20] <- "20 - 24hrs"
drn.posri$todp.class <- factor(todp.class)
boxplot(drn.posri$spd ~ todp.class, horizontal=T, xlab="5-minute
average speed
(km/h)",
las=1, cex.axis=0.8, cex.main=1.0,main="Box Plot of Journey Speed",
col="orange")
abline(v=mean(drn.posri$spd), lty="dashed")
# Adds the mean value to the plot
legend("topright", legend="Grand Mean", lty="dashed",cex=0.8)
#todp_sum <- tapply(drn.posri$spd,drn.posri$todp.class,summary)
tapply(drn.posri$spd,drn.posri$todp.class,summary)
tod.count5 <- ddply(drn.posri, .(day.y,min5.y,spd,todp.class), "nrow")
write.csv(tod.count5, "H:\\R\\stockport2\\tod.count5.csv")
#attach(spd.count)
#coplot(spd~nrow|day.y) # for conditioning plots
#coplot(spd~nrow|todp.class)
#detach(spd.count)
tod.sum15 <- ddply(drn.posri, .(day.y,min15.y,todp.class),
summarise,Vcount=length(VehicleId),
min_spd = min(spd),med_spd = median(spd),
mean_spd = ceiling(mean(spd)),
mean_journey times = ceiling(mean(jtime)))
write.csv(tod.sum15, "H:\\R\\stockport2\\tod.sum15 .csv")
tod.sum15 <- tod.sum15[order(tod.sum15 $todp.class),1:8]
tod.sum15.d3 <- tod.sum15[tod.sum15 $day.y==3,] # change this
accordingly
tod.sum15.d3$Vcount.pct <-
ceiling((tod.sum15.d3$Vcount/sum(tod.sum15.d3$Vcount))*100)
palette <- c("red","yellow","blue","green","orange")
map.class <- tod.sum15.d3$todp.class
```

```
# plot of percentage count by maximum speed (vmax) classification
#plot(avgspd15.d3$mean_spd,avgspd15.d3$Vcount.pct, ylim=c(0,12),
# xlab="15-minute average speed (km/h)",ylab="15-minute daily
# Bluetooth ## function palette adapted from Harris, 2013
count (%)", main="Speed Profile",pch=21,bg=palette[map.class])
#legend("topright",
legend=paste("<",tapply(as.numeric(avgspd15.d3$mean_spd),
legvals <- c(0,7,10,16,20) # cf Harris, 2013
# plot of percentage count by daytime classification
plot(tod.sum15.d3$mean_spd,tod.sum15.d3$Vcount.pct,ylim=c(0,12),
xlab="15-minute average speed (km/h)", ylab="15-minute daily Bluetooth
count(%)",main="Percentage Daily Speed
Distribution",pch=21,bg=palette[map.class])
legend("topright", legend=paste(">=",legvals),pch=21,
pt.bg=palette, pt.cex=1.5, bg="white,title="DayTime classification",
cex=0.8) # code adapted from Harris 2013
legend("right",
legend=paste("<",tapply(as.numeric(tod.sum15.d3$mean_spd),
map.class, max)),pch=21, pt.bg=palette,
pt.cex=1.5, bg="white",
title="DayTime class by Vmax",cex=0.8)
######################################################################
#ftable(drn.posr, row.vars = c(5,7), col.vars = c(11))# count based on
#specified r&c
drn.posri$todp.class <- NULL # to remove the column from the dataframe
#######################################################################
drn.posr.m <- melt(drn.posri, id.vars = 1:9) # to obtain link summary
# drn.posr.m <- melt(drn.posr, id.vars = 1:5) # ditto the above but
with
reduced vars
cst_hrly <- cast(drn.posr.m, day.y ~ hour.y, length) # gives hourly
# daily summary
write.csv(cst_hrly, "H:\\R\\stockport2\\hrly.count.csv",row.names=F)
write.csv(cst_hrly,
"H:\\R\\stockport2\\daily.counthrly.csv",row.names=F)
#write.csv(cst_daily, file="cstp_daily.csv")
avghrly <- cast(drn.posr.m, day.y + hour.y ~ variable, mean)# gives
the mean of
Jtime& speed
write.csv(avghrly, "H:\\R\\stockport2\\avghrly.csv",row.names=F)
daily.counthrly <- cst_hrly
daily.counthrly <- t(daily.counthrly)
colnames(daily.counthrly) <-
c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8","Sun9","Mon10")
#hist(daily.counthrly[,2], ylab="Frequency",
# xlab="Hourly count per day", main="Histogram plot of 3 Sept 2012")
pr <- pairs(daily.counthrly,main="Scatter plot of hourly count for 3-
10 Sept 2012"
,panel=panel.smooth,col.smooth="red",cex.main=1.0)
#pmt <- plot(daily.counthrly[,2],daily.counthrly[,3],main="Scatter
plot of 15-minutes count", xlab="Monday", ylab="Tuesday",cex.main=1.0)
# To obtain the sum total of row and column based on the total daily
#count.
#daily.count <- cast(drn.posr.m, day.y ~ variable, length,
# margins=c("grand_col", "grand_row"))
#write.csv(daily.count,
"H:\\R\\stockport2\\daily.count.csv",row.names=F)
# To obtain the summary based on 15-minute daily count.
daily.count15 <- cast(drn.posr.m, day.y ~ hour.y + min15.y, length)
write.csv(daily.count15,
"H:\\R\\stockport2\\daily.count15.csv",row.names=F)
daily.count15 <- t(daily.count15)
colnames(daily.count15) <-
c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8",
"Sun9","Mon10")
#hist(daily.count15[,2], ylab="Frequency",
# xlab="Daily count (15-minute average)", main="Histogram plot of 3
Sept 2012")
```

```
pr <- pairs(daily.count15, main ="Scatter plot of 15-min count for 3-
10 Sept 2012", panel= panel.smooth, col.smooth="red",cex.main=1.0)
pmt <- plot(daily.count15[,2],daily.count15[,3],main="Scatter plot of
15-minutes count",
xlab="Monday", ylab="Tuesday")
pmt.lm <- lm(daily.count15[,3]~daily.count15[,2])
abline(pmt.lm,col="red")
summary(pmt.lm)
legend("topleft", legend="Adjusted R-sq=0.84", cex=0.6)
legend("left", legend="Sept 3", cex=0.6)
# To obtain the summary based on 10-minute daily count.
daily.count10 <- cast(drn.posr.m, day.y ~ hour.y + min10.y, length)
write.csv(daily.count10,
"H:\\R\\stockport2\\daily.count10.csv",row.names=F)
daily.count10 <- t(daily.count10)
colnames(daily.count10) <-
c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8",
"Sun9","Mon10")
#hist(daily.count10[,2], ylab="Frequency",
# xlab="Daily count (10-minute average)", main="Histogram plot of 3
Sept
2012")
pr <- pairs(daily.count10,main="Scatter plot of 10-min count for 3-10
Sept 2012"
,panel=panel.smooth,col.smooth="red",cex.main=1.0)
pmt <- plot(daily.count10[,2],daily.count10[,3], main="Scatter plot of
10-minutes count",
xlab="Monday", ylab="Tuesday")
# To obtain the summary based on 5-minute daily count.
daily.count5 <- cast(drn.posr.m, day.y ~ hour.y + min5.y, length)
write.csv(daily.count5,
"H:\\R\\stockport2\\daily.count5.csv",row.names=F)
daily.count5 <- t(daily.count5)
colnames(daily.count5) <- c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8",
"Sun9","Mon10")
hist(daily.count5[,2], ylab="Frequency",
xlab="Daily count (5-minute average)", main="Histogram plot of 3 Sept
2012")
pr <- pairs(daily.count5,main="Scatter plot of 5-min count for 3-10
Sept 2012"
,panel=panel.smooth,col.smooth="red",cex.main=1.0)
pmt <- plot(daily.count5[,2],daily.count5[,3],main="Scatter plot of 5-
minutes
count",
xlab="Monday", ylab="Tuesday")
#abline(pmt,col="red")
# computes the vehicle count as well as the mean of time and speed
#cast(drn.posr.m, day.y + hour.y ~ variable, c(length, mean),
# subset = variable %in% c("jtime", "spd"))
####################################################################
# Summarise based on a particular day on a chosen link
drn.posrid <- drn.posri[drn.posri$day.y==3,]# change the index
#accordingly

# Order the data by datetime
drn.posrid <- drn.posrid[order(drn.posrid[,4]),1:11]
# change file in order not to overwrite the previous one
write.csv(drn.posrid,
"H:\\R\\stockport2\\drn.posrid3.csv",row.names=F)
attach(drn.posrid)
drn.posrid$journey times.cut <- cut(as.numeric(jtime),10)
plot(jtime, spd, main="Plot of Journey Time against Speed",
xlab="Time (sec)", ylab="Speed (km/h)", pch="+")
# The follwing demonstrates k-means clustering with R.
tsec <- jtime
#Apply kmeans to the data, and store the clustering result in kc.
#The cluster number is set to 3.
(kc <- kmeans(tsec, 10))
```

```
# Compare the class labels with the clustering result
table(drn.posrid$journey times.cut, kc$cluster)
plot(tsec, col = kc$cluster,main="Clustering of Journey Time")
#points(kc$centers[,c("spd", "jtime")], col = 1:10, pch = 8, cex=2)
# The follwing demonstrates k-means clustering with R.
speed.kmph <- spd
# Apply kmeans to the data and store the clustering result in kc.
# The cluster number is set to 10.
(kc <- kmeans(speed.kmph, 10))
# Compare the class labels with the clustering result
table(drn.posrid$journey times.cut, kc$cluster)
plot(speed.kmph, col = kc$cluster)
boxplot(spd, ylab="Speed (km/h)", las=1, cex.axis=0.8,
main="Box Plot of Journey Speed")
legend("bottomright",legend=c("mean=",round(mean(spd)),
"sd =",round(sd(spd))),cex=0.8)
legend("topright",legend=c("Sept 3, 2012"),cex=0.6)
## savePlot(filename = "Rplot", type = c("pdf"),device=postscript,
# restoreConsole = TRUE)
# dev.copy2pdf(device=postscript, out.type = "pdf")
#summary(spd)
detach(drn.posrid)
####################################################################
## The following performs data summary by first converting a list to a
dataframe
# This section helps to carry out the entire network summary at a go
# reduce the file size
# drn.posr <- lapply(drn_pos,function(drn_pos)drn_pos[c(1:2,8:14,16)])
# drn.posr <- lapply(drn_pos,function(drn_pos)drn_pos[c(1,10:14,16)])
drn.posrd <-ldply(drn.posr)
## To obtain sum (mean or....) use the following
drn.posrd.m <- melt(drn.posrd, id.vars = 1:9)
# drn.posrd.m <- melt(drn.posrd, id.vars = 1:5)
count_daily <- cast(drn.posrd.m, Site.ID.y + day.y ~ hour.y, length) #
#gives daily summary per each station
write.csv(count_daily, "H:\\R\\stockport2\\count_dly.hrly.csv")
#write.csv(cst_daily, file="cst_daily.csv")
mean_daily <- cast(drn.posrd.m, Site.ID.y + day.y + hour.y ~ variable,
mean)# gives the mean of time & speed
write.csv(mean_daily, "H:\\R\\stockport2\\mean_dly.hrly.csv")
# to obtain the sum total of row and column based on the total daily #
#count.
dlymean <- cast(drn.posrd.m, Site.ID.y + day.y ~ variable, mean)
write.csv(dlymean, "H:\\R\\stockport2\\dlymean.csv")
#,margins=c("grand_col", "grand_row"))
attach(avghrly)
bxt <- split(jtime, day.y)
boxplot(bxt, col = "lavender", notch = FALSE, varwidth = TRUE,
main="Boxplot of hourly journey time", ylab="Time(secs)",
xlab="Weekdays (Mon - Mon)",xaxt="n")
mtext(side=1,at=1:8,text=c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8",
"Sun9","Mon10"))
sapply(bxt, sd)
sapply(bxt, mean)
# plot journey speed
bxv <- split(spd, day.y)
boxplot(bxv, col = "grey", notch = FALSE, varwidth = TRUE,
main="Boxplot of hourly speed",xlab="Weekdays (Mon - Mon)",
ylab="Speed (km/h)",xaxt="n")
mtext(side=1,at=1:8,text=c("Mon3","Tue4","Wed5","Thu6","Fri7","Sat8",
"Sun9","Mon10"))
sapply(bxv, sd)
sapply(bxv, mean)
detach(avghrly)
# computes the vehicle summary such as mean time and speed etc
## 15-minute average
sum_dm15 <- cast(drn.posrd.m, Site.ID.y + day.y + hour.y + min15.y ~
variable,
```

```
c(sd, mean), subset = variable %in% c("jtime", "spd"))
write.csv(sum_dm15, "H:\\R\\stockport2\\sum_avg.sd15.csv")
######################################################################
# Models to test if there is any difference between the days
# transpose the data
hrly <-t(cst_hrly) # open and close cst_hrly before running the next
line
hrly <-data.frame(hrly)
colnames(hrly) <-c("Mon", "Tue","Wed", "Thu", "Fri", "Sat", "Sun",
"Mon")
# analysis of variance and model testing
mod1 <- lm(Mon ~ Tue, data=hrly)
summary(mod1)
# Model 2
mod2 <- update(mod1, . ~ . + Wed, data=hrly)
summary(mod2)
# Model 3
mod3 <- update(mod2, . ~ . + Thu, data=hrly)
summary(mod3)
#Model 4
mod4 <- update(mod3, . ~ . + Fri, data=hrly)
summary(mod4)
# Model 5
mod5 <- update(mod4, . ~ . + Sat, data=hrly)
summary(mod5)
# Model 6
mod6 <- update(mod5, . ~ . + Sun, data=hrly)
summary(mod6)
# Model 7
mod7 <- update(mod4, . ~ .- Fri, data=hrly)
summary(mod7)
# An ANOVA to judge if we are supposed to drop Sat and Sun
anova(mod7, mod4)
# Model for the Mon -Thu
mod.4 <- lm(Mon ~ Tue + Wed +Thu, data=hrly)
summary(mod.4)
# Model for the Sat-Sun
mod.2 <- lm(Sat ~ Sun , data=hrly)
summary(mod.2)
#####################################################################
## Perform analysis of variance (AOV)
daytest5 <- read.csv("~/R/daytest5.csv")
head(daytest5)
attach(daytest5)
plot(aov(count~weekdays))
summary(aov(count~weekdays))
#rmv<-weekdays!="Fri7"
summary(aov(count~weekdays,subset=weekdays!="Fri7"))
# remove Friday to Sunday from the data
Page 14
Bluetooth_script
rmv <-(weekdays!="Fri7"&weekdays!="Sat8"&weekdays!="Sun9")
summary(aov(count~weekdays,subset=rmv))
## Remove only Sunday and Saturday to test the significance
rmv<-(weekdays!="Sat8"&weekdays!="Sun9")
summary(aov(count~weekdays,subset=rmv))
summary.lm(aov(count~weekdays)) ##summary based on 5-minute count
aj <- lm(count~weekdays)
## Note that aov summary appears in alphabetical order
summary.lm(aov(count~weekdays,subset=rmv))
## post analysis
an <- aov(count~weekdays)
postan<- TukeyHSD(x=an, 'weekdays', conf.level=0.95)
postan
library(agricolae) # a simplified version of the above
HSD.test(aj, 'weekdays')
# HSD.test(an, 'weekdays') # aliter
#####################################################################
```

```
# Time series analysis, decomposition and classification
# A time series of hourly vehicle count over some days
##################################################################
#daily <- cst_daily # Daily contains the hourly count per day
#daily$Site.ID.y <- NULL
daily <- cst_hrly
#daily=read.csv("H:\\R\\stockport2\\daily.csv",header=T)
daily <- t(daily)
daily <- ts(daily)
#plot(daily[,1],type="b", xaxt="n",ylab="Daily count",xlab="Time
(Hr)",
# xlim=c(1,24), ylim=c(0,450))
plot(daily[,1], xaxt="n",ylab="Hourly count per day",xlab="Time (Hr)",
ylim=c(0,450), main="Plot of hourly count")
for(l in 1:4){
lines(daily[,l],col=l,lty=l)}
xaxislab <- seq(1:25)
axis(1, at=1:25, labels=xaxislab, las=1,cex=0.2)
legend("topleft",lty=c(1:4),col=c(1:4),c("Mon","Tue","Wed","Thu"))
#####################################################################
# Time series analysis using a linear filtering
bt <-read.csv("H:\\R\\bt.altrincham.csv",header=T)
plot(bt[,5],type="l", ylab="Trend",
main="Time series analysis using linear filtering")
bt.1 <- filter(bt[,5],filter=rep(1/5,5))
bt.2 <- filter(bt[,5],filter=rep(1/25,25))
bt.3 <- filter(bt[,5],filter=rep(1/81,81))
lines(bt.1,col="red")
lines(bt.2,col="purple")
lines(bt.3,col="blue")
rm(bt)
#####################################################################
# Daily count
daily5 <-read.csv("H:\\R\\stockport2\\daily.count5.csv",header=T)
daily5 <- t(daily5)
bt <- daily5
plot(bt[,1],type="l", ylab="Trend",
main="Time series analysis using linear filtering")
bt.1 <- filter(bt[,5],filter=rep(1/5,5))
bt.2 <- filter(bt[,5],filter=rep(1/25,25))
bt.3 <- filter(bt[,5],filter=rep(1/81,81))
lines(bt.1,col="red")
lines(bt.2,col="purple")
lines(bt.3,col="blue")
#####################################################################
# Times series analysis
bt.ts <- t(bt)
#bt.ts <- ts(bt,frequency=12,start= c(2011, 10),end=c(2012, 03))
bt.ts <- ts(bt,frequency=12,start= c(2011))
plot(bt.ts[,5], ylab="Trend",main="Time series analysis"
,xaxt="n")
for(m in 1:8){
lines(bt.ts[,m],col=m,lty=m)
}
# plot each profile on a different pan
plot(daily,main="Time series analysis (Mon-Mon)",col=2, cex.main=1.0)
#####################################################################
# exploring the relationships between two (or more) quantitative
#variables.Some ideas from #Stackoverflow
# Interactively choose file bt.altrincham
# bluetooth <- read.csv(file.choose())
#par(mfrow=c(1,1))
#bluetooth <- bt
bluetooth <- data.frame(daily5) # daily5 contains 5-minute count/day
colnames(bluetooth) <- c("Mon", "Tue", "Wed", "Thu","Fri","Sat",
"Sun","Mon")
attach(bluetooth)
# names(bluetooth)
```

```
boxplot(bluetooth[,1:8],col="grey",
notch=T, varwidth=T, las=1, tcl=.5,
xlab=expression("Weekdays"),
ylab=expression("5-minute Bluetooth count"),
main=")
#######################################################################
# analysis btw station pair . This section will be executed if many-
many
# merging is done
# mg2 is a list containing the computed time and speed and other
variables
# within the dataframes
mg1 <- svt[[1]]
pos <-mg1[mg1$jtime>0,]
neg <-mg1[mg1$jtime<0,]
pos <-pos[c(4,12)]
neg <-neg[c(4:5)]
p1 <- as.data.frame(table(pos))
n1 <- as.data.frame(table(pos))
p1 <-p1[p1$day.x==4,]
n1 <-p1[p1$day.x==4,]
# order the data by hour
p1 <- p1[order(p1$hour.y),c(1:3)]
n1 <- n1[order(n1$hour.y),c(1:3)]
# Covariance of two variables
# import link3839
cov(p1$Freq, link3839)
# Correlation of two variables
cor(p1$Freq, link3839)
cor(n1$Freq, link3938)
num <- as.numeric(p1$hour.y)
map.class <- cut(num, 24) #Division into 24 classes (1-24hrs)
pplot <-plot(link3839, p1$Freq, col=c(2:25))
abline(lm(p1$Freq ~ link3839))
nplot <-plot(link3938, p1$Freq)
abline(lm(n1$Freq ~ link3938))
#######################################################################
# The file size is reduced before performing O-D summary. Note that
#duplicates have been removed from the data
# The O-D result is a symmetric matrix. However, if the direction
# of travel is to be considered, then we have to reverse the operation
bt.count <- lapply(bt.rdup,function(bt.rdup) bt.rdup[c(1:4)]) #station
#data
bt.countl <- lapply(drn_pos,function(drn_pos) drn_pos[c(1:4)]) # link
#data
# bt.count1 <- bt.count[[1]] # first element of the list bt.count
ct <- 1 # ct =1 for daily O-D summary else the total summary
n=1:8
if(ct<1){
bt.countx <- bt.count
stn.count <- function(bt.count){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.count$VehicleId))
return(int)
}
count <- lapply(bt.count,stn.count)
ddcount <- ldply(count)
for(n in 1:length(bt.count)){
stn.count <- function(bt.count){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.count$VehicleId))
return(int)
}
count <- lapply(bt.count,stn.count)
ddcount[n] <- ldply(count)
}
colnames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
```

```
"stn8","stn9")
rownames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
"stn8","stn9")
ddcount
write.csv(ddcount, "H:\\R\\stockport2\\ddcount.csv")
} else{
# O-D summary on daily basis
# bt.count <- subset(bt.count, bt.count$day==3)
bt.count <- lapply(bt.count,function(bt.count)
bt.count[bt.count$day==3,])
bt.countx <- bt.count
stn.count <- function(bt.count){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.count$VehicleId))
return(int)
}
count <- lapply(bt.count,stn.count)
ddcount <- ldply(count)
for(n in 1:length(bt.count)){
stn.count <- function(bt.count){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.count$VehicleId))
return(int)
}
count <- lapply(bt.count,stn.count)
ddcount[n] <- ldply(count)
}
colnames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
"stn8","stn9")
rownames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
"stn8","stn9")
write.csv(ddcount, "H:\\R\\stockport2\\ddcount3.csv")
}
ddcount
#paste("Today is", date())
######################################################################
# Determine the number of Ids tracked in both directions
# This gives the number of unique Ids tracked i.e it lists all
# the intersection points. This helps to understand the ids that made
# return journey. x and y are the 2 stations under consideration
#int.xy<-intersect(drn.xy$VehicleId,drn.yx$VehicleId)
######################################################################
# This function lists all the ids that make a return journey
attach(svt1)
int <- function(svt1){
int.xy <- list()
for(i in 3:10){
svtx <- svt1[svt1$jtime<0&svt1$day.x==i,] # opposite direction
svty <- svt1[svt1$jtime>0&svt1$day.y==i,] # forward direction
int.xy <-intersect(svty$VehicleId,svtx$VehicleId)
}
return(int.xy)
}
int.res <- int(svt1)
detach(svt1)
######################################################################
## summary per link for all days. change the index in turn to compute
## the entire network on link basis
#id.count <- ddply(ldnegr, .(Site.ID.x,day.x,VehicleId), "nrow")# get
#the countof variable VehicleId
#write.csv(id.count,file="idcount.csv")
# lines 225-226 take time to run
# create a summary of the data based on the specified variable
#avgtsec <- ddply(ldnegr, .(Site.ID.x, day.x),
summarise,Vcount=length(VehicleId),
```

```
# min_jtime= min(jtime), max_jtime= max(jtime), mean_jtime=
mean(jtime))
var.sum15n <- ddply(ldnegr, .(Site.ID.x, day.x, min15.x),
summarise,Vcount=length(VehicleId),
min_spd = min(spd), mean_spd = ceiling(mean(spd)), med_spd =
median(spd),
max_spd = max(spd), med_journey times = median(jtime),mean_journey
times =
ceiling(mean(jtime)))
write.csv(var.sum15n, "H:\\R\\stockport.neg\\var.sum15n.csv")
#write.csv(var.sum15n, "H:\\R\\wigan.neg\\var.sum15n.csv")
#var.sum15n <- var.sum15n[order(var.sum15n$todp.class),1:8]
var.sum15n.d3 <- var.sum15n[var.sum15n$day.x==3,] # change this
#accordingly
var.sum15n.d3$Vcount.pct <-
round((var.sum15n.d3$Vcount/sum(var.sum15n.d3$Vcount))*100,2)
palette <- c("red","yellow","blue","green","orange")
# map.class <- var.sum15n.d3$todp.class
#library(reshape)
# summarise based on link
drn.negri<-drn.negr[[1]] # change the index in turn according to the
list length
#drn.negri$jtime<- abs(drn.negri$jtime) # to obtain absolute value of
JOURNEY TIMES
# classify the Bluetooth count according to the peak and off-peak
periods
todp.class <- rep("0 - 07hrs", times=nrow(drn.negri))
todp.class[drn.negri$hour.x>=7&drn.negri$hour.x<10] <- "07 - 10hrs"
todp.class[drn.negri$hour.x>=10&drn.negri$hour.x<16] <- "10 - 16hrs"
todp.class[drn.negri$hour.x>=16&drn.negri$hour.x<20] <- "16 - 20hrs"
todp.class[drn.negri$hour.x>=20] <- "20 - 24hrs"
drn.negri$todp.class <- factor(todp.class)
boxplot(drn.negri$spd ~ todp.class, horizontal=T, xlab="Speed (km/h)",
las=1, cex.axis=0.8, main="Box Plot of Journey Speed2")
# Includes options to draw the boxes and labels horizontally
abline(v=mean(drn.negri$spd), lty="dashed")
# Adds the mean value to the plot
legend("topleft", legend="Grand Mean", lty="dashed")
#todp_sum <- tapply(drn.negri$spd,drn.negri$todp.class,summary)
tapply(drn.negri$spd,drn.negri$todp.class,summary)
tod.count5n <- ddply(drn.negri, .(day.x,min5.x,spd,todp.class),
"nrow")
write.csv(tod.count5n, "H:\\R\\stockport.neg\\tod.count5n.csv")
#write.csv(tod.count5n, "H:\\R\\wigan.neg\\tod.count5n.csv")
plot(tod.sum15n.d3$mean_spd,tod.sum15n.d3$todp.class,ylim=c(0,6),
xlab="15-minute average speed (km/h)",ylab="Time of the day class",
main="Daily Speed Classification2",pch=21,bg=palette[map.class])
legvals <- c(0,7,10,16,20)
legend("right", legend=paste(">=",legvals),pch=21,
pt.bg=palette, pt.cex=1.5, bg="white",
title="DayTime classification")
# plot of percentage count by maximum spped (vmax) classification
#plot(tod.sum15n.d3$mean_spd,tod.sum15n.d3$Vcount.pct, ylim=c(0,12),
# xlab="15-minute average speed (km/h)",ylab="15-minute daily
Bluetooth
count (%)",
# main="Speed Profile2",pch=21,bg=palette[map.class])
#legend("topright",
legend=paste("<",tapply(as.numeric(tod.sum15n.d3$mean_spd),
# map.class, max)),pch=21,
pt.bg=palette, pt.cex=1.5, bg="white",
# title="DayTime class by Vmax")
# plot of percentage count by daytime classification
plot(tod.sum15n.d3$mean_spd,tod.sum15n.d3$Vcount.pct,ylim=c(0,12),
xlab="15-minute average speed (km/h)",ylab="15-minute daily Bluetooth
count
(%)",
main="Speed2 distribution over the day",pch=21,bg=palette[map.class])
```

```
legend("topright", legend=paste(">=",legvals),pch=21)
######################################################################
# The file size is reduced before performing O-D summary. Note that
duplicates have been removed from the data
# The O-D result is a symmetric matrix. However, if the direction
# of travel is to be considered, then we have to reverse the operation
bt.count <- lapply(bt.rdup,function(bt.rdup) bt.rdup[c(1:4)]) #station
#data
bt.countp <- lapply(drn_neg,function(drn_neg) drn_neg[c(1:4)]) # link
#data
# bt.count1 <- bt.count[[1]] # first element of the list bt.count
ct <- 1 # ct =1 for daily O-D summary else the total summary
n=1:8
if(ct<1){
#bt.countp <- bt.count ## for station summary
bt.county <- bt.countp
stn.countn <- function(bt.countp){
county <- bt.county[[n]]
intn <- length(intersect(county$VehicleId,bt.countp$VehicleId))
return(intn)
}
countn <- lapply(bt.countp,stn.count)
ddcountn <- ldply(countn)
for(n in 1:length(bt.countp)){
stn.countn <- function(bt.countp){
county <- bt.county[[n]]
intn <- length(intersect(county$VehicleId,bt.countp$VehicleId))
return(intn)
}
countn <- lapply(bt.countp,stn.countn)
ddcountn[n] <- ldply(countn)
}
colnames(ddcountn) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
rownames(ddcountn) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
# colnames(ddcountn) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
# rownames(ddcountn) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
ddcountn
write.csv(ddcountn, "H:\\R\\stockport.neg\\ddcountn.csv")
} else{
# O-D summary on daily basis
# bt.count <- subset(bt.count, bt.count$day==3)
bt.countn <- lapply(bt.countp,function(bt.countp)
bt.countp[bt.countp$day==3,])
bt.county <- bt.countn
stn.countn <- function(bt.countn){
county <- bt.county[[n]]
intn <- length(intersect(county$VehicleId,bt.countn$VehicleId))
return(intn)
}
countn <- lapply(bt.countn,stn.countn)
ddcountn <- ldply(countn)
for(n in 1:length(bt.countn)){
stn.countn <- function(bt.countn){
county <- bt.county[[n]]
intn <- length(intersect(county$VehicleId,bt.countn$VehicleId))
return(intn)
}
countn <- lapply(bt.countn,stn.countn)
ddcountn[n] <- ldply(countn)
}
```

```
colnames(ddcountn) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
rownames(ddcountn) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
```

```
# colnames(ddcountn) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
# rownames(ddcountn) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
write.csv(ddcountn, "H:\\R\\stockport.neg\\ddcountn.csv")
}
ddcountn
#paste("Today is", date())
######################################################################
## Link summary for the forward direction
if(ct<1){
bt.countx <- bt.countl
stn.count <- function(bt.countl){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.countl$VehicleId))
return(int)
}
count <- lapply(bt.countl,stn.count)
ddcount <- ldply(count)
for(n in 1:length(bt.countl)){
stn.count <- function(bt.countl){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.countl$VehicleId))
return(int)
}
count <- lapply(bt.countl,stn.count)
ddcount[n] <- ldply(count)
}
colnames(ddcount) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
rownames(ddcount) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
# colnames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
# rownames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
ddcount
write.csv(ddcount, "H:\\R\\stockport2\\ddcount.csv")
} else{
# O-D summary on daily basis
# bt.count <- subset(bt.count, bt.count$day==3)
bt.countd <- lapply(bt.countl,function(bt.countl)
bt.countl[bt.countl$day==3,])
bt.countx <- bt.countd
stn.count <- function(bt.count){
countx <- bt.countx[[n]]
int <- length(intersect(countx$VehicleId,bt.countd$VehicleId))
return(int)
}
count <- lapply(bt.countd,stn.count)
ddcount <- ldply(count)
for(n in 1:length(bt.countd)){
stn.count <- function(bt.countd){
countx <- bt.countx[[n]]
```

```
int <- length(intersect(countx$VehicleId,bt.countd$VehicleId))
return(int)
}
count <- lapply(bt.countd,stn.count)
ddcount[n] <- ldply(count)
}
colnames(ddcount) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
rownames(ddcount) <-
c("lk12","lk13","lk14","lk15","lk16","lk17","lk18",
"lk19")
Page 27
Bluetooth_script
# colnames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
# rownames(ddcount) <-
c("stn1","stn2","stn3","stn4","stn5","stn6","stn7",
# "stn8","stn9")
write.csv(ddcount, "H:\\R\\stockport2\\ddcountd.csv")
}
ddcount
######################################################################
# Determine the number of Ids tracked in both directions
# This gives the number of unique Ids tracked i.e it lists all
# the intersection points. This helps to understand the ids that made
# return journey. x and y are the 2 stations under consideration
#int.xy<-intersect(drn.xy$VehicleId,drn.yx$VehicleId)
######################################################################
# count of vehicles making a return journey
for(lk in 1:8){
for(di in 3:10){
svtx <- subset(svt[[lk]],svt[[lk]]$jtime<0)
svty <- subset(svt[[lk]],svt[[lk]]$jtime>0)
svtx <- svtx[svtx$day.x==di,] # opposite direction
svty <- svty[svty$day.y==di,] # forward direction
int.xy <- length(intersect(svty$VehicleId,svtx$VehicleId))
print(int.xy)
}
}
## End of the analyis on the opposite direction
######################################################################
## Plot of speed distribution
######################################################################
# Histogram plot of average journey speed
#attach(avghrly)
# import avghrly
#hs <- subset(avghrly,avghrly$day.y<7)
# hs <- hs$spd
hs <- avghrly$spd
#hist(hs,freq=TRUE) # for frequency plot
hist(hs,col="light blue",border="dark blue", freq=FALSE, xlab="Speed
km/h)",
main="Histogram of hourly speed")
legend("right",legend="Hourly average for 8 days", cex=0.6)
hist(hs,col="grey",border="dark blue",main="Histogram of hourly speed"
,cex=0.6, xlab="Speed km/h)", freq=FALSE)
legend("right",legend="Hourly average for 8 days", cex=0.6)
# Add a density curve
lines(density(sort(hs)),col="blue")
# Add a Normal curve
xhs = seq(from=0, to=70, by=0.1)
yhs = dnorm(xhs, mean(hs), sd(hs))
lines(xhs, yhs, lty="dotted",col="red")
rm(xhs, yhs)
legend("topleft", legend=c("density curve","Normal curve"),
lty=c("solid","dotted"),col=c("blue","red"),cex=0.6)
```

```
#####################################################################
## Plotting of points on Google map
#####################################################################
#coords$Easting <- as.numeric(substring(coords[,5],1,6))
#coords$Northing <- as.numeric(substring(coords[,5],7,12))
#write.csv(coords,file="coords2.csv")
coords <- read.csv("H:\\R\\coords2.csv",header=T)
attach(coords)
# A simple plot of point data
#####################################################################
## Plotting X, Y data on Google map
## Load required packages
library(maptools)
library(rgdal)
## Load the data for Bluetooth locations.
#bt.stns<- read.csv(file.choose()) # choose file coords2 interactively
## Inspect column headings
#bt.stns <- coords
#bt.stns <- read.csv(file="Bluetooth_stations.csv",header=TRUE)
BT_ATC_stations <- read.csv("~/R/BT_ATC_stations.csv")
## Inspect column headings
bt.stns <- BT_ATC_stations[,1:5]
head(bt.stns)

## Plot the XY coordinates
attach(bt.stns)
#attach(lonlat2)
# X= Easting
# Y= Northing
plot(X, Y)
#plot(Easting,Northing)
coordinates(bt.stns)<- c("X", "Y")
#coordinates(bt.stns)<- c("Easting", "Northing")
BNG<- CRS("+init=epsg:27700")
p4s <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
bt_wgs84 <- spTransform(bt.stns, CRS= p4s)
writeOGR(bt_wgs84, dsn="sensors.stn.kml", layer= "sites_wgs84",
driver="KML", dataset_options=c("NameField=name"))

detach(bt.stns)# create a simple colour palette which will be used to
split the region
palette <- c("yellow","green","red","purple")
# divide the region into class according to the easting coordinates
map.class <- cut(Easting, 4, labels=FALSE, include.lowest=TRUE)

plot(Easting, Northing, asp=1, main="Map of Bluetooth stations in
Greater Manchester", pch=21, bg=palette[map.class])
text(345000,410000,"Wigan Area")
text(365000,390000,"Altrincham Area")
text(395000,390000,"Stockport Area")
##################################################################
## Plotting Google Static Map
library(RgoogleMaps)
# Choose the coordinates file
bt_stations <- read.csv(file.choose())
# Create a simple colour palette which will be used to split the
region
#palette <- c("yellow","green","red","purple") # All Bluetooth
stations over UK
palette <- c("green","purple","red")
palette <- c("purple","red","green") # Manchester Bluetooth stations
attach(bt_stations)
# divide the region into class accordingly
map.class <- Location
# Plot the map
#plot(Easting, Northing, asp=1, main="", pch=21,
bg=palette[map.class])
#text(345000,410000,"Wigan Area")
```

```
#text(365000,390000,"Altrincham Area")
#text(395000,390000,"Stockport Area")
# Google Static Map plot
MyMap <- MapBackground(lat=latitude, lon=longitude)
PlotOnStaticMap(MyMap, latitude, longitude, pch=21,
bg=palette[map.class])
#legend("bottomright", legend=paste("<",tapply(Easting, map.class,
max)), pch=21,
#       pt.bg=palette, pt.cex=1.5, bg="white", title="Easting coords")
legend("bottomright", c("Birtley","Liverpool","Manchester"), pch=21,
             pt.bg=palette, pt.cex=1.5, bg="white", title="Study
Location")## simple geographical analysis
# Converting the data into a spatial object in R
detach(coords)
coords.xy <- coords
library(sp)
attach(coords.xy)
coordinates(coords.xy) <- c("Easting", "Northing")
# Converts into a spatial object
class(coords.xy)
detach(coords.xy)
# Demonstration of Google motion chart
library(googleVis)
ggmt <- read.csv("H:/R/avgspd.csv",header=T)
# ggmt <- avgspd
gm <- gvisMotionChart(ggmt, idvar="Site.ID.y", timevar="day.y")
plot(gm)
######################################################################
## Plotting data on Google map based on the ideas gained from
##http//:spatialanalysis.co.uk
## Load required packages
library(maptools)
library(rgdal)
## Load the data for Bluetooth locations.
#bt.stns<- read.csv(file.choose()) # choose file coords2 interactively
## Inspect column headings
#bt.stns <- coords
bt.stns <- read.csv(file="Bluetooth_stations.csv",header=TRUE)
## Inspect column headings
head(bt.stns)
## Plot the XY coordinates window.
attach(bt.stns)
# X= Easting
# Y= Northing
plot(X, Y) ## or use plot(Easting,Northing) depending on data format
######################################################################
## Processing and analysis of ANPR data
library(plyr) # advanced aggregation functions
library(lubridate) # datetime function
library(reshape)

MAC1070_2014.03.04_v2 <-
read.csv("V:/val_analysis/Disc_Graham_CeGComputing/Raw Bluetooth Data
A6/MAC1070_2014-03-04_v2.csv")
View(MAC1070_2014.03.04_v2)
MAC1071_2014.03.04_v2 <-
read.csv("V:/val_analysis/Disc_Graham_CeGComputing/Raw Bluetooth Data
A6/MAC1071_2014-03-04_v2.csv")

stn1070 <- MAC1070_2014.03.04_v2
stn1071 <- MAC1071_2014.03.04_v2
rm(MAC1070_2014.03.04_v2,MAC1071_2014.03.04_v2) # to conserve memory

bt.data <- list(stn1070,stn1071)

## function to reduce the file data size as desired
bt.data <- lapply(bt.data,function(bt.data) bt.data[c(1:2,10)])
```

```
head(bt.data[[2]])

## function to order the data by vehicleId
bt.data <-lapply(bt.data, function(bt.data)
bt.data[order(bt.data[,3]),1:3])
head(bt.data[[1]])# to examine part of the data

# apply time format to the list and remove duplicates from data using
# function fdup
#btr.data <- bt.order # assign a new name to the ordered data list

fdup <- function(bt.data){
  for(i in bt.data){
    # tm <- dmy_hms(btr.data$Date)
     tfx <-strptime(bt.data$Date,"%d/%m/%Y %H:%M:%S")
    #tfx <-strptime(bt.data$Date,"%Y-%m-%d %H:%M:%S")

    day <- day(tfx) # retrieve date value from the data

    hour <- hour(tfx) # extract hour component and add to df

    # Extract the minute component of dateTime and add to the df
    min <- minute(tfx)

    sec <- second(tfx) #This extract the seconds part of the data

    # compute time in seconds and add to station data
    tsec <- as.numeric(hour*3600 + min*60 + sec)
    # convert data to vectors to apply unique()
    #hour <- btr.data$hour
    #min <- btr.data$min

    # compute the 15-minute interval summary
    min15 <- floor(as.numeric(min)/15)

    # multiply by 15 for correct minutes format
    min15 <- min15*15   # the 1st 0-15mins is 0

    # compute the 10-minute interval summary
    min10 <- floor(as.numeric(min)/10)

    # multiply by 10 for correct minutes format
    min10 <- min10*10   # the 1st 0-15mins 0

    # compute the 5-minute interval summary
    min5 <- floor(as.numeric(min)/5)

    # multiply by 5 for correct minutes format

    min5 <- min5*5 # the first 0-5mins will be 0

    ## assign new variables to tsec and VehicleId
    y <- bt.data[,3]
    n <- length(y)

    ## compute time difference in seconds between successive points
    # secdif2 <- c(0,as.numeric(abs(tsec[-1]- tsec[-length(tsec)])))
    secdif <- c(0,as.numeric(abs(diff(tsec)))) #same result as above

    ## make a dataframe of the vectors
    bt.data <- data.frame(bt.data,day,hour, min15,min10, min5, secdif)

    ## remove the duplicate records from the data to obtain a subset
    y1<-y[2:n]
    y2<-y[1:(n-1)]
    yc<-as.character(y1)!=as.character(y2)
    bt.data$yc <-c("TRUE",yc)#add "TRUE" to the 1st pt to add up to pt
nos
```

```
    ndup <-
bt.data[(bt.data$yc=="FALSE"&bt.data$secdif>=300)|(bt.data$yc=="TRUE")
,]
    return(ndup)
  }
}
bt.data <- lapply(bt.data,fdup)

head(bt.data[[1]]) # to examine part of the data

# Function to reduce the file size before merging
bt.rdup <- lapply(bt.data,function(bt.data) bt.data[c(1:3)])
head(bt.rdup[[1]])

mg <- merge(bt.rdup[[1]], bt.rdup[[2]], by = "Vehicle.Id", sort=T,all
= FALSE)
head(mg)

dst <- 0.532  # A6, Stockport (1070-1071)

tfx <-strptime(mg$Date.x,"%d/%m/%Y %H:%M:%S")
tfy <-strptime(mg$Date.y,"%d/%m/%Y %H:%M:%S")

## Create time series from the data
wday <- weekdays(tfy)
day <- day(tfy)
hour <- hour(tfy)
min <- minute(tfy)
# compute the 15-minute interval summary
min15 <- floor(as.numeric(min)/15)
# multiply by 15 to obtain the minutes'proper format
min15 <- min15*15

#jtime<- difftime(tfy,tfx,units="secs")
 jtime<- difftime(tfy,tfx,units="auto")
 jtime<- as.numeric(jtime)
## Computation of vehicle speed begins here
tmin <- as.numeric(abs(jtime/60))
tmin <- as.numeric(sprintf("%.2f",tmin))

spd <- ceiling(as.numeric(abs(dst/(as.numeric(jtime/3600)))))

mg <- data.frame(mg,day,hour,min15,wday,jtime,tmin,spd)

# remove point data with different days merged together
mg <-subset(mg,day(tfx)==day(tfy))

# remove vehicles travelling at very low speed and at very high speed
(1st condition)
mg <- subset(mg,spd>5&spd<=120)

######################################################################
Require(openair)
scatterPlot(BTAN2, x = "ANPR_jtime", y = "jtime7170", group=NA,
            type = "default", method="scatter",linear = TRUE, ci =
FALSE,
            xlab="ANPR Journey Time (sec)", ylab="Bluetooth Journey
Time (sec)")


scatterPlot(BTAN2, x = "ANPR_spd", y = "spd7170", group=NA,
            type = "default", method="scatter",linear = TRUE, ci =
FALSE,
            xlab="ANPR Journey Speed (Km/h)", ylab="Bluetooth Journey
Speed (Km/h)")

scatterPlot(BTAN2, x = "ANPR7170N", y = "bt7170N", group=NA,
```

```
            type = "default", method="scatter",linear = TRUE, ci =
FALSE,
            xlab="ANPR flow (veh/15-min)", ylab="Bluetooth flow
(veh/15-min)")

timeVariation(BTAN2, pollutant = "flow_ratio",
              local.time = FALSE, normalise = F,ci =
TRUE,col="green2",
              xlab = c("hour", "hour", "month", "weekday"))

timeVariation(BTAN2, pollutant = "flow_ratio",
              local.time = FALSE, normalise = F,ci =
TRUE,col="green2",
              xlab = c("hour", "hour", "month", "weekday"))

timeVariation(BTAN2, pollutant = "flow_ratio",
              local.time = FALSE, normalise = F,ci =
TRUE,col="green2",
              xlab = c("hour", "hour", "month", "weekday"))

#######################################################################
## O-D Analysis
path.files <- "H:\\R\\wigan\\"
path.files <-"C:\\Users\\Ayodele\\Documents\\R\\wigan\\"
t.data <- lapply(list.files(path = path.files, pattern = ".csv"),
                  function(.file) read.csv(paste(path.files, .file,
                                                sep = ""),header =
TRUE))
#bt1<-bt.data[[1]]
#head(bt1) # to examine part of the data
#######################################################################
## Function to reduce the file data size as desired
bt.data <- lapply(bt.data,function(bt.data) bt.data[c(1:2,15)])
#btr1<-btr.data[[1]]
head(bt.data[[1]])# to examine part of the data
#######################################################################
## Function to order the data by vehicleId
bt.data <-lapply(bt.data, function(bt.data)
bt.data[order(bt.data[,3]),1:3])
head(bt.data[[1]])
#######################################################################
# Apply time format to the list and remove duplicates from data using
# function fdup
fdup <- function(bt.data){
  for(i in bt.data){
    #   tme <- dmy_hms(btr.data$Date)
      tme <- strptime(bt.data$Date,"%d/%m/%Y %H:%M:%S")
    #tme <- strptime(bt.data$Date,"%Y-%m-%d %H:%M:%S")
        day <- day(tme) # retrieve date value from the data

    hour <- hour(tme) # extract hour component and add to df

    # Extract the minute component of dateTime and add to the df
    min <- minute(tme)
        sec <- second(tme) #This extract the seconds part of the data
        # compute time in seconds and add to station data
    tsec <- as.numeric(hour*3600 + min*60 + sec)
    # convert data to vectors to apply unique()
    # hour <- bt.data$hour
    # min <- bt.data$min
    # compute the 15-minute interval summary
    min15 <- floor(as.numeric(min)/15)

    # multiply by 15 to obtain the minutes'proper format
    min15 <- min15*15 +15   # add 15 to make the 1st 0-15mins 15

    # compute the 10-minute interval summary
    min10 <- floor(as.numeric(min)/10)
```

```
    # multiply by 10 to obtain the minutes'proper format
    min10 <- min10*10 +10   # add 10 to make the 1st 0-15mins 15


    # compute the 5-minute interval summary
    min5 <- floor(as.numeric(min)/5)

    # multiply by 5 to give the  minutes a proper format
    min5 <- min5*5 +5 # add 5 so that the first 0-5mins will be 5

    ## assign new variables to tsec and VehicleId
    #x <- bt.data[,8]
    y <- bt.data[,3]
    n <- length(y)

    ## compute time difference in seconds between successive points
    # secdif2 <- c(0,as.numeric(abs(tsec[-1]- x[-length(tsec)])))
    secdif <- c(0,as.numeric(abs(diff(tsec)))) #same result as line 64

    ## make a dataframe of the vectors
    bt.data <- data.frame(bt.data, day, hour, secdif)

    ## remove the duplicate records from the data to obtain a subset
    y1<-y[2:n]
    y2<-y[1:(n-1)]
    yc<-as.character(y1)!=as.character(y2)
    bt.data$yc <-c("TRUE",yc)#add "TRUE" to the 1st pt to add up to pt
nos
    ndup <-
bt.data[(bt.data$yc=="FALSE"&bt.data$secdif>=300)|(bt.data$yc=="TRUE")
,]
    return(ndup)
  }
}
bt.rdup <- lapply(bt.data,fdup)

head(bt.rdup[[1]])
## Function for computing OD using merge option as well as removing
#outliers
#bt.count <- lapply(btc.rdup,function(btc.rdup) btc.rdup[c(1:5)])
#station data

bt.count <- lapply(bt.rdup,function(bt.rdup) bt.rdup[c(1:5)]) #station
#data

# bt.ctod <- lapply(bt.count,function(bt.count)
bt.count[bt.count$day==4,]) # daily

## Import the distance matrix
distM <- read.csv("H:\\R\\wigan_distM2.csv",header=T)

## Interactively choose the distance file
#distM <- read.csv(file.choose())
## Opt 0 or 1 according to whether flow or journey time OD is required
## flow = 0 and JOURNEY TIMES = 1
opt <- 0

if(opt == 0){
##Note: Number of days = 28, 29, 30, 31 depending on the month and
#year
#cycle through the selected hour and days
for(day_selec in 3:10){
  for(hour_selec in 0:23){

    bt.ctod <- lapply(bt.count,function(bt.count)
bt.count[bt.count$day==day_selec & bt.count$hour==hour_selec,]) #
hourly
```

```
    bt.count1 <- bt.ctod
    # k=2 # single (needed to compute between two station pairs)

    ddcount <-list()
    for(k in 1:length(bt.ctod)){
     # kk <- 0 # Initialise kk

      stn.count <- function(bt.ctod){

        countx <- bt.count1[[k]]
       # distM <- read.csv("H:\\R\\wigan_distM2.csv",header=T)

        st1 <- substring(as.character(countx[1,1]),10,13)
        st2 <- substring(as.character(bt.ctod[1,1]),10,13)

        stf <- substring(as.character(distM[,1]),4,7)
        stt <- substring(as.character(distM[,2]),4,7)

        dst <- distM[(as.numeric(st1)==as.numeric(stf))&
                        (as.numeric(st2)==as.numeric(stt)),c(3)]

        m.count <- merge(countx, bt.ctod, by = "VehicleId", sort=T,
all = FALSE)
        #m.count <- merge(countx, bt.ctod, by = "Vehicle.Id", sort=T,
all = FALSE)
        #m.count <- merge(countx, county, by = "VehicleId", sort=T,
all = FALSE)# single

        t.org <-strptime(m.count$Date.x,"%d/%m/%Y %H:%M:%S")
        t.dst <-strptime(m.count$Date.y,"%d/%m/%Y %H:%M:%S")

        #m.count$tdif.od <- difftime(t.dst,t.org,units="secs") # time
diffences btw origins and destinations
        tdif.od <- difftime(t.dst,t.org,units="secs")
# time diffences btw origins and destinations

        ## Computation of vehicle speed begins here
        jt <- as.numeric(tdif.od/3600)
        # thr <- as.numeric(sprintf("%.2f",journey timesme))
        tmin <- as.numeric(abs(journey timesme*60))
        tmin <- as.numeric(sprintf("%.2f",tmin))

        spd <- ceiling(as.numeric(abs(dst/journey timesme)))

        m.count <- data.frame(m.count,tdif.od,tmin,spd)
        #m.count <- data.frame(m.count,tdif.od,tmin)

        # remove vehicles travelling at very low speed and at very
high speed
        m.count <- subset(m.count,spd>5&spd<=120)

        ## remove the vehicles travelling in opposite direction
        m.count <- m.count[m.count$tdif.od>0,]

        # remove outliers from the data i.e. compute the outlier data
points
        ## Rule of thumb to remove outliers (Crawley,2005)

        upquant <- quantile(m.count$tdif.od,0.75) +
1.5*(quantile(m.count$tdif.od,.75)-quantile(m.count$tdif.od,0.25))
        lwquant <- quantile(m.count$tdif.od,0.25) -
1.5*(quantile(m.count$tdif.od,.75)-quantile(m.count$tdif.od,0.25))

        # compute the data range free of outliers
        m.count <-subset(m.count ,
m.count$tdif.od<=upquant&m.count$tdif.od>=lwquant)
```

```
        ## count the number of vehicles
        c.org <- nrow(m.count)

        return(c.org)
      }
      count <- lapply(bt.ctod,stn.count)
      ddcount[k] <- ldply(count)
    }
    ddcount <- ldply(ddcount)

    #Aliter
    #this bit didn't work
    #diag(as.matrix(ddcount))<-0

    #this will make the diagonals zero though
    for(c in 1:length(ddcount[1,])){ddcount[c,c]=0}

    ## Assign column and row names to the variables
    ## Stockport
    #colnames(ddcount) <-
c("stn33","stn34","stn35","stn36","stn37","stn38","stn39","stn40","stn
41")
    #rownames(ddcount) <-
c("stn33","stn34","stn35","stn36","stn37","stn38","stn39","stn40","stn
41")
    ## Wigan
    colnames(ddcount) <-
c("stn12","stn16","stn18","stn21","stn24","stn26","stn29")
    rownames(ddcount) <-
c("stn12","stn16","stn18","stn21","stn24","stn26","stn29")
    ## Trafford
    #colnames(ddcount) <- c("stn1001","stn1002","stn1008","stn1011")
    #rownames(ddcount) <- c("stn1001","stn1002","stn1008","stn1011")

    ## write the results to the specified file
    #res_path="c:/od/result/"
    #res_path <-"C:\\Users\\Ayodele\\Documents\\R\\wigan2\\"
    res_path <- "H:\\R\\wigan_utsg\\"
    fname <-
paste(res_path,"od_d",day_selec,"_h",hour_selec,".csv",sep='')
    write.csv(ddcount,fname,quote=F )# This writes the result to a
#folder

  }}
} else {
  for(day_selec in 3:10){
    for(hour_selec in 0:23){

      bt.ctod <- lapply(bt.count,function(bt.count)
bt.count[bt.count$day==day_selec & bt.count$hour==hour_selec,]) #
hourly

      bt.count1 <- bt.ctod
      # k=2 # single (needed to compute between two station pairs)

      ddcount <-list()
      for(k in 1:length(bt.ctod)){
        # kk <- 0 # Initialise kk

        stn.count <- function(bt.ctod){

          countx <- bt.count1[[k]]
          # distM <- read.csv("H:\\R\\wigan_distM2.csv",header=T)

          st1 <- substring(as.character(countx[1,1]),10,13)
          st2 <- substring(as.character(bt.ctod[1,1]),10,13)

          stf <- substring(as.character(distM[,1]),4,7)
```

```
          stt <- substring(as.character(distM[,2]),4,7)

          dst <- distM[(as.numeric(st1)==as.numeric(stf))&
                         (as.numeric(st2)==as.numeric(stt)),c(3)]

          m.count <- merge(countx, bt.ctod, by = "VehicleId",
sort=T,all = FALSE)
          #m.count <- merge(countx, bt.ctod, by = "Vehicle.Id",
sort=T,all = FALSE)
          #m.count <- merge(countx, county, by = "VehicleId",
sort=T,all = FALSE)# single

          t.org <-strptime(m.count$Date.x,"%d/%m/%Y %H:%M:%S")
          t.dst <-strptime(m.count$Date.y,"%d/%m/%Y %H:%M:%S")

          #m.count$tdif.od <- difftime(t.dst,t.org,units="secs") #
time diffences btw origins and destinations
          tdif.od <- difftime(t.dst,t.org,units="secs") # time
diffences btw origins and destinations

          ## Computation of vehicle speed begins here
          jt <- as.numeric(tdif.od/3600)
          # thr <- as.numeric(sprintf("%.2f",journey timesme))
          tmin <- as.numeric(abs(journey timesme*60))
          tmin <- as.numeric(sprintf("%.2f",tmin))

          spd <- ceiling(as.numeric(abs(dst/jt)))

          m.count <- data.frame(m.count,tdif.od,tmin,spd)
          #m.count <- data.frame(m.count,tdif.od,tmin)

          # remove vehicles travelling at very low speed and at very
high speed
          m.count <- subset(m.count,spd>5&spd<=120)

          ## remove the vehicles travelling in opposite direction
          m.count <- m.count[m.count$tdif.od>0,]

          # remove outliers from the data i.e. compute the outlier
data points
          ## Rule of thumb to remove outliers (Crawley,2005)

          upquant <- quantile(m.count$tdif.od,0.75) +
1.5*(quantile(m.count$tdif.od,.75)-quantile(m.count$tdif.od,0.25))
          lwquant <- quantile(m.count$tdif.od,0.25) -
1.5*(quantile(m.count$tdif.od,.75)-quantile(m.count$tdif.od,0.25))

          # compute the data range free of outliers
          m.count <-subset(m.count ,
m.count$tdif.od<=upquant&m.count$tdif.od>=lwquant)

          ## compute the hourly average journey time in seconds for
the vehicles
          jt.org <- m.count$tdif.od
          #c.org <- round(mean(jt.org),0)
          c.org <- nrow(m.count)
          c.org <- round(c.org*6.0606,0) # multiply the flow by the
inverse of penetratiin rate
          ## compute the hourly average speed for the vehicles
          speed.org <- m.count$spd
          x <- round(mean(speed.org),0)

          c.org <- (602.8 + 50.25*x - 0.7237*x^2 + 0.009258*x^3 -
0.00002583*x^4)*c.org

          return(c.org)
        }
        count <- lapply(bt.ctod,stn.count)
```

```
      ddcount[k] <- ldply(count)
    }
    ddcount <- ldply(ddcount)

    #Aliter

    #This will make the diagonals zero though
    for(c in 1:length(ddcount[1,])){ddcount[c,c]=0}

    ## Assign column and row names to the variables
    ## Stockport
    #colnames(ddcount) <-
c("stn33","stn34","stn35","stn36","stn37","stn38","stn39","stn40","stn
41")
    #rownames(ddcount) <-
c("stn33","stn34","stn35","stn36","stn37","stn38","stn39","stn40","stn
41")
    ## Wigan
    colnames(ddcount) <-
c("stn12","stn16","stn18","stn21","stn24","stn26","stn29")
    rownames(ddcount) <-
c("stn12","stn16","stn18","stn21","stn24","stn26","stn29")
    ## Trafford
    #colnames(ddcount) <- c("stn1001","stn1002","stn1008","stn1011")
    #rownames(ddcount) <- c("stn1001","stn1002","stn1008","stn1011")

    ## write the results to the specified file
    res_path <- "H:\\R\\wigan_utsg\\"
    fname <-
paste(res_path,"od_d",day_selec,"_h",hour_selec,".csv",sep='')
    write.csv(ddcount,fname,quote=F )# This writes the result to a
folder

  }}
}
#################################################################
#### Program to summarise SCOOT data based on 15-minute average flow

## Read in the required SCOOT file(s)
N12643T_3940 <- read.csv("~/R/scoot/N12643T_3940.csv")

N12642F_4039 <- read.csv("~/R/scoot/N12642F_4039.csv")

## Drop column 9 from the data

N12642F_4039[9] <- NULL

## Create the time format for the data
tfx <- strptime(N12642F_4039$Time,"%H:%M:%S")

## Create time series from the data
hour <- hour(tfx)
min <- minute(tfx)

# compute the 15-minute interval summary
min15 <- floor(as.numeric(min)/15)

# multiply by 15 to obtain the minutes in proper format
min15 <- min15*15 +15

N12642F_4039 <- data.frame(hour,min15, N12642F_4039)

## Subset for the complete days

N12642F_4039 <- subset(N12642F_4039, Day!="Mo")

# Create the summary of the data using doBy function
```

```
scoot_4039 <- summaryBy(FLOW+Norm_occ+OCC~Day+hour+min15,
data=N12642F_4039, FUN=c(sum))

#write.csv(cst_hrly, "H:\\R\\stockport2\\hrly.count.csv",row.names=F)
write.csv(scoot_4039, "H:\\R\\scoot\\scoot_4039.csv",row.names=F)
################################################################
## Date Filter Function
##fym <- function(bt.data){
  for(ym in bt.data){
    tfx <- strptime(bt.data$Date,"%Y-%m-%d %H:%M:%S")

    bt.data$year <- year(tfx) # retrieve date value from the data

    bt.data$month <- month(tfx)

    bt.data$day <- day(tfx)
    return(bt.data)
  }
}
bt.data <- lapply(bt.data,fym)
head(bt.data[[1]])
# subsetting for a specific year and month(s)
bt.r <- lapply(bt.data, function(bt.data)
  subset(bt.data,year==2013&month==4))
head(bt.r[[1]])
bt.r <- lapply(bt.r,function(bt.r) bt.r[c(1:3)])
head(bt.r[[1]])
####################################################################
## Function to generate series of date time
# Generate 15-min time series for a 31-day month
mnth <- rep("2013/07/",2976)
dy <- rep(1:31,each=96)
hr <- rep(rep(00:23,each=4),31)
min <- rep(c(00,15,30,45),744)
sec <- rep( 00,2976)
date <- paste(mnth,dy," ",hr,":",min,":",sec,sep="")
###################################################################
## R codes to compute Mahalanobis distance using Bluetooth data by
#E.G. Ayodele, Newcastle University, United Kingdom.
# e.g.ayodele@newcastle.ac.uk. 2016 Edition
## Last modified on 7th December 2016. Codes adapted from: 1) Dr. Jon
#Starkweather, Research and Statistical Support consultant, and 2)
#https://stat.ethz.ch/R-manual/R-
#devel/library/stats/html/mahalanobis.html
require(graphics)
library(rgl)
library(chemometrics)
#Remove the date column and save as another name to preserve file
link0506 <- read.csv("C:/B0925688/Other_Results/link0506.csv")
#link0506 <- na.exclude(link0506)
date <- link0506[,1]
x <- link0506[,c(3:6)] # subsetting for only mean JT and speed
#directional flows
#x <- na.exclude(x)
stopifnot(mahalanobis(x, 0, diag(ncol(x))) == rowSums(x*x))

#use the Moutlier function to compute MDs
md.ratio <- Moutlier(x, quantile = 0.95, plot = FALSE)

#Find the cut-off value
cut.off <- round(md.ratio$cutoff,3)

MD <- round(md.ratio$md,3)

#Summarise result
summary(MD)

#Add the computed MDs to the dataframe
```

```
x <- data.frame(x,MD)
#Add date to preserve time series
x2 <- data.frame(date,x)
#Plot the individual MDs if necessary to visualise
qqplot(MD, x$mean_jt, plot.it = TRUE, xlab = "Mahalanobis' distance",
       ylab = "Mean Journey Time (s)", main = "")

##Remove the outlying data points based on the computed cut-off value
x2.md <- subset(x2,MD <=cut.off )
#Plot the individual MDs if necessary to visualise
qqplot(MD, x2.md$med_jt, plot.it = TRUE, xlab = "Mahalanobis'
distance",
       ylab = "Median Journey Time (s)", main = "")

#summary(x2.md)
## Plot the MDs against Chi Square distribution
x2.md <- x2.md[,c(2:5)]
summary(x2.md)
stopifnot(mahalanobis(x2.md, 0, diag(ncol(x2.md))) ==
rowSums(x2.md*x2.md))
Sx <- cov(x2.md)
D2 <- mahalanobis(x2.md, colMeans(x2.md), Sx)
plot(density(D2, bw = 0.5),
     main="Squared Mahalanobis distances, n=27740, p=3") ; rug(D2)
qqplot(qchisq(ppoints(27740), df = 3), D2,
       main = expression("Q-Q plot of Mahalanobis" * ~D^2 *
                         " vs. quantiles of" * ~ chi[3]^2))

# Compare the Mahalanobis' distances of each data file with simple
#histograms
par(mfrow = c(1,2))
hist(x2$MD, main = "", xlab= "Unfiltered MD")
hist(x2.md$MD, main = "", xlab="Filtered MD")

#Average the filtered data (x2.md) preferrably on daily basis for
#clarity
library("openair", lib.loc="C:/Program Files/R/R-3.0.2/library")
x2.md.plot <- subset(x2,MD <=cut.off )
x2.md.plot <- x2.md.plot[,c(1:5)]
dly.sd <- timeAverage(x2.md.plot, avg.time = "day", statistic = "sd")
dly.mean <- timeAverage(x2.md.plot, avg.time = "day", statistic =
"mean")
#Make a ts data
dly.sd.plot <- ts(dly.sd)
dly.mean.plot <- ts(dly.mean)
#Plot the data
plot(dly.sd.plot[,c(2:3)], plot.type="single",
     #main="Plot of Standard Deviation of Flow",
     ylab="Standard Deviation of Speed", xlab= "Index of Time",
col=c("blue", "red"), lwd=1)
legend(10,8, legend=c("sd_FlowNE","sd_FlowSW"),col=c("blue",
"red"),lty=1,
        cex=0.8, lwd=1, border ="lty", box.col="white")

##NE Directional Ratio
plot(dly.mean.plot[,3],ylab="Speed (km/h)", xlab= "Index of Time",
col="red")
legend(10,51, legend=c("Speed"),col=c( "red"),lty=1,
       cex=0.8, lwd=1, border ="lty", box.col="white")

plot(dly.sd.plot[,3],ylab="Standard Deviation of Speed", xlab= "Index
of Time", col="blue")
legend(220,3, legend=c("sd_Speed"),col=c( "blue"),lty=1,
        cex=0.8, lwd=1, border ="lty", box.col="white")
##SW Directional Ratio
#par(mfrow=c(1,1))
plot(dly.mean.plot[,5],ylab="Journey Time (s)", xlab= "Index of Time",
col="red")
```

```
legend(2,90, legend=c("Journey Time"),col=c( "red"),lty=1,
        cex=0.8, lwd=1, border ="lty", box.col="white")

plot(dly.sd.plot[,5],ylab="Standard Deviation of Flow", xlab= "Index
of Time", col="blue")
legend(220,33, legend=c("sd_Journey Time"),col=c( "blue"),lty=1,
        cex=0.8, lwd=1, border ="lty", box.col="white")
##Total Directional ratio
plot(dly.mean.plot[,4],ylab="Ratio", xlab= "Index of Time", col="red")
legend(20,0.115, legend=c("RatioTotal"),col=c( "red"),lty=1,
        cex=0.8, lwd=1, border ="lty", box.col="white")

plot(dly.sd.plot[,4],ylab="Standard Deviation of Ratio", xlab= "Index
of Time", col="blue")
legend(20,0.057, legend=c("sd_RatioTotal"),col=c( "blue"),lty=1,
        cex=0.8, lwd=1, border ="lty", box.col="white")
######################################################################
timeVariation(x2.md.plot, pollutant = c("Ratio_NE", "Ratio_SW",
"Ratio_Total"),
              local.time = FALSE, normalise = F,ci =
TRUE,col=c("blue","red","orange"),
              xlab = c("hour", "hour", "month", "weekday"), ylab="Flow
Ratio")
######################################################################
# ARIMA Modelling using Bluetooth data by E.G. Ayodele
#(e.g.ayodele@ncl.ac.uk)
## Reference: Data Splitting in R by Jason Brownlee, 2014
## http://machinelearningmastery.com/how-to-estimate-model-accuracy-
#in-r-using-the-caret-package/
# Hyndman, R.J. and Athanasopoulos, G. (2013) Forecasting: principles
#and practice. OTexts. Available at: http://otexts.org/fpp/
#R and Data Mining: Examples and Case Studies by Yanchang Zhao
# http://www.RDataMining.com
library(openair)
library(caret)
library(klaR)
library(forecast)
# read in the data
ts0506 <-
read.csv("C:/b0925688/VDriveCopy220116/trafford2013/ts0506.csv")
##Remove the column containing day, hour and min15 to reduce the data
#size
ts0506 <- ts0506[c(1,5:6)]
# Define an 80%/20% train/test split of the dataset.
split=0.80
trainIndex <- createDataPartition(ts0506$jtime, p=split, list=FALSE)
data_train <- ts0506[ trainIndex,]
data_test <- ts0506[-trainIndex,]
# Convert test data to time series
jt_test <- ts(data_test[,2], start=c(2013, 1), end=c(2013,
12) ,frequency=12)
seasonplot(jt_test, type="b", ylab="Journey Time
(s)",xlab="Time",main="")
# Make daily average from the training data set
jt_train <- timeAverage(data_train, avg.time = "day")
# Convert data to time series
jt_train <- ts(jt_train[,2])#, start=c(2013, 1), end=c(2013, 12),
frequency=12)
#Plot data to explore series
plot(jt_train, type="b", ylab="Journey Time (s)",xlab="Time",main="")
par(mfrow=c(1,2))
Acf(jt_train,main="")
Pacf(jt_train,main="")
acf(log(jt_train),main="")
pacf(log(jt_train),main="")
# Difference and transform the data
acf(diff(log(jt_train)),main="")
pacf(diff(log(jt_train)),main="")
```

```
tsdisplay(diff(jt_train),main="")
# train an arima model
(fit <- arima(log(jt_train), c(0, 1, 1),seasonal = list(order = c(0,
1, 1), period = 12)))
fita <- auto.arima(log(jt_train),seasonal=FALSE)
fit0 <- Arima(log(jt_train),order=c(0,1,1))
fit1 <- Arima(log(jt_train),order=c(1,1,1))
fit2 <- Arima(log(jt_train),order=c(0,1,2))
(fits <- arima(log(jt_train), c(0, 1, 2),seasonal = list(order = c(0,
1, 2), period = 12)))
#par(mfrow=c(1,2))
summary(fit)
summary(fita)
summary(fit0)
summary(fit1)
summary(fit2)
summary(fits)
# Plot the residuals of the chosen model
#plot(residuals(fit), type="b",ylab="Journey time residuals")
plot(residuals(fit), ylab="Residuals of journey time")
# make predictions
pred <- predict(fits, n.ahead = 2*12)
ts.plot(jt_train,2.718^pred$pred, log = "y", lty = c(1,3), col=
c(2,4),ylab= "Journey time (s)")
pred_corr <- 2.718^pred$pred
pred_test <- ts(pred_corr, start=c(2013, 1), end=c(2013,
12) ,frequency=12)
## Aliter
plot(forecast(fit), main="", ylab= "Log of journey time (s)",
xlab="Time")
#plot(fcast <- forecast(fit),main="")
Box.test(residuals(fits), type="Ljung")
#Plot the two series for comparison
val <- cbind(pred_test,jt_test)
write.csv(val,file="H:\\R\\val.csv")
plot(val, plot.type="single",
     main="Plot of training and test data",
     ylab="Journey Time (s)",
     col=c("blue", "red"), lty=1:2)
legend("topleft", legend=c("Train","Test"), col=c("blue",
"red"),lty=1:2)
######################################################################
```