UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Civil Engineering and the Environment

Journey Time Estimation on Bus Routes

By

Jao-Shyan Chen

Thesis for the degree of Doctor of Philosophy

July 2006

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF CIVIL ENGINEERING AND THE ENVIRONMENT

Doctor of Philosophy

JOURNEY TIME ESTIMATION ON BUS ROUTES
By Jao-Shyan Chen

Bus transit operations are impacted by increasing traffic congestion, which results in unreliable bus services and uncertain bus passenger waiting time. Knowledge of bus journey time could provide accurate travel information through Advanced Traveller Information Systems (ATIS) to potential and actual users, so they may make the best travel choices. Alternatively, it enables the evaluation of the bus network for improving bus operation by the dispatcher or for local authorities' initiatives for bus priority.

This research develops bus journey time estimation models using regression and Monte Carlo simulation methodologies, which are built on link and route bases. The data applied for formulating the proposed models were collected on several bus routes in Southampton, UK, including GPS tracking of buses and traffic data from ANPR and SCOOT. The developed models are validated with independent field data.

The results indicated that general travel time, dwell time at bus-stops, control delay at signalised junctions, and delay of deceleration and acceleration due to bus-stop services are the major components of bus journey time. The results showed that regression models can give an acceptable estimate of expected journey time and have the advantage of using only some major independent variables. Monte Carlo models, which can provide information on the distribution of bus journey times due to variability caused by the fluctuation of traffic and passenger demand and signal timing, are shown to give better estimation with greater tolerance when variables have larger deviations. Bus journey time excluding dwell time and acceleration/deceleration delay is estimated to be 1.34 times of that other vehicles' journey time. This study also suggested that the number of stops made by a bus along a bus route, the critical junctions which may have longer signal control delay, and boarding time per passenger are the key factors on the variability of bus journey times, and hence the potential ways to improve them.

Contents

List of Tables

List of Figures

Acknowledgements

I wish to express my deepest thanks to Professor Mike McDonald who not only served as my supervisor but also encouraged and guided me throughout the study. I would like to thank many members from the Transportation Research Group, University of Southampton, who helped in various ways in the production of this thesis: to Mr. Brian Mould for his assistance with the operation of GPS equipment and the instrumented vehicle; to Dr. Yanying Li and Pengjun Zheng for their help of GPS data processing; to Mr. Eamonn Kemmard and Doug Robinson for administering the survey; to Dr. Birendra Shrestha, Andy Richards, Tom Cherrett, and Ben Waterson for their insights in the analysis of SCOOT data; and particularly to Dr. Nick Hounsell for his valuable comments in the initial planning of this study. Sincere thanks are also due to Mr. Doug Robinson and Mrs. Patricia Hollowell for proof reading the manuscript.

I would like to acknowledge the funding provided by the Education Ministry of Taiwan. I am also grateful to the Southampton ROMANSE Centre and in particular to Ray Morris and Simon Chapman for the assistance of SCOOT and ANPR data collection.

Finally, I am especially grateful to my family, and in particular my wife, Yen-Ping Yu, and my daughter, Chi-Chen Chen, for their endless support, encouragement and merriment throughout.

Nomenclature

The nomenclatures used throughout this thesis are as follows:

a	the semi-major axis of biaxial ellipsoids used in the UK (m)
A	the average acceleration rate (m/s/s)
AD	adherence of bus timetable
a_i	acceleration at GPS point i (m/s/s)
$ANPR_{\%}$	the percentage of ANPR journey time per kilometre
b	the semi-minor axis of biaxial ellipsoids used in the UK (m)
BJT_R	total time for a bus travelling along a route (sec)
$Buslane\%$	the percentage of bus lane length
BV	bus average speed along route (km/hr)
BV_i	link bus speed at i link (km/hr)
c	the height above ellipsoid (m)
C	average cycle length (sec)
D	the average deceleration rate (m/s/s)
D_{acc}	acceleration and deceleration delay due to dwells (sec)
DT	total stop times for a bus serves for alighting and boarding passengers (sec)
$D_{waiting}$	dummy variable of additional waiting time at bus-stops
E_i	the Easting of the i point (m)
G	average green time (sec)
i	link number
JT	total time for a bus travelling along a route, which condition is similar to other vehicles (sec)
JT_{ANPR}	ANPR journey time (sec/km)
JT_F	journey time in free flow traffic and unimpeded (sec)
JT_G	bus general travelling time (sec/km)
JT_g	time for a bus travelling along a link measured (sec)
JT_l	total time for a bus travelling along all links excluding dwell times and signal control delay (sec)
JT_T	total journey time of a link (sec)
l	location parameter of distribution
L	route length (km)
L_{link}	Link length (m)

L_{route}	route length (km)
N	number of observations
n	total number of link
N_{ai}	the number of alighting passengers at i stop
N_{bi}	the number of boarding passengers at i stop
$N_{bus-stop}$	the number of stops made by a bus along a route
N_{dist}	the number of disturbances per km
N_i	the Northing of the i point in meters
N_l	the number of left turns per km
N_r	the number of right turns per km
N_{sig}	the number of signalized junctions per km
N_{stop}	the number of bus stops per km
$N_{stopped}$	the number of stopped bus-stops per km
P	the dummy variable of peak/ off-peak period
P_A	Number of alighting passengers
P_B	Number of boarding passengers
$P_{Ei,Ni}$	the i position of easting and nothing (m)
P_R	the probability for a bus stopped by traffic signal
$P(\mathrm{Re}\,d)_i$	probability of a bus encountered a red light at i junction
P_T	Number of total passengers
ΔR	the difference in the ranks between data values in the same pair
R_{ave}	the average duration of red light (sec)
S	the total bus travel distance during deceleration and acceleration (m)
S_a	travel distance during acceleration (m)
S_d	travel distance during deceleration (m)
Sig	total signal control delay along bus route (sec)
Sig_i	signal control delay at i junction (sec)
ST_f	SCOOT flow parameter (veh/5 min)
ST_o	SCOOT occupancy parameter (%)
ST_s	SCOOT speed parameter (km/hr)
$ST_s\%$	the percentage of SCOOT speed parameter
T	the total acceleration and deceleration time per stop (sec)
T_a	time of acceleration (sec)
$T_{a/d}$	delay of deceleration and acceleration (sec)
T_{AD}	total delay of acceleration and deceleration due to dwells (sec)
T_{ai}	alighting time of each alighting passenger at i stop (sec)

T_{ANPR}	ANPR journey time (sec/km)
T_{bi}	boarding time of each boarding passenger at i stop (sec)
T_D	Bus dwell time per stop (sec)
T_d	time of deceleration (sec)
T_{Di}	dwell time at i bus-stop (sec)
t_{i-1}	time stamps at GPS point *i-1*
t_{i+1}	time stamps at GPS point *i+1*
t_i	the time of the i position
T_j	total time delayed at junctions (sec)
T_m	total running time (sec)
T_n	travel time without stopping (sec)
T_p	time saving due to bus priorities (sec)
$T_{\mathrm{Re}d-i}$	duration of red at i junction (sec)
T_s	delay at a signalised control (sec)
T_t	total bus journey time along a route or section of a route (sec)
V_1	the bus cruise speed before deceleration (km/hr)
V_2	the cruise speed after acceleration (km/hr)
V_{bus}	average bus speed along link (km/hr)
V_f	free-flow speed (km/hr)
v_{i-1}	speeds at GPS point *i-1* (km/hr)
v_{i+1}	speeds at GPS point *i+1* (km/hr)
v_i	the speed of the i position (km/hr)
$\overline{V_S}$	sample space mean speed
$\overline{V_t}$	sample time mean speed
x_i	the longitude of the i point in decimal degrees
Y	bus journey time (sec/km)
y_i	the latitude of the i point in decimal degrees
α	shape parameter of distribution
β	scale parameter of distribution
ρ	Spearman's rank order correlation coefficient
σ	standard deviation value
μ	mean value

Abbreviations used

The abbreviations used throughout this thesis are as follows:

A-D	Anderson-Darling
AGTPV	Average Gap Time Per Vehicle
ALOTPV	Average Loop Occupancy Time Per Vehicle
ANOVA	Analysis of Variance
ANPR	Automatic Number Plate Recognition
APC	Automatic Passenger Counting
APTS	Advanced Public Transportation Systems
ATIS	Advanced Traveller Information Systems
AVI	Automatic Vehicle Identification
AVL	Automatic Vehicle Location system
BIPS	Bus Information and Priority System
CB	Crystal Ball
CEP	Circular Error Probability
CfIT	The Commission for Integrated Transport
DETR	Department of Environment, Transport & the Regions
DfT	Department for Transport
DOP	Dilution Of Precision
DRMS	Distance Root Mean Squared
EPE	Estimated Position Error
ETC	Electronic Toll Collection
GIS	Geographic Information System
GLONASS	Global Navigation Satellite System
GPS	Global Positioning System
HCM	Highway Capacity Manual
HDOP	Horizontal Dilution Of Precision
HPE	Estimated Horizontal Position Error
ILD	Inductive Loop Detector
ITS	Intelligent Transportation Systems
K-S	Kolmogorov-Smirnov
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percent Error
MOVA	Microprocessor Optimized Vehicle Actuation

MSL	Mean Sea Level
NMEA	National Marine Electronics Association
OS	Ordnance Survey
OSGB36	Ordnance Survey Great Britain 1936
PDF	Probability Density Function
PDOP	Positional Dilution Of Precision
RMS	Root Mean Square
RMSE	Root Mean Squared Error
ROMANSE	ROad MANagement System for Europe
SA	Sensitivity Analysis
SCOOT	Split Cycle Offset Optimization Techniques
SD	Standard Deviation
SPOT	Signal Progression Optimization Technology
SPRINT	Bus SCOOT, Priority and InforMatics in Public Transport
SPSS	Statistical Package for the Social Sciences
TCQSM	Transit Capacity and Quality of Service Manual
TCRP	Transit Cooperative Research Program
TRB	Transportation Research Board
UTC	Urban Traffic Control
VPE	Estimated Vertical Position Error
VDOP	Vertical Dilution Of Precision
WGS84	World Geodetic System 1984

Chapter 1: Introduction

1.1 Background

Traffic congestion is now seen as a critical problem in British urban areas. The Commission for Integrated Transport [CfIT] (2001) when comparing European transport systems concluded that Britain has the most extensive traffic congestion in Europe. Several indicators confirm this trend, for example, the number of licensed vehicles in GB increased by 68% between 1980 and 2004, total traffic increased by 81%, while the total length of roads increased only by 10% (Department for Transport [DfT], 2004; 2006). Congestion costs had been estimated at about 3.2% of GNP in 1996 in UK, around 6.9 billion pounds for England at 1996 prices (Santos, 2000). It is a key problem for road network reliability. Therefore, some encouraging or restrictive approaches such as public transport and road user charging schemes are adopted for many cities to alleviate congestion.

Encouragement of public transport is one of the most important policies. In order to utilise limited roads more efficiently in the urban area, it provides regular fixed routes, fixed schedule or headway, and reasonable fares to travellers and aims to decrease the use of private vehicles. Among the modes of public transport, buses are the most used form for local journeys and the only public transport alternative to cars in many areas. Nearly two thirds of all public transport journeys in England are by bus (DfT, 2004). Increasing bus usage has been established as a goal in the UK, and local authorities aim to provide better services for buses as an attractive alternative to the car. The Government aims to increase the use of public transport (bus and light rail) in England by more than 12% from the 2000 level by 2010. Thus, how to improve buses service as an attractive alternative to cars is a crucial issue.

The most common suggestion in the public survey of how to improve the bus service is through more reliable and punctual buses except cheaper fares (DfT, 2006). It is also a critical factor that car users will leave their own cars at home (DfT, 2003a). Reliability is concerned with bus on-time performance when a bus schedule is available or regular headway between consecutive buses when buses run at frequent intervals (Transit Cooperative Research Program [TCRP], 2003a). However, it is affected by the increasingly worse traffic congestion. There may be two methods to approach this problem. One is the initiative, which can help buses to get through congestion more easily, such as bus priority. Such a method might be unavailable due to limited road space and/or it may place other vehicles at a disadvantage. The other is to provide passengers with accurate arrival information through Advanced Traveller Information Systems (ATIS). This can assist passengers with better travel decision-making, e.g. to put off the departure time from home, or choose another alternative mode. Such information can be achieved by collecting bus and traffic data, estimating, predicting, and then disseminating to passengers in quasi real-time. In order to obtain a reliable bus arrival time, it is vital to attain effective estimation of bus journey times. The objectives of this research work are outlined below.

1.2 Objectives

The principal aim of this study is to understand the variability of the journey times of buses and their relationship to the journey times of other vehicles for a range of network and traffic conditions in the urban area. The objectives are as follows:

- To analyse the characteristics of the components of bus journey time and their variability;
- To develop a model for estimating bus journey time under non-incident traffic conditions and validate the model by measured data;
- To identify relationships between the journey times of buses and other vehicles;
- To identify potential ways of improving bus journey time.

1.3 Approach

As described in the background, the need to use possible approaches in order to obtain more accurate bus journey time estimation is essential, hence the necessity of this study. The aim and the scope of this research is defined and the objectives are planned. This is followed by a review of available literature, to gain a better understanding of bus journey time related issues, such as research methodologies and journey time data collection methods, in order to ascertain the breadth of this research. Once the research methodologies and data collection method are identified (regression and Monte Carlo simulation; probe vehicles equipped with GPS), a field survey is carried out to identify the accuracy of the proposed collection method for the journey time survey. This is then followed by the collection of data from the field, which is required in the modelling process. A process of integration of GPS data and Geographic Information System software is developed for data processing. The collected data is then used to formulate the main components of bus journey time. Such components and collected data are manipulated to develop estimation models based on regression and Monte Carlo simulation approaches. These include both route-based and link-based modelling processes. The developed models are then validated with independent field data. The validated model is used to determine the sensitivity of the model outcomes to changes of its inputs, and to explore the impact on bus operations under a range of scenarios. An overall discussion is then justified to understand the ability and limitation of this research. Finally, main findings are summed up in conclusion. The sequence of the research processes are presented in Figure 1.1.

Chapter 2: An Overview of Bus Journey Time Related Issues

2.1 Introduction

Chapter 1 described the background, objectives, and intended approach of this research. This chapter describes the related issues with respect to bus journey time from the literature. This overview is essential to obtain the fundamentals of this topic and to ascertain the gap for this research to fill. The chapter starts by defining the journey time and describes the needs of link journey time (section 2.2). It then presents the methods that are available for journey time data collection (section 2.3). Next, methodologies for estimating journey time are illustrated (section 2.4). This is followed by a demonstration of bus usage trend, bus user needs, and potential bus developments (section 2.5). The relationship of journey time of buses and other vehicles is described in section 2.6. After that, the main characteristics of bus journey time are presented (section 2.7). Then, an illustration of bus journey time estimation is described (section 2.8). Finally, this chapter concludes with a summary.

2.2 Journey time

Traffic condition is a dynamic phenomenon with elements of both space and time. Journey time is the most useful measure of traffic performance based on these elements because journey time provides one of various validation checks that a route and/or network is behaving properly. Therefore, collecting and analysing the time traversed along a section could offer a possible solution to explore potential traffic problems. Furthermore, the concept of journey time is easy to understand, being not only accepted by operator and authority but also acknowledged by the public.

2.2.1 Journey time definition

Journey time may be defined as the time which is necessary to traverse a route between two points. It is comprised of running time and stopped delay time (Turnel et al, 1998). The running time means the time when a vehicle is in motion. The stopped delay time means the time when a vehicle stopped. Similarly, the journey time of a bus driving along a route can be divided as Equation 2.1 (Fernández, 1999):

$$T_t = T_m + T_j + T_s \tag{2.1}$$

where,

T_t =total bus journey time along a route or section of a route;

T_m =total running time;

T_j =total time delayed at junctions; and

T_s =total time taken at bus stops.

Journey time can be measured directly by calculating the time gap between departure and arrival. It can also be estimated by assuming the vehicle use at a constant average speed to travel along a roadway segment. Then, the journey time can be estimated as a segment length divided by average speed. In the literature, there are two main methods to calculate the average speed of vehicles and they should be distinguished. The first way is called time mean speed, which is an arithmetic mean speed of vehicles taken over a period of time (Equation 2.2). The other is space mean speed, which is the speed based on the average time taken to cross a specific segment of roadway (Equation 2.3).

$$\overline{V_t} = \frac{\Sigma v_i}{N} \tag{2.2}$$

where,

$\overline{V_t}$ = sample time mean speed;

v_i = speed of the ith vehicle; and

N = number of observations.

$$\overline{V_S} = \frac{D}{\frac{1}{N}\sum t_i} \tag{2.3}$$

where,

$\overline{V_S}$ = sample space mean speed;

D = distance travelled or length of roadway segment;

N = number of observations; and

t_i = travel time of i th vehicle.

It is crucial to distinguish between these two mean speeds. In free flow traffic, the difference between these two is quite small. However, in congested traffic or on a signalised road, the difference will be significant. Turner et al. (1998) indicated the relationship as Equation 2.4:

$$\overline{V}_t = \overline{V}_S + \frac{\sigma_S^2}{\overline{V}_S} \qquad (2.4)$$

Here σ_S^2 is defined as the variance of space speed. On motorways in free flow speed, most vehicles are driving at very high and similar speeds. As a result, σ_S^2 will be small and $\overline{V}_S$ will be large. Therefore, the two mean speeds are almost equal. By contrast, when there is great variability of speeds and their mean speed is relatively small, there will be considerable differences between them. However, to measure average travel speed correctly, the better way is to calculate space mean speed directly (Gartner et al, 1992). With respect to this, a data collection method such as automatic number plate recognition (ANPR) is superior to the spot detector of inductive loop detector, when both data sources are available.

2.2.2 Link journey time

Detailed components of link journey time can be found in Robinson and Polak (2004) suggesting that the journey time of an individual vehicle driving along an urban link can be divided into three parts as Equation 2.5.

$$JT_T = JT_F + T_D + T_S \qquad (2.5)$$

where,

JT_T =total journey time of a link;

JT_F =journey time in free flow traffic and unimpeded;

T_D =delay along the link; and

T_S =delay at a signalised control.

Delay at a signalised junction was studied by Quiroga and Bullock (1999). The total delay that is caused by signalised control includes deceleration, stop, and acceleration delay. Details of the trajectory of distance-time diagram are shown in Figure 2.1. The signalised control delay is calculated as Equation 2.6.

$$Ts=(t5-t1)-(D4-D1)/ V_f \qquad (2.6)$$

Where:

T_S = delay of signalised control

t1 and t5 = the time in which the vehicle passed through D1 and D4

D1 = deceleration which begins when the vehicle approach the junction

D4 = acceleration which ends when the vehicle passed the junction and return to original speed before D1

V_f = free-flow speed

However, in a real situation, it is very difficult to define where and when the deceleration begins (e.g. D1, t1) and the acceleration ends (e.g. D4, t5). The problem can be solved by locating main changes of speed utilising acceleration values (Quiroga and Bullock, 1999). They computed the acceleration by using GPS point speed with the following central difference Equation 2.7.

$$a_i = \frac{v_{i+1} - v_{i-1}}{t_{i+1} - t_{i-1}} \qquad (2.7)$$

where:

a_i = Acceleration at GPS point i;

v_{i+1}, v_{i-1} = Speeds at GPS point $i+1$ and $i-1$; and

t_{i+1}, t_{i-1} = Time stamps at GPS point $i+1$ and $i-1$.

When a vehicle approaches at a junction by a red light, the acceleration, a_i is significantly lower than zero. When the vehicle stops behind the stop-line, the acceleration a_i is zero. By contrast, when the signal turns green and the vehicle starts to leave the junction, the acceleration a_i is significantly higher than zero.

The descriptions above are the general content of an individual vehicle which passed a signalised junction at a red light. However, the similar trajectory may be applicable to the situation when a bus serves a bus stop. Thus, the bus link journey time is derived from the above scenarios, in which a bus serves a bus stop and passes a signalised junction in its red period, and is shown in Figure 2.2. It can be seen that a bus approaches a bus stop at D2 with deceleration during t1 and t2, stops for alighting and boarding passengers during t2 and t3, and leaves the bus stop with acceleration during t3 and t4. Then, the bus approaches a signalised junction at D7 with deceleration during t6 and t7, stops for a red light during t7 and t8, and leaves the junction with acceleration during t8 and t9.

The other critical issue is why a link journey time is required to estimate a route journey time. There are three major reasons for this, namely repeated reduction, ease of prediction, and time reduction. A simplified traffic network in Figure 2.3 is described. This network is comprised of several nodes (e.g. A, B, C, etc.) and links (e.g. $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, etc.) in order to understand the journey time from node A to E. It is easy to measure the time when a vehicle travels through the nodes A and E. The time difference between A and E is the route $\overline{AE}$ journey time. Correspondingly, it might be necessary to measure the journey time when a vehicle passes A and F, if the route $\overline{AF}$ journey time is concerned. However, there are many possible routes between any two nodes in networks. It is impossible to measure all the journey time for all routes between any two nodes, and this will have many repeats, which result in inefficiency and ineffectiveness. Therefore, it is more practical to measure journey time for each link and make necessary link combinations for concerned routes. In other words, the initial stage is to measure all possible links' travel time, which may be used for estimation in traffic network. Then, the route journey time could easily be calculated by adding the successive link journey times along the route. For example, the journey time of $\overline{AE}$ may be calculated as the aggregating of link journey time $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, $\overline{DE}$, or alternatively $\overline{AB}$, $\overline{BC}$, $\overline{CF}$, $\overline{FE}$.

The second reason for using link journey time to calculate route journey time is for ease of prediction. The journey time measured and aggregated as above is past information, that is, the vehicles went through checkpoints along the route at a specific past time. This is historical data, which may have diverse values. For the actual needs of a road user, the most possible route journey time in the near future, namely journey time prediction, may be more important than past information. Moreover, there are many factors which could affect the result of prediction and they may vary depending on spatial or temporal changes. In order to decrease the effects of possible factors, separated links which have simpler characteristics than route may be able to represent predictable units for route journey time.

Reduction of time measurement for longer distances is the third reason (Dunstan, 1997). For instance, if vehicle X takes just 15 minutes to travel from A to B (Figure 2.3), and vehicle Y takes just 25 minutes to travel from B to C, the estimated total journey time from A to C without link separation may require 40 minutes, yet the maximum estimated time is only 25 minutes (maximum link time of link $\overline{AB}$ and $\overline{BC}$) for link estimation separately. It is not necessary to take the additional time for measuring the journey time from A to C, especially when the route length is considerable longer than this.

2.3 Journey time data collection methods

Traditionally, journey times were derived from floating car surveys, with an observer on board recording the cumulative travel times at checkpoints or delays along the route. However, this method is labour intensive and prone to human error. As a result, it was soon replaced by computer aided instruments and advanced technologies.

Several techniques can be used to estimate journey time data. One of these is the inductive loop detector which is the most widely used and accepted traffic detector technology (Dailey, 1993; Sisiopiku and Rouphail, 1994; Westerman, 1994; Dunstan, 1997; Sen, et al., 1997; Petty et al., 1998; D'Angello et al., 1999; Zhang, 1999; Kwon et al, 2000; Xie et al., 2001). Other competing techniques include probe vehicles equipped with GPS, either using transit vehicles or test vehicles (Srinivasan and Jovanis, 1996; Sen, et al., 1997; Lin and Zeng, 1999; Hellinga and Fu, 1999; Li and McDonald, 2002; Bertini and Tantiyanugulchai, 2004); an automatic vehicle location (AVL) system based on dead reckoning technique (Cathy and Dailey, 2002; Cathy and Dailey, 2003); automatic vehicle identification (AVI) technology (Turner, 1996); automatic number plate recognition (ANPR) technology (Li and McDonald); and aerial surveillance (Puget Sound Regional Council [PSRC], 2000). The majority of these techniques used for journey time data collection purposes are either fixed detectors or moving detectors (Miwa and Morikawa, 2003). A review of these techniques is provided in the following subsections and a summary of the main advantages and disadvantages of the main techniques can be found at the end of this section.

2.3.1 Fixed detectors

A fixed detector is defined as a method of data collection that gathers traffic data at pre-selected or fixed points and/or at specific time. The detector is located on or under the pavement, such as inductive loop detector (ILD), or at the roadside such as ANPR, or on an overhead structure such as AVI. With these measurements, at least two detectors are required in a series, or only one (such as ILD) with particular algorithms to calculate journey time and estimate the average speed across the route or link.

An inductive loop detector is a wire embedded into or under the roadway in approximately a square. The loop works on the principle that a magnetic object near an electrical conductor causes an electrical current to be induced. That is to say, a vehicle acts as the magnetic object and the inductive loop as the electrical conductor. A device at the roadside records the signals generated or the signals are sent to central computer for processing. This technology has become the most widely used and accepted traffic detector technology. Split Cycle Offset Optimization Techniques (SCOOT), which uses this technique, is the main example of this. More than 130 cities worldwide have SCOOT system operational or being installed (Peek, 2004). This technique can be divided into two broad classes. One is a single use loop mainly designed to collect vehicle frequency and lane occupancy. Most existing highways are equipped only with this kind of loop. However, this single loop detector is unable to measure speeds directly. If speed is required, then an algorithm involving the assumption of average vehicle length, loop length, and loop location is used. Many methods have been developed to use speed-flow relationships to estimate vehicle speeds from single loop detectors. The other technique is dual use loops, one at each end of the link; this enables the estimation of the link travel time directly and is more accurate. However, it needs two loop detectors in series, increasing the initial cost, hence not many highways have it.

ANPR uses video cameras to collect vehicle license plate numbers and arrival times at various checkpoints. With computerized recognition method, it automatically matches the license plates between adjacent checkpoints and calculates travel times from the difference in arrival times. This technique provides continuous travel time during the data collection period and is able to obtain travel times among vehicles within the traffic stream. Furthermore, the video can be viewed repeatedly for further information. However, the accuracy of license plate recognition depends upon the correct set up of camera, angle of view, configuration, illumination, and weather.

Probe vehicles, which are designed mainly for collecting data in real-time, are already in the traffic stream for different purposes to collect journey time data. These vehicles' main purposes are for specific transport objectives other than data collection of test vehicles (Turner et al., 1998). For example, buses are used for passengers' transportation services; they could be probe vehicles after setting the instruments for data collection. Others such as taxis, commercial vehicles, and private automobiles are all possible probe vehicles. The probe vehicles are equipped with electronic transponders which are encoded with identifications. Antennas which are located on the roadside or on overhead structures emit radio frequency signals to detect the presence of probe vehicles. When the transponder vehicles enter the range of antennas, the capture of unique identification is sent to a roadside reader and attached with date, time, and antenna ID number. These data are then sent to a central computer for storage and processing. Travel times are calculated from the time difference between adjacent antenna locations. AVI technique is widely used in electronic toll collection (ETC), real-time traffic and incident monitoring, and traveller information.

2.3.2 Moving detectors

A moving detector is defined as a method for data collection, which continually gathers traffic data with sensors on board when the vehicle is travelling. Such vehicles may be probe vehicles, which was described in section 2.3.1 or test vehicles, which are for the survey purpose. The logging frequency of a on-board sensor can be less than one second for one record depending on the ability of the detector. Compared with fixed detectors, moving detectors cannot provide continuous data at a certain section, but they can collect data over a large area, and the data may contain a variety of detailed information, such as a vehicle's trajectory, running speeds, accelerations, decelerations, delays, and journey times, which may not be provided by fixed detectors. Available technologies such as a probe vehicle equipped with GPS or cellular telephone can be used for moving detectors. Alternatively, a more extensive traffic condition, such as network traffic distribution and congestion profiles, may be achieved from remote sensor technologies such as a hot air balloon, aerial surveillance (Angel et al., 2003) or satellite (Institute of Transport Research, 2004).

GPS is a system developed by the United States Department of Defence. It currently is made up of 27 satellites. They continuously transmit signals as reference points and are positioned so that receivers on earth may receive signals from six of them almost anywhere and at any time. GPS receiver calculates its position based on triangulation, which is based on the distance between the receiver and three or more satellites. Readers may refer to Aerospace (1999) and Garmin (2004) for more details about GPS. Probe vehicles equipped with GPS are designed to collect real-time positional information from vehicles which is then sent to a control centre for monitoring.

The technique of a probe vehicle equipped with a cellular phone utilizes an existing cellular phone network to collect travel time data. Vehicles with cellular phones on board are potential probe vehicles. The location of the probe vehicle is determined by a combination of lines of bearing and time difference of arrival calculations to locate a vehicle (Turner et al., 1998). However, the low accuracy of this technique is a major issue.

Aerial surveillance uses aircraft with a camera unit mounted on the underside; the video signals are transmitted to a control centre via microwave and processed for traffic management. The aircraft may be a helicopter, spy plane, blimp, satellite, or even a weather balloon. The main applications of this technique in transportation are to help detect and verify incidents, to identify alternative routes during major incidents, and to observe traffic results of incidents. It can also provide continuous traffic conditions over a range of network. Therefore, traffic data such as journey times, delays, and congestion pattern can be captured via the images. However, this technique may not be a cost-effective method for conventional data, and it is currently in developmental or testing stages and has not been extensively field-tested or applied.

2.3.3 Comparative summary of detector technologies

As described above, many technologies are available for journey time data collection. According to the detector position and continuous data provision, they are divided into two groups, namely fixed and moving detector measurements. Each group has several alternatives for journey time data collection purposes, but the state-of-the-art and competing techniques are the inductive loop detector, probe vehicles equipped with GPS, ANPR, probe vehicles equipped with AVI, and aerial surveillance. There is no best technology for all purposes of traffic survey in any situation; each has advantages and disadvantages. The summary of the characteristics for such detectors are provided in Table 2.1.

It can be seen in Table 2.1 that the graphical presentation on the second column shows a possible outcome of spatial and temporal data collected from various detectors. From these graphs, it can be seen that a moving detector can provide more detailed spatial-temporal data than the average data between two detector points of fixed detectors. The estimation of annual operation and maintenance cost on the fourth column is from Li (2004). So far, there is no reference relevant to the cost of aerial surveillance in traffic data collection. For one typical survey, for example using helicopter with cameras, the total cost may be inferred to be very high. However, it might be cost-effective if the total costs are converted into the unit cost of observed vehicles. Journey time and speed data are often used to identify and evaluate congestion patterns and trends. Therefore, it is more important that collected data could represent the position, amount of time, length, intensity, and variation of delay. The fifth column shows such an ability of data representation (Turner et al., 1998). In addition, the main disadvantage of each detector is summarized from Turner et al.(1998), PSRC (2000), and Li (2004).

It can be concluded from Table 2.1 that probe vehicles equipped with GPS are superior in term of cost-effectiveness, detailed spatial-temporal data, well delay representation, and is already proven by former studies. However, the main disadvantage is that samples are limited. This problem may be improved by the increasingly prevalent use of GPS fitting on transit vehicles, commercial fleets, and private cars due to downsizing, decreasing price and the function of route guidance. Consequently, there may be a marked increase in the near future in the use of probe vehicle with GPS collecting traffic data.

2.4 Journey time estimation methodologies

Most of the journey time estimation studies in the literature may be grouped into three categories, namely empirical, analytical, and simulation methodology, which are described in the following subsections.

2.4.1 Empirical methodology

The empirical approach uses simplified representation of transportation system or behaviour, which is based on field data or experiment. It includes an historical profile approach (Shbaklo et al., 1992), an algorithm to convert flow and occupancy into spot speed estimates (Dailey. 1993; Petty et al., 1998; Dailey, 1999; Coifman, 2001), regression method (Wardrop, 1968; Gault and Taylor, 1981; Sisiopiku and Rouphail, 1994; Zhang, 1998; Chakroborty and Kikuchi, 2004; Bertini and Tantiyanugulchai, 2004), and a combination of time series and regression (Waterson, 2005).

2.4.2 Analytical methodology

The analytical approach uses simplified mathematical and logical relationships to represent the complex reality and utilize field data to calibrate and validate. The most well known reference of this type is Highway Capacity Manual [HCM] (TRB, 2000).

Takaba (1991), Anderson and Bell (1998), and Shimuzi (2000) which used queuing theory requiring detailed signal timing and traffic parameters such as traffic flow, saturation flow etc. to estimate journey time. Nelson and Palacharla (1993), Dougherty (1995), and Anderson and Bell (1998) used a neural network method which needs considerable training data for the calibration of the model. Fuzzy logic, which explains vague situations, was used by Li and McDonald (2002). In addition, Chakroborty and Kikuchi (2004), and Bertini and Tantiyanugulchai (2004) used models which were derived from theoretical concepts or drawn from observations and calibrated by field data.

2.4.3 Simulation methodology

The simulation approach uses a computer-based numerical method to mimic a particular transportation facility or environment and represents traffic behaviour over temporal and spatial reality (TRB, 2000). Seneviratne (1988) used the Monte Carlo procedure which is based on the theoretical probability distribution of variables to simulate fixed route bus travel time. Abdelfattah and Khan (1998) used a micro-simulation approach to generate traffic data for a study area, which was used to develop bus delay models. Shrestha (2002) developed a microscopic simulation model to evaluate the performance of bus priority strategies under various scenarios.

2.4.4 Comparison of research methodologies

The summary of the main advantages and disadvantages of journey time estimation methodologies is presented in Table 2.2 (TRB, 2000). There is no single best methodology, which can apply to any circumstance. Thus, the best methodology should depend upon the target of research and the available resources such as data. The main purpose of this research is to estimate bus journey time on a bus route. Thus, the transferability may be a crucial consideration. In addition, the objective of understanding the components and variability of bus journey time may require more comprehensive consideration of possible variables. Furthermore, considerable databases such as SCOOT, ANPR, and AVL based on GPS might be accessible for this study. Therefore, regression approach of empirical methodology, which has the ability to explore the possible factors which may affect bus journey time, is an alternative. In addition, Monte Carlo approach of simulation methodology, which may have a more comprehensive view of possible estimations with probabilities, the ability to understand the outcomes to changes of specific variables, and the transferability might be more relaxed than other simulation approaches such as micro-simulation at a specific bus route, is also an alternative. In fact, mixed methodologies, which use two or all of the above methodologies, may be adopted to carry out model development. For example, simulation models usually utilize former analytical and empirical models to form part of the system components or behaviour (Seneviratne, 1988; Abdelfattah and Khan, 1998). A further discussion of the methodology of bus journey time estimation could be found in section 2.8.3.

2.5 Bus user needs and potential development

2.5.1 The trend of bus usage

Buses are the most commonly used form of public transport choice for local journeys and the only public transport alternative to cars in many areas. Nearly two thirds of all public transport journeys in England are by bus (DfT, 2003a). Increasing bus use has been established as a goal in the UK for many years. Local authorities aim to provide better services for buses as an attractive alternative to the car. The objective in the UK transport ten year plan of 2000 was to increase bus use in England by 10% from the 2000 level by 2010, and meanwhile to improve bus service in terms of punctuality and reliability.

Bus and coach traffic increased from 3.5 to 5.2 billion vehicle kilometres, between 1980 and 2002. However, total traffic increased by 77 percent, from 277 to 490 billion vehicle kilometres over the same period. During 1999/2001 the average person made 1,019 trips per year in total, but only 57 trips (6%) per person per year on local buses, a decline of 22% since 1989/1991 when an average of 73 trips per year were made (DfT, 2003a). Consequently, as may be seen in Figure 2.4, bus passenger journeys have decreased steadily from 1985 to 1993. A plateau may be seen from 1994 to 2000 after which there is a slightly upward trend.

In order to achieve the goal of the UK transport ten-year plan by increasing bus use, a number of areas have made significant progress in delivering bus services. The number of local bus passenger journeys in England increased by 3 percent to 3.9 billion journeys in 2002-2003. However, there were various situations in different areas. For example, in London, because of its network character, there was about 8 percent growth (DfT, 2004).

An average of 17% of bus trips were made each day on weekdays. The main purpose of these trips was commuting and business, education, and shopping related. 13% of a week's trips were made on Saturdays and 5% on Sundays, while most of these trips were for shopping and recreation purposes.

Patterns of bus use during the day varied between weekdays and weekend (Figure 2.5). On weekdays, bus trips peaked during the periods of 8 – 9 am and 3 – 5 pm. The evening peak was wider than the morning one. This is because trips from home to work and school are at similar times in the morning, but trend to be spread out during the evening. At weekends, there is only one peak for bus trips around midday, the periods of 10 am – 2 pm on Saturdays and 10 –12 am on Sundays respectively (DfT, 2003a).

2.5.2 Bus user needs

A survey for understanding public attitudes to transport in England was carried out by Market & Opinion Research International for the Commission for Integrated Transport (CfIT, 2000). It indicated that the most effective policies for reducing car use are as follows: About 26% of car drivers indicated that public transport subsidies to keep fares down would reduce their driving. Additionally, around one in four of car drivers indicated they would drive less if there were improved bus services, more bus services, and more park and ride schemes, followed by about one in five who expressed this attitude if there were charges for city centre access, and greater co-ordination of bus/rail. On the other hand, for bus users, the top priority is reducing fares, followed by improving the frequency and reliability of services, personal security (on buses and while waiting at stops) and increasing the number of places that can be accessed.

It is necessary to make better use of buses to help reduce congestion. Buses services have to achieve following objectives suggested by DfT (2004):

- Attractiveness: to attract car users to use them.
- Mobility: to provide access to any interested destinations that other modes of transport cannot reach.
- Punctuality: to give buses priority in congested locations and improve the fare collection system in order to speed boarding.

- Frequent and reliable services: with real-time information equipment to help operators to run reliable services and to tell passengers when the next buses will arrive at bus stops, and/or with easily obtained information from the internet, mobile phone, etc.
- Seamless: to integrate bus service with other transport modes such as rail, ferry, or flight.
- Safety: to provide safety on board, in bus shelters, at bus stops, and to and from bus stops.

2.5.3 Bus potential development

A scan of the internal and external environment of buses is a crucial part of the planning process. The internal factors of the environment can be classified as strengths (S) and weaknesses (W), and the external factors can be classified as opportunities (O) and threats (T). SWOT analysis, which is usually used in strategic planning of management, is a very effective way to identify the buses resources and capabilities in the competitive environment where they operate (Fleisher and Babette, 2003). A SWOT analysis for buses compared with private automobiles is shown in Table 2.3.

It is vital to understand that a range of schemes is available for improving bus services. However, there is no off-the-rack answer that enables maximum advantages to buses without considering their locations. The most useful scheme for a specific location should depend upon its local conditions. Therefore, in order to understand the problems of bus services, it is crucial to observe their operations involving the identification of problems and opportunities such as specific location of delays, heavily used corridors, high frequency/patronage routes, and an entire bus network approach DfT (2003b).

2.6 Bus journey time in relation to other vehicles

Bus journey time can be estimated either by itself or in relation to another mode (Transit Cooperative Research Program [TCRP], 2003b). The former approach is described in section 2.8 and the latter approach, in relation to other vehicles, particularly automobiles, is presented in the following subsections. The aim of this comparison is to understand the additional time required or possible journey time rate (buses journey time/ automobiles journey time) for buses in order to estimate from other vehicles with available urban traffic control systems. Furthermore, if the additional time is considerable or the journey time rate is high, it may be difficult to attract passengers from their automobiles.

2.6.1 The advantages and disadvantages of bus journey time compared with other vehicles

With respect to journey time, buses lose a great deal of time due to frequent bus-stops in comparison with other vehicles. Only some time can be gained from bus priorities. The advantages and disadvantages of buses journey time compared with other vehicles are summarised in Table 2.4.

2.6.2 Previous research methods and results

There are two major approaches for studying the relationship between the journey time of buses and other vehicles in the literature. One is statistical analysis; the other is the comparison of spatial-temporal trajectory graphs between buses and test vehicles, which is described in following subsections.

2.6.2.1 Statistical approach

Levinson (1983) concluded that car speeds were 1.4 to 1.6 times as fast as bus speeds (including dwell time), which is based on surveys conducted in several U.S. cities. McKnight and Paaswell (1997) developed a regression model for Manhattan CBD Transit that indicated the relationship between the journey time of the buses and other vehicles (McKnight et al., 2004). It identified that bus journey times were 1.75 times slower than other vehicles journey times. TCRP (2000) reported the bus journey times were determined by various stop spacing, dwell times, and the operating environment such as a central business district or arterial roads, and hence various values under different circumstances were provided. Bertini and Tantiyanugulchai (2004) used two data sets, namely buses and test vehicles equipped with GPS, to compare the difference of journey time and speed between them. The result indicated that the test vehicle journey times were 1.38 times the bus journey times when buses were travelling at the recorded maximum speed for entire journey. Chakroborty and Kikuchi (2004) concluded that the average journey time of other vehicles was equal to free-flow journey time plus that bus journey time excluding dwell time multiplied by a 0.14 or 0.18 when less or more frequently congested roads were used. McKnight et al.(2004)

developed a regression model for New Jersey in comparison with the study carried out by McKnigtht and Paaswell (1997). The result indicated that bus journey times were 1.37 times other vehicles journey times, which value was less than the Manhattan study due to various operating environments.

In brief, according to the literature results, bus journey times are 1.37 to 1.75 times slower than other vehicle journey times.

2.6.2.2 Trajectory comparison

Bertini and Tantiyanugulchai (2004) compared the operation between buses and test vehicles using time-distance trajectory. It might be a useful approach to understand where and when the buses gains or losses compared with other vehicles. In addition, by marking the road layout and facilities such as junctions and bus stops associated with the travelling distance, it is possible to identify where the buses are delayed and whether the delay affects only buses or all vehicles. Such an approach is used in a pilot study to explore the possible outcome. The trajectories of a bus, a probe vehicle, and an additional trajectory of bus excluding dwell time are shown in Figure 2.6. The figure of time-distance, time-speed, and time-acceleration/deceleration are presented in the figure from top to bottom respectively. Note that the bus and the probe vehicle did not depart at the same time. In order to make comparison for convenience, the departure times are both set from 00:00:00 (hh:mm:ss). The explanations of these figures are as follows:

- Time-Distance profile

It can be seen from the top diagram of Figure 2.6 that junctions and bus-stops are marked beside the distance axis. The bus journey time was close to the probe vehicle's when the bus journey time excluded the dwell time at bus stops. However, the gap between the trajectory of the probe vehicle and the bus excluding dwell time increased as travel time increased. This is because buses also have the delay of deceleration and acceleration due to stops and average bus speeds are generally slower than the other vehicle's. The average speeds of the bus, the bus excluding dwell time, and the probe vehicle are 19.1, 31.7, and 34.4 kph respectively. Some average running speeds of a particular section of road are marked for comparison. For example, there is a long bridge between distance 1000 and 1500 m with less traffic disturbances, and the average speed over this section of the bus and the probe vehicle were 43 and 48 kph respectively.

This result could support the intuition described above that average bus speeds are generally slower than the other vehicle's.

- Time-Speed profile

It can be seen from the middle diagram of Figure 2.6 that the bus had more stop time with 0 speed including dwell time at bus stops and signal control delay at junctions. The maximum speeds of the bus were under 60 kph; however, the maximum speeds of the probe vehicle were greater than 70 kph.

- Time-Acceleration/Deceleration profile

It can be seen from the bottom diagram of Figure 2.6 that the accelerations of the bus were usually under 1 m/s^2. However, the maximum decelerations were between -1 and -2 m/s^2. The probe vehicle had similar accelerations and decelerations but also had some greater values. In addition, the maximum acceleration of the probe vehicle was greater than the bus; however, the maximum deceleration was smaller to the bus.

2.7 Characteristics of bus journey time

It is recognized that buses have different operating characteristics than other vehicles in traffic (Bertini and Tantiyanugulchai, 2004; Chakroborty and Kikuchi, 2004; McKnight et al., 2004). Other vehicles do not serve customers at bus stops, or decelerate and accelerate around bus stops. Their operating abilities are different due to vehicle characteristics and power performance. The details of advantages and disadvantages of bus journey time compared with other vehicles were described in section 2.6.1. Moreover, bus operations are often affected by schedule adherence and may be impacted by individual driving behaviour (Bertini and Tantiyanugulchai, 2004). Followings are the major characteristics of bus journey time compared with other vehicles, which includes dwell time, deceleration and acceleration due to bus-stops, and the effect of bus priority.

2.7.1 Dwell time

Because of their mission of passenger service, buses take additional time at bus stops compared with other vehicles. Dwell time ranged up to 26% of total bus journey time (Levinson, 1983) and accounted for about half of the journey time between adjacent stops (Lobo, 1997). Consequently, dwell time plays an important role in bus journey time.

2.7.1.1 Dwell time definition

There were several definitions with respect to dwell time. A broad definition, which contains five components, was considered by Levine and Torng (1994). They are the time waiting for boarding passengers to get on bus, the time for boarding, payment, alighting passengers, the time for dealing with equipment, the time waiting at bus-stop for adapting to the timetable, and the time waiting to rejoin traffic. This definition takes in account the entire period when the bus is stopped as well as the waiting time for re-entering the traffic stream. However, most studies accept the definition that dwell time is the total time when a bus stops to serve boarding and alighting passengers plus the time required to open and close doors (TRB, 2000; Rajbhandari et al., 2003; TCRP, 2003a; Dueker et al., 2004; Zhao and Li, 2005). In addition, Guenthner and Hamat (1988) provided a simpler definition that dwell time is the time a bus waits for passengers to alight and board at a bus stop. The popular definition, which was used for most studies, is used in this research.

A bus trajectory for serving a bus stop was shown on Figure 2.2. Point A indicates where the bus starts decelerating and point B where it comes to a complete stop at a bus stop. Point C shows where the bus starts departing and accelerating and point D indicates where the bus returns to original cruise speed. The period between point B and C ($\overline{BC}$ or $\overline{bc}$ or $\overline{t_2 t_3}$) is defined as dwell time which is described above. It can be seen from Figure 2.2 that the total delay for a bus serving a bus-stop includes deceleration delay, dwell time, and acceleration delay, when compared with a vehicle which does not stop at a bus-stop. Detailed bus activities along a bus stop are illustrated on the left hand of Figure 2.7. It can be seen that the broad definition includes bus activities from 3 to 9. However, the definition that is used in this study only includes bus activities from 3 to 8. Note that, some of abnormal doors' opening and closing activities were observed in the field, i.e. the door was opened before the bus stopped, or was closed after it moved. Nonetheless, the bus stop times at bus stops are used as dwell time for this study.

2.7.1.2 Factors affecting dwell time

There are three main factors and their interactions may contribute to the time taken at bus stops as follows:

- Types of buses
- Passengers and bus drivers
- Bus-stop layout and related road facilities
- Interactions between each other or all of them

For example, bus type (low-floor, the number of doors, capacity, etc.), the number of boarding and alighting passengers, special needs passengers (wheelchairs, pushchairs, and bicyclists), mobility on the bus (number of passengers already on board, standees), number of served bus-stops, spacing of bus-stops, fare collection methods, fare types, fare structures, multi-coin fare, interaction between passengers, and interaction between passengers and bus may all influence bus dwell time (Kraft and Deutschman, 1977; Marshall et al., 1990)

2.7.1.3 Results of previous studies

The review of dwell time has been illustrated by Dueker et al.(2004). Some of the results are described here. Levinson (1983) reported that dwell time of each stop is equal to 5 seconds plus 2.75 seconds per boarding or alighting passenger. Guenthner and Sinha (1983) obtained 10~20 seconds for each stop plus 3~5 seconds for each passenger boarding or alighting. The HCM (TRB, 2000) identified that buses take 2~5 seconds generally for door opening and closing and the typical alighting and boarding time per passenger for a conventional bus with one or two available doors is 1.0~2.0 and 1.2~3.0 s respectively. Alternatively, the following values can be adopted if there is no additional information: 60 s for a central business district, main station, key transfer point, or park-and-ride centre; 30 s for primary stops; and 15 s for usual stops. Bertini and EI-Geneidy (2004) found 0.85 seconds for each alighting passenger and 3.6 seconds for each boarding passenger. Dueker et al.(2004) indicated that the average dwell time of each stop is 12.29 seconds with lift operation and 11.84 seconds without it. Furthermore, without lift operation, a base dwell time of 5.14 for opening and closing doors was obtained and each boarding and alighting passenger added 3.48 and 1.70 seconds respectively. On the other hand, with lift operation, the base dwell time for each boarding and alighting passenger was changed to 68.86, 10.21, and 0.51 respectively. In addition to above studies, York (1993) calculated such parameters for different bus types, namely double-deck and single-deck low floor double-deck buses, respectively in a London study. For double-deck buses, the time taken to open and/or close the door(s) and check the traffic was 5.42s, 1.48s for each alighting passenger, and 9.15s for each boarding passenger. For single-deck low floor buses, the time taken to open and/or close the door(s) and check the traffic was 3.55s, 1.99s for each alighting passenger, and 9.18s for each boarding passenger. Shrestha (2002) used similar approach as York, which based on the field data collected in Southampton, UK. The results showed that the time taken for opening and/or closing the door and check the traffic, alighting time per passenger, and boarding time per passenger were 6.85, 1.69, and 9.00s; 3.30, 1.96, and 9.04s for double-deck buses and single-deck low floor buses respectively.

2.7.2 Delay of acceleration/deceleration due to bus-stops

Delay of acceleration/deceleration due to bus-stops may be defined as the total time when a bus is required to serve a bus-stop minus the time if the bus does not stop and minus the dwell time. It may be much clearer to check the distance-time diagram of a bus, which was shown in Figure 2.2. It can be seen that this delay is the sum of the delay of deceleration $\overline{ab}$ and the delay of acceleration $\overline{cd}$. That is, when comparing a non-stop bus, the total delay for serving a bus-stop is $\overline{ad}$, while $\overline{bc}$ is the dwell time which is described in section 2.7.1.1, hence the reminder ($\overline{ad} - \overline{bc}$ or $\overline{ab} + \overline{cd}$) is such delay.

Since it is very difficult to collect successive locations of vehicles without an AVL system such as GPS, only little research has addressed this issue. More often than not, even with such data, it is very difficult to identify where and when the deceleration begins and the acceleration ends in a real situation as described in section 2.2.2. Levinson (1983) indicated that the total acceleration and deceleration time per stop ranged from 11 to 23 seconds and followed Equation 2.8. Bertini and EI-Geneidy (2004) found 26 seconds for each stop.

$$T = 23.4 - 1.53X \quad (2.8)$$

where,

T= the total acceleration and deceleration time per stop; and

X= the number of stops per mile.

2.7.3 Bus priority

It is accepted that buses will remain the main public transport mode for most local journeys. However, buses cannot take an alternative route to get through congestion;

they need assistance from priority measures to break traffic. The aim of authorities is to improve the service of buses not only to keep the existing customers but also to attract car users. For example, a case study for London Bus Initiative phase one showed that over a three-year period (2000-2003), the 27 key routes across London increased their annual number of customers from 163 million to around 200 million, i.e. nearly 22 percent. This was achieved through the installation of more than 1,100 bus priority schemes, while about 200 new or extended bus lanes were implemented, more than 1,400 bus stops were improved, and over 370 traffic signal schemes along with over 320 selective vehicle detection units were delivered (DfT, 2003b). Bus priority schemes can succeed by improving the buses, the infrastructure, priority measures, and real time information by the development of partnerships between local authorities and bus operators (Department of Environment, Transport & the Regions [DETR], 2001). These can significantly improve journey time and reliability and make buses a feasible alternative for the public.

The capacity of roads has not increased in proportion to the growth rate of cars. This causes serious traffic congestion and affects the buses ability to deliver reliable services. It is necessary to consider all modes' requirements for a limited road space and a comprehensive strategy can be established to gain maximum benefit to all. Priority measures may comprise a combination of physical and non-physical facilities.

Physical facilities redistribute the road space to give exclusive right or part of exclusivity to buses. This includes with/contra flow bus lanes, high occupancy vehicle lanes, bus stop improvements, rising bollards, and guided bus way. For instance, a study of DETR (2001) indicated that a fully enforced bus lane could reduce travel time by 7 to 9 minutes along a 10 kilometres highly congested bus route.

- Bus lane

A highway lane primarily for buses, either 24 hours or during specific periods, but sometimes also used by specific vehicles such as ambulances, bicycles and taxis as allowed by authority. On with-flow bus lanes, the buses travel in the same direction as traffic in adjacent lanes. By contrast, on contra-flow bus lanes, the buses travel in the opposite direction to traffic in adjacent lanes.

- High occupancy vehicle lanes

An exclusive traffic lane or facility limited to carrying high occupancy vehicles (e.g. three or more people in one vehicle) and emergency vehicles.

- Bus stop improvements

Including improvements on bus stop consolidations, additional bus berths, bus islands etc.

- Bus boarders

A footway extends into the carriageway around the bus stop. This enables buses to easily access the kerb and alight/board passengers.

- Rising bollards

A type of bus gate prevents access for other vehicles to bus only lanes.

- Guided bus way

A way designed for buses so that buses travelling on their own or steering by guide-wheels on rails or tracks.

Non-physical facilities are primarily using various methods to detect buses and activate traffic lights to give priority for the bus to pass at junctions. Bus Information and Priority System (BIPS), Microprocessor Optimized Vehicle Actuation (MOVA), Bus SCOOT, Priority and InforMatics in Public Transport (SPRINT) and Signal Progression Optimization Technology (SPOT) are usually used for signal priority methods (Department of Transportation, 2002). For example, a DETR Traffic Advisory Leaflet (2001) summarized that Bus SCOOT can reduce bus travel time by 2 to 4 minutes along a 10 kilometres bus route. The variability of travel time improved by about 16 percent and time saving was around 1 to 10 seconds per junction (an average of 4 seconds). Travel time variability was improved by 0 to 20 percent (with an average of 12 percent). The following facilities are bus signal priority and other measures (TCRP, 2000):

- Bus gates

Entry points allow only buses to access bus lanes. The purpose of such control is to ensure the implementation of bus priority and the limitation of other vehicles. Bus gates can be traffic signals actuated by the buses, physical barriers passed only by buses such as rising bollards, or merely signs such as ‘ No Entry Except Local Buses’.

- Bus pre- signals

Traffic signals are situated at the end of bus lanes that permit buses to enter the bus advance area in front of other vehicles.

- Park and ride

Private cars are parked in a particular car park outside the city centre and a frequent bus service operating between car park and the city centre, is provided for drivers.

The focus of bus priorities can consider the main elements of bus journey time described in Equation 2.1. The strategies constitute three parts (Fernández, 1999): 1) link priority: to reduce the journey time of bus movement by segregating buses from the traffic stream, 2) junction priority: to decrease the delays at signalised junctions by adjusting signal settings, 3) stop priority: to lessen dwell time delays by consolidating bus stops or improving bus stops' layout. The most successful approaches are those which can meet the local condition of a particular route or corridor DfT (2003b). They are part of an entire road management scheme that coincides with other traffic control method, road maintenance, and works. Only when these strategies are very well coordinated, delays to all traffic, including buses, can be reduced considerably.

Overall, there is no best measure with best performance for any circumstance. The most appropriate measure in any location will depends upon the local situation in that area and utilizes either one or a combination of the above priority measures (DfT, 2003b). Factors such as traffic conditions, roadway layout, the number and frequency of bus services, residential or commercial areas along the road, etc, must be taken into account in order to establish the most suitable bus priority measure for any particular location.

2.8 Bus journey time estimation

2.8.1 Points of view of bus journey time

Measuring the performance of a transit system is crucial for efficient and effective management. By using field data from available technologies, it is possible to obtain information from the operation of the transit system and compare performance over four different levels, namely system, route, segment, and point level, from day to day and from route to route (Bertini and El-Geneidy, 2003).

There are two diverse views of bus journey time. The Transit Capacity and Quality of Service Manual [TCQSM] (TCRP, 2003b) focuses on measures that reflect transit passengers' point-of-view. Conversely, the Highway Capacity Manual (TRB, 2000) reflects conditions experienced by vehicles using roads. This research concentrates mainly on the latter aspect.

From the view of transit passengers, bus journey time is part of passenger's trip time, i.e. in-vehicle time. The comprehensive trip time is from the origin to the destination, which includes walking time from the origin to a bus stop, waiting time at bus stop, bus journey time, transferring time, and walking time from a bus stop to destination. The TCQSM indicated that bus journey time is an indicator of transit performance covering the aspects of vehicle and passengers and such value can be in relation to other competing modes such as other vehicles (section 2.6), or in comparison with criteria values. Time-related measures are useful for evaluating the service quality of particular trips, which can be used for illustrating the effects of traffic congestion TCRP (2003b).

Highway Capacity Manual indicated that buses in bus lanes in urban areas are influenced by bus stop spacing, dwell times, traffic signals, turning traffic, road geometry, bus lane features, skip-stop operations, and interference caused by other vehicles sharing the lane. In addition to the above factors, buses in mixed traffic situations should be considered more in relation to the interference caused by other traffic. Particularly at junctions, other traffic may block buses from reaching bus stops or halt a bus behind a queue. Moreover, additional delay may occur when buses leave stops and return to the traffic stream. Sometimes buses have to wait for a gap in traffic to merge back into the road if there is no priority facility offered.

2.8.2 Key factors of bus journey time

Bus journey time can be measured by itself directly, e.g. finding a relationship between buses and their affecting factors, or in relation to other competing modes, e.g. finding a relationship between buses and other vehicles or a ratio of these two as described in section 2.6, or both. Several studies have used various factors to estimate bus journey time. Table 2.5 summarizes these factors which have been used in previous studies. The most used factors are the route (or segment) length, the number of bus-stops along the route, the number of bus-stops made by buses, the number of boarding passengers, and other vehicle journey time. Consequently, these key factors play a crucial role in bus journey time estimation and are therefore taken into account in this research.

2.8.3 Methodologies and results of previous studies

Bus journey times in peak hours were reported as 6.0 min/mi in cities, 4.2 min/mi in suburban areas, and 11.5 min/mi in central business districts, while general vehicles usually travelled 1.4 to 1.6 times faster than buses which had about 48~75% in moving, 9~26% at bus-stops, and 12~26 in traffic delays, based on surveys using statistic and regression analysis (Levinson, 1983).

Seneviratne (1988) simulated a fixed route bus journey time by the Monte Carlo procedure, in which variables were assumed to have a specific probability distribution from previous studies or from theoretical function. It indicated that such a simulation model could be used to examine the effects of traffic management schemes, stops consolidation and passenger demand on journey time, and therefore the required buses for service and performances.

McKnigtht and Paaswell (1997) conducted a regression analysis using survey data and concluded that bus travel time rate (minutes per mile) increases with 0.57 times general vehicle journey time rate, each boarding passenger adds 4~5s, and each halt at a bus stop adds about 23s to journey time per mile.

Abdelfattah and Khan (1998) estimated bus journey time by micro-simulation, which calculated from bus speed without delay (with average 40 kph) plus expected delays estimated by developed regression models which accommodated particular scenarios. It concluded that such delay models could be used to improve the reliability of bus arrival time along bus routes.

TCRP (2000) reported bus journey time rate impacted by various bus spacing, dwell time, and operating environment, which was based on empirical evidence.

McKnight et al.(2004) used the regression method based on field data and indicated that bus travel time rate (minutes per mile) increases with 0.73 times general vehicle journey time rate, each boarding passenger adds 3~4s, and each halt at a bus stop adds about 18s to journey time per mile.

Bertini and EI-Geneidy (2004) conducted statistics, regression, and sensitivity analysis using the data from AVL based on GPS and APC and concluded that average non-stop speed of buses was 30 kph, the time lost due to acceleration and deceleration at bus-stops was 26 s per stop, and the dwell time of bus-stop was increased by 0.85s for each alighting passenger and by 3.6s for each boarding passenger.

Chakroborty and Kikuchi (2004) developed a simple predictive function using buses as probe vehicles and their result showed that the average journey time of other vehicles is equal to free-flow journey time plus that bus journey time excluding dwell time multiplied by a 0.14 or 0.18 when less or more frequently congested roads were used.

Overall, the methodologies used in estimating bus journey time in the literature were either statistics combined with regression analysis based on data from field survey of test or probe vehicles, Automatic Vehicle Location (AVL) system with/without Automatic Passenger Counting (APC) on board, or a simulation approach based on known (from previous studies) or theoretical parameters. Then, interpretations of coefficients of model or sensitivity analysis were made to explain the relationship between bus journey time and each factor. With increasingly available traffic management systems on road networks such as SCOOT, ANPR, it is possible to access such traffic data combined with bus data (AVL) and passenger data (APC) in order to have better estimation of bus journey times which are influenced by traffic variations. However, no such study has been found in the available literature. Therefore, this research proposes to fill this gap. In addition to the available data sources, the major model which was developed by previous studies was either by regression or by a simulation approach. According to the comparison of available methodologies, which was discussed in section 2.4.4 and the lack of comparative performance from regression and simulation approaches, this study aims to develop models with these two methodologies in order to ascertain a better method of bus journey time estimation.

2.9 Summary

This chapter has described the related issues of bus journey time involved in this research in detail. The definition of journey time and general elements of it were presented. The concept of link journey time and the needs for such time, i.e. repeat reduction, easy prediction, and time reduction, were discussed. A comparison of available journey time data collection methods was made, encompassing either fixed detectors or moving detectors and the result showed that a probe vehicle equipped with GPS is superior in terms of cost-effectiveness, detail spatial-temporal data, well delay representation, and is already proven by former studies. Several methodologies which were often used for journey time estimation in the literature including empirical, analytical, and simulation approaches, were illustrated.

The trend of bus usage and bus user needs was presented in order to better understand the tendency to allocate the bus for public transport and enhance its service. A SWOT analysis is provided to identify the buses resources and capabilities in the competitive environment in which they operate. Bus journey time can be estimated either by itself or in relation to another mode. The characteristics of bus journey time, e.g. dwell time, delay of acceleration and deceleration due to bus-stops, and bus priorities, and in relation to other vehicles were then illustrated.

Bus journey time can be viewed from the passenger perspective or as reflecting conditions which vehicles experienced on the road. This research focuses on the later aspect. The key factors, which affect bus journey time in the previous studies were found, which includes the route (or segment) length, the number of bus-stops along a route, the number of bus-stops made by buses, the number of boarding passengers, and other vehicle journey time. These key factors play a crucial role in bus journey time estimation and hence will be taken into account in the following research.

The methodologies used for bus journey time estimation in the available literature were either statistics combined with regression analysis based on the data from field survey, or simulation approach based on known or theoretical parameters. With increasingly available traffic management systems on road networks, it is possible to access such traffic data combined with bus and passenger data in order to have better estimation of bus journey times which are influenced by traffic variations. However, no such study has been found in the available literature. Therefore, this research proposes to fill this gap. In addition, according to the advantages and disadvantages of available methodologies and the lack of comparative performance from both regression and simulation approaches, this study aims to develop models with these two methodologies in order to ascertain a better method of bus journey time estimation.

The data accuracy for journey time data collection for this study is presented in the next chapter. The details of the data collection process involved in this research are described in Chapter 4. The description of bus journey time components and the development of the models follow in the subsequent chapters.

Chapter 3: Accuracy of GPS for Journey Time Survey

3.1 Introduction

Many available technologies are available to collect travel time data as described in section 2.3. These include inductive loop detectors which give point measurement of speed; probe vehicles equipped with GPS which can give detailed speed profiles along a route; automatic number plate recognition (ANPR)/automatic vehicle identification (AVI) which can be used to give journey times over a section of road. There is no single best technology which covers all traffic situations. However, probe vehicles with GPS are cost-effective, give detailed spatial-temporal data with delay representation, and have already been proven to be effective (Quiroga and Bullock, 1998a; Turner, 1998; D'Este et al., 1999; PSRC, 2000; and Li, 2004).

Previous studies of the application of GPS for travel time research have either used substantial GPS data from traffic control or fleet management centres, or used stand-alone GPS recording equipment in probe vehicles. In order to save storage memory in large or long duration surveys or keep the equipment portable, logged files have usually contained only date, time, and coordinate data with speed and acceleration calculated during a post analysis phase. The values obtained though this calculated approach are different from those of GPS data measured directly (Zito and Taylor, 1995; Belliss, 2004). Furthermore, some useful GPS output messages were discarded by this restricted data process. Several studies examining GPS accuracy for travel time surveys have used data collected by instrumented vehicle as a basis of comparison (Zito and Taylor, 1994; 1995; Belliss, 2004). In this approach, distance was obtained using a sensor to read the number of wheel rotations and speed was captured by connecting with the reading of the speedometer. However, as wheel circumference is affected by various factors, such as temperature and tyre pressure, this method required frequent calibration to maintain accuracy. Thus, data collected by an instrumented vehicle might be less accurate than GPS in some situations (Ogle et al., 2002). Some studies were made when the Selective Availability (SA), i.e. the artificial error of GPS, was switched on (before 1 May 2000), which gave poor results and are not considered applicable here.

The aim of this chapter is to identify the usefulness and reliability of particular GPS equipment which is used for the collection of bus journey time data in Chapter 4, by comparing its result with more elaborate GPS equipment. This chapter starts with the description of GPS error and accuracy. Then, the requirements of GPS data accuracy for journey time survey are presented. This is followed by conducting a survey to understand the GPS data accuracy in field. Finally, a conclusion sums up this chapter.

3.2 GPS error and accuracy

Nearly all the attention towards GPS accuracy is on position and velocity but not on time. This is because GPS can provide an extremely precise time reference, normally 1μs synchronized GPS time. Thus, such time accuracy is far enough for journey time study hence it is not discussed here. Nevertheless, the accuracy of bus journey time and dwell time are discussed later in sections 9.2.1 and 9.2.2.

There are several factors or disturbances which may affect GPS accuracy. Five types of possible sources of error are illustrated in Figure 3.1 (Wormley, 2004; Garmin, 2005).

(i) GPS satellites: including ephemeris and satellite clock errors, the number of satellites visible, and their geometric distribution.

(ii) Transmission process: these are the main sources of GPS error and include ionosphere, troposphere, and multi-path errors.

(iii)GPS receiver: including clock errors and decoding errors.

(iv)Data recorder: including allowable recording data types and data decimals.

(v) Coordinate transformation from different systems.

The errors shown in Table 3.1 were estimated by Ogle et al. (2002). Although there are prior possible errors, nowadays, the prevailing portable GPS receivers can track up to 12 satellites and have 1 second update rate. The position accuracy which is reported by the manufacturers is 15 meters RMS (Root Mean Square) and velocity accuracy is 0.01 to 0.3 m/s RMS. The official state of GPS accuracy is within 20 m 2DRAMS (with 95% confidence interval) or 10 m RMS (with 50% confidence interval) (Ogle et al, 2002). The accuracy can reach sub-meter when a high-end receiver and/or differential correction is used. In reality, disabled situations occur with an unreliable signal from an insufficient number of satellites, usually inside buildings or tunnels, under over-bridges, around tall buildings in central urban areas, or among trees with dense foliage.

3.3 Requirement of bus journey time accuracy

Bus journey time information is used for transit passengers as traveller information and for transit operators to improve system performance and service. For traveller information, most of the bus real-time information or timetables shown on terminals, bus-stops, and websites are 1 minute accuracy, e.g. the next bus is due at 10.35 (hh:mm) or is 2 minutes late. Thus, such accuracy should be moderate for public users. The positioning accuracy which is derived from the above accuracy may be varied when transforming the time scale to distance scale for different speeds. For instance, if the average bus speed is 40kph, 1 minute accuracy of travelling distance may have 667m tolerance; while in a congested urban area, if the average speed is 20kph, the tolerance may decrease to 167m. Note that additional delays such as bus-stop service and signalized control may result in even shorter tolerance. However, for operational control, more accurate time and position of bus should be required to identify critical sectors for improvement. Therefore, the requirement of bus journey time accuracy may be varied depending upon the intended purpose and the time-space need.

3.4 Journey time data accuracy

3.4.1 Description of survey

3.4.1.1 Survey site

A test route was selected with a range of physical conditions including open motorways, city centre roads with surrounding tall buildings, tree canopy, and over-bridges of shopping mall. The route also contained a variety of traffic conditions to provide a range of operating speed. The test route contained Southampton city centre and part of the M3 motorway in the UK. The various test areas and traffic conditions against travel time are shown in Table 3.2.

3.4.1.2 Types of survey

- Static test

GPS positions were logged for a period of time and the resulting points were dispersed over an area owing to various errors. A survey was undertaken for static test which employed the equipment described in Table 3.3 (set 1). A passive station of the national GPS network of the Ordnance Survey was selected as a known point in Southampton to examine static positional accuracy. The receiver was put on a marked point of the passive station and the GPS data was recorded every second during a one-hour survey.

- Moving test

Three sets of GPS equipment were utilized to log data in a field test. These included a GPS receiver (Garmin 35 PC) and hi-end product (Racelogic VBoxIII). The specifications of each set are shown in Table 3.3. GPS data from these three sets were logged continuously during the field survey. Within Set 1, the position resolution of the Garmin 35 PC receiver was 0.2 m. However, the ability to log position resolution from the prevailing portable data logger was only 2 m and the data accuracy might probably be degraded due to the data logger's limitation. Set 2 used an on-board PC for data logging and the positional accuracy could be maintained as 0.2 m from the receiver, while the positional accuracy of set 3 was 0.01 m.

3.4.1.3 Types of data collection

There were two groups of data collection, namely raw GPS data and signal quality indicators, which were illustrated in Table 3.3 row 3 and 4 for each GPS set respectively.

3.4.2 Data processing

Position, distance, speed, and acceleration data are manipulated to check journey time accuracy in this study. Speed and acceleration are reported usefully to check delays in travel time studies and can be used as indicators to evaluate operational performance by Zito et al. (1995). The following subsections introduce the data processing approaches to access these parameters. The method which is used for the database of each parameter by different GPS sets is shown in Table 3.4.

3.4.2.1 Position

The data obtained from the outputs of survey equipment were all given by the WGS84 (World Geodetic System 1984) coordinate system which showed as latitude and longitude. They were then converted into OSGB36 (Ordnance Survey Great Britain 1936) which were eastings and northings, to fit the digital map on ArcView (GIS software). Grid InQuest software which has a transformation accuracy in the horizontal for 0.1 m RMS and in the vertical for 0.02 m RMS (Ordnance Survey [OS] , 2000b) was used. The detailed processes, which convert the raw data of GPS into correct coordinate system and display on digital map using ArcView are illustrated in Appendix A.

3.4.2.2 Distance

Several calculation methods have been described in the literature to calculate the distance between two GPS points if the distance data of output from GPS equipment were unavailable (sets 1 and 2). Some are quite simple while others are complicated. The general concept can be explained by reference to Figure 3.2, in which Sites A and B are the successive GPS points. The solid line represents the actual trajectory of a vehicle and the dotted line is the straight line. The dotted line does not take into account the additional distance travelled because of the physical curvature of the earth. Intuitively, the dotted line might be similar to the solid line if point A and B are very close. In order to proceed with the calculation of vehicle speed and acceleration, it is essential to select a proper algorithm of which the accuracy is acceptable and easy to calculate.

Five calculation methods were used for comparison. The first simplest calculation is illustrated as Equation 3.1. This method calculates the straight line between two points and is only suitable for Easting and Northing coordinates. Note that the original coordinates obtained from GPS were latitude and longitude of WGS84. The second Equation 3.2 is from Zito (1994). Equation 3.3 and 3.4 are from Meridian World Data (2004) and the last and the most complicated Equation 3.5 is from Adamchuk (2000). Note that Equation 3.2~ 3.5 uses decimal degree data of latitude, longitude, and ellipsoid height coordinates.

- Simple Eastings and Northings Distance (m) $=\sqrt{(E_2-E_1)^2+(N_2-N_1)^2}$ (3.1)
- Zito Decimal Degree Distance (m)$=111296*\sqrt{\cos\frac{(y_2+y_1)}{2}(x_2-x_1)^2+(y_2-y_1)^2}$

(3.2)

- Simple Decimal Degree Distance (m)

$=1609.35*\sqrt{(69.1*(y_2-y_1))^2+(53.0*(x_2-x_1))^2}$ (3.3)

- Improved Decimal Degree Distance (m)

$=1609.35*\sqrt{(69.1*(y_2-y_1))^2+(69.1*(x_2-x_1)*\cos(y_1/57.3))^2}$ (3.4)

- Complicated Decimal Degree Distance (m)

$$=\sqrt{(F_{longitude}*(x_2-x_1))^2+(F_{latitude}*(y_2-y_1))^2} \tag{3.5}$$

$$F_{longitude}\text{(m/degree)}=\left[\frac{a^2}{\sqrt{a^2*\cos\left(\frac{\frac{y_1+y_2}{2}*\pi}{180}\right)^2+b^2*\sin\left(\frac{\frac{y_1+y_2}{2}*\pi}{180}\right)^2}}+c\right]*\cos\left(\frac{\frac{y_1+y_2}{2}*\pi}{180}\right)*\frac{\pi}{180}$$

$$F_{latitude}\text{(m/ degree)}=\left\{\frac{a^2*b^2}{\left[a^2*\cos\left(\frac{\frac{y_1+y_2}{2}*\pi}{180}\right)^2+b^2*\sin\left(\frac{\frac{y_1+y_2}{2}*\pi}{180}\right)^2\right]^{\frac{2}{3}}}+c\right\}*\frac{\pi}{180}$$

where:

E_1, N_1 are the Easting and Northing of the first point respectively in meters;

E_2, N_2 are the Easting and Northing of the second point respectively in meters;

x_1, y_1 are the longitude and latitude of the first point respectively in decimal degrees;

x_2, y_2 are the longitude and latitude of the second point respectively in decimal degrees;

a is the semi-major axis of biaxial ellipsoids used in the UK, which is 6,378,137.000 m (OS, 2000a);

b is the semi-minor axis of biaxial ellipsoids used in the UK, which is 6,356,752.3141 m (OS, 2000a); and

c is the height above ellipsoid.

The data used for comparing the distance accuracy of the above equations was obtained from the inbound direction of Bitterne Road (A3024) in Southampton, UK, approximately 2.5 km long with 1 second GPS data logging frequency. Figure 3.3 illustrates the total distance calculated by comparative equations. It can be seen that the outcome of Equations 3.1, 3.4, and 3.5 are almost on the same line and Equations 3.2 and 3.3 have longer results than others. In addition, the discrepancy among these lines increases when the time increase, i.e. the gap of accumulated distance may increase due to various calculation methods. The results of the above equations are shown in Table 3.5 respectively. It can be seen from the first row that the accumulated distance of Equations 3.1, 3.4, and 3.5 are similar, but Equations 3.2 and 3.3 are not. Theoretically, the Equation 3.5 may have the best accuracy of distance calculation. If the result of Equation 3.5 is taken as the quasi-real travelling distance, the error indicators of Mean Absolute Deviation (MAD), Mean Absolute Percent Error (MAPE), and Root Mean Squared Error (RMSE) are calculated as the second, third, and fourth row respectively. Readers could refer to Toppen and Wunderlich (2003) for details of MAD, MAPE, and RMSE. From Table 3.5, it can be seen that the distance accumulation of Equation 3.1 has the closest value to Equation 3.5 and it has the smallest value of MAD and MAPE and the second smallest value of RMSE. It can be concluded that Equation 3.1 may have the advantage of acceptable accuracy and a straightforward form of calculation. This equation is then utilized for speed and acceleration calculation thereafter.

3.4.2.3 Speed

Two methods are available to obtain speed from GPS data. One is to divide the distance travelled between successive GPS points by the time taken to traverse the distance as Equation 3.6. This approach was used for GPS set 1. The other approach is to access speed data output directly from the GPS receiver which uses the Doppler effect technique. Both sets 2 and 3 were used in this way. Previous studies indicated that speed measured directly by the latter approach is more accurate than those of post-calculation (Zito et al., 1995; Belliss, 2004).

$$Speed = \frac{P_{E2,N2} - P_{E1,N1}}{t_2 - t_1} \tag{3.6}$$

where,

$P_{E1,N1}$, $P_{E2,N2}$ are the first and second position of easting and nothing in metres; and

t_1, t_2 are the time of the first and second position.

3.4.2.4 Acceleration

Normally, acceleration cannot be obtained from GPS receivers directly and must be calculated by dividing the shift in speed by the shift in time as Equation 3.7. This approach was used for both sets 1 and 2. However, acceleration data are available from the direct output of set 3.

$$Acceleration = \frac{v_2 - v_1}{t_2 - t_1} \quad (3.7)$$

where,

v_1, t_1 are the speed and time of the first position respectively; and

v_2, t_2 are the speed and time of the second position respectively.

3.4.3 Results

3.4.3.1 Position

The discrepancies of the GPS points for the static test are shown in Figure 3.4. It can be seen that the middle of the cross shows the position of the passive station in which the GPS receiver was put. There were a large number of discrete positions which was estimated from the GPS. This was due to the limited resolution of the data logger which was described in section 3.4.1.2. GPS points were concentrated mostly on the station position with significant high bar (bar counts 839, 23%) at zero discrepancy of the Easting and Northing coordinates. The easting discrepancy ranges from -2.1 to 3.76 m with a mean of 0.22 m and a standard deviation (SD) of 0.92 m. The northing discrepancy ranges from -5.41 to 5.75 m with a mean of 0.24 m and a SD of 1.86 m. The elevation discrepancy of mean sea level (MSL) ranges from -15.36 to 8.64 m with a mean of -3.41 m and SD of 4.50 m, i.e. the horizontal accuracy of the GPS position is significantly better than the vertical. The horizontal distance to the station position ranges from 0.33 to 6.86 m and the mean length is 1.73 m with a SD of 1.19 m.

The area within which the GPS positions were scattered is called the confidence region (NovAtel, 2003). It can be used to illustrate the specific accuracy with a radius. Several approaches have been used to measure two-dimensional accuracy as described in Table 3.6. They are CEP (Circular Error Probability), DRMS (Distance Root Mean Squared), and 2DRMS with increased probability of containing GPS points of 50%, 65%, and 90% respectively. The horizontal accuracy of the confidence region has been calculated as the last column of Table 3.6. These values indicate that 50% of the horizontal positions should be within 1.61 m radius of a circle (CEP); 65% of horizontal positions should be within 2.07 m radius of a circle (DRMS); and 95% of horizontal positions should be within 4.15 radius of a circle (2DRMS) in the survey.

In brief, it can be concluded that the position accuracy of GPS (set 1), which is used for bus journey time data collection in this study, is within 4.15 m with 95% confidence.

Such accuracy of static position can be used to understand certainty of road facility location, which is discussed later in section 9.2.4.

3.4.3.2 Distance

In addition to the distance estimated from the three GPS sets which was described in Table 3.3, the odometer reading, which was conducted manually in the probe vehicle for several checkpoints during the survey and the ruler function in ArcView (measure distance on digital map with embedded ruler tool) were included for comparison of the total distance travelled. The distance data of set 3 were to be output directly from the equipment. However, the results were quite unreasonable and could not be used for comparison. Therefore, the calculation approach used on set 1 and 2 was adopted on set 3 as well. As can be seen in Figure 3.5, the accumulated distance by set 1, 2, odometer reading, and ruler are clustered (52,344, 53,965, 52,785, and 51,799 m respectively). However, the distance travelled is apparently overestimated by set 3 (73,512 m). Generally, distance measured on the digital map is the lowest because of accumulation of consecutive straight lines between two points, without consideration of lane changing behaviour and drift in driving. Thus, the distance measured by this approach should be less than the real driving distance. GPS errors are included in the calculated distance by using the method which was described as Equation 3.1. As a result, the length calculated in this approach may be greater than the real driving distance. Also, the distance estimated from odometer readings may have errors which result from the instrumented vehicle as explained in section 3.1. The accuracy of the distance is essential as a reference of road geometry to compare the trajectories of vehicles, which was described in section 2.6.2.2. Note that, no filter or corrections were used in the above calculations.

The reason that set 3 overestimated the total travelled distance can be justified by consideration of Figure 3.6. The three figures show the same section of route which were in the city centre with stop-and-start traffic condition. It can be seen from Figure 3.6 that the right figure of set 3 had more significant deviation of adjacent GPS points than the other figures. As a result, the accumulated distance from consecutive points caused overestimation. The greater frequency of set 3 points (100 points per second) may contribute to this error. Therefore, in very congested traffic conditions, a lower updating rate of GPS may be suggested.

In brief, it can be concluded that the distance accuracy of GPS (set 1) and the ruler tool in ArcView, which is used in this study using Equation 3.1 for calculation, is moderate.

3.4.3.3 Speed

Speed comparisons of the three GPS sets are shown in Figure 3.7. It can be seen that they generally fit well. The period during 00:38:00-00:47:00 had the most noise in the data, which was on city centre roads in a congested situation. Means and SD of speeds were similar (mean=25.43, 25.08, 25.07 mph, SD=24.36, 24.51, 24.50 of GPS sets 1, 2, and 3 respectively).

The stated speed accuracy of set 3 (0.1 km/hr), which has the best accuracy, was used as a basis of comparison. The speed discrepancy of sets 1 and 2 against set 3 are illustrated in Figure 3.8. It can be seen that most of the speed discrepancy of set 2 lies in the range of +/-2 mile/hr (the 5 percentile=-1.77 mile/hr and the 95 percentile=1.99 mile/hr) and only some outliers existed at under 30 mph, while the calculated speed discrepancies of set 1 scatter between +/-4 mile/hr (the 5 percentile=-2.61 mile/hr and the 95 percentile=3.86 mile/hr) with some outliers under 30 mph. The speed discrepancies do not increase when the vehicle speed rise. Such a result is the same as that found by Zito et al. (1995). The reason for the speed discrepancy under 30 mph with more noise for both sets is thought to be that this range of speeds occurred on an urban part of the route with more disturbances than on the motorway. This can be explained by the assistance of signal quality indicators in section 3.4.3.5. The GPS speed discrepancy against travel time is shown in Figure 3.9. With travel time referred to in Table 3.2, the specific area and traffic condition can be identified. The first period with the most noise was 00:30:00-00:35:00 and that was on urban roads with dense trees and some traffic constraints. During 00:40:00-00:47:00, there was a few noises within the smaller speed discrepancies in set 1. This was in the city centre roads in stop-and-start traffic conditions. The period of around 00:50:00-01:02:00 had the greatest noise within it and that was on city centre roads with tall buildings and shopping mall over-bridges around. The last period with greater noise was 01:15:00 until the end of the test. This was in the university's campus with tall buildings near by.

Overall, speed data from direct readings of the GPS receiver (set 2) are superior to that from post-calculations (set 1). The speed error would not increase when the vehicle speed increased. The error interval of +/-4 mile/hr (+/- 1.78 m/s) of set 1, which is used in the data collection for this study, is acceptable for the permitted relative error suggested by the Institute of Transportation Engineers (1994) as +/-2mph to +/- 4mph for traffic operation, trend analysis, and economic evaluation purpose. This result is used to estimate the probable accuracy of average acceleration/deceleration rate, which is discussed later in section 9.2.3.

3.4.3.4 Acceleration

The calculated accelerations of sets 1 and 2 compared with the calculated acceleration of set 3 are shown in Figure 3.10. It can be seen that most of the acceleration discrepancy of set 2 lies in the range of +/-2 mile/hr/s (the 5 percentile=-1.33 mile/hr/s and the 95 percentile=1.34 mile/hr/s) and only some outliers existed at under 30 mph, while the acceleration discrepancies of set 1 scatter between +/-4.5 mile/hr/s (the 5 percentile=-4.49 mile/hr/s and the 95 percentile=4.43 mile/hr/s) with some outliers under 30 mph. The acceleration discrepancies do not increase when the vehicle speed rise and the amplitudes of positive and negative acceleration are similar.

In brief, the acceleration error of set 1 which is used in data collection would not increase when the vehicle speed increase and is within an error interval of +/-4.5 mile/hr/s (+/- 2 m/s/s), which is moderate. Such result is used to calculate probable accuracy of average acceleration/deceleration rate, which is discussed later in section 9.2.3.

3.4.3.5 GPS performance indicators

GPS accuracy is sensitive to the available signals from GPS satellites. Some performance indicators are available in NMEA (National Marine Electronics Association) transmitted sentences from the output of the GPS receiver, e.g fix indication and number of satellites in use in $GPGGA; fix type and Positional Dilution Of Precision (PDOP), Horizontal Dilution Of Precision (HDOP), and Vertical Dilution Of Precision (VDOP) in $GPGSA; valid position status in $GPRMC; and Estimated Position Error (EPE), Estimated Horizontal Position Error (HPE), and Estimated Vertical Position Error (VPE) in $PGRME (Garmin, 1999). These indicators can be used to evaluate whether the particular GPS based output is acceptable.

Fixed indication shows the message of unavailable fix, non-differential fix, or differential fix. Theoretically, at least four satellites are necessary to calculate a 3D point on earth and a more accurate position can be achieved if more satellites are available. State-of-the-art GPS receivers can track up to 12 satellites and offer more reliable outputs. Fix type, demonstrates the message of unavailable fix, 2D or 3D fix. DOP is an indicator of satellite geometry; lower values of DOP generally indicates better position accuracy. Zito and Taylor (1995) indicated that the position will be uncertain when the PDOP value is greater than 5 and if the value is less than 3, it should be reliable. Moreover, Ogle et al. (2002) used this factor to filter collected data, that is, unreliable data was discarded when PDOP was greater or equal to 4. DOP consistes of three components, namely PDOP, HDOP, and VDOP, and is calculated as $PDOP^2 = HDOP^2 + VDOP^2$. Valid position status indicates whether or not the GPS point is valid. EPE in meters is the approximated error provided by the GPS calculation, and is also calculated as $EPE^2 = HPE^2 + VPE^2$. Note that the elevation factor has not been discussed in this study. Therefore, only HDOP and HPE have been used for analysis.

Some of the above indicators (number of satellites in use, HPE, HDOP) against travel time in the test are shown in Figure 3.11. It can be seen that the smaller the number of satellites in use, the greater the value of HDOP and HPE, and hence the worse signal condition. The status of the GPS signal was generally satisfactory during the survey. The mean of number of satellites in use, HDOP, and HPE are 7.99, 1.49, and 9.65 respectively and SD are 1.22, 0.58, and 3.60 in a total of 4,566 observations. Travel time corresponding to Table 3.2, GPS signal state of specific location or traffic condition can be obtained. Note that the signal strength is not a constant for a particular place; it might change over time.

An example of GPS performance indicators is demonstrated in Figure 3.12, which indicates the relationship between speed discrepancies and the number of satellites in use. It can be seen from the figure that the speed discrepancies increase when the available satellites decrease. When the number of satellites was 10, the speed discrepancies were between only +/-3 mile/hr, while the discrepancy span increased to -15~ +25 mile/hr when the number of useable satellites fell to 5. It seems to have two asymptotes (showed dotted line) approaching to near zero discrepancy. Similarly, it is possible to relate other GPS signal quality indicators with probable errors of point, distance, speed, and acceleration to attain estimating function.

3.5 Summary

This chapter has described in detail the accuracy of GPS data involved in this research. It began by outlining the problems and methodologies in previous studies and the aim of this chapter. Then, the possible GPS error sources and the prevailing accuracy of portable GPS receiver were presented. Then, the requirement of bus journey time accuracy in respect of the view of users was illustrated. This was followed by a field survey which was conducted to identify the usefulness and reliability of particular GPS equipment which was used for collection of bus journey time data in Chapter 4, by comparing its results such as position, distance, speed, and acceleration with more elaborate GPS equipments. The characteristic of required data and data collection process of this study is described in the next chapter.

Chapter 4: Data Collection

4.1 Introduction

In Chapter 2, methodologies for estimating bus journey time and methods for collecting journey time data were reviewed and described. It is clear that regression and stochastic simulation approaches are suitable for formulating bus journey time estimation models and probe vehicles equipped with GPS are superior for data collection. Furthermore, Chapter 3 illustrated the data accuracy of journey time in terms of possible errors during the field survey of GPS data. As a result, correct types of data are required to represent bus operation associated with traffic conditions and the road environment. This chapter aims to characterise the data required in the modelling process and describes the data collection in the field.

This chapter starts with the description of data requirements (section 4.2). Then, data collection planning (section 4.3), concerning where and how data has been collected, is illustrated. It will be followed by the main survey (section 4.4) carried out for data collection. After that, a process to examine possible errors after data collection (section 4.5) is explained. Finally, a summary is made for this chapter (section 4.6).

4.2 Data requirement

In order to formulate the proposed models, the following data are required to explore possible factors of bus journey time. They are thought to be related directly or indirectly to bus journey time and are feasible to collect from the selected sites in the field. They are grouped according to their characteristics.

4.2.1 Bus data

This is comprised of two sets of data, namely buses' data of themselves and operation results. The former data includes available bus services, bus timetables, bus routes, bus types, and so on. The latter data is related with bus journey time, which represents the profile in space and time of buses travelling along links or routes. This includes journey time, dwell time, deceleration and acceleration time.

4.2.2 General traffic data

This data identifies traffic conditions during a specific period at a particular sector of road. They are general vehicles' journey time, signal timing data of each link, traffic flows, speeds, occupancies, and so on. All vehicles may be included except motorcycles and pedal cycles.

4.2.3 Road geometry and facilities

This data is required to define the physical characteristics of road links and vehicle routes. It includes lengths, the locations of road facilities such as bus-stops, junctions, loop detectors, stop lines, the number of lanes, bus lane length, the number of signalised junctions, the number of turns along bus routes, the number of bus stops, and so on.

4.3 Data collection planning

The above section described the data required for formulating bus journey time models. Similarly, it is necessary to have the data for later stage of validation. Thus, the data collection plan was developed in order to gather data for both stages, namely model formulation and validation. The concept of the data manipulation of this study is illustrated as Figure 4.1. Part of data (e.g. routes 1, 2, and 3 in the figure) is used for model formulation and validation (Chapter 6 and section 7.2) and the remainder part of the data (e.g. route 4) is used for independent validation (section 7.3). A further discussion of data manipulation can be seen latter in section 9.5.1.3.

4.3.1 Survey site

It is essential to understand the data requirements described in section 4.2 before site selection. It is found that Southampton in the UK has a comprehensive bus network and several radial bus corridors, which are suitable for this research. In addition, most of the main junctions are coordinated with the SCOOT (Split Cycle Offset Optimisation Technique) system and considerable traffic parameters are accessible from the traffic control centre of the ROMANSE (ROad MANagement System for Europe) office. Furthermore, travel time information, which is provided for drivers along the main corridors using Automatic Number Plate Recognition (ANPR) technology, is also an invaluable data source for this study.

At the initial stage of planning, several factors for choosing feasible survey routes were taken into account:

- Main corridors
- Bus routes which also have travel time information for car drivers
- Junctions along the routes controlled by the SCOOT system
- Fixed routes and frequent bus services
- Considerable number of passengers alighting and boarding along bus routes
- At least one end of bus route to or from city centre.

In fact, no single route could match all the above requirements. Thus, routes were selected with as many of the above factors as possible. As a result, four bus routes were selected in Southampton. They were The Avenue (A33), Portswood Road, Bitterne Road (A3024), and Portsmouth Road (A3025), which are from the city centre to Chilworth Roundabout, Swaythling, Botley Road via Northam Bridge, and Pound Road via Itchen Bridge respectively. The location of these routes in Southampton is shown on the map in Figure 4.2.

In order to obtain acceptable observations for analysis, generally 30 observations for each route are necessary due to statistic sample requirements (Kleinbaum et al., 1998). Additional discussion of sample size can be seen in section 9.5.2.1. On checking the bus services along these routes (Table 4.1), it was found that the least service is two buses per hour. As a result, a minimum of 2 surveyors and 18 hours for each route were required for the survey when considering that some misses might happen. These 18 hours were distributed over 3 days, i.e. 6 hours a day. The survey dates were selected to be during the school term on weekdays.

4.3.2 Data collection methods

Various data collection methods are manipulated to collect those data which were described in section 4.2 as follows.

- Bus data

Due to inaccessibility of database such as GPS based AVL system (this is discussed later in section 9.3.2), a designed approach which included a portable GPS receiver and data logger as well as a recording sheet is used for surveyors' ride check (Appendix B). The recording sheet refers the method described in the Transit Capacity and Quality of Service Manual (TCRP, 2000). The data collected in this method includes bus position and their time logs for every second from GPS data, and bus direction (inbound/outbound), bus service number, bus type, boarding and alighting time, arrival and departure time of bus-stops, the number of alighting and/or boarding passengers, and notes from surveyors' recording sheets. These data are then used for calculating bus journey time, dwell time, and identifying bus data.

- General traffic data

General traffic data is used to identify the specific traffic condition at a particular time at an explicit location. There are two sources of general traffic data which could be obtained from Southampton traffic control centre – ROMANSE office. They are ANPR journey time, which provides current journey time for car drivers for specific routes on main corridors, and traffic parameters from the SCOOT system. The archived ANPR journey times are the average journey time for 5 minutes of all vehicles travelling on individual route, for which number plates are recognised and matched by cameras and software developed by SIEMENS. Traffic data includes collected automatically by inductive loop detectors at available SCOOT links such as traffic flow, speed and occupancy, and the signal timing comes from the SCOOT messages UO7, MO2, and M37.

- Road geometry and facilities

The data needed for this part was the location, length of link and route, and the type of layout on the road. A compact, wrist-type GPS receiver, Garmin Foretrex 201 was used to locate the position of bus-stops, junctions, inductive loop detectors, stop lines, pedestrian crossings, and ANPR cameras. After data collection from the field, all positional coordinates of facilities are transformed and presented on a digital map. These locations are used to correspond with bus GPS data which described in above subsection. This is because that data collected from GPS of the on-board survey are merely coordinates and time tags. They could make sense only when these data relate their coordinates with corresponding physical positions such as bus stops or junctions on the map. The lengths of links and routes were measured manually using a Trumetre measuring wheel.

A pre-test was carried out on 12 January 2005 on part of the Bitterne route to make sure all the data were acceptable for data requirements and model formulation. All positional data are converted to National Grid, i.e. Easting and Northing coordinates, using Grid InQuest software. The output of coordinates were imported into ArcView (GIS software), and superimposed onto a 1:1250 digital map of the Southampton road network. The details of above processes are described in Appendix A.

Some observations and refinements for processing the data were made during this initial test. Following are the lessons learned from which improvements were made to the ensuing main survey.

- A bus might stop longer at a bus stop to synchronise with the timetable, even though passenger alighting and boarding had finished. A point was made of mentioning to surveyors to make a note for such special events on the recording sheets.
- During the boarding time at bus stops, drivers took a lot of time to deal with passengers who bought tickets with cash and required change. This fare collection method is thought to be the largest time consumption factor for dwell time. Thus, an additional survey of fare collection methods during dwell time was planned as part of the main survey.
- Distance measurement survey using Trumetre measuring wheels is labour-intensive and time-consuming work. It was found that an alternative method of using the distance measuring tool of ArcView on the digital map gave a similar result to the former approach, which was described in section 3.4.3.2. It was decided to use the latter method for convenience in the main survey.
- The available GPS signals play an important role in collecting bus GPS data. The GPS receiver must be positioned as close to clear sky as possible for tracking available satellites.

4.4 Main Survey

As described in section 4.3, the survey can be achieved in three parts. The first and second parts, namely bus and general traffic data, must collect data in the same period. However, the data of road geometry and facilities could be done before the main survey. The main survey was conducted during 11 am to 6 pm on 13th to 15th of September 2005. The details of these data collections are described in the following subsections.

4.4.1 Bus data

The on-board bus survey was carried out by noting manually on the designed recording sheet accompanied by automatic position logging using portable GPS equipment. All surveyors are requested to note down the departure and arrival time at every bus stop for serving customers as well as the number of alighting and/or boarding passengers. The start and end points of the survey route are also recorded with respect to the boarding and alighting time for journey time calculations. In addition, some useful information about bus type and operation such as direction (inbound/outbound), bus service number, and bus type (whether the bus is low-floor; whether the bus is double-decker, and how many doors did the bus use) are included. In all 284 runs and 1409 dwells were collected. Individual runs and dwells at bus-stops for each direction of each route are shown in Table 4.2.

4.4.2 General traffic data

As described in section 4.2, two sets of data are collected automatically in the traffic control centre and are archived afterwards. The general vehicles' journey time from the ANPR data are continuous 5-minutes averages of journey time from 08:00 on the 13^{th} to 24:00 on the 15^{th}. They contain time stamps, travel times, the number of matches, the number of plates in, and the number of plates out for each ANPR route. The available ANPR routes for this survey and their lengths between two cameras are shown in Table 4.3 respectively. The Avenue route is between Charlotte Place Roundabout in the city centre and Chilworth Roundabout near the M27 motorway; the Bitterne route is between the junction of Brinton's Road and Six Dials near the city centre and the junction of Botley Road/Bursledon Road; and the Portsmouth route is between Saltmarsh Roundabout at the city centre and the junction of Pound Road and Portsmouth Road. The schematics of the above three routes are shown on Figure 4.3. Note that there is no ANPR data available along the Portswood route.

The archived traffic data are during 07:00 to 19:00 of the survey dates. The available SCOOT links and loop detectors for each route are presented in Table 4.4. These detectors were checked with the ROMANSE office before the survey to confirm they were working properly. UO7 and MO2 are the 5-minutes averaged data; however, M37 is signal timing data, which is not at a fixed interval. It is diverse, depending upon changeable cycle length according to the fluctuation of traffic demand.

4.4.3 Road geometry and facilities

This data was collected at the initial stage of the data collection process. The details of the route data are illustrated in Table 4.5. It includes route length, the number of bus stops, the number of signalised junctions, the number of signalised pedestrian crossings, the number of zebra crossings, the number of roundabouts, the number of give-way signs, the number of right turns, the number of left turns, the number of bus lanes, and total length of bus lane of each route. In addition, an example of facility locations on a map is shown on Figure 4.4. It illustrates the location of road facilities such as detectors, stop lines, bus stops, and bus lane on the digital map. Again, these locations are used to check relevant positions for bus GPS data.

4.5 Data checks

It is essential to check for possible errors in the data after collection to ensure that there are no errors in data processing. The checking processes for the data collected in section 4.4 are described in the following subsections.

4.5.1 Bus data recorded by surveyors

This data was collected manually by surveyors in an on-board survey. A total of 284 recording sheets were obtained for all routes. All data are entered into spreadsheets of MS Excel format. Bus journey times are calculated from the time difference of boarding and alighting, and dwell time of each dwell is calculated from the time difference of arrival and departure. Notes marked by surveyors, which are invaluable information for special events, are also included. Special attention is paid to significantly longer or shorter journey time or dwell times to ensure they are error free. A total of 1409 dwells at bus stops was recorded for all sheets. An example of such data checked by the above processes is shown in Table 4.6. It can be seen that the columns of 'Boarding time' and 'Alighting time' are the time when surveyor got on and off a bus and the times are the same when surveyor was on the same bus journey. The column of 'Journey time' is calculated from the time difference of these two figures. The column of 'Total dwell time' is calculated the total dwell times of a bus journey, which are in the 'Dwell time' column, e.g. the first figure of 142s of 'Total dwell time' is calculated from the sum of 25, 37, 8, 35, 35, 9, 17, 11s in the 'Dwell time' column. The column of 'Arrival time' and 'Departure time' are the times when a bus just arrives at a bus-stop after deceleration and departs from a bus-stop before acceleration respectively. The 'Dwell time' column is calculated from the time difference between these two figures.

4.5.2 Bus GPS data

This data was collected automatically in a portable data logger which was carried by surveyors on board the bus, as described in section 4.3.2. A total of 24 files were obtained from 8 surveyors during 3 survey days. Bus positions were logged for every second. As a result, each file may have 25,200 records if the GPS data logger recorded all periods of the survey per single day (7 hours*60 minutes*60 seconds). An example of this raw data is shown in Table 4.7. These data are then presented on a digital map to check the data correctness, which is processed by the approach illustrated in Appendix A, which is described in section 4.3.2. This can be achieved through decomposing a file, i.e. dividing a file into individual bus runs, by referring to the time tags on the recording sheets (section 4.5.1 and Appendix B). Some bus runs have comprehensive GPS coverage of trajectories along the bus routes. However, others lost part or entirely of their trajectory, depending upon the available GPS signals at that time and at the location, as described in Chapter 3. As a result, a total of 284 bus runs were checked visually and data conditions were marked for future reference. Above processes are discussed later in section 9.3.3.

4.5.3 ANPR data

This data is collected and journey times calculated automatically, as described in section 4.3.2. An example of the raw data archived from the ROMANSE office is shown in Table 4.8. It can be seen that the data is 5 min interval in the 'Timestamp' column. The 'Travel time' column is the ANPR journey time. The column of 'Plates in' and 'Plates out' are the number of plates which can be recognized by the ANPR system for entering and exiting vehicles. The column of 'Matches' is the number of the same plates can be matched for entering and exiting vehicles. This stage is to check unreasonable data and discard them from the following analysis. The absurd data is defined as that for which average speeds (route length divided by journey time) are greater than 50 mph (the maximum posted speed limit along any of the routes) or journey times are longer than 1 hour. Generally, the rate of matched plates for calculating journey time was not high. It is thought to be that some cameras are not installed at suitable locations. Note that this ANPR system was launched in the middle of 2005. The locations of some cameras were still under refinement when the survey was carried out.

4.5.4 SCOOT data

This data was collected automatically from the inductive loop detectors of SCOOT as described in section 4.3.2. The data are given special attention in regard to missing values and disabled link arms. The possible error has to be cautiously checked before model formulation. For example, some speed values in the UO7 message have a default value of 50 mph, which are not suitable for use.

The UO7 message from the SCOOT system includes time stamp, loop detector ID, average speed, flow, occupancy, average gap time per vehicle (AGTPV), and average loop occupancy time per vehicle (ALOTPV). The MO2 message includes time stamp, number of vehicle stops, delay, flow, and congestion. M37 includes signal timing of the UTC stage, preceding inter-green, green length, and total length. The content of these messages, definitions and units of each value are available from the SCOOT User Guide (SIEMENS, 1999). The raw data of the UO7, MO2, and M37 messages are shown in Tables 4.9, 4.10, and 4.11 respectively.

4.5.5 Road geometry and facilities

This data was collected manually as described in section 4.3.2. The highlight is put on the possible entering errors and missing values in database. Location of road facilities can be checked by coincidence with the road geometry on the digital map. For example, if a position is away from the road and its relative location with other facilities is wrong, it is thought to be an error and is required to be re-located. The number of facilities, such as bus stops or signalised junctions can be calculated from the map when the above process is completed. In addition, there may be equal numbers of facilities along each direction (inbound/outbound) normally, as long as their driving route is the same. Any difference between them may highlight a possible error.

4.6 Summary

A detailed description of the data collection processes was illustrated in this chapter. Three groups of data, namely bus data, general traffic data, and road geometry and facilities data were required in this study. The data collection was accomplished for main bus corridors in Southampton, UK. Both manual and automatic data collection approaches were used, including surveyors' riding checks, road characteristics surveys, GPS and ANPR technologies, and the SCOOT system. All the collected data were checked to avoid various possibilities of errors. These data are now ready to proceed with the data processing of bus journey time components, which is addressed in the next chapter.

Chapter 5: Bus Journey Time Components

5.1 Introduction

The characteristics of bus journey time were reviewed in section 2.7. Bus journey time has particular features of dwell time, delay of deceleration and acceleration for bus stops, and bus priorities compared with other vehicles. Chapter 3 described and identified the accuracy of journey time data collection using GPS. The required data for modelling were collected and checked in Chapter 4. This chapter aims to formulate the key parts of bus journey time using the collected data.

Bus journey time is separated into three major parts, namely, general travel time, dwell time, and delay of acceleration and deceleration due to passenger services. According to the survey data, bus general travel times, which are similar to other vehicles' travelling situation without bus-stop services, ranged from 64% to 96% of bus journey time with a mean of 84%. Dwell time ranged from 1% to 28% of bus journey time with a mean of 11%. Delay of acceleration and deceleration ranged from 1% to 12% of bus journey time with a mean of 5%. The main factors which affect such deviations are traffic conditions and passenger demands.

This chapter begins with descriptions of bus general travel time (section 5.2) which illustrate buses operation that is similar to other vehicles, i.e. no bus stop services. Then, dwell time (section 5.3) concerning alighting and boarding passengers is explained. After that, particular attention is paid to the delay of bus stops due to deceleration and acceleration (section 5.4). In addition to these main components, other factors which affect bus journey time are described in section 5.5. Finally, a brief conclusion is made for this chapter. Note that this chapter is based on the distinction between buses and other vehicles. Some key factors of journey time delay such as control delay of signalized junctions, which influence both buses and other vehicles, are not discussed in this chapter.

5.2 General travel time

General travel time, which has a similar travelling scenario to other vehicles, i.e. excluding the passenger services at bus-stops, is the main part (64%~96%) of bus journey time. Although buses may not serve bus-stops and have the same traffic environment as other vehicles, their journey time is generally greater than other vehicles as described in section 2.6. The factors that affect such general travel time include vehicle characteristics, lane selection, and timetable adherence. The vehicle characteristics of the bus and the passenger load impede their operation more than other vehicles whether in acceleration, deceleration, turn, or roundabout. Buses usually travel on the near side lane when two or more lanes are available. Travelling speeds in the near side lane are usually slower than other lane(s) owing to turning movement, roadside parking, or slow-speed vehicles. In addition, in order to make the bus timetable reliable, there must be some flexibility for the fluctuation in passenger demand and traffic conditions, This necessitates a longer gap between two check points rather than shorter in printed timetable.

General travel time data is obtained by subtracting dwell time (section 5.3) and delay of deceleration and acceleration due to bus-stop service (section 5.4) from bus journey time. They can be divided by ANPR journey time at the same period to achieve a general concept of relationship based on other vehicles journey time. A histogram of the percentage of general travel time compared with ANPR journey time is illustrated in Figure 5.1. It can be seen that the line displayed on the histogram is the normal distribution line for reference. It can be seen that the percentage distribution is similar to normal. This can be supported by the Kolmogorov-Smirnov (K-S) test of normality illustrated in Table 5.1 (p-value=0.2 > α= 0.05, accepted the hypothesis that there is no significant evidence to suggest that the distribution is not normal). Descriptive statistics are shown in Table 5.2, the percentages rangeing from 0.81 to 1.78 with a mean of 1.34.

5.3 Dwell time

A definition of dwell time, the main factors affecting dwell time, as well as a review of previous studies were described in section 2.7.1. Again, dwell time is the time a bus waits at a bus stop for passengers to alight and board the bus. Dwell time data collected in chapter 4, which is part of the bus data is used in this section. This section starts with passenger distribution along a bus route (section 5.3.1). Then, the dwell time of each stop is explained (section 5.3.2). After that, the dwell time of each passenger is illustrated in section 5.3.3. This is followed by a statistical test to identify the key factors in dwell time (section 5.3.4). Finally, a model formulation of dwell time (section 5.3.5) is presented.

5.3.1 Passenger distribution along bus route

The major difference between buses journey time and other vehicles' is bus-stop service for boarding and alighting passengers. It is essential to understand passenger demand along a bus route in order to estimate how many bus-stops are made by a bus and hence bus dwell times. However, it is not simple. The demographic distributions and possible passengers vary among different bus routes. The boarding and alighting passengers at each stop may be closely related to time factors such as time of day, the day of the week etc. Therefore, it is both impossible and unnecessary to have a common equation to applicable to all bus routes. In practice, there are two approaches which can be used for this issue. One is a theoretical assumption; the other is a field survey for the proposed route. These are described in the following.

Most of the studies in the literature use theoretical assumption. This is because the real demand of passengers involves too many factors, which are difficult to obtain and may be changeable over time. Thus, considerable assumptions such as passenger arrival distribution, the average gap between buses, and the maximum capacity of the bus etc. were made for this approach (Guenthner and Snha, 1983; Lobo, 1997; Horbury, 1999).

The most practical way to understand passenger demand is to measure it directly. If an automatic passenger counting system is not available for this purpose, then a riding check survey, which was used in previous studies, is required. Such an approach was used in data collection in section 4.3.2. The distribution of number of passengers per stop along bus routes are shown in Table 5.3. The average value was used in this table and was calculated from the total runs of each direction of each route. It can be seen from the Table 5.3 that distributions are different among different routes, and that even different directions on the same route are not similar. In order to compare the difference between the data collected and the theoretical distribution which was suggested by previous studies, the values in Table 5.4 are averaged to calculate the probability of the number of passengers at each stop. The theoretical distributions, which are used for comparison with survey data, are Poisson and negative binomial distribution, which were suggested by previous studies (Guenthner and Snha, 1983; Lobo, 1997; Horbury, 1999). The results are shown in Table 5.4. It can be seen that when the number of passengers at each stop is relatively low, say smaller than 6, Poisson distribution can be used acceptably. However, if the number of passenger at each stop is greater then 6, then a negative binomial distribution could be a better alternative. A further description of distribution with respect to the use of discrete or continuous distribution for a better fit is presented in section 6.3.2.2 and section 9.5.3.1.

5.3.2 Dwell time at each bus stop

Originally, a total of 1,083 dwells were collected in the survey. Among these records, a total of 128 records contained notes, which took longer or shorter dwell time than usual or indicated a specific situation during the data collection process. The notes and the number of records for these unusual activities are shown in Table 5.5. For some notes, for example, bus stopped over bus-stop, equipment operation, and so on, these events are not usual and such activities do not involve passenger service at bus-stops. Therefore, these records (93 records) were excluded (with excluding mark '*'on the last column in Table 5.5) from the analysis afterwards. Overall, 990 records are left.

The mean and standard deviation (SD) of dwell time for each stop were 16.51 and 18.13 s respectively. A frequency histogram is shown in Figure 5.2. Some dwell times are greater than 150 s, but most of them are under 50 s. The line shown on this figure is the normal distribution line for reference. It can be seen that dwell time frequency distribution is far from normal distribution, which is pointed (kurtosis=15.50) and piles up on the left (Skewness=3.33). This can be supported by the test of normality of K-S test (p-value<0.001 less than α=0.05). The best fit distribution using Crystal Ball [CB] software is lognormal distribution which is shown on Figure 5.3.

Previous studies concluded that the time used for boarding and alighting activities are different and generally average alighting times are less than boarding times (Guenthner and Sinha, 1983; Guenthner and Hamat, 1988; TRB 2000; Rajbhandari et al., 2003; Dueker et al., 2004; Patnaik et al., 2004; Zhao and Li, 2005). Thus, the dwell times of each stop are divided into three groups, namely boarding passengers only, alighting passengers only, and combined boarding and alighting passengers. Among these 990 dwells, 347 dwells (35%) are only boarding passengers, 366 dwells (37%) are only alighting passengers, and 277 dwells (28%) are combined boarding and alighting passengers. This category also supported by ANOVA (Analysis of variance) test that the mean value of these three groups are not all the same (F=100.82, p-value<0.001 less than $\alpha=0.05$). In addition, the post hoc test of Tamhane is used for multiple comparisons of each pair. The result showed that combined boarding and alighting passengers at each stop have the greatest dwell time due to at least 2 people, followed by boarding passengers, and then alighting ones and there were significant differences between them.

5.3.3 Dwell time of each passenger

The mean and SD of dwell time of each passenger are 8.20 and 6.70s respectively. A frequency histogram is shown in Figure 5.4. Some dwell times of passengers are greater than 50 s, but most of them are less than 30s. The thick line shown on this figure is the normal distribution line for reference. It can be seen that such frequency distribution is far from normal distribution which is 'pointy' (kurtosis=17.60) and piles up on the left (skewness=3.00) distribution. This can be supported by K-S normality test (p-value<0.001 less than $\alpha=0.05$, hence rejected the hypothesis) which indicates that it is not a normal distribution.

The average time and accumulated percentage for a passenger to board or alight a bus in dwell activities of boarding, alighting, and combined boarding and alighting are illustrated on Figures 5.5 to 5.7 respectively. It can be seen from these figures that they all have extreme values beyond the main histograms on the right hand. A method, which excludes the effect of extreme observations, is adopted for a generalized analysis as following (Zhao and Li, 2005). A further discussion of treating outliers can be seen latter in section 9.5.2.2. Table 5.6 illustrates the representative sample percentage of these three activities. Accumulated percentages below the maximum dwell time are shown on the first column. The second column shows that when the maximum boarding dwell time of 21s is used, this criterion covers 90 percent of cases. In addition, when maximum boarding dwell time of 24, 32s is used, the criterion covers 95, 99 percent cases respectively. The third and fourth column are similar to the second column. It can be seen that 95% accumulated percentage could have an acceptable sample percent and without the extreme dwell times which are far away most of samples. Therefore, dwell times which are beyond these values, i.e. boarding time greater than 24s; alighting time greater than 14s; and both boarding and alighting time greater than 15s, are discarded in the following analysis. Table 5.7 illustrates the difference between the average dwell time with and without extreme values. It can be seen that in the without extreme values scenario only a few observations are discarded from the observations. However, there are many improvements on decreasing the deviation of SD. For example, the reduced observations of average boarding time per passenger are 17 (observations are reduced from 347 to 330), however, SD decreased from 7.56 to 5.26s (30%). Furthermore, the maximum value reduced from 76 to 24s. Similarly, SD of the average alighting time and both boarding and alighting time per passenger are decreased from 4.51 to 2.58s (43%), 5.93 to 3.27s (45%) respectively.

The frequency distribution for the average dwell time of each passenger of boarding, alighting, and combined boarding and alighting are illustrated on Figures 5.8 to 5.10 respectively. Four possible theoretical distributions, namely normal, lognormal, gamma, and Weibull distributions are used to fit the above three activity distributions respectively. The probability density functions (PDF) of these distributions are as follows:

Normal: $$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \quad (5.1)$$

Lognormal: $$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-\left[\frac{(\ln x-\mu)^2}{2\sigma^2}\right]} \quad (5.2)$$

Gamma: $$f(x;\alpha,\beta)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}(x-l)^{\alpha-1}e^{-\frac{x-l}{\beta}} \quad (5.3)$$

Weibull: $$f(x;\alpha,\beta)=\frac{\alpha}{\beta^{\alpha}}(x-l)^{\alpha-1}e^{-(\frac{x-l}{\beta})^{\alpha}} \quad (5.4)$$

Where
- x = average dwell time per passenger, measured in seconds;
- μ = mean value of x;
- σ = standard deviation value of x;
- α = shape parameter(α >0);
- β = scale parameter (β >0); and
- l =location parameter.

It seems somewhat difficult to decide which distribution is the best fit for each activity distribution. Thus, the automatic distribution fit function in Crystal Ball and the K-S test are used to examine whether the dwell time of each passenger follow above distributions. As a result, the best-fit function of each activity is Gamma (K-S value=0.04, location=1.53, scale=2.31, and shape=5.35), Weibull (K-S value=0.08, location=0.50, scale=4.95, and shape=1.77), and Lognormal (K-S value=0.07, mean=5.93, and SD=3.66) distribution for boarding, alighting, and combined boarding and alighting respectively.

5.3.4 Factors affecting dwell time

Several types of data were recorded on the recording sheets of the on-board bus survey as described in section 4.3.2. These data, including different routes, directions, dates, periods, various bus services, and bus types, are used to explore the factors affecting dwell time of each passenger using the ANOVA test. Eight factors are examined in the test. They are 1) routes, 2) directions (inbound/outbound), 3) survey dates, 4) period (peak/off-peak), 5) bus services, 6) whether buses are low floor 7) whether buses are double-decker, and 8) how many doors are used. The dwell time data of three routes for formulation are used for the test. Hypotheses, results, and descriptive statistics of these tests are shown in Table 5.8.

The results show that the dwell times of each passenger in the factor of 1), 2), 4), and 5) were not all the same, i.e. there may be significant differences between them. The results of these factors also demonstrate that the dwell times of each passenger in the Portswood route are far from those on the Bitterne route, but they may be similar to the Portsmouth route. That means different routes may have dissimilar or similar passenger demand in some way. It also supports the view that passenger demand should be independent of each bus route. For the direction factor, they were not all the same between inbound and outbound direction. It implies that passenger demand in the inbound and outbound direction of each route might be different. However, this study collected dwell time data during 11:00 to 18:00 excluding the morning peak hours and this may result in a biased conclusion. It can be seen from the last column of the descriptive statistics of Table 5.8 that the mean dwell times of each passenger in outbound, which may be affected more by afternoon peak hours, are less than inbound (the mean values of outbound and inbound are 6.99, and 7.60s respectively). It is thought that passengers might be in a hurry in the peak hour or that this period contained more young passengers such as students or passengers with more commuters that used un-cashed tickets. There were also significant differences relating to peak and off-peak times. It can be seen from the last column of the descriptive statistics that the mean dwell times of each passenger in peak time were smaller than off-peak time (the mean values of peak and off-peak are 6.50, and 8.04 s respectively), which may support the inference above. Although, there is a significant difference between various bus services, when more details of bus services belonging to a particular bus route were checked, it can be found that the bus services in the same bus route are similar, but they may have dissimilar or similar dwell time in different bus routes.

By contrary, the test results (Table 5.8) also demonstrate that there is no significant evidence to support the difference in factors of 3), 6), 7), 8) of dwell time. In other words, different survey dates, various bus types such as low floor and double-decker, and number of doors used for passenger service are not the major influence on bus dwell time, which were based on the collected data in this study. The survey dates- Tuesday, Wednesday, and Thursday- are all weekdays and may have a similar pattern as described in section 2.5.1. For the bus types, the average number of passengers of each stop is quite small and passengers are less than half of the bus capacity (50 seats) most of the time. There are only a few standees on very few buses even in the peak hour. It is thought that these two factors might have more influence when more passengers are boarding and alighting bus. It is interesting to understand the effect of the number of door used in dwell time. Because of few passengers (average 2 passengers) alighting and boarding at each stop in the survey and only 28% of total dwells with both alighting and boarding activities, this factor is not significant. Although some buses along the Bitterne route have two doors, they did not use the rear door. Therefore, the factor used in the test is the number of doors used rather than the number of doors the bus has. In addition, when two doors are used, passengers should follow the rule of getting on the bus using the front door and getting off by the rear door. However, it frequently happens that some alighting passengers use the front door rather than the rear door on the Portswood route.

5.3.5 Model development

1) Data description

The dependent variable is the bus dwell time at each stop measured in seconds. The independent variables consist of the number of alighting passengers, the number of boarding passengers, and the total number of passengers. The descriptions of the above variables are listed below:

T_D	Bus dwell time per stop (seconds)
P_A	Number of alighting passengers
P_B	Number of boarding passengers

P_T Number of total passengers

2) Relationship between variables

Correlations between each variable are shown in Table 5.9. It can be seen in the T_D column that P_B has the highest correlation (0.70) with T_D and P_T have a high correlation coefficient (0.59) with T_D as well. All the signs of predictors were positive, i.e. a greater number of passengers will result in a longer dwell time, which matches the expectation. However, P_T also had high correlation with P_A (0.78) and P_B (0.75). This suggests a risk of multi-collinearity, which could cause a regression model with inaccurate coefficients as discussed latter in section 9.5.2.2. Thus, it is avoided by using P_T along with P_A and P_B at the same time in the following process.

3) Model formulating methods

Simple and multiple regression methods are used to model dwell time data. Many linear and non-linear regression models were tried. The relationship between bus dwell times per stop against independent variables are illustrated on Figure 5.11 (A)~ (C) respectively. It can be seen from the above figures that scatters might not follow a simple trend and that their variances are not all the same. Thus, more types of regression models should be taken into account. The modelling process began with a simple linear regression model of dwell time T_D as a function of each independent variable. Next, more variables were considered and explored including transformed data such as square (P_A^2), cubic (P_A^3), log, and interaction ($P_A * P_B$, $P_A^2 * P_B$, $P_A * P_B^2$) transformation, which are known as polynomial models (Fox, 1997; Kleinbaum, 1998). Then, logarithm and exponential were also tried.

4) Model selection procedures

The rules, which are used to keep potential models under consideration in this study, are described as follows:

- The model has to be significant from statistic test, i.e. the F value is higher than the critical value or the P value is less than the critical alpha level ($P<\alpha$). Note that α is 0.05 in this study and the selection of significance level is discussed in section 9.6.1.
- The sign of coefficients (positive or negative) of the model has to be reasonable and the coefficients have to be significant from the t test.

- The higher R-square value, the better model it is, but it is not necessarily the highest.
- A compact model including only several key independent variables is better than that with considerable variables of a complicated model which has slightly higher R-square value.

5) Summary of models

The main models under the above rules are summarized in Table 5.10. Each number shows a model of bus dwell time at each stop. There are 4 linear models at the top of the table, followed by 5 polynomial models, then 2 logarithm models, and 2 exponential models on the bottom. The columns illustrate the constants and coefficients of the independent variables of each model. R-square value, F value and/or p value, critical value, and observations are displayed on the right of the Table. All constants and coefficients are significant at the 95% level ($\alpha=0.05$). Note that the dependent variable is the bus dwell time at a bus-stop measured in seconds.

The preferred model is the No 4 of linear model, which interpreted 63 percent of the variation in bus dwell time, which is shown as Equation 5.5. This means that the dwell time has a basic value of 5.07s plus 1.19s for each alighting passenger and 8.88s for each boarding passenger when a bus stops at a bus-stop. This model does not have the highest R square, but it has the advantage of being easily understood with only two predictors, namely the number of alighting passengers and the number of boarding passengers, easily estimated, and comparable with previous studies. Such data is also available from data collection.

$$T_D = 5.07 + 1.19\,P_A + 8.88\,P_B \tag{5.5}$$

5.4 Delay of deceleration and acceleration due to bus-stops

Delay of deceleration and acceleration due to bus-stops was defined in section 2.7.2, which is the time difference between when a bus has a dwell on a bus-stop and when it does not. In short, it is the time loss due to serving the bus-stops.

5.4.1 Decomposing delay of deceleration and acceleration

Delay of deceleration and acceleration can be calculated by adding the time of deceleration and acceleration together and subtracting the time travelled without stopping, which is shown as Equation 5.6.

$$T_{a/d} = T_a + T_d - T_n \tag{5.6}$$

where,

$T_{a/d}$ = delay of deceleration and acceleration;

T_a = time of acceleration;

T_d = time of deceleration; and

T_n = travel time without stopping.

The components of this delay can be determined as

$$T_a = \frac{V_2}{A} \tag{5.7}$$

$$T_d = \frac{V_1}{D} \tag{5.8}$$

$$T_n = \frac{S}{(\frac{V_1 + V_2}{2})} \tag{5.9}$$

where,

V_1 = the bus cruise speed before deceleration;

V_2 = the cruise speed after acceleration;

A = the average acceleration rate;

D = the average deceleration rate; and

S = the total bus travel distance during deceleration and acceleration.

In Equation 5.9 the total travel distance S is determined as

$$S = S_a + S_d \tag{5.10}$$

where,

S_a = travel distance during acceleration; and

S_d = travel distance during deceleration

In Equation 5.10 these distances can be determined as

$$S_a = \frac{V_2^2}{2A} \tag{5.11}$$

$$S_b = \frac{V_1^2}{2D} \tag{5.12}$$

In fact, the crucial factors which determine the delay are the rate of deceleration and deceleration, and cruise speed before and after bus dwell. Owing to the development of data collection technology such as GPS, speed profiles are available for less than one second updating frequency. These data are invaluable for estimating such delay, which was previously not possible.

Speed profiles are obtained from the successive bus GPS data which were collected in section 4.4. The cruise speeds before deceleration (V_1) are identified when bus speeds have significant decrease without apparent increase and drop to 0. Similarly, the cruise speeds after acceleration (V_2) are determined when the bus speeds increased from 0 to a speed which is relatively stable without significant increase. One example for identifying such speeds (V_1 and V_2) is illustrated in Figure 5.12. These two cruise speeds are then used to calculate in Equations 5.7 and 5.8. Then, average rates of acceleration and deceleration are required for calculating the time for deceleration and acceleration (T_a and T_d). The concept of average acceleration and deceleration rate is shown in Figure 5.13. It can be seen that the bold line is the bus trajectory, i.e. bus speed is V_1 initially, then the bus starts to decrease speed at t_1 and drops to 0 at t_2. The period between t_2 and t_3 is the dwell time at the bus stop. After that, the bus begins acceleration at t_3 and increases its speed from 0 to V_2 at t_4. As a result, A and D are the average acceleration and deceleration which are used by Equations 5.7 and 5.8. In addition, the travel time without stopping (T_n) is calculated by the total distance travelled during deceleration and acceleration divided by the average speed of V_1 and V_2. The total distance S is calculated by Equations 5.11 and 5.12.

5.4.2 Model development of acceleration/deceleration rate

In order to estimate the delay of deceleration and acceleration, it is essential to estimate acceleration/deceleration rate accurately. Because the rates are heavily impacted by the cruise speeds before deceleration and after acceleration, it is proposed to formulate a function, which is explained by such speeds.

1) Acceleration rate

A total of 259 acceleration records were obtained using a speed profile check which was described in section 5.4.1. The frequency histogram is shown in Figure 5.14. It can be seen that most acceleration rates are between 0.5 and 1.0 (m/s/s). The line displayed on the histogram is normal distribution line for reference. It can be seen that the acceleration distribution is not similar to normal. This can be supported by the K-S test of normality illustrated in Table 5.11 (p-value<0.001 less than α=0.05). The descriptive statistics are shown in Table 5.12 that the acceleration rates range from 0.21 to 1.79 with a mean value of 0.84 (m/s/s).

Acceleration rates compared with the bus cruise speeds after acceleration are shown in Figure 5.15. It can be seen that there is a positive relationship between them, i.e. acceleration rates increase when the speeds rise. However, it is not thought to be the case when speed exceeds 50 kph. Thus, the model of best fit of regression function might not be a simple linear one. After several trials of different regression models, it was found that exponential function is the best fit function which is displayed in the Figure 5.15, but its explanatory power is quite low (R-square=0.08), hence it requires looking for the other function. A method, which uses the function of SPSS (Statistical Package for the Social Science) for curve estimation, is tried. The result showed that cubic function, which is shown as Equation 5.13, is the best fit; however, the R-square value (0.10) is still low. Therefore, a further analysis for better estimating the acceleration rate is required.

$$A = 0.12 + 0.46 \times V_2 - (8.1\,E-04) \times V_2^2 + (3.78\,E-06) \times V_2^3 \tag{5.13}$$

The percentage of speeds range over 50 kph is only 10% of collected records and their trend and variations are quite different from the other part (0 ~50kph), hence the attempt to discard them from the following analysis. The best curve of excluding such speeds is shown as Equation 5.14, which has an R-square value of 0.21. This result also cannot be accepted due to low explanatory power and discarding some data.

$$Ln(A) = 0.66 - \frac{8.57}{V_2} \tag{5.14}$$

A classification of speed range is then attempted in order to look for a better estimation function. The speed range was divided into 5 categories, namely less than 20, 20-30, 30-40, 40-50, and greater than 50 (kph). The best fit function and its R-square value of each category is illustrated in Table 5.13. It can be seen that the R-square value in each speed category is smaller than Equation 5.14 which has 0.21, except the speed category under 20 kph (R-square=0.40). As a result, an alternative using transforming data was adopted.

Transformations can often assist in modelling of data (Fox, 1997). The dependent variable of acceleration rates are transformed as Equation 5.15. Then, a linear regression model with R-square value of 0.97, which is shown as Equation 5.16, was obtained. The scatter plot of transformed acceleration rates against bus speeds are shown in Figure 5.16. It can be seen that this function fits the data very well. Thus, this function requires conversion of the transformed acceleration rates (Equation 5.15) into original acceleration rate, which is illustrated as Equation 5.17. This model's estimates compared with observed acceleration rates are shown on Figure 5.17. Its residual of model estimates, which are presented as observed values subtracting model estimates, are plotted against bus speeds in Figure 5.18. Although the residual distribution is not perfect and it may underestimate the rate when bus speeds greater than 50kph, it is thought that the acceleration rate estimation model (Equation 5.17) is moderate.

$$A' = \frac{(100 * A)^{0.1}}{\sqrt{V_2}} \qquad (5.15)$$

$$A' = -0.14 \times Ln(V_2) + 0.78 \qquad (5.16)$$

$$A = \frac{(-0.14 \times Ln(V_2) \times V_2^{0.5} + 0.78 \times V_2^{0.5})^{10}}{100} \qquad (5.17)$$

2) Deceleration rate

Similar processes, which were described in the above section (acceleration rate), are conducted for estimating deceleration rate as well. A total of 299 deceleration records were obtained using a speed profile check which was discussed in section 5.4.1. The frequency histogram is shown in Figure 5.19. It can be seen that most acceleration rates are between -1.3 and -0.5 (m/s/s). The line displayed on the histogram is the normal distribution line for reference. It can be seen that the acceleration distribution is not similar to normal. This can be supported by the K-S test of normality illustrated in Table 5.14 (p-value<0.001 less than α=0.05). Descriptive statistics are shown in Table 5.15 and the deceleration rates ranges from -2.3 to -0.27 with a mean of-1.03 (m/s/s).

The dependent variable of acceleration rates were transformed using Equation 5.18. Then, a linear regression model with R-square 0.97, which is shown as Equation 5.19, was obtained. The scatter plot of transforming acceleration rates against bus speeds is shown in Figure 5.20. It can be seen that this function fits the data very well. Thus, this function requires the transformed function (5.19) to be converted into original acceleration rate, which is illustrated as Equation 5.20. This model's estimates against observed acceleration rates are shown in Figure 5.21. Its residual of model, which are presented as observed values subtracting model estimates, is plotted against bus speeds in Figure 5.22. It can be seen from this figure that the residual distribution is not perfect and may overestimate the rate when bus speeds greater than 55kph. However, it is thought that the acceleration rate estimation model (Equation 5.20) is moderate.

$$D' = \frac{-(100 * |D|)^{0.1}}{\sqrt{V_1}} \quad (5.18)$$

$$D' = 0.13 \times Ln(V_1) - 0.75 \quad (5.19)$$

$$D = \frac{-(0.13 \times Ln(V_1) \times V_1^{0.5} + 0.75 \times V_1^{0.5})^{10}}{100} \quad (5.20)$$

5.5 Other factors

5.5.1 Bus priorities

Detailed description of bus priority was presented in section 2.7.3. According to the collected data, one type of bus priority in terms of bus lane can be used to explore its effect on bus journey time for this study. Three SCOOT links with various length of bus lane in the Bitterne route are taken into account. They are N10121D, N10214D, and N07221E links, which link lengths are 127, 174, and 322 m and the length of bus lane are 127, 136, and 93 m respectively.

The concern is paid to how much journey time could save due to bus lane. Journey time is associated with travelled distance, hence the comparison of average speeds or journey time per km between buses travelling on a bus lane and other vehicles on normal lanes can be used to check the performance of bus lane. It is clear that bus lane cannot reveal any advantage for the buses over other vehicles when traffic flow is relatively low. Bus lanes can only work when traffic reaches a significant level. Then, buses can drive through the bus lane faster than other vehicles on other lanes. However, this should have a premise that the interference among buses is not serious, i.e. waiting for bus berth at bus-stop or blocking by other buses along bus lane happens rarely. Such phenomenon was not observed during data collection period.

The relationship of SCOOT occupancy (%) against bus speeds and other vehicle speeds is shown in Figure 5.23. It can be seen that when traffic condition of occupancy is greater than 23%, the bus lane has the advantage of travel speed over other vehicles. However, there is no particular effect on bus lane below such value. Nonetheless, the observations of occupancy over 23% were quite few in this case, hence additional study is essential to obtain a robust conclusion.

5.5.2 Schedule adherence

The reliability of the bus timetable is one of the most important indicators of bus service performance. Thus, it is thought that bus drivers may do their best to match the scheduled timetable at checkpoints, as displayed on issued leaflets. Schedule adherence is used to identify this effect on bus journey time. In this study, departure adherence, which calculates the difference between the timetable and the actual departure time of a bus at the departure point of bus route in the survey, is used to relate with the bus journey time. Intuitively, the bus driver might drive more slowly if the bus departs ahead of the timetable and faster when the bus departs behind schedule.

On-time buses are defined as buses which depart within one minute of the scheduled timetable either earlier or later. For example, departure adherence is identified as 5 when buses departed after the timetable for 6 to 7 minutes and is labelled as -3 when buses left before the timetable for 4 to 5 minutes and so forth.

A total of 202 observations were identified. All the data are averaged for each group of departure adherence, which are ranged from -3 to 25. Bus journey times per km against such departure adherence are presented in Figure 5.24. It can be seen that when a bus is on time, the bus journey time is 172s per km. Bus journey times are similar when departure adherence is between -1 and 3, i.e. before timetable 2 minutes and after 4 minutes. However, bus journey time per km decreased gradually when departure adherence rose exceeded 4 and declined significantly when departure adherence increased greater than 15. By contrast, when departure adherence dropped to less than -1, bus journey time per km increased significantly.

A simple linear regression model is initially obtained by weighting the count of each departure adherence value, which is presented as Equation 5.21 with R-square 0.70. It can be seen that the slope is negative and contributes to 2.88 s per km for per unit of departure adherence. Furthermore, a higher R-square (0.84), which fits polynomial function as Equation 5.22 is obtained, which is shown as a fitted line in Figure 5.24.

$$Y = 175.81 - 2.88X \text{, } R^2 = 0.70 \qquad (5.21)$$

$$Y = -0.02X^3 + 0.76X^2 - 10.18X + 185.14 \text{, } R^2 = 0.84 \qquad (5.22)$$

Where,
Y= Bus journey time (seconds per km)
X= Departure adherence

In order to achieve the reliability of bus service, some strategy such as bus holding is used to enhance on-time adherence. Such approach is discussed later in section 9.5.1.2.

5.6 Summary

A detailed explanation of bus journey time components was presented in this chapter. Three main parts, namely general travel time, dwell time, and delay of acceleration and deceleration due to stop services, were discussed individually according to the data collected in Chapter 4. In addition, the effect of bus priority and schedule adherence were illustrated as well. Some fundamental functions were obtained of each component, which are used to formulate a bus journey time estimation model, which is addressed in the next chapter. A further discussion of key components and influential variables of bus journey time is discussed in section 9.4.1.

Chapter 6:

Development of Bus Journey Time Estimation Models

6.1 Introduction

Methodologies for estimating bus journey time were reviewed in sections 2.4 and 2.8 from which it was understood that a regression approach would have the ability to explore the possible factors which may affect bus journey time, but that Monte Carlo simulation enables to manipulate the stochastic variables to have a more complete view of possible estimations with probabilities. Both approaches have been used in this research. The accuracy of journey time data was illustrated in Chapter 3. The data required to develop models was collected and checked as reported in Chapter 4 and the main components of bus journey time were clarified in Chapter 5. The research described in this chapter formulated and verified bus journey time models from the database.

The chapter starts with the development of bus journey time using regression approach (section 6.2), which includes the methodology for development and formulates the models built on route and link basis. The model development using Monte Carlo simulation is described in section 6.3 with similar processes. The chapter is concluded with a summary.

6.2 Regression Approach

Regression methodology has been the main approach used in previous studies as described in detail in section 2.8.3. It has the superiority of manageable interpretation and application.

6.2.1 Procedure

In this study, a procedure is built to conduct regression analysis, which is presented in Figure 6.1. Followings are the details of the processes in the procedure.

(1) Identify model structure

Bus journey time estimation model is separated into several main parts. Each main part is identified and defined

(2) Identify dependent and independent variables

The dependent variable such as bus journey time and potential independent variables, which were collected in the survey, are identified and described.

(3) Explore data

The correlation among variables including between dependent and independent variables and between independent variables are explored to understand the importance of potential variables and their trends, and the possibility of colinearity.

(4) Criteria for selecting a model

The criteria used to keep the potential models in this study as illustrated in section 5.3.5, are again described in the follows:

- The model has to be significant from statistic test, i.e. the F value is higher than the critical value or the P value is less than the critical alpha level ($P<\alpha$). Note that α is 0.05 in this study.
- The sign of coefficients (positive or negative) of the model has to be reasonable and the coefficients have to be significant from the t test.
- The higher R-square value, the better model it is, but it is not necessary the highest.
- A compact model including only several key independent variables is better than that with considerable variables of a complicated model which has slightly higher R-square value.

(5) Strategies for selecting independent variables

This is concerned with determining how many variables and which particular variables should be in the model during the process of analysis. Generally, backward elimination procedure is used for conducting the analysis. However, if the results are not acceptable, then other alternatives such as forward selection, and stepwise procedures are tried. Readers can refer to Fox (1997), Kleinbaum et al. (1998) and Field (2005) for more details about these strategies.

(6) Conduct analysis and model selection

SPSS is used for conducting model formulation using the strategies in (5) and the model selection is based on (4).

(7) Probable preferred models including critical independent variables

This is concerned with determining which critical variables should be in the model when alternative models are available. Such variables enable the model to represent the outcome of bus journey time, can facilitate the variation of traffic fluctuation, and the potential ability to estimate on other bus routes. For example, the length of route/link, the number of alighting and boarding passengers, and the variables which can represent traffic conditions, are the essential variables. A further discussion about selection for preferred model can be seen in section 9.5.2.2.

(8) Examine the selected models

The models obtained from (6) is examined with (7) to achieve a model which is preferred and valid under the statistic test in (4).

(9) Preferred model

This is the model which is validated by (8), but might not be the best performance of regression indicators such as R-square.

In this study, two mechanisms were used to approach the estimation of bus journey time, namely route-based and link-based methods. Most of empirical models of bus journey time found in the literature, which used a regression approach, were route-based method, because the availability of related data for separated links was hard or laborious to access. This research has benefited from the development of travel time data collection technique such as ANPR, GPS and the inductive loop detectors of SCOOT system. Such techniques make it possible to obtain more detailed and various data from individual section of road and therefore, more elaborate link-based approach was also tried to formulate bus journey time.

6.2.2 Route based

6.2.2.1 Model structure

The total route journey time (BJT_R) taken by a bus to travel a road section can be separated into two main parts: time spent in travelling which may be similar to that of other vehicles (JT) and time spent at bus stops (DT), represented by Equation 6.1.

$$BJT_R = JT + DT \qquad (6.1)$$

where,

BJT_R = total time for a bus travelling along a route;

JT = total time for a bus travelling along a route excluding dwell times; and

DT = total stop times for a bus served for alighting and boarding passengers.

$$JT = L \times JT_G$$

where,

JT_G = bus general travelling time per kilometre; and

L =route length measured in kilometres.

Each of these times is related to its explanatory factors which can be modelled individually. The modelling of DT was illustrated in section 5.3.5 and JT_G is modelled as follows.

6.2.2.2 Data description

The dependent variable (response) is bus route journey time per kilometre (JT_G) measured in seconds. The independent variables (predictors) includes the number of bus stops per km (N_{stop}), the number of signalized junctions per km (N_{sig}), the number of disturbances per km (N_{dist}), the number of actual bus stops per km ($N_{stopped}$), the number of left turns per km (N_l), the number of right turns per km (N_r), the percentage of bus lane length ($Buslane\%$), adherence of bus timetable (AD), ANPR journey time per km (JT_{ANPR}), SCOOT speed parameter (ST_s), SCOOT flow parameter (ST_o), SCOOT occupancy parameter (ST_f), the dummy variable of direction factor (D), and the dummy variable of peak/off-peak period (P). More detailed descriptions of the above variables are described as follows:

N_{stop}	The number of bus stops per km	Total number of bus stops along the bus route divided by route length
N_{sig}	The number of signalized junctions per km	Total number of signalised junctions along the bus route divided by route length
N_{dist}	The number of disturbances per km	Total number of disturbances including signalised junctions, pedestrian crossings, roundabouts, and give way junctions along the bus route divided by route length
$N_{stopped}$	The number of stopped bus-stops per km	Total number of stopped bus-stops for which bus stopped to serve passengers along bus route divided by route length
N_l	The number of left turns per km	Total number of left turns along the bus route divided by route length
N_r	The number of right turns per km	Total number of right turns along the bus route divided by route length
$Buslane\%$	The percentage of bus lane length	Total length of bus lane divided by the bus route length
AD	Adherence of bus timetable	Departure adherence of timetable which was explained in section 5.5.2
JT_{ANPR}	ANPR journey time per km	All vehicles journey time obtained from ANPR system divided by ANPR route length
ST_s	SCOOT speed parameter (kph)	The speed value of traffic parameter obtained from SCOOT system
ST_o	SCOOT occupancy parameter (%)	The occupancy value of traffic parameter obtained from SCOOT system
ST_f	SCOOT flow parameter (vehicle per 5 minutes)	The flow value of traffic parameter obtained from SCOOT system
D	The dummy variable of Direction factor	Inbound (toward city centre) or outbound (away city centre) (inbound: 1; outbound: 0)
P	The dummy variable of peak/ off-peak period	Peak: 1; off-peak: 0

An averaged SCOOT parameter (ST_s, ST_o, and ST_f) is used in this study, i.e., each SCOOT link parameter is averaged during a journey of the bus journey time. These SCOOT links along the bus route are then averaged to obtain a final ST_s, ST_o, and ST_f.

For example, There was a bus journey time of 20 minute 05 second from 14:45:57 to 15:06:02 pm on the outbound Bitterne route on 13/10/05. This period contained 5 slots of data including 14:45:00~14:50:00, 14:50:00~14:55:00, 14:55:00~15:00:00, 15:00:00~15:05:00, and 15:05:00~15:10:00 related to the UO7 SCOOT message. Data for these five slots are averaged for each parameter (speed, flow, and occupancy) of each link. Then, all SCOOT links along outbound Bitterne route are also averaged for each parameter to achieve final ST_s, ST_o, and ST_f values.

As it is crucial to understand the possible interrelationships between independent dependent variables to prevent colinearity before conducting the modelling processes. The following subsection demonstrates relationship between response and predictors, and between predictors.

6.2.2.3 Relationship between variables

Correlation coefficient (r) between each variable are given in Table 6.1 and key points are described below. It has been expected that the higher r has the higher strength of the straight-line relationship with bus journey time JT_G, i.e. when r is close to 1, a predictor with a high (low) value will likely have a high (low) value for response. However, if r is close to zero, there is little linear association between predictors and response, but perhaps there is high non-linear relationship between them (Kleinbaum et al., 1998).

For the correlation between response and predictors, it can be seen from Table 6.1 in JT_G column that the number of disturbances N_{dist} has the highest correlation (0.66), and then follows the number of signalized junctions N_{sig} (0.62), and the number of bus stops $N_{stopped}$ (0.58). It also has been expected that JT_G would increase with the increase of the number of left turns (N_l), the number of right turns (N_r), and SCOOT flow parameter (ST_o). However, the negative signs on the coefficients indicated the opposite. This is thought to be the interaction between predictors or constrained data for specific variables may cause this problem.

For the correlation between predictors, it has been expected that the higher r between two predictors might have the risk of colinearity if both of them are in a regression model. It can be seen from Table 6.1 that the number of signalized junctions N_{sig} and the number of disturbances N_{dist} (0.82), SCOOT occupancy parameter ST_o and speed parameter ST_f (0.81), and N_{dist} and the number of right turns N_r (-0.80), have high r.

This gives the caution in modelling process that these pairs cannot put in a model or should have additional intersection predictor to reduce the risk.

JT_G compared with all independent variables are plotted on Figure 6.2. The lines showed in each figures are the linear trend of scatter plots for reference. It can be obtained from the comparison between the representative of reference line and the spread of scatters that whether or not a particular independent variable can be fit with linear association and the slope magnitude which can refer to the values of correlations which were illustrated in Table 6.1. Key points are described below:

- The scatter plots of most figures are thought to be not continuous due to the lack of comprehensive data in between. Generally, for a true underlying straight-line model, these trends of the independent variables provide moderate help in estimating response in terms of bus journey time.
- The relationship between JT_G and other traffic parameters such as ANPR journey time (JT_{ANPR}) and SCOOT parameters of speed, occupancy, and flow (ST_s, ST_o, and ST_f), which are presented on Figure 6.2-I, J, K, and L respectively, might not be represented with a linear model well.
- The dummy variables of direction and time period factors compared with JT_G are illustrated on Figure 6.2-M and N. It can be seen from Figure 6.2-M that the outbound (0) has higher bus journey time than inbound (1). This is thought to be the data collection period (11:00-18:00) including more on afternoon hour which have more traffic away from city centre hence higher JT_G. There is no significant difference between the peak (1) and off-peak factor (0) on Figure 6.2-N.

6.2.2.4 Selection of regression models

The major multiple regression models are summarized in Table 6.2. Model selection methods was developed as described in section 6.2.1. A model of bus route general journey time (JT_G) is shown in each row. R-square values, adjusted R-square values, P-values, F values, critical values, observations, and the number of predictors in each model are displayed at the right hand side of the table. All the constants and coefficients shown were found to be significant at the 95% level except with brackets. The independent variable was bus route general journey time (JT_G) which excluded dwell time measured in seconds. The following comments are relevant to the modelling process.

- For each route, some contribution of predictors towards response are fixed such as the number of bus stops (N_{stop}),the number of signalized junctions (N_{sig}), the number of disturbances (N_{dist}), and the percentage of bus lane length ($Buslane\%$), e.g. a bus route has certain number of bus-stops, signalised junctions, and fixed length of bus lane; these constant parameters showed constant estimates in the model. Thus, the ability for bus route general journey time predicting the traffic variance depends upon other predictors such as adherence of bus timetable (AD), ANPR journey time (JT_{ANPR}), and SCOOT traffic parameters (ST_s, ST_o,and ST_f). Among these key explanatory factors of traffic fluctuations, the contribution from AD is limited. Therefore, the availability of ANPR journey time and SCOOT parameters are crucial to achieve better estimates of bus journey time.
- Some predictors such as the number of left turns (N_l),the number of right turns (N_r), direction factor of D, and time period factor of P are not shown on selected models in Table 6.2. It is considered that these predictors are not the key factors on bus general journey time for collected data.
- The numbers of predictors in selected models are generally 4 or 5, which are suitable for simplified model and easy for estimating afterwards.

All the signs of coefficients were checked carefully to make sure that they are expected, e.g. when the number of bus-stops increase, the response of bus journey time would increase, hence this variable with positive sign (+). As the models in Table 6.2 were all significant, consideration was given to the best performance of error representation, which uses mean error, mean absolute error, mean absolute percent error, and root mean square error to calculate above models' errors. Such results are shown in Table 6.2-1. It can be seen that model 3 has the lowest value of above error measures, which use SCOOT occupancy parameter. In addition, in order to accommodate some road network with ANPR system, model 8 is also selected for alternative. They could be used as alternatives depending upon which of the SCOOT or ANPR journey time parameters is available. These two preferred models are illustrated as follows.

The preferred model using SCOOT parameters is no. 3 in Table 6.2 and gives as Equation 6.2. Predicted value against observed data and residuals plot are shown in

Figures 6.3 and 6.4 respectively. From Figure 6.3 it can be seen that the predicted value generally fitted the observed data moderate (shown as a 45° straight line) except towards the upper end where values are underestimated. There is no hint of any systematic trend of residual plot in Figure 6.4, which indicates that this linear pattern may fit this model. The characteristics of this preferred model can be used as features to be considered in application.

$$JT_G = 55.5 + 70.81\,N_{stop} + 13.45\,N_{stopped} - 1.34\,AD - 2.6\,ST_s + 0.72\,ST_o \qquad (6.2)$$

The preferred model for ANPR journey time is no. 8 in Table 6.2 and gives as Equation 6.3. Predicted values plotted against observed data and residual values are shown in Figures 6.5 and 6.6 respectively. It can be seen from Figure 6.5 that predicted values generally fit the observed data moderate (shown as a 45° straight line). However, the available data for values of JT_G over 200 is sparse due to limitations in the available ANPR data. There is no hint of any systematic trend of residual plot in Figure 6.4, which indicates that this linear pattern may fit this model.

$$JT_G = 58.43 + 31.15\,N_{dist} + 22.61\,N_{stopped} - 365.4\,Buslane\% + 0.22\,JT_{ANPR} \qquad (6.3)$$

From Equation 6.1, BJT_R equals to JT plus DT, the comprehensive bus route journey time model requires to add Equation 5.5 of DT, which requires multiplying with the number of stopped bus-stops ($N_{bus-stop}$). Therefore, the complete models are shown as Equations 6.4 and 6.5 for available data from SCOOT and ANPR respectively.

$$\begin{aligned} BJT_R & \text{(using SCOOT data)} \\ &= JT + DT \\ &= L * (55.5 + 70.81\,N_{stop} + 13.45\,N_{stopped} - 1.34\,AD - 2.6\,ST_s + 0.72\,ST_o) + \\ &\quad (5.07\,N_{bus-stop} + 1.19\,P_A + 8.88\,P_B) \qquad (6.4) \end{aligned}$$

$$\begin{aligned} BJT_R & \text{(using ANPR data)} \\ &= L * (58.43 + 31.15\,N_{dist} + 22.61\,N_{stopped} - 365.4\,Buslane\% + 0.22\,JT_{ANPR}) + \\ &\quad (5.07\,N_{bus-stop} + 1.19\,P_A + 8.88\,P_B) \qquad (6.5) \end{aligned}$$

6.2.2.5 *Summary*

Bus route journey time model consisted of general travel time and dwell time. Detailed description of formulating general travel time model including explanation of variables, relationship among variables, and selection of models were achieved. Two preferred models, which were based on the availability of ANPR journey time or SCOOT traffic parameters, were obtained. An alternative of link based approach is described in the following subsection.

6.2.3 Link based

6.2.3.1 Model structure

Bus journey time is composed of several blocks of journey time. The separations of such block journey time could be physical facilities or non-physical characteristics. The physical facilities include road junctions, signalized control junctions, or bus stops. The non-physical characteristics involve fixed length, fixed time travelling, or fixed time sampling according to data collection methods such as GPS. Links separate by SCOOT links are used for this study. A sketch of such separations is illustrated in Figure 6.7. Total journey time can be divided into several link journey times, i.e. $link_{i-1}$, $link_i$, and $link_{i+1}$. Each link contains a single SCOOT link which between detector and stop line. It can be seen from Figure 6.7 that the length of a SCOOT link is less than the proposed link. Such a gap in knowledge has been filled with the extension of SCOOT link, i.e. traffic parameters such as average speed not only used on SCOOT $link_i$ but also extended to the proposed $link_i$. If a SCOOT link is unavailable for the proposed link, an adjacent link with similar environment is adapted to estimate the link journey time. However, if the adjacent SCOOT link is unavailable or the road environment of proposed link varies with the adjacent SCOOT link, an individual link should be identified. Consequently, bus journey time (BJT_L) can be decomposed into four main parts: time spent as other vehicles (JT_l), time spent at bus stops (DT), time spent due to signalized control (Sig), and delay due to deceleration and acceleration for stays (D_{acc}), as presented on Equation 6.6. In addition, JT_l is the total sum of the general travel time of all links (JT_g) along a bus route, as presented as Equation 6.7.

$$BJT_L = JT_l + DT + Sig + D_{acc} \qquad (6.6)$$

Where,

JT_l = total time for a bus travelling along all links excluding dwell times and signal control delay;

DT = total stop times for a bus served for alighting and boarding passengers along all links;

Sig = signalized control delay; and

D_{acc} = acceleration and deceleration delay due to dwells.

$$JT_l = \sum_{i=1}^{n} \left(JT_g\right)_i \qquad (6.7)$$

Where,

JT_g = time for a bus travelling along a link measured in seconds;
i= link number; and
n= total number of links.

Each component of BJT_L can be modelled individually. The modelling of DT was described in section 5.3.5. JT_g, Sig, and D_{acc} are modelled as follows.

6.2.3.2 Link general travel time

• Data description

The dependent variable (response) is bus link general journey time (JT_g) measured in seconds. The independent variables (predictors) includes link length (L_{link}), SCOOT speed (ST_s), SCOOT occupancy (ST_o), and SCOOT flow (ST_f). Descriptions of these variables are listed below:

L_{link}	Link length (m)	SCOOT link separation illustrated on Figure 6.6.
ST_s	SCOOT speed parameter (kph)	Speed value of traffic parameter obtained from SCOOT system
ST_o	SCOOT occupancy parameter (%)	Occupancy value of traffic parameter obtained from SCOOT system
ST_f	SCOOT flow parameter (vehicle per 5 minutes)	Flow value of traffic parameter obtained from SCOOT system

Averaged SCOOT parameters for each variable (ST_s, ST_o, and ST_f) are used in this study as described in section 6.2.2.2. The following process is to check interrelationship between response and predictors, and between predictors before conducting the modelling process.

• Relationship between variables

The method of interpretation in this section is similar to section 6.2.2.2, which aims to understand the strength of linear association between response and predictors, identify high values of correlation between predictors to prevent using both of them in regression model from the risk of colinearity, and be aware of the fitness of linear trend for scatter plot of response against each predictor. Key points are described below.

For the correlation between response and predictors, it can be seen from Table 6.3 in JT_g column that L_{link} has the highest correlation value (r) (0.92) and that other predictors have quite low values. It has been accepted that the increase of traffic occupancy and the decrease of travel speed would result in the increase of bus journey time. However, the signs which are shown in the SCOOT parameter of occupancy and speed (ST_o and ST_s) show the opposite. This is thought that the interaction with other predictors and limited data may cause this problem.

For the correlation between predictors, the r between predictors are relative low (less than 0.54), hence little colinearity would happen if all predictors use in a regression model.

Bus link general journey time JT_g compared with all predictors are plotted in Figure 6.8. It can be seen that there are no significant linear association between them except link length (L_{link}) of figure (A). It is thought to be that these diagrams contain the factor of link length which strength might be stronger than those of SCOOT parameters. Therefore, transformation of JT_g to JT_g per km may be essential before examining their actual relationship with JT_g.

Bus link general journey time per km (JT_g per km) compared with SCOOT parameters are shown in Figures 6.9-A, B, and C respectively. It can be seen clearly from Figure 6.8-A that there is a slightly flat of negative slope between SCOOT speed parameter and bus link general travel time per km. However, the linear trend is not clear between SCOOT flow parameter and JT_g per km in Figure 6.9-B. There is an insignificant positive slope between SCOOT occupancy parameter and JT_g per km, which is shown in Figure 6.8-C. Note that insufficient data in high vehicle flow (ST_f >100 veh/5 min) and occupancy (ST_o > 30%) may result in the incomplete interpretation of traffic conditions.

- Selection of regression models

The major regression models are summarized in Table 6.4. Model selection method was described in section 6.2.1. The display in table is similar to the Table 6.2. The dependent variable is bus link general travel time measured in seconds. All the constants and coefficients shown were found to be significant at the 95% level.

The simplest model with only one predictor of link length (L_{link}) interprets 94 percent (84% with constant in model 1) of the variation in response of model 2. R-square values are all greater then 0.94 when models have more than one predictor. The following are relevant to the modelling process.

- Link length (L_{link}) is a crucial predictor in bus link general travel time model, the additional explanatory power contributed from other predictors constrained.
- Only one parameter of SCOOT can be presented in the model. If two or more parameters of SCOOT attend in one regression model, insignificant coefficients of predictors will be caused.
- The entering of ST_s into a model has always an incorrect sign (+).

All the signs of coefficients were checked cautiously to make sense and all models were found to be significant at 95% level. Therefore, an additional identification is required to select a superior model. As the models were significant, consideration was given to preference. All the models were compared with the predicted value of models against observed data which was used in formulating the models, as well as scatter plots of residual (observed values subtracted predicted values). During the visual check process, the following observations were made:

- The models without the predictor of link length (L_{link}) (models 3 and 4 in Table 6.4) are seriously underestimated bus link general travel time (JT_g) and have greater variances when observed JT_g over 40 s.
- The models have difficulties to estimate JT_g when observed data are over 70 s. The percentage of observed JT_g which were greater than 70s is 2.7 percent.
- The variances of predicted values increase when observed data of JT_g increase.

The preferred model is visually selected with better fit of predicted values, which is the model of no. 5 in Table 6.4, which is presented as Equation 6.8. Predicted values against observed data and residuals plot are shown in Figures 6.10 and 6.11 respectively. From Figure 6.10 it can be seen that the predicted value generally fitted the observed data well (shown as a 45° straight line) except towards the upper end where values are underestimated, i.e. this model might underestimate bus link general journey times when they are greater than 70 s. In addition, the predicted values might have greater variances when JT_g increase. Such situation is especially significant when JT_g increase from less than 10 s to 30 s. These features also can be identified with residual plot in Figure 6.11. It can be seen from Figure 6.11 that residuals on the positive part are greater than the bottom of negative part, especially when residuals are greater than 20, i.e. the magnitude of underestimation might be greater than overestimation. The characteristics of this preferred model can be used as features to be considered in application.

$$JT_g = 0.086\, L_{link} + 0.038\, ST_f \quad (6.8)$$

6.2.3.3 Link signalized control delay

The delay of signalized control of each link depends upon traffic condition at specific time point and particular junction. Even at the same junction, the signal timing changes frequently due to that SCOOT is an adaptive system according to the traffic fluctuation and coordinates operation to adjacent junctions. Therefore, it is impossible and is not necessary to obtain a general model which can apply to all junctions. However, in order to estimate bus journey time by link basis approach, this delay has to take into account due to contributing considerable variations to link journey time. For link journey time of this study, link lengths range from less than 100 m to 700 m. Assume that a bus travelling a link of 100m with an average speed of 40 kph (11 m/s). It may take 9 s to travel this link. If the bus encounters red light and stops at stop line for 60 s, the total journey time would be 69s to complete the link. However, if the bus does not encounter red light, the total journey time is merely 9s. Consequently, signal delay contributes considerably on the variability of link journey time.

A combined approach, which includes analytical model and empirical data, are used for estimating link signal delay of this study. The link signal delay (Sig) is determined by multiplying the probability for a bus stopped by traffic signal (P_R) by the average duration of red light (R_{ave}), which is shown as Equation 6.9.

$$Sig = P_R \times R_{ave} \quad (6.9)$$

The probability of a bus stops by traffic signal is determined as Equation 6.10.

$$P_R = \frac{C - G}{C} \quad (6.10)$$

where,

C= average cycle length; and
G= average green time.

and the duration of red light (x) is determined as Equation 6.11, which uniform bus arrivals are assumed.

$$P(x) = \begin{cases} \frac{1}{a} & \text{for } 0 \le x \le a \text{, and } a = C - G \\ 0 & \text{for } x < 0 \text{, and } x > a \end{cases} \tag{6.11}$$

Mean and variance of Equation 6.11 are $\frac{a}{2}$ and $\frac{a^2}{12}$ respectively. As a result, link signal delay (Sig) can be calculated as Equation 6.12.

$$Sig = \frac{C-G}{C} \times \frac{a}{2} = \frac{C-G}{C} \times \frac{C-G}{2} = \frac{(C-G)^2}{2C} \tag{6.12}$$

Signal timing data are extracted from SCOOT M37 message of collected data which were described in sections 4.4 and 4.5. The green time and the length of each stage were averaged individually for available period (7:00-19:00) of each signal junction of each day. Then, the average length of each stage are calculated together as an average cycle length of each day. Similarly, the average green times are averaged using above approach and are identified as bus approaching direction of each junction. These average cycle times and green times of each day are also averaged for survey days to obtain a final average cycle length (C) and average effective green time (G). This approach was adopted by Shrestha (2002) on all traffic signals along a study bus route.

6.2.3.4 Acceleration/deceleration delay

This delay is obtained by multiplying the number of stopped bus stops with the delay value of each stop. Each acceleration/deceleration delay of a bus stop was explained in section 5.4, which is determined with the factors of bus cruise speed before deceleration (V_1), cruise speed after acceleration (V_2), average acceleration rate (A, Equation 5.17), and average deceleration rate (D, Equation 5.20). The number of stopped bus stops ($N_{stopped}$) is dependent on individual bus route character in terms of passenger demand, which might be different to each other. Thus, field survey of passenger demand for the proposed bus route is required to understand this variable properly; otherwise, a suitable assumption of passenger demand is needed.

An assumption of equality of V_1 and V_2 is made, i.e. the bus speeds are assumed the same before and after a bus serves a bus-stop. The link bus speed is determined as link length (L_{link}) divided by link general travel time (JT_g) which is obtained from Equation 6.8. Due to the difficulties to determine where the stopped bus-stops are at specific links, this study focuses on the total delay of acceleration/deceleration from the total number of stopped bus-stops ($N_{stopped}$) without identifying which bus-stop at which link. Therefore, an aggregate average speed from each link along bus route is essential to obtain an average speed for the bus route in order to calculate this delay. Bus average speed (BV) along a bus route is calculated as Equation 6.13.

$$BV = \frac{\sum_{i=1}^{n} BV_i}{n} \qquad (6.13)$$

Where,

BV_i = link bus speed at i link; and

n = total number of links along bus route.

Consequently, the delay of acceleration/deceleration in Equation 5.6 can be converted to as follows:

$$T_{a/d} = T_a + T_d - T_n = \frac{V_2}{A} + \frac{V_1}{D} - \frac{\frac{V_2^2}{2A} + \frac{V_1^2}{2D}}{(\frac{V_1 + V_2}{2})}, \text{ (where } BV = V_1 = V_2 \text{ is assumed)}$$

$$= \frac{BV \times (A + D)}{2 \times A \times D}$$

6.2.3.5 Link based bus journey time

As described in Equation 6.6, bus journey time (BJT_L) is comprised of four components, namely bus general travel time (JT_l), dwell time (DT), signal control delay (Sig), and deceleration/acceleration delay (D_{acc}). Thus, a comprehensive bus link journey time model requires to combine them together, which is illustrated as Equation 6.14.

$$BJT_L = \underset{\text{(Equation 6.8)}}{JT_l} + \underset{\text{(Equation 5.5)}}{DT} + \underset{\text{(Equation 6.12)}}{Sig} + \underset{\text{(Equation 5.6)}}{D_{acc}}$$

$$=\sum_{i=1}^{n}(0.086\mathrm{L}_{\mathrm{link}}+0.038\mathrm{ST}_{\mathrm{f}}\ )_i+(5.07\,N_{stopped}+1.19\ \mathrm{P_A}+8.88\ \mathrm{P_B})+$$

$$(\sum_{i=1}^{n}\frac{(C_i-G_i)^2}{2C_i})+(N_{stopped}*\frac{BV\times(A+D)}{2\times A\times D})\qquad(6.14)$$

6.2.3.6 Summary

Bus link journey time model consisted of general travel time, dwell time, signal control delay and acceleration/deceleration delay. Detailed description of formulating link general travel time model including explanation of variables, relationship between variables, and selection of a preferred model were achieved. Dwell time and acceleration/deceleration delay were discussed in section 5.3 and 5.4 respectively. In addition, a combined approach for estimating link signal delay, which included analytical model and empirical data were illustrated. An alternative approach for bus journey time estimation of Monte Carlo Simulation is illustrated in the following sections.

6.3 Monte Carlo Simulation

In the Monte Carlo approach, sampling experiments are performed with stochastic variation from probability distribution (Rubinstein, 1981). However, due to the lack of availability of field data, most previous studies conducted this stochastic simulation process using theoretical probability distributions. Such distributions required a considerable number of assumptions to simplify the practical situation. Advances in computing technology and commercial tool in Monte Carlo simulation such as Crystal Ball® (Decisioneering, 2005), can provide a comprehensive scenarios with uncertainty identified and evaluated, and thereby get closer to understanding what might happen in reality (Rhodes, 2005).

6.3.1 Procedure

In the Monte Carlo simulation of bus journey time, the individual elements of general travel time, dwell time and acceleration/deceleration delay are expressed as separate modules. Each module has independent variables which have particular probabilistic distribution obtained from field data, modular structure in terms of logical and/or mathematical representation, and output. The procedure of this mechanism is developed as shown in Figure 6.12. Followings are the details of the processes in the procedure.

(1) Identify model structure and subsequent modules

Bus journey time estimation model is separated into several main parts in terms of modules which are identified and defined.

(2) Formulate individual modules

Each module is formulated with the input variables such as fixed values or stochastic variables and is presented as an algorithm or equation.

(3) Simulation settings

Monte Carlo analysis of this study is based on spreadsheet of Mirosoft® Excel and Monte Carlo tool of Crystal Ball. Simulation trials of 10,000 are set and simulator may stop running only when precision control limit reached 95% confidence level. The sampling method for generating random numbers is Monte Carlo mechanism which has more random in generation. An alternative of Latin Hypercube mechanism is an option when the simulation requires more even sampling. The difference between them can be found in Rubinstein (1981), which is beyond this research. An example of outcome of these two methods is illustrated in Figure 6.13, which is similar after 10,000 trials.

(4) Model verification

During above model development stage, it is desired that the simulation model includes all the necessary components and ensures it can work. At this stage, the developed model is not just to run but insure that the model operates as intended (Chung, 2004). The procedure of verification is illustrated in Figure 6.14, which is composed of two verification levels, namely modular level and assembled model level. The mathematical representations of the modules should clarify the relationship between response and input variables. This can be identified by checking the result of modular output for a single step of simulation before trying considerable trials at once, when a module initially becomes executable. Such approach is to ensure each run of simulation must represent a scenario which actually happens in reality (Vose, 2000). Then, the simulation result of each module is verified in terms of probable range and distribution with hypothetical or/and the field data. Refinements of modular structure should be done during the iterations of input, output, and feedback. Thus, it is easier to build and check a model by individual modules than to formulate a complete model directly (McKinney and Engfer, 2004).

The above processes are the modular level verifications. When the results of all modules are acceptable as intended, all modules can be assembled to a comprehensive model and verified at the model level. Such an approach includes the processes derived from the modular level to examine whether or not the developed model can work properly. In addition, verification at the model level may require consideration of the relationship between modules and their possible interactions. The iterations of the verification process at the model level or modular level are ended only when the model is built correctly. Here, verification of the model was conducted by visual checking the output from the series of test runs using hypothetical or/and field data.

(5) Monte Carlo simulation model

This is the model based on modular formulation in (2) and is verified with hypothetical or/and field data in (4).

Assumptions were made for this study that components among bus journey time, variables in each component, and observations showed in Monte Carlo output were all independent. Such assumptions are owing to no interactions were made between each of them by this approach. Crystal Ball software is used to enhance the fitness of particular probability distribution of each variable, execute simulations, and perform simulation results. Both route-based and link-based mechanisms, which are similar to regression analysis, are used for formulating bus journey time models and are illustrated as following subsections.

6.3.2 Route based

6.3.2.1 Model structure

The concept of bus route journey time is derived from time difference between buses and other vehicles. Generally, buses take longer travel time than other vehicles even without bus-stop services, which was described in section 2.6. Buses do have dwell times and acceleration/deceleration delay due to dwells of bus-stops. In addition, they might have some bus priorities such as bus lane and signal priority, which may decrease bus journey time under some conditions. Timetable adherence also contributes adjustments towards journey time. Consequently, bus route journey time (BJT_R) is comprised of six main components, namely general travel time (JT), dwell time (DT), acceleration/deceleration delay (T_{AD}), benefit of priorities (T_p), and schedule adherence

(S_a), which is shown as Equation 6.15. These components are modelled in individual modules in following subsections. Due to constrained data of T_p which was discussed in section 5.5.1 and a difficulty to define S_a of random observations which are generated by simulator, these two modules are excluded in the following analysis.

$$BJT_R = JT + DT + T_{AD} - T_p + S_a \quad (6.15)$$

where;

BJT_R = total time for a bus travelling along a route measured in seconds;
JT = total time for a bus travelling along a route, which condition is similar to other vehicles;
DT = total stop times for a bus serves for alighting and boarding passengers;
T_{AD} = total delay of acceleration and deceleration due to dwells;
T_p = time saving due to bus priorities; and
S_a = journey time adjustment due to schedule adherence.

6.3.2.2 Modules formulation

Each module contains structure, input variables, and output response, which is formulated as follows:

- General travel time (JT)module

The concept is that bus general travel time is a percentage of other vehicles' journey time in terms of ANPR journey time, which can be determined as Equation 6.16.

$$JT = ANPR_{\%} * T_{ANPR} * L_{route} \quad (6.16)$$

where,

$ANPR_{\%}$ = percentage of ANPR journey time per kilometre, which is obtained from bus journey time per km, which excludes dwell time and acceleration/deceleration delay, divided by ANPR journey time per km;
T_{ANPR} = ANPR journey time per km; and
L_{route} = route length measured in km.

There are three input parameters in Equation 6.16. L_{route} is a fixed value for each route. $ANPR_{\%}$ is a probabilistic variable which can be derived from collected data. A detailed description including a fitted distribution, distribution parameters, statistic tests of distribution (Anderson-Darling [A-D], Chi-square, and Kolmogorov-Smirnov [K-S]), and descriptive statistics of this probabilistic variable is illustrated in Appendix C-1. This value can be used to understand the relationship of journey time of buses in relation to other vehicles, which is discussed later in section 9.4.2. A further discussion about goodness-of-fit statistics can be seen in section 9.5.3.1.

T_{ANPR} can be obtained from the historical data of ANPR system. In this study, the differences between survey dates are not significant under statistical test, hence an average value of these three days of each period (5 min) during 11am to 7pm for each direction (inbound, outbound) is used. Then, each direction is fitted with a proper distribution for T_{ANPR}. Some records in the collected ANPR data had extreme values. These data were cleaned using the minimum and maximum possible values which are described in the following. The minimum value is defined as travelling with the posted speed limit for the whole route, for example, a speed limit of 40 mph on a 5 km route may have a minimum 281s (4min 41s) journey time. By contrast, the maximum value is defined as travelling with a speed of 5 mph, i.e. the maximum journey time for a 5 km route is 2,252s (37min 32s). Thus, if the ANPR data are beyond these two values, they are considered to exclude from the following analysis. The distribution of T_{ANPR} of the proposed route, which is dependent on particular route, is described in model validation (section 7.2.3).

- Dwell time module (DT)

The expected dwell time is determined as the number of stopped bus-stops, the total number of alighting passengers, and the total number of boarding passengers, which is determined as Equation 6.17. An assumption is made that alighting time of each passenger at a stop is equal and every passenger's boarding time is the same at the same bus-stop.

$$DT = \sum_{i=1}^{N_{stopped}} (N_{ai} \times T_{ai}) + \sum_{i=1}^{N_{stopped}} (N_{bi} \times T_{bi}) \tag{6.17}$$

Where,

$N_{stopped}$ = the number of stopped bus stops along bus route;
N_{ai} = the number of alighting passengers at i stop;
N_{bi} = the number of boarding passengers at i stop;
T_{ai} = alighting time of each alighting passenger at i stop;
T_{bi} = boarding time of each boarding passenger at i stop;

There are five input parameters in dwell time module, namely $N_{stopped}$, N_{ai}, N_{bi}, T_{ai}, and T_{bi}. They are all probabilistic variables which can be derived from collected data.

Note that passenger demand is different depending upon the individual route characteristics, hence $N_{stopped}$, N_{ai}, and N_{bi} are identified for study route, which is discussed in model validation (section 7.2.3). However, the remainder variables (T_{ai} and T_{bi}) are assumed no significant difference between different routes and different bus-stops. All the collected dwell time data, which includes abnormal notes which were excluded in section 5.3.2 and extreme values which were excluded in section 5.3.3, are considered in these two probabilistic variables for the reality. The detailed descriptions of probabilistic variables of T_{ai}, and T_{bi} are illustrated in Appendix C-2.

- Acceleration/deceleration delay module (T_{AD})

The expected acceleration/deceleration delay (T_{AD}) is determined as the number of stopped bus-stops multiplying by acceleration/deceleration delay per stop which is described in sections 5.4 and 6.2.3.4. An assumption of equality of V_1 and V_2 is also made, hence T_{AD} is determined as Equation 6.18, i.e. the bus speeds are assumed the same before and after a bus serves a bus-stop. The bus speed (BV) is determined as the percentage of SCOOT speed parameter ($ST_s\%$) multiplying by the average SCOOT speed parameter (ST_s) along bus route, which is presented as Equation 6.19.

$$\begin{aligned} T_{AD} &= N_{stopped} \times T_{a/d} \\ &= N_{stopped} \times (T_a + T_d - T_n) \\ &= N_{stopped} \times \left(\frac{BV \times (A + D)}{2 \times A \times D}\right) \end{aligned} \qquad (6.18)$$

$$BV = ST_s\% \times ST_s \qquad (6.19)$$

$ST_s\%$ is derived from bus journey time (BJT_R), which excludes dwell time (DT) and acceleration /deceleration delay ($T_{a/d}$), dividing by the route length (L_{route}). Then, such value divides by the average SCOOT speed parameter (ST_s), which is shown as Equation 6.20.

$$ST_s\% = \frac{\dfrac{(BJT_R - DT - T_{a/d})}{L_{route}}}{ST_s} \qquad (6.20)$$

As a result, there are five probabilistic variables as input parameters, namely $N_{stopped}$, $ST_s\%$, ST_s, A and D. They can be derived from collected data. As discussed in dwell time module, $N_{stopped}$ is dependent on the proposed route. In addition, SCOOT speed parameter (ST_s) is also unique to each direction of individual route. Therefore, they will be discussed when specific route is used for validation (section 7.2.3). The detailed descriptions of other stochastic variables ($ST_s\%, A$ and D) are presented in Appendix C.3.

6.3.2.3 Model verification

The procedure of verification was illustrated in section 6.3.1. The probable outcome including the minimum and maximum values, ranges, and distribution were visually checked by the series of test runs of possible inputs from hypothetical or/and field data. Verification started with checking general travel time module (JT) with intended output for test data. Since the percentage of ANPR journey time ($ANPR_{\%}$) was obtained in Appendix C.1, ANPR journey time (T_{ANPR}) and route length (L_{route}) are inputs for tests.

The outputs of dwell time module (DT) was checked with possible inputs of the number of stopped bus-stops ($N_{stopped}$), the number of alighting passengers (N_{ai}) and the number of boarding passengers (N_{bi}). The alighting time per alighting passenger (T_{ai}) and boarding time per boarding passenger (T_{bi}) were obtained in Appendix C.2. When the module outcome was checked with field data, it was found that the distribution of module output is similar to field data ; however, the module output had longer dwell times of a bus journey (greater than 300s) at the right tail, e.g. an example of Bitterne inbound route, which is shown in Figure 6.15. This is due to the assumption of equality of alighting time or boarding time at a bus-stop in Equation 6.17. For example, the maximum number of alighting and boarding passengers at a stop in this study route are both 7, if the alighting and boarding time of each passenger also have the maximum value in the field in terms of 55 and 94 s respectively, this stop would have the dwell time of 7*55+7*94=1,043s. Of course, it is impossible that passengers all have the maximum alighting and boarding time at a stop in real world. Therefore, if a higher value of alighting or boarding time is obtained in the simulation, in the meanwhile, more passengers get on and off at this stop, the stop would have considerable dwell time. The solution towards this problem may generate the alighting time and boarding time for each passenger instead of above assumption. Such approach may help prevent dwell time with higher values on the right tail of the distribution. Nevertheless, the process of a further programming would be required, which is beyond the function of Crystal Ball. Therefore, setting a filter, which limits the maximum value of dwell time, might be an alternative to solve the problem. For instance, the maximum dwell time value of field data is 326 s in above case; it may be acceptable to set the valid value in the range of 0 to 350 s in this case. The result with such filter is presented in Figure 6.16. It can be seen that the module with filter has better match of the distribution than Figure 6.15. Consequently, a proper filter for the extreme dwell times might be required for this module.

The acceleration/deceleration delay (T_{AD}) was checked outputs with possible inputs of the number of stopped bus-stops ($N_{stopped}$) and SCOOT speed parameter (ST_s), since $ST_s\%$, A and D were achieved in Appendix C.3.

Once above modules had been verified, they are combined together into a comprehensive model and are checked in model verification with the minimum and maximum values, ranges, and distribution of bus journey time outcome.

6.3.2.4 Summary

The components of Monte Carlo model of bus route journey time were described. Three modules were included in this study, namely general travel time, dwell time, and acceleration/deceleration delay modules. Structure, input parameters, and output response of each module were clarified. Then, all modules and model were verified to ensure that they work as intended. A alternative approach of link-based Monte Carlo modelling is illustrated in the follows.

6.3.3 Link based

The definition of link separations is similar to the link-based regression approach , which was described in section 6.2.3.1, which is divided by the available SCOOT links along bus route.

6.3.3.1 Model structure

The structure of link-based Monte Carlo model is similar to section 6.2.3.1. Bus journey time (BJT_L) can be expressed as the total sum of following four modules: time spent as other vehicles (JT_l), time spent at bus stops (DT), time spent due to signalized control (Sig), and delay due to deceleration and acceleration for dwells (D_{acc}), which were presented as Equation 6.6.

6.3.3.2 Modules formulation

Structure, input variables, and output response of each module are described as follows.

- General travel time (JT_g)module

The concept of bus link general travel time is a bus travelling along a link without dwells and signal delay. The scenario is similar to other vehicles' condition. However, buses might generally have slower speed than other vehicles as described in section 2.6. JT_g can be determined as Equation 6.21.

$$JT_g = \frac{L_{link}}{V_{bus}} \tag{6.21}$$

Where,

L_{link} = link length measured in meters; and

V_{bus} = average bus speed along link.

The bus speed (V_{bus}) is determined as a function of SCOOT speed parameter (ST_s), which use regression approach with filtered 796 records of unimpeded (without bus-stop services and signal control delay) bus link travel time data. Each link journey time is achieved by calculating the time difference between entering and leaving time stamps of particular link of bus GPS data which are associated with digital map. Then, average bus speed (V_{bus}) are obtained by dividing link length with above link journey time. Note that, if there are dwells or signal delays nearby stop-line in link journey times, these data are not used for this model formulation due to such times are modelled individually in this research. ST_s is obtained by checking SCOOT UO7 message when buses pass the specific link from above GPS time stamps. ST_s is average speed (3 survey dates) of each period (5 minutes) during 07:00-19:00. An example of above data is shown in Table 6.5.

A scatter plot of V_{bus} against ST_s is illustrated on Figure 6.17. It can be seen that there is a positive relationship between them, i.e. bus speeds increase when SCOOT speeds rise. Significant deviations are seen within it though. Thus, regression model of best fit might not be simple linear formulation. A transformation is made for the response as Equation 6.22. Then, an exponential regression model with R-square value of 0.98, which is presented as Equation 6.23, is obtained. The scatter plot of transformed bus speed (V_{bus}') against SCOOT speed is shown in Figure 6.18. It can be seen that this exponential function fitted the transformed data of bus speed very well. Consequently, this function requires converting the transformed Equation 6.21 into original V_{bus}, which is illustrated in Equation 6.24. Residuals of this model, which is presented as observed values subtracting model estimates, are plotted against SCOOT speeds in Figure 6.19. Although the residual distribution is not perfect, it is thought that bus speeds estimation function (Equation 6.24) is moderate. A comparison of bus speed estimates against SCOOT speeds is presented on Figure 6.20. It can be seen that when SCOOT speeds are less than 20 kph, the estimates are overlapping with SCOOT speeds. However, the gap is increasing when the SCOOT speeds increase.

$$V'_{bus} = \frac{V_{bus}^{0.1}}{ST_s^2} \tag{6.22}$$

Where,

ST_s = speeds from SCOOT UO7 message measured in kph.

$$V'_{bus} = 1.10 \times ST_s^{-1.93} \tag{6.23}$$

$$V_{bus} = 2.62 \times ST_s^{0.65} \tag{6.24}$$

- Dwell time (DT) module

This module is the same as described in route-based model (section 6.3.2.2).

- Signal control delay (Sig) module

The expected signal control delay along bus route is determined as the total signal delay at each signalized junction as Equation 6.25. Each single signalized junction delay is determined as the probability for a bus encounters red signal at a signalized junction multiplying the probable duration under assumption that buses are uniform arrivals and no initial queue, which is shown as Equation 6.26.

$$Sig = \sum_{i=1}^{n} Sig_i \qquad (6.25)$$

Where,

Sig = total signal control delay along bus route; and

Sig_i = signal control delay at i junction.

$$Sig_i = P(\mathrm{Re}\,d)_i \times T_{\mathrm{Re}\,d-i} \qquad (6.26)$$

Where.

$P(\mathrm{Re}\,d)_i$ = probability of a bus encountered a red light at i junction; and

$T_{\mathrm{Re}\,d-i}$ = duration of red at i junction.

- Acceleration/deceleration delay (D_{acc}) module

The expected acceleration/deceleration delay (D_{acc}) is determined as the number of stopped bus-stops multiplying by acceleration/deceleration delay per stop which is similar to the acceleration/deceleration delay module in route-based model (Equation 6.18 in section 6.3.2.2). The same assumption is also made for the quality of V_1 and V_2. However, the bus speed BV_i in Equation 6.19 is replaced by V_{bus} which was described above (Equation 6.24).

6.3.3.3 Model verification

The procedure of verification was explained in section 6.3.1. The probable outcome including the minimum and maximum values, ranges, and distribution were visually checked by the series of test runs of possible inputs from field and/or hypothetical data. Verification started with modular level test and followed by model level test, which are presented as follows:

- Modular level verification

(1) General travel time (JT_g)module

There are two input variables in this module, namely individual link lengths (L_{link}) and SCOOT speed parameter (ST_s), which use the Equations 6.21 and 6.24. L_{link} for each link is fixed value and ST_s is probabilistic variable, hence deviation of this module is derived from only this ST_s variable. When the module outcome was checked with field data, it was found that the distribution of module output is similar to field data ;

however, the module output overestimates JT_g, e.g. an example of Bitterne inbound route, which is shown in Figure 6.21, the module overestimates JT_g about 200s. As a result, further refinement of this module is essential. Note that the observed data includes signal delay due to the difficulties for separating it from field data. In order to compare this data, the estimated general travel time requires adding signal delay.

It was thought that the shift is due to the underestimation of V_{bus} in Equation 6.21. While such value is determined by the ST_s in Equation 6.24. This can be support by checking Figure 6.20 that the estimated V_{bus} is always smaller than SCOOT speed and the gap increases when SCOOT speeds rise. However, it was found that the bus speeds of field data may be greater than SCOOT speeds when checked on Figure 6.19. For instance, when SCOOT speeds is 40 kph, the bus speeds ranges from 10 to nearly 60 kph. As a result, Equation 6.24 underestimates V_{bus}, hence the modular output overestimates the JT_g.

An approach to improve Equation 6.24 is considered by dividing bus speeds of field data into 5 groups of SCOOT speeds, namely less than 20, 20-30, 30-40, 40-50, and more than 50 kph, i.e. when SCOOT speeds are less than 20 kph, all bus speeds, which SCOOT speeds are less than 20, are grouped into together, and so on. The bus speed of each group is determined by its corresponding SCOOT speeds, hence a regression model for each group was obtained. However, the result of this approach also showed markedly pointy distribution than observed data.

An alternative using regression method is achieved with predictors of link length (L_{link}) and SCOOT speed (ST_s), which is shown as Equation 6.27. The distribution of this revised modular output against observed data is shown in Figure 6.22. It can be seen that the modular distribution is similar to the observed. Therefore, this module is acceptable for link general travel time estimation.

$$\mathrm{Ln}(V_{bus})=0.199\,\mathrm{Ln}(L_{link}) + 0.638\,\mathrm{Ln}(ST_s),\ \text{R-square}=0.97 \qquad (6.27)$$

(2) Dwell time (DT) module

This module is similar to the route dwell time module, which is verified in section 6.3.2.3.

(3) Signal control delay (Sig) module

The signal delay of observed data are not available due to the difficulties to separate this part from the general travel time (JT_g). Thus, the adequacy of modular structure (Equations 6.25 and 6.26) is crucial. The concept of the probability of a bus encountered a red light at junction ($P(\mathrm{Re}\,d)_i$) is similar to Anthony (2001). Duration of red light ($T_{\mathrm{Re}d-i}$) is under an assumption that buses are uniform arrivals and no initial queue. The multiplication of above variables represented a simplified scenario of control delay at signalized junctions. However, the aim of estimating control delay of this study is to obtain a average time on a basis of whole bus route, which is not to focus on an individual junction. Therefore, this module might be thought to be moderate.

(4) Acceleration/deceleration delay (D_{acc}) module

This module is similar to the route acceleration/deceleration delay module, which is verified in section 6.3.2.3.

- Model level verification

Once above modules had been verified, they are combined together into a comprehensive model and checked in model verification with the minimum and maximum values, ranges, and distribution of bus journey time outcome. Generally, this model works as proposed according to the verification.

6.3.3.4 Summary

The components of link-based Monte Carlo model of bus route journey time were described. Four modules were included in this model, namely general travel time, dwell time, signal delay, and acceleration/deceleration delay modules. Structure, input parameters, and output response of each module were clarified. Then, all modules and model were verified to ensure that they work as intended.

6.4 Summary

This chapter has described the development process of bus journey time estimation models. Regression and Monte Carlo approaches were used to formulate such models. Both route and link based mechanisms were illustrated. After the modelling processes, verifications were carried out to make sure the models developed were as maybe expected. It was then concluded that these models can work as intended to estimate bus journey time under the bus routes conditions described in the database. As a result, these verified models are then used to validate with field data, which is illustrated in Chapter 7.

Chapter 7: Model Validation

7.1 Introduction

Chapter 6 illustrated the process of model development, based on the components of bus journey time in chapter 5 and data collection in chapter 4. This chapter describes the approaches adopted for checking the developed models to ensure they perform sufficiently accurately. The first part of this chapter describes validation of the models, by checking the model output against the field data collected for formulating the model. The second part presents further validation of the models which are validated in the first part by comparing the model estimates against field data collected independently. The differences of the field data sets manipulated in these two parts were described in Figure 4.1.

7.2 Validation

Validation is defined as the process of ensuring that a model represents reality at a given confidence interval, i.e. the proposed model can represent reasonably the actual system (Chung, 2004). The purpose of validation is to demonstrate that the proposed model performs sufficiently accurately for a specific purpose (Robinson, 2004). However, the significant problem with such validation is that there may not be any accurate real-world data against which the model may be compared. Thus, a validation which uses approximate real-world data from limited samples may not provide absolute confidence on the proposed model, but it could help to increase confidence (Robinson, 2004).

7.2.1 Validation method

There are three comparison approaches which can be used: (1) graphical comparison of data, (2) confidence intervals, and (3) hypothesis tests (Sargent, 1998). These are described below.

(1) Graphical comparison of data

The model estimates and field data are plotted graphically to determine whether the model's outputs have sufficient accuracy for its intended purpose. Two types of figures are used: scatter plots and histograms. For regression model validation, each estimate has corresponding observed data; the number of samples are equal in both sets. Thus, scatter plots are used for a visual check of the regression models. By contrast, the simulation samples are significantly greater than observed data (typically 10,000 against dozens), hence probability values of histograms are used in the Monte Carlo model's validation.

(2) Confidence intervals

Confidence intervals were suggested by Law and Kelton (1991), which contain richer information for validating complex simulation systems. This method checks not only the average levels of both model and field data, but also the comparison of their spread. This can be achieved by matching how closely the mean values differ between the model and the real world, and by comparing the distribution of the data visually. Such an approach can be used as the model range of accuracy for model validation. Readers can refer to Sargent (1998) for more details.

(3) Hypothesis tests

Hypothesis tests can be used to compare the means, variances, and distributions of the output estimates of models and the observed field data. Such an approach determines whether the model's output has an acceptable range of accuracy. There are four different types of test which may be applied to validate models. The selection of the suitable comparison test depends upon whether or not the data is of normal distribution, paired, independent, and similar in variance. A flow chart (Figure 7.1) can help to choose an appropriate hypothesis test (Chung, 2004). Readers can refer to Sargent (1998), Chung (2004), Robinson (2004), and Toledo and Koutsopoulos (2004) for more details about these hypothesis tests. In this study, the normality test in Figure 7.1 is relaxed due to the robustness of the two-sample t procedure, as discussed in Moore and McCabe (2006), and the relatively large samples of this study (the minimum sample is 36). In addition to the test for the valid centre measurement, a two-sample Kolmogorov-Smirnov test was used to test the null hypothesis that the model estimates have the same distribution as the field data. Readers can refer to Field (2005) for more details about this test.

The above methods are used to validate the models which were developed in Chapter 6. Such validation is carried out by comparing the output (journey time, dwell time etc.) of each bus as given by the model with that obtained from the field. The field data which was used for formulating models on the Bitterne route inbound direction is used for validation. The details of this route, which is 6.8 km long towards city centre, were described in section 4.3.1. The following subsections contain the details of the validation process, including the components of the model such as dwell time and a comprehensive model of bus journey time.

7.2.2 Regression model validation

7.2.2.1 Route-based regression model

- Journey time validation

The journey time validation is conducted by comparing the journey time of each bus journey estimated by the model with that from the observed data. There are two route-based regression models, namely SCOOT (Equation 6.4) and ANPR (Equation 6.5) models, which were described in section 6.2.2.4. They were validated in the following ways respectively:

(1) Graphical comparison

The model estimates of the SCOOT and ANPR models against field data are shown in Figures 7.2 and 7.3 respectively. The straight line of a 45° on the figure is the observed data for reference. It can be seen on Figure 7.2 that the SCOOT model can estimate journey time well when less than 1200s, and had acceptable variation. However, it may underestimate journey time due to more scattering below the reference line, and may not properly estimate journey times more than 1400s in this study route. Figure 7.3 of the ANPR model shows the similar situation to the SCOOT model.

(2) Confidence interval

Confidence intervals of field data, SCOOT model and ANPR model are presented on Figure 7.4. They are separate due to that the numbers of field data for SCCOT and ANPR model are different (SCOOT-38 and ANPR-36). It can be seen that the length of intervals are similar. It is thought that this is due to the same number of samples and similar variance. The mean value of the SCOOT model is included in the 95% confidence interval of the field data, and the confidence interval of the SCOOT model is slightly lower than the field data. By contrast, the ANPR model's mean value is about the lower bound of 95% confidence interval of the field data and its mean is significant less than that of the field data. According to above comparison, it may by concluded that the SCOOT model may estimate journey time better than the ANPR model within a

95% confidence interval. However, such a conclusion requires further evidence in order to support a robust result, hence the following statistical test.

(3) Hypothesis test

As described in section 7.2.1, the flow chart of Figure 7.1 was used to select the most appropriate hypothesis test. Again, the normality requirement is relaxed due to the relatively large samples of this study. The paired t-test is selected for this examination. The null hypothesis is that the model is valid for the acceptable range of accuracy ($H_0 : \mu_{model} - \mu_{field} = 0$) and the alternative hypothesis is that the model is invalid for the acceptable range of accuracy ($H_1 : \mu_{model} - \mu_{field} \neq 0$). The significance level (α) of 0.05 was used in this study. That is, if the statistical significance (P-value) is smaller than α, the difference is significant, hence we reject the null hypothesis, i.e. the model is invalid for estimation.

The results of the paired sample test of both models are shown in Table 7.1. It can be seen that the P-values are in the last column (2–tailed Sig.). The test result of the SCOOT model showed insignificant (P-value=0.23 > α) difference from the field data. However, there is significant evidence (P-value=0.04 < α) to reject the null hypothesis of the ANPR model. This confirms the validity of the SCOOT model results with the field data.

- Dwell time validation

The dwell time validation is carried out by comparing the dwell time of each bus journey estimated by the model with that from the field data. Both route-based and link-based regression models used the same dwell time model (Equation 5.5), which was described in section 5.3.5.

(1) Graphical comparison

The model estimates against field data which excludes unusual and extreme observations as described in sections 5.3.2 and 5.3.3, are shown in Figure 7.5. The straight line of a 45° on the figure is field data for reference. It can be seen in Figure 7.5 that the dwell time model can generally estimate journey time well.

(2) Confidence interval

Confidence intervals of model estimates against field data excluding unusual and extreme observations are presented on Figure 7.6. It can be seen that the length of intervals of model estimates is slightly shorter than field data, i.e. the variance in model estimates is smaller. The mean value of the dwell time model is included in the 95% confidence interval of the field data, and the mean position of the model is similar to the field data, which supports the description in the graphical comparison. Nonetheless, such deduction requires further evidence in order to support a robust result, hence the following statistical test.

(3) Hypothesis test

Again, the normality requirement is relaxed in this study. The paired t-test is selected for this trial. The results are shown in Table 7.2. It can be seen that there is no significant difference (P-value=0.624 greater than α) between model estimates and the field data, hence we accept the null hypothesis. This confirmed the validity of the dwell time regression model results with the field data.

In brief, the estimates of the SCOOT journey time model are valid in this case study as well as its sub-model of dwell time. The ANPR journey time model proves to be invalid. The invalid result is refined which is discussed latter in section 9.6.2.

7.2.2.2 Link-based regression model

The link-based regression model (Equation 6.14) is illustrated in section 6.2.3.5. The journey time validation is conducted by comparing the journey time of each bus journey estimated by the model with that from the observed data.

(1) Graphical comparison

The model estimates against field data are shown in Figure 7.7. It can be seen that this model can estimate journey time well only when journey times are less than about 1200s. However, it is thought to significantly underestimate journey time beyond this value. When checking the possible reason for this, it is found that signal delay component of this model cannot accommodate the possible changes in the real world. In other words, the fixed estimate of signal delay of 97.36s is unable to represent the variable conditions in reality. Theoretically, a bus may have a minimum signal delay of 0 if it encounters green at all signalised junctions. By contrast, it might have a maximum signal delay of 453.82 s (in this case) if it encounters all red. If the above two extreme cases of signal delay are taken into account in this model, an adjusted journey time model with an upper bound and lower bound can be obtained. Such results are shown on Figure 7.8. It can be seen that all the field data fall between the two bounds.

In addition, the shortest journey times of the field data are located at the lower bound which has minimum signal delay, and the longest journey times are located at the upper bound which has maximum signal delay. Therefore, this adjusted model enables the presentation the possible journey time of the two extremes; however, it is unable to provide any estimates in between.

(2) Confidence interval

Confidence intervals of the field data against model outputs are presented in Figure 7.9. It can be seen that the length of the model's interval is significantly shorter than that of the field data, i.e. the variance of journey time of the field data was greater than that of the model estimates. The mean value of the model is excluded from the 95% confidence interval of the field data, and the position of model is significantly less than the field data, which shows the possibility of underestimation.

(3) Hypothesis test

The results of the paired sample test of both data sets are shown in Table 7.3. The test result of the link-based model shows significant difference (P-value<0.001 which is smaller than α) from the field data, hence the rejection of the null hypothesis of this model.

Note that the dwell time model, which is used in this link-based journey time model, is the same as in route-based journey time model. In brief, the estimates of the link-based journey time model are invalid in this case study. This result is further discussed latter in section 9.6.2.

7.2.3 Monte Carlo model validation

The procedure of simulation model validation is similar to the verification procedure in Figure 6.14, which was illustrated in section 6.3.1, which is also composed of two levels of validation, namely modular level and assembled model level. Whereas verification is the continuous process to ensure that the developed models operate as intended, validation is the process to ensure that they represent reality reasonably (Chung, 2004). Therefore, the main focus of validation is the outcome of model output instead of model itself in the verification process. In order to obtain an accurate representation of bus journey time, considerable time was spent checking the modular input data, range, and probability distribution of each probabilistic variable with proper statistic distribution.

Then, the simulation result of each module is validated with mean value, range, and distribution with the available field data.

The above processes are the validations at the modular level. When all modules are acceptable at a given confidence level, they can be then assembled to a comprehensive model and checked at the model level in terms of bus journey time. Such an approach includes the comparison model output with real data as per the modular level validation. The validation process in modular and model levels are to ensure the outputs enable to represent the real world scenarios in both accurate and realistic ways (Vose, 2000).

7.2.3.1 Route-based simulation model

The model is illustrated in section 6.3.2. The additional variables required for the Monte Carlo route-based model, which are dependent on the bus route, are passenger demand and traffic data, i.e. $N_{stopped}$, N_{ai}, and N_{bi} in the dwell time module; T_{ANPR} in the general travel time module; and $N_{stopped}$ and ST_s in the acceleration/deceleration delay module. These probabilistic variables are presented respectively in Appendix C.4.

For modular level validation, the simulation result of each module was compared to the observed data by means of the processes described in section 7.2.1. The following are the results of the general travel time module (JT), dwell time module (DT), and acceleration/deceleration delay module (T_{AD}) validation.

- General travel time module (JT) validation

(1) Graphical comparison

The probability of modular estimates against field data is shown in Figure 7.10. It can be seen that the observed data only have a range of general travel time between 800 and 1200s. However, the simulation results show a greater range than the observed data, hence the flatter distribution. This is because the input data of simulation model has comprehensive period (5 minutes interval during 07:00 to 19:00), while the field data is limited with only 21 observations. Thus, the field data may not represent a comprehensive distribution of traffic conditions in the real world.

(2) Confidence interval

Confidence intervals of the field data against the model outputs are presented in Figure 7.11. It can be seen that the length of the modular estimate's interval is significantly shorter than that of the field data. This is due to the considerable number of trials (10,000), which decreases the standard error, hence the range of confidence interval is shorter. The mean value of the module estimate is contained with the 95% confidence interval of the field data, and the position of the model mean is slightly lower than the field data mean.

(3) Hypothesis test

As described in section 7.2.1, the flow chart in Figure 7.1 is used to select the most appropriate hypothesis test. Again, the normality requirement is relaxed in this study. Due to the independent sampling of the simulation and significantly larger sample sizes compared with the field data, it is obvious that the paired t-test cannot be used. In addition, the variance equality assumption may be unrealistic in the context of traffic simulation (Toledo and Koutsopoulos, 2004). Consequently, the Smith-Satterthwaite test, similar to the approximate t-solution procedure which was discussed in Toledo and Koutsopoulos (2004), is selected for this test. The null hypothesis is similar to the test for the regression model, i.e. that the model is valid for the acceptable range of accuracy ($H_0 : \mu_{model} - \mu_{field} = 0$) and the alternative hypothesis is that the model is invalid for the acceptable range of accuracy ($H_1 : \mu_{model} - \mu_{field} \neq 0$). The significance level (α) of 0.05 is also used. That is, if the statistical significance (P-value) is smaller than α, the difference is significant, hence we would reject the null hypothesis, i.e. the model is invalid for estimation.

The result of the Smith-Satterthwaite test is shown in Table 7.4. It can be seen that the variance between these two datasets is significant (P-value=0.002 < α) for Levene's test for equality of variances (on the second column of Table 7.4), which confirmed the assumption of unequal variance above. The mean test result of this module showed that there is not significant evidence (P-value=0.185 > α) to reject the null hypothesis. This confirms the validity of the general travel time module results with the field data.

In addition to the test for the valid measure of centre, a two-sample Kolmogorov-Smirnov [K-S] test is used to test the null hypothesis that the two samples have the same distribution. The result of the K-S test is shown in Table 7.5. It can be seen that the difference between these two data is not significant (P-value=0.09 > α), hence the acceptance of the null hypothesis, i.e. there is not significant evidence to reject that the simulation results and field data have the same distribution. This again confirms the validity of the general travel time module results with the field data, not only the means but also their spread.

- Dwell time module (DT) validation

(1) Graphical comparison

The probability of modular estimates of dwell time against field data is shown in Figure 7.12. It can be seen that the distribution of these two data sets is quite similar. The highest probability of dwell time is between 100 to 200s in field data as well as in the simulation results.

(2) Confidence interval

Confidence intervals of the field data and the modular outputs are presented in Figure 7.13. It can be seen that the length of dwell time module estimate's interval is significantly shorter than that of the field data. Again, this is due to the considerable number of trials (10,000), which decreases the standard error, hence the shorter range. The mean value of module is included with the 95% confidence interval of the field data; and the position of module is significantly higher than the field data; however, the discrepancy is less than 20s.

(3) Hypothesis test

The result of the test is shown in Table 7.6. It can be seen that the variance between these two datasets is not significant (P-value=0.37 > α) for Levene's test for equality of variances. Thus, the independent t-test with equal variance is used. The mean test result of this module shows that there is not significant evidence (P-value=0.10 > α) to reject the null hypothesis. This confirms the validity of the dwell time module results with the field data.

The result of the K-S test is shown in Table 7.7. It can be seen that the difference between these two datasets is not significant (P-value=0.30 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and field data may have a similar spread. This again confirms the validity of the dwell time module results with the field data, both in the measure of centre and distribution.

- Acceleration/deceleration delay module (T_{AD}) validation

(1) Graphical comparison

The probability of the modular estimates of the T_{AD} module against field data is shown in Figure 7.14. It can be seen that the range of simulation results is significantly greater than the field data, which is 25 to 125s, on two tails. It is thought to be that the module simulates scenarios which might actually happen in reality. Such scenarios included the extreme possibilities in both the shortest and longest delays. However, it is very possible that these extreme possibilities may not be collected in the field data, hence the shorter range.

(2) Confidence interval

Confidence intervals of the field data and the modular outputs are presented in Figure 7.15. It can be seen that the length of the modular estimate's interval is significantly shorter than the field data. Again, this is due to the very numerous trials. The mean value of the module is included within the 95% confidence interval of the field data, and the position of the module mean is close to that of the field data.

(3) Hypothesis test

The result of the test is shown in Table 7.8. It can be seen that the variance between these two datasets is not significant (P-value=0.66 > α) for Levene's test for equality of variances, hence the independent t-test with equal variances is used. The mean test result of this module showed that there is not significant evidence (P-value=0.68 > α) to reject the null hypothesis. This confirmed the validity of the acceleration/deceleration module results with the field data.

The result of the K-S test is shown in Table 7.9. It can be seen that the difference between these two datasets is not significant (P-value=0.76 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and field data may have the same distribution. This again confirms the validity of the acceleration/deceleration delay module results with the field data, both in the measure of centre and distribution.

For model level validation, the simulation result of the bus route journey time (BJT_R) model is compared to the observed data as per the procedure for the modular level.

(1) Graphical comparison

The probability of the model estimates of bus journey time against the field data is shown in Figure 7.16. It can be seen that the range of simulation results is significantly greater than the field data, which is 900 to 1500s on two tails. Again, the relatively limited field data (38 observations) may not form a comprehensive distribution of the real world.

(2) Confidence interval

Confidence intervals of the field data and the model outputs are presented in Figure 7.17. It can be seen that the length of model estimate's interval is significantly shorter than the field data. Again, this is due to the considerable number of simulation trials. The mean value of the module is included within the 95% confidence interval of the field data, and the position of the model mean is very close to the field data.

(3) Hypothesis test

The result of the test is shown in Table 7.10. It can be seen that the variance between these two datasets is significant (P-value=0.005 < α) for Levene's test for equality of variances, hence the Smith-Satterthwaite test with unequal variances is used. The mean test result of this module shows that there is not significant evidence (P-value=0.73 > α) to reject the null hypothesis. This confirms the validity of the bus journey time model results with the field data.

The result of the K-S test is shown in Table 7.11. It can be seen that the difference of spread between these two datasets is not significant (P-value=0.19 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and field data may have the same distribution. This again confirms the validity of the journey time model results with the field data, not only in the measure of centre but also their spread.

In brief, the route-based Monte Carlo model is found to be valid both at the modular level and model level for the field condition of study route, within accepted statistical significance level. Following is the validation of link-based Monte Carlo model, which uses the same processes.

7.2.3.2 Link-based simulation model

The model was illustrated in section 6.3.3. The procedure of validation was explained in the former part of this section. The Bitterne inbound route is also used for the link-based model validation. This route is separated into 14 links comprising 13 SCOOT links and 1 city centre segment which is not a SCOOT link. The details of each link, including link length and SCOOT link ID are presented in Table 7.12. Input parameters of each module are illustrated on Appendix C.5. The outputs of each module and model are validated with field data.

For modular level validation, the simulation result of each module is compared to observed data using the methods described in section 7.2.1. Due to the unavailability of field data for individual general travel time and signal delay, the two modular results are combined together to be compared with the available field data which encompassed both datasets. Following are the validation results of general travel time (JT_g) combined with the signal control delay module (Sig), dwell time module (DT), and acceleration/deceleration delay module (D_{acc}).

- General travel time (JT_g) combined with signal control delay module (Sig)

(1) Graphical comparison

The probability of modular estimates against field data is shown on Figure 7.18. It can be seen that the range of simulation estimates are similar to the observed data except for the additional left tail with some lower values. The distribution of simulation results is also similar to the field data; however, a slightly 'pointy' spread of field data is shown in the figure.

(2) Confidence interval

Confidence intervals of the field data and the combined modular outputs are presented in Figure 7.19. It can be seen that the length of the modular estimate's interval is significantly shorter than the field data. Again, this is due to considerable simulation trials (10,000), which decrease the standard error, hence the shorter range. The mean value of the module is included within the 95% confidence interval of the field data, but the position of this combined modular mean is significantly lower (about 30s) than the field data.

(3) Hypothesis test

As described in section 7.2.1, the flow chart of Figure 7.1 is also used to select the most appropriate hypothesis test. Again, the normality requirement is relaxed in this study. The result of the mean test is shown in Table 7.13. It can be seen that the variance between these two data is significant (P-value=0.005 < α) of Levene's test for equality of variances. Thus, Smith-Satterthwaite test with unequal variances is used for the mean test. The result of this combined module shows that there is not significant evidence (P-value=0.085 > α) to reject null hypothesis. This confirms the validity of the combined module results with the field data.

In addition to the test for the valid measure of centre, the two-sample K-S test is also used to test the null hypothesis that the two samples have the same distribution. The result of the K-S test is shown in Table 7.14. It can be seen that the difference between these two datasets is not significant (P-value=0.053 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and field data may have similar spread. This again confirms the validity of this combined module results with the field data, both in the mean and distribution.

- Dwell time module (DT)

(1) Graphical comparison

The probability of the modular estimates against the field data is shown in Figure 7.20. It can be seen that the range and distribution of dwell time estimates obtained from the simulation is similar to the observed data.

(2) Confidence interval

Confidence intervals of the field data and the modular outputs are presented in Figure 7.21. It can be seen that the length of the modular estimate's interval is significantly shorter than the field data. Again, this is due to the numerous simulation trials. The mean value of the module is included within the 95% confidence interval of the field data and the position of the modular mean is very close to the field data.

(3) Hypothesis test

The result of the mean test is shown in Table 7.15. It can be seen that the variance between these two data sets is not significant (P-value=0.233 > α) for Levene's test for equality of variances, hence the independent t-test with equal variances is used for the mean test. The test result of this module shows that there is not significant evidence (P-value=1 > α) to reject the null hypothesis. This confirms the validity of the dwell time module results with the field data.

The result of the K-S test is shown in Table 7.16. It can be seen that the difference of distribution between these two datasets is not significant (P-value=0.86 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and field data may have similar spread. This again confirms the validity of these dwell time module results with the field data, not only in the measure of centre but also their distribution.

- Acceleration/deceleration delay (D_{acc}) module

(1) Graphical comparison

The probability of the modular estimates against the field data is shown in Figure 7.22. It can be seen that the range of the simulation is greater than the field data, especially at the right tail in terms of longer delay. There was no significant difference in the distribution between these datasets.

(2) Confidence interval

Confidence intervals of the field data and the modular outputs is presented in Figure 7.23. It can be seen that the length of the modular estimate's interval is significantly shorter than the field data. Again, this is due to the considerable trials of simulation. The mean value of the module is included in the 95% confidence interval of the field data, and the position of the modular mean is close to the field data.

(3) Hypothesis test

The result of the mean test is shown in Table 7.17. It can be seen that the variance between these two data sets is not significant (P-value=0.149 > α) for Levene's test for equality of variances, hence the independent t-test which assumed equal variances is used for the mean test. The result of this module showed that there is not significant evidence (P-value=0.365 > α) to reject the null hypothesis. This confirms the validity of the acceleration/deceleration delay module results with the field data.

The result of the K-S test is shown in Table 7.18. It can be seen that the difference between these two datasets is not significant (P-value=0.26 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and field data may have similar distribution. This again confirms the validity of these acceleration/deceleration module results with the field data, both in the mean and distribution.

For model level validation, the simulation result of the bus route journey time (BJT_L) model is compared to the observed data as per the procedure for the modular level.

(1) Graphical comparison

The probability of model estimates of bus journey time against the field data is shown in Figure 7.24. It can be seen that the range of simulation results is greater than the field data on the right tail in terms of longer journey time. Again, this is due to the limited field data (38 observations) which may not form a comprehensive description of the real world. In addition, the distribution of these two datasets looks similar. Both of them have a higher probability of journey time of between 1000 to 1200s.

(2) Confidence interval

Confidence intervals of the field data and the model outputs are presented in Figure 7.25. It can be seen that the length of the model estimate's interval is significantly shorter than the field data. Again, this is due to the considerable number of trials of the simulation model. The mean value of the module is included within the 95% confidence interval of the field data, and the position of model mean is very close to the mean of field data.

(3) Hypothesis test

The result of the test is shown in Table 7.19. It can be seen that the variance between these two datasets is not significant (P-value=0.21 > α) for Levene's test for equality of variances, hence the independent t-test which assumed equal variance is used. The mean test result of this module shows that there is not significant evidence (P-value=0.84 > α) to reject the null hypothesis. This confirms the validity of the link-based journey time model results with the field data.

The result of the K-S test is shown in Table 7.20. It can be seen that the difference between these two datasets is not significant (P-value=0.63 > α), hence the acceptance of the null hypothesis, i.e. the simulation results and the field data may have the same distribution. This again confirms the validity of the link-based journey time model results with the field data, not only in the mean but also the spread.

In brief, the link-based Monte Carlo model is found to be valid both at the modular level and at the model level for the field condition of study route, within accepted statistical significance level.

7.2.4 Validation result

Validation was carried out by comparing the output estimated by the model with that from the field data for the 4 models developed. The results showed that only the route-based SCOOT model was valid for regression models under the case study. By contrast, both route-based and link-based models were valid in the Monte Carlo simulation models, as well as their individual modules. Therefore, a further independent validation on these valid models are required for increased confidence of bus journey time estimation.

7.3 Independent validation

Independent validation of the valid models is conducted to ensure the performance of models enables them to reflect the scenarios in the field properly. This is achieved by comparing the estimates or outputs of the models with data collected independently from that used for model formulation. Models validated in the above section (7.2) are applied. The validation methods described in section 7.2.1 are used.

The field data for independent validation was collected on the Avenue route in Southampton, which was described in section 4.3. An inbound route towards the city centre is used for validation, which is 5.5 km long and has 14 bus stops within it. Both ANPR journey time and SCOOT data are available for this route. Fifteen links which are separated by available 14 SCOOT links and 1 non-SCOOT link are used. Bus frequency along this route is 5 services per hour in peak hours. A total of 39 bus trips and 200 dwells of on-board bus survey were surveyed between 11:00 ~ 18:00 for 3 weekdays. The input parameters of the model and outputs of simulation, which were named 'assumptions' and 'forecasts' in the output report of Crystal Ball, are shown as route-based and link-based Monte Carlo models respectively in Appendix D. Details of the validation results are described in the following subsections.

7.3.1 Journey time validation

The journey time independent validation is conducted by comparing the bus journey times estimated by the models with that from the observed independent field data. The validated two Monte Carlo simulation models, namely the route-based and link-based models (which were formulated in section 6.3.2 and 6.3.3, and validated in section 7.2.3.1 and 7.2.3.2 respectively), are validated as follows.

(1) Graphical comparison

The probability of these two models' estimates against field data is shown in Figure 7.26. It can be seen that these two models generally match the field data well. The distribution of the route-based model is a little flatter than the observed data; however, the link-based model is a little 'pointier' than the distribution of the field data. The highest bars of these data sets in terms of the maximum probability of bus journey times are between 700 and 800s.

(2) Confidence interval

Confidence intervals of field data against model outputs are presented in Figure 7.27. It can be seen that the length of model estimate's intervals are both significantly shorter than the field data. This is due to the considerable number of trials (10,000) which decreases the standard error, hence the shorter range. The mean value of the models are both included within the 95% confidence interval of the field data, and the mean position of the models are both close to but slightly higher than the field, e.g. the mean journey times of the field data, route-based model, and link-based model were 789, 803, and 806s respectively.

(3) Hypothesis test

As described in section 7.2.1, the flow chart of Figure 7.1 is used to select the most appropriate hypothesis test. Again, the normality requirement is relaxed in this study. The null hypothesis is similar to the test conducted above. The null hypothesis is that the model is valid for the acceptable range of accuracy ($H_0 : \mu_{model} - \mu_{field} = 0$), and the alternative hypothesis is that the model is invalid for the acceptable range of accuracy ($H_1 : \mu_{model} - \mu_{field} \neq 0$). The significance level (α) of 0.05 is also used.

The result of the hypothesis test is shown in Table 7.21. It can be seen that the variance between the field data and route-based model is not significant (P-value=0.49 > α) for Levene's test for equality of variances, hence the independent t-test which assumes equal variances is used. The mean test result of the route-based model shows that there is not significant evidence (P-value=0.49 > α) to reject the null hypothesis. This confirmed the validity of the route-based model results with the field data. For the link-based model, the variance between the field data is significant (P-value<0.01 smaller than α) for Levene's test for equality of variances, hence the Smith-Satterthwaite t-test is used. The mean test result of the link-based model shows that there is not significant evidence (P-value=0.42 > α) to reject the null hypothesis. This also confirms the validity of the link-based model results with the field data. Therefore, both models are valid in the mean test.

In addition to the test for the valid measure of centre, a two-sample K-S test is also used to test the null hypothesis that the two samples have the same distribution. The results of the K-S test are shown in Table 7.22. It can be seen that the differences of distribution between the field data and these models' estimates are not significant (P-value=0.40 of route-based model, P-value=0.06 of link-based model, both are greater than α), hence the acceptance of the null hypothesis, i.e. the simulation results of these two models and the field data may have the same distribution. This again confirms the validity of these two models' results with the field data, not only for the measure of centre but also their spread.

7.3.2 Dwell time validation

The dwell time independent validation is conducted by comparing the dwell times estimated by the model with that from the observed independent data. The validated two Monte Carlo simulation models, namely the route-based and link-based models, which both have the same dwell time model, are validated as follows.

(1) Graphical comparison

The probability of dwell time model estimates against the field data is shown in Figure 7.28. It can be seen that the distributions of these two data sets are similar. The model has greater range than the field data, i.e. there are some dwell time estimates greater than 200s, which were not collected in the limited observed data. Most of the dwell times of each bus run are between 0 to 125s, either in the field data or model estimates.

(2) Confidence interval

Confidence intervals of the field data against the model outputs are presented in Figure 7.29. It can be seen that the length of dwell time model estimate's interval is significantly shorter than the field data. Again, this is due to the considerable trials of simulation. The mean value of the model is included within the 95% confidence interval of the field data, and the position of model mean is very close to the field data.

(3) Hypothesis test

The result of the hypothesis test is shown in Table 7.23. It can be seen that the variance between the field data and dwell time model is not significant (P-value=0.22 > α) for Levene's test for equality of variances, hence the independent t-test with equal variances is used. The mean test result of the route-based model shows that there is not significant evidence (P-value=0.88 > α) to reject the null hypothesis. This confirms the validity of the dwell time model results with the field data.

The results of the K-S test are shown in Table 7.24. It can be seen that the difference of distribution between the field data and dwell time model estimates is not significant (P-value=0.53 > α), hence the acceptance of the null hypothesis, i.e. the simulation results of the dwell time model and the field data may have same distribution. This again confirms the validity of the dwell time model results with the field data, both in the measure of centre and distribution.

7.3.3 Independent validation result

Independent validation was carried out by comparing the outputs estimated by validated models in section 7.2 with the field data collected independently. Both route-based and link-based Monte Carlo models were validated. The results showed that these models were valid both in journey time and dwell time under the condition of the study route.

7.4 Summary

This chapter described the validation process conducted for the developed models to make sure that they can work properly with sufficient accuracy as proposed. The process began with validation for the developed regression models and Monte Carlo models. The valid models were then followed by independent validation. Validation data were collected in the field. Results showed that the Monte Carlo models were valid under the field conditions of the study route. The validated models can therefore be used to explore different inputs and their impacts on the bus journey time outputs, in order to identify potential ways to improve bus journey time. The sensitivity analysis is described in Chapter 8.

Chapter 8: Sensitivity Analysis

8.1 Introduction

Chapter 7 presented the validation process of developed models. The validated model is then used to determine the sensitivity of the model outcomes to changes of its parameters. Such analysis is established by ranking the key parameters and understanding the effect of each key variable's magnitude on the strength of model outcomes. The aim of sensitivity analysis (SA) is to increase the confidence in the model and its estimates by understanding how the model outputs respond to alterations in the inputs. In addition, it is possible to explore the impact of changes on bus operations that might not be possible to conduct in the field.

This chapter begins with identifying critical variables by ranking the key parameters in the model. Then, parametric analysis, which measures the magnitude of each key variable on the response while all other variables are holding constant, is presented. This is followed by reverse sensitivity analysis. After that, simulation is conducted with various scenarios to explore possible improvements on bus journey time. Finally, this chapter is concluded with a summary.

Note that the validated model of the link-based Monte Carlo model is used in this chapter, and the field data which was collected on the Avenue route (inbound direction) and was used for independent validation in section 7.3 is used for analysis.

8.2 Identifying critical variables

The overall sensitivity of model response to a predictor is a combination of two factors, namely the model sensitivity and the predictor's uncertainty (Decisioneering, 2005). The model sensitivity is the effect that a change in a predictor produces in a response. Such effect is determined only by the formula of the model. This chapter focuses on this model sensitivity only, while readers can refer to Vose (2000) for details about predictor uncertainty.

There are a considerable number of input variables in the model. It is not possible to consider all of them without thinking about their relative importance. Due to the constraint on resources (time, money, etc.) in the real world, it is crucial to devote more resources to the key variables in order to arrive at the best estimates (Velez-Pareja, 2006).

8.2.1 Method

The Spearman's rank order correlation coefficient (ρ) is used to identify critical variables. This method uses the ranking of the data rather than the actual values, thus it is independent of the distribution of data sets (Vose, 2000). The value of ρ is between -1 and +1; a high ρ means that the input variable has a significant impact on the model output. Positive correlations indicate that an increase in input variable is associated with an increase in the model output and negative coefficients indicate the opposite situation. ρ is calculated as Equation 8.1.

$$\rho = 1 - \left(\frac{6\sum(\Delta R)^2}{n(n^2 - 1)} \right) \quad (8.1)$$

Where,
ΔR = the difference in the ranks between data values in the same pair; and
n = the number of data pairs

A method to help interpret the rank correlations provided by Crystal Ball is called "Contribution to Variance", which shows the percentage of the variance in the model output impacted by each input variable (Decisioneering, 2005). Note that such a percentage is only an approximation and is not a precise value of variance.

8.2.2 Results

The contribution to variance using above method is shown in Figure 8.1. Variables having more than 5% contribution towards bus journey time are presented on the figure. It can be seen that a total of six key variables had 67.4% impact on the output variations. The other variables contributed the remainder (32.6%). The number of bus-stops in the dwell time module ($N_{stopped}$) has the highest percentage (27.4%) and the same variable in the acceleration/deceleration delay module has the fourth ranking (8.1%). Thus, the total effect (35.5%) of this input variable on the bus journey time is remarkable. Such variability is derived from the passenger demand along the bus route. Passenger demand can be obtained through historical data, field survey, or assumed distribution. Therefore, understanding this variable more accurately can greatly help the estimate of the dwell time and acceleration/deceleration delay of the bus journey time.

The other 4 critical variables are all associated with signal delay at junctions. For example, the red period on N07311I link has the second rank (9.9%) of impact on journey time variance. It is found that this junction has the longest red time (86 s) along the bus route, and has the highest probability (0.82) of encountering a red light. As a result, this might be the reason that they contribute so much variation (0~86 s) to journey time variance.

Therefore, the number of bus-stops that a bus may serve along the bus route, and the critical junctions which may have longer control delay, are the key variables on the variance of bus journey time. For a good estimate, it is thus essential to have accurate inputs, especially for the critical variables.

8.3 Parametric analysis

This method measures the magnitude of each key variable on the response while all other variables are holding constant (Decisioneering, 2005).

The result of this analysis is shown in Figure 8.2 and Table 8.1. Only the top 10 variables are shown, and percentiles of 5% and 95% of each variable were used to calculate the estimates of upside and downside. It can be seen from Figure 8.2 that each variable illustrates the swing between the maximum and minimum journey time values. The median value of each variable is used to calculate the base case. The variable causing the largest swing which have the most effect on the outputs shows at the top and the variable causing the smallest swing which have the least effect appears at the bottom. For variables that have a positive effect on output, the upside of the variable (shown in black) is to the right of the base case and the downside of the variable (shown in gray) is to the left side of the base case. For variables which have inverse relationship with the output, the bars are reversed. The order of this figure is somewhat different (after the second) with the order in Figure 8.1 due to its assumption of holding all other variables constant. The number of bus-stops (both in DT module and D_{acc} module) is found to be the most important variable on bus journey time. From Table 8.1, the detailed values of input and output can be seen. The variable which caused the largest swing shows at the top row and the variable which caused the smallest swing appears at the bottom. Each row contains the downside, upside, and base case columns both in bus journey time estimates and inputs. For example, the journey time estimates ranged from 753 to 1025s when the inputs of number of bus-stops stopped (DT module) are 1 to 9.

The possible range of bus journey time caused by this variable is 272s. It is interesting that the same variable in the D_{acc} module with the same input value (1 to 9) produced a smaller range of journey times (73s; 853 to 926s). Therefore, it is useful to check the key variables to see what the effect on output is when the input changes.

In addition to the above variable, some signal timings of critical junctions are the key variables in this case, for instance, red period and probability of red signal on N07331I, N0311E and N04141E links. Again, these junctions had a longer red period and greater probability of red signal compared with other junctions, hence greater variances may occur. Also, the variable of boarding time for a passenger (T_{bi}) ranged from 4 to 34 s. As a result, it caused a variance of 30 s on journey time. The variable of bus speed (V_{bus}), which showed in the key variables, was due to the longest link (743 m), hence longer journey time. Therefore, the response may have greater variance if the input variables have a longer delay on bus journey time.

In fact, the assumption for removing other variables' effect on output is not practical in reality. This is called a major weakness of parametric analysis by Velez-Pareja (2006). In addition, this method depends significantly upon the particular base case (the median value was used in this study) used for the variable, that is, a different base case might have a different result (Decisioneering, 2005). Several trials of alternative base cases may be required to examine whether they have a significant difference.

8.4 Reverse sensitivity analysis

In the previous sections, the methods used were sensitivity analysis, in which input parameters were changed to calculate the corresponding change of output estimates. With reverse sensitivity analysis, the desired value of output is determined rather than the critical variable, in order to check the key parameters required to reach that desired value of output (Velez-Pareja, 2006). For instance, we may be interested in a particular long journey time (say 1200s in this case) and would like to understand the value of input variables required

An example of this analysis for a bus journey time between 1200 and 1300s is shown in Table 8.2. It can be seen that the particular journey time is presented on the header row and each row shows an input parameter. D1, D2....D11 (Di) shows the order of deceleration rate for 1st, 2nd...and 11th bus-stops as well as Ai, Nai, Nbi, Tai, and Tbi. For example, when the bus journey time was 1201.93s, the required input parameters were D1=1.52, A1=1.20,..., $N_{stopped}$=8,..., and so on. According to the critical variables described in the above sections, attraction focuses on the key inputs. When the bus journey times were between 1200 and 1300s, the number of bus-stops which a bus requires stopping were more than 8 (from a total of 14 bus stops along the bus route), and need to stop at least 2 junctions from the 3 critical junctions (N07331I, N03111E, and N04141E links) due to red light. Thus, using this method, the effect of the combination of input values to a particular response can be identified.

8.5 Scenario analysis

8.5.1 Effect of passenger demand

The data of the field study case showed that the passenger demand is very low. A total of 200 dwells in 39 runs were collected in the 3 days' survey. In average, only 4 passengers boarded and 4 passengers alighted each bus, and each bus stopped at 5 bus-stops from a total of 14 bus-stops along the study route. Each dwell-period at a bus-stop was for less than 2 passengers. The major effect of low passenger demand is that the dwell time and acceleration/deceleration delay due to dwells are less, hence shorter bus journey time.

According to sections 8.2 and 8.3, the most critical input variable is the number of bus-stops that a bus served. Thus, if an increase of passenger demand causes an increase of bus-stops stopped and no other variables changes, what is the possible change in bus journey time? In addition, if the increase of passenger demand causes the increase of boarding and alighting passengers and the other variables hold constant, what is the probable alteration? The increase of passenger demand might be due to the improvement of the bus service such as reliability, punctuality, availability and accessibility, or limitations to private cars such as congestion charging and/or park-and-ride scheme, which attract more passengers to the bus.

8.5.1.1 Scenario description

Two scenarios were conducted to compare the base case (with no changes). Details of these two scenarios are described as follows:

(1) Increase of bus-stops stopped

An assumption is made that the number of bus-stops which a bus stopped doubles in the mean and follows a normal distribution ($N(\mu,\sigma^2) = N(10, 2.25)$). Then, each value is rounded down to an integer to match the character of discrete distribution.

(2) Increase of boarding and alighting passengers

An assumption is made that the number of boarding and alighting passengers doubles, both in the mean and maximum values, and follows a lognormal distribution. In other words, the alighting passengers of each stop has mean=2, standard deviation=2.2, and values between 0 and 12; the boarding passengers of each stop has mean=2, standard deviation=1.6, and values between 0 and 8. Then, each value is rounded down to an integer to match the character of discrete distribution. The selection of lognormal distribution is due to such distribution is better fit of the field data.

8.5.1.2 Simulation result

The comparison results of the above 2 scenarios, along with base case, is shown in Table 8.3. It can be seen that the doubling in the number of bus-stops stopped (scenario (1)) caused increases both in dwell time and acceleration/deceleration delay. The average increase in dwell time was 93% and the acceleration/deceleration delay was 91%. The total effect on bus journey time was 14%. In the increased alighting and boarding passengers scenario (2), only the dwell time increased by 143%, whereas the effect on bus journey time was the same as scenario (1) (14%). There was no change in signal delay in both scenarios.

Detailed profiles containing the distributions and probabilities of the above scenarios, along with the base case, are shown in Figures 8.3 to 8.5 for dwell time, acceleration/deceleration delay, and bus journey time respectively. It can be seen from Figure 8.3 that the spread and range of dwell time changed significantly according to the scenarios. With the increase of passengers in scenario (2), its distribution was flat and its right tail was extended to more than 400s with 5% probability. Thus, the effect on the longer dwell time is quite significant. From Figure 8.4 it can be seen that the distribution of scenario (1) was 'pointy'. This was thought to be due to the assumption of a normal distribution, which was different from the historical data shape (with higher probability on its left tail) in the base case. According to Figure 8.5, although the effects on bus journey time mean value were similar (14%) in both scenarios, their spread and range were quite different. The distribution of scenario (2) had a wider range, with two longer tails, hence a flat spread. However, their most probable journey times were both between 800 and 1000s.

8.5.2 Effect of signal timing change

The signalized control delay in the field-study case was about 13% of bus journey time (104s), with a total of 4 junctions and 7 pelican crossings along the bus route. The maximum possible delay of a junction was 86s at N07331I link. It is clear that a bus might have a 0 control delay when no red was encountered. Conversely, a total of 400 s control delay might be encountered if a bus encountered red at all junctions and pelican crossings for the entire red period. Actually, the simulation result showed that the 0 delay was probably happening; however, there was no control delay greater than 300 s obtained in simulation. It was thought to be that some of the signals might coordinate together to achieve progressing for vehicles, hence little chance to have continuous red at adjacent junctions.

According to sections 8.2 and 8.3, the second most critical input variable was the red period of N7331I link. Thus, the concern is the possibility of reducing the control delay of this junction. It was observed that this link was very congested during peak hours. It is supposed that the authority should attempt to adjust the signal timing by increasing green time in order to alleviate the jam. In such a situation, it is interesting to explore the probable impact on bus journey time.

8.5.2.1 Scenario description

The details of signal staging and timing of the intended junction (N7331I) are shown in Table 8.4. The target stage is stage 2, which is on London Road towards the city centre. From Table 8.4 it can be calculated that the original setting of green time of the target stage is 19.03s and cycle length is 105.36s (total length of stage 1, 2, and 3). The assumption is made that the green time of the target stage is increased by 20s, i.e. total green time is 39.03s and cycle time is 125.36s, and other settings hold constant. This scenario is intended to explore the effect on bus journey time due to this change.

8.5.2.2 Simulation result

The comparison result between the above scenario and the base case is shown in Table 8.5. It can be seen that the increase of green time of the critical junction caused a small decrease in signal delay. The average decrease in signal delay was 5s (5%). The effect on bus journey time was 0.64%. The difference of distribution of these two data sets was negligible, and there was no change in dwell time nor acceleration/deceleration delay.

8.5.3 Effect of change in boarding time

Bus-stop service is the mission of bus operation. Total time of such service could contribute up to 26% towards bus journey time (Levinson, 1983). In this study, dwell time of the study case was 10% due to low passenger demand, as described in section 8.5.1. Thus, effective and efficient management of such time is essential for bus operation.

The comparison of dwell time components of this study and the literature are presented on Table 8.6, which is determined as Equation 8.1 (similar to Equation 5.5). Note that the dwell time used in the regression model of this study exclude extreme events, as described in section 5.3, whereas all dwell time data were included in Monte Carlo model.

$$T_D = a + b \times P_A + c \times P_B \qquad (8.1)$$

Where,

T_D = bus dwell time per stop (s);
P_A = the number of alighting passengers;
P_B = the number of boarding passengers;
a = door opening and closing time (s);
b = passenger alighting time (s/p); and
c = passenger boarding time (s/p).

It can be seen from this table that the values of door opening and closing time (a) and passenger alighting time (b) in the regression model of this study are similar to previous studies. However, passenger boarding time (c=8.88 s) is significantly greater than in the literature except York (1993) and Shrestha (2002), which studies were both in the UK. In addition, the mean values of c in the Monte Carlo model (14s) are even greater than the regression model's. Therefore, it is interesting to explore the possible reasons for such higher boarding time for the UK bus services, in order to improve bus service.

There are five main factors which may affect dwell time (TCRP, 2003a). Among them fare collection methods, vehicle types, and in-vehicle circulation are thought to be the major influences. The average time required to pay a fare is the critical factor. Different types of fare collection methods may result in diverse boarding times (Guenthner and Hamat, 1988; Marshall et al., 1990). Vehicle types consider how many doors are available for alighting and boarding, whether a bus is low-floor for easy access and whether a bus is double-decker, affecting dwell time. In-vehicle circulation means that it may take much time to enter the bus when standees occupy the fare collecting area. Note that the loading and unloading of wheelchairs could heavily impact the dwell time and result in a significantly longer dwell time. However, this activity did not happen in the survey, and hence it is negligible in the following.

An on-board survey was conducted to understand the probable reasons of longer boarding time on the Bitterne route (Figure 4.2) on 15 September 2005, between 11:00~18:00. The recording sheet is shown as Table 8.7 and an example of this collected data is illustrated in Table 8.8. A total of 22 bus runs were surveyed, including 174 dwells with 255 alighting and 180 boarding passengers. Among these dwells, 11 had extra waiting time with no passenger activity and were excluded. It was found during the survey that about half of the buses had two doors available, though they didn't use the rear door for passenger services, and no standing occurred, hence these two factors were discarded. As a result, a statistical test was carried out to examine the following 2 cases: whether dwell times were similar under different bus types (low-floor/ non-low-floor, double-decker/ non-double-decker) and whether boarding times were similar under various fare collection methods.

With the 1st case, the impacts on dwell time of these two factors were not significant (low-floor/ non-low-floor: p-value=0.28 > α; double-decker/ non-double-decker: p-value=0.82 > α), hence no significant differences were concluded. With the 2nd case, particular attention is paid on the boarding passengers in order to simplify the scenario, and 3 fare collection methods, namely pre-paid ticket (swift card, pass), cash without change, and cash with change, were examined. The descriptive statistics of these 3 groups is shown in Table 8.9. It can be seen that the mean boarding time of a passenger with a pre-paid ticket (7.82s) was less than the cash-without-change case (14.65s) and the cash-with-change case (16.13s) took the greatest time to complete. These results were supported by the one-way analysis of variance (ANOVA), that they were not all the same (F=5.46, p-value=0.01 < α). Therefore, it can be found according to this result that the boarding time per passenger of the study route is significantly greater than the value suggested for the literature in Table 8.6, for either the pre-paid case or cash dealings. Thus, any improvement which can help reduce such time such as prepayment, is essential in order to have a positive impact on the quality of bus service (TCRP, 2003a). Readers could refer to TCRP (2003c) for details of fare collection methods.

8.5.3.1 Scenario description

An assumption is made that the boarding time per passenger of the study route has been improved due to effective fare collection methods being adopted, such as encouragement of prepayment. Suppose that the average boarding time is reduced to 5s (origin mean value =14s) with standard deviation of 10s (origin S.D.= 10.73s) and follows a lognormal distribution as used in the simulation models. Such distribution of boarding time was also indicated by Guenthner and Hamat (1988). Other variables are held constant. The aim of this scenario is to explore the effect on bus journey time due to the decrease of boarding time.

8.5.3.2 Simulation result

The comparison result of the above scenario along with base case is shown in Table 8.10. It can be seen that the change in the boarding time per passenger causes a very dramatic decrease in dwell time. The average decrease in dwell time is 44%. The total effect on the reduction of bus journey time is 4%. There is no change in signal delay and acceleration/deceleration delay in this scenario. A detailed profile containing the distribution and probability of the above scenario along with base case is shown in Figure 8.6. It can be seen that the spread of dwell time changed significantly according to the scenario. With the decrease of boarding time, its distribution is 'pointy' and has a pile-up on the left.

8.5.4 Effect of bus priority

There is no bus priority along the study case route at the moment. However, this might be changed if the traffic condition gets worse and the demand of the bus service increases. If this is the case, it is essential to identify the range of problems and opportunities by reviewing bus services. The probable strategies of bus priority constitute 3 groups (Fernández, 1999): 1) link priority: to reduce the journey time of bus movement by segregating buses from the traffic or by limiting vehicles entering a particular area; 2) junction priority: to decrease the delays of signalized junctions by adjusting signal settings; 3) stop priority: to reduce dwell time by consolidating bus stops or improving bus stops' layout. Suppose that the local authority has the intention of improving the bus service by reducing journey time on the study route. When checking the available data, it was found that bus general travel time (excluding dwell time, signal control delay, and acceleration/deceleration delay) ranged to more than 70% of bus journey time. Thus, particular attention is paid on the critical section of road for link priority.

The method described in section 8.2 is used for identifying the critical link. The contribution of each link to the variance of bus general travel time is shown in Figure 8.7. As the length of each link is fixed, the variation of link travel times derives from the fluctuation of speeds. Thus, the variables shown in Figure 8.7 are the speeds of particular links. It can be seen in Figure 8.7 that the speed variable V_{bus} contributed most (36.4%) of the general travel time variance, then N07331I-STs was the second which contributed 12.8%, and so on. Since bus speed in the link-based simulation model of this study is subject to link length and SCOOT speed on that link, as described in Equation 6.27 in sections 6.3.3.2 and 6.3.3.3. According this equation, a long link length and high SCOOT speed could result in a high bus speed and therefore result in a short link journey time. The details of each link, including the percentage of contribution to variance from Figure 8.7, link length, and the mean value of SCOOT speed of each link are shown in Table 8.11. It can be seen that for the Chilworth link(speed variable= V_{bus}), which contributed 36.4% to general travel time variance, link length is 743 m and the mean speed of SCOOT was 45 kph. Note that this link, which is far away from city centre, does not have a SCOOT link and the traffic condition was free flowing most of the time according to field observation, hence the assumption of bus mean speed of 45 kph is made. Therefore, it is clear that such a link is not a critical link for the priority purpose. Then, the second-most contribution towards general travel time of 12.8% for N07331I link (London Road towards city centre) was checked. This link is located in the city centre and is a very busy commercial road, with roadside parking on both sides and a high frequency of buses passing through it. Its length is 351 m. The traffic condition in the peak hours is very congested, hence the mean speed of SCOOT was only 27 kph which is significantly lower than other links. Thus, these characteristics of this link may match the requirement for link priority.

Several bus priority strategies are available for selection (TCRP, 2003a; DETR, 2001). However, there is no best measure with best performance for any circumstance. The most appropriate measure in any location will depend upon the local situation in that area and utilizes either one or a combination of the priority measures (DfT, 2003b). Suppose that an improvement scheme is made for the above link (N07331I), which prohibited all vehicles entering except buses, taxis, and bikes. Actually, such scheme is similar to the following link (Above Bar Street), which has been in place since 2005. As a result, it could help to alleviate the congestion and increase the bus speed, hence shortening journey time of this link.

8.5.4.1 Scenario description

An assumption is made that the average bus speeds on link N07331I have been improved due to above scheme of link priority. Suppose that the average bus mean speed is increased to 40 kph (origin mean value =27 kph) with standard deviation of 5 kph (origin SD= 4.45 kph) and follows a normal distribution. Other variables are held constant. The aim of this scenario is to explore the effect on bus journey time due to the scheme of link priority for buses.

8.5.4.2 Simulation result

The comparison result of the above scenario along with base case is shown in Table 8.12. It can be seen that the change in the acceleration/deceleration delay causes a slight increase, since the higher speeds of the bus requires a little bit of time for deceleration and acceleration. The average increase in such delay is 3% (1.63s). However, the total effect on the reduction of bus journey time is 2% (15s). Such a decrease could great help buses through the intended link (N07331I). There was no change in signal delay and dwell time in this scenario.

8.6 Summary

This chapter has described the sensitivity of the model outcomes to changes of its parameters. The link-based Monte Carlo model was used to conduct the simulation on the Avenue inbound bus route. The sensitivity analysis started by ranking the key parameters in the model to identify critical variables. Next, parametric analysis, which measures the effect of each dependent variable on the independent variable while removing the effects of other dependent variables, was presented. Then, reverse sensitivity analysis was illustrated, which enables us to check the key input parameters that are required to reach the desired value of output. Afterward, simulations were conducted under various scenarios such as an increase in passenger demand, change in signal timing, reduction in boarding time, and establishment of a bus priority scheme, to explore the possible impacts or improvements on bus journey time. The results from these simulations were compared to the base case of field data. The overall discussion of this study is described in Chapter 9.

Chapter 9: Discussion

9.1 Introduction

Chapter 7 presented the validation process of developed models. The validated model was then used to determine the sensitivity of the model outcomes to changes of its parameters in Chapter 8. The aim of this chapter is to better understand the underlying assumption, limitation, and potential improvement of this study which were not discussed in previous chapters or which required additional explanations.

This chapter generally follows the work flow on Figure 1.1. It begins with the understanding of the accuracy of bus journey time data in this study. Then, the difficulty in data collection and processing are described. This is followed by a discussing of the possible factors in bus journey time components. Next, a further discussion relates to the underlying assumptions, limitations, and potential approaches of model development. After that, arguments are made with reference to the validation process and potential ways to refine invalid models. Finally, a brief conclusion finishes this chapter.

9.2 Accuracy of bus journey time data

Both GPS and on-board survey using recording sheets were used in this study, which was illustrated in section 4.4.1. The accuracy of GPS for the journey time survey was described in chapter 3, while, the possible error of recording time which resulted from the surveyors, though it might be diverse, will not be discussed here. However, the journey time collected from these two approaches can be checked against each other to avoid error. Following are the discussion of accuracy with regards to bus journey time as well as dwell time, acceleration/deceleration rate, and the road facility location which was used to pinpoint the location of GPS points.

9.2.1 Bus journey time

As described in section 2.2.1, journey time is the time required to traverse a route (or link) between two successive points such as departure and arrival checkpoints. Thus, with GPS data, journey time is calculated as the time difference between two such points. The possible error which may occur in identifying the two points, is shown in Figure 9.1. It can be calculated from Figure 9.1 that actual journey time is T2-T1, however, the GPS points might not be the exact position of these two points and hence the journey time is replaced by t2-t1. Thus, the journey time error is calculated as (T2-T1)-(t2-t1) or (t1-T1)+(t2-T2). With a one second updating frequency of GPS data, the possible maximum error between T1 and t1 (or T2 and t2) is less than 1 second. Consequently, the maximum error of a journey time calculation is less than 2 seconds. It is clear that if the journey time is quite long, e.g. minimum 510 s for a minimum 3.2 km bus route in this study, the error can be negligible. However, if the journey time is very short, e.g. 6 s for a minimum 85 m link length, the effect on the link journey time is significant. Nonetheless, the error on the aggregate journey time is also insignificant. Note that the possible error of GPS points, which was presented in Chapter 3, is not taken into account in above discussion.

9.2.2 Dwell time

As described in section 2.7.1.1, dwell time is the time when a bus stops to enable passengers to alight and board at a bus stop. Thus, with GPS data, dwell time is calculated as the time difference between two successive points of which one is the start stop after deceleration and the other is the end stop just before acceleration. However, the problem is that bus GPS points were not always at the same point during the stop period due to the position error which was described in section 3.4. Furthermore, it is not easy to identify the exact points of those two according to the changes of speeds associated with the points' sequence of GPS data. Therefore, referring the dwell time data recorded by surveyors was necessary in order to reduce such error.

9.2.3 Acceleration/deceleration rate

The estimation of average acceleration/deceleration rate (A/D) for bus-stops was described in section 5.4 which used Equation 9.1. When a bus is accelerating (A), V_1=0 and t_1 is the time before acceleration; V_2 is the bus cruise speed after acceleration and t_2 is the time when it reached the speed V_2. When a bus is decelerating (D), V_1 is the bus cruise speed before deceleration and t_1 is the time of speed V_1; V_2 =0 and t_2 is the time when the bus stopped at a bus-stop. Thus, the accuracy of A/D is affected by the speed reading of V_1 (or V_2) and the identification of t_1 and t_2 as well as the time gap between t_1 and t_2.

$$A/D = \frac{V_2 - V_1}{t_2 - t_1} \tag{9.1}$$

The speed accuracy of GPS was described in section 3.4.3, which was less than +/-4 mph (+/-1.78 m/s). The possible error of identification of t_1 and t_2 may be assumed to be similar to the journey time, which was described above, as a maximum 2 seconds. The time gap between t_1 and t_2 ranged from 1 to 37 s with a mean of 7.68 s in this study. Therefore, the possible error of A/D in this study can be roughly calculated as following of ± 0.31 m/s/s.

$$\text{Possible error of A/D} = \frac{Max.\ speed\ error}{(Time\ gap\ of\ t_1\ and\ t_2) - (Max.\ identificaton\ error\ of\ t_1\ and\ t_2)}$$

$$= \frac{\pm 1.78}{7.68 - 2} = \pm 0.31\ \text{(m/s/s)}$$

9.2.4 Road facility location

As described in section 4.3.2, a compact and wrist-type GPS of Garmin Foretrex 201 was used to locate the position of bus-stops, junctions, inductive loop detectors, stop lines, pedestrian crossings, and ANPR cameras. The position accuracy of such GPS is the same as that of Garmin 35 PC, while the Garmin 35 PC GPS was used in chapter 3 for understanding the position accuracy of this study. Thus, the position accuracy, which was illustrated in section 3.4.3 as within 4.15 m with 95% confidence, may be applied to this part.

9.3 Data collection

9.3.1 With regard to transit performance measures

There were 3 levels of analysis used in this study, namely route level, link level, and point level, in which physical facilities on road such as inductive loop detectors and dwells at bus-stops were used to separate collected data. In route level analysis, only part of bus routes was used in the field survey. The route lengths and locations were detailed and presented in Table 4.5 and Figure 4.2. In fact, if the data is available, it would be better to include the whole bus route from departure to the end of bus service.

Such whole route analysis might be more comprehensive than that used in this study for route level analysis. In link level analysis, the consideration of affecting factors must be more detailed and specific. For example, in regression model development, the control delay of signalized junctions was merely an independent variable (the number of signalized junctions) in the route-based model, whereas, it was an independent formulation in the link-based model. This is because signal delay contributes significantly to the link journey time variation, which was described in section 6.2.3.3. Dwell time and schedule adherence were used in point level analysis. Readers may refer to Bertini and El-Geneidy (2003) for details of transit performance measures in terms of system, route, segment, and point level.

9.3.2 Accessibility of data source

There were three groups of data collected for this study, namely bus data, general traffic data, and road geometry and facilities data, which were described in section 4.4. Among them, only ANPR and SCOOT data were accessed from the ROMANSE office. Other data required collection from the field, which was constricted by the scale and duration due to labour-intensive survey. Consequently, it is crucial to access a reliable and considerable data source such as bus GPS data and automatic passenger counting (APC), which might be available in the near future, rather than limited field survey.

9.3.3 Data processing

There were detailed descriptions of the procedure for integrating GPS and GIS as described in Appendix A. The aim of displaying GPS points on a digital map showed on GIS software such as ArcView can be achieved via these steps. However, considerable manual processes were still in existence in the above process. For instance, it was necessary to identify a specific position for bus GPS data such as the start and the end point of each bus run of GPS data for each route for route-based analysis or each link for link-based analysis. In addition, in order to achieve acceleration/deceleration data around bus-stops, considerable time was spent on observing change of speeds against the leaving/approaching bus-stop and separating such data from other sites such as signalized junction or congested situation which also had the situation of acceleration/deceleration. Such manual approaches may be manageable with only a few routes and several dozens of observations. However, if a large quantity of data is available, then some automatic processing of data such as designing a program must be carried out in order to speed the data processing and reduce the possibility of human error within it.

9.4 Bus journey time components and comparison with other vehicles

9.4.1 Key components

There are four major components of bus journey time, namely general travel time, dwell time at bus-stops, delay of deceleration and acceleration for bus-stops, and control delay at signalized junctions, which were described in chapter 5. Considerable variables were taken into account for modelling, which were detailed and illustrated in chapter 6. The independent variables which were used in each result model are summarized in Table 9.1. In general, route (or link) length, traffic condition in terms of other vehicles' speed and bus priority such as bus lanes may affect general travel time. The number of stopped bus-stops, bus speed, and acceleration/deceleration rate may influence the delay due to deceleration and acceleration for bus-stops. The control delay on signalized junctions may be affected by the number of signalized junctions and cycle length and the green time of each junction. Dwell time at bus-stops is determined by the number of stopped bus-stops, the number of alighting and boarding passengers, and the alighting and boarding time of each passenger. These considered variables seem to reflect the variability of bus journey time. However, some factors including road layout such as the number of lanes, roundabout, and roadside parking, the type of bus stop, time series such as the day of the week, nonrecurring incidents such as road maintenance and accidents, and so on, might influence journey time but are not taken into account in this study owing to data constraints or the limited scale of the test bed.

9.4.2 Relationship of journey time of buses in comparison with other vehicles

As described in section 6.3.2.2, the percentage of ANPR journey time (bus journey time /ANPR journey time) was modelled in Monte Carlo models, which bus journey time excluded dwell time and acceleration/deceleration delay due to dwells. The percentage which was illustrated in Appendix C.1 ranged from 0.81 to 1.78 with a mean value of 1.34. This value can therefore be used to compare previous studies which were described in section 2.6.2. The result is shown in Table 9.2. It can be seen that the percentage value of this study is similar to previous studies (1.37~1.75) but slightly lower. This is due to the fact that bus journey time of this study subtracted acceleration/deceleration delay which was not taken into account in other studies hence a slightly lower value was obtained.

9.5 Model development

9.5.1 General issues (independent of models)

9.5.1.1 Time series

Clearly, bus journey time is determined by traffic conditions and passenger demand which data are both series of observations ordered in time, i.e. it is time series data. Such factors were not taken into account in this study because the aim of this research is to understand the overall journey time estimation which focus was not put on the estimation of specific time or prediction on any time scale (short-time or long term). However, it might be interesting to understand the probability distribution against not only the overall journey time estimation but also the time series along a bus route, which can be shown as a three dimensional diagram in Figure 9.2.

9.5.1.2 Bus holding

Holding involves an early bus waiting at a bus-stop for a period of time, in order to match the timetable. It has the advantages of maintaining adherence to the bus timetable and minimizing the waiting time for the downstream at-stop passengers. By contrast, it also has the disadvantages of increasing in-vehicle delay and resulting in the idle usage of the bus. Thus, some studies addressed this trade-off problem (Barnett, 1974; Fu and Yang, 2002; Sun and Hickman, 2004). Due to the considerable variation of traffic and passenger demand, it is more difficult for operations to maintain the bus timetable. Furthermore, a timetable which is not continually updated and improved may also result in similar problems.

In this study, the factor of such a strategy was not modelled due to the difficulty in separating the holding time from the dwell time of collected data. Only some description regarding the effect of schedule adherence on bus journey time (section 5.5.2) was discussed. Actually, such strategy was observed in the field survey. About 24% of bus runs had additional waiting time with no passenger activities, which were marked on the recording sheets. Although these extra waiting times may not be all for the holding, they indeed contribute considerable variation to dwell time as well as bus journey time. Thus, taking into account this factor might be an approach to refining models. Such consideration is discussed in section 9.6.2.

9.5.1.3 Alternative data manipulation

As indicated on Figure 4.1, three of the total four bus routes of field data (route 1, 2, and 3) were used for model development and the remaining one route data (route 4) was used for independent validation. In fact, there are four possible combinations for formulating a model in the collected four routes, namely (1, 2, 3), (1, 2, 4), (1, 3, 4), and (2, 3, 4), and the result of each model might be similar or diverse. Due to the considerable work involved in data processing and categorizing, especially GPS data, alternative combinations were not carried out. However, it might be interesting to conduct a comparison among alternative models from different combinations.

The above approach is one of the subset strategies (split-sample or hold-out method) for conducting the modelling process. The other strategy, which could use all four routes data but randomly select half (or a specific percentage) of the data for modelling and the remaining half for validation, might also be applicable. In addition to this subset strategy, which excludes the validation data for model developing, alternative re-sampling strategy such as cross-validation or the bootstrapping method might be applicable. Readers may refer to Shao (1993), Shao and Tu (1995), Zhu and Rohwer (1996), Kohavi (2004) for more details.

9.5.2 Regression model

9.5.2.1 Sample size

If required data cannot be accessed from the traffic database and may be expensive for field survey, it is necessary to consider the cost-effectiveness of the necessary sample size for an acceptable regression model. Thus, it is essential to understand the minimum required sample size for the intended model before gathering data. Considerable literature has discussed the sample size requirement for travel time studies (Robertson et al., 1994; Quiroga and Bullock, 1998b; Turner et al., 1998). Basically, there are two types of formulations used for estimating the minimum sample size, namely sample ranges and sample standard deviations, which are referred to by Quiroga and Bullock (1998b). However, such approaches essentially require substantial set of initial runs to obtain reliable estimates for mean (μ) and standard deviation (σ), then these estimates are used to calculate the minimum sample size using the above formulations. Thus, a rule of thumb which was introduced by Kleinbaum et al. (1998) might be quite useful in circumstances where such trials are neither possible nor practical. Such an approach is based on the minimum requirement of approximately 10 error degrees of freedom in regression model, i.e. the minimum sample size is determined as Equation 9.2.

$$n - k - 1 \geq 10 \tag{9.2}$$

or $n \geq k + 11$

where, n is the number of observations and k is the number of independent variables in the regression model. Alternatively, a rule of thumb was suggested for regression to have 5 observations per independent variables, namely n is determined as Equation 9.3.

$$n \geq 5k \tag{9.3}$$

9.5.2.2 Additional considerations in the modelling process

- Treating outliers

There may be some debate on excluding the effect of extreme observations of dwell time using the approach by Zhao and Li (2005), which was described in sections 5.3.2 and 5.3.3. Apparently, the presence of such extreme values can significantly affect the fitting of least-squares regression model and result in the model failing to capture important features of the data (Fox, 1997; Kleinbaum et al., 1998). However, it is not necessary to discard those observations unless they were owing to mistakes. In some cases, they may imply particular circumstances that require additional investigation. There are several residual statistics such as standardized, studentized, and jackknife residuals which can be used to detect outliers and some methods such as leverage, Cook's distance can be used to measure the influences of extreme observations on the model. Readers can refer to Field (2005) for more details about treating outliers.

- Collinearity

The high relationship between explanatory variables could cause a regression model with inaccurate coefficients as described by Kleinbaum et al. (1998), "*the estimated regression coefficients in the best model may be highly unstable(i.e. high variance) and may be quite far from the true parameter values*". Some independent variables such as the number of signalized junctions per km (N_{sig}) and the number of disturbances per km (N_{dist}), and SCOOT parameters (ST_s, ST_o, ST_f) are very likely to have collinearity as described in sections 6.2.2.3 and 6.2.3.2. Thus, consideration must be given to the selection of appropriate independent variables involved to avoid the above problem. Additionally, there are several approaches such as centering, variance inflation factor (VIF), eigenvalues, etc., which can help decrease or diagnose collinearity (Fox, 1997; Kleinbaum et al., 1998).

- Selection for preferred model

It is better to specify the probable preferred models which include the most important independent variables before conducting regression analysis as flow chart of Figure 6.1 in section 6.2.1. This is because the best model which showed the best value of criterion such as R-square might not be the preferred model most of the time. Thus, the process for selecting an appropriate model not only manipulates the mechanisms of statistics but also requires professional understanding of the intended purposes in terms of which of the variables should be in the model. Especially, when several alternative regression models are available, having similar performance of criterion (say R-square value) and all are accepted (the acceptance of model F test and coefficients of t test), the selection derived from such a view is substantial.

In addition, particular consideration must be given to R-square value. It is clear that such value is a basic indicator for deciding which model is best. However, this value might have 3 possible disadvantages which were indicated by Kleinbaum et al. (1998) as follows. Therefore, it is not suggested that this criterion solely should be used, but that it should be accompanied by the criteria proposed in section 6.2.1.

> *"R-square has three potentially misleading characteristics. First, it tends to overestimate the corresponding population value. Second, adding predictors, even useless ones, can never decrease R-square. In fact, adding variables invariably increase R-square, at least slightly. Finally, R-square is always largest for the maximum model, even though a better model may be obtained by deleting some (or even many) variables. The reduced model may be better because it may sacrifice only a negligible amount of predictive strength while substantially simplifying the model."*

9.5.3 The Monte Carlo model

9.5.3.1 Selecting input probability distributions

- Data sources and alternative approaches

There are basically two sources of information which can be used to quantify the uncertainty of input variables within the Monte Carlo model. One is available data and the other is expert opinion (Vose, 2000). If the intended data has never been collected in the past or if it is expensive to achieve, the latter source may be an alternative. If it is possible to collect data whether it is from the field or from literature, it can be used in one of the following approaches to identify input probability distributions. Detailed comparisons of (1), (2), and (3) were discussed by Law and Kelton (1991):

(1) The field data values themselves are used directly, e.g. variables of the number of stopped bus-stops ($N_{stopped}$), number of boarding and alighting passengers (N_{bi}, N_{ai}) in dwell time module (section 7.3.1); or

(2) The field data values themselves are used to fit an empirical distribution function such as regression model; or

(3) The field data is used to fit a theoretical distribution function, e.g. the percentage of ANPR journey time per km ($ANPR_{\%}$) in general travel time module (section 6.3.2.2); or

(4) Taken from literature.

- Discrete or continuous variable?

"*A variable that is discrete in nature is usually, but not always, best fitted to a discrete distribution.*" (Vose, 2000). In some circumstance, discrete distribution might be approximated very similarly by continuous distribution. If this is the case, the discrete distribution can be modelled by continuous distribution for convenience, and its separate character can be converted back by rounding the decimals. This approach can be found in the variable of the number of stopped bus-stops ($N_{stopped}$) in dwell time module of section 7.2.3.1.

- How representative is the fitted data?

Many heuristic procedures and goodness-of-fit statistics may be available (Law and Kelton, 1991; Vose, 2000), but Chi-square (χ^2) and Kolmogorov-Smirnoff (K-S) are the most commonly used. In addition to these two tests, the Anderson-Darling (A-D) statistic, which is a sophisticated version of the K-S test that weights the differences between the two distributions at their tails greater than at their mid-ranges (Decisioneering, 2005), is also used in this study. Such test results are shown in the description of input variables, e.g. the statistics column of Appendix C.1.

9.5.3.2 Interpretation of simulation outcome

The interpretation of simulation outcome is not easy. The answer to a question cannot be easily provided by its result as described by Rubinstein (1981) "*Simulation is an imprecise technique. It provides only statistical estimates rather than exact results, and it only compares alternatives rather then generating the optimal one*". The outcome of simulation is nothing but the distribution which has possible values with probabilities. The probable answer to a question requires professionals to interpret according to the underlying resources or conditions of the question. In addition, when interpreting the outcome from simulation, special consideration must be given to the validity of the model and the underlying assumptions which were made (Robinson, 2004).

9.6 Model validation

9.6.1 Selection of significance level (α)

Hypothesis test was used in model validation for the comparison of means, variances, and distributions to determine whether the models' output has an acceptable range of accuracy. There are two types of possible error in testing hypothesis. The first, type I error, is rejecting the null hypothesis H_0 when it is true and the second, type II error, is failing to reject the null hypothesis when it is false. The probability of the type I error, α, is called the model builder's risk, and the probability of the type II error, β, is called the model user's risk (Sargent, 1998). These two errors are related. When the sample size is constant, a decrease in the probability of one type of error always results in an increase of the other. The only way to reduce both errors is to increase the sample size (Montgomery and Runger, 2003). With the purpose of journey time estimation of this study, the application of the model is very important. It is better to keep the model user's risk as small as possible. Therefore, consideration must be given to both type I and type II errors when using the hypothesis test for model validation.

9.6.2 Potential methods to improve regression model

It was concluded in chapter 7 (model validation) that both regression models were invalid through the process of validation. However, such a result might be due to some variation of key variables which cannot be explained in the proposed models. Thus, a potential way to improve the above regressions is discussed as follows.

- Refinement of link-based regression model

In section 7.2.2.2, an adjusted model which was considered the variation of signal delay on signalized junctions was suggested for the link-based regression model. Such an adjusted model enables us to estimate the possible journey time of the two ends, namely upper and lower bounds; however, it was unable to provide any estimates in between. Actually, the crucial problem in the link-based regression model is that the estimates failed to reflect the longer journey time as shown on Figure 7.7. Among the journey time components, general travel time (JT_l) contributed the most (about 70%) to the total journey time. Such general travel time was the total sum of link general travel time (JT_g), which is used in the Equation 6.8 with link length (L_{link}) and the SCOOT flow parameter (ST_f) independent variables. It is clear that each link length is fixed, hence all variations rely upon the fluctuation of flow derived from the SCOOT parameter. Since the average SCOOT flow parameter of each link was used in this study, i.e. if the probable bus journey time is 20 minutes (say 13:27~13:47), then an average flow of SCOOT UO7 message with 5 period (13:25~13:30, 13:30~13:35, 13:35~13:40, 13:40~13:45, 13:45~13:50) was calculated for the input value. This was due to the difficulty of identifying the passing time of each bus on a specific detector. Therefore, such flow value was averaged twice. The first was averaged to 5 minutes for SCOOT UO7 message and the second was averaged for a bus journey time period. Consequently, it is difficult for such a parameter to represent the variation of traffic. Nonetheless, it is possible to locate a bus when passing each detector with GPS and hence the exactly passing time can be determined. Such approach may require additional GPS data processing which is associated with GIS to identify when and where a bus passes a specific detector. In addition, a small period of SCOOT message such as UO6 (30 s) is suggested to reduce the negative effect of the average.

- Refinement of route-based regression model

For the route-based regression model, the ANPR journey time model was valid in the validation, but its sub-model of dwell time failed due to underestimation with longer dwell time. The SCOOT journey time model also showed invalid owing to the underestimation of the greater journey time, which was concluded in section 7.2.2.1. It is likely that their sub-model of dwell time may be the crucial factor for underestimation of bus journey time. Thus, a refined dwell time model is required to improve route-based regression models. Bus holding and treating outliers were discussed in above sections (sections 9.5.1.2 and 9.5.2.2). Such factors may actually affect dwell time in the real world. Therefore, a revised route dwell time model (DT) shown as Equation 9.4, which is derived from the original model (T_D, Equation 5.5) and adds a dummy variable of additional waiting time at bus-stops ($D_{waiting}$), is recommended. Note that the data used in this regression model is the same as the original model, whereas, they are based on route dwell time rather than for each stop. The additional waiting time involves no passenger activities but may not be exactly the bus holding strategy.

$$DT = 11.03 + \sum_{i=1}^{N_{stopped}} T_{Di}(Equation\ 5.5) + 48.6 * D_{waiting} \quad , R^2 = 0.63 \qquad (9.4)$$

Where, $D_{waiting}$ =0, if there is no additional waiting time at bus-stops; $D_{waiting}$ =1, if there is one or more additional waiting time at bus-stops.

This revised model has a constant of 11.03 s which is extra dwell time compared with the origin, and has an additional waiting time of 48.6 s if there is any waiting at bus-stops. The revised model and original model against field data are shown in Figure 9.3. It can be seen that the revised model significantly improved the problem of underestimation. This model is then used to replace the original model for the revalidation of the dwell time model and SCOOT and ANPR route-based journey time models. The result showed not only that the revised dwell time model is valid (p-value=0.84 > α) but also that the two route-based regression models are valid (p-value=0.50 > α for SCOOT model; p-value=0.86 > α for ANPR model). Thus, the improvement of the dwell time model can considerably enhance the quality of the route-based journey time regression model.

- Performance comparison between link- and route-based regression models

According to the regression model development in sections 6.2.2 and 6.2.3, validation in section 7.22, and above refinement, the main difference between link- and route-based models are that link-based model formulated signal control delay and delay of acceleration and deceleration independently. However, the deviation of signal control delay and the characteristics of average of data source and approach cannot represent the reality properly by link-based model, which were discussed above. Thus, the route-based model, which is valid for the field data, is suggested for the regression approach.

In principle a link-based formulation would be preferred if the formulation is robust. However, at this stage of development, the link-based model would need to be improved for it to be recommended.

9.6.3 Application of developed models

An important aspect of a model's application is its transferability, which shows how easily a model could be applied to a different location with minimal remodelling or recalibration (Robinson and Polak, 2004). For this application, different bus routes may have various characteristics especially in different cities, which are site specific, and the deviations in bus operations, passenger demand, and traffic condition might be diverse. It is not practical to include all possible variables in a model. Nonetheless, the concepts of building blocks such as link-based analysis (section 2.2.2), modelling structure such as the main components of bus journey time (section 9.4.1), influential variables of bus journey time (section 9.4.1), modelling approaches (regression and Monte Carlo simulation), and data sources for formulating such as bus operation data, general traffic data, and road geometry and facilities data (section 4.2), can be used to develop elaborate models in different areas.

The models, which were developed in chapter 6 and validated in chapter 7, were based on the data collected in the field at the specific bus routes described in chapter 4. It is suggested that these models might be applicable only to those which have similar circumstances to the study routes. Such circumstances include the types of bus route, traffic condition, road layout, passenger demand, and bus operation etc. For example, the types of bus route may radiate from the city centre, go through or inside or in-between city centre etc., which are illustrated on Figure 9.4. Passenger demand includes the amount of passengers and their distribution along the bus route including space and time. Bus operation includes bus frequency or headway, bus stop spacing, and bus types etc.

The application of regression models would require recalibration with local data due to the coefficients of the models being dependent on the particular routes. By contrast, the Monte Carlo simulation models requires stochastic input variables for each route, hence such approach may be more applicable than regression method. In addition, particular attention should be paid to the underlying assumptions for this research which were described as follows:

- Control delay at signalised junctions (sections 6.2.3.3 and 6.3.3.2): the uniform bus arrivals and no initial queue were assumed.
- Acceleration and deceleration delay (section 6.2.3.4): equality of V_1 and V_2 was made, i.e. the bus speeds are assumed the same before and after a bus serves a bus-stop.
- Formulation of Monte Carlo approach (section 6.3.1): an assumption was made for this study that modules between bus journey time, the variables in each module, and observations showed in Monte Carlo output were all independent. This is a simplification of this research.
- Dwell time module of Monte Carlo simulation (section 6.3.2.2): alighting time of each passenger at a stop is equal and every passenger's boarding time is the same at the same bus-stop.

9.7 Summary

This chapter has described the overall discussion in exploring the possible limitation and alternative approaches of this study, which was derived from the underlying assumptions and probable error, in order to better understand the ability of this study. Discussion was started with the accuracy of bus journey time, dwell time, acceleration/deceleration rate, and road facility location, which were obtained form GPS data. Then, alternative methods for accessing abundant data sources and processing such data were mentioned. Additional consideration of affecting variables of bus journey time for more comprehensive research was also addressed. Next, some factors such as time series, bus holding, and alternative data manipulation which may influence the model performance were illustrated. In addition, supplementary discussions, i.e. sample size, additional consideration in modelling process, selecting input probability distributions and interpretation of simulation outcome, were stated for the regression and the Monte Carlo model respectively. Finally, there were discussions of the test risk of validation and of potential methods to improve invalid models.

Chapter 10: Summary and Conclusion

This chapter summarises the achievements carried out during the research work and derives conclusions from the study. The aim of this research was to understand the variability of the journey times of buses and their relationship to the journey times of other vehicles using models developed within this research. The study began with a review of available literature on journey time estimations, data collection methods, characteristics of bus journey times, relationships with other vehicles and modelling methodologies. It was clear from the literature review that an elaborate model is necessary to meet the requirement of traffic fluctuation and passenger demand, incorporating increasingly advanced traffic control and traveller information systems in the urban area. In addition, the use GPS for data collection was selected and the accuracy of the survey data was investigated. The data collection was then conducted for four main bus routes in Southampton, UK. Multiple regression and Monte Carlo simulation models based on link and route separations were developed according to the field data, and then validated. The valid model was used to determine the sensitivity of the model outputs to changes of its parameters and explore the impact of changes on bus operations. The underlying assumption, limitation, and potential improvement of this study were then discussed for better understanding. The key findings of the research are given in this chapter. This chapter starts with the key finding of this research, followed by potential applications and possible areas for future research.

10.1 Main findings of the research

- General travel time (71%), in which the driving situation is similar to other vehicles but generally longer, dwell time at bus-stops (14%), control delay at signalised junctions (9%), and delays of deceleration and acceleration due to serving bus-stops (6%), are major components of bus journey time (sections 5.1 and 9.4.1). (The values shown in brackets are the journey time of each component compared with total journey time, which may be slightly different of different bus routes. Above values are the case of the Bitterne inbound route.)
- Using probe vehicles equipped with GPS to collect bus journey time could provide a detailed temporal-spatial profile with delay information which enables analysis of the

delay of deceleration and acceleration. The accuracy of GPS for the journey time survey, which used a portable GPS receiver with a 1-second updating rate and data-logger in this study, is as follows (sections 3.4.3 and 9.2.3):

- Position: within 4.15 meters with 95% confidence
- Speed: +/- 6.4 kph (1.78 m/s)
- Acceleration/deceleration: +/- 0.31 m/s/s.

- The number of stops made by a bus along a bus route and the critical junctions which may have longer signal control delay are the key variables on the variances of bus journey time. The former variable contributes the most variation (35% in this study) towards bus journey time. Such variability was derived from passenger demand along the bus route, while passenger demand can be obtained through historical data, field survey, or assumed distribution. Therefore, understanding this variable more accurately can greatly help the estimate of the dwell time and acceleration/deceleration delay, and hence bus journey time (sections 8.2.2 and 8.3).
- Signal control delay plays an important role in bus link journey time estimation and is the major source of deviation, hence needs isolating from the general travel time to identify its effect (sections 6.2.3.3, and 7.2.2.2).
- A dwell time model of each bus stop is recommended, as follows. Dwell time with a basic value of 5.07s for opening and closing doors plus 1.19s for each alighting passenger and 8.88s for each boarding passenger when a bus stops at a stop. The comparison of this result and the literature was presented in Table 8.6 (sections 2.7.1, 5.3.5, and 8.5.3).

$$T_D = 5.07 + 1.19\,P_A + 8.88\,P_B$$

Table 10.1: Description of proposed dwell time model of each stop

Symbol	Meaning	Unit	Range of data
T_D	Bus dwell time per stop	sec	3 ~ 154
P_A	Number of alighting passengers	n. of people	0 ~ 13
P_B	Number of boarding passengers	n. of people	0 ~ 12

Additionally, a revised dwell time model for route estimation, which consider the additional waiting time at bus stops e.g. due to a holding strategy, is presented as follows (section 9.6.2).

$$DT = 11.03 + \sum_{i=1}^{N_{stopped}} T_{Di}(Equation\ 5.5) + 48.6 * D_{waiting}$$

Table 10.2: Description of proposed route dwell time model

Symbol	Meaning	Unit	Range of data
DT	Bus dwell time per stop	sec	$0 \sim 337$
$N_{stopped}$	Number of stops made by a bus along a route	n. of stops	$0 \sim 15$
T_{Di}	Bus dwell time at *i* bus-stop	sec	$3 \sim 154$
$D_{waiting}$	Dummy variable of additional waiting time at bus-stops	n.a.	0, 1

- The models of acceleration/deceleration rate (A/D) due to bus-stops are recommended as follows, which were modelled from successive GPS points of buses: The key variable is the bus cruise speed before deceleration (V_1) and the bus cruise speed after acceleration (V_2)(sections 2.7.2, and 5.4.2).

$$A = \frac{(-0.14 \times Ln(V_2) \times V_2^{0.5} + 0.78 \times V_2^{0.5})^{10}}{100}$$

$$D = \frac{-(0.13 \times Ln(V_1) \times V_1^{0.5} + 0.75 \times V_1^{0.5})^{10}}{100}$$

Table 10.3: Description of proposed average acceleration/deceleration rate model

Symbol	Meaning	Unit	Range of data
A	the average acceleration rate	m/s/s	$0.21 \sim 1.79$
V_2	the cruise speed after acceleration	km/hr	$7 \sim 72$
D	the average deceleration rate	m/s/s	$-0.27 \sim -2.30$
V_1	the bus cruise speed before deceleration	km/hr	$8 \sim 72$

- Regression models have been found to give an acceptable estimate of expected journey time and have the advantage of using only some major independent variables, i.e. distance, the number of bus-stops per km, the number of bus-stops made by a bus per km, timetable adherence, traffic condition (e.g. speed and occupancy of SCOOT parameters), the number of disturbances per km, the percentage of bus-lane length against total length, other vehicles' journey time (i.e. ANPR journey time), the number of alighting passengers, and the number of boarding passengers, which is recommended as follows (sections 6.2.2.4. and 9.6.2):

- Model for SCOOT data

L *(55.5+70.81 N_{stop} + 13.45 $N_{stopped}$ -1.34 AD -2.6 ST_s + 0.72 ST_o)+

(11.03+5.07 $N_{bus-stop}$ + 1.19 P_A + 8.88 P_B+48.6 $D_{waitting}$)

- Model for ANPR data

L *(58.43+31.15 N_{dist} + 22.61 $N_{stopped}$ - 365.4 $Buslane\%$ + 0.22 JT_{ANPR}) +

(11.03+5.07 $N_{bus-stop}$ + 1.19 P_A + 8.88 P_B+48.6 $D_{waitting}$)

Table 10.4: Description of proposed regression model

Symbol	Meaning	Unit	Range of data
L	route length	km	3.2 ~ 6.8
N_{stop}	the number of bus stops per km	n. of bus stops /km	2.05 ~ 2.69
$N_{stopped}$	the number of stopped bus-stops per km	n. of stops/km	0 ~ 2.69
AD	adherence of bus timetable	min	-3 ~ 25
ST_s	SCOOT speed parameter	Km/hr	24.95 ~ 39.3
ST_o	SCOOT occupancy parameter	%	27.5 ~ 104.49
$N_{bus-stop}$	the number of stops made by a bus along a route	n. of stops	0 ~ 15
P_A	Number of alighting passengers	n. of people	0 ~ 13
P_B	Number of boarding passengers	n. of people	0 ~ 12
$D_{waitting}$	dummy variable of additional waiting time at bus-stops	n.a.	0, 1
N_{dist}	the number of disturbances per km	n. of disturbances/km	1.61 ~ 5.27
$Buslane\%$	the percentage of bus lane length	percentage	0 ~ 0.16
JT_{ANPR}	ANPR journey time per km	sec/km	81.05 ~179.27

- Monte Carlo models, which yield information on the distribution of bus journey time due to variability caused by the fluctuation of traffic and passenger demand and signal timing, were shown to give better estimations, with greater tolerance when variables had larger deviations. Both route-based and link-based Monte Carlo models are valid under the independent validation of the study route. Such models require more independent variables for each independent bus route than regression models, hence the approaches of Monte Carlo model may be more applicable (sections 6.3.2, 6.3.3, 7.2.3, 7.3, and 9.6.3).

- Bus journey time which is excluded dwell time and acceleration/deceleration delay due to serving bus-stops is estimated to be 1.34 times of that other vehicles' journey time. This result is similar to those found from previous studies, but slightly lower. This may be because the acceleration/deceleration delay had been subtracted from the bus journey time of this study, which was not taken into account in other studies, hence a slightly lower value was obtained (sections 2.6.2 and 9.4.2).
- The number of bus-stops made by a bus along the bus route and the critical junctions which may have longer signal control delay are the key factors on the variances of bus journey time. Thus, the potential ways to improve bus journey time should focus on these crucial aspects. The consolidation of bus-stops and adjustment of signal timing at critical junctions are the most essential improvements (sections 8.2.2, 8.3, 8.5.1, and 8.5.2).
- The boarding time per passenger of the study routes is significantly greater than the value suggested by the literature, hence improvement of such time, such as through simplification of fare collection methods, might be a cost-effective approach compared with bus priority schemes to decrease dwell time as well as bus journey time (sections 2.7.1.3, 5.3.5, and 8.5.3).

10.2 Potential applications

10.2.1 Further applications of the current model

Bus journey time influences service attractiveness, operation cost, and system efficiency (Bertini and El-Geneidy, 2004). Understanding the key factors affecting bus journey time and planning potential strategies to improve services and operations are essential. Followings are the potential applications using developed models.

- To identify critical places

The most useful aspect of this study is the ability to use the estimation model to identify critical points or links for improvement, e.g. bus-stops consolidation or bus priority scheme, using sensitivity analysis approaches, which were described in chapter 8.

- To evaluate intended strategies

In addition to the strategies for enhancing bus service such as changes in passenger demand, fare collection methods, and bus priority which was discussed in section 8.5, the developed model may be used to evaluate the intended strategies which might not be feasible to test in the field before putting into practice. For example, a local authority may plan to install a signalised control at a junction which has had several accidents and increasing traffic demand. Bus operators may be concerned about the effect of bus delay on the schedule and dispatching if the junction is located on a busy bus route. Therefore, the delay time due to the installation of the signals can be assessed through simulation.

- To explore the requirement of a bus fleet for a potential bus route or the adjustment of an existing bus route

When a new bus route or adjustment of an existing bus route is planned and after the survey of potential passenger demand along the route, it is not easy to understand the probable journey time dependent upon the traffic conditions, and hence the required number of buses and the schedule for service. With the help of the current model, it may be possible to get a general idea at the initial stage.

10.2.2 Bus journey time application

As the developed models aim to estimate bus route journey time, their sub-models such as dwell time and delay of acceleration/deceleration models used a route level for estimation, i.e. the models didn't estimate dwell time and delay of acceleration/deceleration at a specific location for a specific time. In addition, prediction was not taken into account in the research, which was discussed in section 9.5.1.1. Therefore, in their current forms, they are more suitable for off-line journey time analyses which were described in section 10.2.1. Nevertheless, the link-based models, which can estimate journey time on a specific link, may make modifications for real-time applications. Such modifications require the ability to locate actual position of a bus and coordinate with real-time database. For example, the traffic parameters at the first detector can be used to calculate general travel time for the first link, dwell time can be calculated using the data from on-board passenger counting, and delay of acceleration/deceleration can be estimated by whether a bus made a stop along the first link. The journey time obtained for the first link can be added to the start time to obtain a new start time for the second link. The measurement of traffic parameters associating with the new start time can be determined and the journey time of the second link is calculated. The journey time for the following links which are composed of a route are calculated in the same approach.

In order to improve the service of public transport to attract more potential users, authorities aim to provide traveller information as one of the most important improvements, which is within the context of Intelligent Transportation Systems [ITS](FHA, 2005). Advanced Public Transportation Systems (APTS) are designed to collect, process, and disseminate real-time data of operational individual vehicles via advanced location and communication technologies. Such data can be coupled with traffic data from a traffic management system to help inform travellers of possible delays through Advanced Traveller information Systems (ATIS) and for quick response to transit operation by dispatchers. The data flow in these systems is shown in Figure 10.1. The key element and requirement of such a system is the ability to estimate transit arrival time, which is based on timely and accurate data from APTS such as GPS, Automatic Vehicle Location (AVL), Automatic Passenger Counting (APC) and traffic management systems such as SCOOT and ANPR, in order to provide real-time, accurate information to travellers, so that they may make better travel decisions. This information can be then disseminated as traveller information, e.g. en-route information such as displays on board buses and pre-trip information such as visual display at kiosks, terminals, transfer points, and bus-stops, and operational information for the dispatcher to improve transit operation and maintenance.

With competition between GPS manufacturers rapidly enhancing technology and performance, the creation of new products by integrating existing equipment such as vehicle navigation and mobile phone, and the reduction of prices, these indeed help the spread and increase the proportion of potential probe vehicles on the road network, and hence the more comprehensive traffic data can be achieved. In-vehicle navigation GPS is expected to grow from about 4 million sold in 2006 to 30 million sold by 2020; in the meanwhile, mobile phones enabled with GPS are expected to increase in number at a surprisingly rate from around 200,000 sold in 2006 to as many as 2 billion sold annually by 2020 (GPS World, 2006). In addition, the competition between the constellation providers, namely the GPS of the USA, the Global Navigation Satellite System (GLONASS) of the Russian Federation, and Galileo of the European Union, should do the same (Ashjaee, 2006). Therefore, a high accuracy of vehicle location and high proportion of probe vehicles on the road network are expected. This is a very good niche to manage traffic with a comprehensive view of the network.

10.3 Future work

The considered variables in this study seem to be able to reflect the variability of bus journey time for the studied bus routes. However, some factors including road layout such as the number of lanes, roundabouts, and roadside parking, bus stop features such as bus stop consolidation, bus island, time factors such as time of the day, day of the week, nonrecurring incidents such as road maintenance and accidents, and additional bus priority facilities such as bus gating, signal priority, and pre-signal, and so on, might influence journey time but were not taken into account in this study due to limited data or limited scale of bus routes. Therefore, an elaborate model may be required to accommodate various circumstances.

It can be expected that more accurate and comprehensive data sources will be available and accessible in the near future as described above. Thus, it is essential to integrate all the possible data sources into a platform which is accessible for potential applications. Then, these data can be processed through data mining, data fusion, or estimation techniques to become useful information and be disseminated to users in a timely manner for better decision making.

In the future, with the increase of GPS installed in vehicles or embedded in mobile phones, a high density of probe vehicles on road networks and a high percentage of 'probe people' among travellers is expected. Thus, the movement of vehicles and people across a wide range of areas can be studied more comprehensively. In addition, a more complete travel study including when, where, and how a traveller travels from departure to destination, in terms of revealed traveller behaviour, will be achievable.

TABLES

Table 2.1: Summary of competing detectors

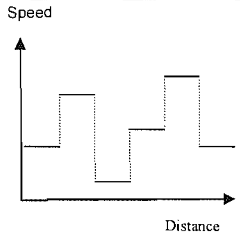
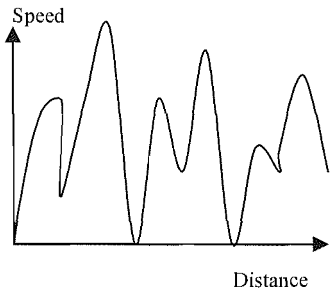
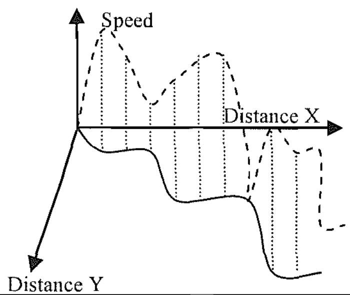
Detector Group	Graphical Presentation	Detector	Annual operation and maintenance cost for each station	Representing Delay	Main Disadvantage
Fixed detectors		Inductive loop detector	£150-£200	Average time of delay between two detectors	Lane closure for installation and maintenance
		ANPR	£350	Average time of delay between two detectors	Performances affected by illumination, and weather
		AVI	£350	Average time of delay between two detectors	Limited samples and constrained by infrastructure
Moving detectors		Probe vehicles equipped with GPS	£150-£200	Duration and intensity	Limited samples
		Aerial surveillance	No reference	Duration, intensity, extent, and reliability	Limited records and constrained studies

Table 2.2: Main advantages and disadvantages of journey time estimation methodologies

Methods / advantages and disadvantages	Empirical approach	Analytical approach	Simulation approach
Advantages	• Enable to explore possible factors • Easy understanding and application, and transferability to other sites	• Enable to transfer to other sites	• Enable to understand the effect of specific variables by controlling the predictors • Enable to vary traffic flow, bus frequency, and passenger demand over time and space • Several performance indicators are available and cost less • A more comprehensive view of possible estimations which may not be available from the field data
Disadvantages	• Main predictors are decided by the researchers, some important variables may be unconsidered • Considerable field data are required to formulate models	• Substantial assumptions are required to simplify the reality • Considerable mathematical and logical forms of equations are not easy to understand.	• Require considerable factors and data • Require substantial verification, calibration, and validation • Results might not be consistent for each execution and limited transferability

Table 2.3: A SWOT analysis for buses compared with private automobiles

Strength: • Congestion is thought to be even worse due to limited road space • Bus priority facilities are available in most cities • Public transport is regarded as the primary trend for authorities in most metropolitan areas • Lower cost owing to government subsidy against keeping a car and other expenditure such as fuel, parking fee, etc. • Congestion charging is carried out or under planning.	**Weakness:** • Reliability may be getting worse due to more congested traffic • Privacy and safety issues • Not door-to-door service • It is uncomfortable when seat is unavailable or bus is crowded.
Opportunity: • Increasing bus use has been established as a policy. • More bus priority schemes are under planning. • New technologies such as AVL, AVI which is associated with personal mobile communication could provide real-time information for passengers, fleet management, and local authority for enhancing bus service.	**Threat:** • Other competing public transport mode such as rapid transit, rail • New technologies such as vehicle guidance can help cars get through congestion • The overall cost of motoring has remained at or below its 1980 level • Road redistribution or bus priority measures might not reach a general consensus by the public • Do nothing for the buses by local authorities

Table 2.4: The advantages and disadvantages of buses journey time compared with other vehicles

Advantages	Disadvantages
Bus priority methods (TCRP, 2003) Roadway treatment: • Exclusive bus lane Site-specific treatments: • Queue jumps • Curb extensions • Boarding islands Traffic signal treatments: • Signal priority • Bus gate Bus operation treatments: • Bus stop relocation • Bus stop consolidation • Skip- stop operation Others: • Yield-to-bus laws • Parking restrictions • Turn restriction exemption	• Delay due to bus-stops service (deceleration, stop for boarding and alighting passengers, acceleration, returning to traffic) • Bus route might not be the economical path, e.g. the path might not be the shortest path or take least time owing to the consideration of business to serve more passengers. That is, with the same origin and destination, other vehicles may not drive the same route as buses. Buses may usually leave the urban corridor and go into more developed area such as high streets, shopping malls, employment centres, and residential areas) • Although congestion ahead is known, buses can not take any alternative to avoid the jam due to the fixed route. • Buses usually travel on the curb lane of an urban corridor when two and more lanes are available. The leftmost lane is usually slower than other lanes. • The ability of operation (start, acceleration, deceleration, and turn) is not as good as automobiles because of the vehicle characters. • Bus holding for schedule adherence • Preparing to stop at bus stop (finding potential passengers when there are people waiting at bus stop) may affect driving speed compared with other vehicles. • Roadside parking and temporary parking may jam buses or slow them down. • If bicycle, motorbike, taxi or other vehicles are allowed to use the bus lane, this may affect bus driving

Table 2.5: Summary of key factors influencing bus journey time from literature

Factors		1983 Levinson	1988 Seneviratne	1997 McKnigtht and Paaswell	1998 Abdelfattah and Khan	2003b TCRP	2004 McKnight , et al.	2004 Bertini, EI-Geneidy	2004 Chakfoborty and Kikuchi
Road geometry	Route/segment length		√	√	√	√	√	√	√
Road and traffic control facilities	N of bus stops (stop spacing)	√	√	√	√	√	√		
	N of signalized junctions		√		√	√			
Dwell time related	N of bus-stops made by bus				√			√	√
	Dwell time								√
	N of boarding passengers			√		√	√	√	
	N of alighting passengers					√		√	
	Total N. of dwell passengers	√	√						
	Acceleration / deceleration	√						√	
Traffic parameters (congestion, flow, speed, density)	Congestion					√			
	Flow, speed, or density				√				
	Other vehicle journey time			√			√		√
	Turning movement				√				
Bus operation	Schedule adherence					√			
	Vehicle characteristics					√			
Others	Service areas					√			

Table 3.1: Possible sources of error

Error source	Error	
	RMS*	2DRMS**
Ionosphere	7 m	13.6 m
Clock and ephemeris	3.6m	6.98 m
Average geometry of satellites	2 m	3.9 m
Receiver noise	1.5 m	2.9 m
Multipath	1.2 m	2.3 m
Troposphere	0.7 m	1.4 m
Total errors	16 m	31 m

RMS*:Root Mean Square, with 50% confidence interval.
2DRMS**: Twice Distance Root Mean Square, with 95% Confidence interval.

Table 3.2: Test areas and traffic conditions against travel time

Travel time	Area	Traffic condition
00:00:00 ~ 00:06:46	urban roads	Only slightly restricted manoeuvres.
00:06:47 ~ 00:28:37	Motorway	Some restricted manoeuvres and slight queuing for work zone of one location.
00:28:38 ~ 00:29:53	rural roads	Only slightly restricted manoeuvres.
00:29:54 ~ 00:37:20	urban roads with dense trees around	Some restricted manoeuvres.
00:37:21 ~ 00:47:35	city centre roads	Extremely low speed, stop-and-start condition.
00:47:36 ~ 00:53:16	city centre roads	Restricted by the traffic with some delays.
00:53:17 ~ 01:03:37	city centre roads with some tall buildings or shopping mall over-bridge around	Restricted by the traffic with some delays
01:03:38 ~ 01:09:42	urban roads	Some restricted manoeuvres.
01:09:43 ~ 01:16:44	urban roads	Slightly restricted manoeuvres.

Table 3.3: GPS Equipment Specifications

GPS sets	1	2	3
GPS Receiver	Garmin 35 PC	Garmin 35 PC	Racelogic VBox III
Data Recorder	Potable data logger (Model: DGPS-XM4-ALT)	On-board PC	On-board PC
Raw data logged	Date, time, latitude, longitude, ellipsoid height	Date, time, latitude, longitude, ellipsoid height, velocity, heading	Date, time, latitude, longitude, ellipsoid height, distance, velocity, acceleration, heading
Signal Quality indicators logged		Number of satellites in use, horizontal dilution of precision (HDOP), vertical dilution of precision (VDOP), position dilution of precision (VDOP), estimated horizontal position error (HPE), estimated vertical position error (VPE), and estimated position error (EPE)	Number of satellites in use
Performance	Update rate: 1 Hz Time accuracy: 1 second Position accuracy: 15 m RMS	Update rate: 1 Hz Time accuracy: 1 second Position accuracy: 15 m RMS Velocity accuracy: 0.2 m/s RMS	Update rate: 100 Hz Time accuracy: 0.01 second Position accuracy: 3 m 95% Circle of error probable (CEP) Distance accuracy: 0.05% (<50 cm per Km) Velocity accuracy: 0.1 Km/h Acceleration accuracy: 0.5%

Table 3.4: Data processing method for accessing parameters of each data set

Approach		Parameters			
		Position	Distance	Speed	Acceleration
GPS set	1	Measured	Calculated	Calculated	Calculated
	2	Measured	Calculated	Measured	Calculated
	3	Measured	Measured	Measured	Measured

Measured: data was obtained from GPS receiver directly.
Calculated: data was achieved by a post calculation.

Table 3.5: Distance accuracy comparison for five algorithms

	Equation 3.1	Equation 3.2	Equation 3.3	Equation 3.4	Equation 3.5
Distance accumulation (m)	2444.8413	3040.4888	2747.5847	2441.4113	2445.9080
MAD	0.00228	1.273192	0.645989	0.009629	-
MAPE	0.00046	0.310573	0.159412	0.002286	-
RMSE	0.03861	2.025238	1.036526	0.014768	-

Table 3.6: 2D position accuracy measures

Accuracy Measures	Formula	Probability	Definition	Calculated confidence region (m)
CEP	$0.62\sigma_x + 0.56\sigma_y$ (Accurate when $\sigma_y / \sigma_x > 0.3$)	50%	The radius of a circle contains the position with probability of 50%.	1.6093
DRMS	$\sqrt{\sigma_x^2 + \sigma_y^2}$	65%	The square root of the average of the squared horizontal position errors.	2.0735
2DRMS	$2\sqrt{\sigma_x^2 + \sigma_y^2}$	95%	Twice the DRMS of the horizontal position errors.	4.1469

Note: σ_x and σ_y (0.9151 and 1.8606) are the standard deviations of estimated coordinates (x,y) of GPS points.

Table 4.1: Bus service for bus routes

Survey Route	Outbound (bus company: bus service)	Inbound (bus company: bus service)	Bus frequency (peak time)
Bitterne Road (A3024)	First: 5A, 78 SolentBlueLine: M27	First:5, 5A, 78	5A: 3 services per hour 78: 1 service per hour M27:1 service per hour
Portsmouth Road (A3025)	First: 72, 80	First: 72, 80	72: 1 service per hour 80: 1 service per hour
Portswood Road	First:11A SolentBlueLine: BlueStar 2 Unilink: U6	First:11A SolentBlueLine: BlueStar 2 Unilink: U6	11A: 3 service per hour BlueStar 2: 4 services per hour U6: 2 services per hour
The Avenue (A33)	SolentBlueLine: BlueStar 1, 44/44A	SolentBlueLine: BlueStar 1, 44/44A	BlueStar 1: 4 services per hour 44/44A: 1 service per hour

Table 4.2: The number of survey runs and dwells of each direction of each route

Survey Route		Survey Runs	Dwells
Avenue (A33)	Outbound	41	220
	Inbound	39	200
Bitterne (A3024)	Outbound	38	272
	Inbound	38	226
Portsmouth(A3025)	Outbound	27	118
	Inbound	27	70
Portswood	Outbound	38	159
	Inbound	36	144
Total	Outbound	144	769
	Inbound	140	640

Table 4.3: Available ANPR routes for survey and their lengths

ANPR Route		Distance between cameras (m)
Avenue (A33)	Outbound	4,577
	Inbound	4,558
Bitterne (A3024)	Outbound	6,084
	Inbound	6,058
Portsmouth (A3025)	Outbound	4,890
	Inbound	5,102

Table 4.4: Available inductive loop detectors by route

Route	Number of detectors	Loop detector ID
Avenue (A33)	13	N07331, N04154, N04155, N04156, N04157, N04158, N04141, N04144, N03112, N03111, N03311, N03244, N03214
Bitterne (A3024)	13	N07211, N07221, N10221, N10214, N10231, N10241, N10111, N10121, N10321, N10331, N10341, N10351, N10361
Portsmouth (A3025)	4	N12134, N12111, N12121, N12164
Portswood	9	N05111, N05142, N06121, N06125, N06111, N06134, N06211, N06231, N06232

Table 4.5: Road geometry and facilities by route

Route / facilities	Avenue		Bitterne		Portsmouth		Portswood	
	I	O	I	O	I	O	I	O
Length (m)	5520	5370	6829	6566	6664	6821	3225	3346
Number of bus-stops	14	13	17	16	16	14	8	9
Number of signalised junctions	4	5	16	17	5	5	9	9
Number of signalised pedestrian crossings	7	7	4	1	2	3	7	7
Zebra crossings	1	1	1		1		1	1
Roundabout	2	2	1		2	2		
Give-Way signs					1	1		
Right turns	1		3	5	4	4		1
Left turns		1	5	4	3	4	1	
Bus lane: number of places, total length (m)		2, 580	6, 1123	2, 392	1, 85	1, 210		

I: Inbound; O: Outbound

Table 4.6: An example of recording sheet data of buses after error check

No.	Route	Direction	Date	Bus Number	Low floor	Double_decker	Num_doors	Boarding time (hh:mm:ss)	Alighting time (hh:mm:ss)	Journey time (s)	Total dwell time (s)	Bus stops	Arrival time (mm:ss)	Alighting passengers	Boarding passengers	Departure time (mm:ss)	Dwell time (s)	Note
1	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	1047		1	1112	25	
2	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	1340		6	1417	37	
3	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	1541	2		1549	8	
4	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	1722		3	1757	35	
5	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	1931	3		1940	9	
6	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	2230	6		2247	17	
7	Bitterne	O	13	5A	N	N	1	80645	82927	1362	142	7	2413	3		2424	11	
8	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	3539		1	3542	3	
9	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	3750		1	3804	14	traffic congested
10	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	3904	2	2	4001	57	
11	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	4052	2	1	4101	9	
12	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	4145	1		4149	4	
13	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	4301	3	3	4327	26	
14	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	4523		1	4618	55	waiting
15	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	4728		1	4737	9	
16	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	5058		1	5128	30	
17	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	5402	6		5411	9	
18	Bitterne	I	13	5A	N	Y	2	83428	85620	1312	223	11	5528	2		5535	7	
19	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	1155		1	1215	20	
20	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	1448		1	1510	22	elder
21	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	1602		2	1625	23	elder
22	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	2007	3	3	2045	38	elder
23	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	2153		1	2158	5	
24	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	2428		1	2449	21	wait for a while
25	Bitterne	O	13	5A	N	N	1	90923	92809	1126	132	7	2649	1		2652	3	
26	Bitterne	I	13	5A	N	Y	2	93621	95505	1124	214	6	3726		1	3733	7	
27	Bitterne	I	13	5A	N	Y	2	93621	95505	1124	214	6	3906		4	3936	30	
28	Bitterne	I	13	5A	N	Y	2	93621	95505	1124	214	6	4157		1	4216	19	
29	Bitterne	I	13	5A	N	Y	2	93621	95505	1124	214	6	4257	4		4306	9	
30	Bitterne	I	13	5A	N	Y	2	93621	95505	1124	214	6	4342	7	2	4602	140	wait for about 2 min
31	Bitterne	I	13	5A	N	Y	2	93621	95505	1124	214	6	5432	4		5441	9	
32	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	1318	3		1327	9	
33	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	1415		1	1425	10	
34	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	1540		1	1600	20	
35	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	1750		1	1803	13	
36	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	1943		1	2002	19	
37	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	2126	4	9	2246	80	
38	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	2434		2	2500	26	
39	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	2820	1		2829	9	
40	Bitterne	O	13	5A	N	Y	1	100839	103007	1288	193	9	2855	2		2902	7	
41	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	4822	1	1	4845	23	
42	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	5149	1		5156	7	
43	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	5244	2		5336	52	buggy, 2 children
44	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	5512	6	3	5557	45	
45	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	5813		1	5914	61	equipment operation
46	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	10120		1	10130	10	
47	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	10324	3	2	10348	24	
48	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	10625	1		10629	4	
49	Bitterne	I	13	5	N	N	1	104615	110914	1379	245	9	10749	2		10808	19	elder

Table 4.7: An example of raw data from GPS data logger

Valid position	Latitude (ddmm.mmm)	North or South	Longitude (ddmm.mmm)	East or West	Time (hhmmss)	Date (ddmmyy)	Altitude (metre)
A	5054.537	N	120.819	W	130343	130905	53
A	5054.534	N	120.813	W	130344	130905	53
A	5054.531	N	120.807	W	130345	130905	54
A	5054.528	N	120.801	W	130346	130905	54
A	5054.525	N	120.795	W	130347	130905	54
A	5054.522	N	120.788	W	130348	130905	55
A	5054.519	N	120.782	W	130349	130905	55
A	5054.516	N	120.776	W	130350	130905	55
A	5054.513	N	120.77	W	130351	130905	55
A	5054.51	N	120.764	W	130352	130905	56

Table 4.8: An example of raw ANPR data

Date	Timestamp	Link operatinal	Timestamp	Link Operational	Period	Travel Time	Samples	Plates In	Plates Out	Matches	Matches Calculation	Validity
13	10:00	09:58	13/09/2005 10:00	13/09/2005 09:58	600	09:19	204	56	75	4	4	Non-fatal Error/Start Node
13	09:55	09:52	13/09/2005 09:55	13/09/2005 09:52	600	10:01	262	69	78	9	8	Non-fatal Error/Start Node
13	09:50	09:48	13/09/2005 09:50	13/09/2005 09:48	600	09:34	240	64	88	8	8	Non-fatal Error/Start Node
13	09:45	09:43	13/09/2005 09:45	13/09/2005 09:43	300	09:55	165	28	49	5	5	Non-fatal Error/Start Node
13	09:40	09:37	13/09/2005 09:40	13/09/2005 09:37	600	09:28	258	59	65	5	5	Non-fatal Error/Start Node
13	09:35	09:33	13/09/2005 09:35	13/09/2005 09:33	600	10:04	231	73	64	6	6	Non-fatal Error/Start Node
13	09:30	09:27	13/09/2005 09:30	13/09/2005 09:27	600	12:30	274	66	83	8	8	Non-fatal Error/Start Node
13	09:25	09:22	13/09/2005 09:25	13/09/2005 09:22	300	13:25	185	25	48	4	4	Non-fatal Error/Start Node
13	09:20	09:18	13/09/2005 09:20	13/09/2005 09:18	600	12:54	281	84	105	5	5	Non-fatal Error/Start Node
13	09:15	09:12	13/09/2005 09:15	13/09/2005 09:12	600	11:02	292	74	106	9	9	Non-fatal Error/End Node
13	09:10	09:08	13/09/2005 09:10	13/09/2005 09:08	300	11:07	205	38	52	7	7	Non-fatal Error/End Node
13	09:05	09:02	13/09/2005 09:05	13/09/2005 09:02	600	14:23	353	83	104	8	8	Non-fatal Error/End Node
13	09:00	08:58	13/09/2005 09:00	13/09/2005 08:58	300	13:04	185	38	47	5	5	Non-fatal Error/End Node
13	08:55	08:53	13/09/2005 08:55	13/09/2005 08:53	300	12:55	218	36	50	8	8	Non-fatal Error/End Node
13	08:50	08:47	13/09/2005 08:50	13/09/2005 08:47	600	17:23	400	79	125	7	7	Non-fatal Error/End Node
13	08:45	08:43	13/09/2005 08:45	13/09/2005 08:43	600	14:38	352	86	111	11	11	Non-fatal Error/End Node
13	08:40	08:37	13/09/2005 08:40	13/09/2005 08:37	300	14:13	253	45	57	8	8	Non-fatal Error/End Node
13	08:35	08:33	13/09/2005 08:35	13/09/2005 08:33	600	13:44	361	88	112	7	7	Non-fatal Error/End Node
13	08:30	08:28	13/09/2005 08:30	13/09/2005 08:28	300	12:23	219	46	56	4	4	Non-fatal Error/End Node
13	08:25	08:23	13/09/2005 08:25	13/09/2005 08:23	300	14:15	212	47	53	6	6	Non-fatal Error/End Node

Table 4.9: An example of SCOOT UO7 message (13 September 2005)

Week date	Time	Message	Link node		Speed (mph)		Speed (kph)		Flow (vehicles/5 min)		Occupancy (%)		AGTPV		ALOTPV
Tu	17:05:00	U07	N10121D	MPH	16	KPH	25	FLOW	11	OCC	9	HR	4375	SR	375
Tu	17:05:00	U07	N10121E	MPH	27	KPH	43	FLOW	60	OCC	10	HR	2538	SR	196
Tu	17:10:00	U07	N10121D	MPH	16	KPH	25	FLOW	10	OCC	8	HR	4629	SR	370
Tu	17:10:00	U07	N10121E	MPH	28	KPH	45	FLOW	44	OCC	8	HR	2956	SR	186
Tu	17:15:00	U07	N10121D	MPH	16	KPH	25	FLOW	10	OCC	8	HR	4625	SR	375
Tu	17:15:00	U07	N10121E	MPH	24	KPH	38	FLOW	43	OCC	10	HR	2535	SR	221
Tu	17:20:00	U07	N10121D	MPH	16	KPH	25	FLOW	22	OCC	7	HR	5479	SR	370
Tu	17:20:00	U07	N10121E	MPH	27	KPH	43	FLOW	53	OCC	10	HR	2445	SR	200
Tu	17:25:00	U07	N10121D	MPH	16	KPH	25	FLOW	13	OCC	8	HR	4995	SR	386
Tu	17:25:00	U07	N10121E	MPH	25	KPH	40	FLOW	37	OCC	7	HR	3849	SR	224

Table 4.10: An example of SCOOT MO2 message (13 September 2005)

Week date	Time (hh:mm:ss)	Message	Link node		Period (s)		Vehicle stops (N. of vehicle stops/hr)		Delay (vehicle-hours/hr)		Flow (vehicles/hr)		SCOOT Congestion (intervals/hr)		Detector Congestion (intervals/hr)		Faulty links
Tu	17:02:28	M02	N10121R	PERIOD	300	STP	0	DLY*10	0	FLO	0	CONG	0	RAW	0	FLTS	0
Tu	17:02:28	M02	N10121Q	PERIOD	300	STP	0	DLY*10	0	FLO	0	CONG	0	RAW	0	FLTS	0
Tu	17:02:28	M02	N10121E	PERIOD	300	STP	252	DLY*10	11	FLO	415	CONG	0	RAW	0	FLTS	0
Tu	17:02:28	M02	N10121D	PERIOD	300	STP	199	DLY*10	23	FLO	262	CONG	0	RAW	0	FLTS	0
Tu	17:07:28	M02	N10121R	PERIOD	300	STP	0	DLY*10	0	FLO	0	CONG	0	RAW	0	FLTS	0
Tu	17:07:28	M02	N10121Q	PERIOD	300	STP	0	DLY*10	0	FLO	0	CONG	0	RAW	0	FLTS	0
Tu	17:07:28	M02	N10121E	PERIOD	300	STP	333	DLY*10	9	FLO	515	CONG	0	RAW	0	FLTS	0
Tu	17:07:28	M02	N10121D	PERIOD	300	STP	165	DLY*10	16	FLO	169	CONG	0	RAW	0	FLTS	0
Tu	17:12:28	M02	N10121R	PERIOD	300	STP	0	DLY*10	0	FLO	0	CONG	0	RAW	0	FLTS	0
Tu	17:12:28	M02	N10121Q	PERIOD	300	STP	0	DLY*10	0	FLO	0	CONG	0	RAW	0	FLTS	0
Tu	17:12:28	M02	N10121E	PERIOD	300	STP	225	DLY*10	9	FLO	330	CONG	0	RAW	0	FLTS	0
Tu	17:12:28	M02	N10121D	PERIOD	300	STP	148	DLY*10	17	FLO	183	CONG	0	RAW	0	FLTS	0

Table 4.11: An example of SCOOT M37 message (13 September 2005)

Week date	Time (hh:mm:ss)	Message	Node			UTC stage		Inter Green (s)		Green length (s)		Length (s)
We	17:01:25	M37	N10121	UTC	STG	A	IG	7	GN	94	LEN	101
We	17:01:43	M37	N10121	UTC	STG	C	IG	7	GN	11	LEN	18
We	17:02:11	M37	N10121	UTC	STG	A	IG	7	GN	21	LEN	28
We	17:02:37	M37	N10121	UTC	STG	B	IG	7	GN	19	LEN	26
We	17:03:13	M37	N10121	UTC	STG	C	IG	5	GN	31	LEN	36
We	17:04:13	M37	N10121	UTC	STG	A	IG	7	GN	53	LEN	60
We	17:04:47	M37	N10121	UTC	STG	C	IG	7	GN	27	LEN	34
We	17:05:46	M37	N10121	UTC	STG	A	IG	7	GN	52	LEN	59
We	17:06:22	M37	N10121	UTC	STG	C	IG	7	GN	29	LEN	36
We	17:07:32	M37	N10121	UTC	STG	A	IG	7	GN	63	LEN	70

Table 5.1: Test of normality for the percentage of general travel time against ANPR journey time

	Kolmogorov-Smirnov(a)		
	Statistic	df	Sig.
Bus general travel time/ANPR JT	.065	58	.200

Table 5.2: Descriptive statistic of percentages of general travel time against ANPR journey time

		Statistic	Std. Error
Bus general travel time/ANPR JT	Mean	1.3401	.03269
	Median	1.3194	
	Variance	.062	
	Std. Deviation	.24895	
	Minimum	.81	
	Maximum	1.78	
	Range	0.97	
	Skewness	-.009	.314
	Kurtosis	.124	.618

Table 5.3: Passenger distribution along bus routes

N of passenger per stop	Bitterne O	Bitterne I	Portsmouth O	Portsmouth I	Portswood O	Portswood I	Avenue O	Avenue I
0	7.76	9.82	9.32	13.24	4.35	3.65	7.41	8.82
1	3.61	3.39	2.48	1.68	1.76	2.38	3.00	3.21
2	2.05	1.66	1.12	0.44	1.22	0.73	1.26	1.28
3	1.03	0.68	0.32	0.28	0.38	0.51	0.62	0.38
4	0.53	0.61	0.32	0.20	0.46	0.38	0.33	0.18
5	0.18	0.26	0.16	0.08	0.30	0.14	0.15	0.05
6	0.39	0.26	0.16	0.00	0.22	0.11	0.10	0.03
7	0.11	0.11	0.08	0.00	0.16	0.05	0.00	0.00
8	0.11	0.08	0.00	0.04	0.05	0.03	0.03	0.03
9	0.08	0.11	0.00	0.00	0.03	0.00	0.03	0.00
10	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
11	0.03	0.00	0.00	0.00	0.05	0.00	0.00	0.00
12	0.08	0.03	0.04	0.00	0.00	0.00	0.00	0.00
13	0.03	0.00	0.00	0.04	0.00	0.00	0.00	0.00
14	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00

Note: O-Outbound, I-Inbound

Table 5.4: The distribution of survey data of the number of passengers per stop of the field data compared with theoretical distribution

N of passenger per stop	Data collected		Theoretical distribution	
	Average	Probability	Poison	Negative Binomial
0	8.0462	0.6016	0.5444	0.7866
1	2.6875	0.2009	0.3310	0.0928
2	1.2194	0.0912	0.1006	0.0429
3	0.5253	0.0393	0.0204	0.0247
4	0.3753	0.0281	0.0031	0.0157
5	0.1656	0.0124	0.0004	0.0105
6	0.1588	0.0119		0.0072
7	0.0633	0.0047		0.0051
8	0.0446	0.0033		0.0037
9	0.0296	0.0022		0.0027
10	0.0034	0.0003		0.0020
11	0.0100	0.0008		0.0015
12	0.0182	0.0014		0.0011
13	0.0083	0.0006		0.0008
14	0.0067	0.0005		0.0006
15	0.0032	0.0002		0.0005
16	0.0000	0.0000		0.0004
17	0.0064	0.0005		0.0003
18	0.0000	0.0000		0.0002
19	0.0000	0.0000		0.0002
20	0.0000	0.0000		0.0001
21	0.0032	0.0002		0.0001

Table 5.5: Notes and number of records of dwells on recording sheets

Note	Description	Number of records	Excluding mark (*)
1	1 infant	11	
2	1 infant and buggy	4	
3	2 infants	6	
4	2 infants and buggy	1	
5	Bus stopped over the bus_stop	1	*
7	Elder	5	
8	Equipment operation	1	*
9	Bus go and stop again	1	*
10	Just waiting (no passenger alighting or boarding)	5	*
12	Non bus_stop	7	*
13	Passenger query	8	
15	Passenger tried to find money	1	*
17	Queuing (waiting for bus berth)	5	*
18	Return to traffic congested	16	*
19	Waiting at bus_stop	56	*
Total		128	93 (excluded records)

Table 5.6: Alternative accumulated percentages for various options for maximum dwell times per passenger

Accumulated percentage (%) below the max. dwell time	Max. dwell time of Boarding (s)	Max. dwell time of Alighting (s)	Max. dwell time of Combined (s)
90 %	21	10	12
95 %	24	14	15
99 %	32	25	30

Table 5.7: Comparison of descriptive statistics of dwell time with/ without extreme values

Descriptive Statistic	Av. boarding time		Av. alighting time		Av. boarding and alighting time	
	With	Without	With	Without	With	Without
Mean	11.90	10.82	5.72	4.91	6.82	5.92
Median	10.00	10.00	5.00	4.42	5.33	5.00
S.D.	7.56	5.26	4.51	2.58	5.93	3.27
Minimum	1	1	0.67	0.67	0.83	0.83
Maximum	76	24	34	14	55	15.29
Observations	347	330	366	348	277	265

Table 5.8: ANOVA test results of possible factors in dwell time

Statistic Test categories	Dwell time per passenger			
	Hypothesis	Result (α=0.05)	Note	Descriptive statistics
1. Routes	$H_0 : \mu_1 = \mu_2 = \mu_3$ H_a : not all of the μ_i are equal	F=6.419, P-value=0.002, reject H_0, this means these routes are not all the same. When check the post hoc test of Tamhane, $\mu_1 \neq \mu_2$ and μ_1 may equal to μ_3 are found. Test of homogeneity of variance[HOV] (levene statistic) shows that these three routes' variance are not all the same (p-value=0.000).	μ_1, μ_2, and μ_3, represent the mean of dwell time per passenger of Portswood, Bitterne, and Portsmouth route respectively.	μ_1=7.77, SD_1=4.10, n_1=283; μ_2=6.73, SD_2=4.74, n_2=482; μ_3=7.90, SD_3=5.31, n_3=178
2. Inbound/Outbound	$H_0 : \mu_1 = \mu_2$ H_a : not all of the μ_i are equal	P-value=0.048, reject H_0, this means of inbound and outbound are not all the same. Test of HOV shows that their variances may be similar (p-value=0.099).	μ_1, and μ_2 represent the mean of dwell time per passenger of Inbound and Outbound respectively.	μ_1=7.60, SD_1=4.92, n_1=427; μ_2=6.99, SD_2=4.50, n_2=516
3. Different survey date	$H_0 : \mu_1 = \mu_2 = \mu_3$ H_a : not all of the μ_i are equal	P-value=0.434, there is not significant evidence to support they are different. Test of HOV shows that their variances may be similar (p-value=0.215).	μ_1, μ_2, and μ_3 represent the mean of dwell time per passenger of 13^{th}, 14^{th}, and 15^{th} respectively.	μ_1=7.49, SD_1=4.80, n_1=295; μ_2=7.30, SD_2=4.72, n_2=343; μ_3=7.00, SD_3=4.58, n_3=305
4. Different period	$H_0 : \mu_1 = \mu_2$ H_a : not all of the μ_i are equal	P-value=0.000, reject H_0, this means peak and off-peak period are not all the same. Test of HOV shows that their variances are significant different (p-value=0.004).	μ_1, and μ_2 represent the mean of dwell time per passenger of Peak (15.00-18.00) and Off-peak period (11.00-15.00 and 18.00-19.00) respectively.	μ_1=6.50, SD_1=4.43, n_1=476; μ_2=8.04, SD_2=4.85, n_2=467

Table 5.8: ANOVA test results of possible factors in dwell time (continue)

Statistic Test categories	Dwell time per passenger			
	Hypothesis	Result (α=0.05)	Note	Descriptive statistics
5. Bus services	$H_0: \mu_1 = \mu_2 = ... = \mu_9$ H_a : not all of the μ_i are equal	P-value=0.001, reject H_0, this means that these 7 bus services are not all the same. It can be seen from the post hoc test of Tamhane that 80 and 5A, 11A and 5A are significant different. Test of HOV shows that their variances are not all the same (p-value=0.000).	μ_1, μ_2, ... μ_6 and μ_7 represent the mean of dwell time per passenger of bus service 5, 72, 80, 11A, 5A, U6C, and U6H respectively.	μ_1=7.79, SD_1=5.57, n_1=43; μ_2=7.22, SD_2=5.43, n_2=88; μ_3=8.56, SD_3=5.14, n_3=90; μ_4=8.49, SD_4=4.54, n_4=119; μ_5=6.63, SD_5=4.65, n_5=439; μ_6=7.08, SD_6=3.28, n_6=75; μ_7=7.39, SD_7=3.99 n_7=89
6. Whether the bus is low floor?	$H_0: \mu_1 = \mu_2$ H_a : not all of the μ_i are equal	P-value=0.753, there is not significant evidence to support they are different. Test of HOV shows that there is not significant different in their variances (p-value=0.214).	μ_1, and μ_2 represent the mean of dwell time per passenger of low-floor buses and non-low-floor buses respectively.	μ_1=7.20, SD_1=4.44, n_1=359; μ_2=7.30, SD_2=4.86, n_2=584
7. Whether the bus is double-decker?	$H_0: \mu_1 = \mu_2$ H_a : not all of the μ_i are equal	P-value=0.526, reject H_0, there is not significant difference between them. Test of HOV shows that there is not significant different in their variances (p-value=0.118).	μ_1, and μ_2 represent the mean of dwell time per passenger of double-decker buses and non-double-decker buses respectively.	μ_1=7.38, SD_1=4.92, n_1=370; μ_2=7.19, SD_2=4.56, n_2=573
8. How many doors does the bus use?	$H_0: \mu_1 = \mu_2$ H_a : not all of the μ_i are equal	P-value=0.113, there is not significant evidence to support they are different. However, Test of HOV shows that there is significant different in their variances (p-value=0.002).	μ_1, and μ_2 represent the mean of dwell time of buses with one door and two doors respectively.	μ_1=7.13, SD_1=4.90, n=724; μ_2=7.70, SD_2=3.94, n_2=219

Table 5.9: Correlations among dwell time variables

		T_D	P_A	P_B	P_T
T_D	Pearson Correlation	1	.231(**)	.697(**)	.592(**)
	Sig. (2-tailed)		.000	.000	.000
	N	990	643	624	990
P_A	Pearson Correlation	.231(**)	1	.229(**)	.778(**)
	Sig. (2-tailed)	.000		.000	.000
	N	643	643	277	643
P_B	Pearson Correlation	.697(**)	.229(**)	1	.754(**)
	Sig. (2-tailed)	.000	.000		.000
	N	624	277	624	624
P_T	Pearson Correlation	.592(**)	.778(**)	.754(**)	1
	Sig. (2-tailed)	.000	.000	.000	
	N	990	643	624	990

** Correlation is significant at the 0.01 level (2-tailed).

Table 5.10: Summary of bus dwell time models

Linear models (Y=dwell time of each stop)

No.	Const.	P_A	P_B	P_T	R^2	F-value	Critical value	P-value	Obs.
1	4.12	1.94			0.28	132.52	<3.81	<0.001	348
2	0.56		10.05		0.64	586.84	<3.81	<0.001	330
3	2.82			5.34	0.45	782.04	<3.81	<0.001	943
4	**5.07**	**1.19**	**8.88**		**0.63**	**218.86**	**<3.81**	**<0.001**	**265**

Polynomial models (Y=dwell time of each stop)

No.	Const	P_A	P_B	P_T^2	$P_A * P_B$	P_B^2	P_B^3	R^2	F-value	Critical value	Obs.
5	10.76			0.47				0.37	962	<3.07	943
6		3.25	11.56		-1.00			0.65	341	<2.45	265
7			12.07		-0.32			0.56	525	<2.17	265
8			12.28		-0.8	0.46	-1.12	0.62	314	<2.45	265
9	12.37	1.68				1.10		0.46	199	<2.45	265

Logarithm models (Y=dwell time of each stop)

No.	Y	Const	$\log(P_T)$	$\ln(P_A)$	$\ln(P_B)$	R^2	F-value	Critical value	Obs.
10	log(Y)	0.83	0.78			0.39	6549	<3.07	943
11	ln(Y)	2.56		0.13	0.75	0.53	3780	<2.68	265

Exponential model (Y=dwell time of each stop)

$Y = Const + a \times (\Pr edictor)^b$

No.	Const	precictor	a	b	R^2	F-value	Critical value	Obs.
12		P_T	6.25	0.93	0.45	784	<2.68	943
13	5.78	P_B	7.32	1.14	0.62	916	<2.68	330

Table 5.11: Test of normality for acceleration rates

	Kolmogorov-Smirnov(a)		
	Statistic	df	Sig.
Average Acc/Dcc (m/s/s)	.104	259	.000

Table 5.12: Descriptive statistic of acceleration rates

		Statistic	Std. Error
Acceleration rates (m/s/s)	Mean	.8380	.01751
	Median	.8092	
	Variance	.079	
	Std. Deviation	.28183	
	Minimum	.21	
	Maximum	1.79	
	Range	1.58	
	Skewness	.803	.151
	Kurtosis	.789	.302

Table 5.13: Estimated function of acceleration rate for different speed categories

Speed range (kph)	Best fit function	R-square
0-20	$A = 0.048 V_2^{0.9295}$	0.3980
20-30	$A = -0.3416 Ln(V_2) + 1.9213$	0.0330
30-40	$A = 1.4082 V_2^{-0.1395}$	0.0015
40-50	$A = 0.0131 V_2 + 0.3089$	0.0169
50 and more	$A = 1.545e^{-0.0136 v_2}$	0.0410

Table 5.14: Test of normality for deceleration rates

	Kolmogorov-Smirnov(a)		
	Statistic	df	Sig.
Average Deceleration	.096	299	.000

Table 5.15: Descriptive statistic of deceleration rates

		Statistic	Std. Error
Deceleration rates (m/s/s)	Mean	-1.0269	.02273
	Median	-.9424	
	Variance	.154	
	Std. Deviation	.39300	
	Minimum	-2.30	
	Maximum	-.27	
	Range	2.02	
	Skewness	-.911	.141
	Kurtosis	.603	.281

Table 6.1: Correlation between bus route journey time variables

	JT_G	N_{stop}	N_{sig}	N_{dist}	$N_{stopped}$	N_l	N_r	Buslane%	AD	JT_{ANPR}	ST_s	ST_o	ST_f	D	P
JT_G	1.00														
N_{stop}	0.55	1.00													
N_{sig}	0.62	0.74	1.00												
N_{dist}	0.66	0.74	0.82	1.00											
$N_{stopped}$	0.58	0.44	0.65	0.57	1.00										
N_l	-0.43	-0.49	-0.30	-0.72	-0.26	1.00									
N_r	-0.52	-0.34	-0.45	-0.80	-0.36	0.57	1.00								
Buslane%	-0.33	-0.04	0.10	-0.27	-0.01	0.33	0.70	1.00							
AD	-0.38	0.03	-0.14	-0.19	-0.14	0.06	0.41	0.61	1.00						
JT_{ANPR}	0.14	-0.15	-0.17	-0.17	-0.18	-0.04	-0.11	-0.13	-0.11	1.00					
ST_s	-0.61	-0.22	-0.40	-0.67	-0.43	0.67	0.73	0.59	0.50	-0.19	1.00				
ST_o	-0.01	-0.07	0.12	-0.39	0.10	0.69	0.54	0.38	-0.04	0.05	0.29	1.00			
ST_f	0.30	0.02	0.28	-0.05	0.31	0.31	0.14	0.09	-0.25	0.17	-0.26	0.81	1.00		
D	-0.27	0.10	-0.04	0.08	-0.12	0.08	-0.10	0.30	0.52	-0.04	0.43	-0.28	-0.41	1.00	
P	0.07	0.02	0.01	0.01	0.05	-0.02	0.01	0.01	0.05	0.25	-0.05	0.24	0.39	-0.01	1.00

Table 6.2: Summary of bus route general journey time (JT_G) models

Response: route bus general journey time (second)

No	Constant	N_{stop}	N_{sig}	N_{dist}	$N_{stopped}$	N_l	N_r	Buslane%	AD	JT_{ANPR}	ST_s	ST_f	ST_o	D	P	R-square	Adjusted R-square	P-value	F	Critical value	Obser-* vations	N of predictors
		+	+	+	+	+	+	-	-	+	-	+	+	-	+							
1	176.46		17.08		12.75			-137.4			-1.68					0.59	0.58	<0.001	70.76	<2.45	202	4
2	72.31	68.42			15.48				-1.47		-2.65					0.62	0.61	<0.001	79.62	<2.45	202	4
3	(55.5)	70.81			13.45				-1.34		-2.6		0.72			0.63	0.62	<0.001	65.76	<2.29	202	5
4	106.26		19.06		13.49			-212.32					0.96			0.59	0.58	<0.001	70.47	<2.29	202	4
5	94.66		15.80		21.26			-190.54		0.22						0.53	0.52	<0.001	28.61	<2.53	105	4
6	(33.28)	33.14	18.61					-212.82					1.44			0.58	0.57	<0.001	66.94	<2.45	202	4
7	61.74	29.69	20.21					-157.22	-1.02							0.55	0.54	<0.001	60.62	<2.45	202	4
8	58.43			31.15	22.61			-365.40		0.22						0.54	0.52	<0.001	28.78	<2.53	105	4
9	72.31	68.42			15.48				-1.47		-2.65					0.62	0.61	<0.001	79.62	<2.45	202	4
10	(14.93)	36.67		10.66				-71.72	-1.1				2.08			0.60	0.59	<0.001	58.72	<2.29	202	5

* There were no ANPR data available on one route (Portswood route) hence less observations.
Coefficients in ( ) are not significant at the 95% level.

Table 6.2-1: Evaluation of alternative regression models (route general journey time)

Measures	*model 1*	*model 2*	*model 3*	*model 4*	*model 5*	*model 6*	*model 7*	*model 8*	*model 9*	*model 10*
Mean Error	-0.0091	-0.0390	**-0.0088**	-0.0127	-2.3634	0.0144	0.1102	-1.6220	-0.0390	-0.0487
Mean Absolute Error	16.1760	15.7959	**15.4284**	15.8098	14.4604	16.1536	17.1071	13.1168	15.7959	15.8014
Mean Absolute Percent Error	0.0971	0.0963	**0.0940**	0.0946	0.0978	0.0975	0.1035	0.0869	0.0963	0.0955
Root Mean Square Error	21.2344	20.4917	**20.2576**	21.2599	18.7843	21.5805	22.1936	16.9660	20.4917	20.9729

Table 6.3: Correlation between bus link general journey time variables

	JT_g	L_{link}	ST_s	ST_f	St_o
JT_g	1				
L_{link}	0.92	1			
ST_s	0.16	0.26	1		
ST_f	0.19	0.24	0.21	1	
St_o	-0.02	-0.05	-0.53	0.54	1

Table 6.4: Summary of bus route general journey time (JT_g) models

No	Response	Constant	L_{link} +	ST_s -	ST_f +	St_o +	LN(L_{link})	Ln(ST_s)	R-square*	P-value	F	Critical value	Obser-vations	N of predictors
1	Y	3.31	0.083						0.84	<0.001	2316.69	<3.02	446	1
2	Y		0.092						0.94	<0.001	6854.57	<3.02	446	1
3	Y				0.38				0.63	<0.001	749.32	<3.02	446	1
4	Y					1.21			0.54	<0.001	521.10	<3.02	446	1
5	Y		0.086		0.038				0.94	<0.001	3566.71	<3.07	446	2
6	Y		0.086			0.16			0.94	<0.001	3715.79	<3.07	446	2
7	Ln(Y)						0.56		0.98	<0.001	24228.77	<3.02	446	1
8	Ln(Y)						0.918	-0.538	0.99	<0.001	24063.88	<3.07	446	2

* For regression without constant, R-square measured the proportion of the variability in predictor about the origin explained by regression. This value cannot be compared to those models including constant.

** Coefficients are all significant at the 95% level

Table 6.5: An example of link bus speed and SCOOT speed data

Order	Route	Surveyor	I/O	Date	Link	length (m)	Entry Time	Leave time	Bus link JT	***BV (kph)***	N of Stops	Dwell time	**ST_s (kph)**
846	Bitterne	Jason	I	13	N07211D	146	130912	130957	45	**11.68**	0	0	**35**
862	Bitterne	Sansaka	I	13	N07211D	146	122437	122524	47	**11.18**	0	0	**35**
866	Bitterne	Sansaka	I	13	N07211D	146	171324	171338	14	**37.54**	0	0	**35**
874	Bitterne	Sansaka	I	15	N07211D	146	165714	165820	66	**7.96**	0	0	**35**
850	Bitterne	Jason	I	13	N07211D	146	173940	174131	111	**4.74**	0	0	**37**
852	Bitterne	Jason	I	14	N07211D	146	132207	132222	15	**35.04**	0	0	**37**
855	Bitterne	Jason	I	14	N07211D	146	160922	160937	15	**35.04**	0	0	**37**
863	Bitterne	Sansaka	I	13	N07211D	146	133330	133414	44	**11.95**	0	0	**37**
851	Bitterne	Jason	I	14	N07211D	146	122200	122243	43	**12.22**	0	0	**38**
858	Bitterne	Jason	I	15	N07211D	146	124947	125048	61	**8.62**	0	0	**38**
854	Bitterne	Jason	I	14	N07211D	146	150902	150940	38	**13.83**	0	0	**40**
857	Bitterne	Jason	I	14	N07211D	146	182357	182410	13	**40.43**	0	0	**40**
868	Bitterne	Sansaka	I	14	N07211D	146	130859	130950	51	**10.31**	0	0	**41**
193	Bitterne	Jason	O	14	N10111B	299	173930	174006	36	**29.90**	0	0	**14**
212	Bitterne	Sansaka	O	15	N10111B	299	172012	172124	72	**14.95**	0	0	**19**
197	Bitterne	Jason	O	15	N10111B	299	165440	165525	45	**23.92**	0	0	**20**
200	Bitterne	Sansaka	O	13	N10111B	299	143350	143446	56	**19.22**	0	0	**20**
182	Bitterne	Jason	O	13	N10111B	299	113311	113346	35	**30.75**	0	0	**22**
198	Bitterne	Sansaka	O	13	N10111B	299	115244	115354	70	**15.38**	0	0	**24**

Table 7.1: Result of paired sample test of route-based regression models

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference: Lower	Upper			
Pair 1	Observed - SCOOT	21.84421	109.79603	17.81127	-14.24485	57.93327	1.226	37	.228
Pair 2	ANPR - Field	-36.88395	101.28264	16.88044	-71.15306	-2.61483	-2.185	35	.036

Table 7.2: Result of paired sample test of dwell time regression models

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference: Lower	Upper			
Pair 1	Observed_dwell_time - Model_dwell_time	-2.20000	27.45505	4.45380	-11.22425	6.82425	-.494	37	.624

Table 7.3: Result of paired sample test of link-based regression models

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Model - Observed	-112.592	129.77996	21.05309	-155.250	-69.93476	-5.348	37	.000

Table 7.4: Result of Smith-Satterthwaite test of route-based general travel time simulation module

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
General travel time	Equal variances assumed	9.817	.002	.699	10019	.484	27.00470	38.62487	-48.70781	102.71720
	Equal variances not assumed			1.371	20.327	.185	27.00470	19.69498	-14.03601	68.04540

Table 7.5: Result of Kolmogorov-Smirnov test of route-based general travel time simulation module

		General travel time
Most Extreme Differences	Absolute	.272
	Positive	.272
	Negative	-.156
Kolmogorov-Smirnov Z		1.246
Asymp. Sig. (2-tailed)		.090

Table 7.6: Result of independent test of route-based dwell time simulation module

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Dwell time	Equal variances assumed	.822	.365	-1.631	10036	.103	-15.61462	9.57591	-34.38532	3.15608

Table 7.7: Result of Kolmogorov-Smirnov test of route-based dwell time simulation module

		Dwell_time
Most Extreme Differences	Absolute	.159
	Positive	.023
	Negative	-.159
Kolmogorov-Smirnov Z		.978
Asymp. Sig. (2-tailed)		.295

Table 7.8: Result of independent test of route-based acceleration/deceleration simulation delay module

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Ac/dec delay	Equal variances assumed	.194	.659	-.418	10019	.676	-2.14922	5.13915	-12.22298	7.92455

Table 7.9: Result of Kolmogorov-Smirnov test of route-based acceleration/deceleration delay simulation module

		Acc/dec delay
Most Extreme Differences	Absolute	.147
	Positive	.099
	Negative	-.147
Kolmogorov-Smirnov Z		.671
Asymp. Sig. (2-tailed)		.759

Table 7.10: Result of Smith-Satterthwaite test of route-based bus journey time simulation delay model

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
JT	Equal variances assumed	7.744	.005	-.229	10036	.819	-7.01854	30.64141	-67.08184	53.04477
	Equal variances not assumed			-.349	37.659	.729	-7.01854	20.12566	-47.77290	33.73583

Table 7.11: Result of Kolmogorov-Smirnov test of route-based bus journey time simulation model

		JT
Most Extreme Differences	Absolute	.177
	Positive	.177
	Negative	-.162
Kolmogorov-Smirnov Z		1.089
Asymp. Sig. (2-tailed)		.187

Table 7.12: Details of Bitterne inbound links

Link No.	SCOOT link ID	Link length (m)	Accumulated length (m)	Note
1	N10361A	652	652	
2	N10351E	278	930	
3	N10341D	681	1,611	
4	N10331E	116	1,727	
5	N10321D	82	1,809	
6	N10121D	805	2,614	
7	N10111G	792	3,406	
8	N10241A	318	3,724	
9	N10231E	440	4,164	
10	N10221F	512	4,676	
11	N10214D	211	4,887	
12	N07221E	550	5,437	
13	N07211D	183	5,620	
14		1,209	6,829	No SCOOT link available

Table 7.13: Result of Smith-Satterthwaite test of link-based general travel time combined with signal delay module

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
JTg+ Sig	Equal variances assumed	7.816	.005	2.842	10019	.004	35.61410	12.53235	11.04817	60.18002
	Equal variances not assumed			1.815	20.034	.085	35.61410	19.62370	-5.31575	76.54394

Table 7.14: Result of Kolmogorov-Smirnov test of link-based general travel time combined with signal delay module

		JTg+Sig
Most Extreme Differences	Absolute	.295
	Positive	.295
	Negative	-.084
Kolmogorov-Smirnov Z		1.349
Asymp. Sig. (2-tailed)		.053

Table 7.15: Result of independent t-test of link-based dwell time module

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Dwell time	Equal variances assumed	1.421	.233	.000	9583	1.000	.00359	11.87240	-23.26883	23.27601

Table 7.16: Result of Kolmogorov-Smirnov test of link-based dwell time module

		Dwell_time
Most Extreme Differences	Absolute	.098
	Positive	.098
	Negative	-.097
Kolmogorov-Smirnov Z		.601
Asymp. Sig. (2-tailed)		.863

Table 7.17: Result of independent t-test of link-based acceleration/deceleration delay module

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Acc_dec delay	Equal variances assumed	2.080	.149	-.906	10019	.365	-3.88543	4.28837	-12.29149	4.52063

Table 7.18: Result of Kolmogorov-Smirnov test of link-based acceleration/deceleration delay module

		acc_dec_delay
Most Extreme Differences	Absolute	.221
	Positive	.081
	Negative	-.221
Kolmogorov-Smirnov Z		1.010
Asymp. Sig. (2-tailed)		.259

Table 7.19: Result of independent t-test of link-based journey time model

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Journey time	Equal variances assumed	1.561	.212	.197	10036	.844	3.49882	17.78584	-31.36499	38.36264

Table 7.20: Result of Kolmogorov-Smirnov test of link-based journey time model

		Journey time
Most Extreme Differences	Absolute	.122
	Positive	.105
	Negative	-.122
Kolmogorov-Smirnov Z		.750
Asymp. Sig. (2-tailed)		.628

Table 7.21: Result of test of route and link based journey time model

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Route-based journey time	Equal variances assumed	.473	.492	-.689	10037	.491	-14.47620	20.99871	-55.63788	26.68548
	Equal variances not assumed			-.684	38.292	.498	-14.47620	21.16388	-57.30950	28.35710
Link-based journey time	Equal variances assumed	21.398	.000	-1.400	10037	.162	-17.43575	12.45495	-41.84995	6.97845
	Equal variances not assumed			-.825	38.102	.415	-17.43575	21.13753	-60.22269	25.35119

Table 7.22: Result of Kolmogorov-Smirnov tests of route- and link-based journey time model

		Route-based	Link-based
Most Extreme Differences	Absolute	.143	.215
	Positive	.098	.093
	Negative	-.143	-.215
Kolmogorov-Smirnov Z		.892	1.342
Asymp. Sig. (2-tailed)		.404	.055

Table 7.23: Result of independent test of dwell time model

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Dwell Time	Equal variances assumed	1.535	.215	.148	10037	.882	1.32988	8.97247	-16.25797	18.91772
	Equal variances not assumed			.176	38.422	.861	1.32988	7.53646	-13.92138	16.58114

Table 7.24: Result of Kolmogorov-Smirnov test of dwell time model

		DT
Most Extreme Differences	Absolute	.130
	Positive	.130
	Negative	-.066
Kolmogorov-Smirnov Z		.809
Asymp. Sig. (2-tailed)		.529

Table 8.1: The key variables in parametric analysis

Variable	Bus journey time (s)			Input value		
	Downside	*Upside*	*Range*	*Downside*	*Upside*	*Base Case*
$N_{stopped}$ (DT module)	753.09	1025.49	272.4	1	9	5
N07311I Red Period (s)	850.44	928.14	77.70	4.32	82.01	43.17
N03111E Red Period (s)	850.56	928.02	77.45	4.30	81.76	43.03
$N_{stopped}$ (D_{acc} module)	852.85	925.73	72.87	1	9	5
N04141E Red Period (s)	863.71	914.87	51.16	2.84	54.00	28.42
N07331I Probability of Red	846.13	889.29	43.17	0	1	1
N03111E Probability of Red	846.26	889.29	43.03	0	1	1
Tbi	881.82	912.18	30.36	3.63	33.99	11.10
N04141E Probability of Red	860.87	889.29	28.42	0	1	1
V_{bus}	905.99	878.60	27.39	35.13	54.87	45

Table 8.2: Example of reverse sensitivity analysis (bus journey time between 1200 and 1300 s)

Bus journey time(s)		1201.93	1205.41	1210.99	1216.35	1234.80	1246.42
Input Parameters	D1	1.52	1.27	0.93	1.17	1.59	1.86
	A1	1.20	0.53	0.67	0.59	0.59	0.85
	D2	1.36	1.78	0.64	0.93	1.00	1.40
	A2	1.07	1.37	0.76	0.74	0.86	1.10
	D3	0.98	0.86	1.28	1.66	1.78	1.25
	A3	0.89	1.07	0.93	0.51	1.30	0.87
	D4	2.04	1.28	0.87	1.04	1.55	1.27
	A4	0.85	1.17	0.98	1.34	1.16	0.95
	D5	1.17	1.01	0.60	1.08	1.89	1.75
	A5	1.33	1.01	0.82	0.74	0.59	0.93
	D6	1.12	0.79	1.12	0.92	1.81	2.39
	A6	0.79	0.89	1.36	0.60	0.49	0.86
	D7	1.10	1.56	0.65	0.78	0.85	0.83
	A7	0.99	0.88	1.39	1.14	1.03	0.74
	D8	1.60	1.18	1.68	1.35	0.84	1.40
	A8	1.16	0.74	0.67	0.98	1.17	1.06
	D9	1.26	0.80	0.77	1.45	0.91	0.91
	A9	0.65	0.84	0.64	0.71	1.03	0.64
	D10	1.66	0.64	1.56	1.05	1.53	1.79
	A10	0.86	0.81	0.65	1.08	1.12	1.31
	D11	1.91	1.58	2.98	0.94	1.05	1.43
	A11	0.80	0.66	0.76	0.57	0.63	0.93
	Nstopped	8.00	11.00	11.00	9.00	8.00	9.00
	Na1	0.00	3.00	0.00	0.00	1.00	0.00
	Nb1	1.00	1.00	1.00	1.00	0.00	3.00
	Ta1	5.23	5.15	3.69	1.63	3.13	2.65
	Tb1	3.44	3.45	23.90	6.52	20.89	81.71
	Na2	2.00	0.00	1.00	1.00	0.00	3.00
	Nb2	0.00	2.00	2.00	0.00	2.00	0.00
	Ta2	7.17	4.06	4.72	3.48	2.42	4.44
	Tb2	6.59	17.21	11.83	14.65	7.03	9.05
	Na3	0.00	0.00	0.00	1.00	0.00	0.00
	Nb3	3.00	3.00	1.00	0.00	1.00	1.00
	Ta3	2.00	2.84	4.50	14.60	1.73	10.48

Table 8.2: Example of reverse sensitivity analysis (bus journey time between 1200 and 1300 s) – continued

Bus journey time(s)		1201.93	1205.41	1210.99	1216.35	1234.80	1246.42
Input Parameters	Tb3	9.04	21.81	10.01	21.54	15.19	13.73
	Na4	1.00	0.00	0.00	2.00	2.00	2.00
	Nb4	0.00	1.00	2.00	0.00	1.00	0.00
	Ta4	3.84	4.12	4.25	5.45	2.96	20.98
	Tb4	5.11	14.83	7.61	21.87	22.86	9.45
	Na5	3.00	1.00	3.00	1.00	1.00	1.00
	N5	0.00	0.00	0.00	1.00	0.00	2.00
	Ta5	12.41	1.27	4.35	12.15	4.96	2.75
	Tb5	11.15	8.73	38.61	2.96	8.27	20.83
	Na6	0.00	6.00	1.00	6.00	1.00	1.00
	Nb6	1.00	1.00	1.00	1.00	1.00	0.00
	Ta6	3.76	1.54	6.01	11.22	3.91	5.41
	Tb6	27.80	18.68	5.75	38.35	15.70	14.39
	Na7	1.00	1.00	1.00	1.00	0.00	2.00
	Nb7	0.00	0.00	1.00	0.00	1.00	1.00
	Ta7	6.13	1.84	2.26	2.76	6.12	4.01
	Tb7	5.23	22.20	11.62	7.52	12.31	12.20
	Na8	1.00	3.00	1.00	1.00	1.00	0.00
	N8	0.00	3.00	1.00	3.00	0.00	1.00
	Ta8	7.79	1.52	3.93	3.60	2.07	3.82
	Tb8	2.26	7.58	21.33	57.08	22.34	8.29
	Na9	1.00	0.00	0.00	1.00	1.00	0.00
	Nb9	0.00	1.00	2.00	1.00	0.00	1.00
	Ta9	8.02	6.95	4.19	14.06	5.94	5.31
	Tb9	5.50	20.12	18.96	4.82	8.94	11.66
	Na10	0.00	0.00	1.00	0.00	0.00	1.00
	Nb10	1.00	2.00	0.00	1.00	1.00	0.00
	Ta10	5.79	10.37	2.90	1.59	7.84	7.66
	Tb10	19.62	27.70	12.56	4.27	22.44	18.59
	Na11	1.00	0.00	1.00	0.00	1.00	0.00
	Nb11	0.00	1.00	1.00	1.00	0.00	1.00
	Ta11	1.92	20.36	1.57	11.12	10.69	2.48
	Tb11	16.45	82.68	3.56	11.63	6.75	15.34
	Vbus	36.42	37.79	44.91	33.80	40.33	46.39
	N03111E-STs	36.39	32.66	38.36	38.27	32.11	32.60
	N03112C-STs	24.93	24.59	21.44	23.45	25.44	23.92
	N03214A-STs	43.11	38.15	38.79	45.72	33.96	42.34
	N03244A-STs	1.91	49.96	4.25	44.76	1.98	53.32
	N03244E-STs	62.00	55.33	40.26	65.88	48.88	39.22
	N03311J-STs	52.82	57.67	48.23	46.81	54.77	52.63
	N03311L-STs	47.40	50.26	46.34	46.34	44.50	32.36
	N04141E-STs	28.59	38.80	34.44	33.90	36.02	31.31
	N04144C-STs	40.56	41.99	39.46	38.16	37.70	38.15
	N4155A-STs	30.91	41.31	37.22	35.99	32.52	34.64
	N04157K-STs	20.79	24.59	21.19	28.99	21.32	18.88
	N04158K-STs	35.82	28.97	29.95	28.36	38.25	37.52
	N04158XA-STs	28.48	37.47	31.40	33.35	32.78	25.83
	N07331I-STs	21.03	27.81	21.25	25.44	28.54	27.38
	N03214A Probability of Red	0.00	0.00	1.00	0.00	0.00	1.00
	N03214A Red Period (s)	2.66	6.99	3.95	3.59	9.94	8.47
	N03244A Probability of Red	0.00	0.00	0.00	0.00	0.00	0.00
	N03244A Red Period (s)	3.19	13.62	12.60	4.74	7.17	9.14
	N03311L Probability of Red	0.00	1.00	0.00	0.00	0.00	0.00
	N03311J Red Period (s)	5.49	12.25	8.87	11.47	15.55	2.60
	N03111E Probability of Red	1.00	1.00	0.00	1.00	1.00	0.00
	N03111E Red Period (s)	57.41	39.01	63.00	66.01	33.26	48.55
	N03112C Probability of Red	0.00	0.00	1.00	0.00	0.00	1.00
	N03112C Red Period (s)	3.75	26.65	33.77	5.99	17.32	34.89
	N04144C Probability of Red	0.00	0.00	1.00	0.00	0.00	0.00
	N04144C Red Period (s)	12.66	1.56	11.05	13.91	10.24	4.17
	N04141E Probability of Red	0.00	1.00	1.00	1.00	1.00	1.00
	N04141E Red Period (s)	3.24	38.54	17.07	51.31	54.16	34.26
	N04157K Probability of Red	0.00	0.00	0.00	0.00	0.00	1.00
	N04157K Red Period (s)	9.41	2.27	17.10	15.08	7.51	16.80
	N04155A Probability of Red	0.00	0.00	0.00	1.00	1.00	0.00
	N04155A Red Period (s)	14.35	16.79	14.29	1.67	7.19	9.29
	N04158X Probability of Red	0.00	0.00	0.00	0.00	0.00	0.00
	N04158X Red Period (s)	8.26	5.76	14.49	3.83	14.46	1.10
	N07331I Probability of Red	1.00	1.00	1.00	1.00	1.00	1.00
	N07311I Red Period (s)	1.74	67.86	73.29	65.28	56.78	73.01

Table 8.3: Comparison of changes in passenger demand scenarios

	Mean value per run (s)			
Scenarios	Dwell time	Acc/dec delay	Signal delay	Total bus journey time
Base case	78.64	46.53	103.85	812.89
(1)	+73.05 (+93%)	+42.56 (+91%)	-	+115.22 (+14%)
(2)	+112.47 (+143%)	-	-	+112.62 (+14%)

Table 8.4: Signal staging and timing of junction N07331 (junction of London Road and Brunswick Place)

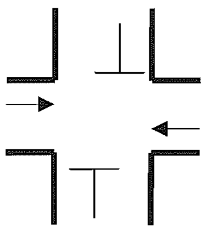
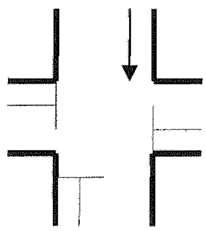
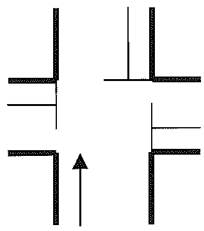
	1			2			3		
Signal Staging				 Target stage					
Signal Timing (s)	Green time	Inter-green time	Length	Green time	Inter-green time	Length	Green time	Inter-green time	Length
Origin setting	43.62	12.97	56.59	19.03	9.00	28.03	10.74	10.00	20.74
Scenario	43.62	12.97	56.59	**39.03**	9.00	28.03	10.74	10.00	20.74

Table 8.5: Comparison of change in signal timing scenario

	Mean value per run (s)			
Scenarios	Dwell time	Acc/dec delay	Signal delay	Total bus journey time
Base case	78.64	46.53	103.85	812.89
Extending green time	-	-	-5.01 (-5 %)	-5.23 (-0.64 %)

Table 8.6: Comparison of dwell time components

		a	b	c
Literature	Levinson (1983)	5	2.75*	
	Guenthner and Sinha (1983)	10~20	3~5*	
	York (1993)	3.55~5.42	1.48~1.99	9.15~9.18
	HCM(2000)	2~5	1.0~2.0	1.2~3.0
	Shrestha (2002)	3.30~6.85	1.69~1.96	9~9.04
	TCRP(2003)	2~5	1.4~3.7	2.25~4.3
	Bertini and El-Geneidy (2004)	N/A	0.85	3.6
	Dueker et al.(2004)	5.14	1.7	3.48
This study	Regression model	5.07	1.19	8.88
	Monte Carlo model	N/A	Lognormal distribution (μ=6.07, σ=4.98), range 0.67~55 s	Lognormal distribution (μ=13.99, σ=10.73), range 1~94 s

Note *: per boarding or alighting passenger

Table 8.7: Recording sheet of passenger boarding time survey

Dwell Time and Passenger Data Sheet

Sheet No:

- **Direction**:____(I: to City Centre; O: away from City Centre)
- **Bus Number**:______.
- **Bus Type**: Is bus low floor? Yes/No; Is bus double-decker? Yes/No; How many doors: ____.
- **Boarding Time**: (hh:mm:ss); **Alighting time**: .

Stop order	Arrival time (mm:ss)	Off	On				Departure time (mm:ss)	Note
			Show card or receipt	Buy ticket no change	Buy ticket with change	Others		
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								

Table 8.8: Example of collected data of passenger boarding time survey

					Bus types			Dwell time & Number of alighting and boarding passengers					Fare collection methods				
No.	Route	Direction	Date	Bus Number	Low floor	Double_decker	Num_doors	Stop time	N of alighting	N of boarding	Start time	Dwell time	Show card or receipt	Buy ticket no change	Buy ticket with change	others	Note
1	Bitterne Rd	O	15	5A	Y	N	1	749		1	753	4	1				
2	Bitterne Rd	O	15	5A	Y	N	1	901		1	917	16	1				
3	Bitterne Rd	O	15	5A	Y	N	1	1034	1	1	1040	6	1				
4	Bitterne Rd	O	15	5A	Y	N	1	1223		1	1240	17		1			
5	Bitterne Rd	O	15	5A	Y	N	1	1318	2	2	1338	20		2			
6	Bitterne Rd	O	15	5A	Y	N	1	1449		2	1520	31		2			
7	Bitterne Rd	O	15	5A	Y	N	1	1633	2		1644	11					
8	Bitterne Rd	O	15	5A	Y	N	1	1754	1		1757	3					
9	Bitterne Rd	O	15	5A	Y	N	1	1841	5	4	1916	35	1	2	1		
10	Bitterne Rd	O	15	5A	Y	N	1	2230	1		2239	9					

Table 8.9: Descriptive statistic of boarding time per passenger for different fare collection methods

	Pre-paid ticket	*Cash without change*	*Cash with change*
Mean	7.82	14.65	16.13
Standard Deviation	4.96	6.84	10.33
Sample Variance	24.59	46.78	106.73
Minimum	3	7	8
Maximum	23	24	31
Count	19	10	4
Confidence Level (95.0%)	2.39	4.89	16.44

Table 8.10: Comparison of change in boarding time scenario

	Mean value per run (s)			
Scenarios	Dwell time	Acc/dec delay	Signal delay	Total bus journey time
Base case	78.64	46.53	103.85	812.89
Decrease in boarding time	-34.48 (-44%)	-	-	-34.48 (-4 %)

Table 8.11: Details of link general travel time data

Links	Contribution to variance %	Speed variable	Link length (m)	Mean of SCOOT speed (KPH)	Note
Chilworth	36.4	Vbus	743	45*	no SCOOT link available
N07331I	12.8	N07331I-STs	351	27	SCOOT link
N04144C	12.6	N04144C-STs	1,467	39	
N03244E	7	N03244E-STs	315	51	
N03311L	6.2	N03311L-STs	131	42	
N03244A	5.1	N03244A-STs	622	49	
N04157K	4.9	N04157K-STs	349	22	
N03111E	3.8	N03111E-STs	326	38	
N03214A	2.3	N03214A-STs	192	42	
N04141E	2.1	N04141E-STs	290	35	
N04158X	1.8	N04158X-STs	145	32	
N04155A	1.7	N04155A-STs	138	34	
N04158K	1.7	N04158K-STs	193	35	
N03112C	1.3	N03112C-STs	163	24	
N03311J	0.2	N03311J-STs	85	56	

* Assumed bus mean speed

Table 8.12: Comparison of change for bus priority

	Mean value per run (s)			
Scenarios	Dwell time	Acc/dec delay	Signal delay	Total bus journey time
Base case	78.64	46.53	103.85	812.89
Bus priority	-	+1.63 (+3%)	-	-15.12 (-2 %)

Table 9.1: Summary of independent variables used in the result models

Approach / Key Component	Regression		Monte Carlo simulation	
	Route-based	Link-based	Route-based	Link-based
General travel time	• Number of disturbances per km • Number of bus-stops per km • Number of stopped bus-stops per km • Percentage of bus lane length • SCOOT occupancy and speed parameters • ANPR journey time per km • Timetable adherence	• Link length • SCOOT flow parameters	• Route length • Percentage of ANPR journey time per km • ANPR journey time per km	• Link length • SCOOT speed parameter
Acc/dec delay		• Number of stopped bus-stops per km • Bus cruise speed before deceleration • Bus cruise speed after acceleration • Acceleration rate • Deceleration rate	• Number of stopped bus-stops per km • Percentage of SCOOT speed parameter • SCOOT speed parameter • Acceleration rate • Deceleration rate	• Number of stopped bus-stops per km • SCOOT speed parameter • Acceleration rate • Deceleration rate
Signal delay		• Average cycle length • Average green time	This part was included in general travel time.	• Number of signalized junctions • Probability of a bus encountered a red signal at a junction • Duration of red at a junction
Dwell time	• Number of stopped bus-stops • Number of alighting passengers • Number of boarding passengers	• Number of stopped bus-stops Number of alighting passengers • Number of boarding passengers	• Number of stopped bus-stops • Number of alighting passengers at a stop • Number of boarding passengers at a stop • Alighting time of each alighting passenger at a stop • Boarding time of each boarding passenger at a stop	• Number of stopped bus-stops • Number of alighting passengers at a stop • Number of boarding passengers at a stop • Alighting time of each alighting passenger at a stop • Boarding time of each boarding passenger at a stop

Table 9.2: Comparison of bus journey time rate with other vehicles

Bus journey time = **k*** (other vehicle journey time)	
Studies	k value
Levinson (1983)	1.4-1.6
McKnight and Paaswell (1997)	1.75
McKnight et al. (2004)	1.37
This study	1.34

FIGURES

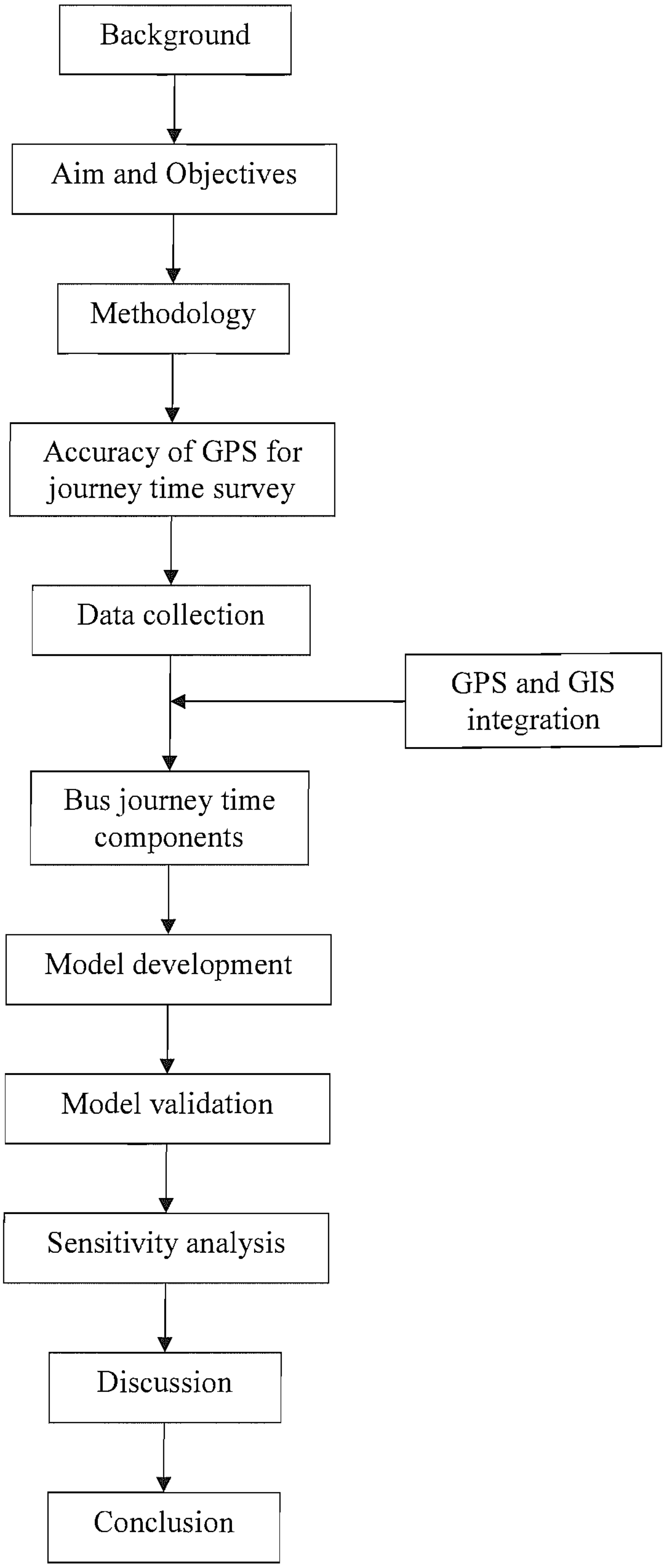


Figure 1.1: Outline of the research

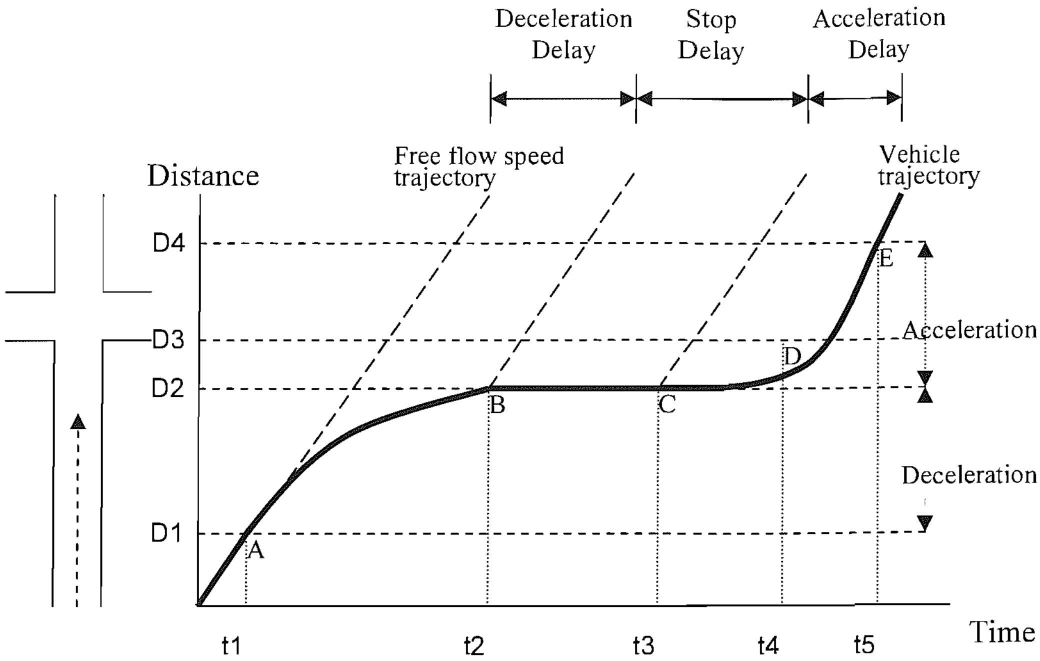


Figure 2.1: Distance-time diagram at signalised junction

(source from Quiroga and Bullock (1999))

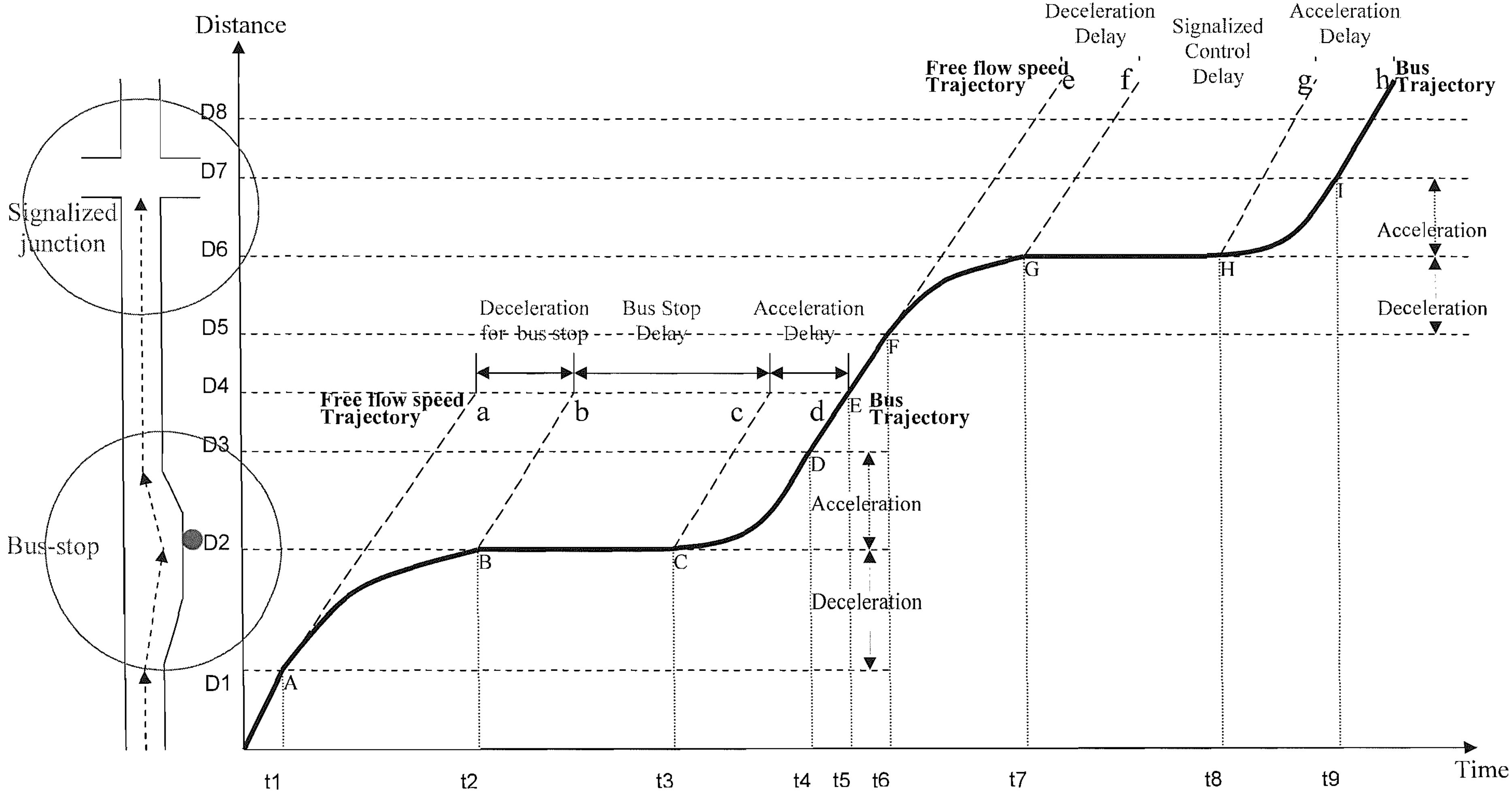


Figure 2.2: Distance-time diagram of a bus travelling along a bus stop and a signalised junction

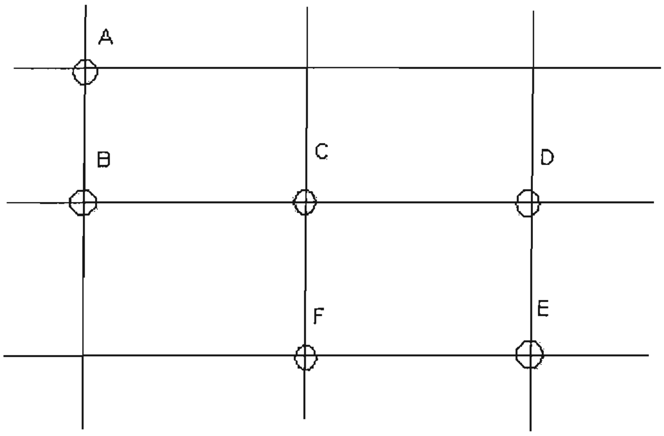


Figure 2.3: A simplified traffic network for explaining link journey time

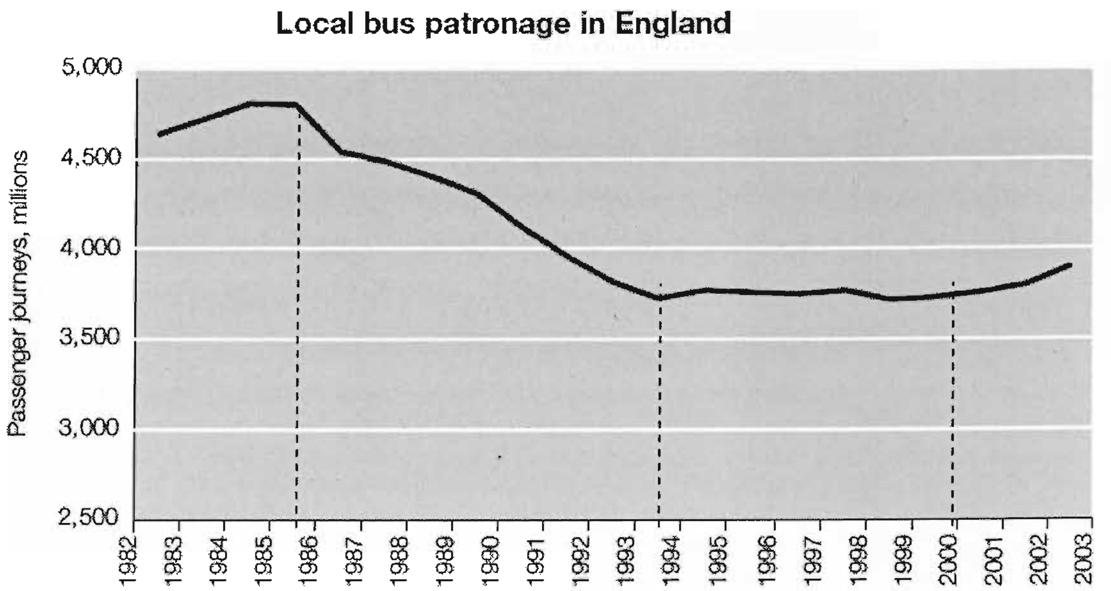


Figure 2.4: Local bus passengers in England (source from Dft (2003))

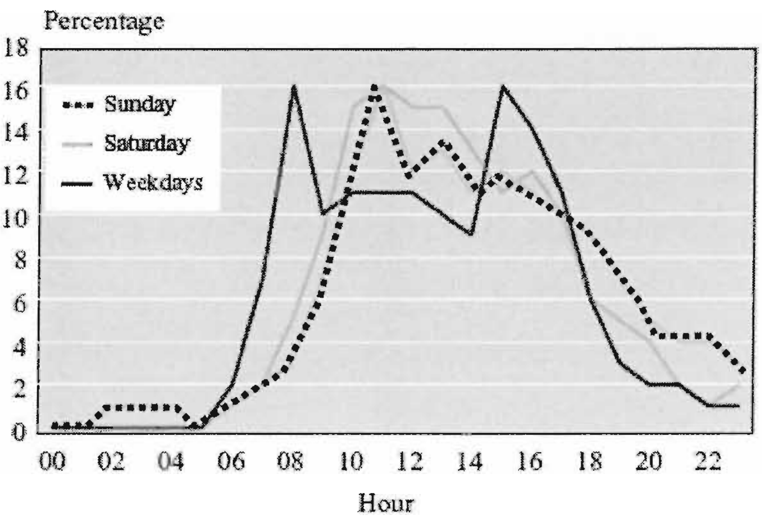


Figure 2.5: Patterns of bus use during weekdays and weekend
(source from DfT(2003))

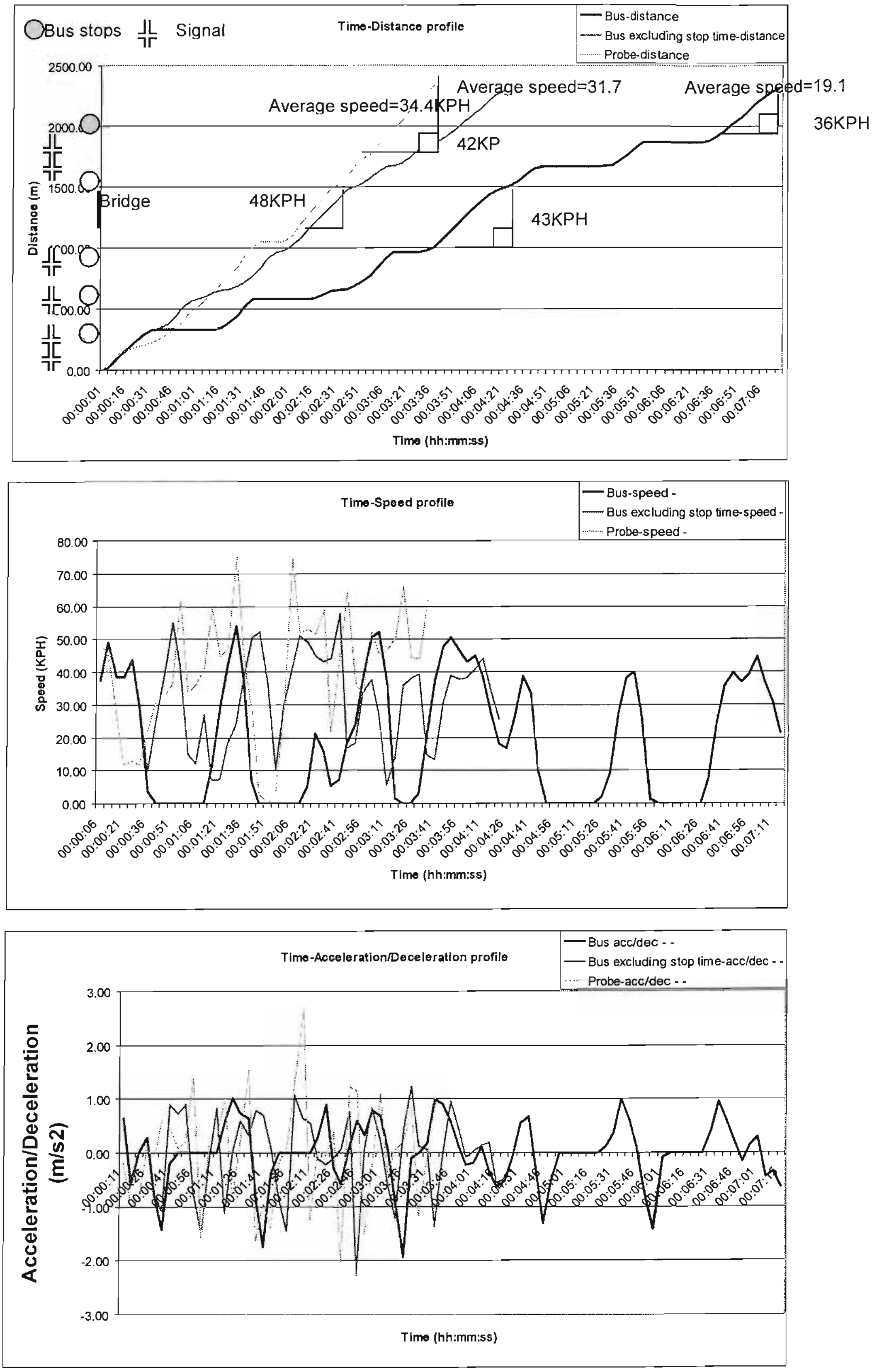


Figure 2.6: Time against distance, speed, and acceleration/deceleration profile
(data from the First 5/5A bus and the probe vehicle on Bitterne outbound road in Southampton on 12 January 2005, bus start at 7:27:16 and probe vehicle start at 7:22:20 respectively)

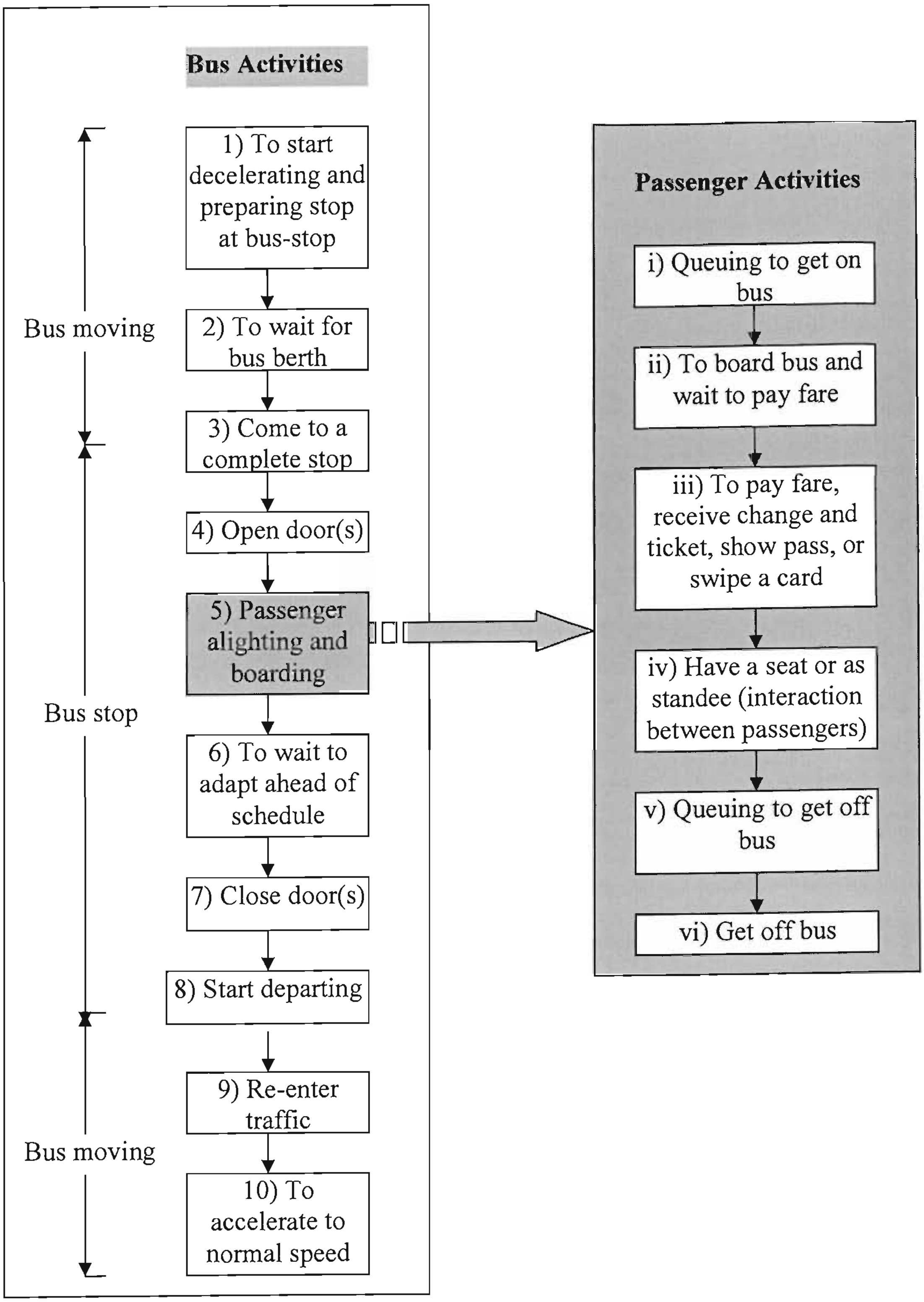


Figure 2.7: Bus activities along a bus-stop and passenger activities at a bus-stop

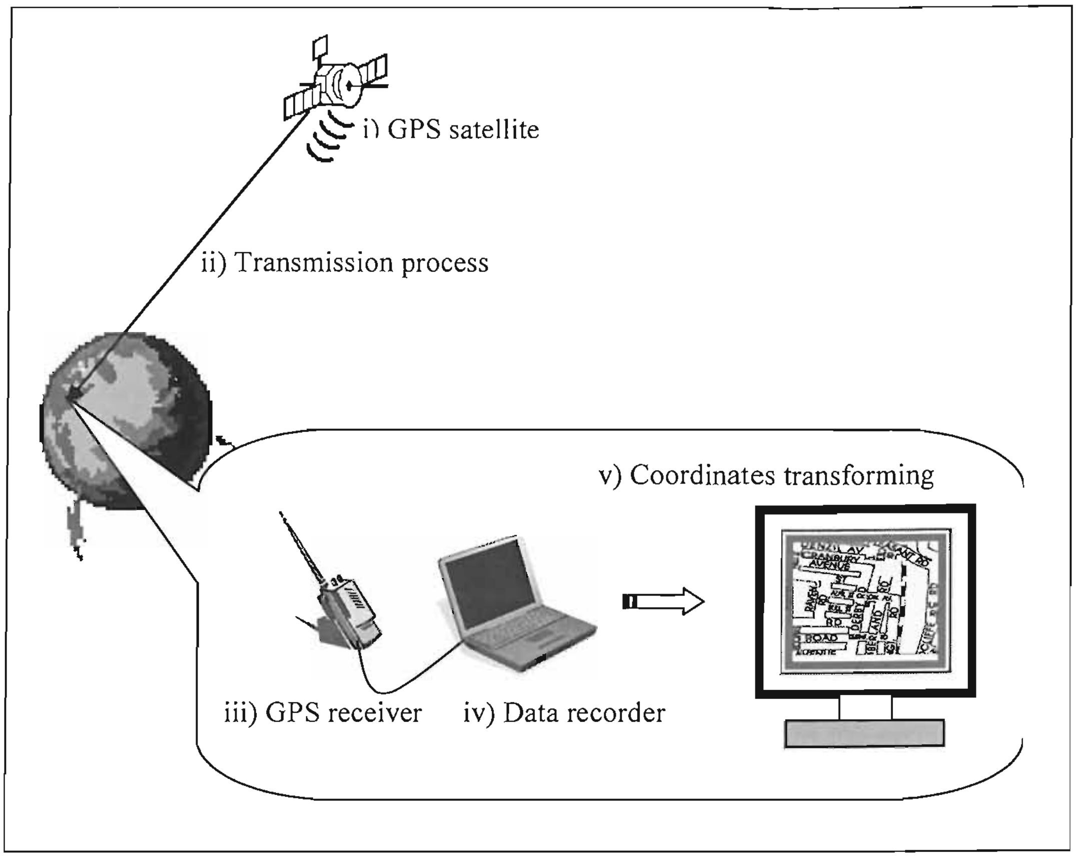


Figure 3.1: Error sources of GPS data

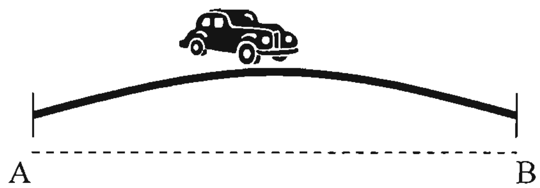


Figure 3.2: The difference of distance of straight line and curve between A, B

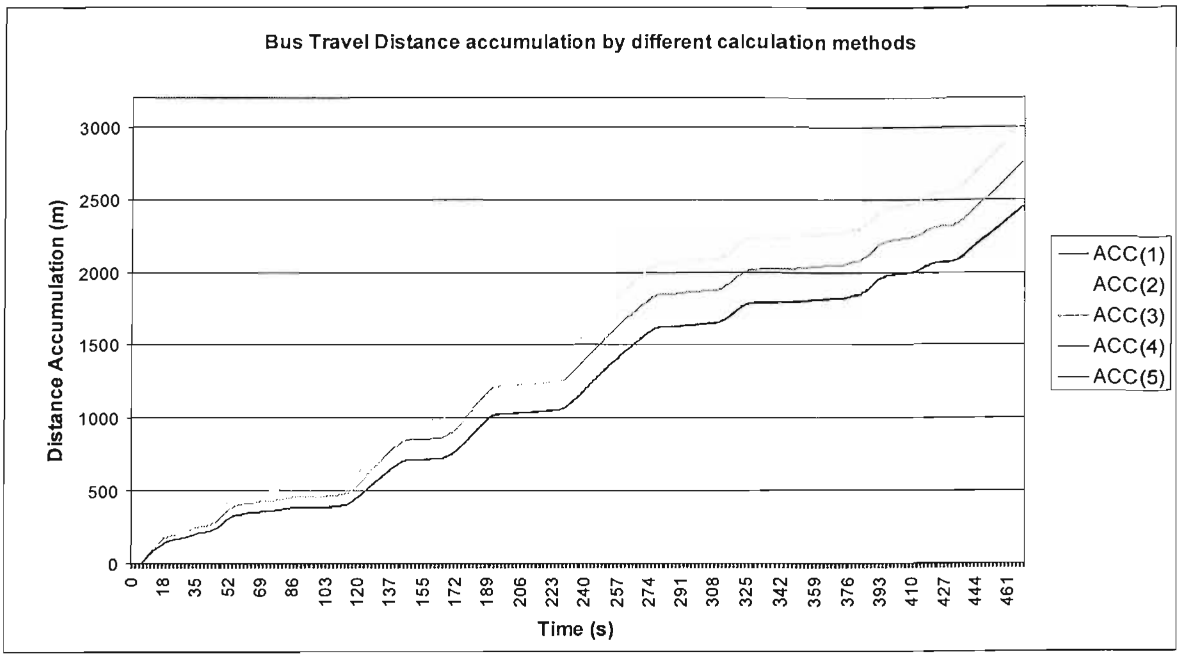


Figure 3.3: Total distance calculated by comparative equations

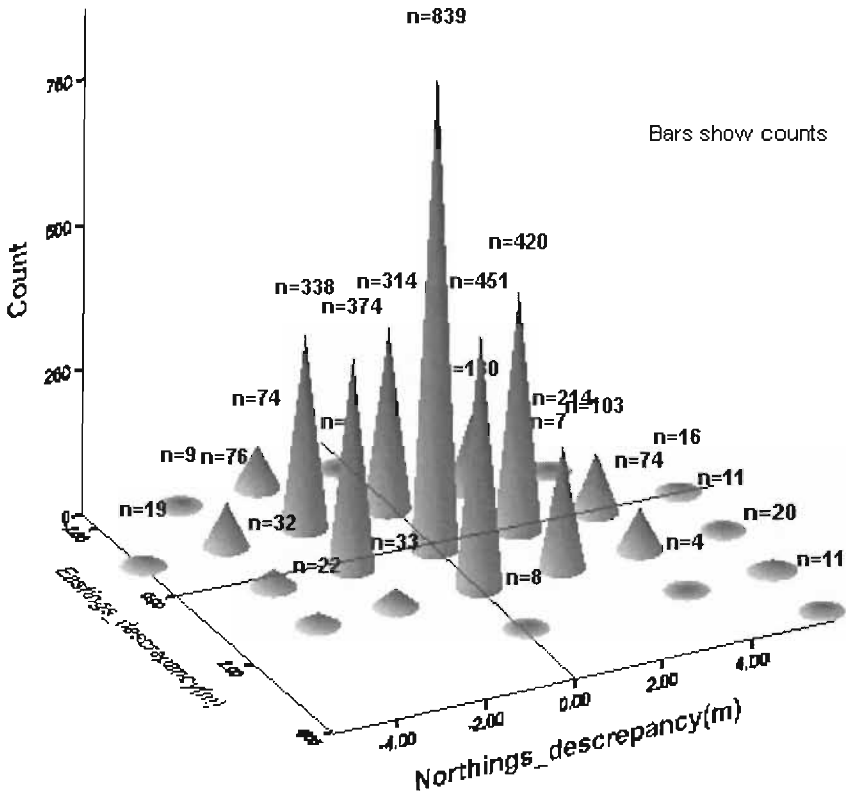


Figure 3.4: Scatters of Eastings and Northings discrepancy and frequency counts

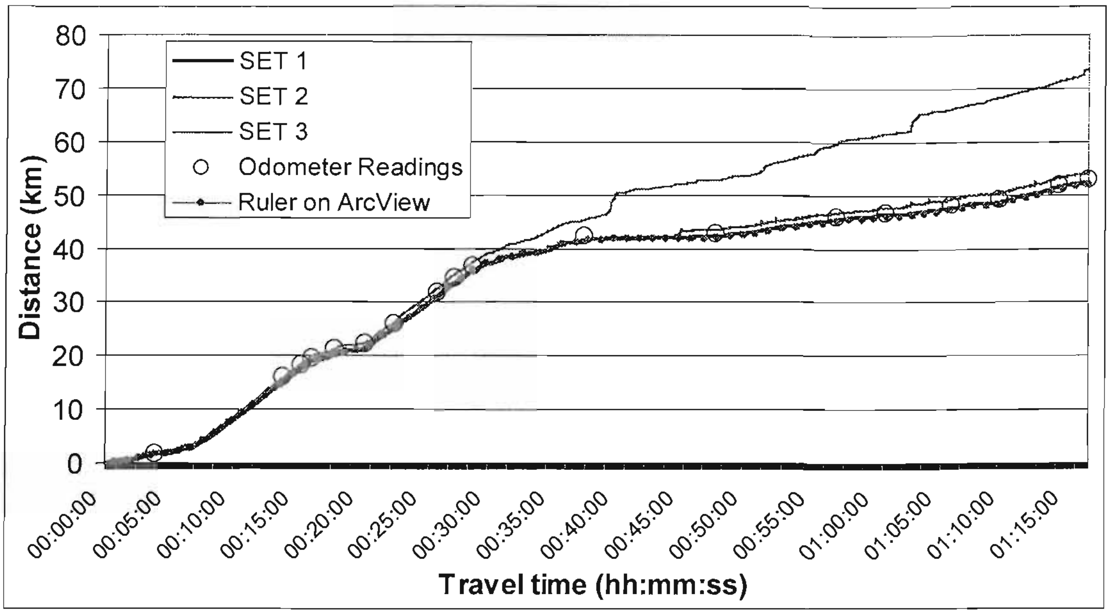
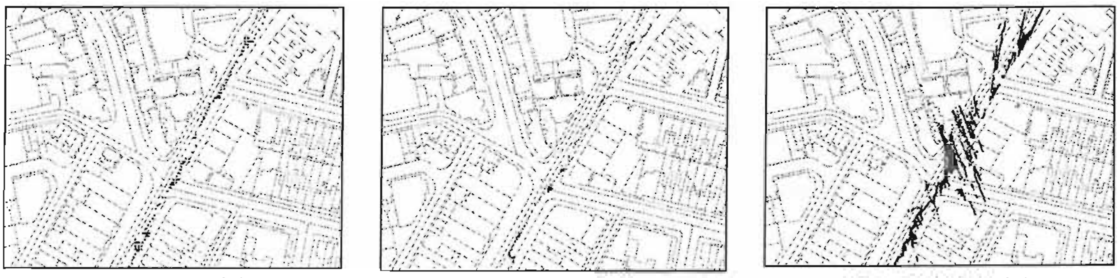


Figure 3.5: Accumulated distance comparison of various approaches



Set 1 (1Hz) Set 2 (1Hz) Set 3 (100Hz)

Figure 3.6: GPS points on digital map for different GPS sets (logging frequency was shown in round brackets)

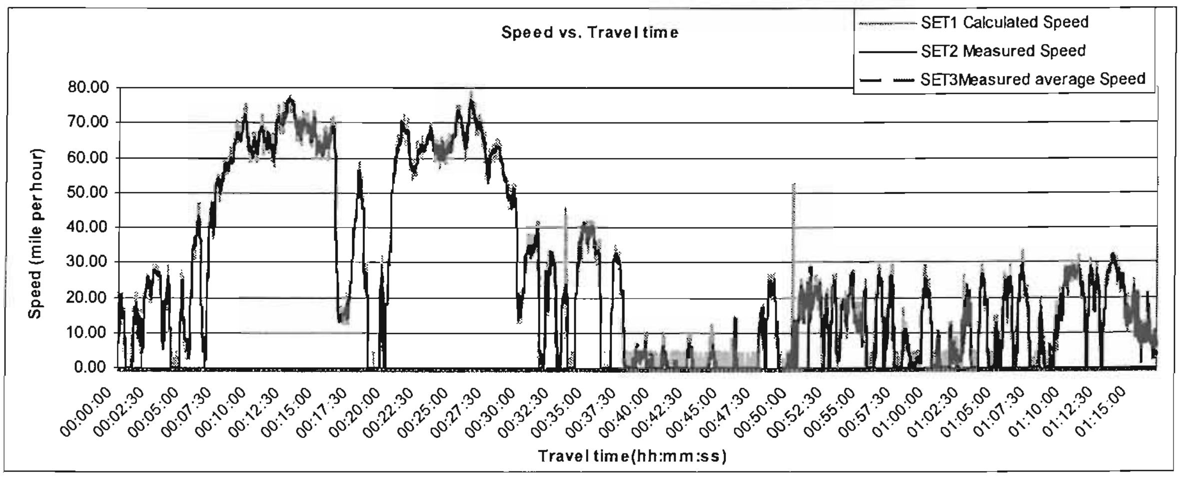


Figure 3.7: Speeds comparison of different GPS sets

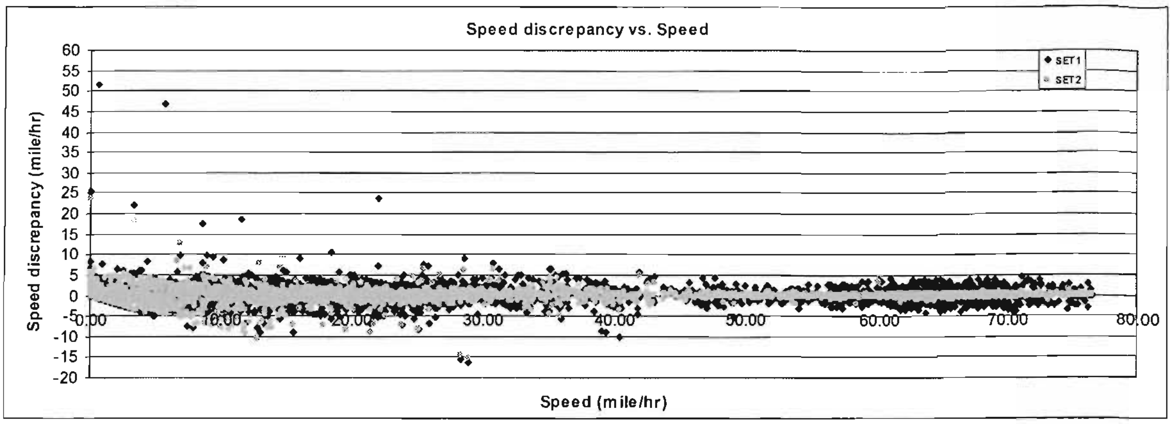


Figure 3.8: GPS speed discrepancy vs. vehicle speed (set 3 as standard)

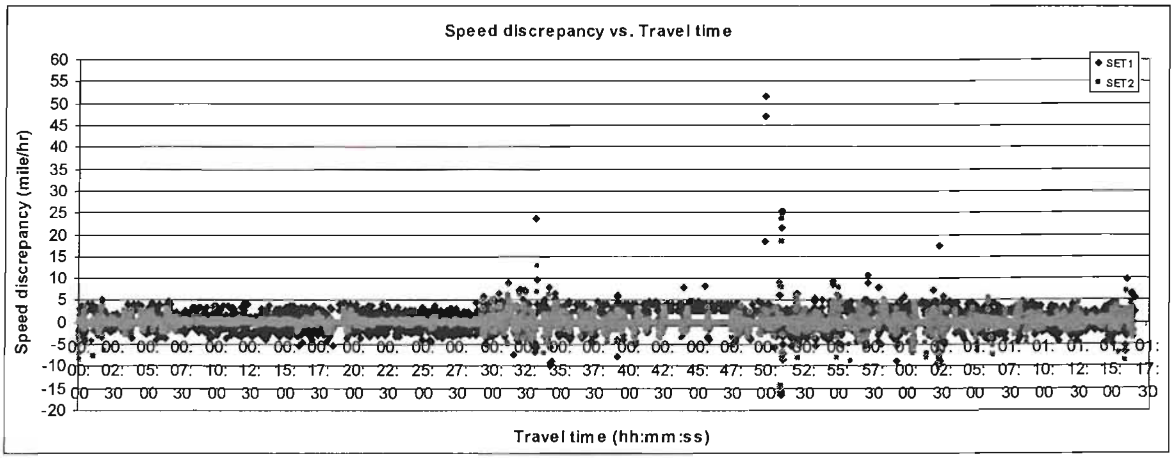


Figure 3.9: GPS speed discrepancy vs. travel time (set 3 as standard)

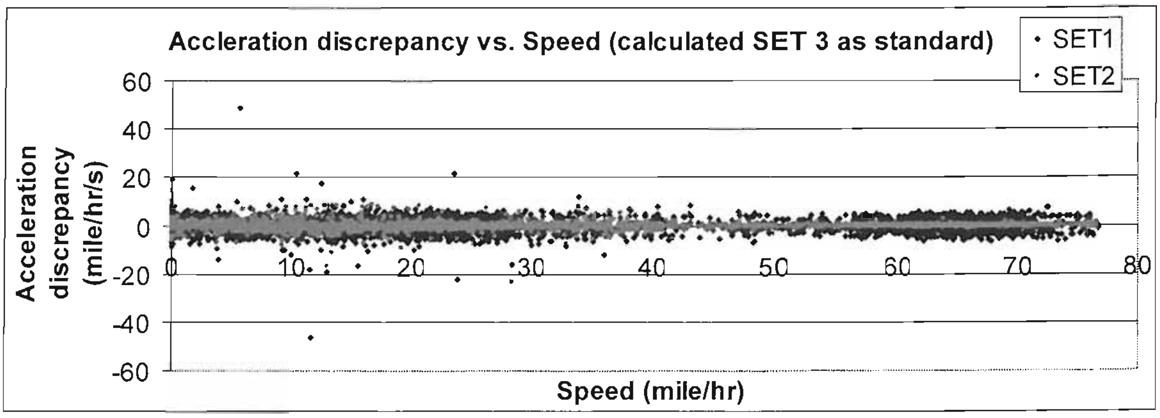


Figure 3.10: Acceleration vs. travel speed (set 3 as standard)

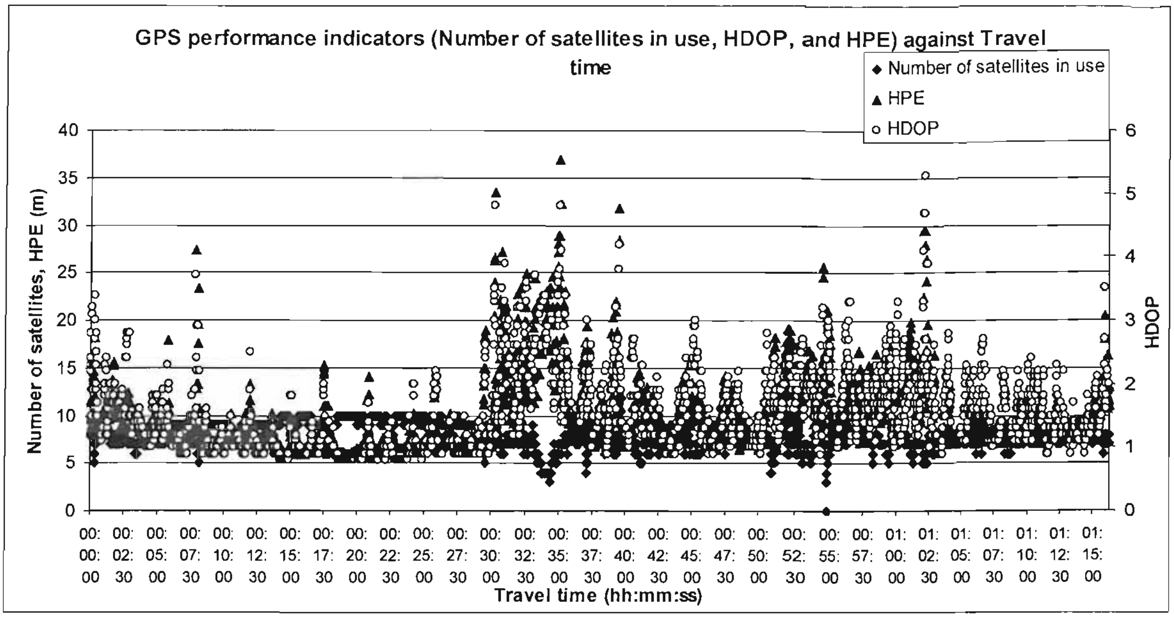


Figure 3.11: GPS performance indicators against travel time

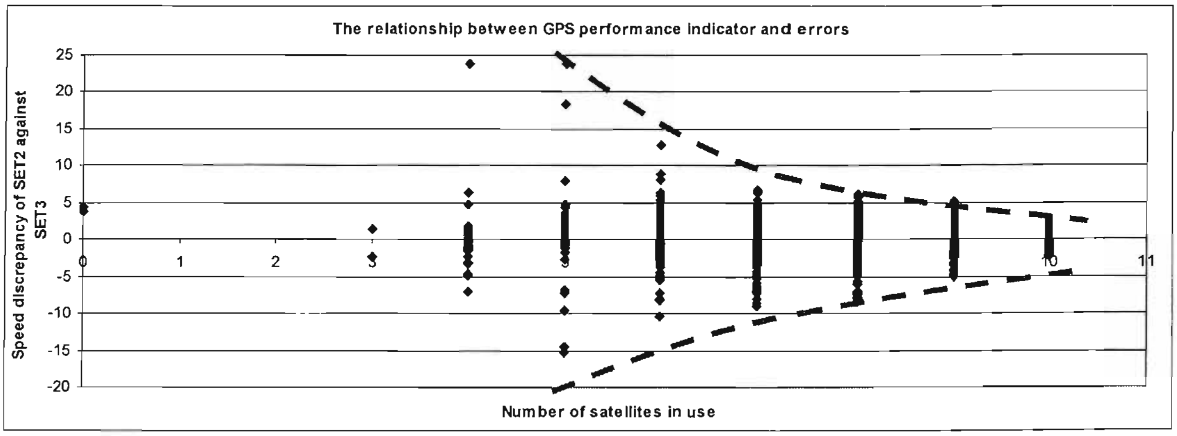


Figure 3.12: The relationship between GPS performance indicator and error

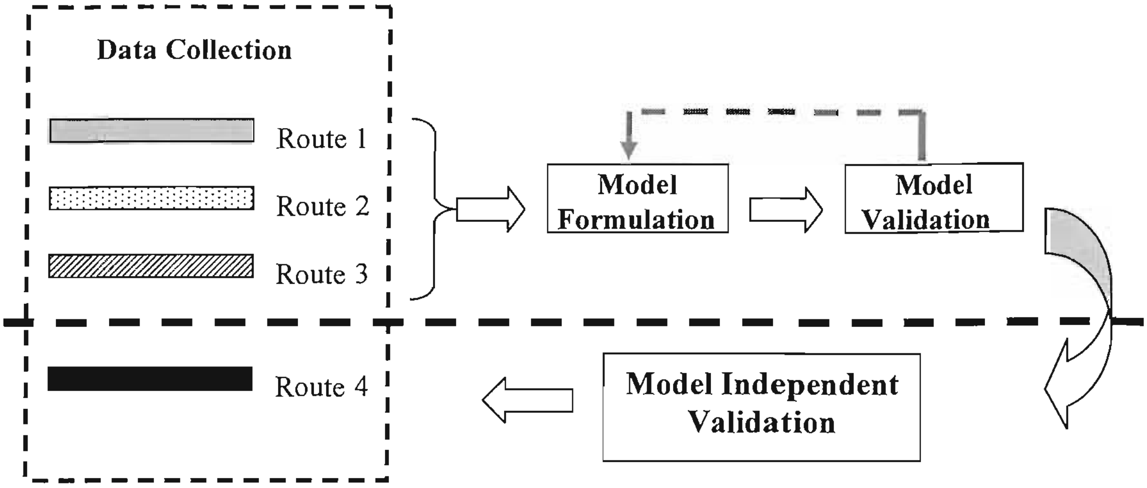


Figure 4.1: Data manipulation of this study

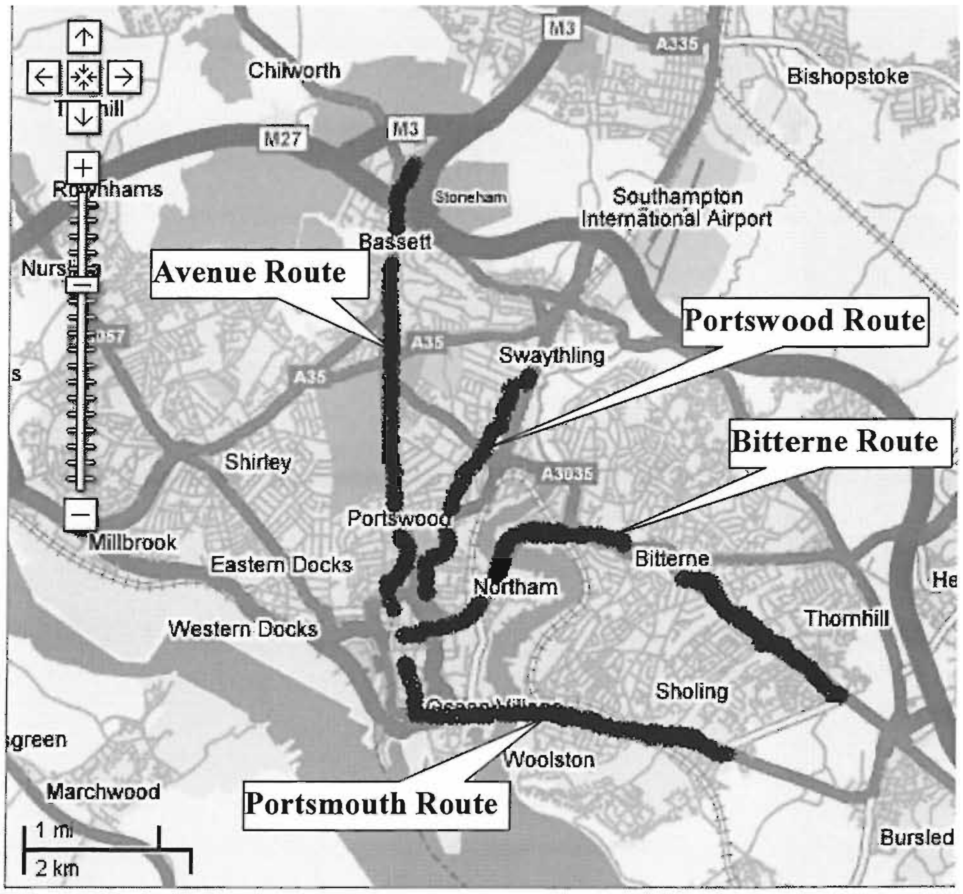


Figure 4.2: Four bus routes of survey sites in Southampton

(Accessed from the Google maps on http://maps.google.com/)

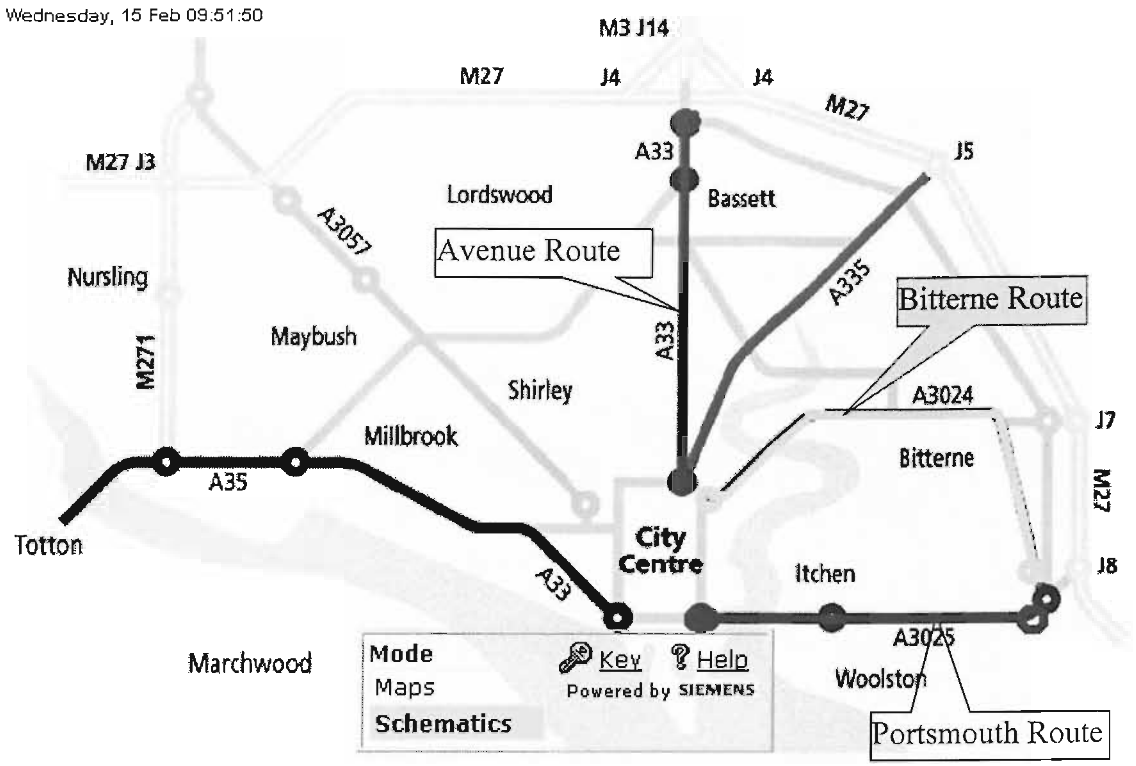


Figure 4.3: Available routes of ANPR journey time for this study
(Accessed from the ROMANSE website on http://southampton.romanse.org.uk/)

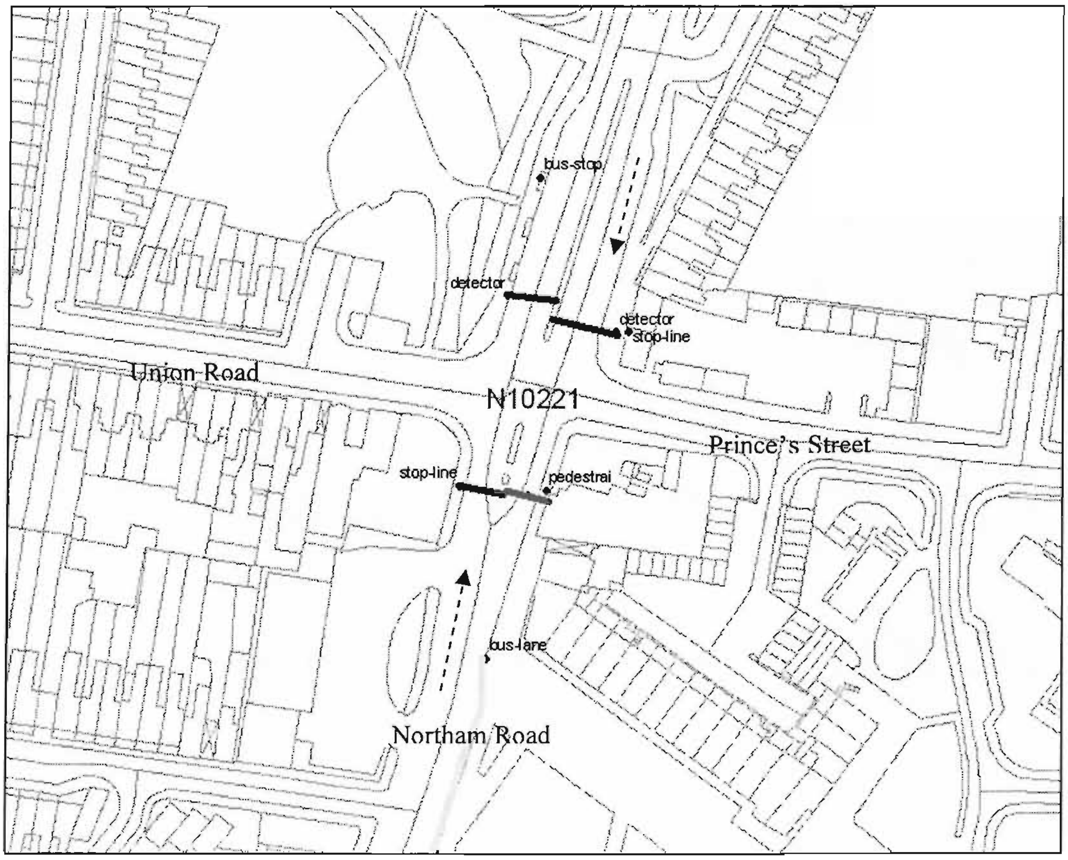


Figure 4.4: Example of road facilities locations at the junction of Northam Road and Union Road on the Bitterne route

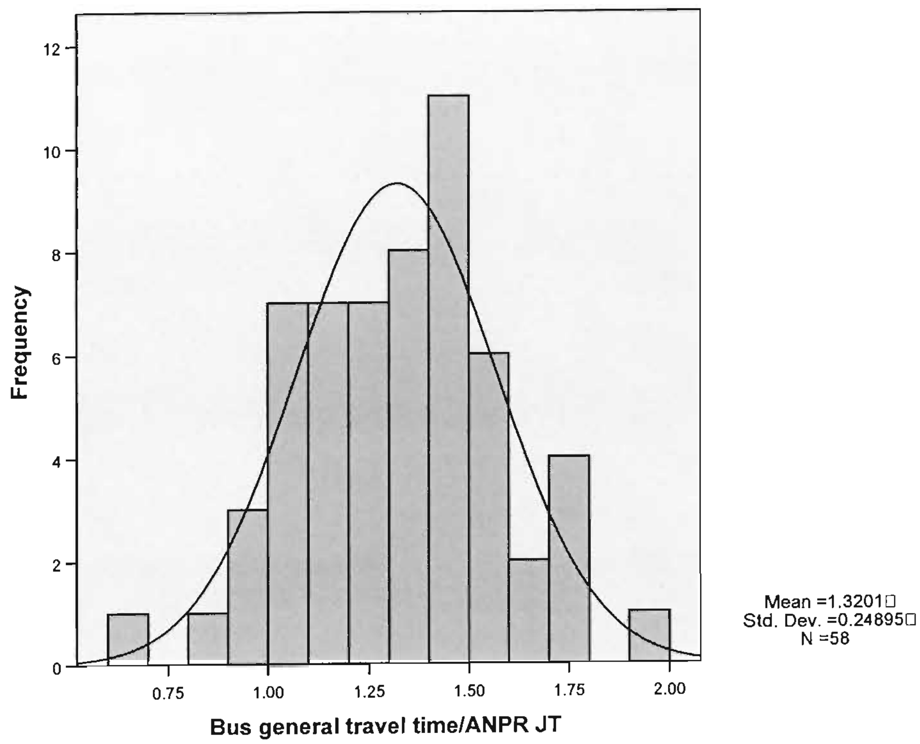


Figure 5.1: The frequency histogram of percentage of bus general travel time against ANPR journey time

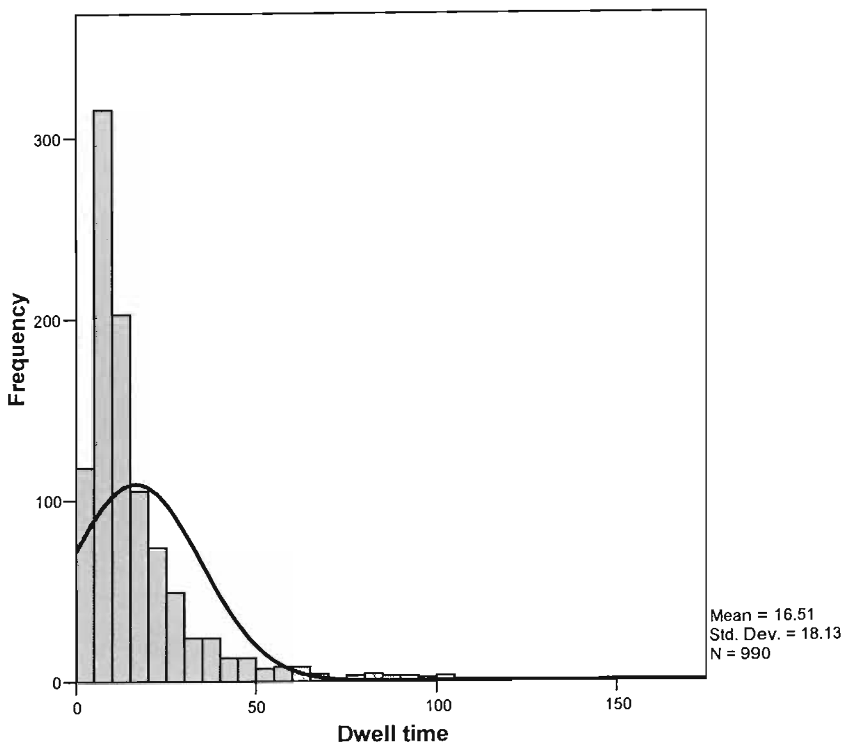


Figure 5.2: The frequency histogram of dwell time for each stop

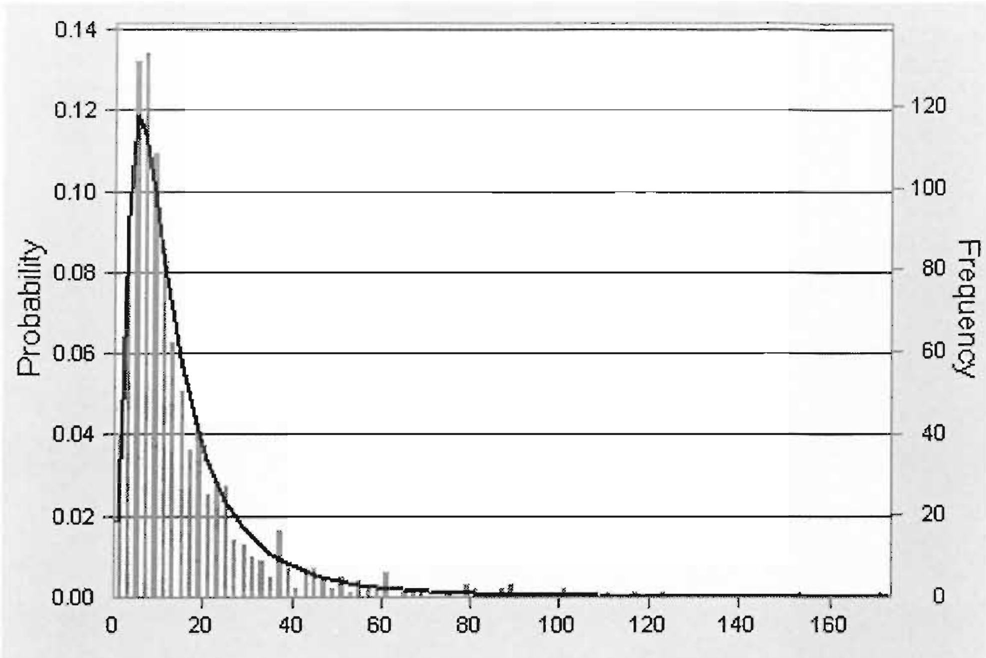


Figure 5.3: Lognormal distribution fit of dwell time at each stop (mean=16.22, SD=16.93)

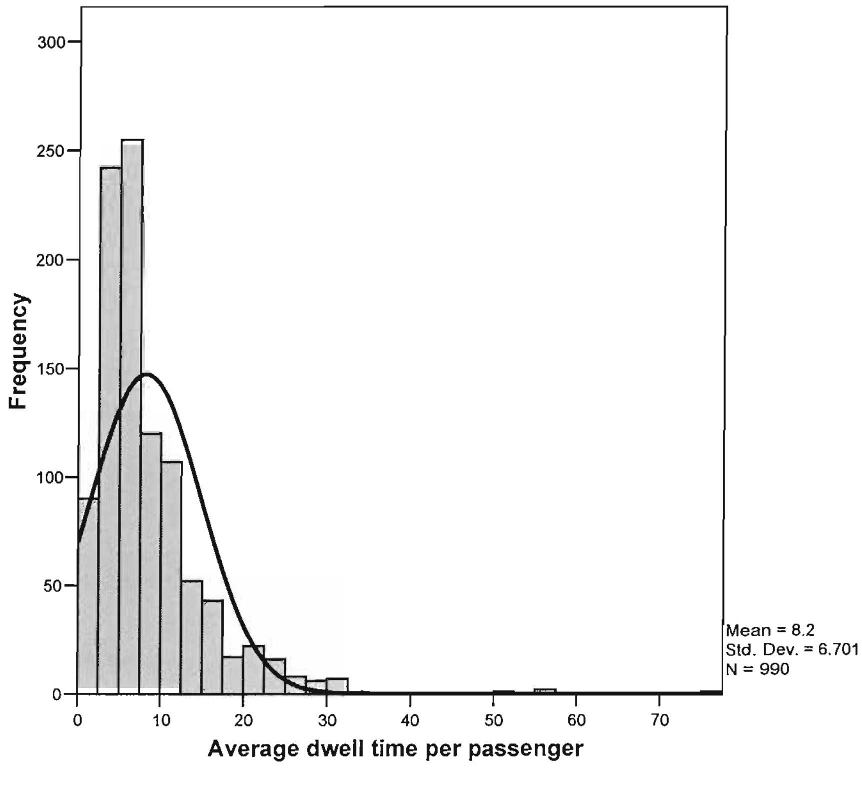


Figure 5.4: The frequency histogram of dwell time for each passenger

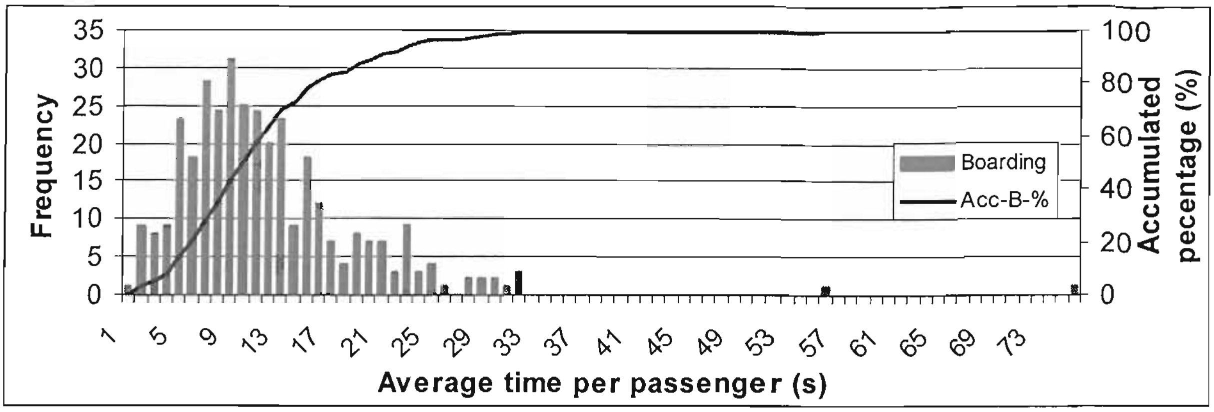


Figure 5.5: Average boarding time per passenger and accumulated percentage

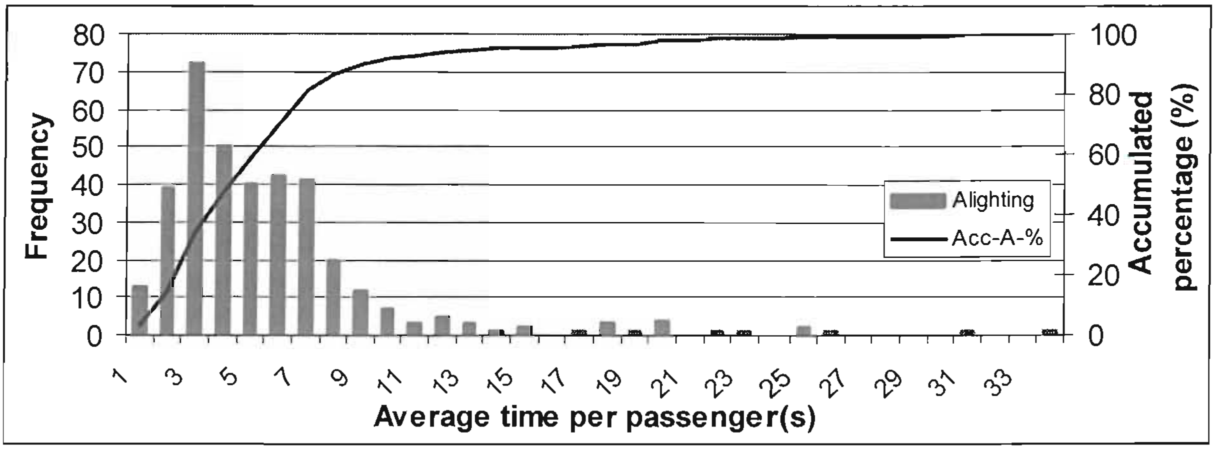


Figure 5.6: Average alighting time per passenger and accumulated percentage

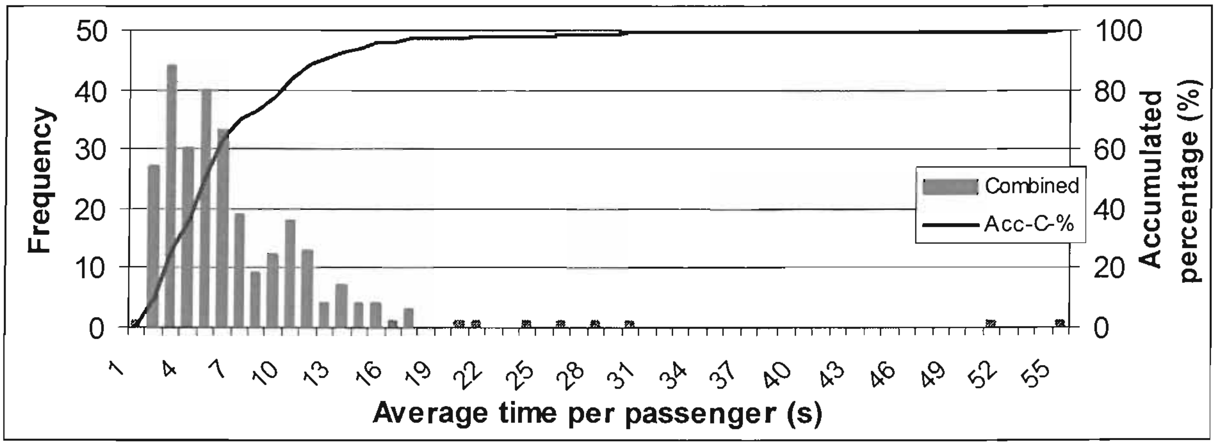


Figure 5.7: Average time per passenger of combined boarding and alighting and accumulated percentage

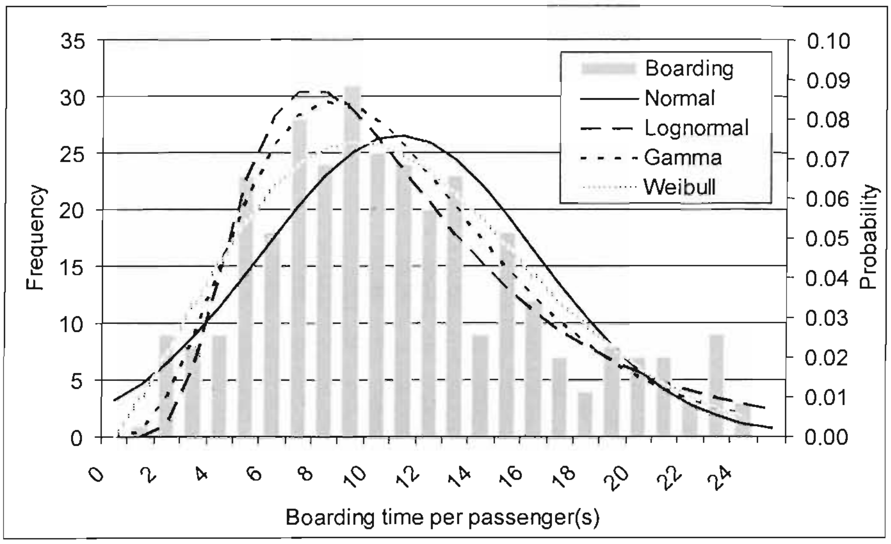


Figure 5.8: Average boarding time per passenger

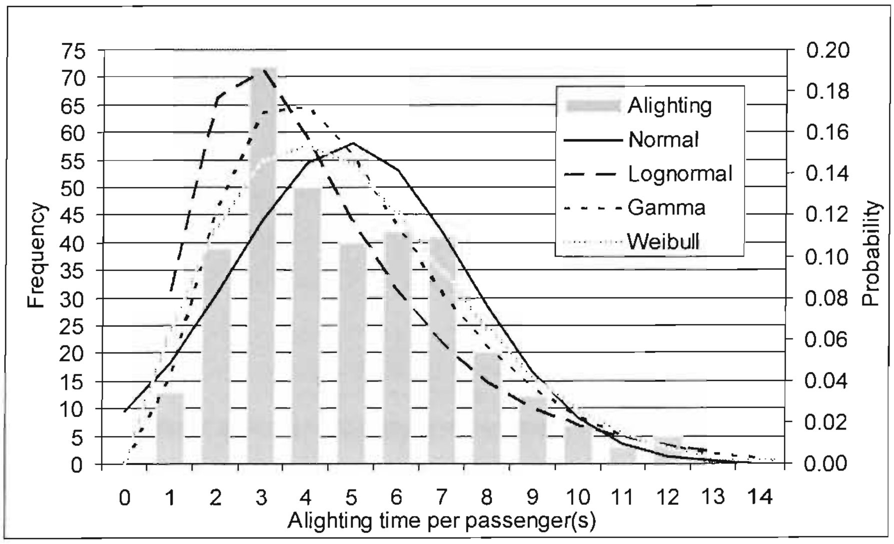


Figure 5.9: Average alighting time per passenger

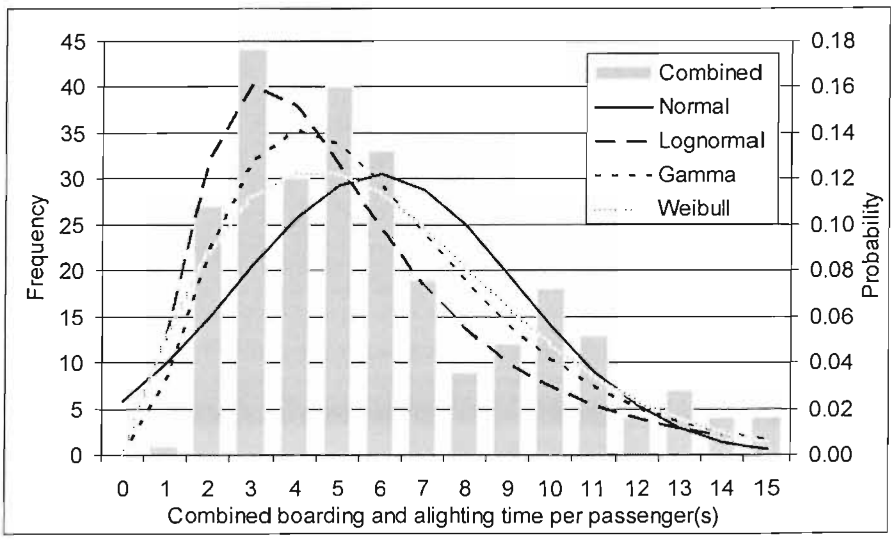


Figure 5.10: Average time per passenger of combined boarding and alighting

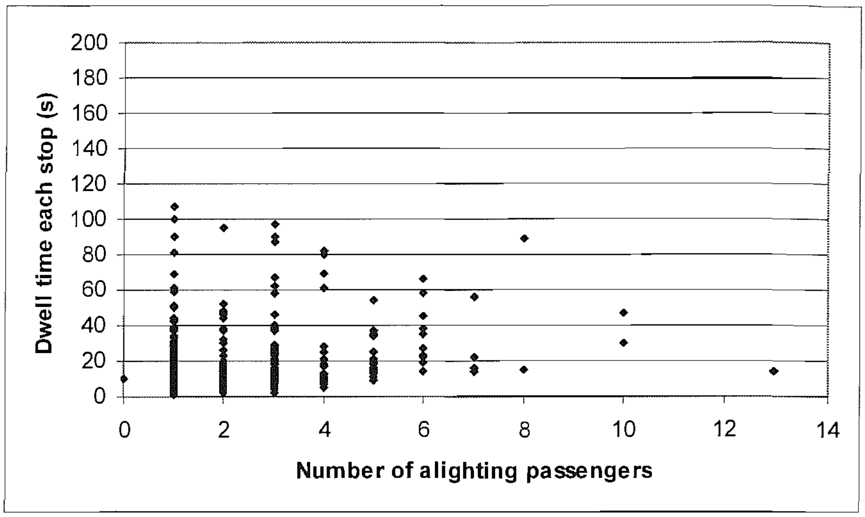


Figure 5.11 (A): Dwell time at each stop against the number of alighting passengers

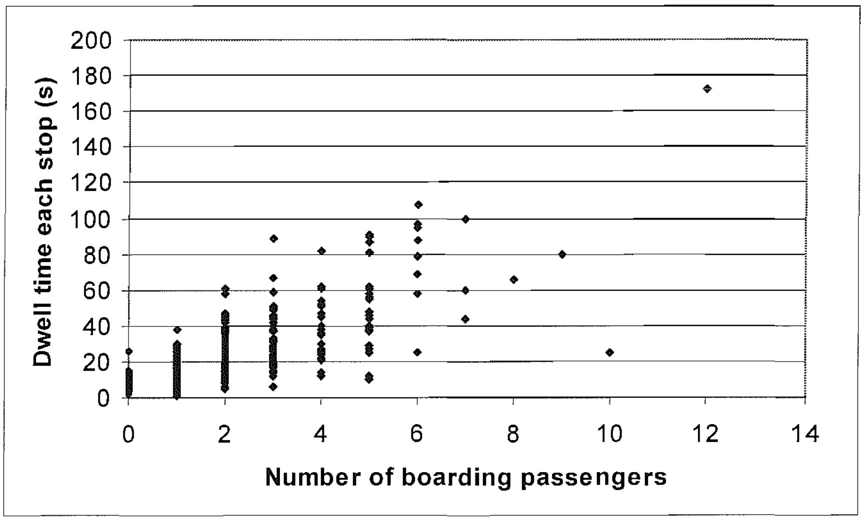


Figure 5.11 (B): Dwell time at each stop against the number of boarding passengers

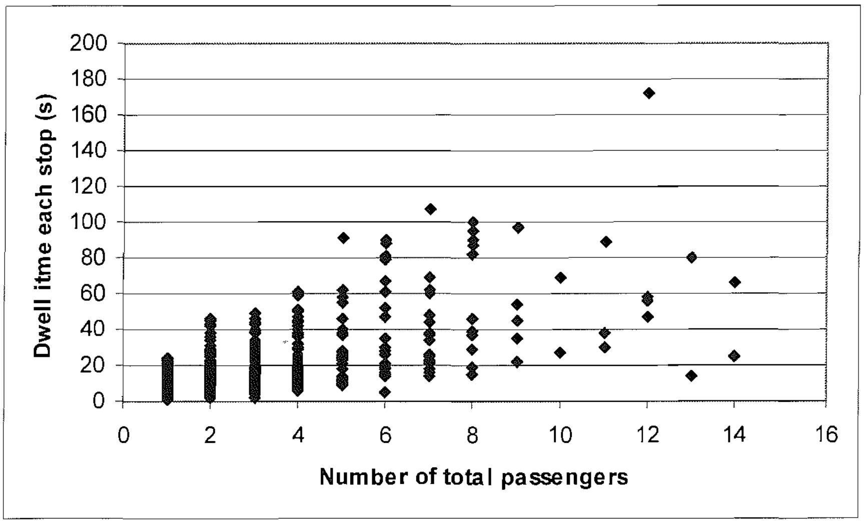


Figure 5.11 (C): Dwell time at each stop against the number of total passengers

Time (hhmmss)	Speed (m/s)	Speed (kph)
142939	10.94	39.39
142940	10.88	39.17
142941	10.09	36.31
142942	10.94	39.38
142943	11.17	40.20
142944	10.05	36.17
142945	10.71	38.57
142946	10.72	38.57
142947	9.35	33.66
142948	8.23	29.61
142949	7.28	26.21
142950	4.71	16.96
142951	3.52	12.65
142952	2.32	8.35
142953	1.20	4.30
142954	0.00	0.00
142955	0.00	0.00
142956	0.00	0.00
142957	0.00	0.00
142958	0.00	0.00
142959	2.32	8.36
143000	3.52	12.65
143001	4.71	16.96
143002	5.84	21.01
143003	7.03	25.31
143004	8.23	29.61
143005	10.55	37.97
143006	10.55	37.97
143007	10.55	37.97
143008	12.87	46.32
143009	11.89	42.81
143010	12.99	46.76

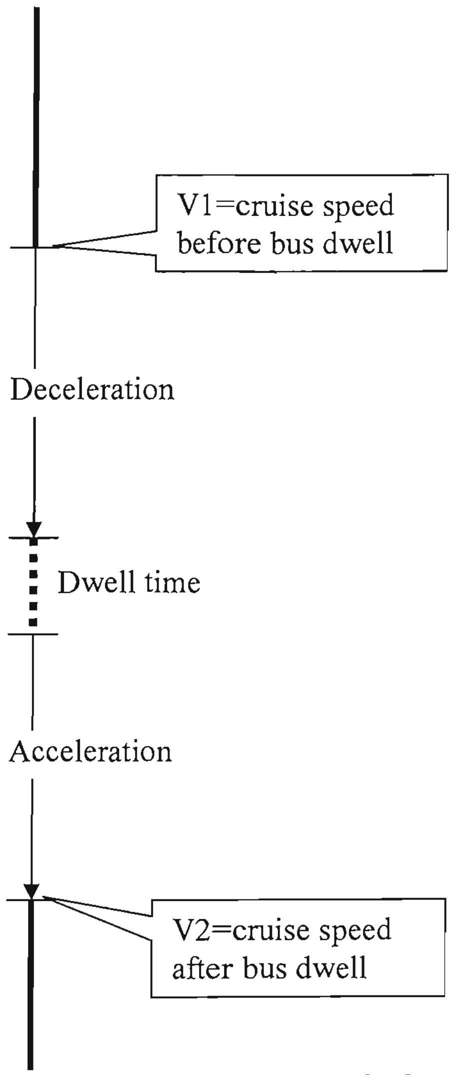


Figure 5.12: An example of identifying the cruise speed before and after bus dwell (data from inbound of Bitterne route on 13/09/2006)

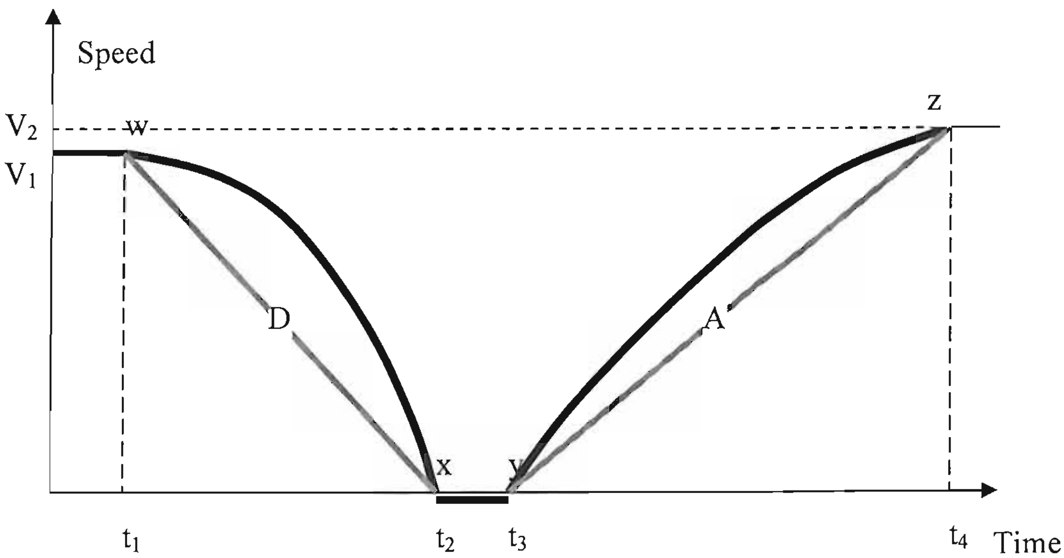


Figure 5.13: Average acceleration/deceleration rate calculation

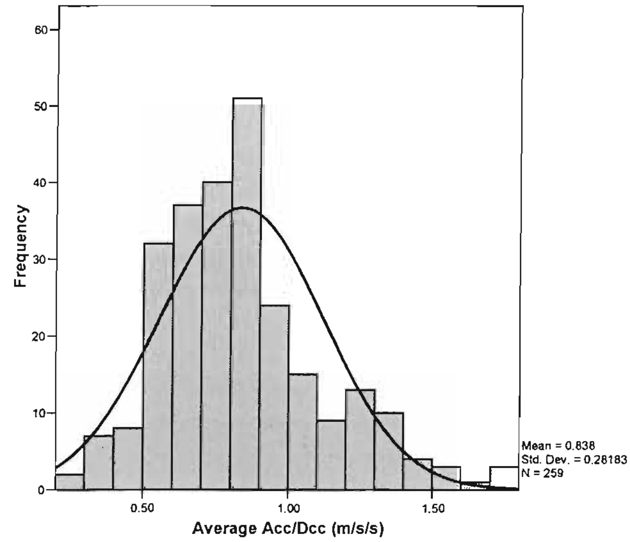


Figure 5.14: Acceleration rate frequency histogram

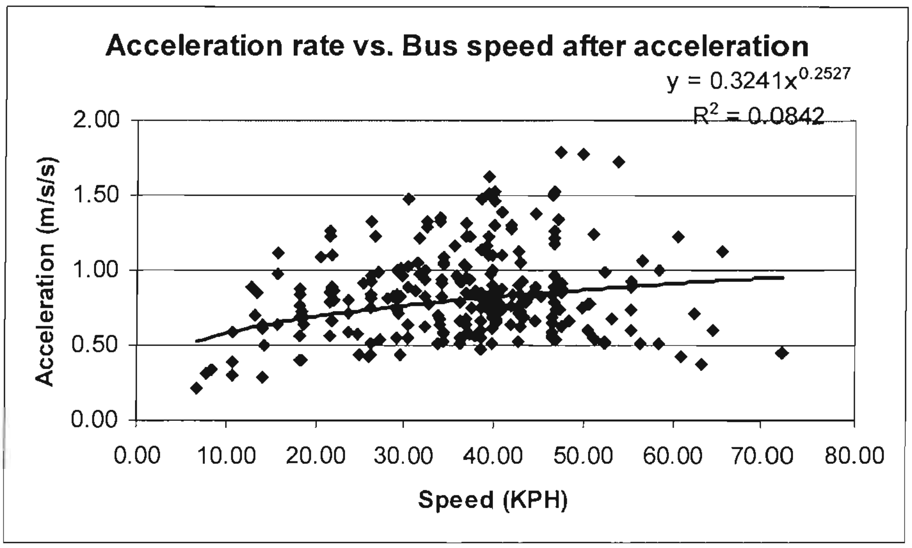


Figure 5.15: Acceleration rates against bus cruise speed after acceleration

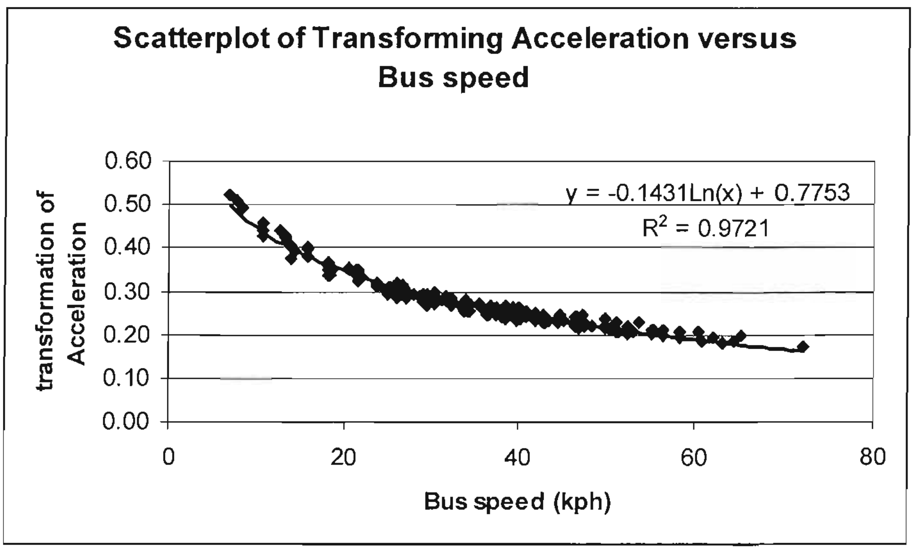


Figure 5.16: Scatter plot of transforming acceleration rates against bus speeds

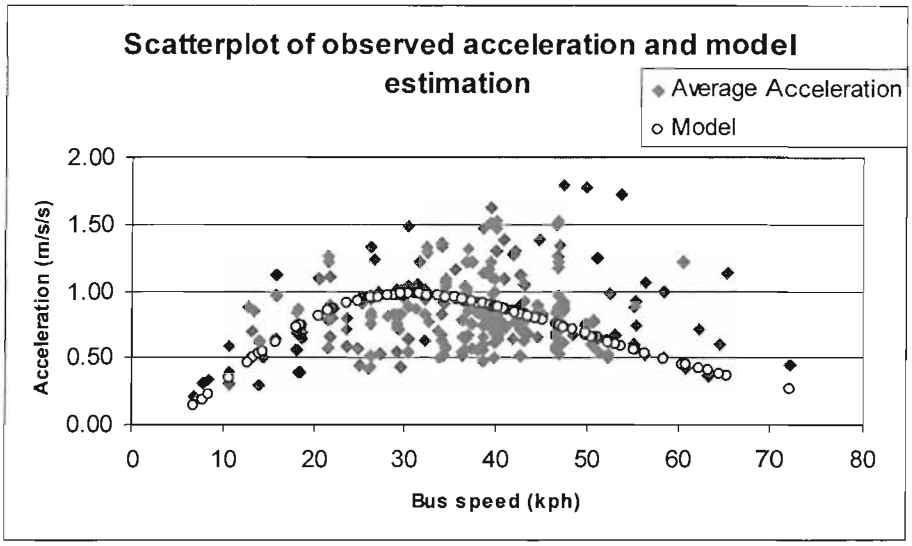


Figure 5.17: Scatter plot of observed acceleration rates against model estimates

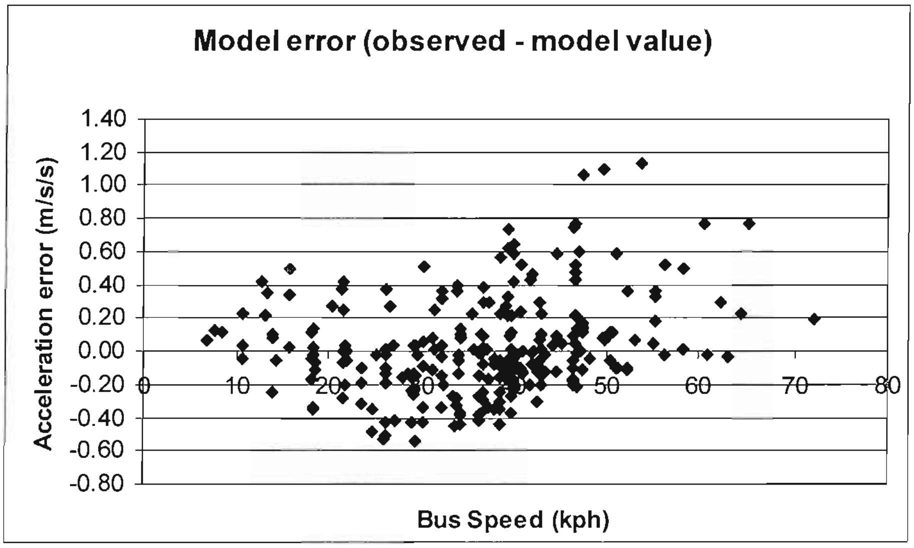


Figure 5.18: Scatter plot of model residuals against bus speed

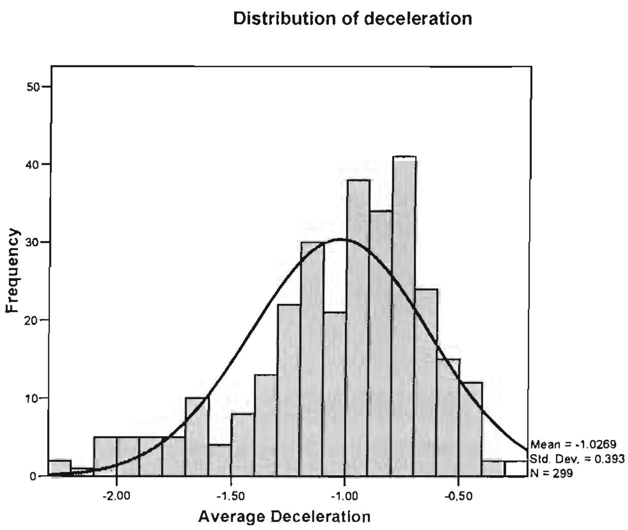


Figure 5.19: Deceleration rate frequency histogram

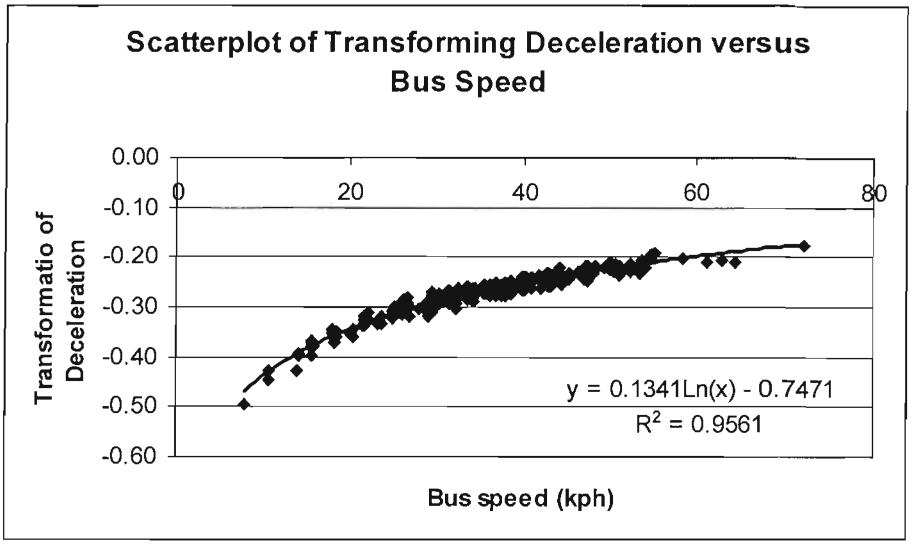


Figure 5.20: Scatter plot of transforming deceleration rates against bus speeds

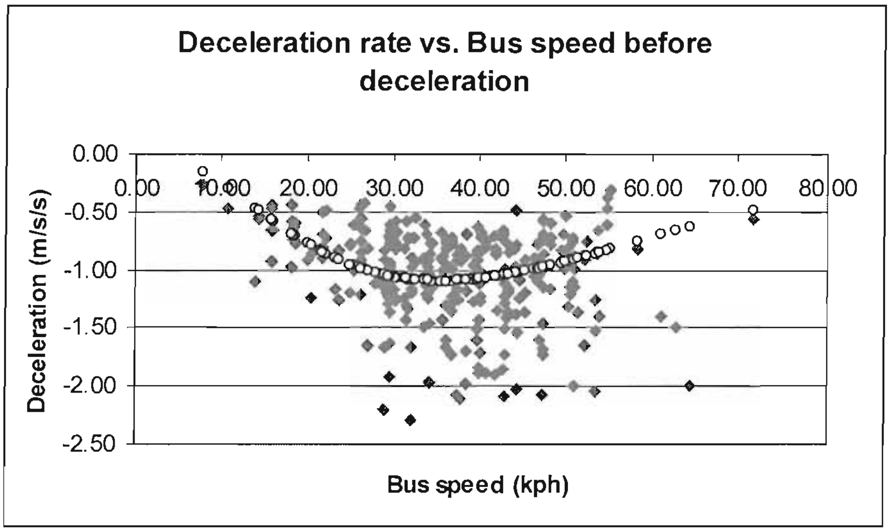


Figure 5.21: Scatter plot of observed deceleration rates against model estimates

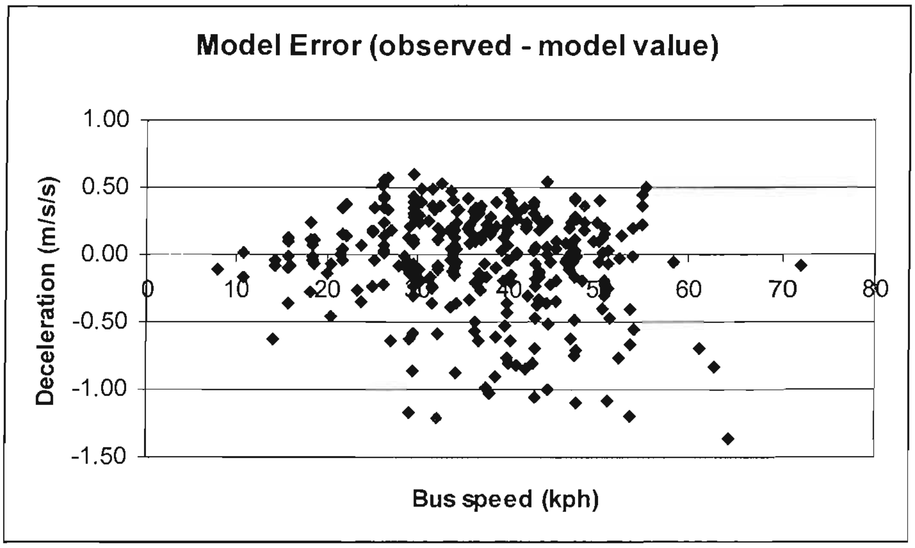


Figure 5.22: Scatter plot of model residuals against bus speed

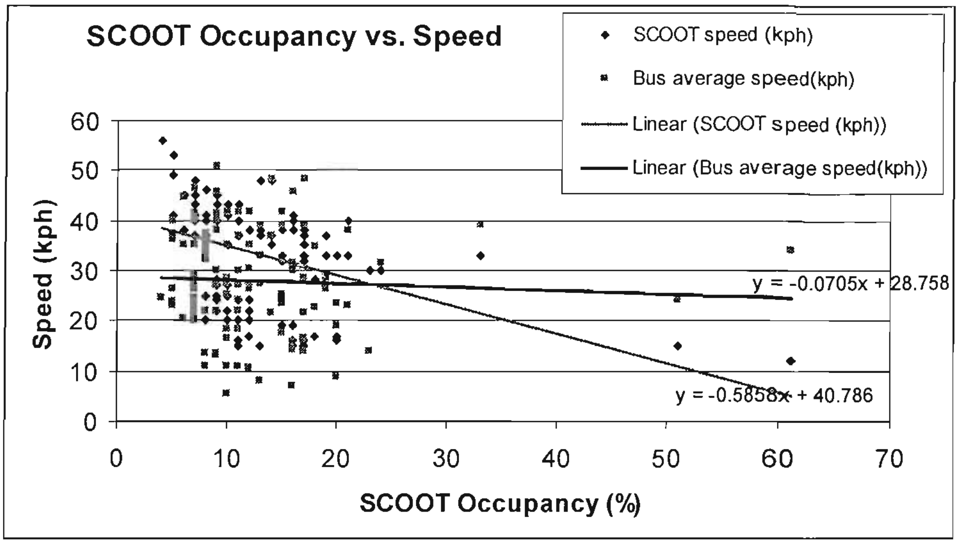


Figure 5.23: SCOOT occupancy against Bus and other vehicles speed

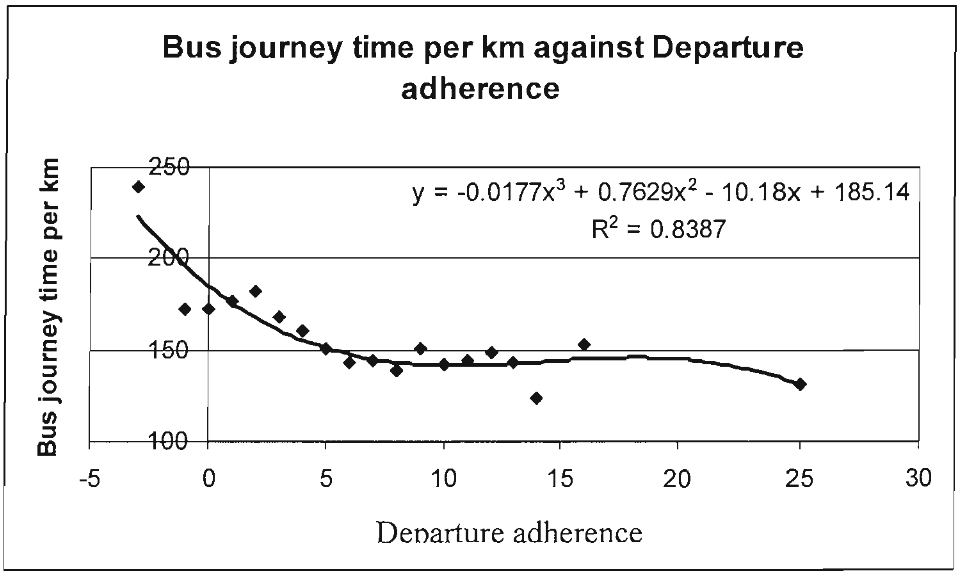


Figure 5.24: Schedule adherence effect on bus journey time

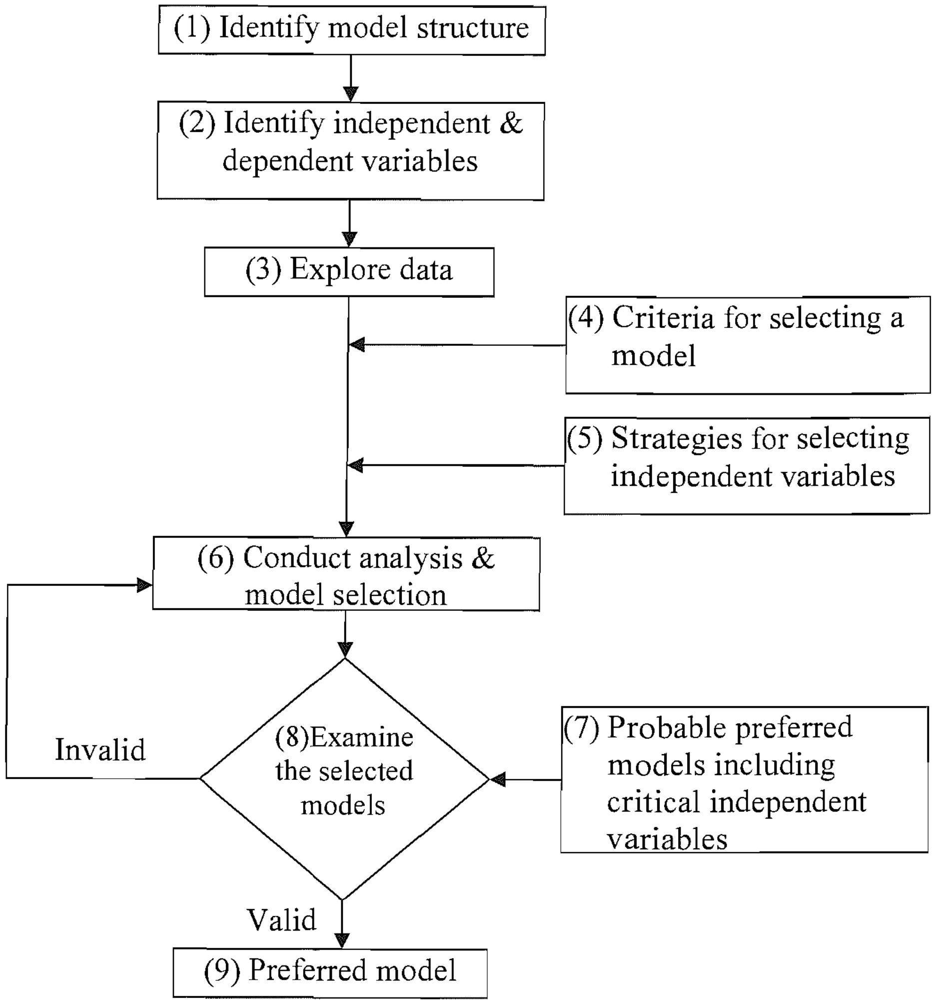


Figure 6.1: A procedure for conducting regression analysis

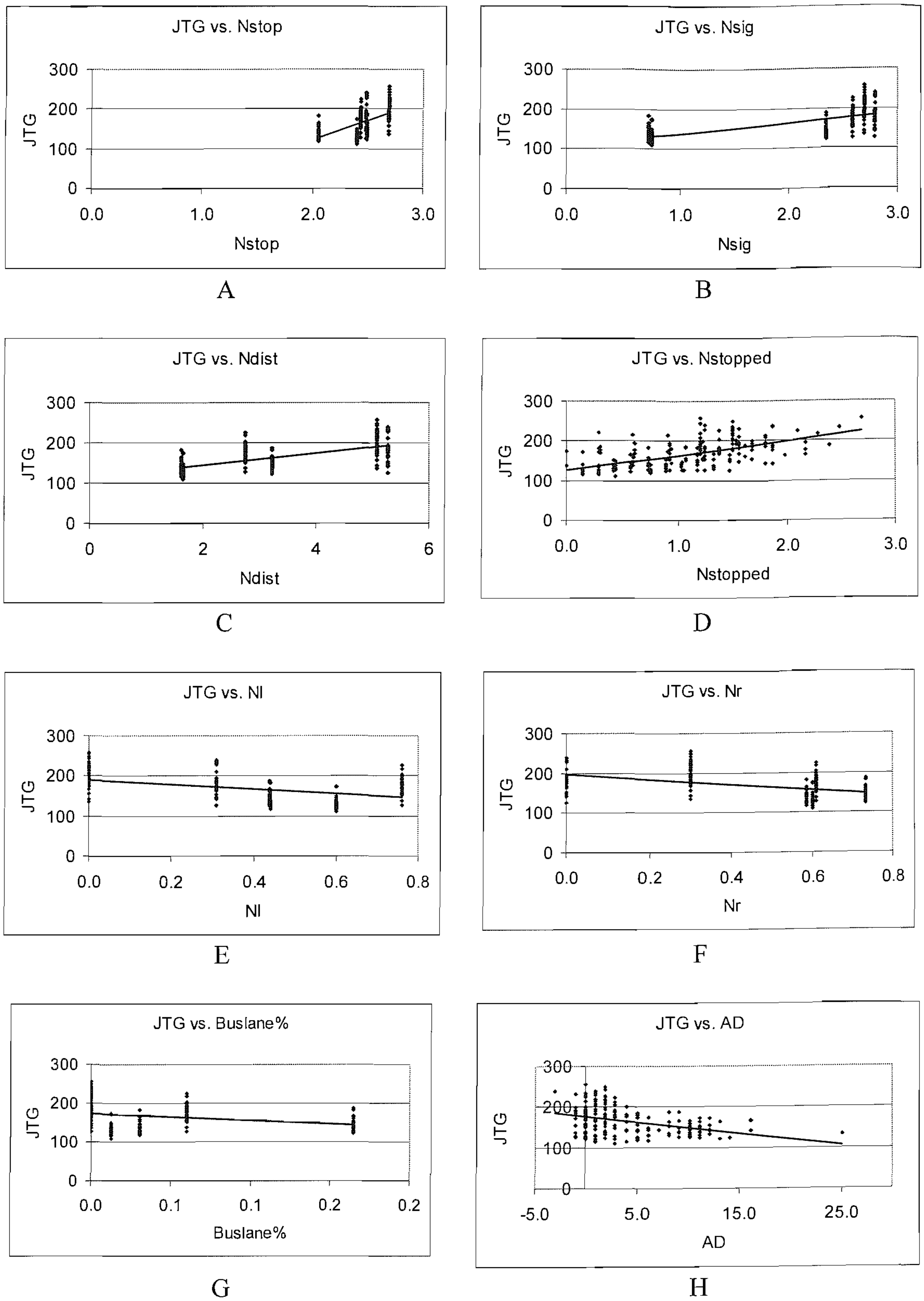


Figure 6.2: Scatter plot of bus route general travel time JT_G against predictors

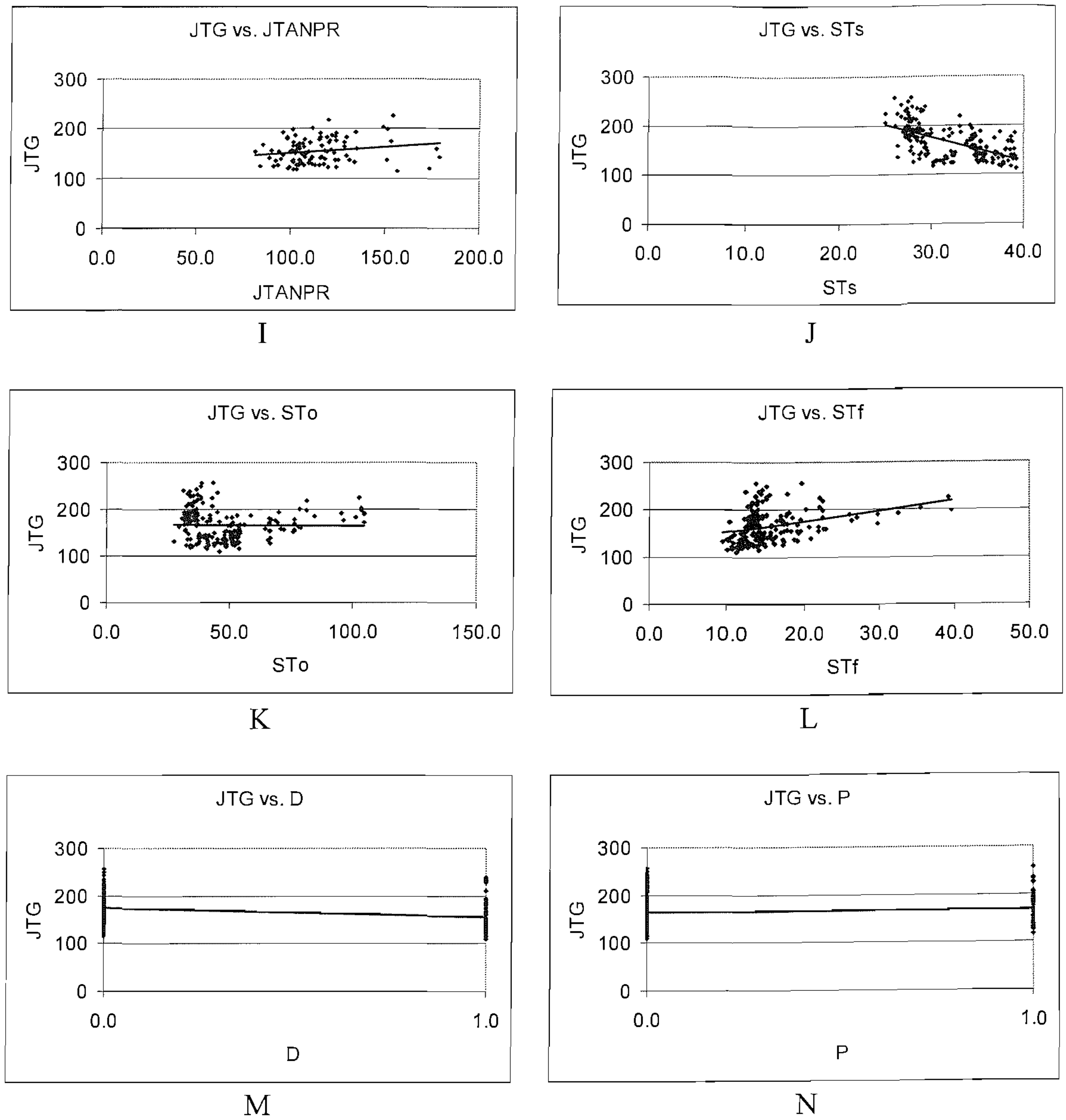


Figure 6.2: Scatter plot of bus route general travel time JT_G against predictors (continued)

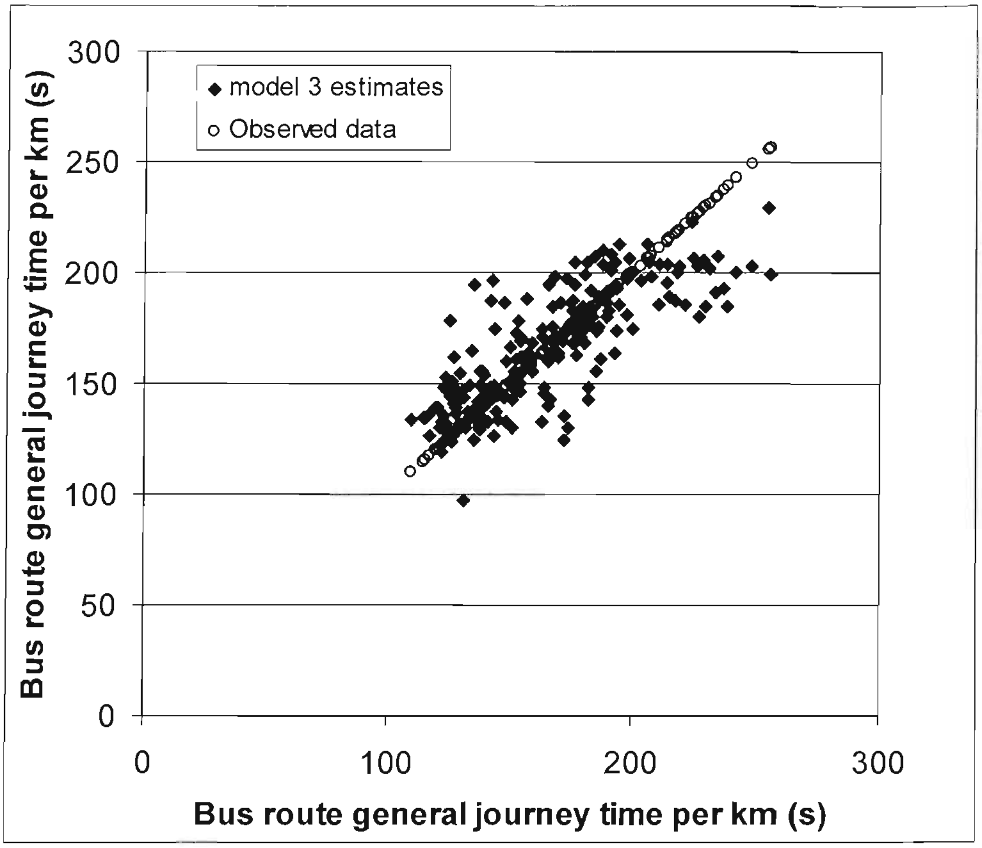


Figure 6.3: Scatter plot of route general journey time model predicted values against observed data (SCOOT model)

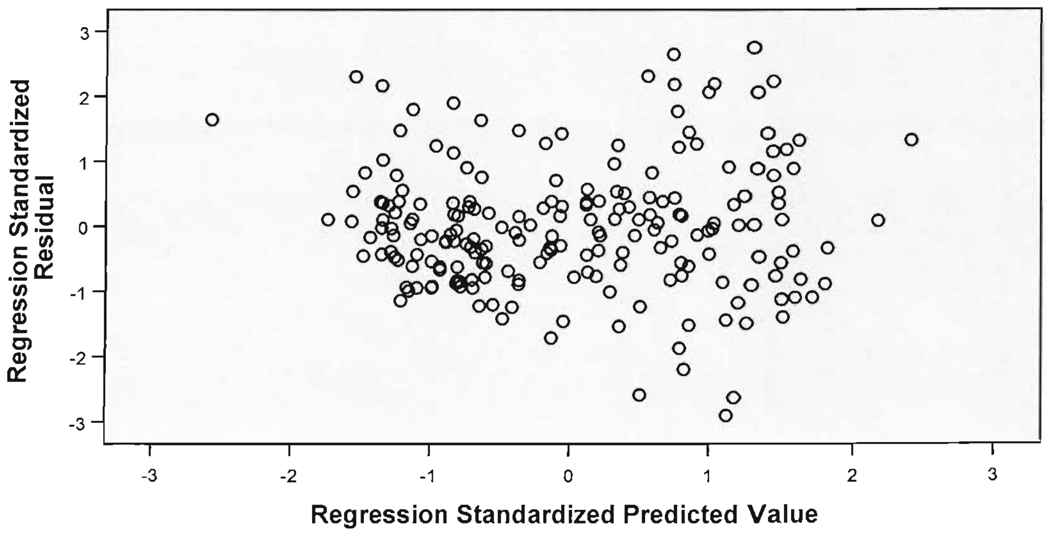


Figure 6.4: Scatter plot of route general journey time model residual against observed journey time (SCOOT model)

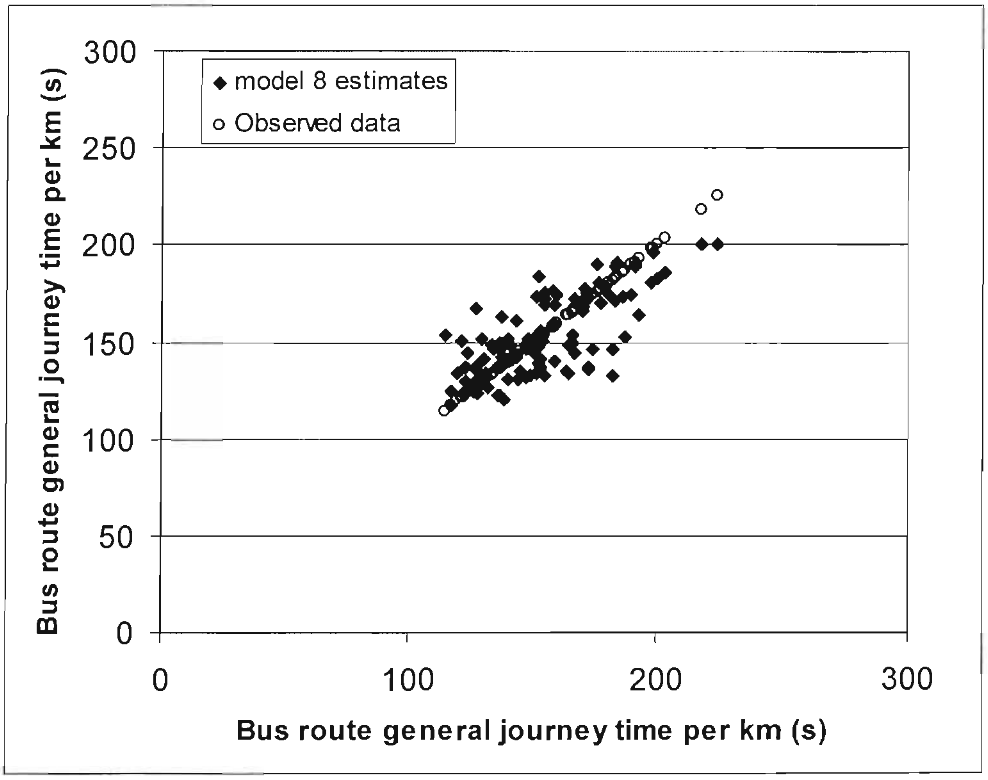


Figure 6.5: Scatter plot of route general journey time model predicted values against observed data (ANPR model)

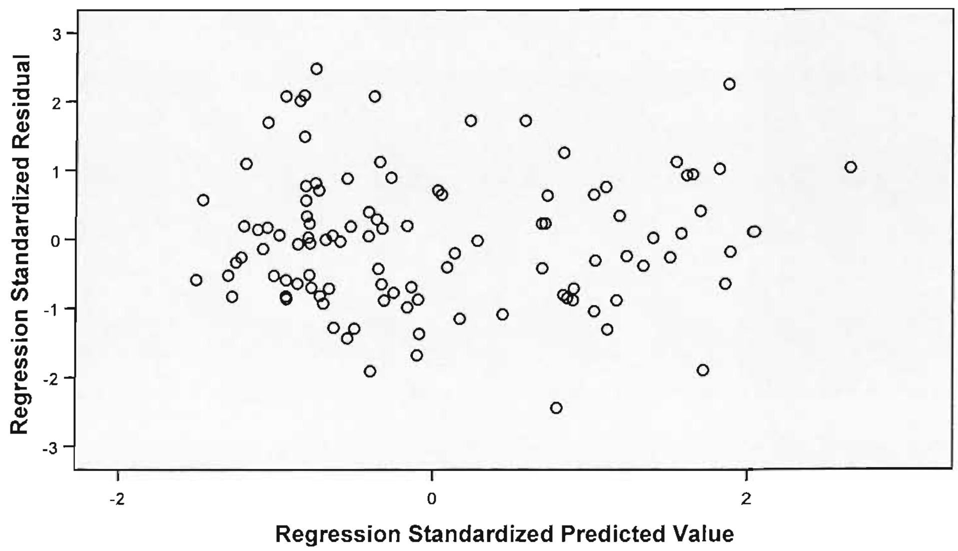


Figure 6.6: Scatter plot of route general journey time model residual against observed journey time (ANPR)

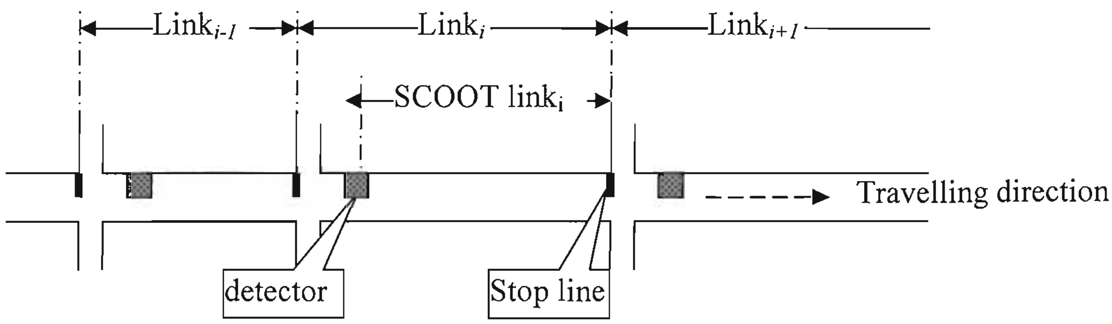


Figure 6.7: A sketch of link based journey time approach

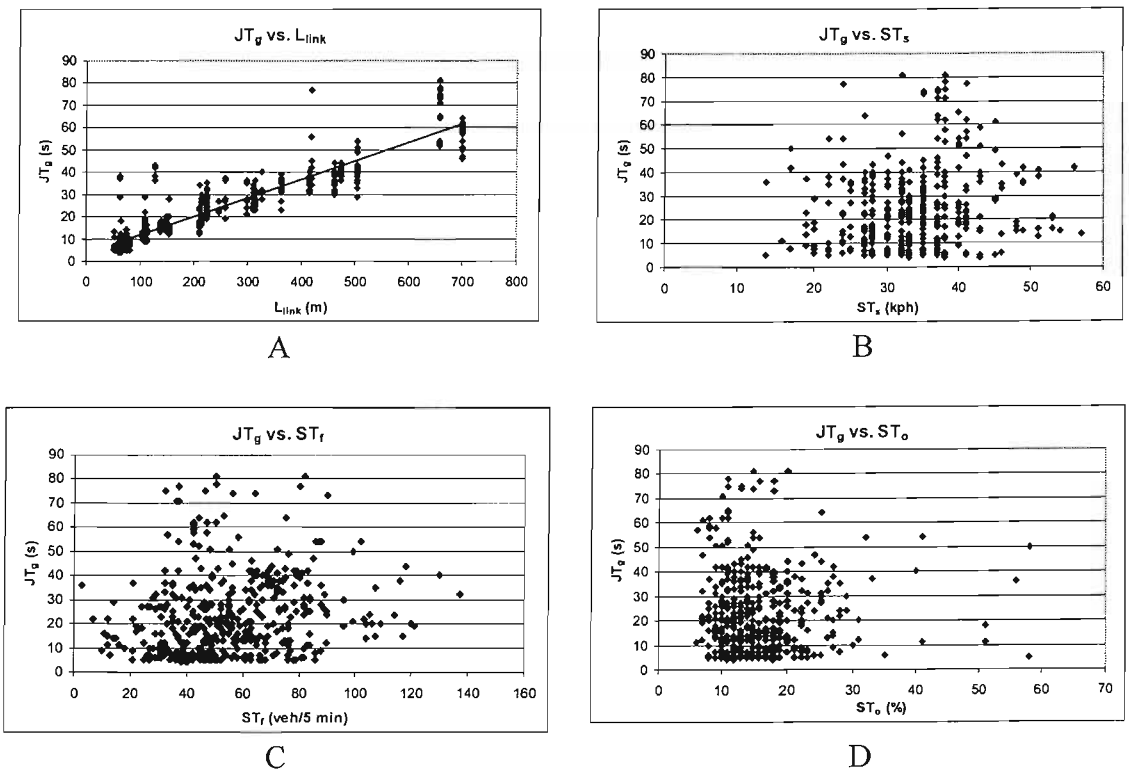
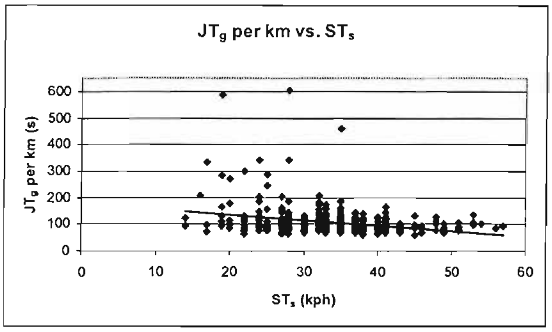
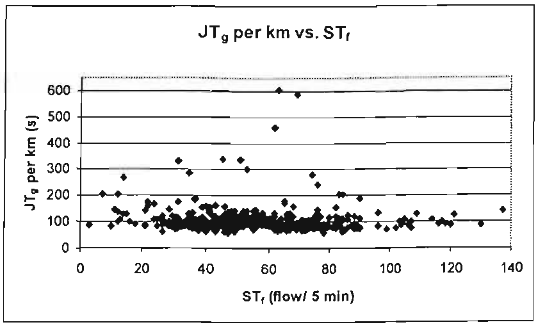


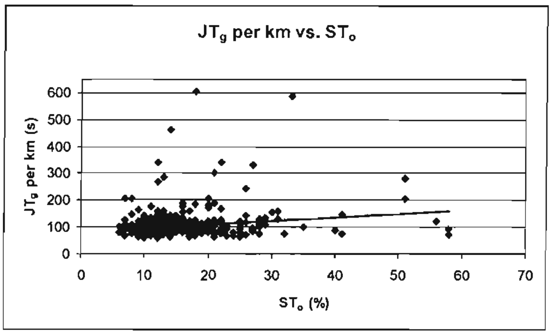
Figure 6.8: Scatter plot of bus link general travel time JT_g against predictors



A



B



C

Figure 6.9: Scatter plot of bus link general travel time JT_g per km against SCOOT parameters

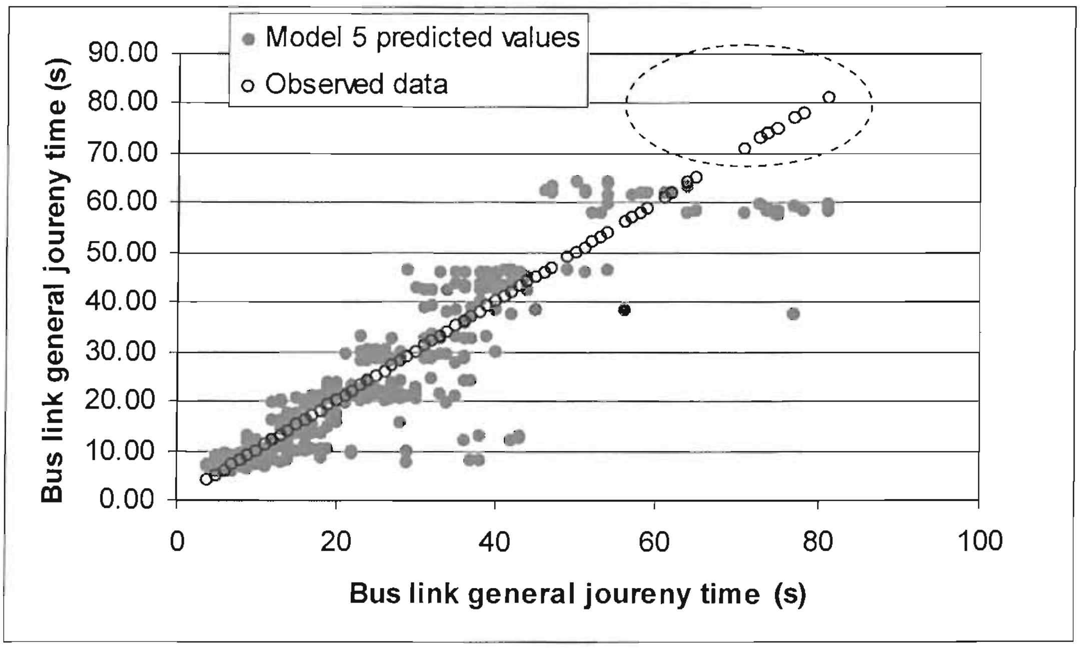


Figure 6.10: Scatter plot of link general travel time model predicted values against observed data

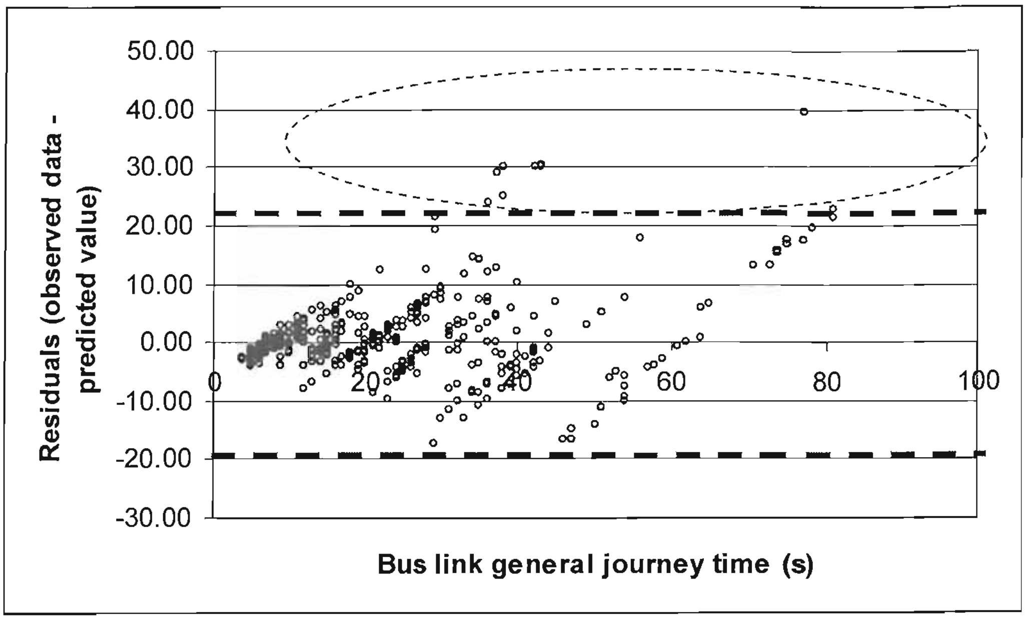


Figure 6.11: Scatter plot of link general travel time model residual against observed journey time

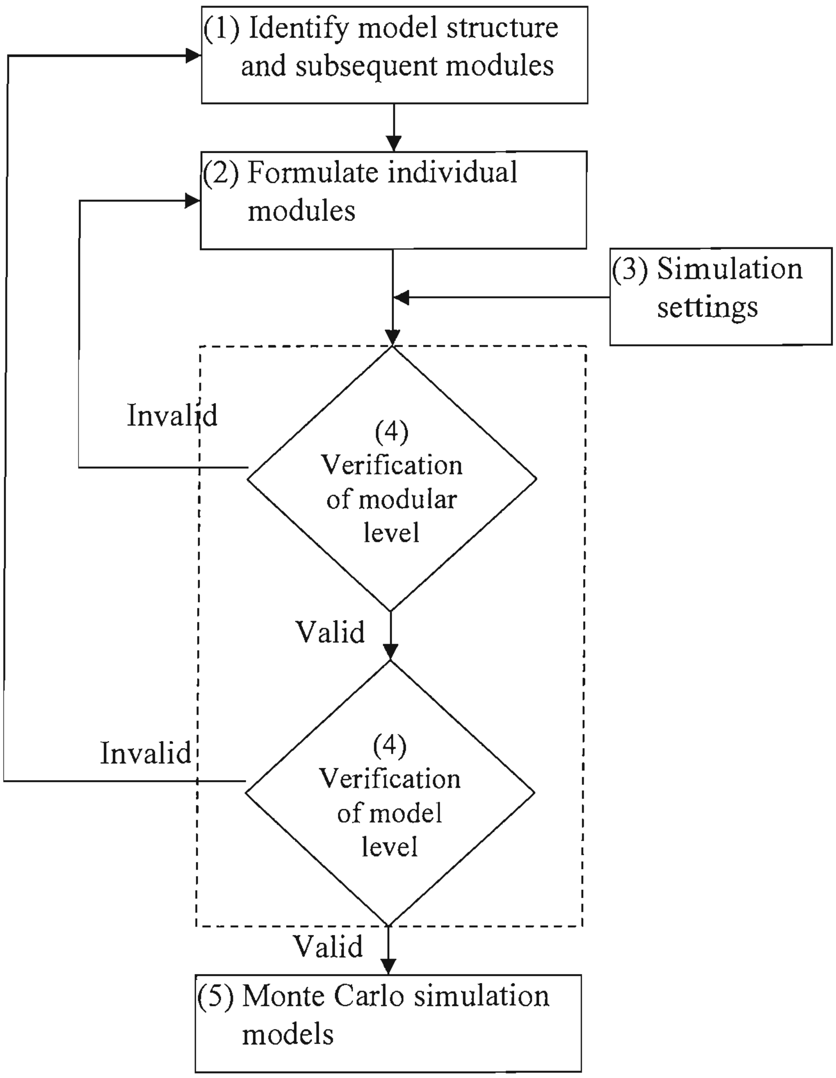


Figure 6.12: A procedure for conducting Monte Carlo simulation analysis

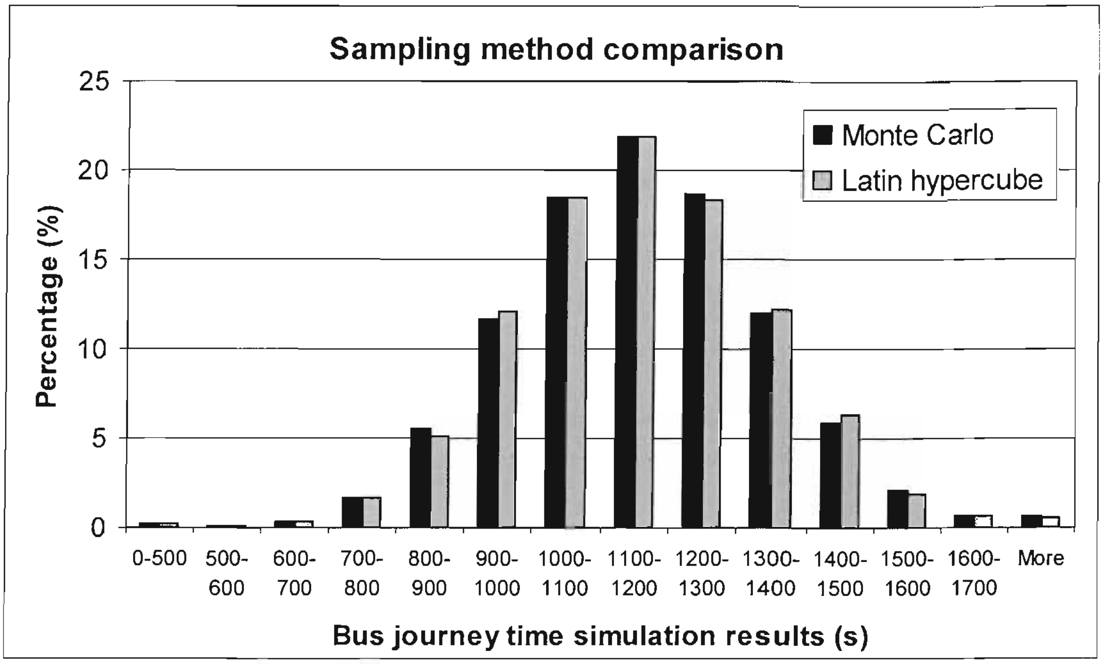


Figure 6.13: Sampling method comparison of Monte Carlo simulation results

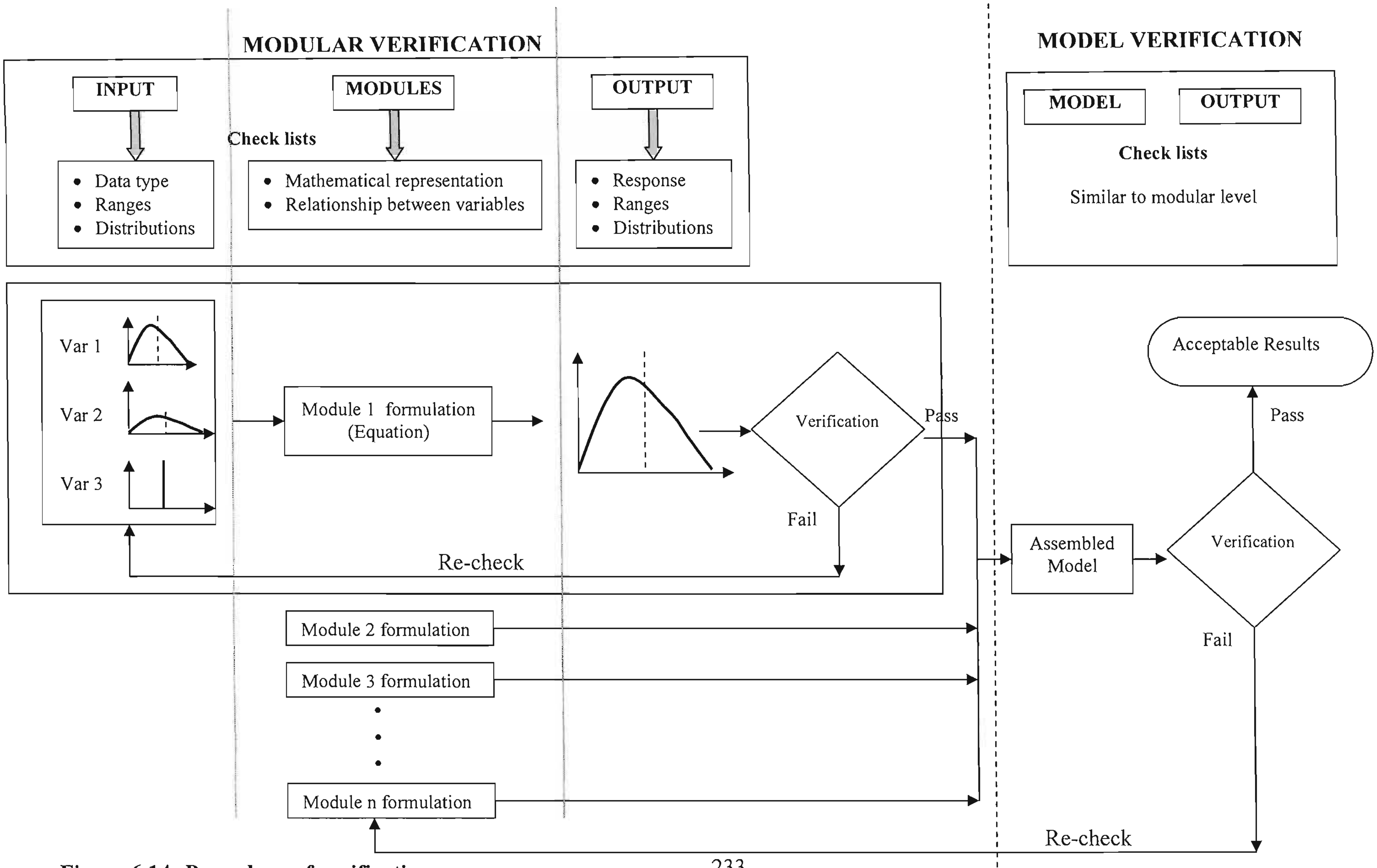


Figure 6.14: Procedure of verification

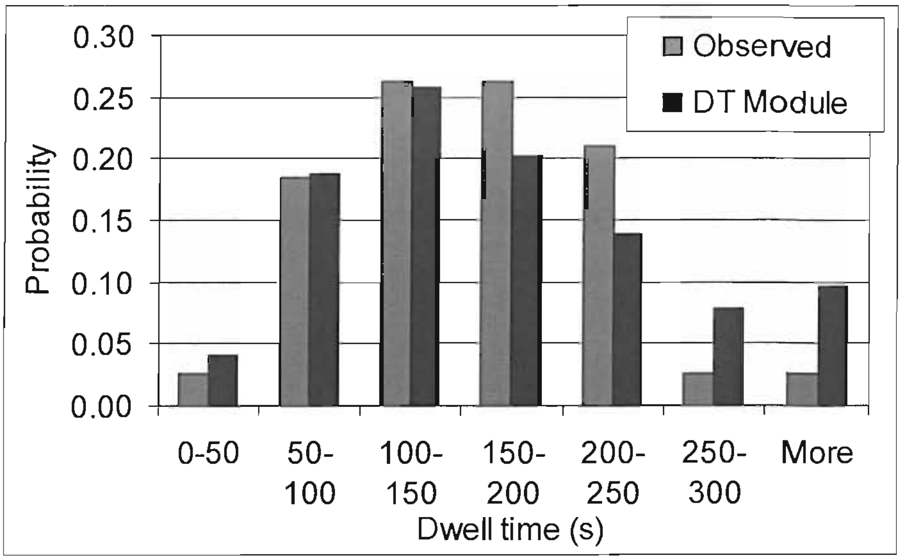


Figure 6.15: Comparison of link-based dwell time (DT) module result against observed data

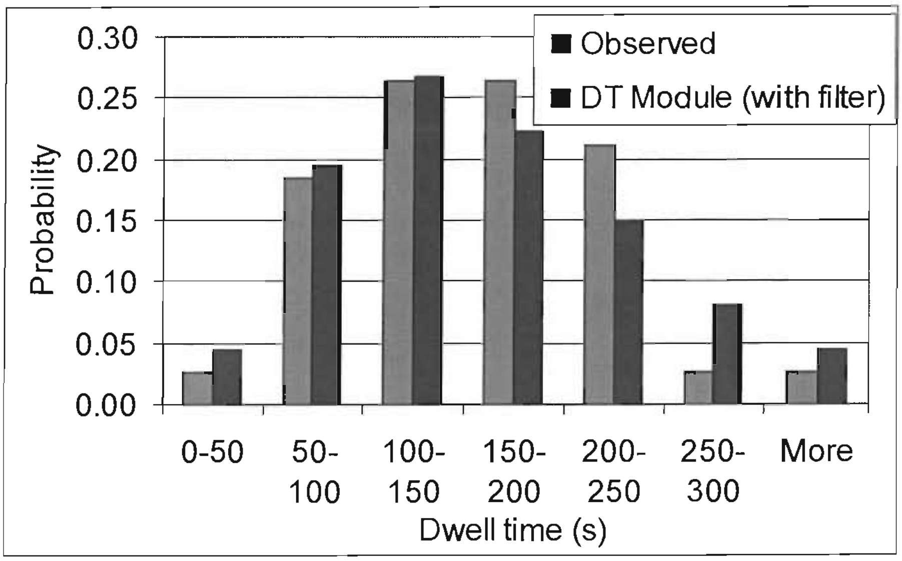


Figure 6.16: Comparison of link-based dwell time (DT) module result with output filter result against observed data

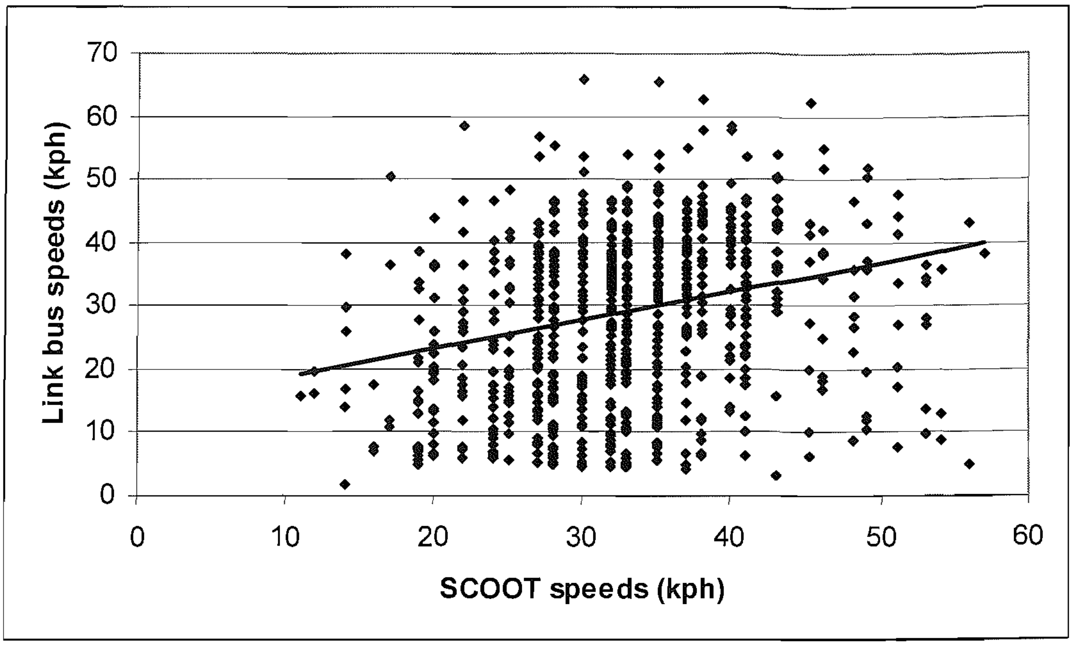


Figure 6.17: Scatter plot of link bus speed (BV) against SCOOT speed (ST_s)

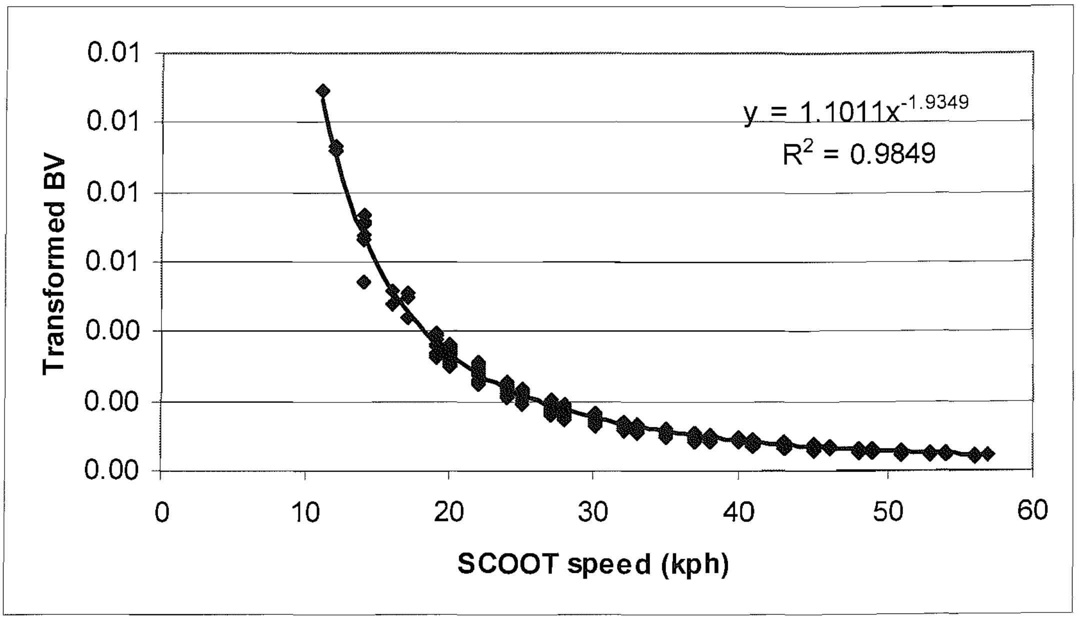


Figure 6.18: Scatter plot of transformed link bus speeds against SCOOT speeds

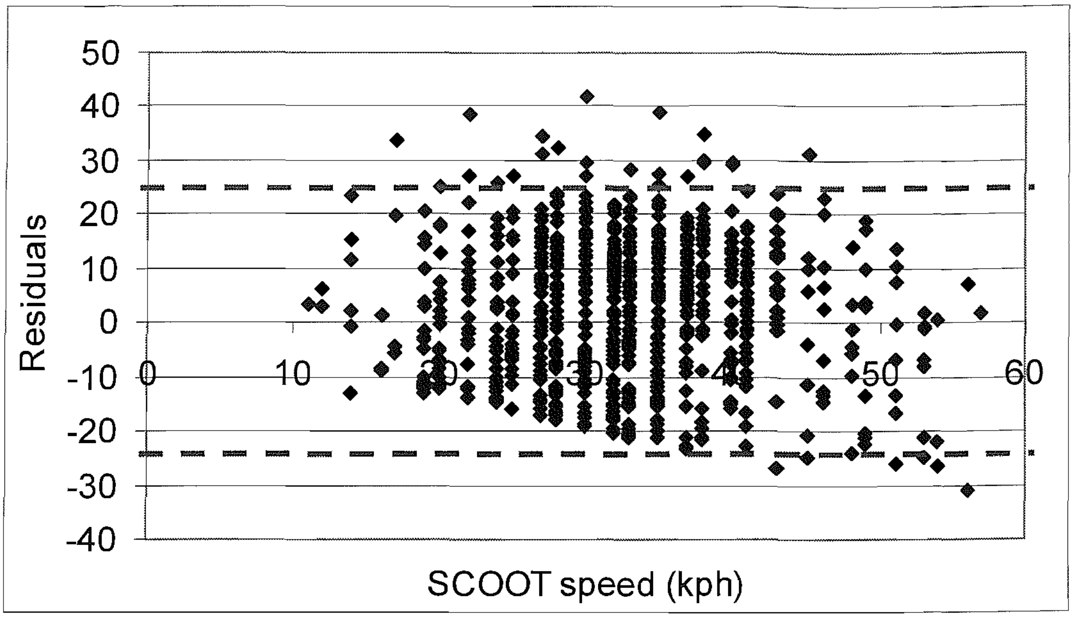


Figure 6.19: Scatter plot of residuals against SCOOT speeds

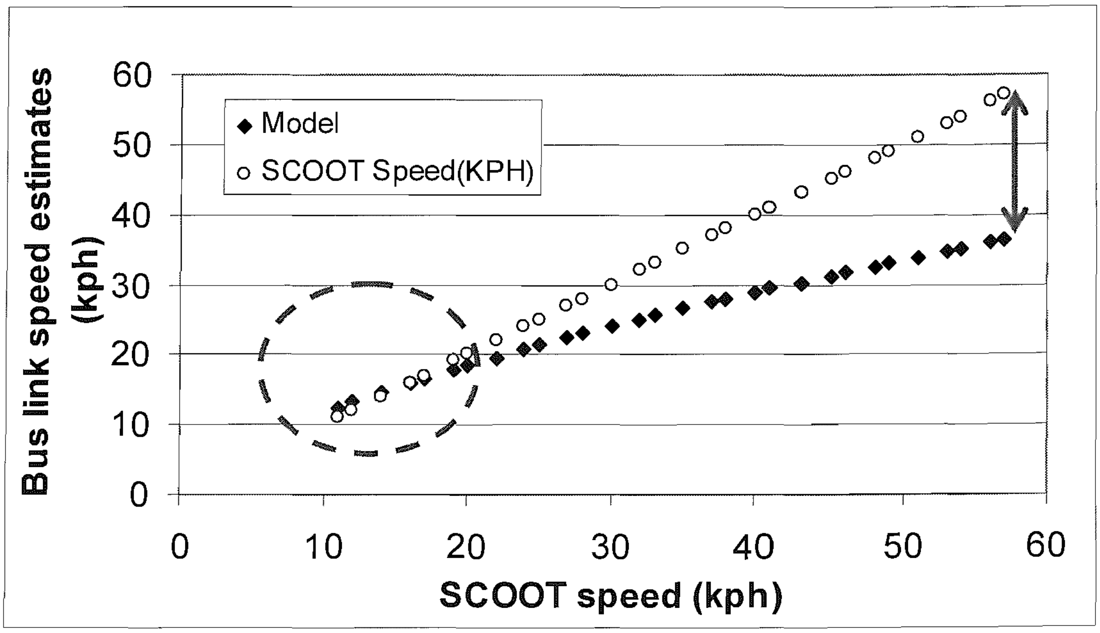


Figure 6.20: Comparison of bus link speed estimates against SCOOT speeds

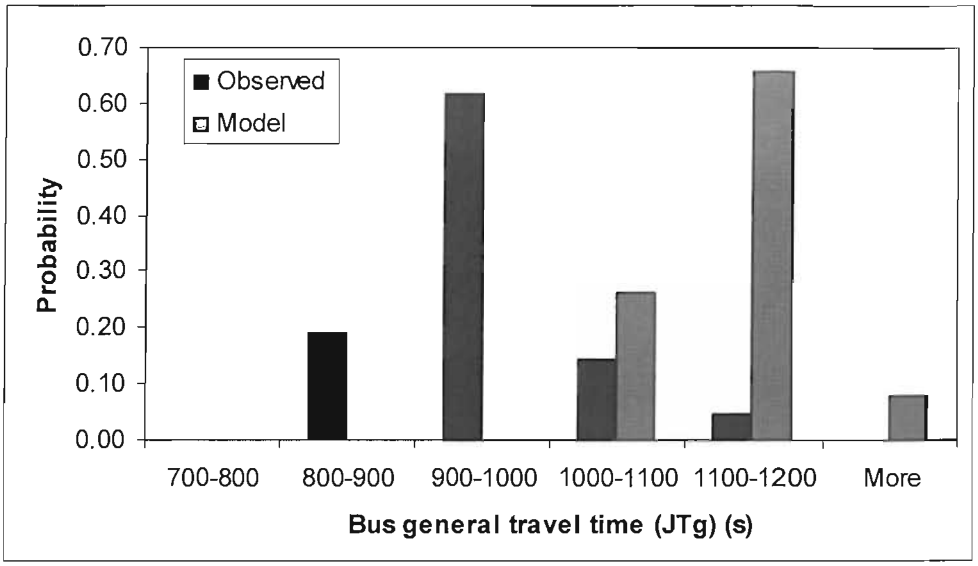


Figure 6.21: Comparison of link-based general travel time (JT_g) module result against observed data

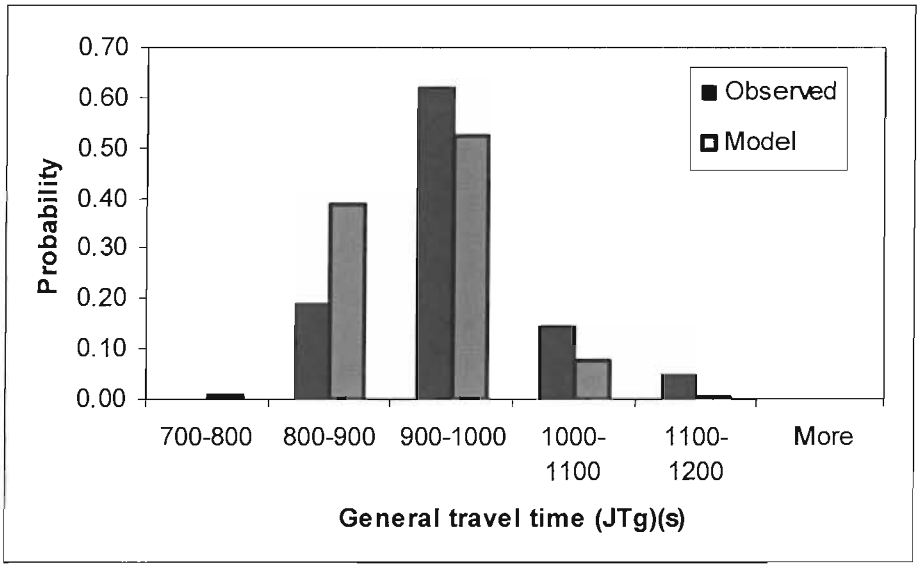


Figure 6.22: Revised link-based general travel time (JT_g) module result against observed data

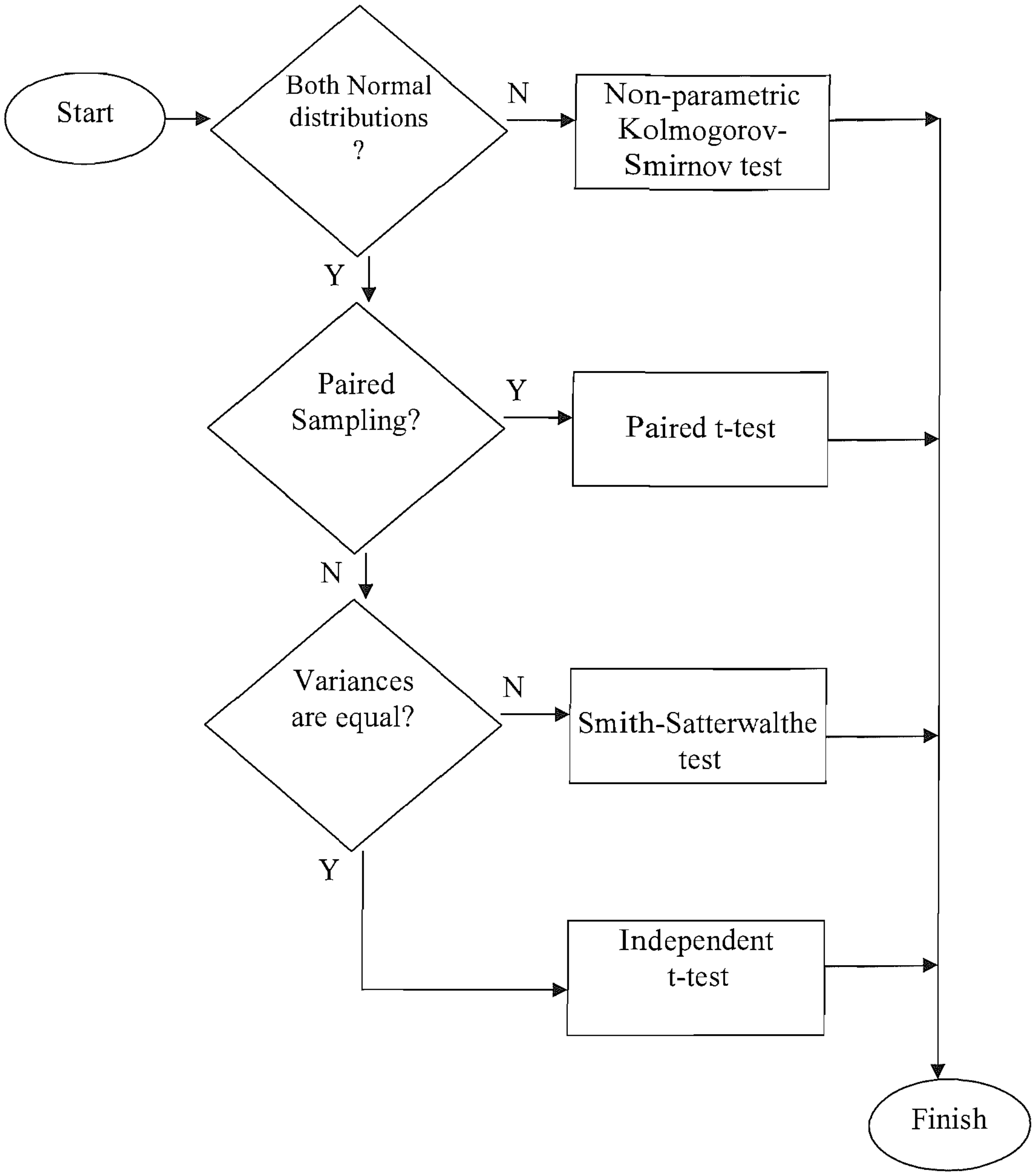


Figure 7.1: Hypothesis test flow chart

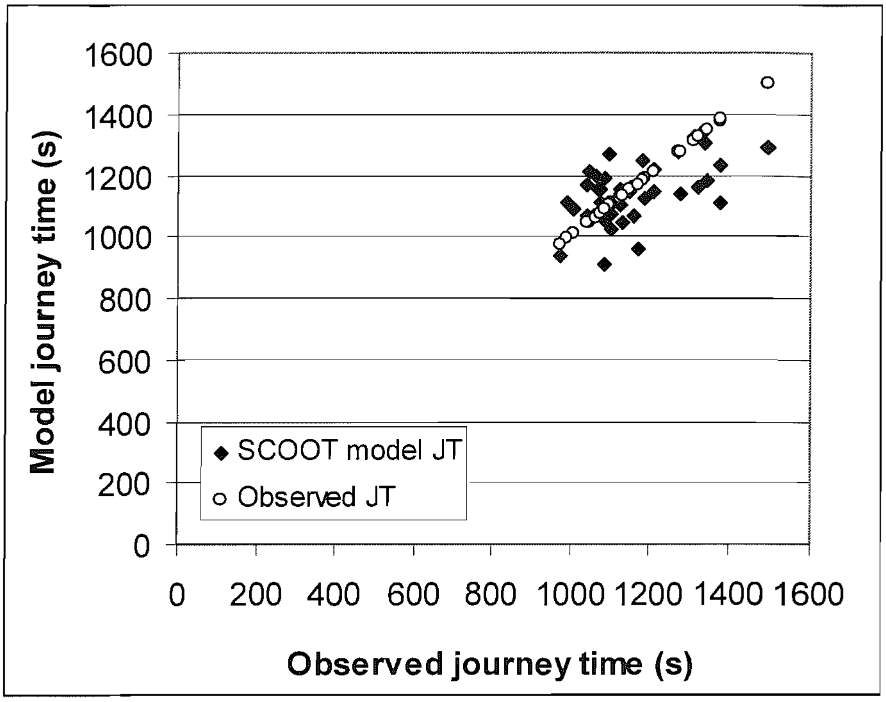


Figure 7.2: Route-based SCOOT regression model estimates against field data

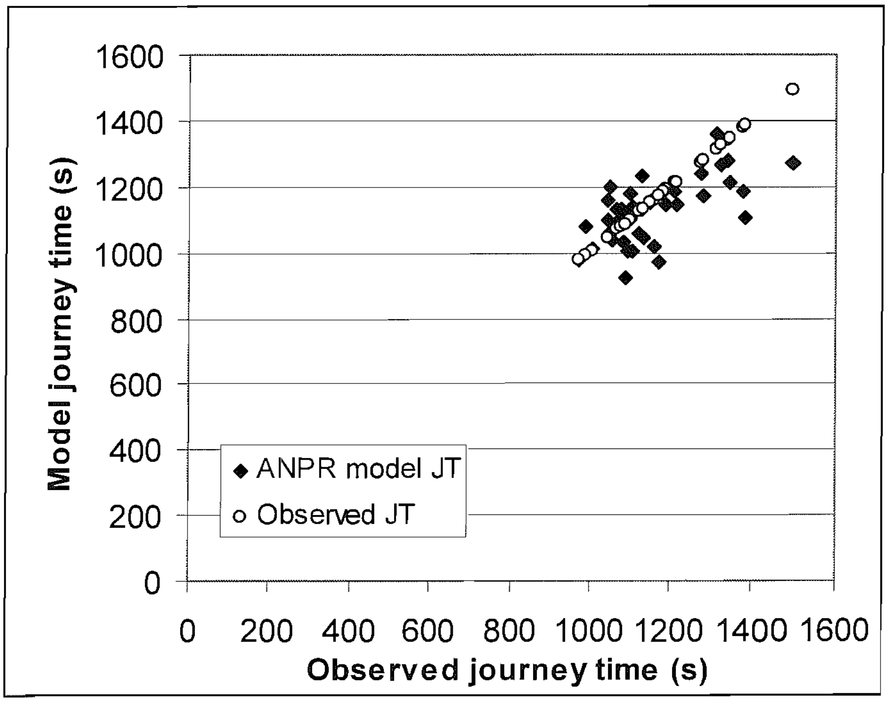
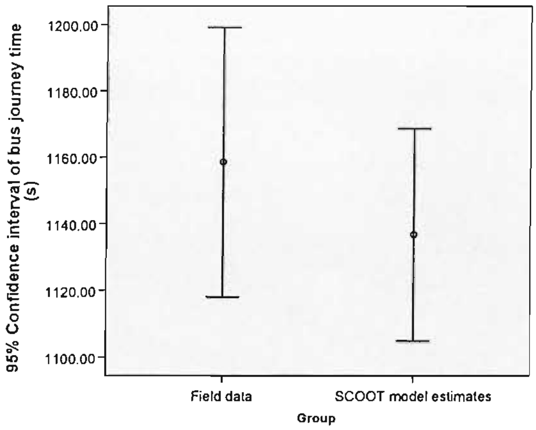


Figure 7.3: Route-based ANPR regression model estimates against field data



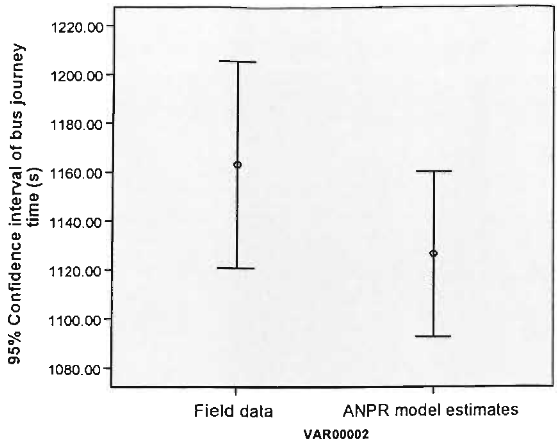


Figure 7.4: Confidence intervals of field data and route-based models estimates

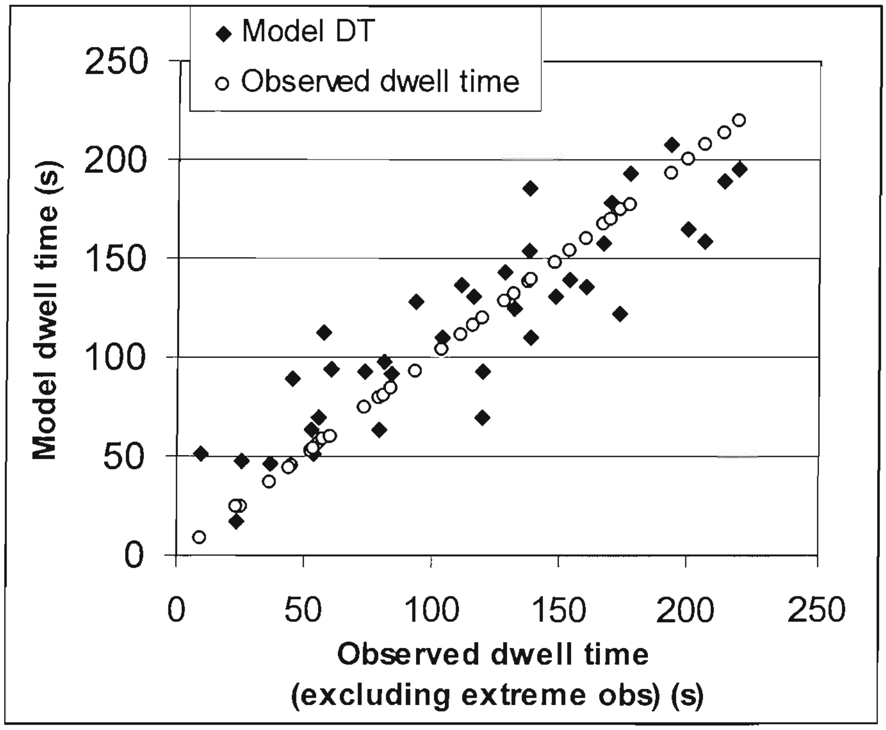


Figure 7.5: Dwell time model estimates against field data

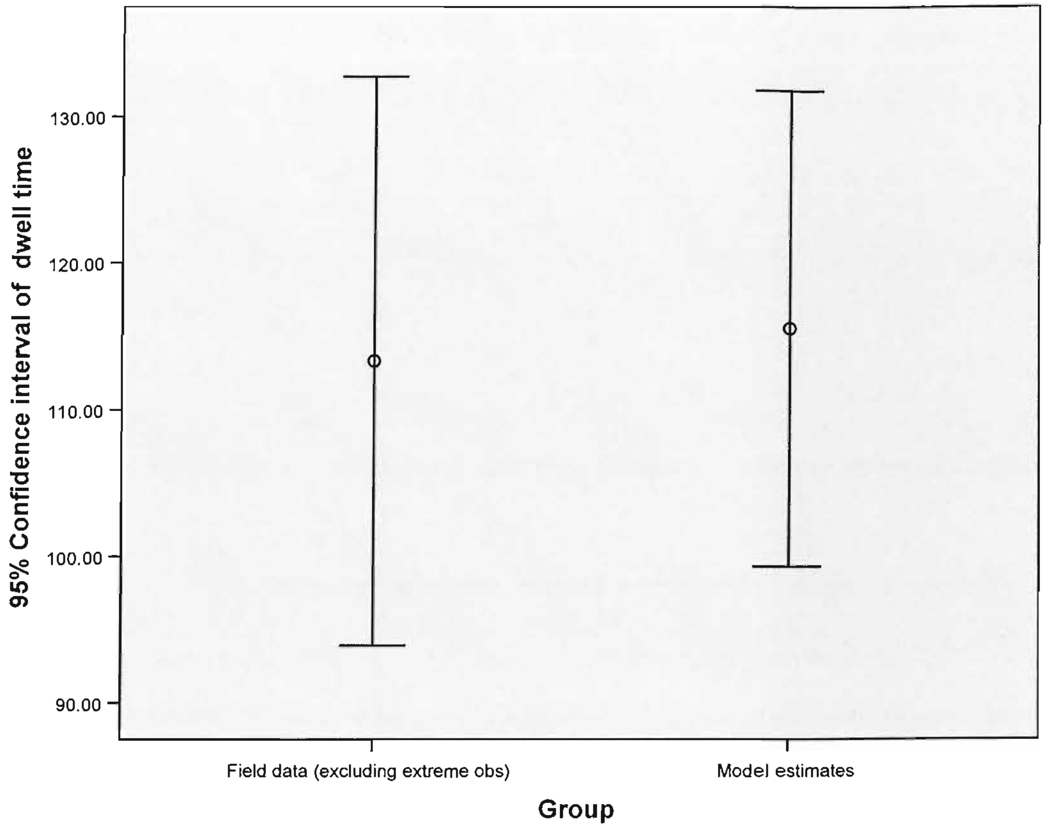


Figure 7.6: Confidence intervals of dwell time model estimates against field data

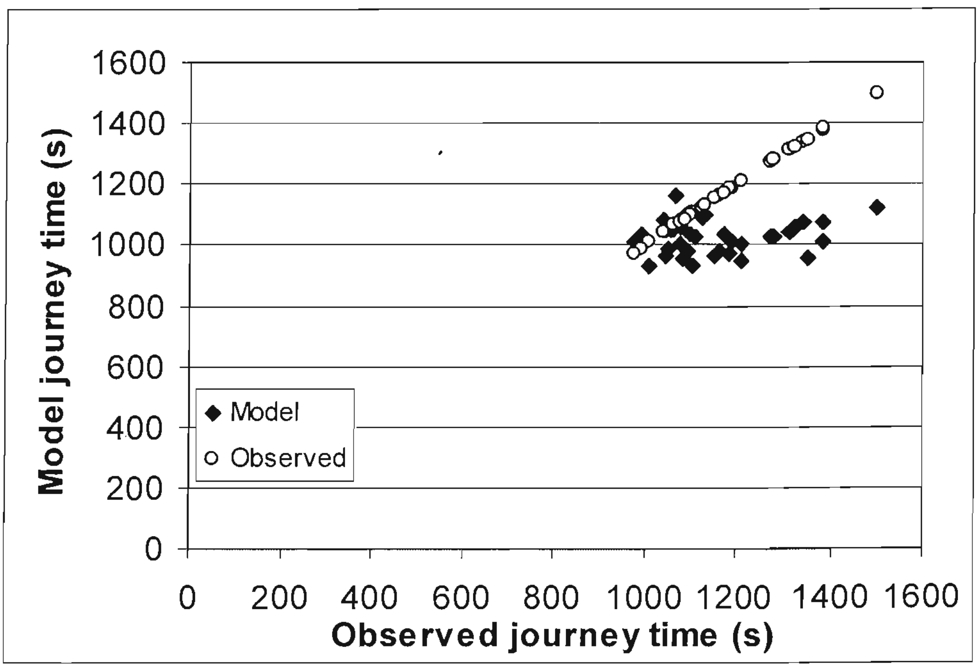


Figure 7.7: Link-based regression model estimates against field data

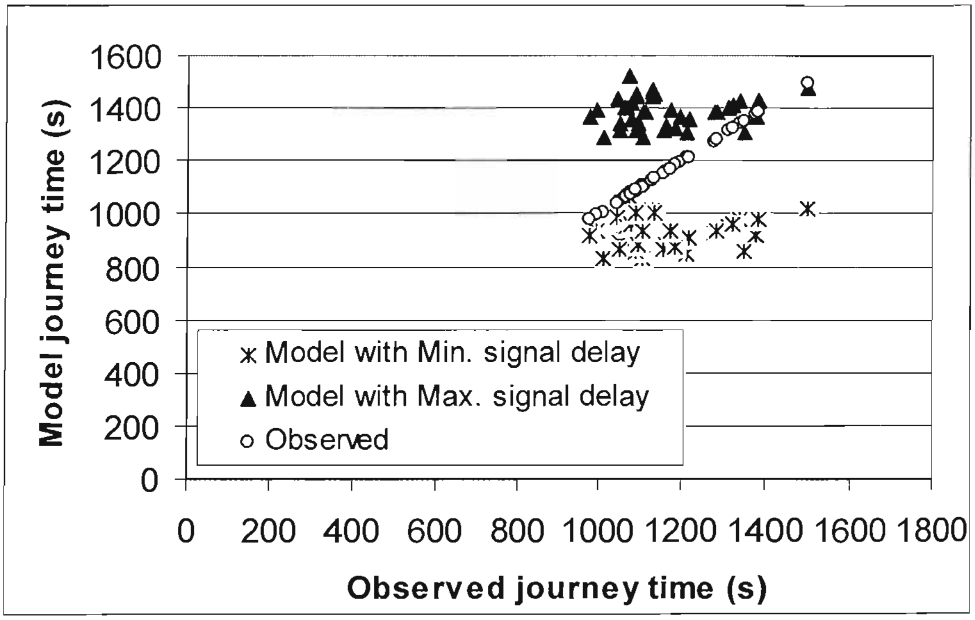


Figure 7.8: Adjusted link-based regression model estimates against field data

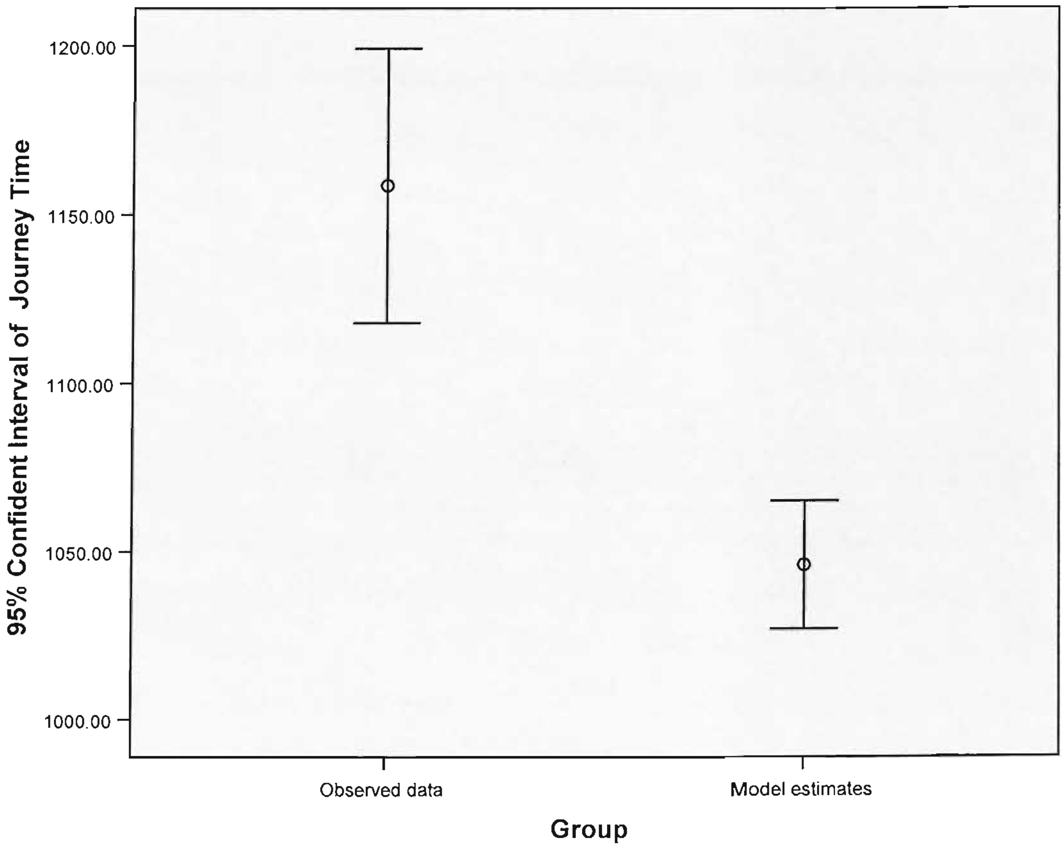


Figure 7.9: Confidence intervals of field data and link-based model estimates

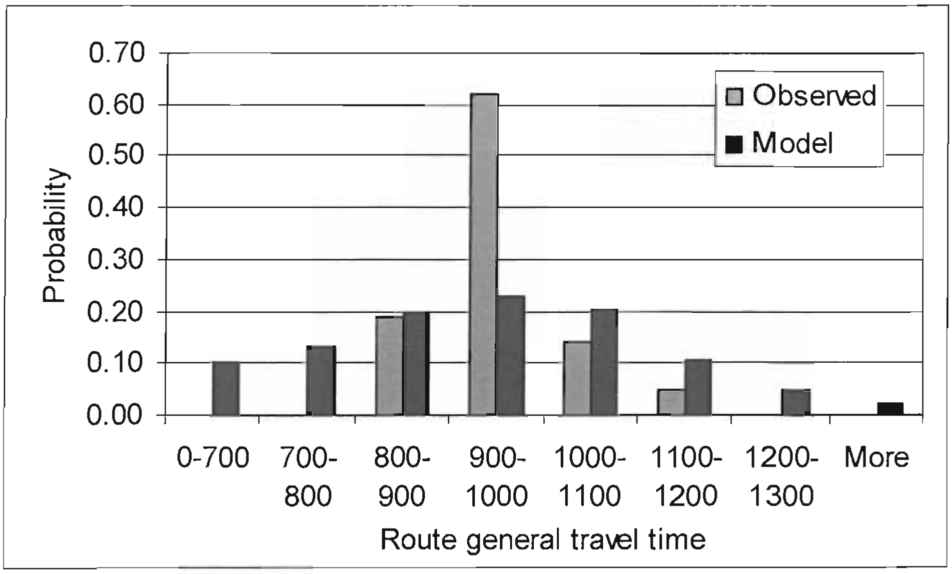


Figure 7.10: The probability of modular estimates of general travel time against field data

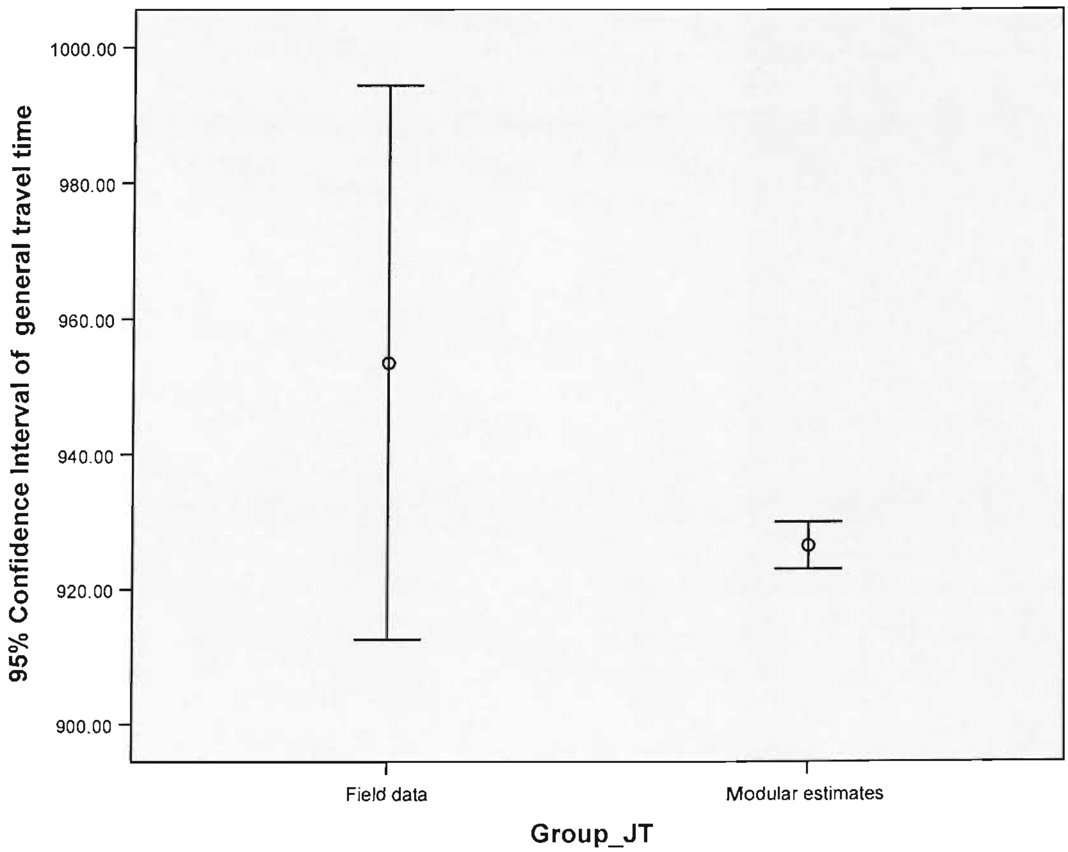


Figure 7.11: Confidence intervals of route-based general travel time modular estimates against field data

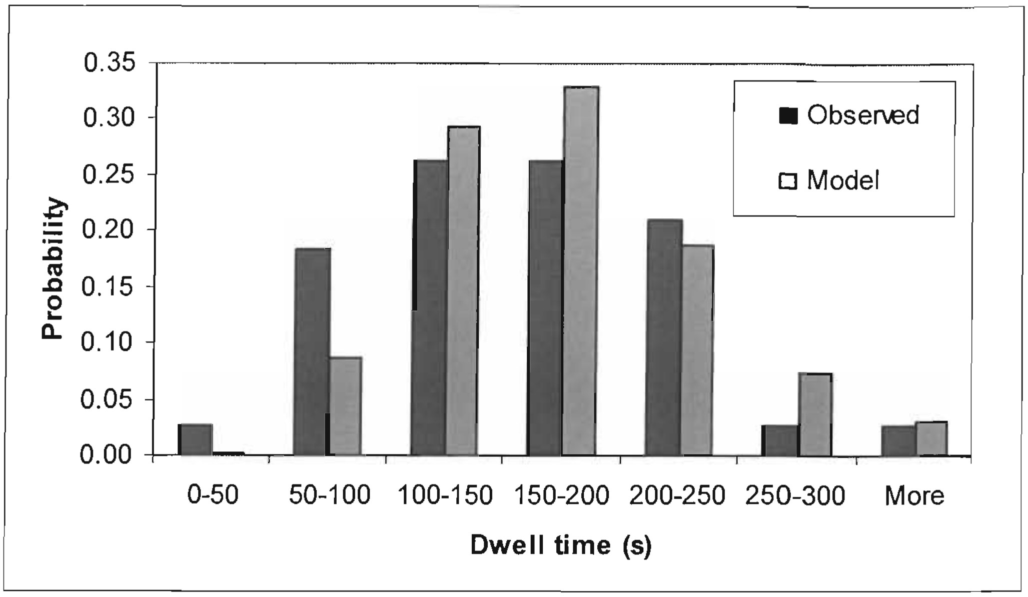


Figure 7.12: The probability of modular estimates of dwell time against field data

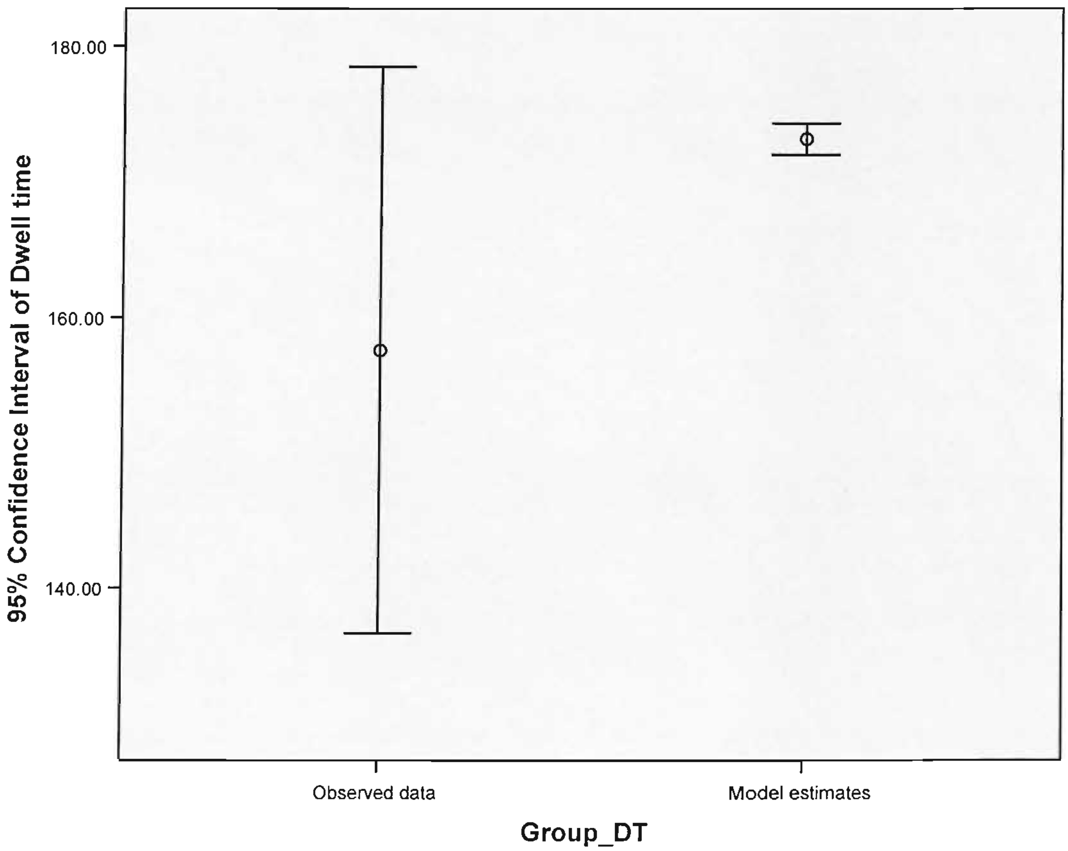


Figure 7.13: Confidence intervals of route-based dwell time modular estimates against field data

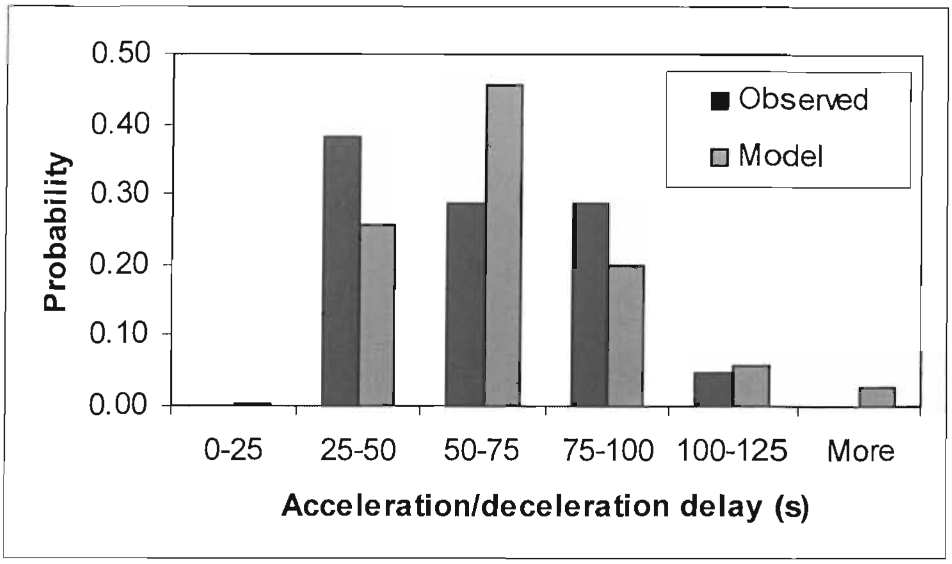


Figure 7.14: The probability of modular estimates of acc/dec delay against field data

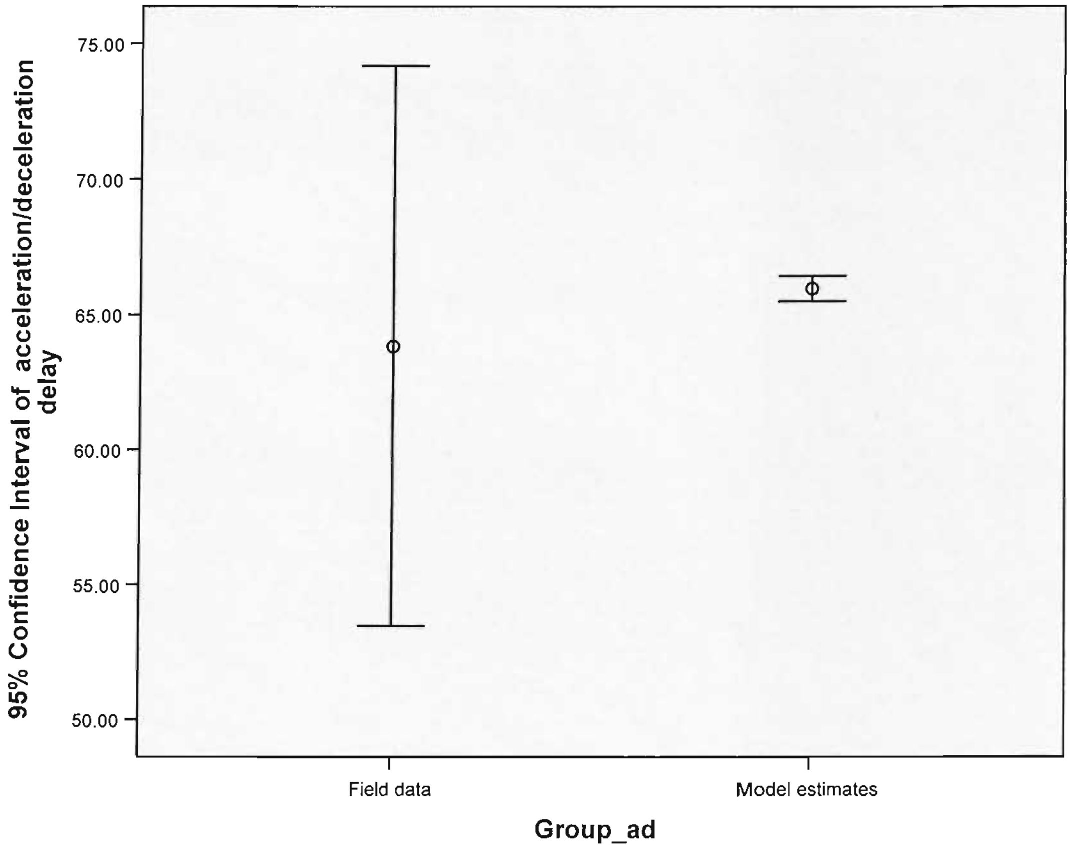


Figure 7.15: Confidence intervals of route-based acceleration/deceleration delay modular estimates against field data

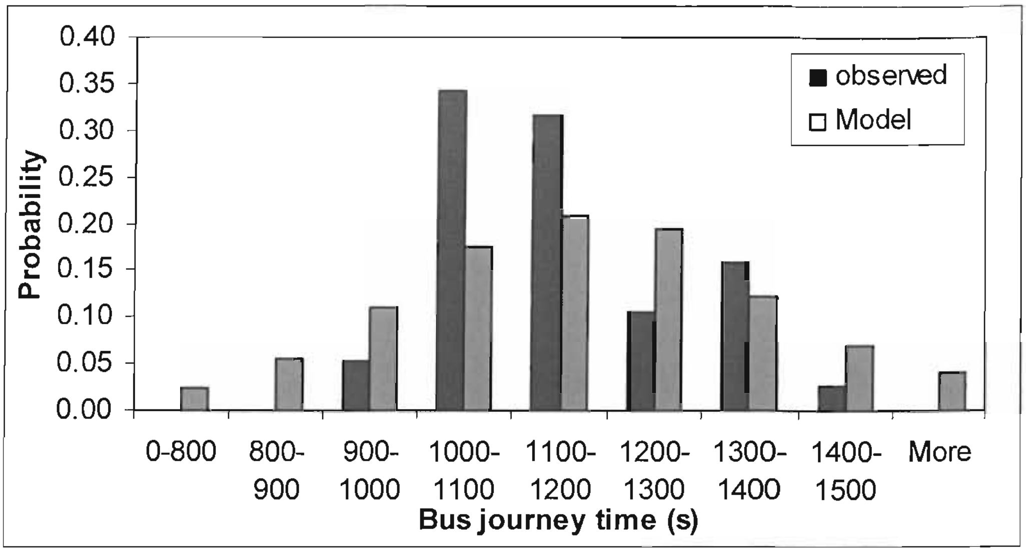


Figure 7.16: The probability of bus journey time route-based model estimates against field data

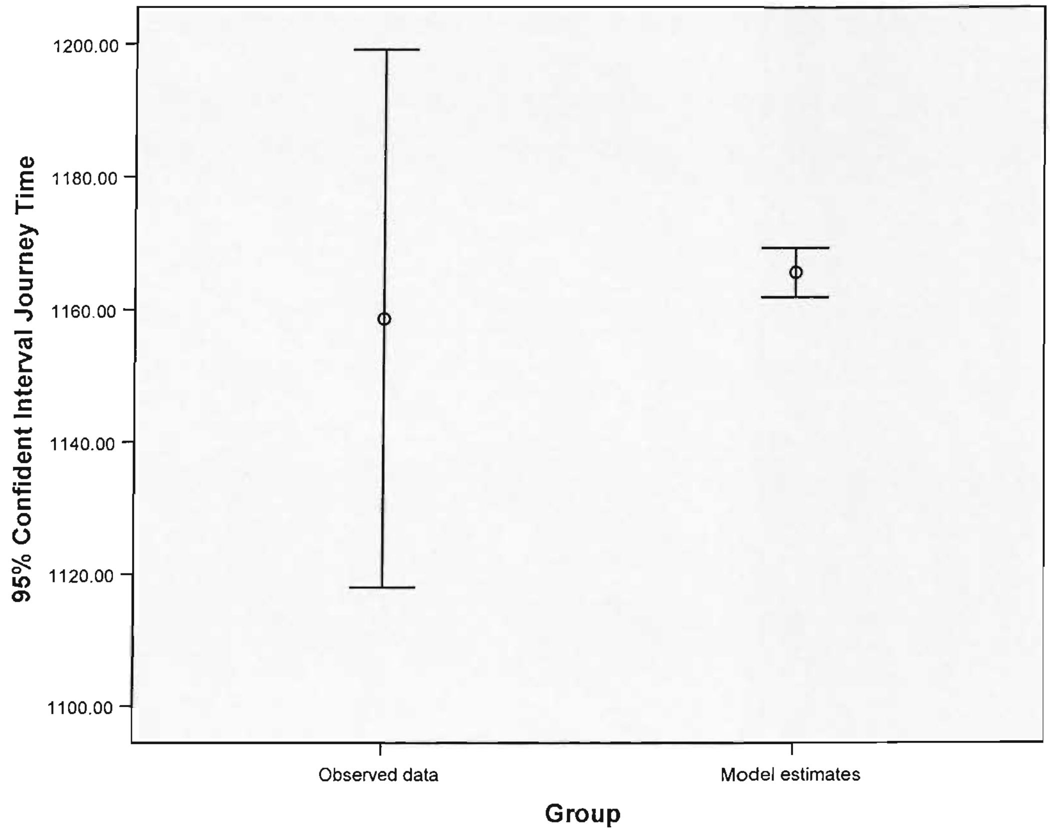


Figure 7.17: Confidence intervals of bus journey time route-based model estimates against field data

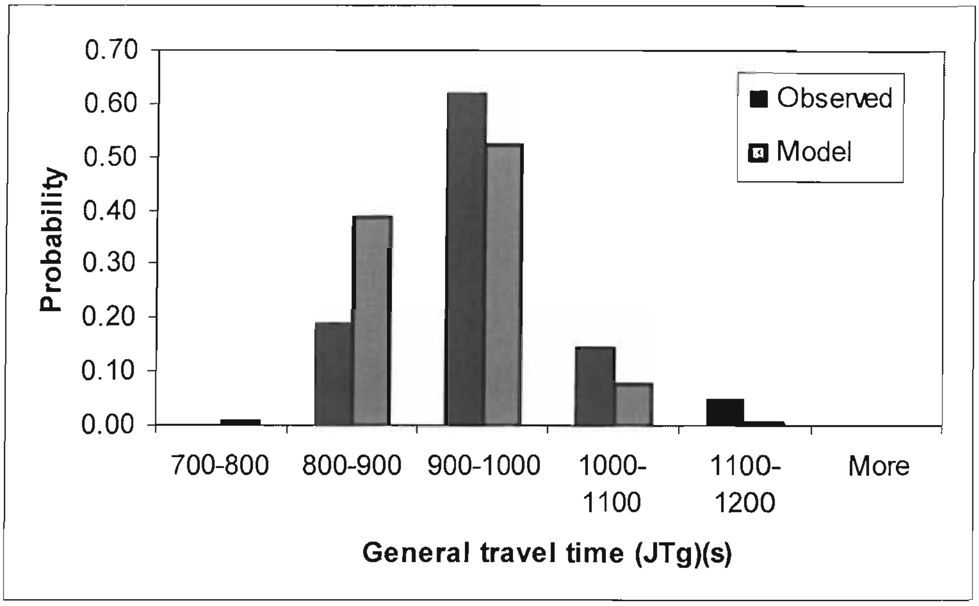


Figure 7.18: The probability of link-based general travel time combined with signal delay modular estimates against field data

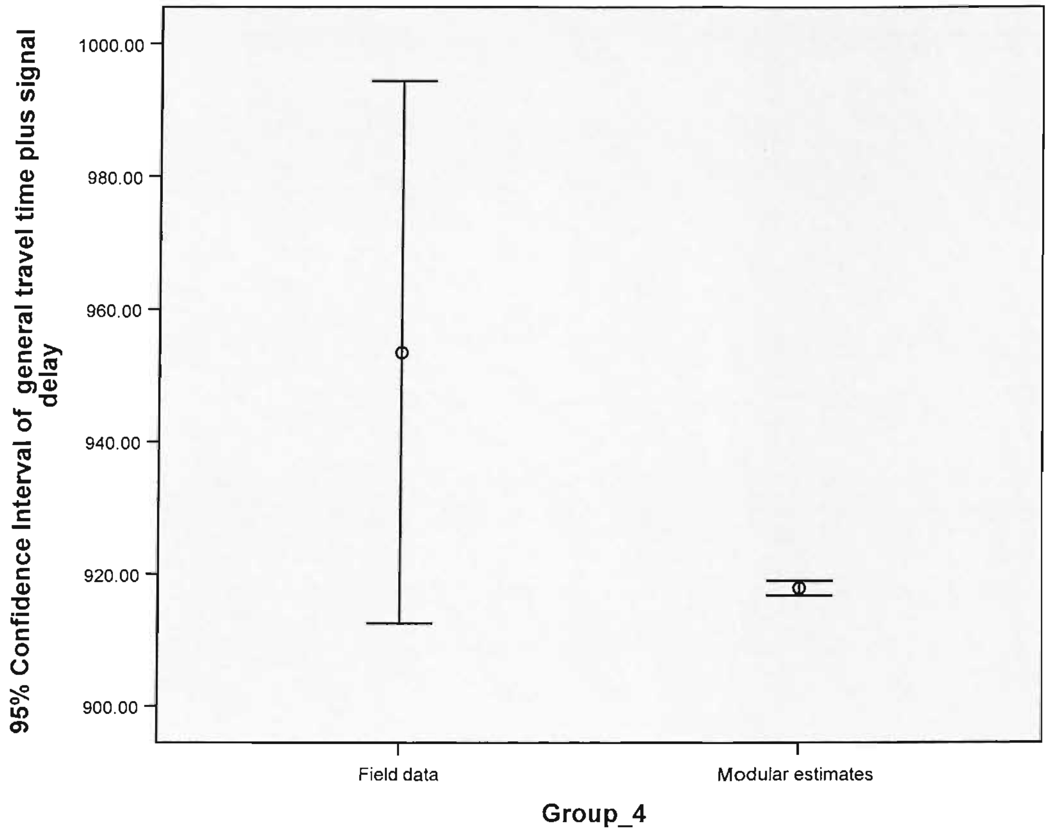


Figure 7.19: Confidence intervals of link-based general travel time combined with signal delay modular estimates against field data

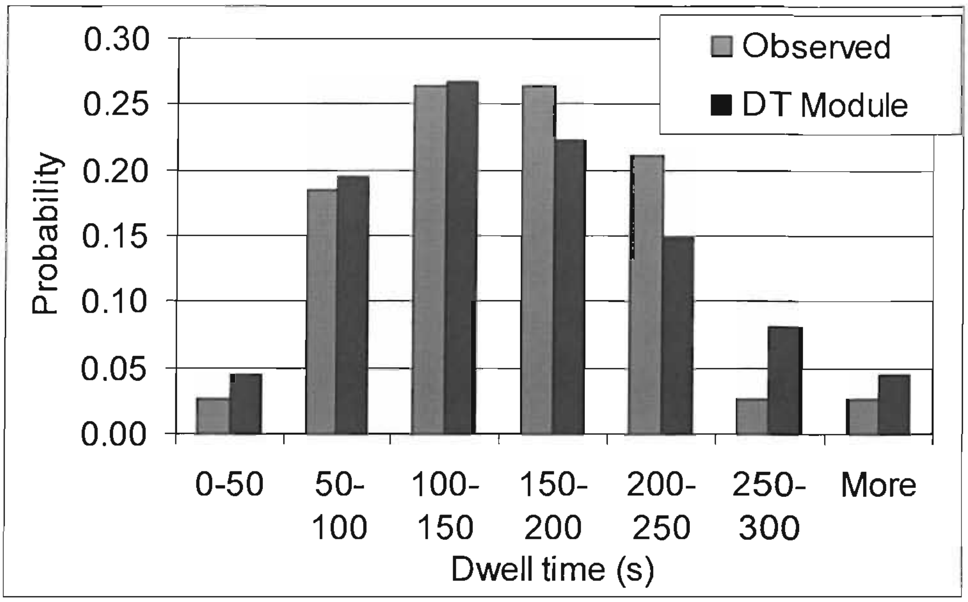


Figure 7.20: The probability of link-based dwell time module estimates against field data

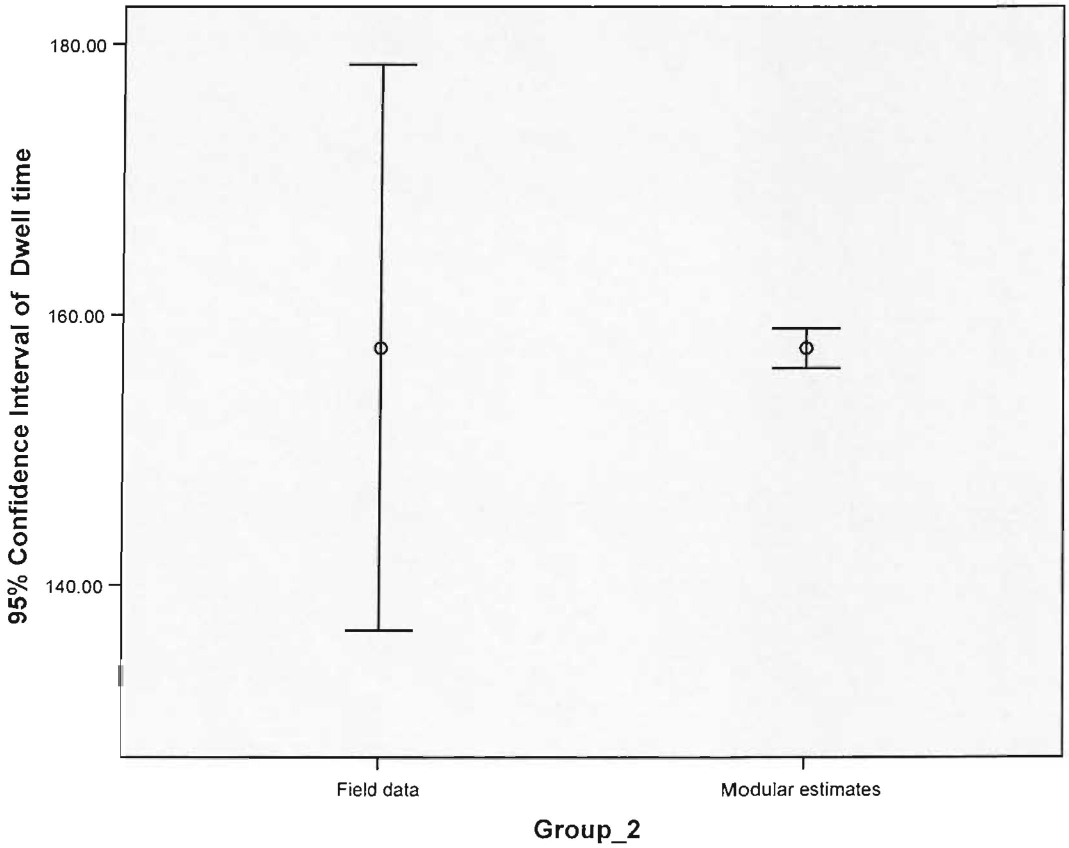


Figure 7.21: Confidence intervals of link-based dwell time module estimates against field data

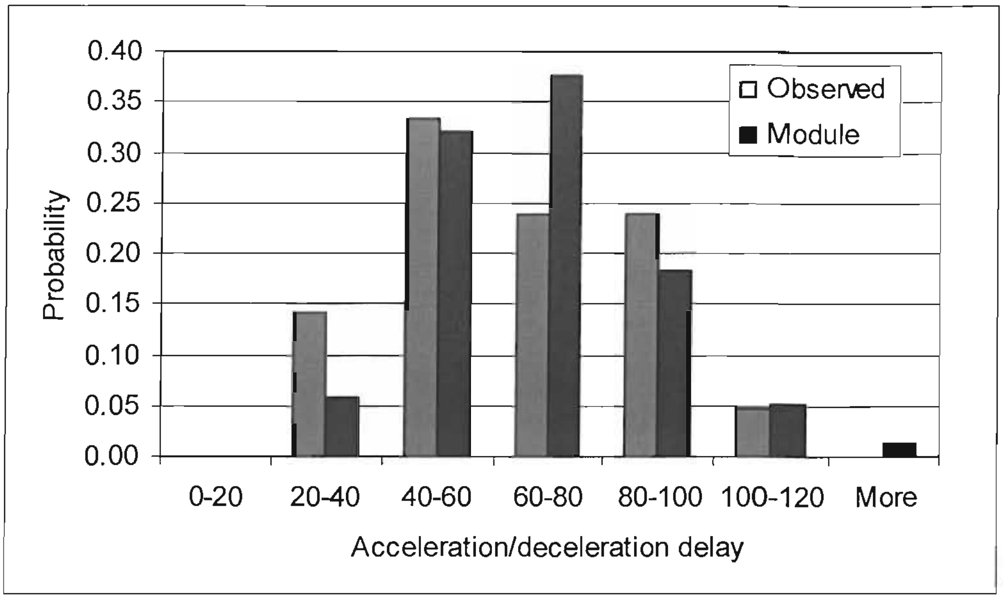


Figure 7.22: The probability of link-based acceleration/deceleration module estimates against field data

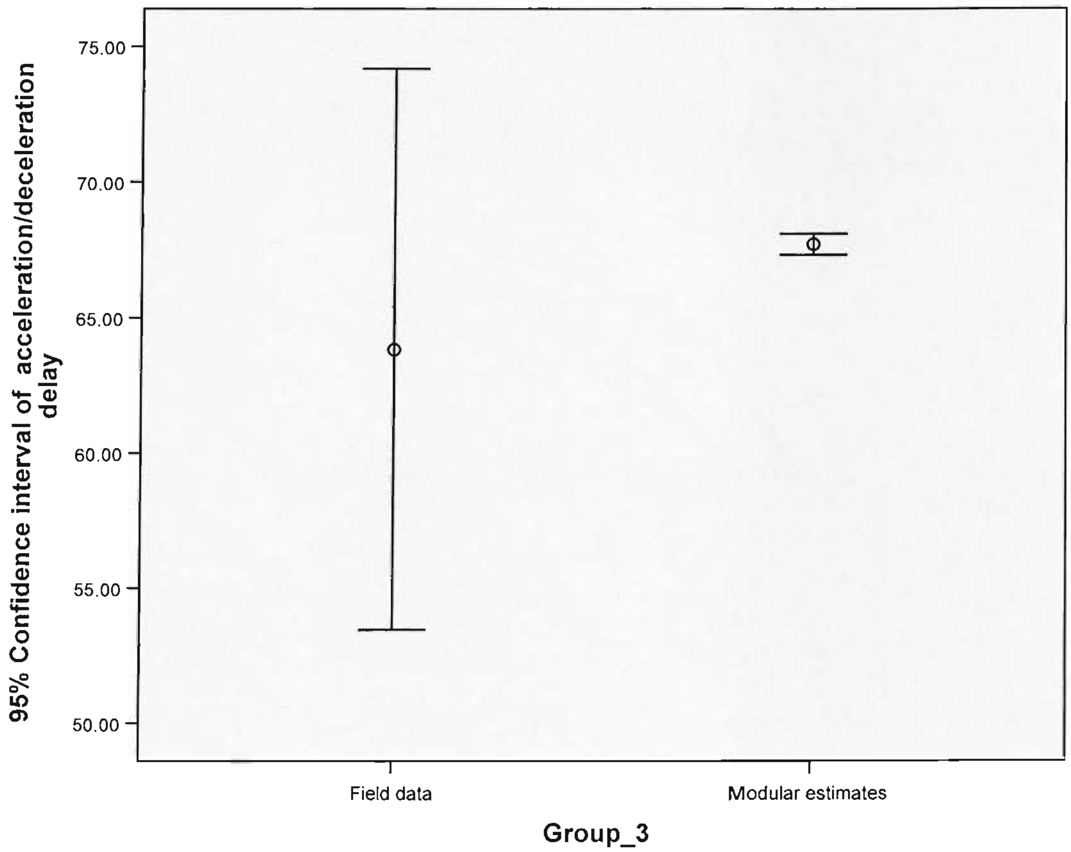


Figure 7.23: Confidence intervals of link-based acceleration/deceleration module estimates against field data

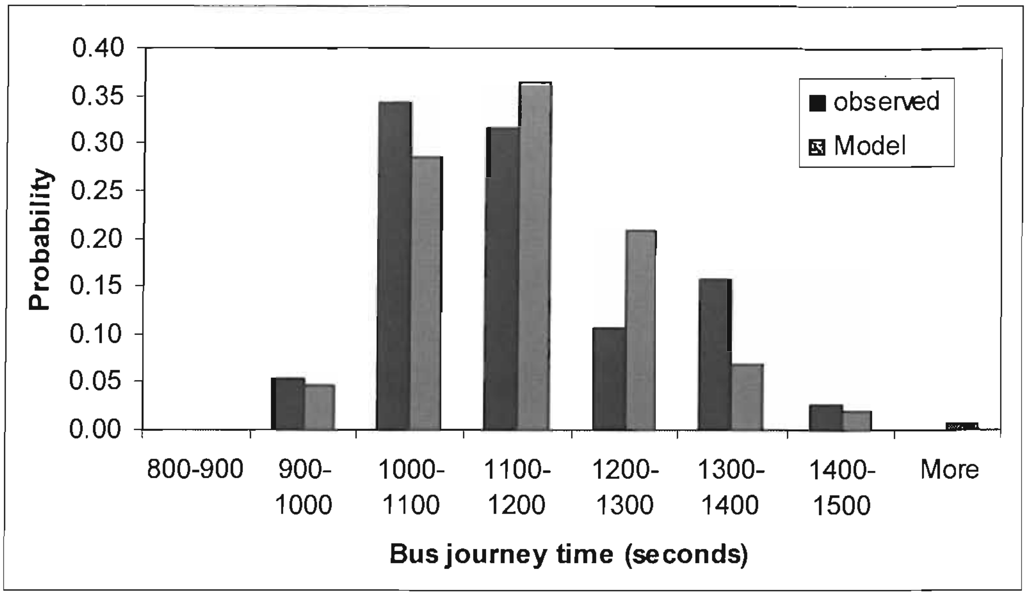


Figure 7.24: The probability of link-based bus journey time model estimates against field data

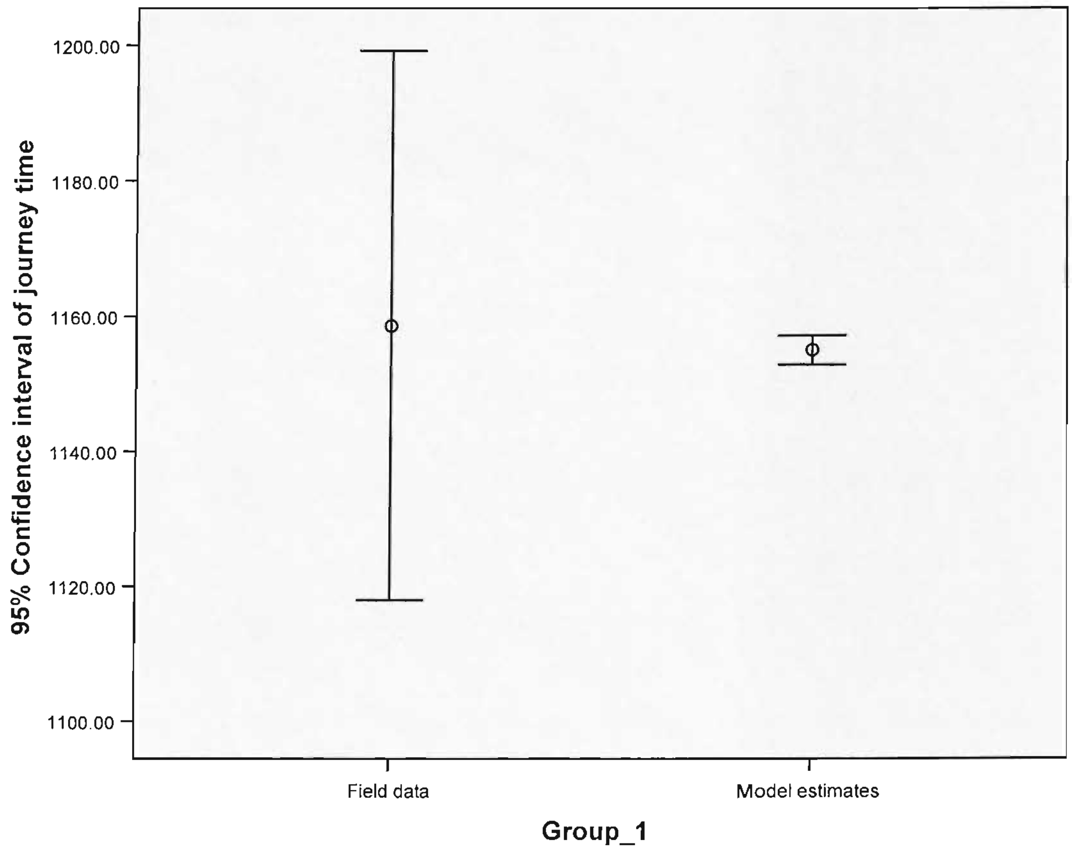


Figure 7.25: Confidence intervals of link-based journey time model estimates against field data

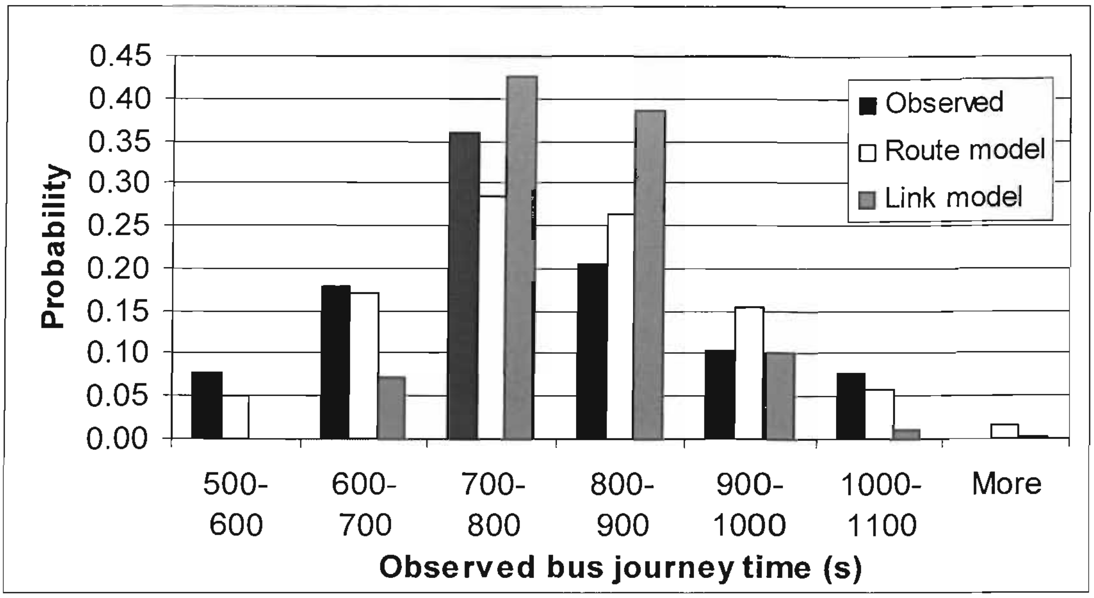


Figure 7.26: The probability of route- and link-based journey time model estimates against field data

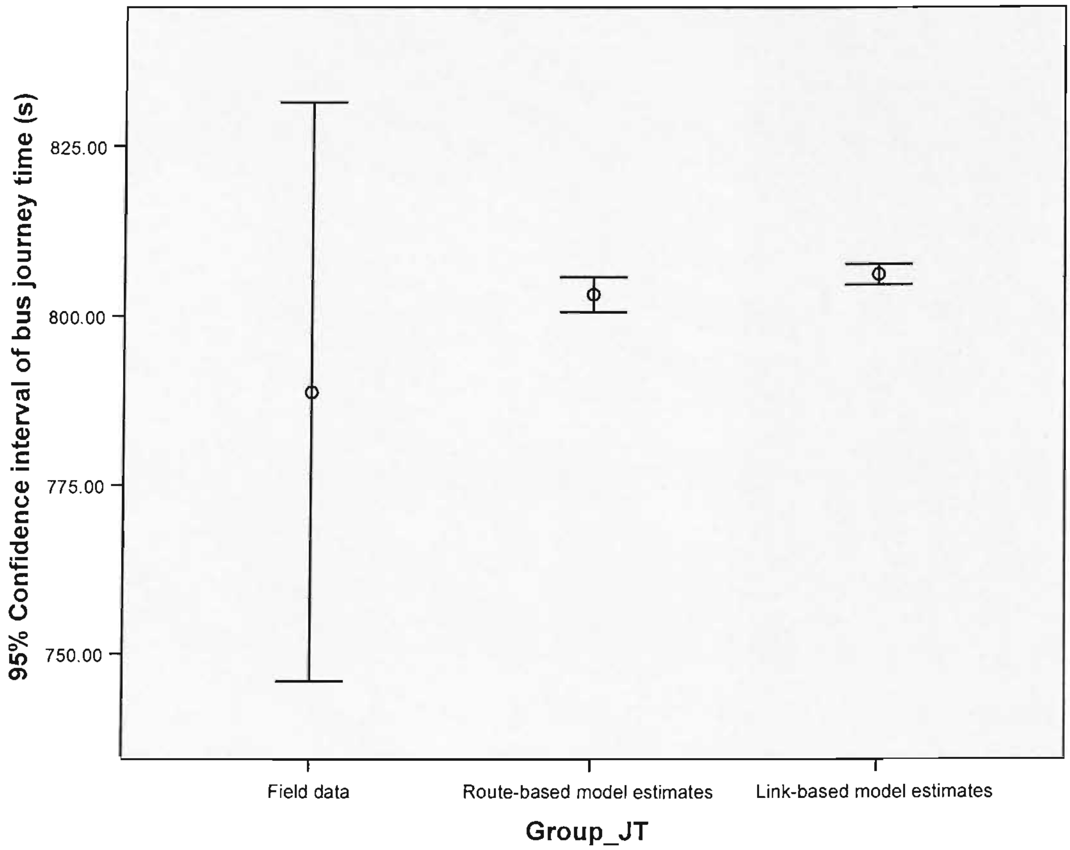


Figure 7.27: Confidence intervals of route- and link-based journey time model estimates against field data

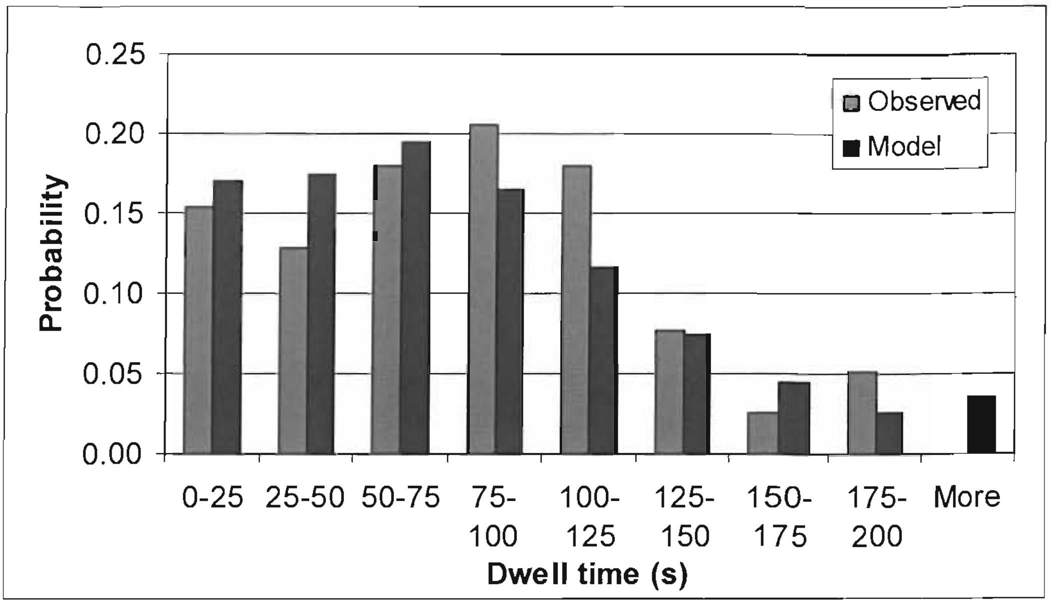


Figure 7.28: The probability of dwell time model estimates against field data

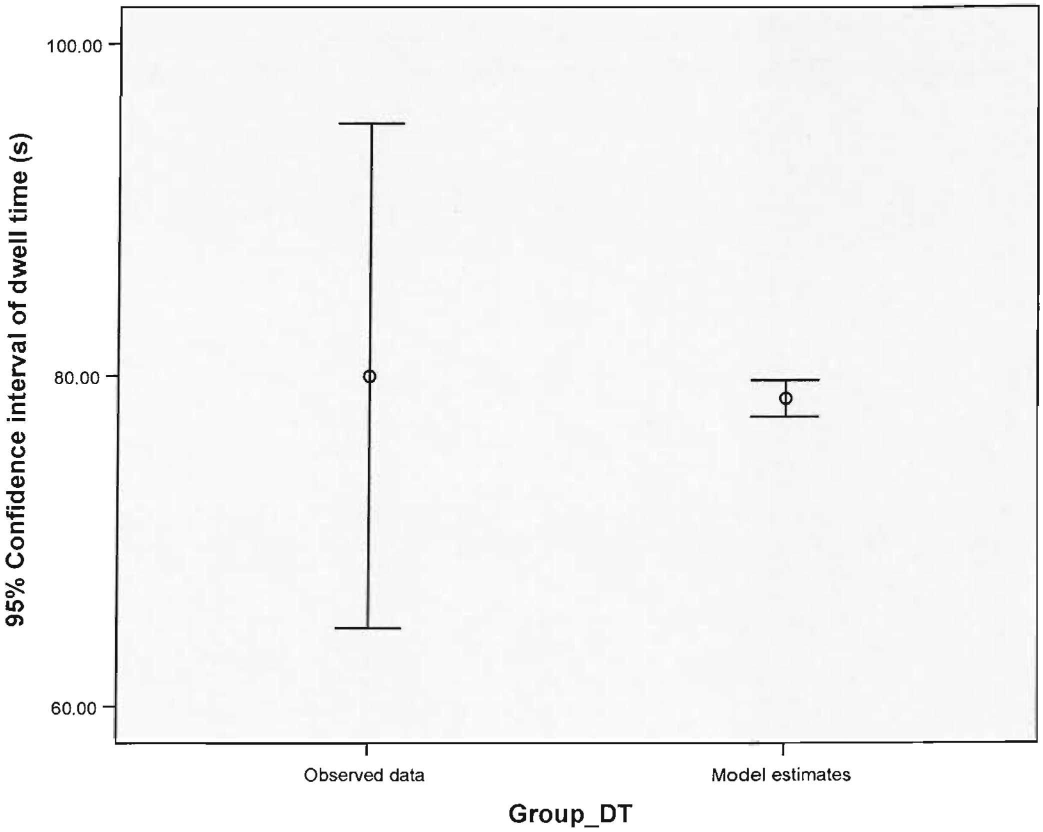


Figure 7.29: Confidence intervals of dwell time model estimates against field data

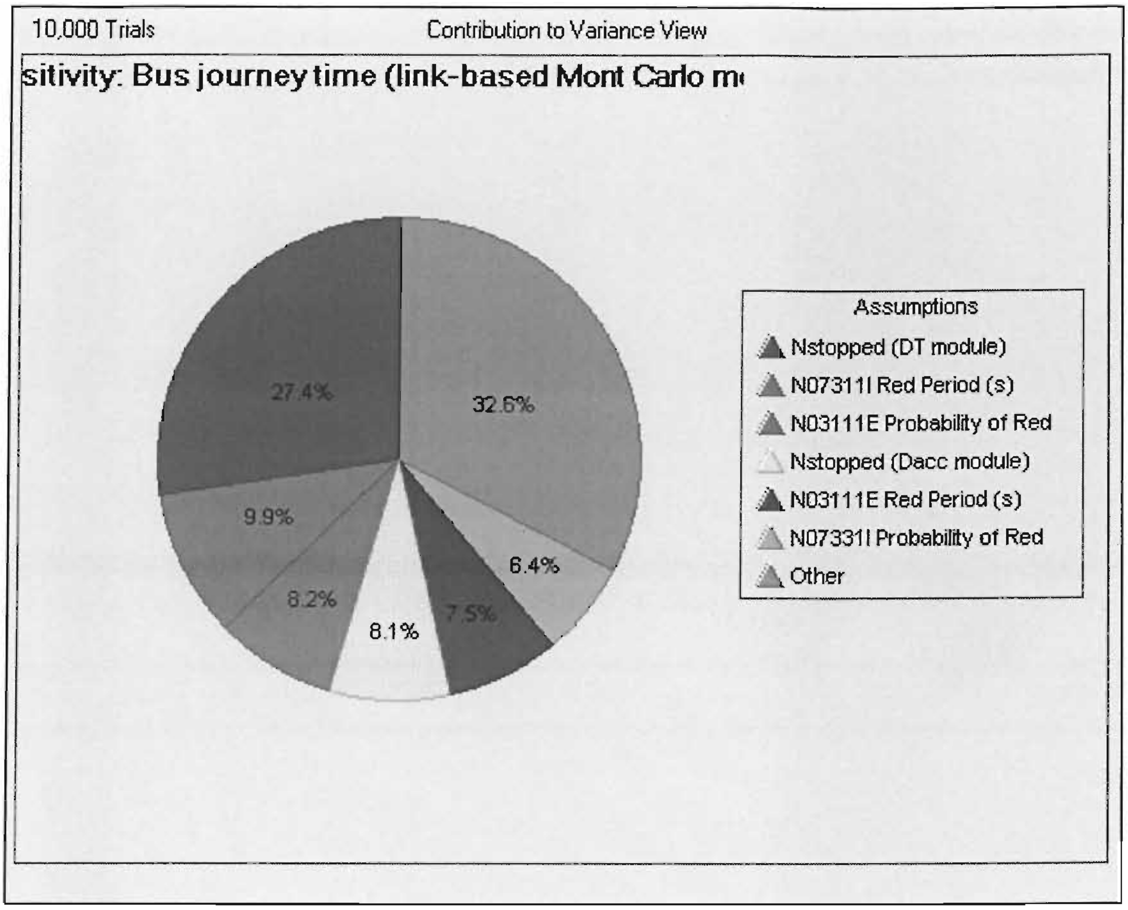


Figure 8.1: Percentage of variance contribution towards bus journey time of critical variables

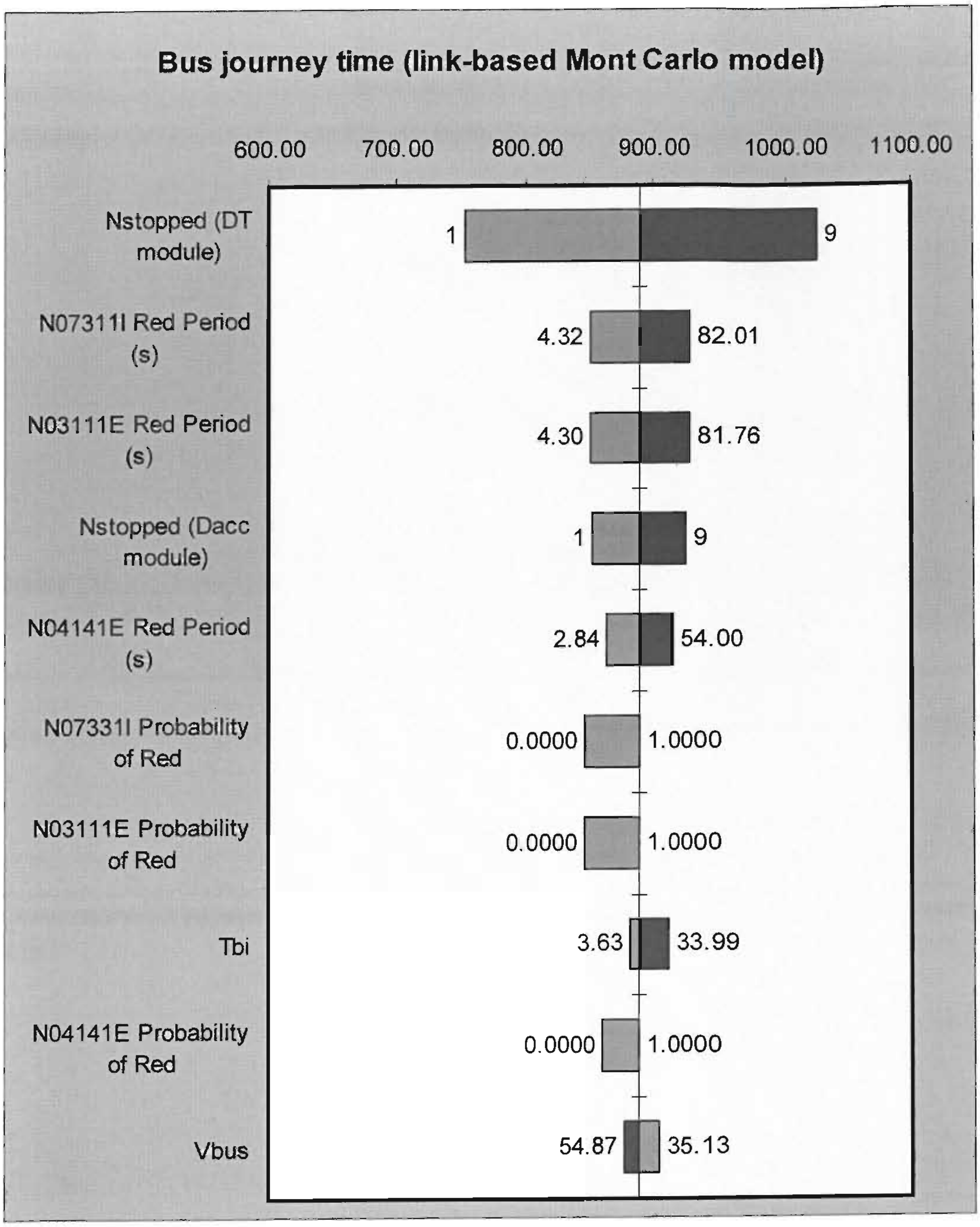


Figure 8.2: The key variables in parametric analysis

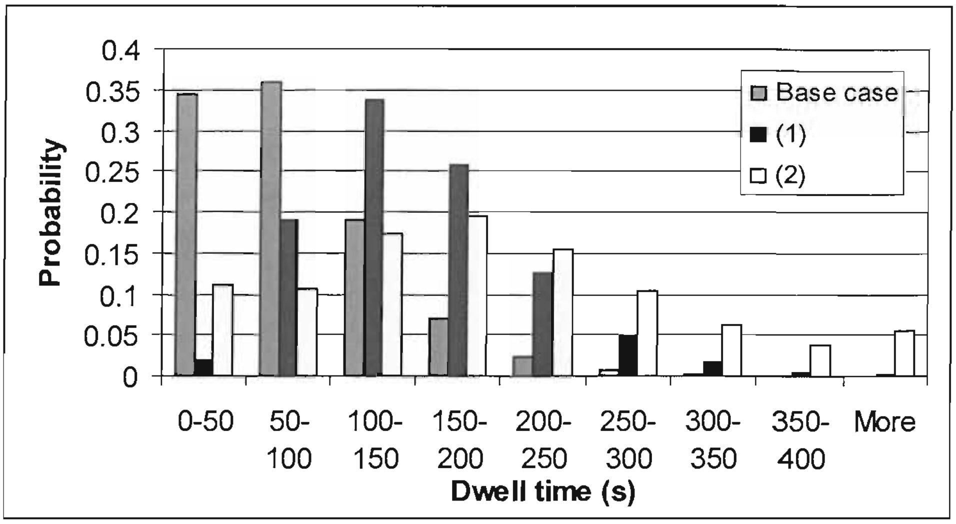


Figure 8.3: Comparison of changes in dwell time in passenger demand scenarios

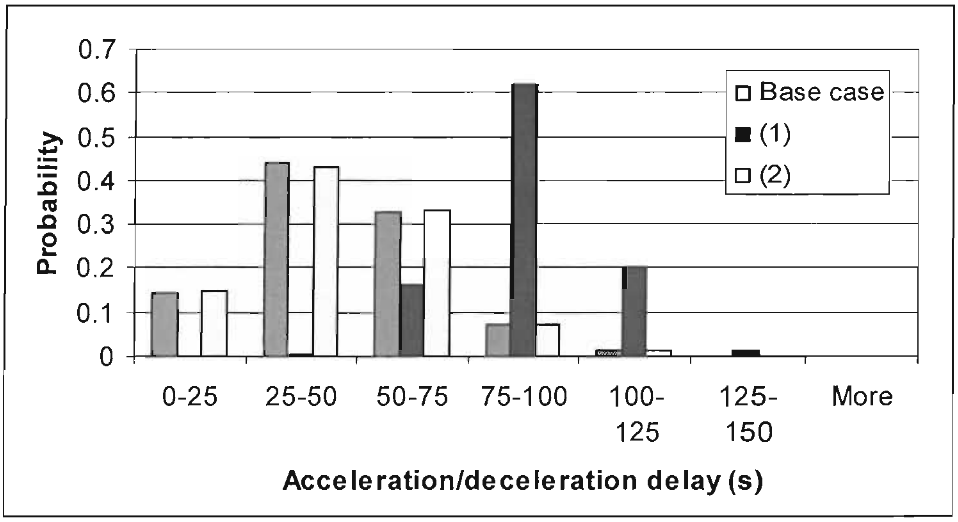


Figure 8.4: Comparison of changes in acceleration/deceleration delay in passenger demand scenarios

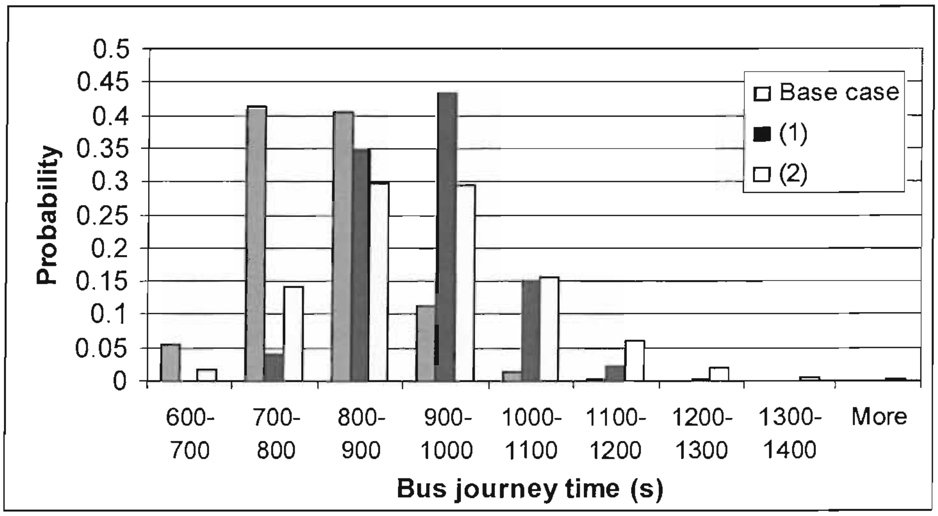


Figure 8.5: Comparison of changes in bus journey time in passenger demand scenarios

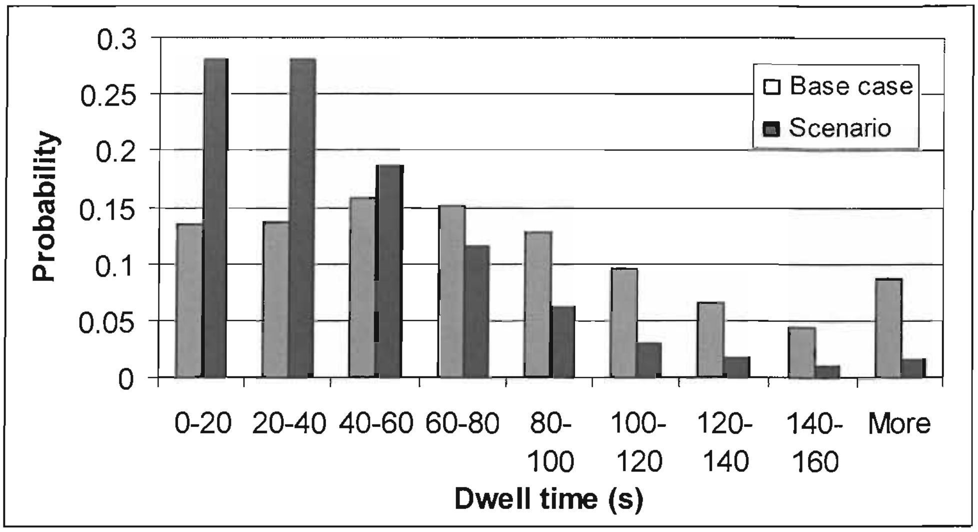


Figure 8.6: Comparison of changes in dwell time changing boarding time scenario

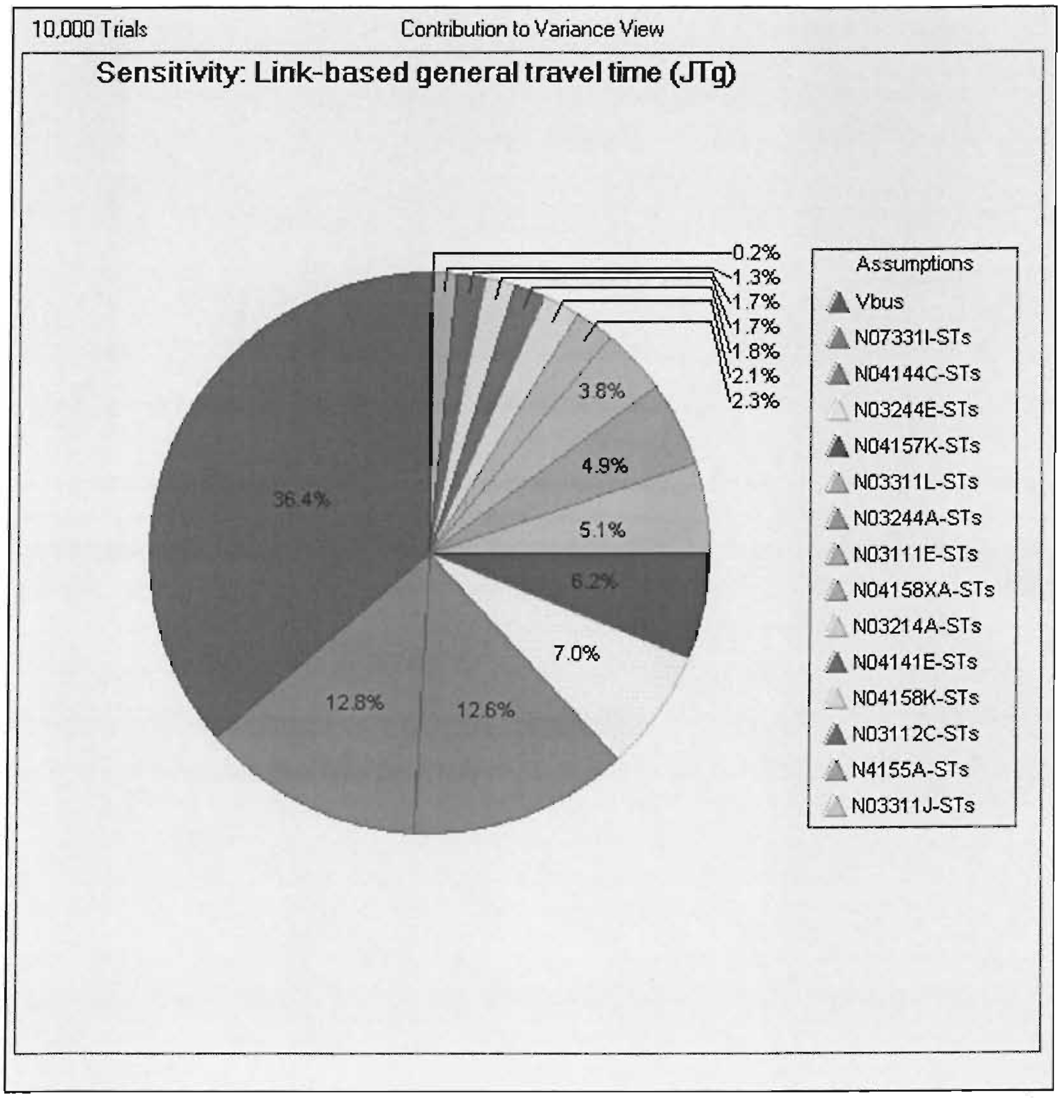
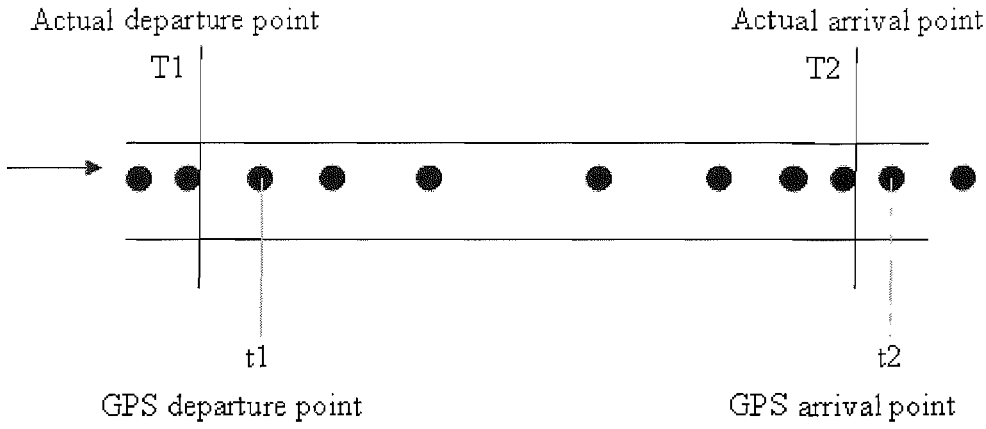


Figure 8.7: Percentage of variance contribution of each link towards bus general travel time



*The black dots indicate GPS points

Figure 9.1: The possible error on the identifying the departure and arrival points of GPS data

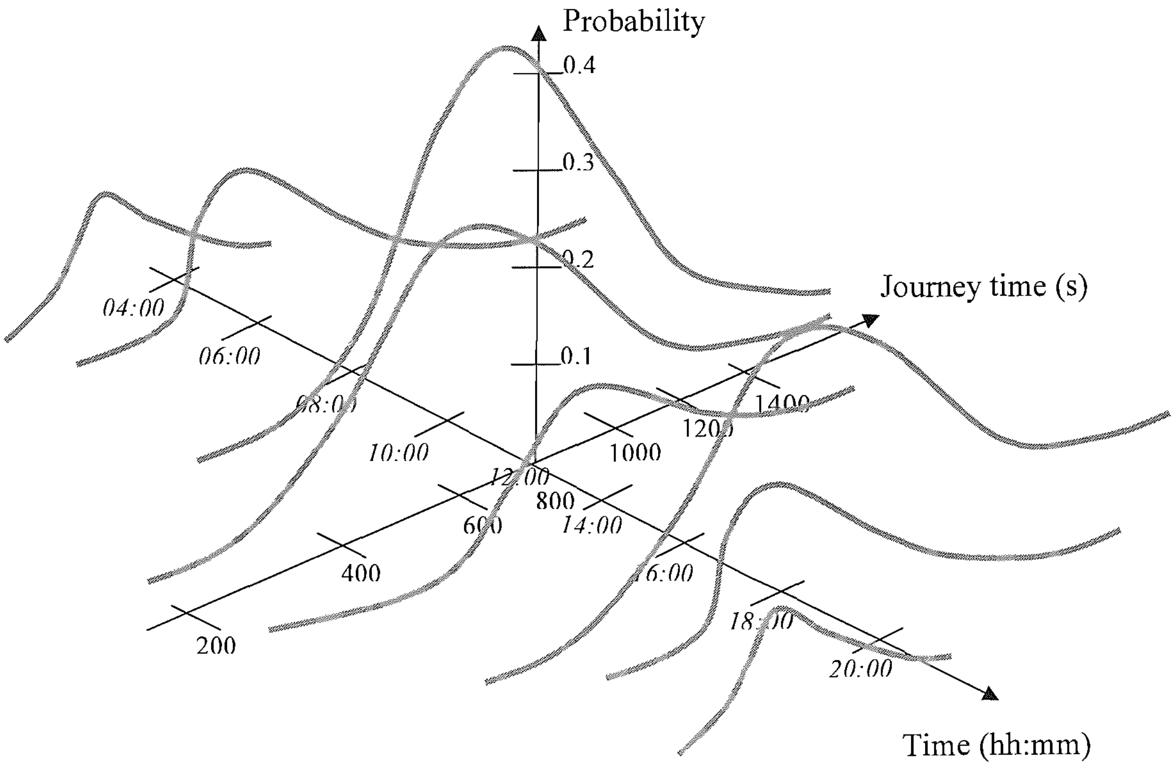


Figure 9.2: Example of 3D diagram of probability distribution against journey time estimation and time series.

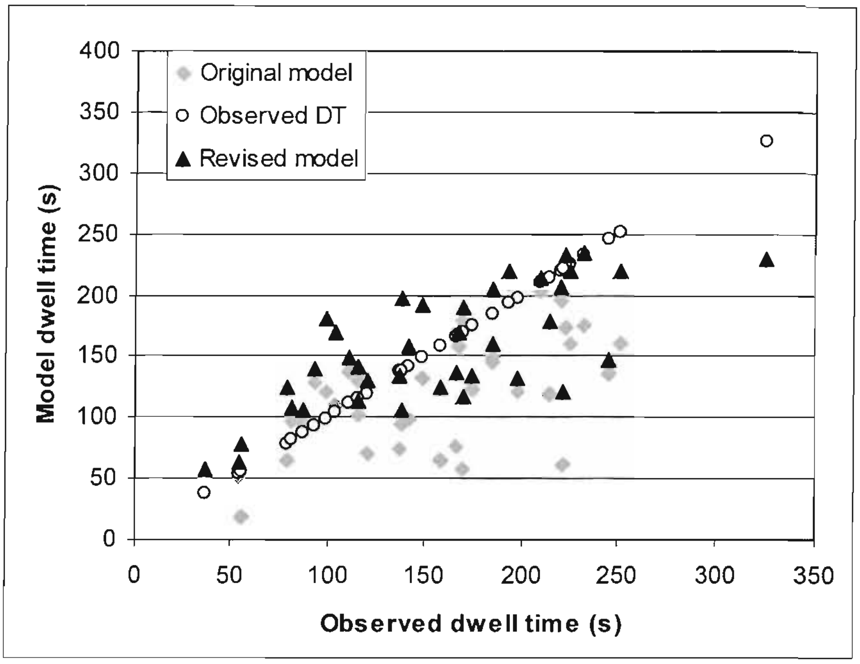
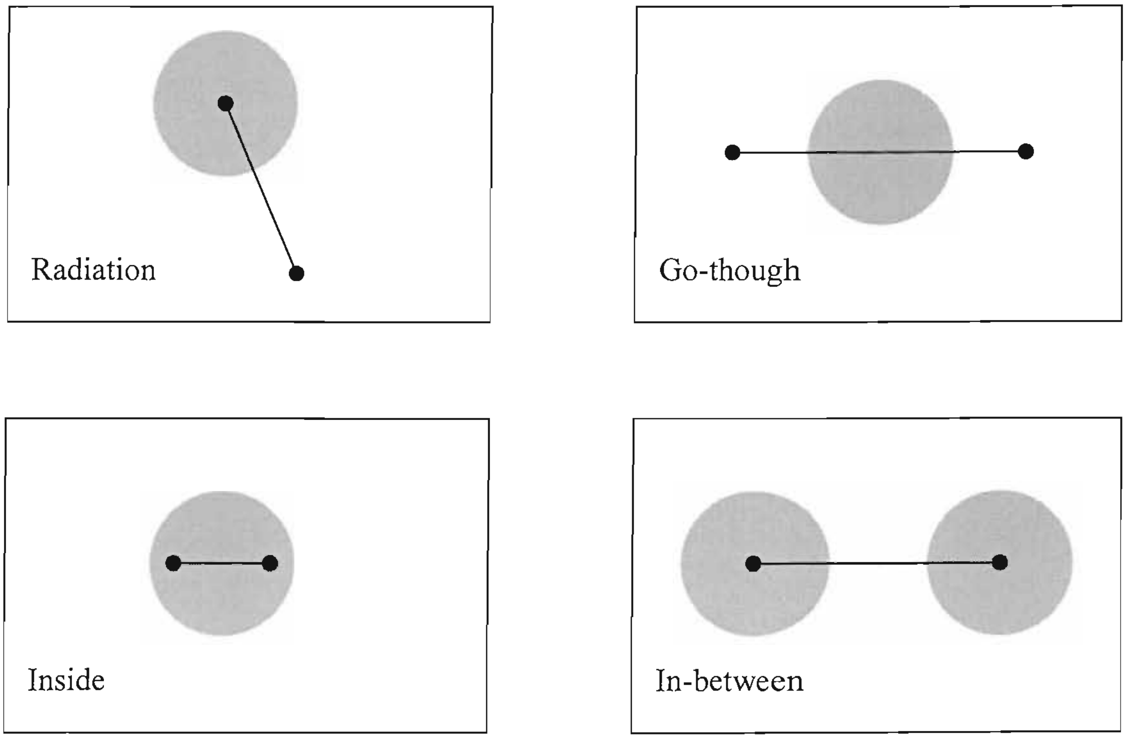


Figure 9.3: Revised dwell time model and original model against field data



A grey circle indicates a city centre, the line shows the bus route, and the black spot indicates departure or end point.

Figure 9.4: Different types of bus route

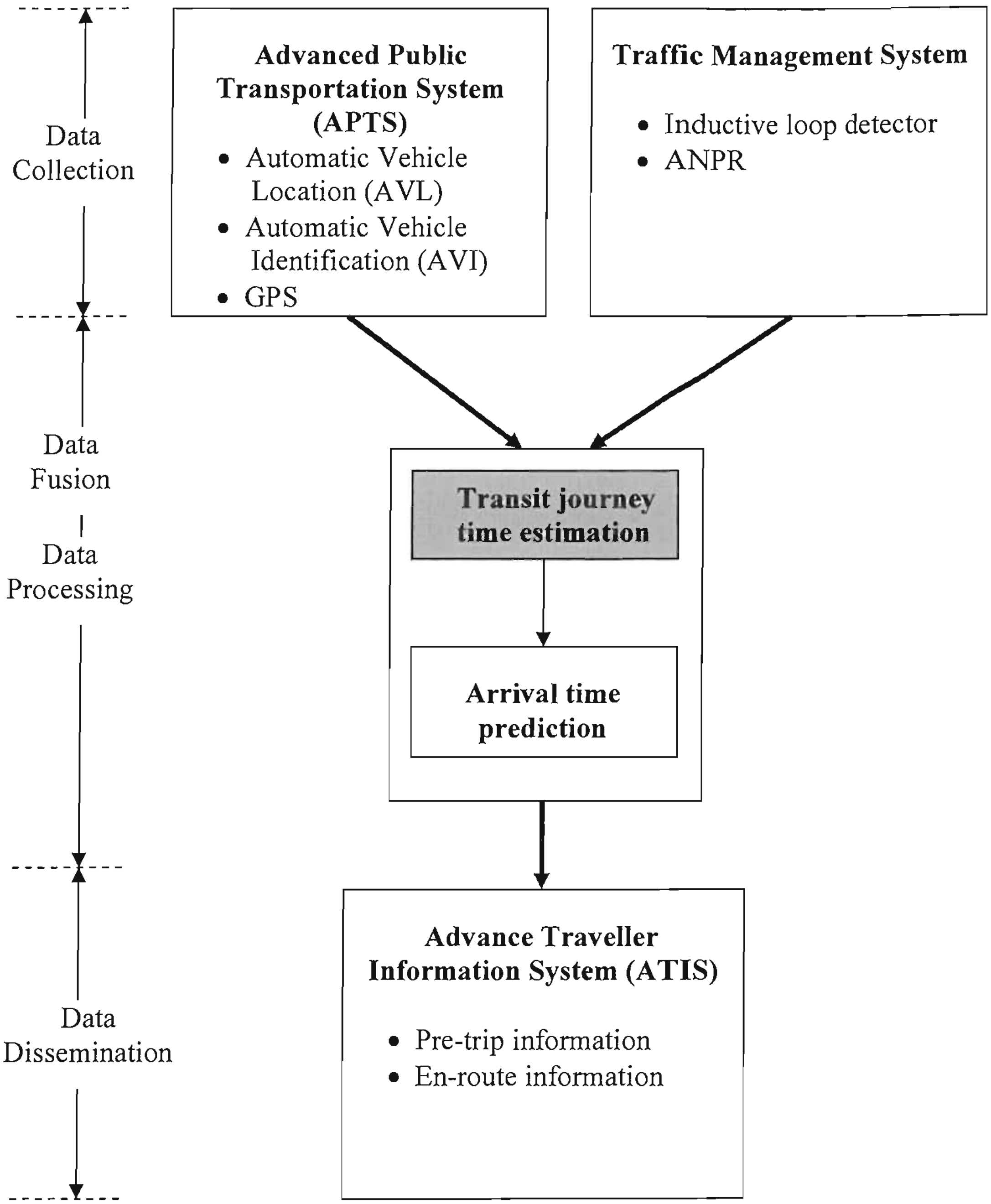


Figure 10.1: Data flow of bus journey time in the context of ITS

Appendix A

Process of GPS data integrated with GIS (ArcView)

Contents

Before analyzing the collected data, it is essential to show where GPS points on digital map are using GIS software, i.e. a trajectory of a bus can be displayed on particular survey road and a roughly idea of vehicle's speed according to the data frequency and density of GPS points can be achieved. Figure A.1 illustrates this concept for a bus driving along a route from left to right on map. It can be seen that there are two sites that have dense GPS points. On the left, it is a delay before signalized control junction, the bus might have a delay after a queue with deceleration first and then acceleration through the junction. On the right, there is a bus stop. The bus might serve passengers at the bus stop hence stop for a while with more dense GPS points at the stop position nearby.

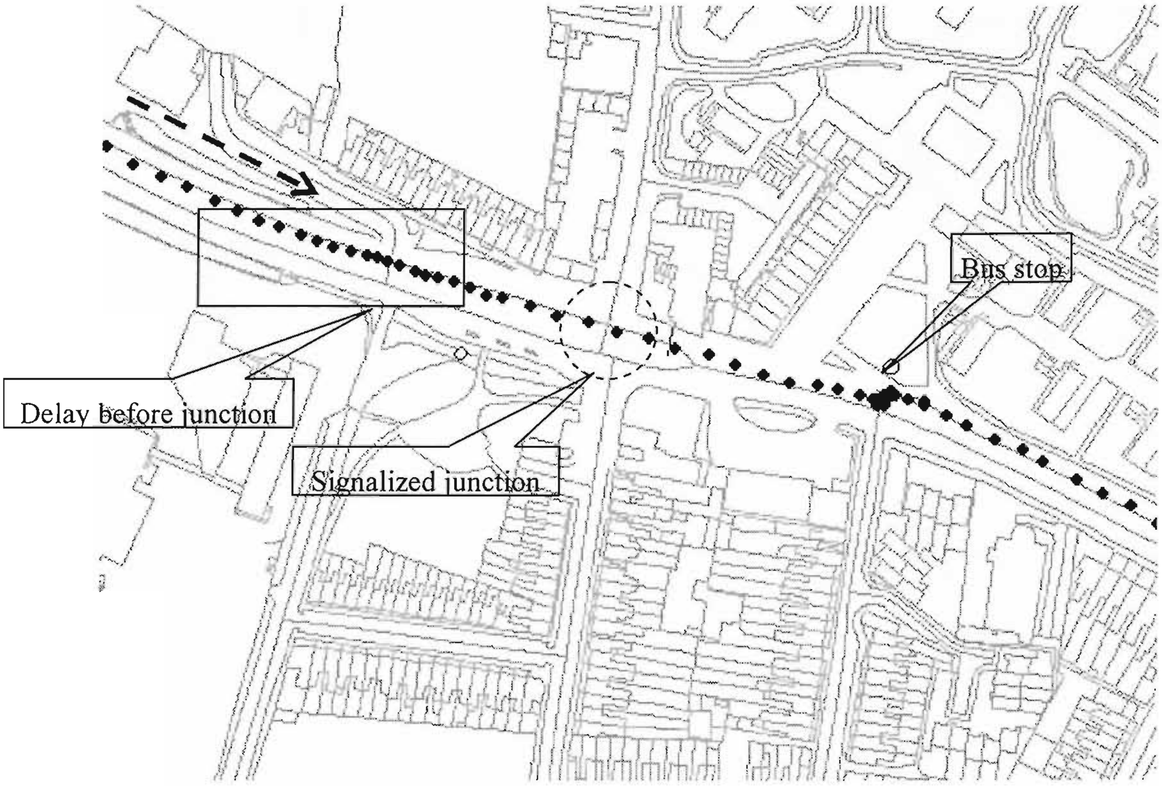


Figure A.1: GPS successive points of a bus on digital map (First 5/5A inbound service on Northam Road, Southampton on 12 January 2005)

This section is divided into three parts that deals with transforming GPS data, processing digital map for GIS, and importing GPS data into GIS.

A.1 Transforming GPS data

Many coordinates systems can be used to define a position anywhere on the earth. They are latitude, longitude, and ellipsoid height, Cartesian coordinates, easting and northing and so on. Whatever the coordinate is used, it is required to define the spatial relationship of the coordinate system to the Earth, this concept is generally called datum (Ordnance Survey [OS], 2000). It is possible to use approximate models to convert coordinates from one coordinate system to another, but not exactly formulate applying in perfect geometry because of the real situation on the ground. If the need of accuracy is about 5 to 10 meters, the transforming approach is simple and easy. Otherwise, more processes are required to convert into the different coordinate system (OS, 2000).

The datum used for GPS data is WGS84 (World Geodetic System 1984) or ETRS89 (European Terrestrial Reference System 1989), the differences between these two are negligible here. It uses a three dimensional coordinate system. Therefore, WGS84 positions can be showed as either X Y Z Cartesian coordinates or latitude, longitude, and ellipsoid height coordinates. The digital map used in this study is OS map. The coordinates of all features illustrated on this map are determined concerning to a TRF (Terrestrial Reference Frame) called National Grids or OSGB36 (Ordnance Survey Great Britain 1936) (OS, 2000). Therefore, it is essential to transform WGS84 (or ETRS89) coordinate system to OSGB36 in order to integrated GPS data into digital map.

Different transformations are required in different parts of the country because of the distortions of OSGB36 TRF. OS has developed a transformation that can deals with Easting and Northing coordinates shifts between WGS84 and OSGB36 cover Britain with transformation accuracy in horizontal of 0.1m and in vertical of 0.02m (OS, 2000). This is called National Grid Transformation OSTN02. It is available using software such as Grid InQuest (available at OS website: http://www.gps.gov.uk/additionalinfo/questDeveloper.asp).

The interface protocol design of GPS is based on NMEA0183 (the National Marine Electronics Association). This interface standard defines electrical signal requirements, data transmission protocol and time, and specific sentence formats for a 4800-baud serial data bus (NMEA, 2003). The data stored in the log including latitude, longitude, altitude, time, date, etc. is extracted from the $GPGGA and $GPRMC sentences of NMEA. An example of raw data of one GPS record is shown as follows.

A,	**5056.0260,**	**N,**	**123.6290,**	**W,**	**153809,**	**70904,**	**50**
Status, A=Valid position; V=receiver warning,	Latitude (ddmm.mmmm),	North or South,	Longitude (ddmm.mmmm),	West or East,	Time (UTC) hhmmss	Date ddmmyy	Altitude-meters

The transformation of WGS84 GPS data into OSGB36 format is comprised of two processes. The first converts GPS logged data to appropriate format and the next is coordinates transformation. The following are the detailed processes.

1. Convert logged GPS data into appropriate format

Use MS Excel to import the logged GPS data. Then, change the format of records to fit the requirements of converter. Here the software package of Grid InQuest is used for converting. Latitude and longitude data of GPS are presented as ddmm.mmmm as described above. However, they require converting into decimal degrees or degrees, decimal minutes or degrees, minutes, seconds as Equation A.1, A.2, and A.3.

- Decimal degree:

$$dd.dddd = \left\{\frac{ddmm.mmmm}{100}\right\} + \frac{ddmm.mmmm - \left\{\frac{ddmm.mmmm}{100}\right\} * 100}{60} \qquad (A.1)$$

- Degrees, decimal minutes:

$$dd, mm.mmmm = \left\{\frac{ddmm.mmmm}{100}\right\}, ddmm.mmmm - \left\{\frac{ddmm.mmmm}{100}\right\} * 100 \qquad (A.2)$$

- Degrees, minutes, and seconds:

$$dd, mm, ss.ssss = \left\{\frac{ddmm.mmmm}{100}\right\}, \left\{ddmm.mmmm - \left\{\frac{ddmm.mmmm}{100}\right\} * 100\right\} \qquad (A.3)$$
$$\cdots\cdots\cdots\cdots\cdots, (ddmm.mmmm - \{ddmm.mmmm\}) * 60$$

Note: the brackets { } denote an integer.

An example of format of a WGS84 coordinates record in Degrees, Decimal minutes is shown as:

Text, N, 52, 56.0260, W, 1, 23.6290, 250.1234

The fields are comma-delimited and arbitrary length. Thus, it does not matter whether integers are one or two characters, or how many decimal places should give in the real-valued fields. However all records in a file must be in the same format. Because of the converter accepts merely a plain text (ASCII) file in a simple format, it is suggested to save such file in the format of Comma Separated Value file (.CSV).

2. To transform the coordinates of WGS84 into OSGB36

Run GridInQuest and select File\Convert text file (as Figure A.2) to import data from step 1. Next, select the type of coordinate that are contained in the files: choose in ETRS89 and Latitude, longitude and ellipsoidal height. Then, define the field deliminator of the data with comma and convert into ETRS89 Easting and Nothing. After that, assign the individual column to latitude, longitude, height, point ID and the coordinates format which used in data. Finally, select the output coordinate system with OSGB36 (Eastings, Northings, orthometric height) and the output file, then press the Next to execute, the transformation file can be obtained after convertion.

Above steps are the details for converting GPS raw data from WGS84 coordinates into OSGB36. After this process, a digital map is required before importing such data in GIS.

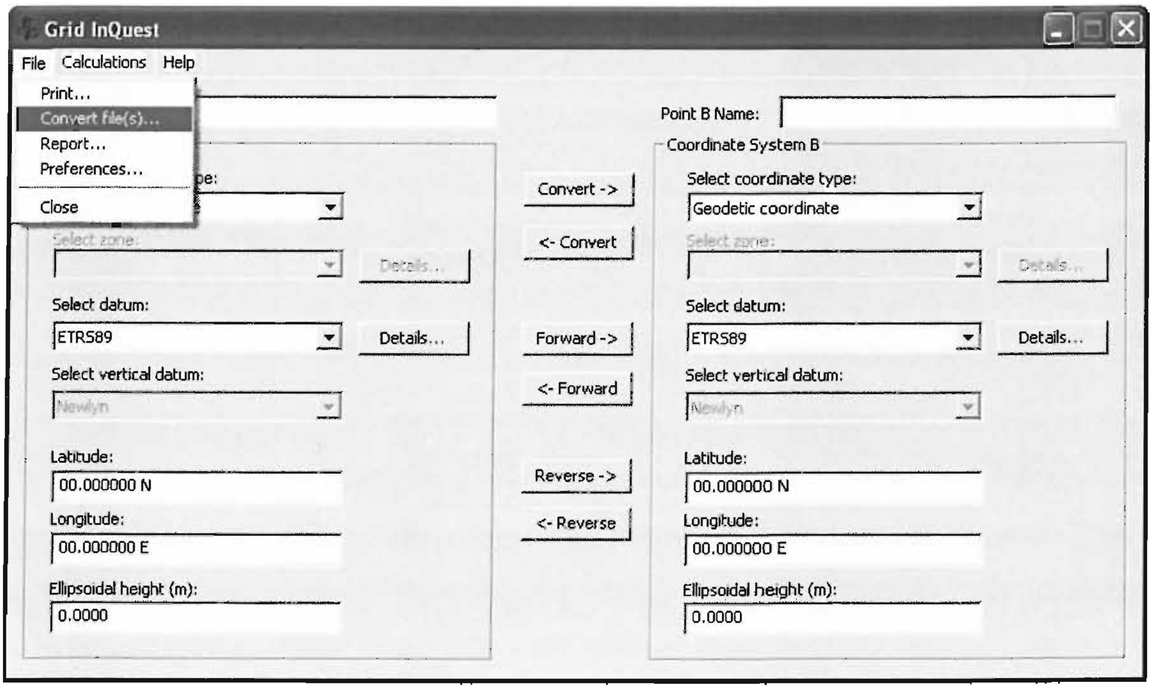


Figure A.2: To transform the coordinates of WGS84 into OSGB36 using GridInQuest

A.2 Digital map for GIS

Generally, there are three ways to obtain digital map for use. Firstly, the method is to purchase from commercial map. This is the easiest method but may cost a lot. Many digital-map products are available for the particular need of application such as street map for vehicle navigation. Secondly, transforming an original paper map into a digital map may work. However, such approach is the most difficult and time consuming work including considerable processes such as scanning, transforming, or using digital panel and combining different format maps. Lastly, the method is to download digital map from accessible website. This approach may require subscribing to the available websites and needing some processing. However, such approach has the superiority of low cost and less processing techniques than the second method. Therefore, the following description focuses on this approach.

It is essential to understand the requirements of map scale and the data format before downloading the digital map from available websites. For the map scale, it should consider many factors such as GPS data frequency, vehicle speed, the available scale of digital map, and requirement of road geometry and facilities for study. The vehicle running distance depending upon different vehicle speed and updating frequency of GPS is shown in Table A.1. It can be seen that the first value of 0.56 m means under vehicle running speed 5 mph and 0.25 s (4 Hz) updating frequency of GPS, the distance between two successive GPS points is 0.56 m. Generally, the average running speed of bus is around 10-20 mph in urban area. If the updating frequency is every 1 s, the distance between two GPS points is about 4~8 meters. It means the resolution of digital map may require up to 10 meters as well to associate with the similar resolution of GPS points. When check the real distance for 4~8 meters on various map scales in Table A.2 , it can be concluded that a map scale which is not greater than 1:10,000 is required for bus journey time study.

There are four data formats available on the available website (http://edina.ac.uk/digimap/). They are National Transfer Format (NTF, British Standard 7567), Drawing Exchange Format (DXF), Tagged Interchange File Format (TIFF or TIF), and Lempel-Ziv-Welch (LZW) compressed. If a special application of graphic software such as AutoCAD is used, it may use the DXF format. Otherwise, for common GIS software (e.g. ArcView, MapInfo, Arc/Info), the NTF data format is required.

Table A.1: Vehicle running distance in different speed and updating rate (meter)

GPS frequency / Vehicle speed	0.25s	0.5s	1s	2s	5s	10s	30s	1min	5min
5 mph	0.56	1.11	2.22	4.44	11.11	22.22	66.67	133.33	666.67
10 mph	1.11	2.22	4.44	8.89	22.22	44.44	133.33	266.67	1333.33
20 mph	2.22	4.44	8.89	17.78	44.44	88.89	266.67	533.33	2666.67
30 mph	3.33	6.67	13.33	26.67	66.67	133.33	400.00	800.00	4000.00
40 mph	4.44	8.89	17.78	35.56	88.89	177.78	533.33	1066.67	5333.33
50 mph	5.56	11.11	22.22	44.44	111.11	222.22	666.67	1333.33	6666.67

Table A.2: Real distance on various map scales

Map scale	1:1,250	1:2,500	1:10,000	1:50,000	1:2,500,000
Actual length (m) in real when on map is 1 mm	1.25	2.5	10.00	50	2500

The steps for downloading an appropriate digital map for this study are described as followings:

1. The available website for downloading digital map: Edina Digimap

After login, select the data collection of Ordnance Survey maps and data. Next, select the data download service for advanced users, which is on the bottom of the page. Then, choose Digimap Download button for Ordnance Survey digital map data tiles for use in GIS or CAD (Figure A.3). At this stage, there are four methods to retrieve data on the page with a flow chart on the left. They are Ordnance Survey tile name, A Map of Britain, National Grid co-ordinates, and A Gazetteer. In the HELP, there are detailed introduction and instructions to guide for the each process. Here National Grid coordinates (Ordnance Survey [OS], 2000), which is also called Eastings and Northings coordinates, is used for downloading data.

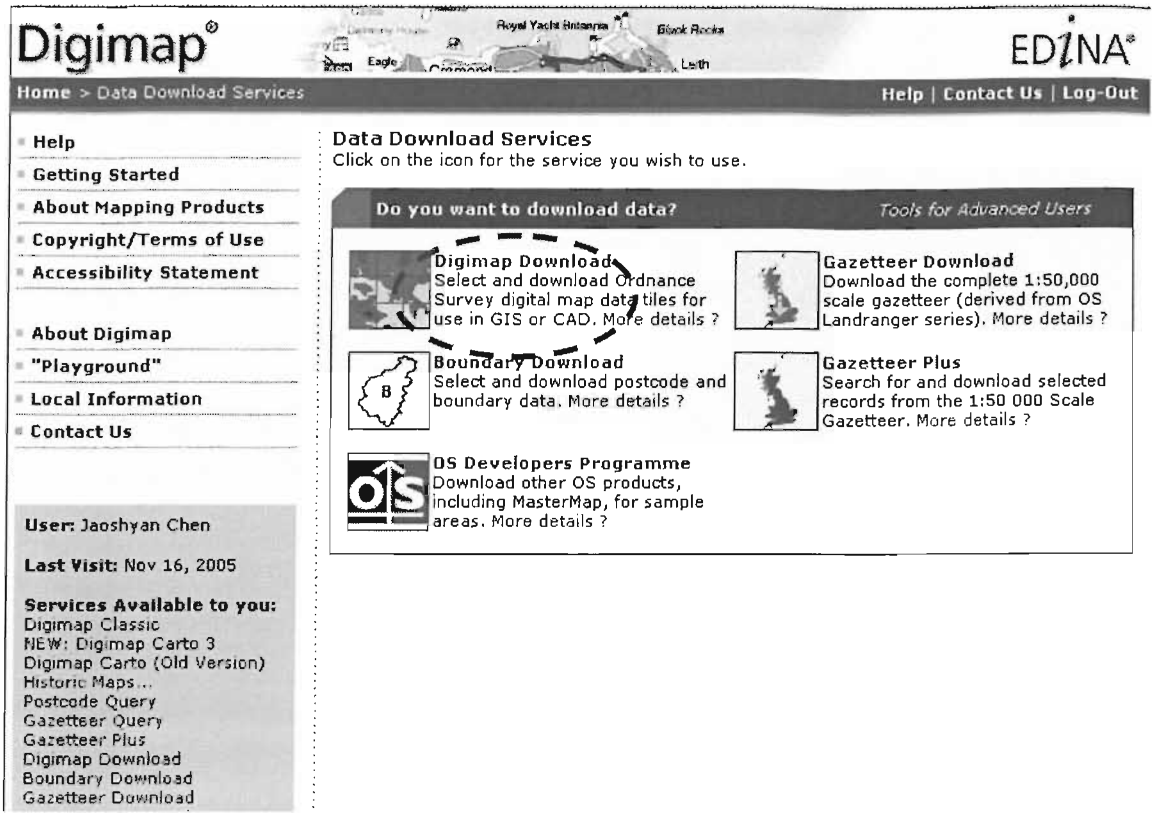


Figure A.3: Choose the Digimap Download from the available services

2. To obtain easting-northing coordinates

Before carry on doing next step, it is necessary to identify where the map for the proposed location is and how big it is. A free Streetmap website (http://www.streetmap.co.uk/) was used for this study to attain the location of the two corners (southwest and northeast) coordinates on the bottom message row. For example, the initial survey of this study is A3024 between city centre and Westend Road. The coordinate of the southwest corner around city centre is (442250, 112250) (Figure A.4) and the northeast corner around Westend Road is (444750, 113250).

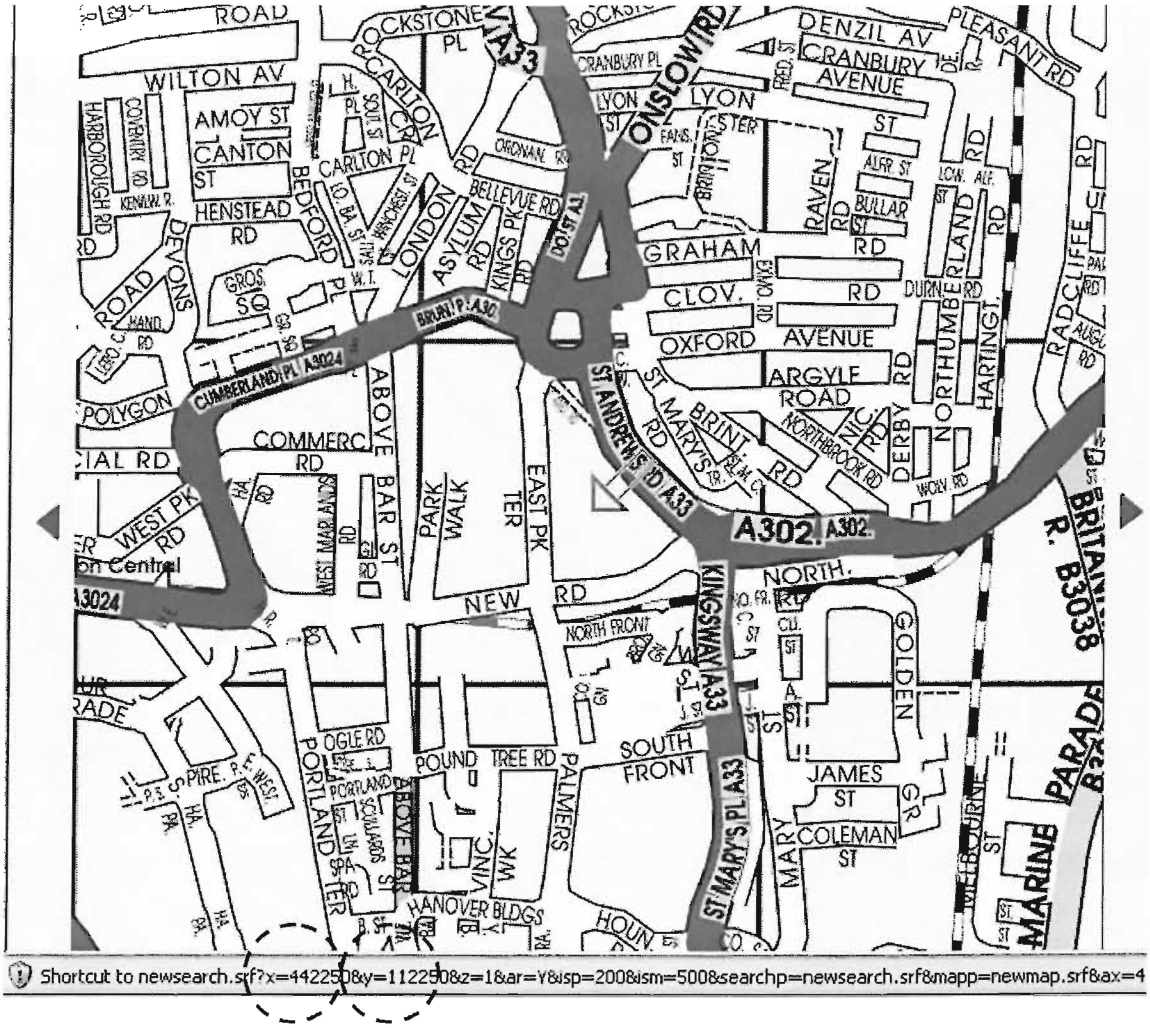


Figure A.4: Check the corner coordinates using Streetmap

3. To identify by National Grid coordinate

With the above the two corners of southwest and northeast coordinates, they can be entered into the 6 figures of bottom left and top right of the National Grid coordinates to specify the area of interest. Next, select the Landline-Plus for search. The accuracy of Land-line-Plus in urban is 0.4 m for 1:1,250 scales and in rural is 0.9-1.2 m for 1:2,500 scales (http://digimap.edina.ac.uk/ Help page). Then, press the View Tile List button. There are 18 tiles on the list with the tile name and the data size. Select all tiles and press Continue button, then choose the file format for downloading. Note that Landline-Plus is only available in NTF format for vector data. Thus, select such method for archiving and compression and choose zip archive for windows systems, and press Extract Data button. There are three files after the files extracted. Remember to save these files for later conversion.

4. To use ESRI MapManager 6.2 to convert NTF file into shapefile for ArcView

1) Unzip the saved file from above step (3).

2) Run MapManager 6.2; Press File\Select files and select unzipped files, and then press OK. There are three new windows opened, including Output Format, Data Schema, and Selected Files, which can change the output file directory, function settings, and check each file status (Figure A.5). Note that the output format of Shape File Format have to be selected. Then, press the Convert Files button to process conversion from each NTF format file to shape file individually.

3) To append shape files (OSANNO.SHP, OSLINE.SHP, and OSPOINTS.SHP) from the last output folder using the function of Appends Shape File in MapManager 6.2. After that, a folder is produced which is named Appended with a total of 9 files (.DBF, .SHP, and .SHX).

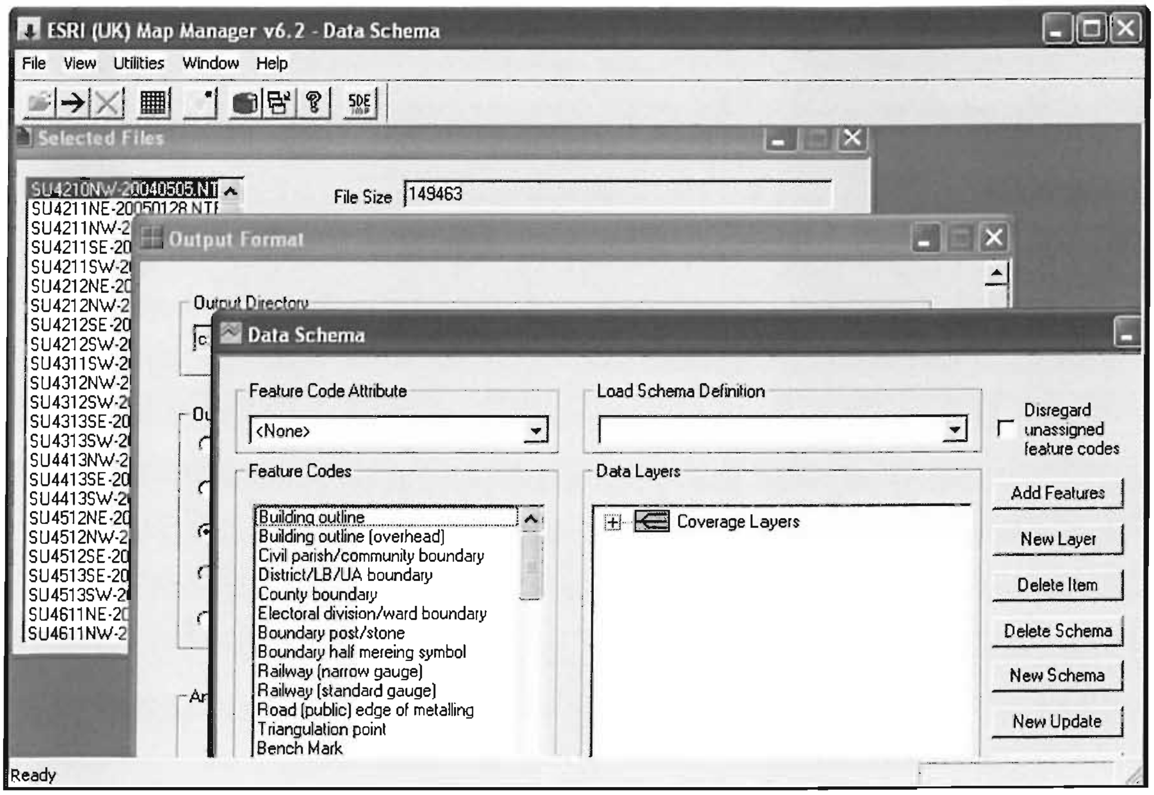


Figure A.5: To convert NTF file into shapefile using ESRI MapManager

5. To import converted shapefile into ArcView

1) Run ArcCatalog in the GIS software (here the ESRI ArcView 8.2 is used for GIS software). In ArcCatalog, identify the location of the Appended folder from above step (4) and use the Preview function. When the folder is opened, it can be seen that

there are three types of data, namely OSANNO, OSLINES, and OSPOINTS. The intended map can be seen from the Preview of OSLINES data (Figure A.6).

2) Run ArcMap in the ArcView, and select a new empty map. Go back to ArcCatalog window, click and hold the OSLINES catalog and pull it into ArcMap window. Then, a digital-map as Figure A.7 can be seen. Remember to save the map for later use.

Above section introduced the process to achieve a digital map from available website and import into GIS software such as ArcView. Next, integration of GPS and GIS is completed by importing GPS data from section A.1.

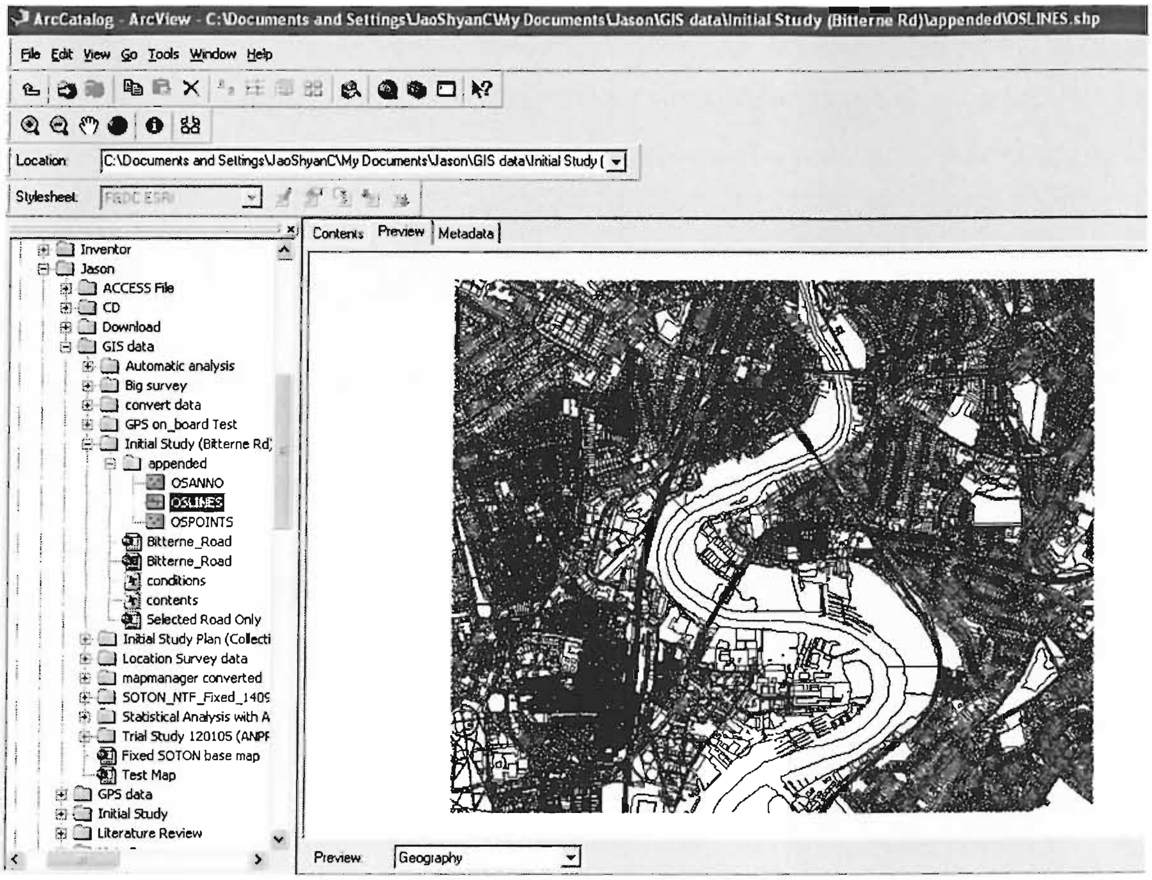


Figure A.6: To import converted shapefile into ArcView using ArcCatalog

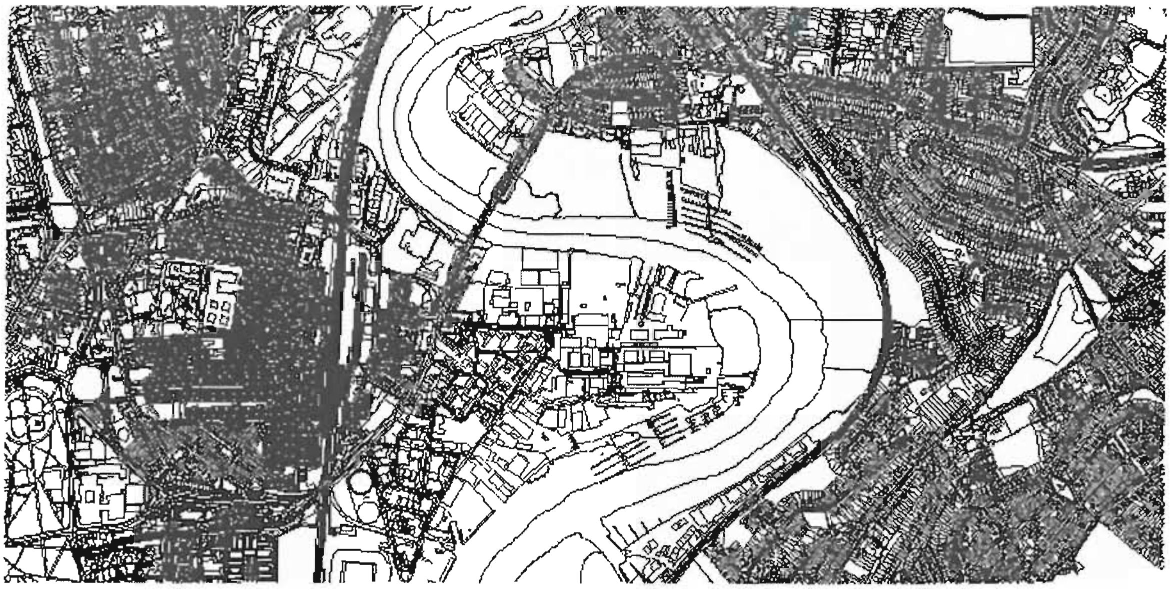


Figure A.7: Example of the initial study digital map.

A.3 Import GPS data into GIS

The GPS data file with National Grid coordinates which obtained from section A.1 is a text file (.CSV). It is required to convert such file into the dBase Format (.DBF) file before importing into ArcView of GIS. This can be achieved using MS Excel with Save As function. Followings are the steps for importing GPS data into ArcView:

1) Run ArcMap and open an existing map from section A.2.

2) Select Tools\Add X,Y data. From the new window of Add X,Y data, choose the table from converted .dbf file and specify for the X and Y coordinate with Eastings and Northings coordinates. Next, identify spatial reference with Edit and select OSGB 1936.prj from Europe in Geographic Coordinate Systems. It can be finished by pressing the OK button. Finally, the digital map with GPS data like Figure A.1 is achieved. Again, Remember save the file for later data analysis.

Appendix B

Bus On-board Survey

Contents

The on-board bus survey is carried out by noting the designed recording sheet manually and accompanied by logging position using portable GPS equipment automatically. All surveyors are requested to note down the departure and arrival time at every bus stop for serving patronages as well as the number of alighting and/or boarding passengers. The start and end point of survey route are also recorded as the boarding and alighting time for journey time calculations. In addition, some useful information about bus type and operation such as direction (inbound/outbound), bus service number, and bus types (whether the bus is low floor; whether the bus is double-decker, and how many doors did the bus use) are included.

An induction of survey is held for each surveyor before the main survey to make sure that every surveyor could understand the aim, safety, equipments used, approach, duration, and dates of survey. Following are the induction content, survey routes and durations, equipments, survey tips, recording sheet, and related documents of the bus on-board survey

B.1 Induction content of bus on-board survey

Bus Journey Time and Dwell Time Survey Induction

1. Bus journey time survey information (B.2)
2. Safety briefing for bus surveyors (B.3)
3. Survey equipments (B.4)
4. Survey tips (B.5) and recording sheet (B.6)
5. Where is your bus route (B.7)? Where are the bus stops for you to board and alight buses and the available Bus Services for Survey (B.8)? Timetables of available bus services.
6. Please come to collect your survey stuff (GPS receiver, data logger, battery, and recording sheets) on 11.00AM (13^{th}, 14^{th}, and 15^{th} of September, 2005) at Faraday Building R9003 and return them after survey each single day.

B.2 Bus journey time survey information

This sheet is used as an information leaflet to pass to anyone who asks.

BUS JOURNEY TIME SURVEY INFORMATION

WHEN ....?

- This survey is being carried out between 11:00 and 18:30 13th to 15th September 2005

WHERE ...?

- The Avenue (A33) between City Centre and Chilworth Roundabout
- Bitterne Road (A3024) between City Centre and Botley Road via Northam Bridge
- Portsmouth Road (A3025) between City Centre and Pound Road via Itchen Bridge
- Portswood Road between City Centre and Swaythling

WHY ....?

- The aim of this survey is to better understand the journey time and dwell time of buses of these routes, to help provide travel time information in the future.

WHAT ....?

- The equipment that you see is GPS receiver to receive the signal from satellites and calculate the position of this receiver.
- The recording sheet is used to note the number of passengers alighting and boarding the bus and the duration concerning to the bus dwell time.
- The recordings do not relate to schedules or any measure of bus performance and cannot and will not be used for that purpose.
- All recorded information is restricted to researchers and will not be held beyond the end of the study.

WHO ....?

- The survey is being carried out by the Transportation Research Group of the University of Southampton.

For more information, please contact....

- Professor Mike McDonald, (023) 8059 2192, or email mm7@soton.ac.uk

B.3 Safety briefing for bus surveyors

SAFETY BRIEFING FOR BUS SURVEYORS

1. Your safety and the safety of other people is the greatest importance.
2. You should have a seat after boarding a bus and keep yourself stable before doing the survey.
3. Do not cross the road unless essential. Please use existing pedestrian crossing facilities whenever possible.
4. Data logger and battery should be kept dry at all times.
5. You should have information sheets to pass to anyone who asks.
6. Try not to distract other passengers.
7. Bus companies and University of Southampton have been notified and agreed to the survey taking place.

Any incidents or potential incidents should be brought to the attention of Jason as soon as possible.

Jason : 0797 404 7986

B.4 Survey equipments

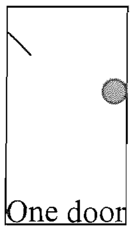
- GPS receiver: Garmin GPS 35-PC, 
- GPS data logger: DGPS-XM4-ALT , 
- Battery: Yuasa rechargeable lead-acid battery, 12V, 2.1Ah

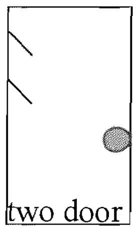
B.5 Survey tips

1. Before departure, you have to make sure that you bring all followings: GPS receiver, data logger, battery, recording sheets, watch, pen, and your money (£5.5) for bus ticket (Solent day travel ticket).

2. Before the survey, you have to make sure all the equipments connecting and fixing properly and let GPS receiver enable to receive enough satellites' signal until the LED indicator flashing twice at a time. It may take more than 5 minutes to do so depending upon your environment. Therefore, please keep your GPS receiver face to the clear sky at the initial stage. When you have a twice flash on the LED, it means data has been received and saved and you are ready to conduct the survey.

3. To take on the selected bus and try to find a seat close to big clear window where you can clearly see the alighting and boarding passengers and might put the GPS receiver on the frame (may use with temporary tape to stick). The suggested seat is at right side as followings, but they depend on the real situation on bus. In case of many standees on bus, the left side should be select to prevent blocking the view from standees. If there is no seat close to the window available, please ask someone to move as you can as possible. Otherwise, this run will be useless.





4. When you are on board, please record the "Arrival Time" when the bus comes to a complete stop; count and record the number of passengers alighting and the number of passengers boarding; record the time "Departure time" when the bus starts to move; Note any special circumstances especially for taking extra time, for example, wheelchair and pushchair movement, bus waiting at bus-stop for adherence the schedule .

5. To take off the bus at specific stop; check the recording sheet is complete with all the required data; check the LED still flash twice a time.

6. Go on doing procedure 3, 4, 5, until the end of survey. The survey period must have 5 hours at least.

7. If you need to go to toilet during the survey, please find somewhere in city centre. When you are away the survey, please pull off the power connection, and when you start again, please follow the procedure from 2.
8. Each route has two surveyors. Surveyors do not take on the same bus at the same time (first come first go).
9. When you finish the survey, please return all equipments and sheets to Room 9003, Faraday Building.

Note. Please hold the GPS receiver, keep it face to the clear sky all the time, and check the LED condition (flashing twice at a time) after each run.

Contact number: 023 8059 5386 (Uni 25386), 0797 404 7986 Jason

B.6 Recording sheet

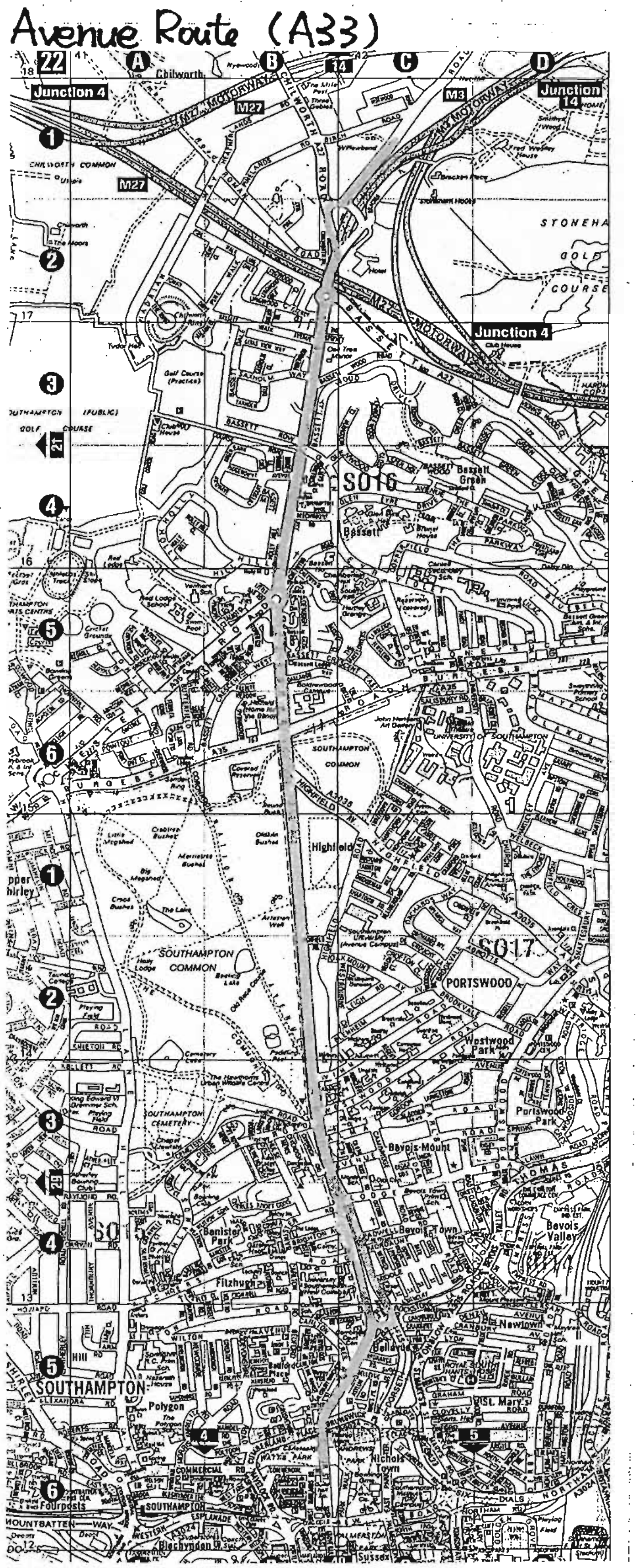
Dwell Time and Passenger Data Sheet

Sheet No:

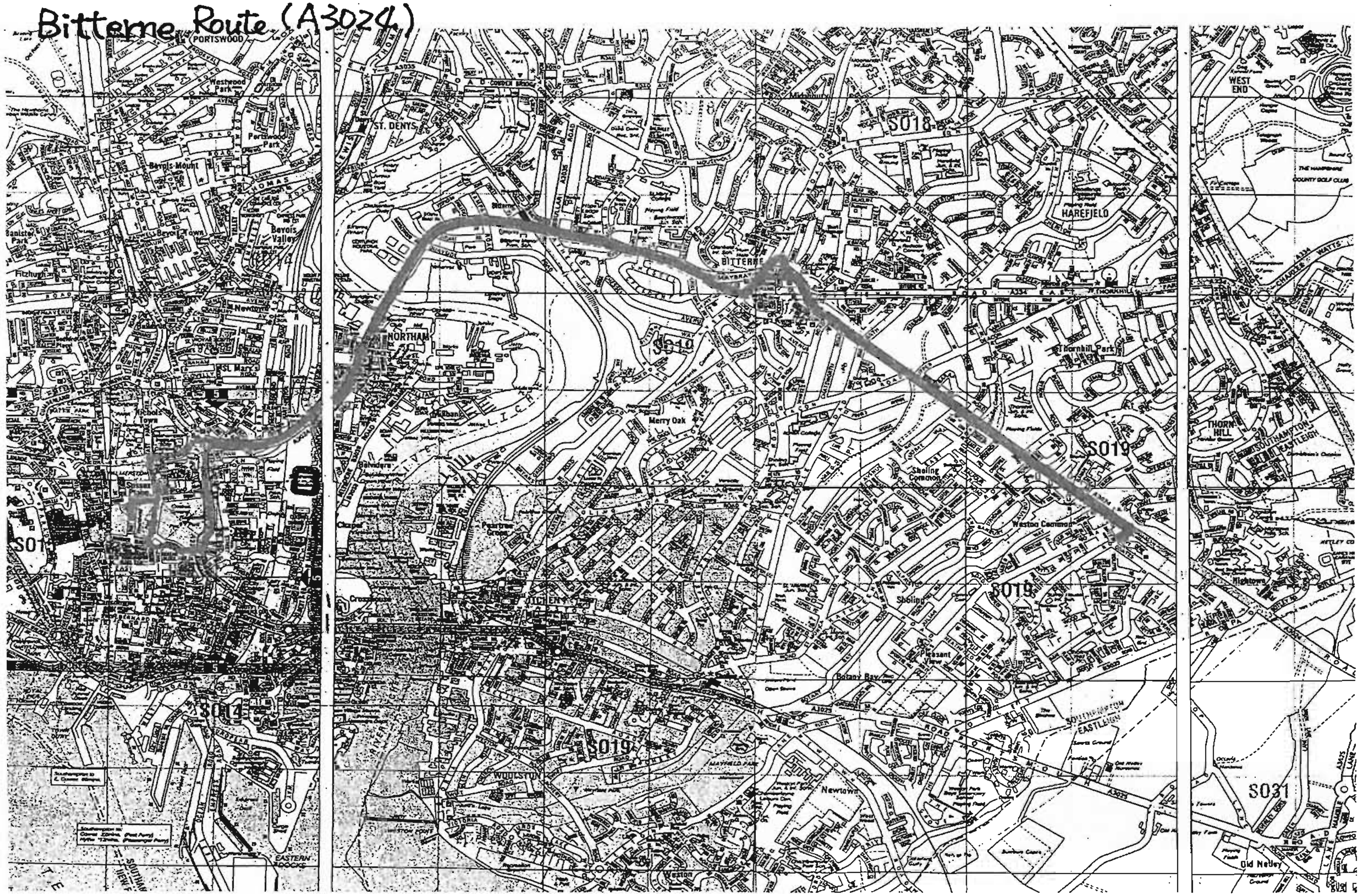
- **Direction**:____(**I**: to City Centre; **O**: away from City Centre)
- **Bus Number**:______.
- **Bus Type**: Is bus low floor? Yes/No; Is bus double-decker? Yes/No; How many doors: ____.
- **Boarding Time**:_______(hh:mm:ss); **Alighting time**: _________ .

Stop order	Arrival time (mm:ss)	Passengers Alighting	Passengers Boarding	Departure time (mm:ss)	Note
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

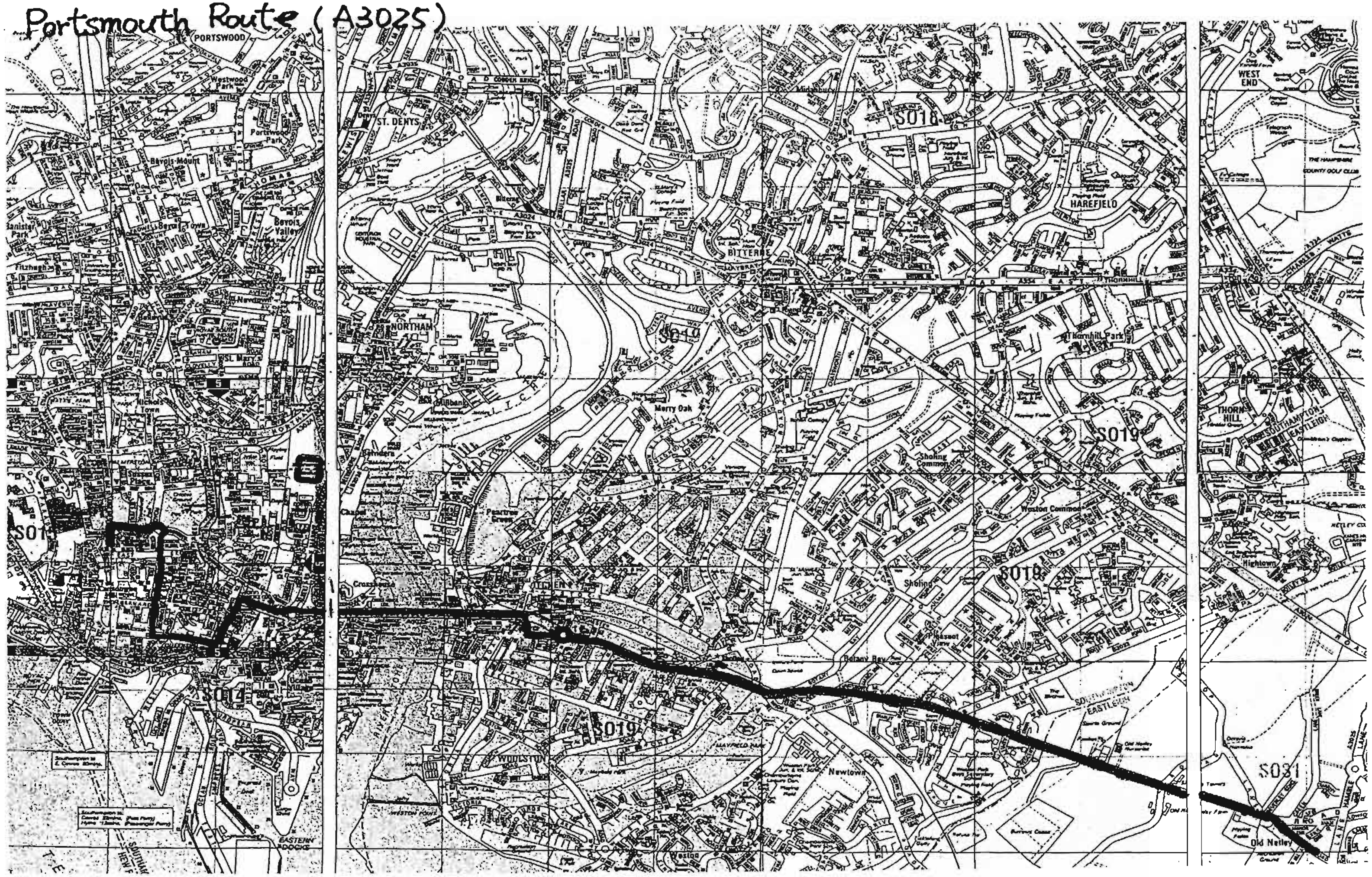
B.7.1 Survey route map- Avenue route (A33)



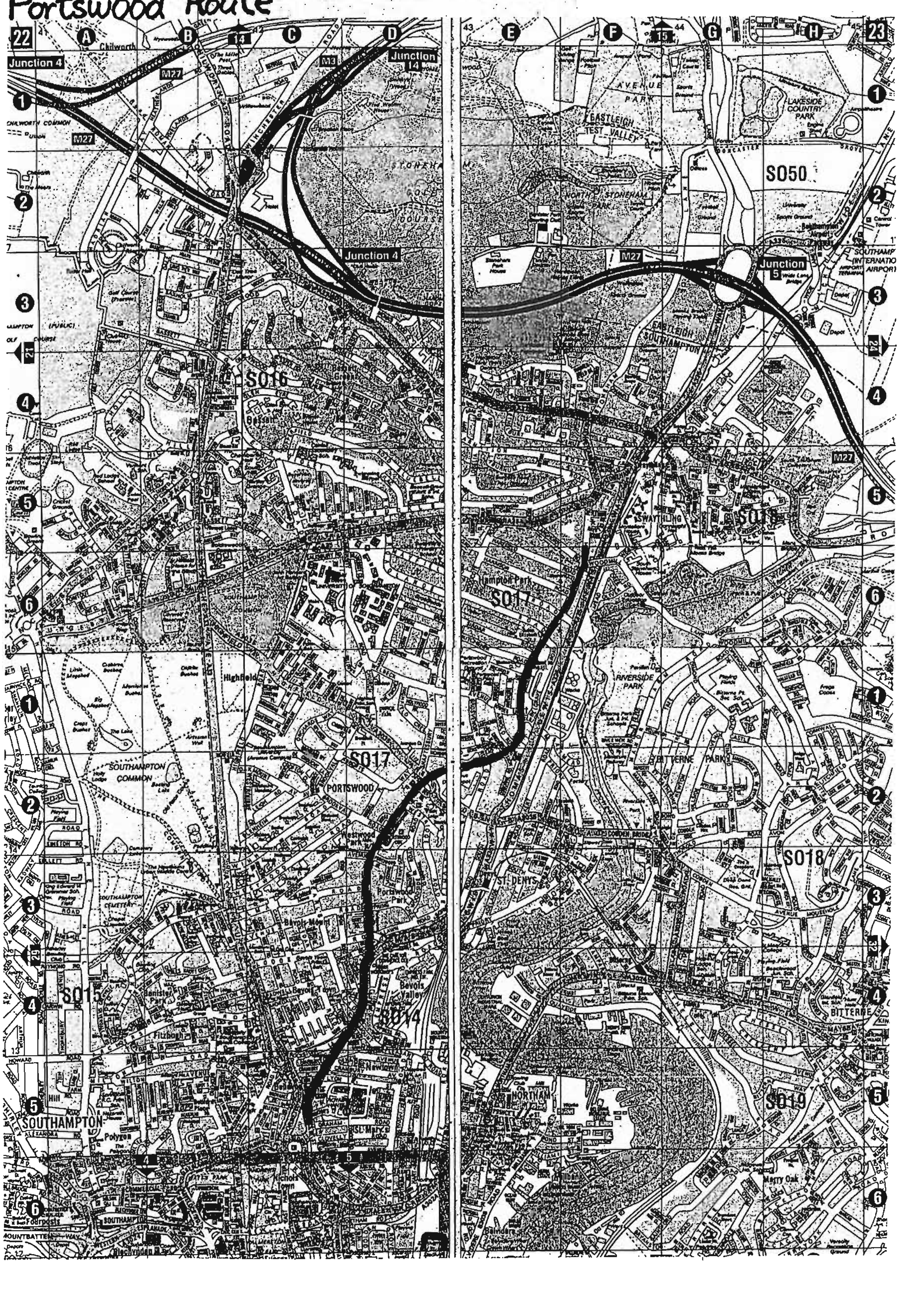
B.7.2 Survey route map- Bitterne route (A3024)



B.7.3 Survey route map- Portsmouth route (A3025)



B.7.4 Survey route map- Portswood route



B.8 Available bus service for service by route

Available Bus Service for Survey

Survey Route	Outbound	Outbound get off stop	Inbound	Inbound get off stop	Bus frequency (peak time)
Avenue (A33) (between City Centre and Chilworth Roundabout)	Solentblueline: BlueStar 1, 44/44A	After Chilworth Roundabout	Solentblueline: BlueStar 1, 44/44A	Watts Park (Above Bar Street)	BlueStar 1: 4 services per hour 44/44A: 1 service per hour
Bitterne (A3024) (between City Centre and Botley Road via Northam Bridge)	First: 5A, 78 Solentblueline: Solent shuttle (Southampton-Portsmouth express)	(1) Orpen Road (2) After Hamble Motor Peugeot	(1) First: 5/5A (2) First: 78 (Solent shuttle?)	City Center Pound Tree Road	5A: 3 services per hour 78: 1 service per hour Solent Shuttle:1 service per hour
Portsmouth (A3025) (between City Centre and Pound Road via Itchen Bridge)	First: 72, 80	Pound Road (Old Netley)	First: 72, 80	City Centre In front of Bargate	72: 1 service per hour 80: 1 service per hour
Portswood (between City Centre and Swaythling)	First:11A Unilink: U6	After B&Q before Woodmill Road	First:11A Unilink: U6	St. Mary Road (R.H.S Hospital, just before the Round about of Charlotte Place)	11A: 3 service per hour U6: 2 services per hour

B.9 An example of survey letter for surveyor for someone who may concern

TRG
Transportation
Research Group

School of Civil Engineering and the Environment
Professor Mike McDonald, Director TRG
University of Southampton
Highfield
Southampton
SO17 1BJ United Kingdom
Tel +44 (0)23 8059 2192
Fax +44 (0)23 8059 3152
Email mm7@soton.ac.uk
Web www.trg.soton.ac.uk

7th September 2005

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

Jaoshyan Chen

The above named individual is conducting a survey of the accuracy of a range of GPS systems to estimate vehicle movements. A simple commercial device is being used which receives and records signals from satellites.

If you have any queries please contact Miss Melanie Hallford on 8059 2192.

Yours faithfully,

M Hallford

p.p. Mike McDonald

The Transportation Research Group is based within the School of Civil Engineering and the Environment, University of Southampton
The Transportation Research Group is a member of Rail Research UK

B.10 An example of survey letter to inform a bus company

TRG
Transportation
Research Group

School of Civil Engineering and the Environment
Professor Mike McDonald, Director TRG

University of Southampton
Highfield
Southampton
SO17 1BJ United Kingdom

Tel +44 (0)23 8059 2192
Fax +44 (0)23 8059 3152
Email mm7@soton.ac.uk
Web www.trg.soton.ac.uk

7th September 2005

Uni-Link,
The Travel Centre,
Building 57,
University of Southampton,
Highfield,
Southampton.

Dear Sir/Madam,

One of our PhD students, Jayshyan Chen, is studying the potential use of GPS data for journey time estimation in an urban network. He is working closely with staff in the ROMANSE traffic control centre and my colleague, Dr. Nick Hounsell. This letter is to inform you of surveys which will be taking place.

Over the next week, he will be undertaking a survey which will include location measurements on selected bus routes in Southampton. The surveyors will pay fares as usual, will record some data in an unobtrusive way and should not be noticed by drivers or passengers. However, they may make longer or more complicated journeys than is normal. The data will be used to compare with equivalent journey times by cars on the same road sections to try to determine statistical relationships. The recordings do not relate to schedules or any measure of bus performance and cannot and will not be used for that purpose. The survey staff will have a letter of introduction to be used if necessary.

If you have any queries, please contact Jayshyan, Nick or myself at the University or Ray Morris at the ROMANSE office, who are providing the parallel traffic data.

Many thanks,

Yours faithfully,

M Hallford

for Mike McDonald

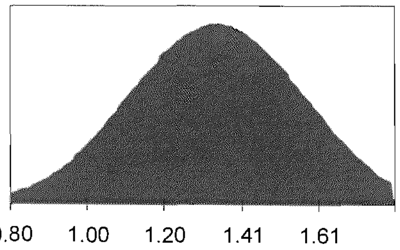
The Transportation Research Group is based within the School of Civil Engineering and the Environment, University of Southampton
The Transportation Research Group is a member of Rail Research UK

Appendix C

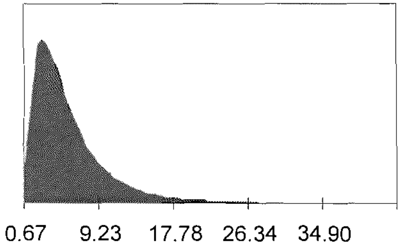
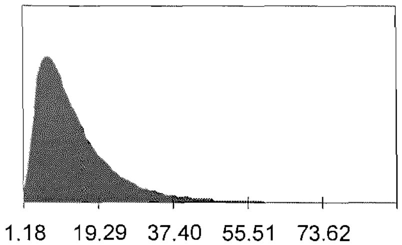
Input parameters of Monte Carlo models

Content

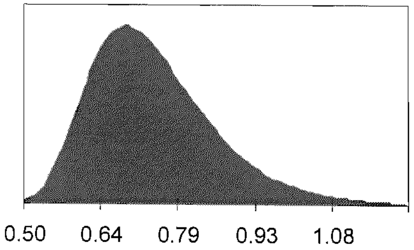
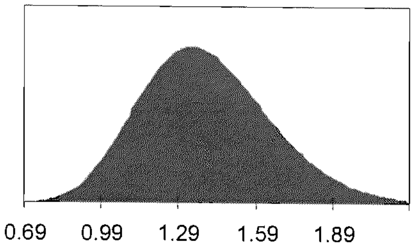
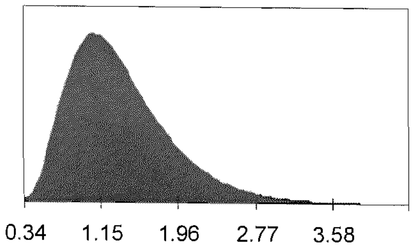
C.1: Description of input chance variables of $ANPR_{\%}$ in general travel time (JT) module

Input parameter	Fitted distribution	Distribution parameters	Statistics	Note
$ANPR_{\%}$	Normal distribution  	Mean 1.34 S.D. 0.23	Test: A-D 0.22 Chi-square 2.24 K-S 0.09 Descriptive statistics: Mean 1.34 Minimum 0.81 Maximum 1.78 Obs. 58	

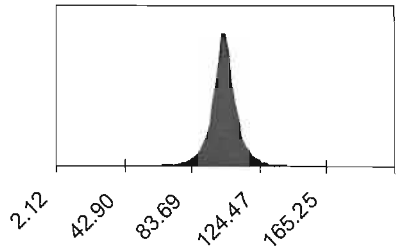
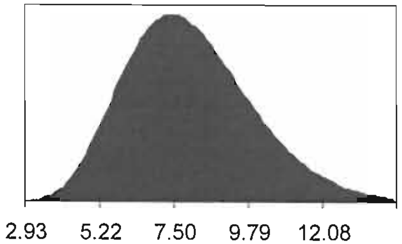
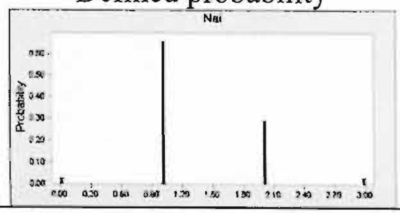
C.2: Description of input chance variables in dwell time (DT) module

Input parameter	Fitted distribution	Distribution parameters	Statistics	Note
T_{ai}	Lognormal distribution  	Mean 6.07 S.D. 4.98	Test: A-D 3.64 Chi-square 273.59 K-S 0.09 Descriptive statistics: Mean 6.34 Minimum 0.67 Maximum 55 Obs. 480	
T_{bi}	Lognormal distribution  	Mean 13.99 S.D. 10.73	Test: A-D 1.56 Chi-square 30.21 K-S 0.07 Descriptive statistics: Mean 14.14 Minimum 1 Maximum 94 Obs. 383	

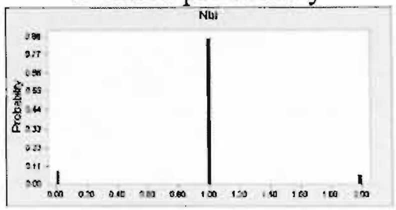
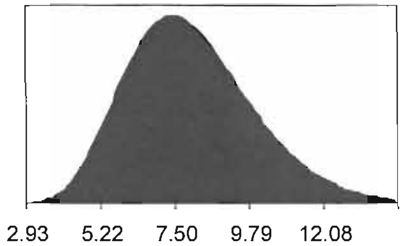
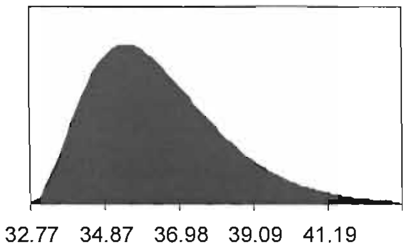
C.3: Description of input chance variables in acceleration/deceleration delay (T_{AD}) module

Input parameter	Fitted distribution	Distribution parameters	Statistics	Note
ST_s%	Maximum extreme distribution 	Likeliest 0.69 Scale 0.10	Test: A-D 0.29 Chi-square 4.36 K-S 0.04 Descriptive statistics: Mean 0.74 Minimum 0.53 Maximum 1.06 Obs. 107	
A	Gamma distribution 	Location 0.29 Scale 0.06 Shape 18.5321	Test: A-D 0.75 Chi-square 31.04 K-S 0.04 Descriptive statistics: Mean 0.88 Minimum 0.10 Maximum 1.99 Obs. 387	
D	Lognormaldistribution 	Mean 1.37 S.D. 0.57	Test: A-D 1.48 Chi-square 52.13 K-S 0.06 Descriptive statistics: Mean 1.35 Minimum 0.31 Maximum 2.99 Obs. 464	

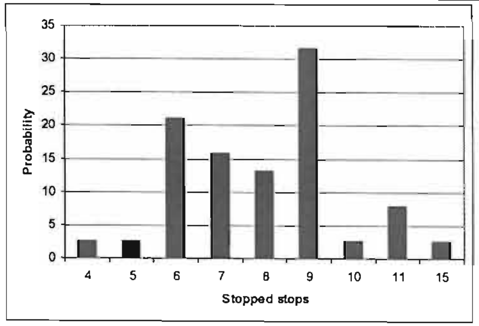
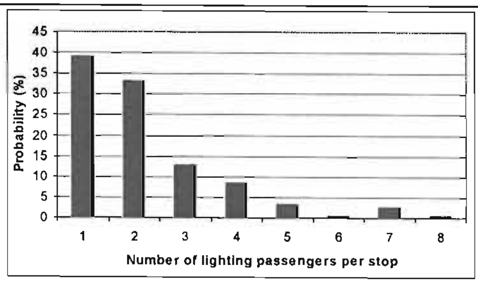
C.4: Description of variables for Bitterne inbound route validation

Module	Input parameter	Fitted distribution	Distribution parameters	Statistics	Note
JT	T_{ANPR}	Student's t distribution 	Midpoint 103.06 Scale 5.72 Deg. Freedom 2.21450	Test: A-D 0.36 Chi-square 9.09 K-S 0.06 Descriptive statistics: Mean 103.39 Minimum 81.05 Maximum 131.95 Obs. 95	5 minutes interval of ANPR journey time from 11:00 to 19:00, averaged from 3 days
DT	$N_{stopped}$	Gamma distribution 	Location 1.20 Scale 0.58 Shape 11.70785	Test: A-D 0.96 Chi-square 23.05 K-S 0.16 Descriptive statistics: Mean 8 Minimum 4 Maximum 15 Obs. 38	The same in T_{AD} module.
	N_{ai}	Defined probability 	Value / Probability 0.00 / 2.63 1.00 / 65.79 2.00 / 28.95 3.00 / 2.63		

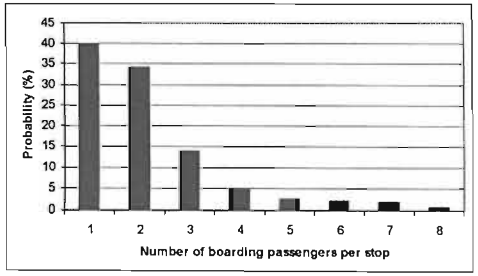
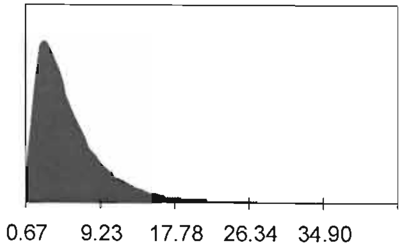
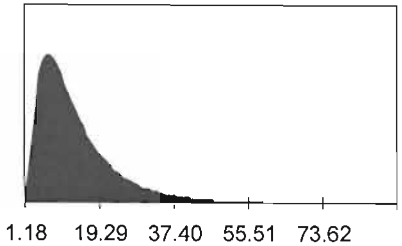
C.4: Description of variables for Bitterne inbound route validation (continued)

Module	Input parameter	Fitted distribution	Distribution parameters	Statistics	Note
	N_{bi}	Defined probability 	Value / Probability 0.00 / 7.89 1.00 / 86.84 2.00 / 5.26		
T_{AD}	$N_{stopped}$	Gamma distribution 	Location 1.20 Scale 0.58 Shape 11.70785	Test: A-D 0.96 Chi-square 23.05 K-S 0.16 Descriptive statistics: Mean 8 Minimum 4 Maximum 15 Obs. 38	The same in *DT* module.
	ST_s	Gamma distribution 	Location 32.53 Scale 0.96 Shape 4.09767	Test: A-D 1.17 Chi-square 13.08 K-S 0.09 Descriptive statistics: Mean 36.45 Minimum 33.2 Maximum 40.63 Obs. 96	5 minutes interval of SCOOT UO7 message from 11:00 to 19:00, averaged from 3 days

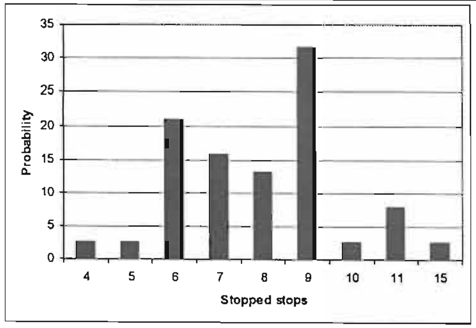
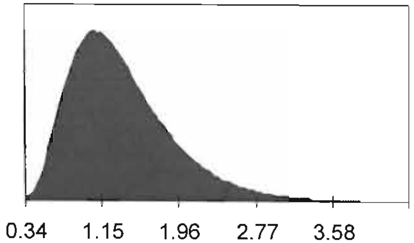
C.5: Input parameters of each module of Bitterne inbound case link-based model

Module	Input parameter	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
JT_g	L_{link}	Fixed values			Each length of link was shown in Table 7.12
	ST_s	Fitted distribution from historical data of each link			Shown in Appendix C.4
DT	$N_{stopped}$		Value / Probability 4.00 / 2.63 5.00 / 2.63 6.00 / 21.05 7.00 / 15.79 8.00 / 13.18 9.00 / 31.58 10.00 / 2.63 11.00 / 7.89 15.00 / 2.63	Descriptive statistics: Mean 8 Minimum 4 Maximum 15 Obs. 38	The same as D_{acc} module
	N_{ai}		Value / Probability 0 / 39.15 1 / 33.10 2 / 12.81 3 / 8.54 4 / 3.20 5 / 0.36 6 / 2.49 7 / 0.36	Descriptive statistics: Mean 1 Minimum 0 Maximum 7 Obs. 281	

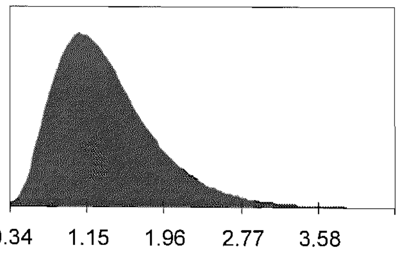
C.5: Input parameters of each module of Bitterne inbound case link-based model (continued)

Module	Input parameter	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
DT	N_{bi}		Value Probability 0 39.86 1 34.16 2 13.88 3 4.98 4 2.49 5 2.14 6 1.78 7 0.71	Descriptive statistics: Mean 1 Minimum 0 Maximum 7 Obs. 281	
	T_{ai}	Lognormal distribution 	Mean 6.07 S.D. 4.98	Test: A-D 3.64 Chi-square 273.59 K-S 0.09 Descriptive statistics: Mean 6.34 Minimum 0.67 Maximum 55 Obs. 480	
	T_{bi}	Lognormal distribution 	Mean 13.99 S.D. 10.73	Test: A-D 1.56 Chi-square 30.21 K-S 0.07 Descriptive statistics: Mean 14.14 Minimum 1 Maximum 94 Obs. 383	

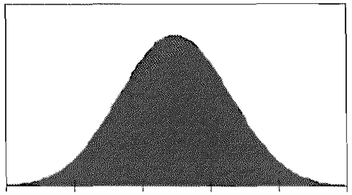
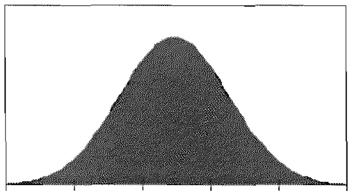
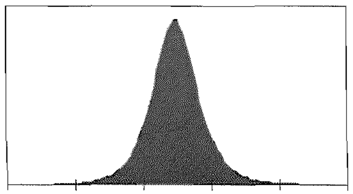
C.5: Input parameters of each module of Bitterne inbound case link-based model (continued)

<table>
<tr><th>Module</th><th>Input parameter</th><th>Value/ Function/ Fitted distribution</th><th>Distribution parameters</th><th>Statistics</th><th>Note</th></tr>
<tr><td rowspan="2">Sig</td><td>$P(\mathrm{Re}\,d)_i$</td><td rowspan="2">Fitted distribution from link average signal timing</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2">Shown in Appendix C.6</td></tr>
<tr><td>$T_{\mathrm{Re}d-i}$</td></tr>
<tr><td rowspan="3">D_{acc}</td><td>$N_{stopped}$</td><td>
</td><td>Value / Probability
4.00 / 2.63
5.00 / 2.63
6.00 / 21.05
7.00 / 15.79
8.00 / 13.18
9.00 / 31.58
10.00 / 2.63
11.00 / 7.89
15.00 / 2.63</td><td>Descriptive statistics:
Mean 8
Minimum 4
Maximum 15
Obs. 38</td><td>The same as DT module</td></tr>
<tr><td>BV</td><td>$BV = \frac{\sum_{i=1}^{n} V_{bus-i}}{n}$ (Equation 6.13),
$V_{bus-i} = 2.62 \times ST_s^{0.65}$ (Equation 6.24)</td><td></td><td></td><td>L_{link} from Table 7.12, ST_s is shown in Appendix C.7</td></tr>
<tr><td>D</td><td>Lognormal distribution

</td><td>Mean 1.37
S.D. 0.57</td><td>Test:
A-D 1.48
Chi-square 52.13
K-S 0.06
Descriptive statistics:
Mean 1.35
Minimum 0.31
Maximum 2.99
Obs. 464</td><td></td></tr>
</table>

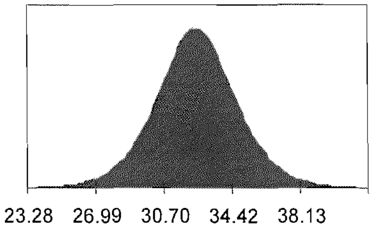
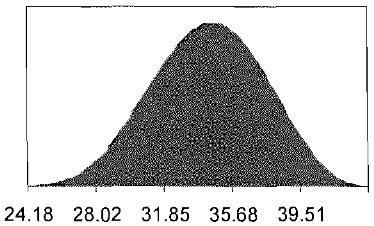
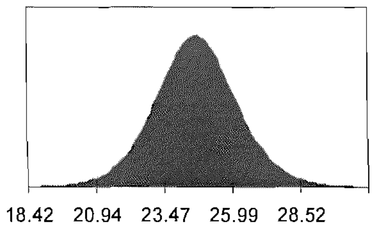
C.5: Input parameters of each module of Bitterne inbound case link-based model (continued)

Module	Input parameter	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
	A	Lognormal distribution  	Mean 1.37 S.D. 0.57	Test: A-D 1.48 Chi-square 52.13 K-S 0.06 Descriptive statistics: Mean 1.35 Minimum 0.31 Maximum 2.99 Obs. 464	

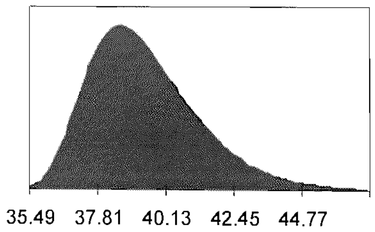
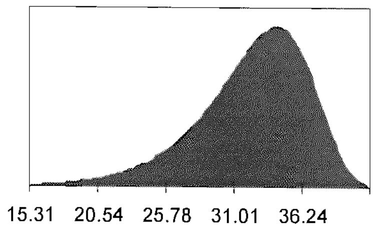
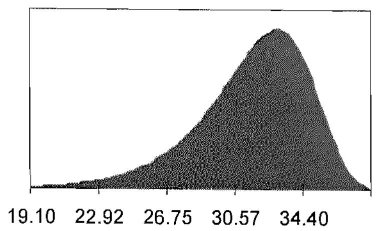
C.6: SCOOT speed distribution of Bitterne inbound links

Link No.	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
1. N10361A	Normal distribution 	Mean 33.79 S.D. 3.03	Test: A-D 0.35 Chi-square 9.11 K-S 0.05 Descriptive statistics: Mean 33.79 Minimum 24 Maximum 41.5 Obs. 144	
2. N10351E	Normal distribution 	Mean 33.79 S.D. 3.03	Test: A-D 0.36 Chi-square 9.11 K-S 0.05 Descriptive statistics: Mean 33.79 Minimum 24 Maximum 41 Obs. 144	
3. N10341D	Student's t distribution 	Midpoint 33.00 Scale 1.51 Deg. Freedom 3.89475	Test: A-D 0.54 Chi-square 15.07 K-S 0.08 Descriptive statistics: Mean 33.06 Minimum 24.67 Maximum 38.33 Obs. 144	

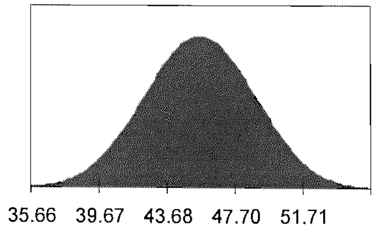
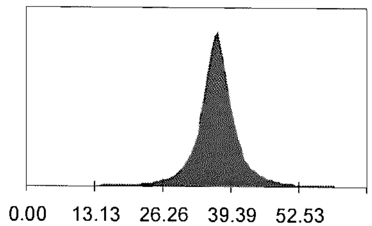
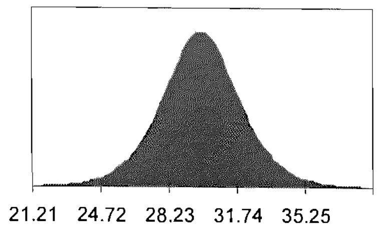
C.6: SCOOT speed distribution of Bitterne inbound links (continued)

Link No.	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
4. N10331E	Logistic distribution 	Mean 32.47 Scale 1.33	Test: A-D 0.54 Chi-square 20.49 K-S 0.06 Descriptive statistics: Mean 32.49 Minimum 27.33 Maximum 39.33 Obs. 144	
5. N10321D	Beta distribution 	Minimum 21.74 Maximum 44.47 Alpha 6.55855 Beta 5.26367	Test: A-D 0.41 Chi-square 13.08 K-S 0.07 Descriptive statistics: Mean 34.35 Minimum 26 Maximum 41 Obs. 144	
6. N10121D	Student's t distribution 	Midpoint 24.67 Scale 1.44 Deg. Freedom 8.82276	Test: A-D 0.72 Chi-square 7.31 K-S 0.08 Descriptive statistics: Mean 24.87 Minimum 26.67 Maximum 43.33 Obs. 144	

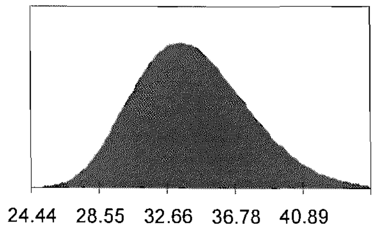
C.6: SCOOT speed distribution of Bitterne inbound links (continued)

Link No.	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
7. N10111G	Maximum extreme distribution  	Likeliest 38.56 Scale 1.59	Test: A-D 0.50 Chi-square 19.94 K-S 0.11 Descriptive statistics: Mean 39.45 Minimum 35.67 Maximum 44.33 Obs. 144	
8. N10241A	Minimum Extreme distribution  	Likeliest 34.29 Scale 3.58	Test: A-D 1.68 Chi-square 30.42 K-S 0.10 Descriptive statistics: Mean 32.12 Minimum 18.67 Maximum 40 Obs. 144	
9. N10231E	Minimum Extreme distribution  	Likeliest 32.97 Scale 2.62	Test: A-D 0.85 Chi-square 10.19 K-S 0.08 Descriptive statistics: Mean 31.56 Minimum 25.33 Maximum 38.33 Obs. 144	

C.6: SCOOT speed distribution of Bitterne inbound links (continued)

Link No.	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
10.N10221F	Normal distribution 	Mean 45.59 Scale 3.21	Test: A-D 0.58 Chi-square 7.49 K-S 0.05 Descriptive statistics: Mean 45.59 Minimum 39 Maximum 54.67 Obs. 144	
11.N10214D	Student's t distribution 	Midpoint 36.67 Scale 2.66 Deg. Freedom 1.75891	Test: A-D 0.67 Chi-square 19.76 K-S 0.06 Descriptive statistics: Mean 36.24 Minimum 16 Maximum 47.33 Obs. 144	
12.N07221E	Logistic distribution 	Mean 29.90 Scale 1.26	Test: A-D 0.40 Chi-square 12.36 K-S 0.07 Descriptive statistics: Mean 29.93 Minimum 24 Maximum 37.33 Obs. 144	

C.6: SCOOT speed distribution of Bitterne inbound links (continued)

Link No.	Value/ Function/ Fitted distribution	Distribution parameters	Statistics	Note
13.N07211D	Gamma distribution 	Location 18.76 Scale 0.80 Shape 19.46803	Test: A-D 0.50 Chi-square 9.47 K-S 0.08 Descriptive statistics: Mean 34.26 Minimum 24.67 Maximum 50.33 Obs. 144	
14.City centre	Fixed speed of 20 kph is assumed			

C.7: Average signal timing of Bitterne inbound links

Order	Junction node	Green time(s)	Red time - T_{Red-i} (s)	Cycle time(s)	Green time per cycle (%)	$P(Red)_i$
1	N10361A	24.8	57.43	82.23	0.3016	0.6984
2	N10351E	54.18	42.11	96.29	0.5627	0.4373
3	N10341D	38.85	44.23	83.08	0.4676	0.5324
4	N10331E	51.78	31.45	83.23	0.6221	0.3779
5	N10321D	63.02	22.75	85.77	0.7348	0.2652
6	N10121D	42.46	35.37	77.83	0.5455	0.4545
7	N10111G	76.53	33.32	109.85	0.6967	0.3033
8	N10241A	72.19	31.12	103.31	0.6988	0.3012
9	N10231E	83.28	17.74	101.02	0.8244	0.1756
10	N10221F	41.49	50.86	92.35	0.4493	0.5507
11	N10214D	179.5	25	204.5	0.8778	0.1222
12	N07221E	31.13	29.82	60.95	0.5107	0.4893
13	N07211D	33.67	65.25	98.92	0.3404	0.6596

Appendix D

Simulation report of independent validation

Contents

D.1: Independent validation: route-based Monte Carlo model

Forecasts

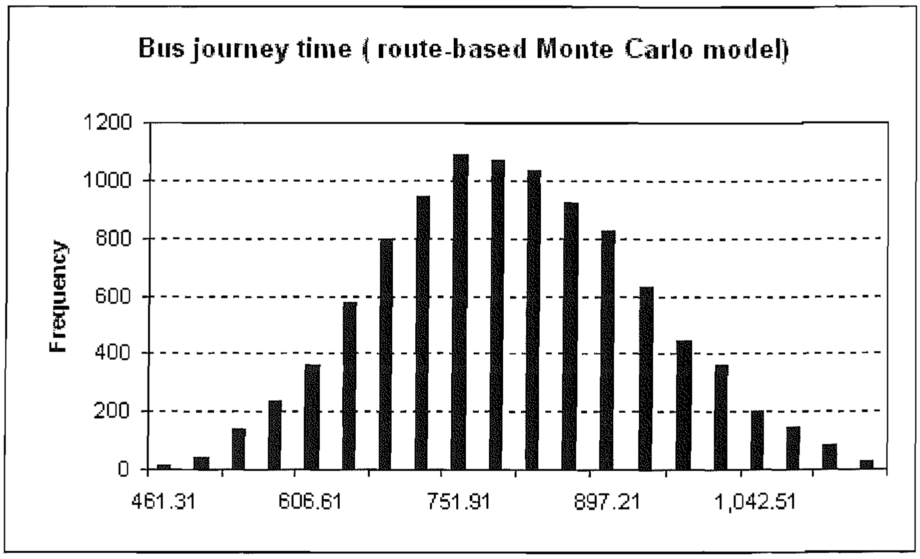
Forecast: Bus journey time

Summary:

Entire range is from 443.14 to 1,481.76

Base case is 812.75

After 10,000 trials, the std. error of the mean is 1.31

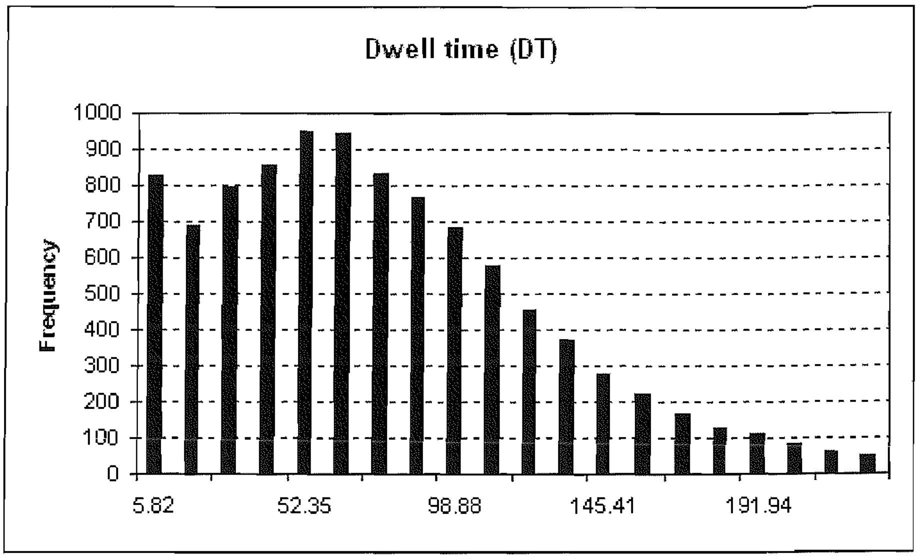


Statistics:	Forecast values
Trials	10,000
Mean	803.19
Median	796.90
Mode	---
Standard Deviation	130.88
Variance	17,129.04
Skewness	0.26987
Kurtosis	3.05
Coeff. of Variability	0.16295
Minimum	443.14
Maximum	1,481.76
Range Width	1,038.62
Mean Std. Error	1.31

Percentiles:	Forecast values
0%	443.14
10%	639.00
20%	690.62
30%	730.15
40%	763.43
50%	796.90
60%	831.57
70%	870.41
80%	912.64
90%	976.66
100%	1,481.76

Forecast: Dwell time (DT)

Summary:
Entire range is from 0.00 to 550.53
Base case is 34.05
After 10,000 trials, the std. error of the mean is 0.55

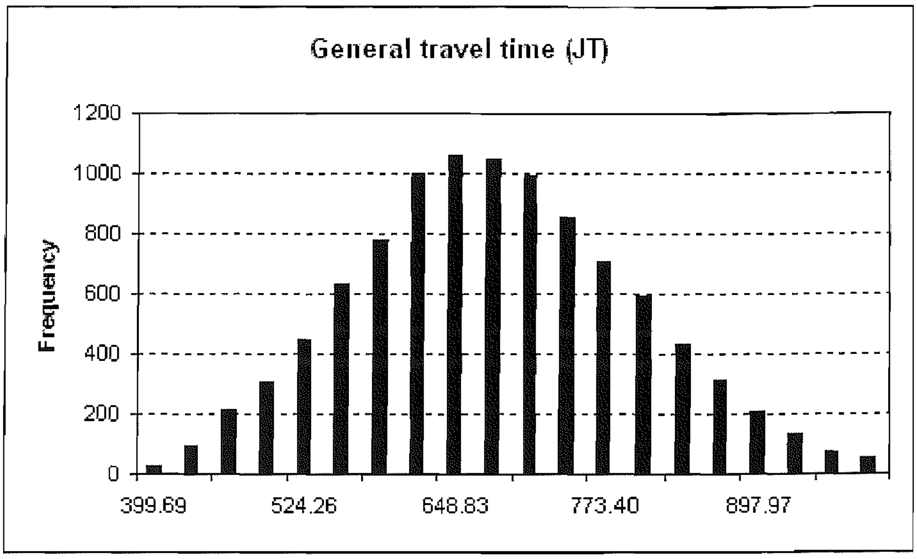


Statistics:	Forecast values
Trials	10,000
Mean	77.92
Median	68.90
Mode	0.00
Standard Deviation	55.26
Variance	3,053.92
Skewness	1.20
Kurtosis	5.71
Coeff. of Variability	0.70919
Minimum	0.00
Maximum	550.53
Range Width	550.53
Mean Std. Error	0.55

Percentiles:	Forecast values
0%	0.00
10%	14.61
20%	30.40
30%	44.16
40%	56.68
50%	68.90
60%	82.93
70%	99.00
80%	117.93
90%	149.52
100%	550.53

Forecast: General travel time (JT)

Summary:
Entire range is from 384.11 to 1,134.48
Base case is 734.16
After 10,000 trials, the std. error of the mean is 1.16

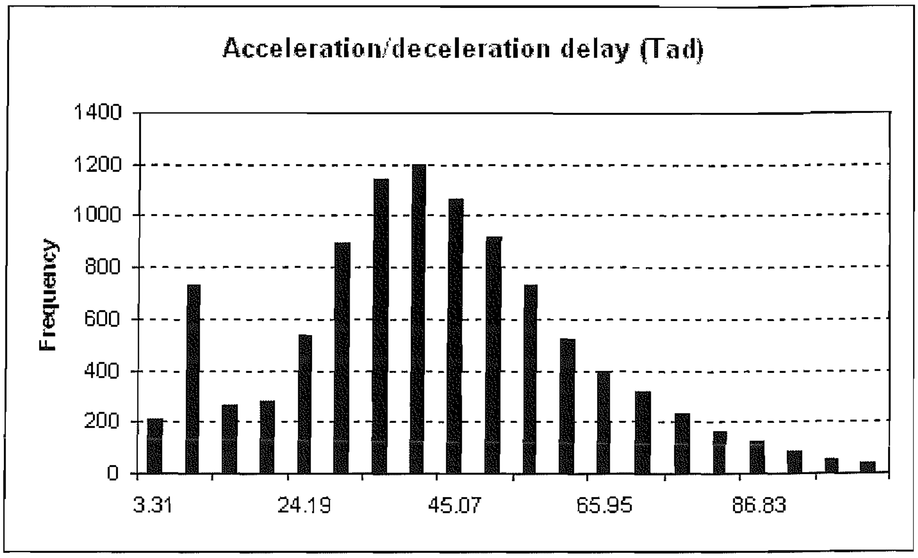


Statistics:	Forecast values
Trials	10,000
Mean	682.40
Median	676.81
Mode	---
Standard Deviation	115.92
Variance	13,437.04
Skewness	0.22104
Kurtosis	2.83
Coeff. of Variability	0.16987
Minimum	384.11
Maximum	1,134.48
Range Width	750.37
Mean Std. Error	1.16

Percentiles:	Forecast values
0%	384.11
10%	534.36
20%	582.60
30%	619.06
40%	648.02
50%	676.81
60%	707.26
70%	740.65
80%	780.09
90%	836.29
100%	1,134.48

Forecast: Acceleration/deceleration delay (Tad)

Summary:
Entire range is from 0.70 to 342.00
Base case is 44.54
After 10,000 trials, the std. error of the mean is 0.22



Statistics:	Forecast values
Trials	10,000
Mean	42.87
Median	41.32
Mode	---
Standard Deviation	22.23
Variance	494.02
Skewness	1.31
Kurtosis	12.44
Coeff. of Variability	0.51846
Minimum	0.70
Maximum	342.00
Range Width	341.30
Mean Std. Error	0.22

Percentiles:	Forecast values
0%	0.70
10%	11.87
20%	26.63
30%	32.43
40%	36.99
50%	41.32
60%	45.92
70%	51.48
80%	58.34
90%	70.06
100%	342.00

End of Forecasts

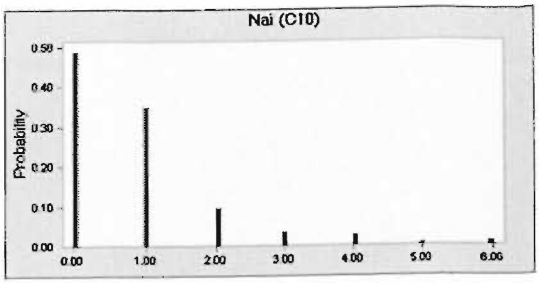
Assumptions

DT module

Assumption: Nai

Custom distribution with parameters:

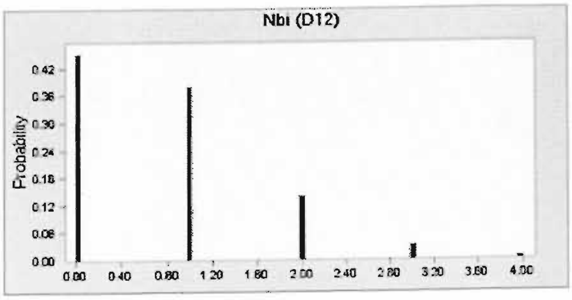
Value	Probability
0.00	0.49
1.00	0.35
2.00	0.09
3.00	0.03
4.00	0.02
5.00	0.00
6.00	0.01



Assumption: Nbi

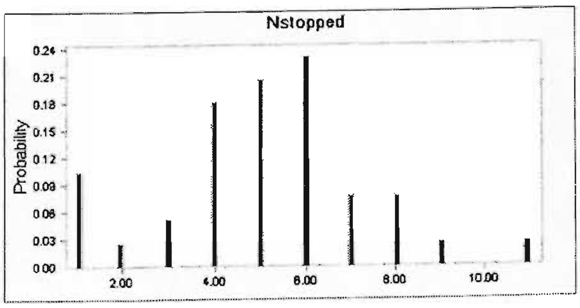
Custom distribution with parameters:

Value	Probability
0.00	0.45
1.00	0.38
2.00	0.14
3.00	0.03
4.00	0.00



Assumption: Nstopped

Value	Probability
1.00	0.10
2.00	0.03
3.00	0.05
4.00	0.18
5.00	0.21
6.00	0.23
7.00	0.08
8.00	0.08
9.00	0.03
11.00	0.03

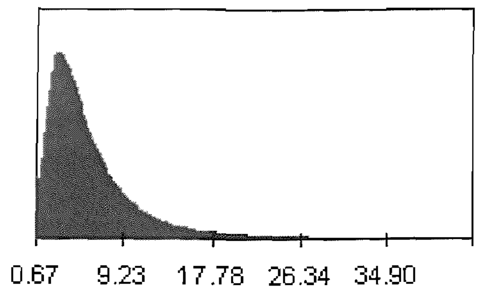


Assumption: Tai

Lognormal distribution with parameters:

Mean	6.07
Std. Dev.	4.98

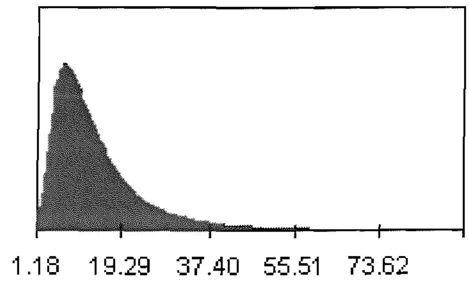
Selected range is from 0.67 to 55.00



Assumption: Tbi

Lognormal distribution with parameters:

Mean	13.99
Std. Dev.	10.73



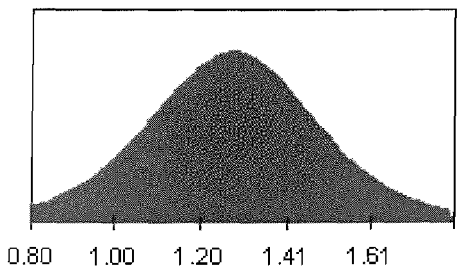
JT module

Assumption: ANPR%

Logistic distribution with parameters:

Mean	1.28
Scale	0.13

Selected range is from 0.80 to 1.80

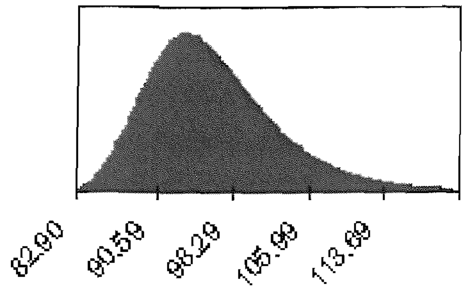


Assumption: TANPR

Maximum Extreme distribution with parameters:

Likeliest	93.51
Scale	5.49

Selected range is from -Infinity to 121.00



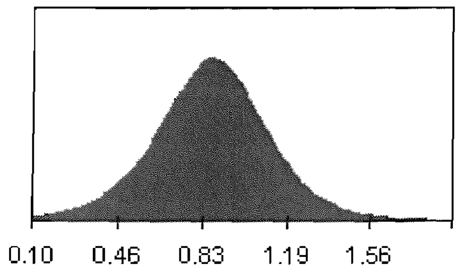
Tad module

Assumption: A

Logistic distribution with parameters:

Mean	0.87
Scale	0.15

Selected range is from 0.10 to 1.99

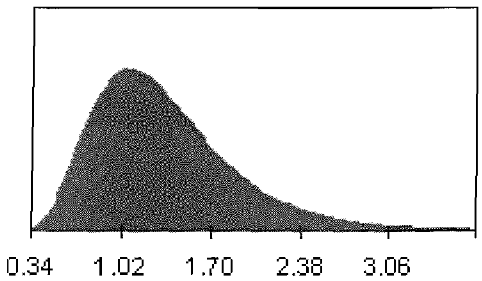


Assumption: D

Lognormal distribution with parameters:

Mean	1.37
Std. Dev.	0.57

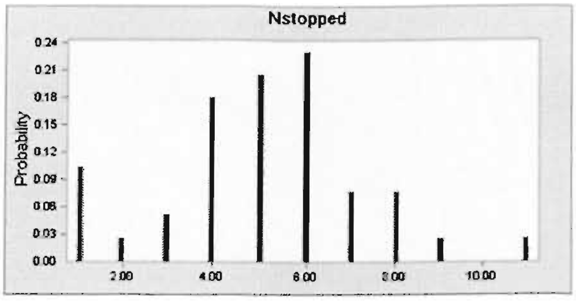
Selected range is from 0.31 to 3.71



Assumption: Nstopped

Custom distribution with parameters:

Value	Probability
1.00	0.10
2.00	0.03
3.00	0.05
4.00	0.18
5.00	0.21
6.00	0.23
7.00	0.08
8.00	0.08
9.00	0.03
11.00	0.03

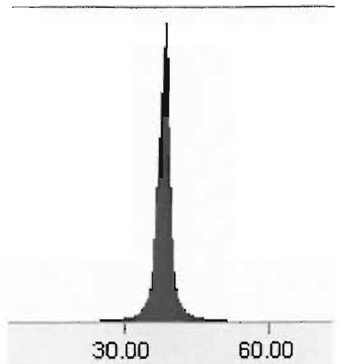


Assumption: STS

Student's t distribution with parameters:

Midpoint	38.02
Scale	0.90
Deg. Freedom	1.252905572

Selected range is from 1.00 to Infinity

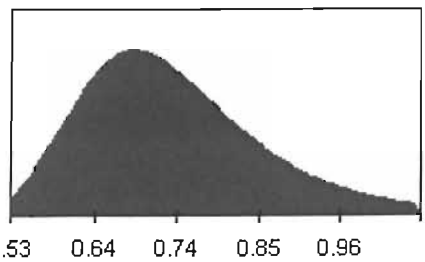


Assumption: STs%

Maximum Extreme distribution with parameters:

Likeliest	0.69
Scale	0.10

Selected range is from 0.53 to 1.06



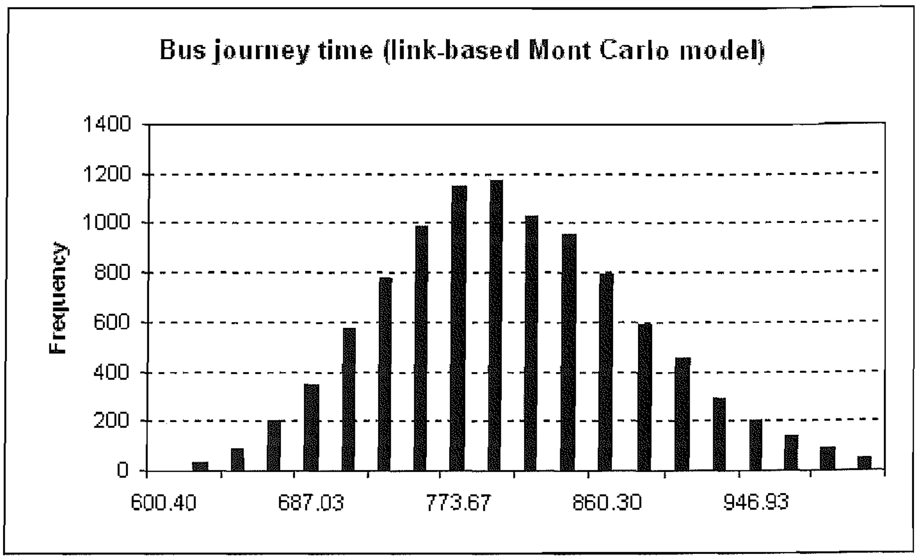
End of Assumptions

D.2: Independent validation: link-based Monte Carlo model

Forecasts

Forecast: Bus journey time

Summary:
Entire range is from 587.84 to 1,194.56
Base case is 1,015.44
After 10,000 trials, the std. error of the mean is 0.77

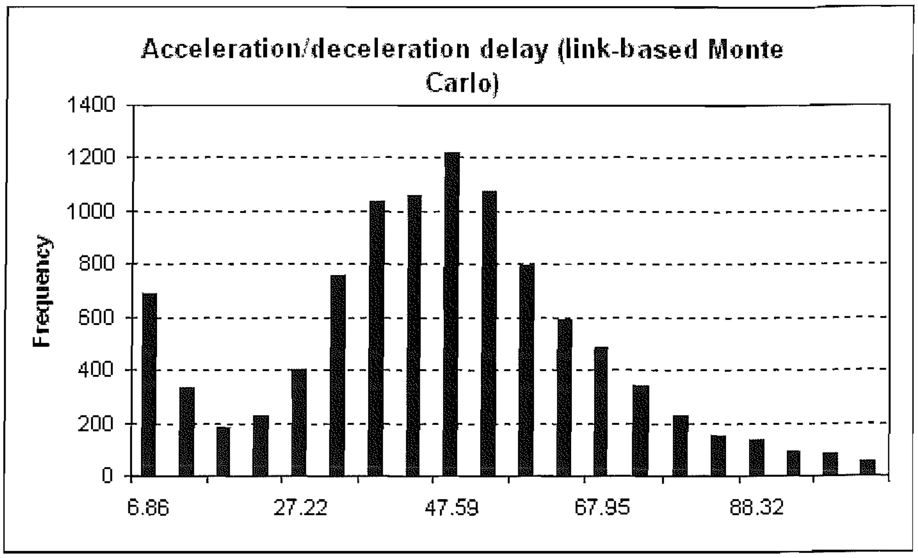


Statistics:	Forecast values
Trials	10,000
Mean	806.15
Median	800.15
Mode	---
Standard Deviation	77.35
Variance	5,983.18
Skewness	0.47040
Kurtosis	3.45
Coeff. of Variability	0.09595
Minimum	587.84
Maximum	1,194.56
Range Width	606.72
Mean Std. Error	0.77

Percentiles:	Forecast values
0%	587.84
10%	711.80
20%	740.63
30%	762.69
40%	781.70
50%	800.15
60%	820.29
70%	842.41
80%	867.95
90%	907.13
100%	1,194.56

Forecast: Acceleration/deceleration delay

Summary:
Entire range is from 4.31 to 128.88
Base case is 57.83
After 10,000 trials, the std. error of the mean is 0.21

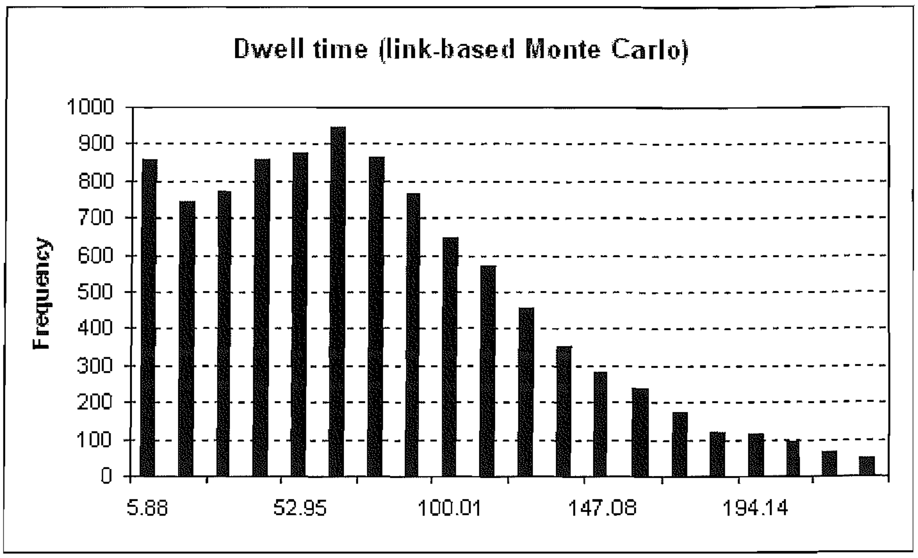


Statistics:	Forecast values
Trials	10,000
Mean	46.53
Median	46.43
Mode	---
Standard Deviation	21.29
Variance	453.22
Skewness	0.27008
Kurtosis	3.40
Coeff. of Variability	0.45756
Minimum	4.31
Maximum	128.88
Range Width	124.57
Mean Std. Error	0.21

Percentiles:	Forecast values
0%	4.31
10%	14.06
20%	31.21
30%	37.00
40%	41.92
50%	46.43
60%	50.71
70%	55.40
80%	62.18
90%	72.73
100%	128.88

Forecast: Dwell time

Summary:
Entire range is from 0.00 to 502.35
Base case is 204.30
After 10,000 trials, the std. error of the mean is 0.56

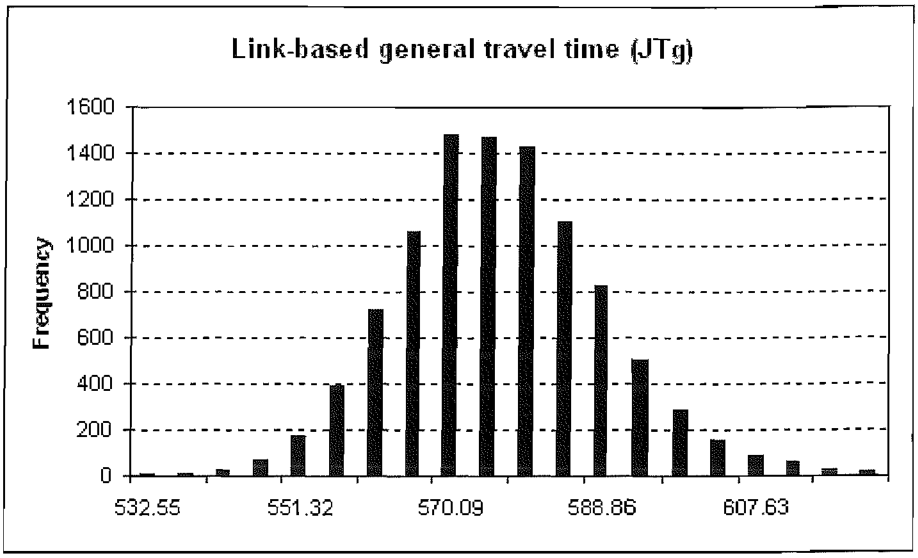


Statistics:	Forecast values
Trials	10,000
Mean	78.64
Median	69.86
Mode	0.00
Standard Deviation	55.96
Variance	3,131.02
Skewness	1.07
Kurtosis	4.77
Coeff. of Variability	0.71150
Minimum	0.00
Maximum	502.35
Range Width	502.35
Mean Std. Error	0.56

Percentiles:	Forecast values
0%	0.00
10%	13.51
20%	29.55
30%	43.89
40%	57.61
50%	69.86
60%	83.39
70%	99.49
80%	119.96
90%	152.97
100%	502.35

Forecast: Link-based general travel time (JTg)

Summary:
Entire range is from 523.72 to 945.67
Base case is 544.64
After 10,000 trials, the std. error of the mean is 0.17

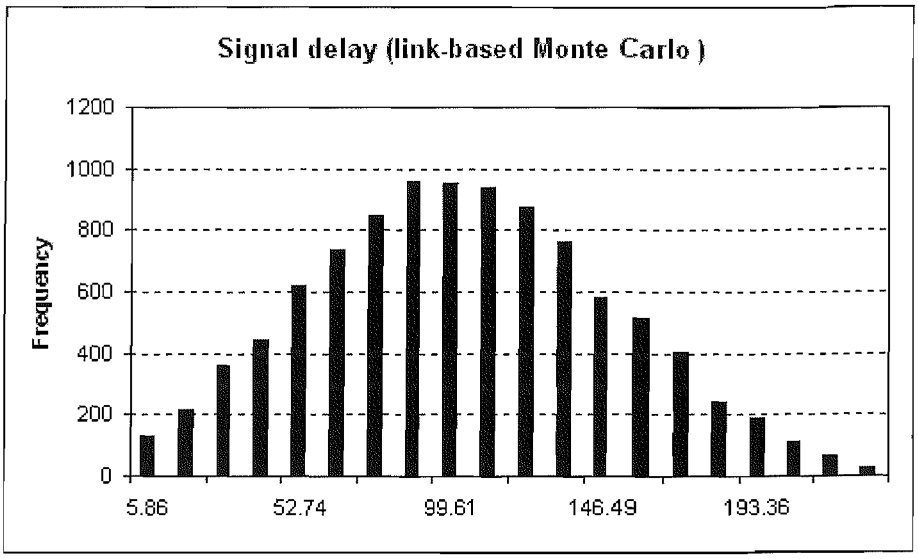


Statistics:	Forecast values
Trials	10,000
Mean	577.13
Median	575.81
Mode	---
Standard Deviation	16.76
Variance	280.83
Skewness	5.37
Kurtosis	85.17
Coeff. of Variability	0.02904
Minimum	523.72
Maximum	945.67
Range Width	421.95
Mean Std. Error	0.17

Percentiles:	Forecast values
0%	523.72
10%	560.88
20%	565.97
30%	569.57
40%	572.63
50%	575.81
60%	578.99
70%	582.59
80%	586.80
90%	593.22
100%	945.67

Forecast: Signal delay

Summary:
Entire range is from 0.00 to 279.76
Base case is 208.67
After 10,000 trials, the std. error of the mean is 0.47



Statistics:	Forecast values
Trials	10,000
Mean	103.85
Median	102.18
Mode	0.00
Standard Deviation	46.62
Variance	2,173.06
Skewness	0.20260
Kurtosis	2.69
Coeff. of Variability	0.44886
Minimum	0.00
Maximum	279.76
Range Width	279.76
Mean Std. Error	0.47

Percentiles:	Forecast values
0%	0.00
10%	43.05
20%	62.31
30%	77.67
40%	90.28
50%	102.18
60%	114.79
70%	127.90
80%	143.81
90%	165.99
100%	279.76

End of Forecasts

Assumptions

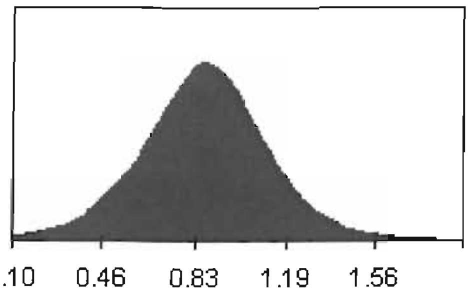
Dacc module

Assumption: A

Logistic distribution with parameters:

Mean	0.87
Scale	0.15

Selected range is from 0.10 to 1.99

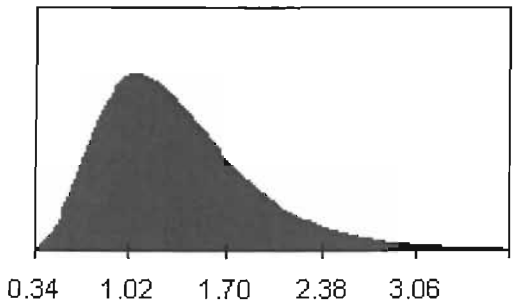


Assumption: D

Lognormal distribution with parameters:

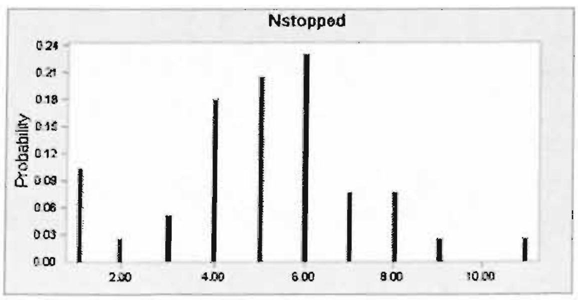
Mean	1.37
Std. Dev.	0.57

Selected range is from 0.31 to 3.71



Assumption: Nstopped

Value	Probability
1.00	10.26
2.00	2.56
3.00	5.13
4.00	17.95
5.00	20.51
6.00	23.08
7.00	7.69
8.00	7.69
9.00	2.56
11.00	2.56

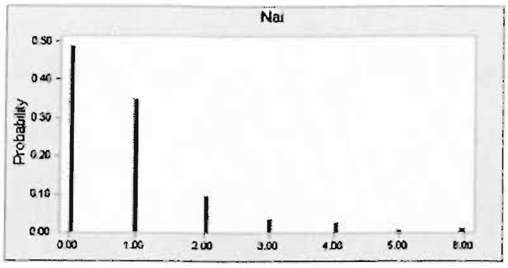


DT module

Assumption: Nai

Custom distribution with parameters:

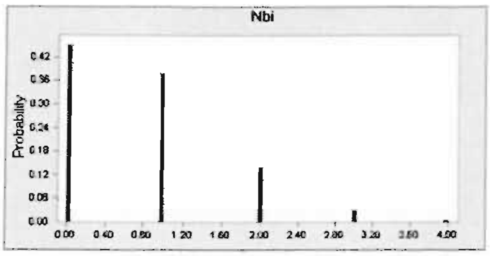
Value	Probability
0.00	0.49
1.00	0.35
2.00	0.09
3.00	0.03
4.00	0.02
5.00	0.00
6.00	0.01



Assumption: Nbi

Custom distribution with parameters:

Value	Probability
0.00	0.45
1.00	0.38
2.00	0.14
3.00	0.03
4.00	0.00

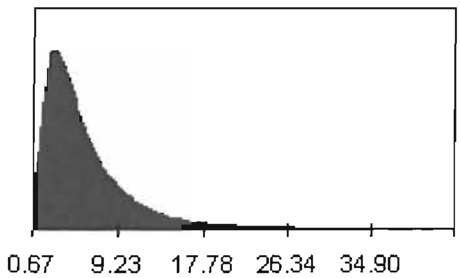


Assumption: Tai

Lognormal distribution with parameters:

Mean	6.07
Std. Dev.	4.98

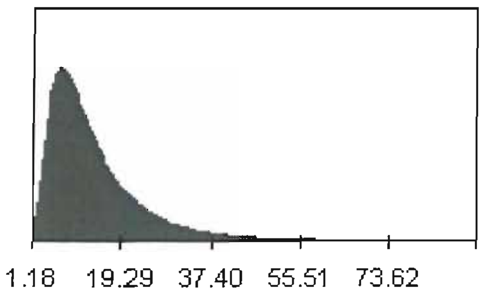
Selected range is from 0.67 to 55.00



Assumption: Tbi

Lognormal distribution with parameters:

Mean	13.99
Std. Dev.	10.73



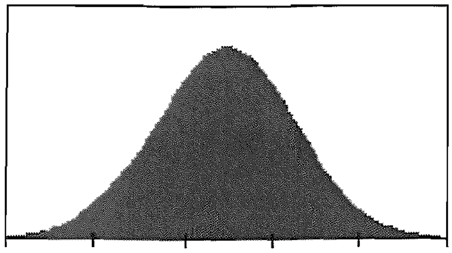
JTg module

Assumption: Vbus of Chilworth link

Normal distribution with parameters:

Mean	50.00
Std. Dev.	5.00

Selected range is from 1.00 to Infinity



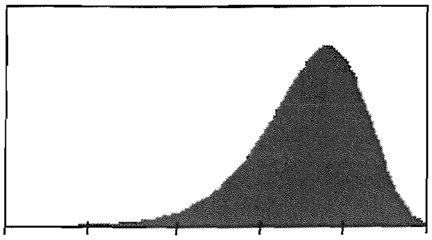
JTg module

Assumption: N03111E-STs

Weibull distribution with parameters:

Location	-9.41
Scale	48.83
Shape	15

Selected range is from 1.00 to Infinity

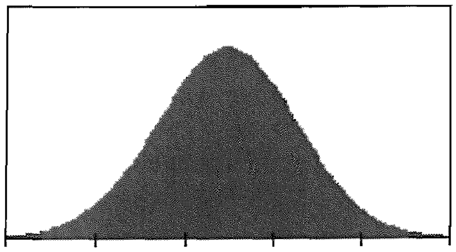


Assumption: N03112C-STs

Normal distribution with parameters:

Mean	24.02
Std. Dev.	2.09

Selected range is from 1.00 to Infinity

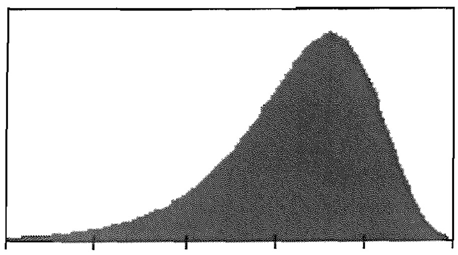


Assumption: N03214A-STs

Minimum Extreme distribution with parameters:

Likeliest	44.75
Scale	4.34

Selected range is from 1.00 to Infinity

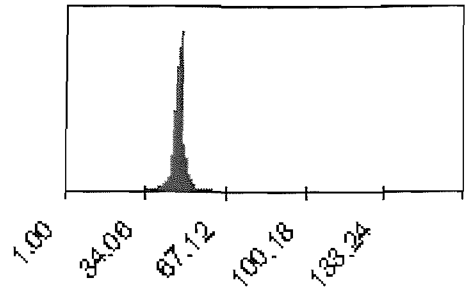


Assumption: N03244A-STs

Student's t distribution with parameters:

Midpoint	48.33
Scale	1.16
Deg. Freedom	1.072185932

Selected range is from 1.00 to Infinity



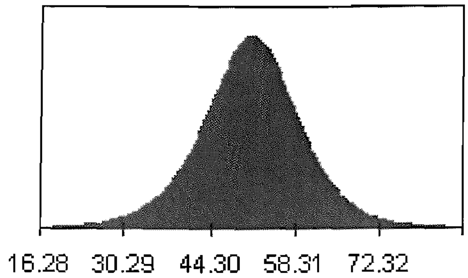
JTg module

Assumption: N03244E-STs

Logistic distribution with parameters:

Mean	50.95
Scale	5.02

Selected range is from 1.00 to Infinity

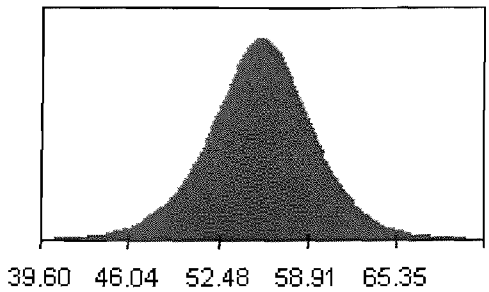


Assumption: N03311J-STs

Logistic distribution with parameters:

Mean	55.53
Scale	2.31

Selected range is from 1.00 to Infinity

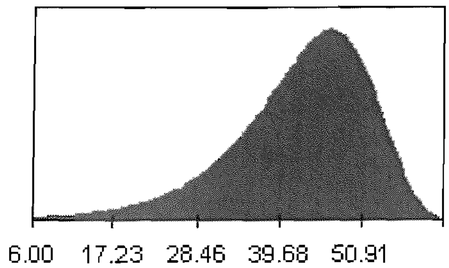


Assumption: N03311L-STs

Minimum Extreme distribution with parameters:

Likeliest	46.72
Scale	7.69

Selected range is from 1.00 to Infinity

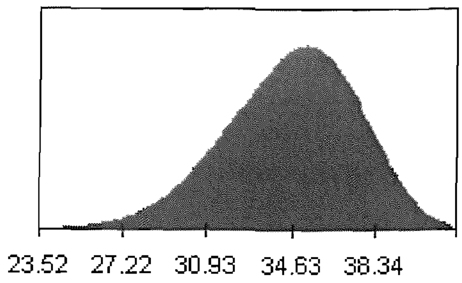


Assumption: N04141E-STs

Weibull distribution with parameters:

Location	20.73
Scale	15.11
Shape	5.449114758

Selected range is from 1.00 to Infinity



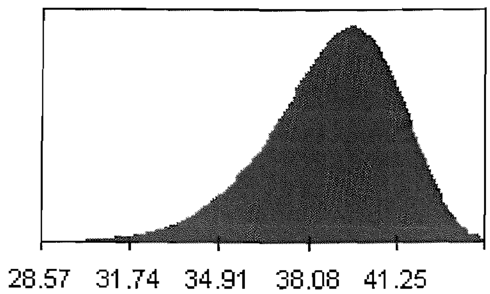
JTg module

Assumption: N04144C-STs

Weibull distribution with parameters:

Location	23.64
Scale	16.27
Shape	7.705393451

Selected range is from 1.00 to Infinity

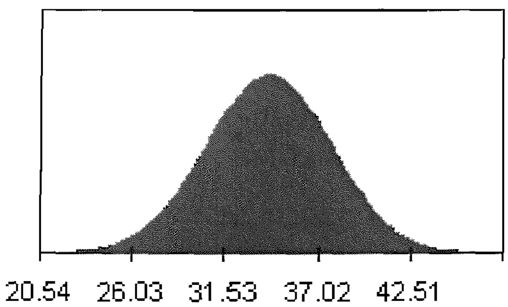


Assumption: N4155A-STs

Beta distribution with parameters:

Minimum	-18.38
Maximum	86.65
Alpha	100
Beta	100

Selected range is from 1.00 to Infinity

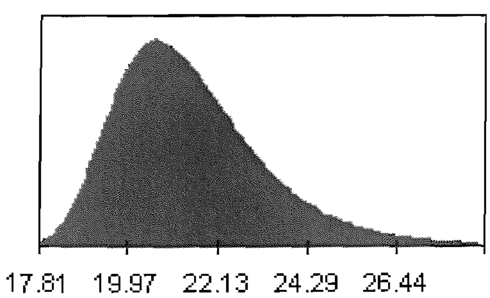


Assumption: N04157K-STs

Maximum Extreme distribution with parameter

Likeliest	20.67
Scale	1.48

Selected range is from 1.00 to Infinity

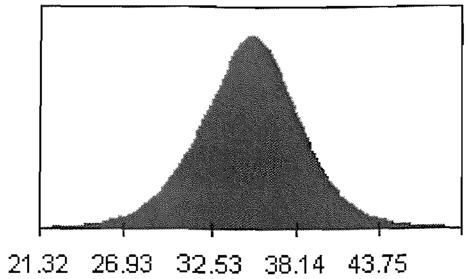


Assumption: N04158K-STs

Logistic distribution with parameters:

Mean	35.20
Scale	2.01

Selected range is from 1.00 to Infinity



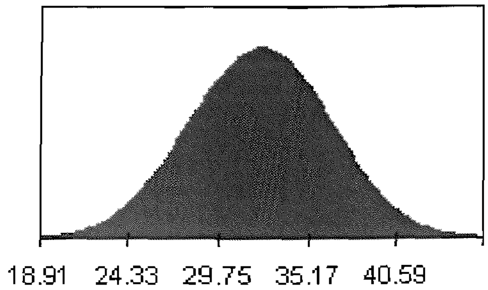
JTg module

Assumption: N04158XA-STs

Normal distribution with parameters:

Mean	32.33
Std. Dev.	4.34

Selected range is from 1.00 to Infinity

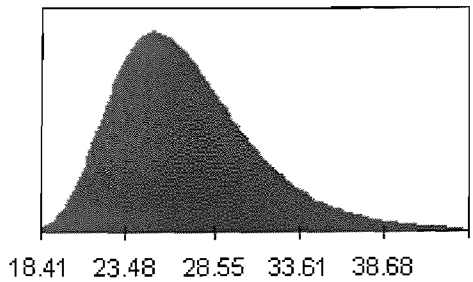


Assumption: N07331I-STs

Maximum Extreme distribution with parameters:

Likeliest	25.12
Scale	3.47

Selected range is from 1.00 to Infinity

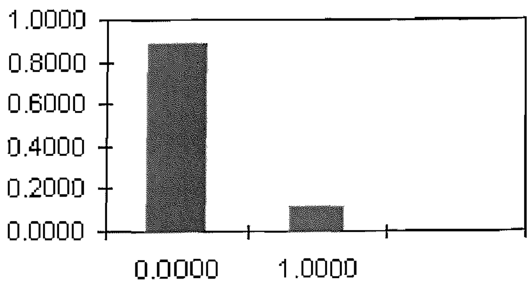


Sig module

Assumption: N03214A-Probability of Red

Yes-No distribution with parameters:

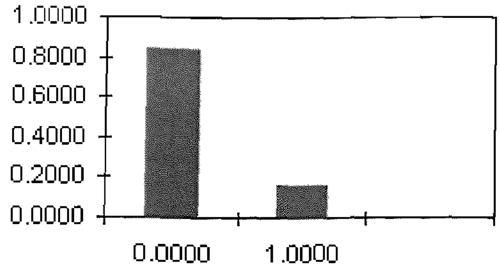
Probability of Yes(1)	0.1125



Assumption: N04157K-Probability of Red

Yes-No distribution with parameters:

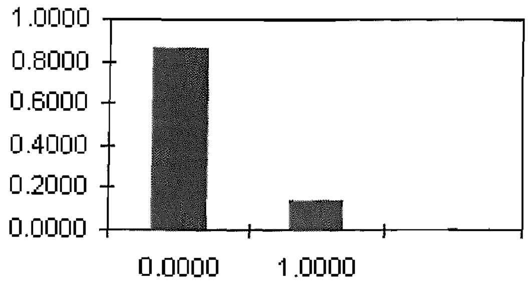
Probability of Yes(1) 0.1631



Assumption: N04155A-Probability of Red

Yes-No distribution with parameters:

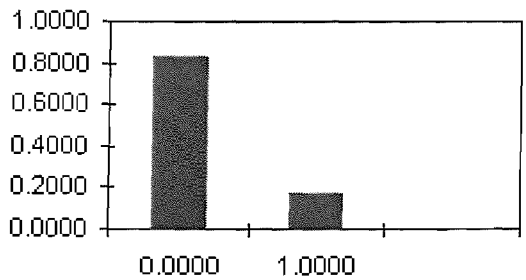
Probability of Yes(1) 0.1338



Assumption: N04158X-Probability of Red

Yes-No distribution with parameters:

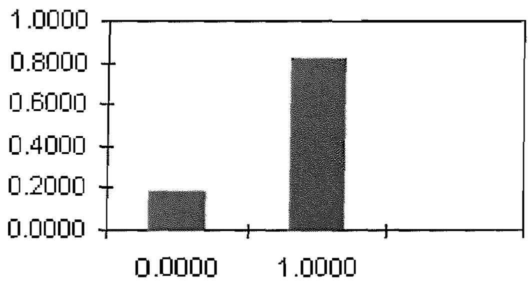
Probability of Yes(1) 0.1702



Assumption: N07331I-Probability of Red

Yes-No distribution with parameters:

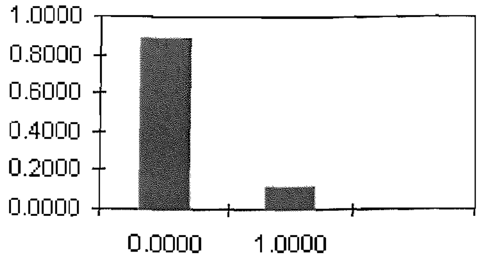
Probability of Yes(1) 0.8194



Assumption: N03244A-Probability of Red

Yes-No distribution with parameters:

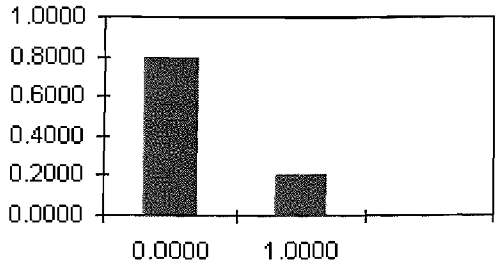
Probability of Yes(1) 0.1138



Assumption: N03311J-Probability of Red

Yes-No distribution with parameters:

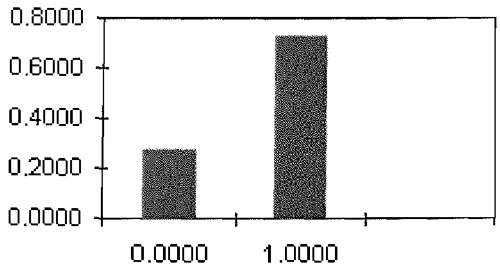
Probability of Yes(1) 0.2011



Assumption: N3111E-Probability of Red

Yes-No distribution with parameters:

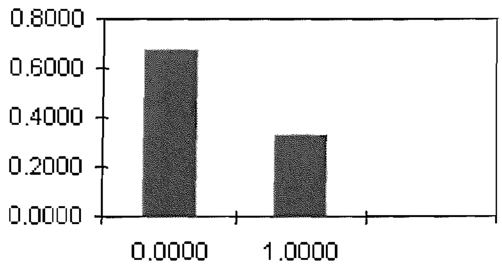
Probability of Yes(1) 0.7298



Assumption: N3112C-Probability of Red

Yes-No distribution with parameters:

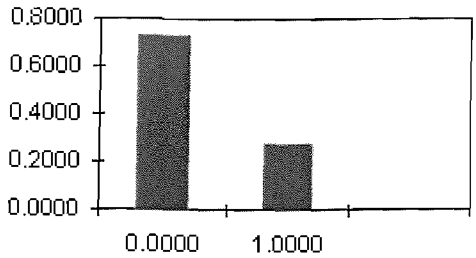
Probability of Yes(1) 0.3266



Assumption: N04144C-Probability of Red

Yes-No distribution with parameters:

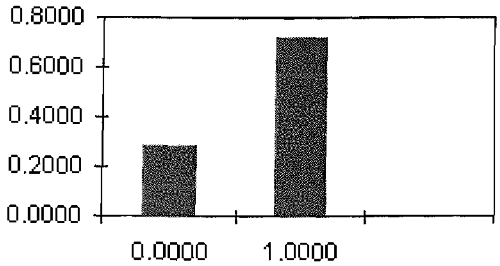
Probability of Yes(1)	0.2726



Assumption: N04141E-Probability of Red

Yes-No distribution with parameters:

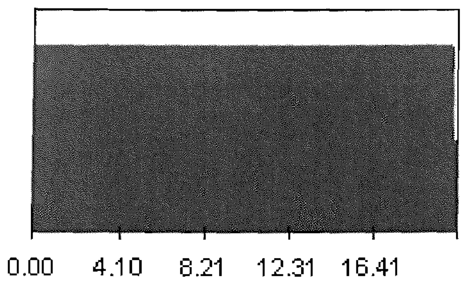
Probability of Yes(1)	0.7157



Assumption: N04157K - Red period

Uniform distribution with parameters:

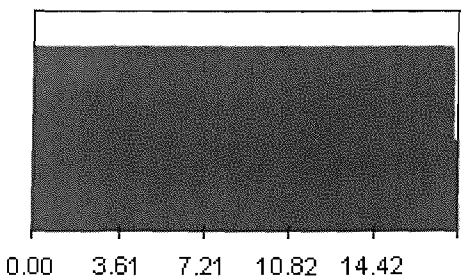
Minimum	0.00
Maximum	20.31



Assumption: N04155A - Red period

Uniform distribution with parameters:

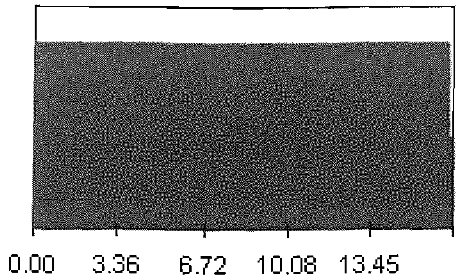
Minimum	0.00
Maximum	17.85



Assumption: N04158X - Red period

Uniform distribution with parameters:

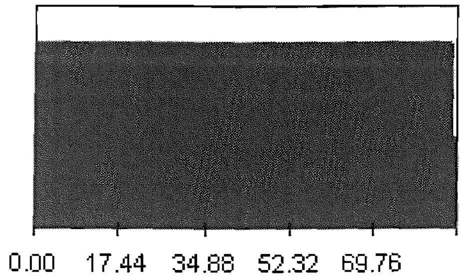
Minimum	0.00
Maximum	16.64



Assumption: N07331I - Red period

Uniform distribution with parameters:

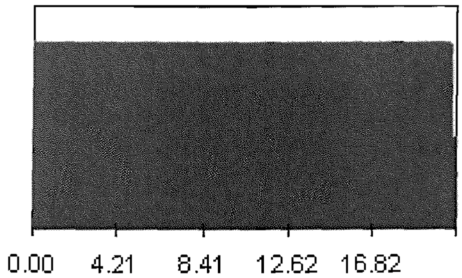
Minimum	0.00
Maximum	86.33



Assumption: N03244A - Red period

Uniform distribution with parameters:

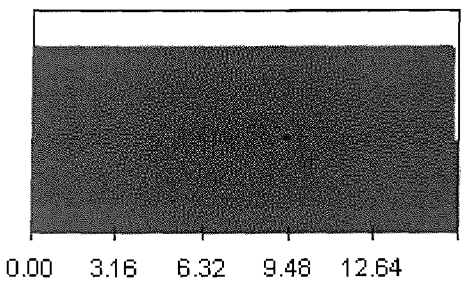
Minimum	0.00
Maximum	20.82



Assumption: N03311J - Red period

Uniform distribution with parameters:

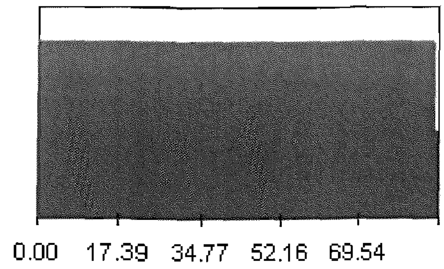
Minimum	0.00
Maximum	15.64



Assumption: N03111E - Red period

Uniform distribution with parameters:

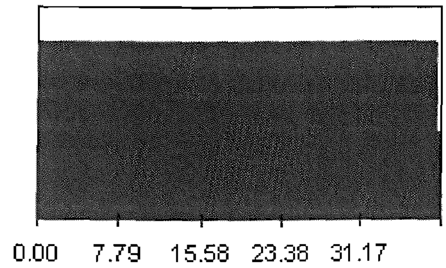
Minimum	0.00
Maximum	86.06



Assumption: N03112C - Red period

Uniform distribution with parameters:

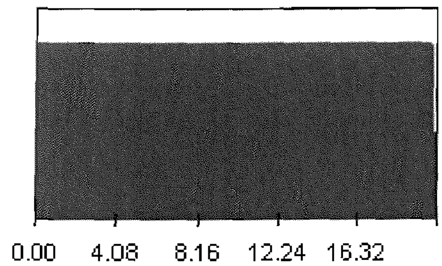
Minimum	0.00
Maximum	38.57



Assumption: N04144C - Red period

Uniform distribution with parameters:

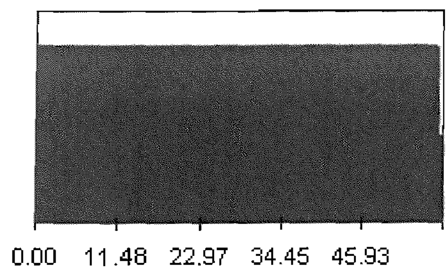
Minimum	0.00
Maximum	20.20



Assumption: N04141E - Red period

Uniform distribution with parameters:

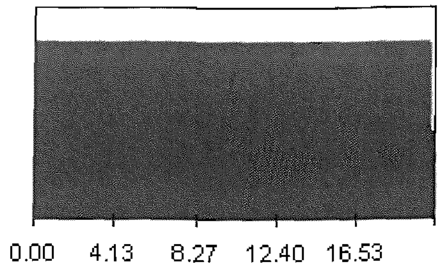
Minimum	0.00
Maximum	56.84



Assumption: N03214A - Red period

Uniform distribution with parameters:

Minimum	0.00
Maximum	20.46



End of Assumptions

References

Abdelfattah, A. M., and Khan, A. M., 1998, Models for predicting bus delays, *Transportation Research Record* 1623, Transportation Research Board, National Research Council, Washington DC, pp.8-15.

Adamchuk, V., 2000, Longitude and latitude conversion, Available from: http://pasture.ecn.purdue.edu/~abegps/web_ssm/web_GPS_eq.html, Accessed on 22 September 2004.

Aerospace Corporation Education, 1999, Available from: http://www.aero.org/publications/GPSPRIMER/index.html. Accessed on 25 August 2004.

Anderson, J. and Bell, M., 1998, Travel time estimation in urban road networks, *IEEE*, pp.924-929.

Angel, A., Hickman, M., Mirchandani, P., and Chandnani, D., 2003, Methods of analyzing traffic imagery collected from aerial platforms, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, No. 2, pp.99-107.

Anthony, A. S., 2001, Model for determining optimum bus-stop spacing in urban areas, *Journal of Transportation Engineering*, May/June.

Ashjaee, J, 2006, Galileo, GLONASS, and GPS, GPS World, Febuary. Available from: http://www.gpsworld.com/gpsworld/article/articleDetail.jsp?id=300303, Accessed on 6 March 2006.

Barnett, A., 1974, On controlling randomness in transit operations, *Transportation Science*, Vol.8, No.2, pp.101-116.

Belliss, G.2004, Detailed speed and travel time surveys using low cost GPS equipment, *IPENZ Transportation Group Technical Conference*, New Zealand.

Bertini, R. L., and El-Geneidy, A., 2003, Generating transit performance measures with archived data, *Transportation Research Record* 1841.

Bertini, R. L., and El-Geneidy, A., 2004, Modelling transit trip time using archived bus dispatch system data, *Journal of Transportation Engineering* , ASCE Jan/Feb.

Bertini, R. L., and Tantiyanugulchai, S., 2004, Transit buses as traffic probes: empirical evaluation using geo-location data, *Transportation Research Board 2004 Annual Meeting* CD-ROM.

Cathey, F. W., and Dailey, D. J., 2002, Transit vehicles as traffic probe sensors, *Transportation Research Record* 1804.

Cathey, F. W., and Dailey, D. J., 2003, Estimating corridor travel time by using transit vehicles as probes, *Transportation Research Record* 1855.

Chakroborty, P., and Kikuchi, S., 2004, Estimating travel times on urban corridors using bus travel times data, *Transportation Research Board 2004 Annual Meeting* CD-ROM.

Chung, C. A., 2004, Simulation modeling handbook: a practical approach, *CRC press LLC.*

Coifman, B., 2001, Improved velocity estimation using single loop detectors, *Transportation Research Part A*, 35, pp.863-880.

Dailey, D. J., 1993, Travel-time estimation using cross-correlation techniques, *Transportation Research Part B*, 27B, No.2, pp.97-107.

Dailey, D. J., 1999, A statistical algorithm for estimating speed from single loop volume and occupancy measurements, *Transportation Research Part B*, Vol. 33B, No. 5, pp. 313-22.

D'Angello, M. P., Al-Deek, H. M., and Wang, M. C. 1999, Travel-time prediction for freeway corridors, *Transportation Research Record* 1676, Transportation Research Board, Washington, D.C., pp.184-191.

Decisioneering, 2005, Crystal Ball 7.2 user manual, Decisioneering Inc.

Department of Environment, Transport & the Regions, 2001, Traffic Advisory Leaflet 6/01. Bus priority. Traffic Advisory Unit, UK.

Department of Transportation, 2002, Effectiveness of bus signal priority- final report, Report No. NCTR-416-04, Washington D.C.

Department for Transport, 2003a, Transport trends summary, UK.

Department for Transport, 2003b, Bus priority: the way ahead, UK.

Department for Transport, 2004, The future of transport – a network for 2030, UK.

Department for Transport, 2006, Transport trend: 2005 edition, UK.

D'Este, G. M., Zito, R., and Taylor, M. A. P., 1999, Using GPS to measure traffic system performance, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 14(4). pp.255-265.

Dougherty, M., 1995, A review of neural networks applied to transport, *Transport Research part C*, Vol. 3, No. 4, pp. 247-260.

Dueker, K. J., Kimpel, T. J., and Strathaman, J. G., 2004, Determinants of bus dwell time, *Journal of Public Transportation*, Vol. 7, No. 1, pp.21-40.

Dunstan, S., 1997, Travel time data is at hand, *Traffic Technology International*, pp.152-156.

Federal Highway Administration, 2005, ITS user service document, *Department of Transportation*, Washington D.C.

Fernández, R., 1999, Design of bus-stop priorities, *Traffic Engineering and Control*, 40(6), 335-340.

Field, A., 2005, Discovering statistics using SPSS, second edition, *SAGE Publication.*

Fleisher, C. S., and Babette E. B., 2003, Strategic and Competitive Analysis: Methods and Techniques for Analyzing Business Competition, *Prentice Hall.*

Fox, J., 1997, Applied regression analysis, linear models, and related models, *SAGE Publications.*

Fu, L., and Yang, X., 2002, Design and implementation of bus-holding control strategies with real-time information, *Transportation Research Record* 1791, Transportation Research Board, Washington, DC.

Garmin, 1999, GPS 35/36 TracPak™ Technical specification, Garmin Corporation.

Garmin, 2005. What is GPS? Available from: http://www.garmin.com/aboutGPS/, Aaccessed on 23 October 2005.

Garmin, n.d., Available from: http://www.garmin.com/aboutGPS/, Accessed on 8 April 2004.

Gartner, N. H., Messer C. J., Rathi A., Cunard R., Lieu H., and Mahmassani H., 1992, Traffic flow theory- a state-of-the-art report, Transportation Research Board.

Gault, H. E., and Taylor, I. G., 1981, The use of the output from vehicle detectors to assess delay in computer-controlled area traffic control system, University of Newcastle, *Transport Operations Research Group*, Research Report no. 37.

GPS world, 2006, GPS phones will boom, but hurdles loom, April, Available from: http://www.gpsworld.com/gpsworld/article/articleDetail.jsp?id=317551, Accessed on 2 May 2006.

Guenthner, R. P., and Hamat, K.,1988, Transit dwell time under complex fare structure, *Journal of Transportation Engineering*, Vol. 114, No. 3, pp.367-379.

Guenthner, R. P., and Sinha, K. C., 1983, Modeling bus delays due to passenger boardings and alightings, *Transportation Research Record* 915, pp.7-13.

Hellinga, B. and Fu, L., 1999, Assessing expected accuracy of probe vehicle travel time reports, *Journal of Transportation Engineering*, ASCE, 125(6), pp.524-530.

Horbury, A. X.,1999, Using non-real-time automatic vehicle location data to improve bus services. *Transportation Research Part B,* Vol.33, pp.559-579.

Institute of Transportation Engineers, 1994, Manual of transportation engineering studies - Travel time and delay studies, Washington D.C., pp.52-68.

Institute of Transport Research, 2004, Space borne traffic data collection (RAVE) project, Available from: http://www.dlr.de/vf/forschung/projekte/space_borne_traffic_data_collection/rave, Accessed on 17 April 2005.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., and Mizam, A., 1998, Applied regression analysis and other multivariable methods, three edition, *Brooks/Cole Publishing Company.*

Kohavi, R., 2004, What are cross-validation and bootstrapping? Available from: http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-12.html, Accessed on 3 May 2006.

Kraft, W. H., and Deutschman, H., 1977, Bus passenger service-time distributions. *Transportation Research Board* 625, pp.37-43.

Kwon, J., Coifman, B., and Bickel, P., 2000, Day-to-day travel-time trends and travel-time prediction from loop detector data, *Transportation Research Record* 1717, Transportation Research Board, Washington, D.C., pp.120-129.

Law, A. M., and Kelton, D. K.,1991, Simulation modeling & analysis, second edition, *McGraw-Hill*, Inc.

Levine, J. C., and Torng, G. W.,1994, Dwell time effects of low-floor bus design, *Journal of Transportation Engineering*, Vol. 120, No. 6, pp.914-929.

Levinson, H. S., 1983, Analyzing transit travel time performance, *Transportation Research Record* 915, Transportation Research Board, Washington, D.C., pp.1-6.

Li, Y., 2004, Journey time estimation and incident detection using GPS equipped probe vehicle, PhD thesis, University of Southampton, Southampton, UK.

Li Y., and McDonald, M., 2002, Link travel time estimation using GPS equipped probe vehicle, *The IEEE 5th International Conference on Intelligent Transportation Systems* , 3-6 September 2002, Singapore.

Li Y., and McDonald, M., n.d., Link travel time estimation using GPS equipped probe vehicle, unpublished document.

Lin, W-H., and Zeng, J., 1999, An experimental study on real time bus arrival time prediction with GPS data, *Transportation Research Record* 1666, Transportation Research Board.

Lobo, A. X., 1997, Automatic vehicle location technology: application for buses, PhD thesis, University of London, UK.

Marshall, L. F., Levinson, H. S., Lennon, L. C., and Cheng, C., 1990, Bus service times and capacities in Manhattan, *Transportation Research Record* 1266, Transportation Research Board, pp.189-196.

McKinney, B. L., and Engfer, D. R., 2004, Formulating risk into research and engineering projects, *Proceedings of the 2004 Crystal Ball user conference.*

McKnight, C. E., et al., 2004, The impact of traffic congestion on bus travel time in Northern New Jersey, *Transportation Research Board 2004 Annual Meeting* CD-ROM.

McKnight, C. E., and Paaswell, R. E., 1997, Impact of congestion on New York bus service, UTRC for MTA New York City Transit.

Meridian World Data, 2004, Distance calculation- how to calculate the distance between two points on the earth, Available from: http://www.meridianworlddata.com/distance-calculation.asp, Accessed on 4 June 2004.

Miwa, T., and Morikawa, T., 2003, Analysis on Route Choice Behavior Based on Probe-Car Data, *10th World Congress ITS* CD, 16-20, November, Madrid, Spain.

Montgomery, D. C., and Runger, G. C., 2003, Applied statistics and probability for engineers, third edition, *John Wiley & Sons*, Inc.

Moore, D. S., and McCabe, G. P.,2006, Introduction to the practice of statistics, fifth edition, *W. H. Freeman and Company*, New York.

Nelson, P., and Palacharla, P., 1993, A neural network model for data fusion in ADVANCE, *Pacific Rim Transportation Technology Conference Proceedings*, Vol. I, pp. 237–243

NovAtel, 2003, GPS position accuracy measures, Available from: http://www.novatel.com/ Documents/ Bulletins/apn029.pdf, Accessed on 11 August 2005.

Ogle, J., Guensler, R., Bachman, W., Koutsak, M., and Wolf, J., 2002, Accuracy of Global Positioning System for determining driver performance parameters, *Transportation Research Record* 1818, Transportation Research Board, pp.12-24.

Ordnance Survey, 2000a, A guide to coordinate systems in Great Britain, Available from: http://www.gps.gov.uk/guidecontents.asp, Accessed on 18 April 2005.

Ordnance Survey, 2000b, National GPS network information, Available from: http://www.gps.gov.uk/faq.asp#faq15, Accessed on 18 April 2005.

Patnaik, J., Chien, S., and Bladikas, A., 2004, Estimation of bus arrival times using APC data, *Journal of Public Transportation*, Vol. 7, No. 1, pp.1-20.

Peek traffic Ltd., 2004, SCOOT - Split Cycle Offset Optimisation Technology, Available from: http://www.peek-traffic.co.uk/data/PDF/sales/SCOOT.pdf, Accessed on 20 June 2005.

Petty, K. F., Bickel, P., Ostland, M., Rice, J. Schoenberg, F., Jiang, J. and Ritov, Y., 1998, Accurate estimation of travel times from single-loop detectors, *Transportation Research Part A*, Vol.32, Issue 1 , pp.1-17.

Puget Sound Regional Council, 2000, Travel Time Data Collection Using Global Position System Technology. BRW Inc., Texas Transportation Institute KDD and Associates Innovative Transportation Concepts.

Quiroga, C. A. and Bullock, D., 1998a, Travel time studies with global positioning system and geographic information systems: an integrated methodology, *Transportation Research Part C*, Vol. 6, Issue 1-2, pp.101-127.

Quiroga, C. A. and Bullock, D, 1998b, Determining of sample sizes for travel time studies, *ITE Journal*, August 1998, pp. 92-98.

Quiroga, C. A., and Bullock, D., 1999, Measuring control delay at signalized intersections, *Journal of Transportation Engineering*, July/August, pp.271-280.

Rajbhandari, R., Chien, S. I., and Daniel, J. R., 2003, Estimation of bus dwell times with automatic passenger counter information, *Transportation Research Record* 1841, pp.120-127.

Rhodes, J. A., 2005, Engineering estimates for environmental liabilities using Crystal Ball®, *Proceedings of the 2005 Crystal Ball User Conference.*

Robertson, H. D. et al., 1994, Travel time and delay studies, ITE Manual of Transportation Engineering Studies. Englewood Cliffs, N.J., USA: Prentice Hall, pp.52-68.

Robinson, S., 2004, Simulation: the practice of model development and use, *John Wiley & Sons.*

Robinson, S., and Polak, J., 2004, Some new perspectives on Urban Link Travel Time Models: Is the k-nearest neighbours approach the solution? *The University Transport Study Group* Conference, Newcastle Upon Tyne, January 2004.

Rubinstein, R. Y., 1981, Simulation and the Monte Carlo method, *John Wiley & Sons.*

Santos, G., 2000, Road pricing on the basis of congestion costs: consistent results from two historic UK towns, *the 79th Annual Meeting of the Transportation Research Board*, Washington DC, January 9-13.

Sargent, R. G., 1998, Verification and validation of simulation models, *Proceedings of the 1998 Winter Simulation Conference.*

Sen, A., Soot, S., Ligas, J., and Tian, X., 1997, Arterial travel time estimation: probes, detectors and assignment-type models, *Transportation Research Board 76th Annual Meeting*, 12-16, January.

Seneviratne, P. N., 1998, Simulation of fixed route bus travel time, *Journal of Advanced Transportation Engineering*, Vol.22, No.1, pp.39-53.

Shao, J.,1993, Linear model selection by cross-validation, *Journal of the American Statistical Association*, 88(422), pp.486-494.

Shao, J., and Tu, D., 1995, The jackknife and bootstrap, *Springer-Verlag*, New York.

Shbaklo, S., et al., 1992, Short-term travel time prediction, ADVANCED Project Report TRF-TT-01.

Shimizu, H., Kobayashi, M., and Yonezawa, Y., 2000, An analysis of mean link travel time for urban traffic networks, *IEEE 51st Vehicular Technology Conference Proceedings*, VTC Spring Tokyo.

Shrestha, B. P., 2002, Simulating advanced bus priority strategies at traffic signals. PhD thesis, University of Southampton, UK.

SIEMENS Traffic Controls Limited, 1999, SCOOT user guide, Issue.19, HF16940,UK.

Sisiopiku, V. P., and Rouphail, N. M., 1994, Toward the use of detector output for arterial link travel time estimation: a literature review. *Transportation Research Record* 1457, Transportation Research Board, pp.158-165.

Srinivasan, K. K. and Jovanis, P. P., 1996, Determination of number of probe vehicles required for reliable travel time measurement in urban network, *Transportation Research Record* 1537, Transportation Research Board, pp.15-22.

Sun, A., and Hickman, M., 2004, The holding problem at multiple holding stations, 9th *International Conference on Computer-Aided Scheduling of Public Transport*, August 9-11, California, USA.

Takaba, S., Morita, T., Hada, T., Yamaguchi, M., and Usami, T., 1991, Estimation and measurement of travel time by vehicle detectors and license plate readers, *Proceeding of Vehicle Navigation and Information Systems*, Vol.1, pp.257-267.

The Commission for Integrated Transport, 2000, Public attitudes to transport in England- a survey carried out by Market & Opinion Research International for the commission for integrated transport, Available from: http://www.cfit.gov.uk/reports/mori/index.htm, Accessed on 5 October 2004.

The Commission for Integrated Transport, 2001, European best practice in delivering integrated transport – key findings, Research reports.

Toledo, T., and Koutsopoulos, H. N., 2004, Statistical validation of traffic simulation models, *TRB 2004 Annual Meeting* CD-ROM, Transportation Research Board, National Research Council, Washington D.C.

Toppen, A., and Wunderlich, K., 2003, Travel time data collection for measurement of advanced traveler information systems accuracy, project 0900610-D1, *Federal Highway Administration*. Available from: http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13867.html, Accessed on 22 May 2005.

Transit Cooperative Research Program, 2000, Operational Analysis of Bus Lanes on Arterials: Application and Refinement, RRD-38, Transportation Research Board, National Research Council, Washington D.C.

Transit Cooperative Research Program, 2003a, Transit capacity and quality of service manual, Report 100, 2nd Edition, Transportation Research Board, National Research Council, Washington D.C.

Transit Cooperative Research Program, 2003b, A guidebook for developing a transit performance-measurement system, Report 88, Transportation Research Board, National Research Council, Washington D.C.

Transit Cooperative Research Program, 2003c, Bus rapid transit Vol.2: implementation guidelines, Report 90, Transportation Research Board, National Research Council, Washington D.C.

Transportation Research Board, 2000, Highway Capacity Manual 3rd Edition, Special report 209, Washington D.C.

Turner, S. M., 1996, Advanced techniques for travel time data collection, *Transportation Research Board 75th Annual Meeting*.

Turner, S. M., Eisele, W. L., Benz, R. J. and Holdener, D. J., 1998, Travel time data collection handbook, Texas Transportation Institute, FHWA U.S. DOT.

Velez-Pareja, I.,2006, Practical sensitivity analysis, Social Science Research Network, Available from: http://ssrn.com/abstract=881055, Accessed on 18 April 2006.

Vose, D., 2000, Risk analysis: a quantitative guide, second edition, *John Wiley and Sons*.

Wardrop, J. G., 1968, Journey speed and flow in central urban areas, *Traffic Engineering and Control*, (9), pp.528-539.

Waterson, B. J.,2005, Data fusion for travel time estimation in urban traffic networks, PhD thesis, University of Southampton, UK.

Westerman, M., 1994, Estimation of travel times for ATMS and ATIS, The University Transportation Study Group, Leeds.

Wormley, S. J., 2004, GPS errors and estimating your receiver's accuracy- what's the difference between repeatability and accuracy? Available from: http://www.edu-observatory.org/gps/gps_accuracy.html, Accessed on 23 October 2005.

Xie, C., Cheu, R. L., and Lee, D-H., 2001, Calibration-free arterial link speed estimation model using loop data. *Journal of Transportation Engineering*, ASCE, 1276, pp.507-514.

York, I. O., 1993, Factors affecting bus-stop times, TRL Project Report 2, T1/25, Transport Research Laboratory, Crowthorne, UK.

Zhang, H. M., 1999, Link-journey-speed model for arterial traffic, *Transportation Research Record* 1676, Transportation Research Board, pp.109-115.

Zhang, M. and He, J. C., 1998, Estimating arterial travel time using loop data, Public Policy Centre, University of Iowa.

Zhao, F. and Li, M-T, 2005, Calibration of highway /transit speed relationships for improved transit network modeling in FSUTMS, draft final report, Lehman Center for Transportation Research Department of Civil & Environmental Engineering, Florida International University, US.

Zhu, H., and Rohwer, R., 1996, No free lunch for cross-validation, *Neural Computation*, 8, pp.1421-1426.

Zito, R., D'este, G., and Taylor, A. P., 1995, Global Positioning systems in the time domain: How useful a tool for Intelligent Vehicle-Highway systems? *Transportation Research Part C*, Vol. 3, Issue 4, pp.193-209.

Zito, R., and Taylor, A. P., 1994, The use of GPS in travel-time surveys, *Traffic Engineering and Control*, Vol.35, pp.685-690