TRAVEL TIME ESTIMATION IN URBAN AREAS USING NEIGHBOUR LINKS DATA

by

MOHAMED ELESAWEY

B.Sc., Ain Shams University, 2002
M.Sc., Ain Shams University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2010

© Mohamed Elesawey, 2010
ABSTRACT

Travel time is a simple and robust network performance measure that is perceived and well understood by the public and politicians. However, travel time data collection can be costly especially if the analysis area is extensive. This thesis proposes a solution to the problem of limited network sensor coverage caused by insufficient sample size of probe vehicles or inadequate numbers of fixed sensors. The approach makes use of travel time correlation between nearby (neighbour) links to estimate travel times on links with no data using neighbour links travel time data. A framework is proposed that estimates link travel times using available data from neighbouring links. The proposed framework was validated using real-life data from the City of Vancouver, British Columbia. The travel time estimation accuracy was found comparable to the existing literature. The concept of neighbour links travel time estimation was extended and applied at a corridor level. Regression and Non-Parametric (NP) models were developed to estimate travel times of one corridor using data from another corridor. To analyze the impact of the probes’ sample size on the accuracy of the proposed methodology, a case study was undertaken using a VISSIM microsimulation model of downtown Vancouver. The simulation model was calibrated and validated using field traffic volumes and travel time data. The methodology provided reasonable estimation accuracy even using small probe samples. The use of bus travel time data to estimate automobile travel times of neighbour links was explored. The results showed that bus probes data on neighbour links can be useful for estimating link travel times in the absence of vehicle probes. The fusion of vehicle and bus probes data was analyzed. Using transit data for neighbour links travel time estimation was shown to improve the accuracy of estimation at low market penetration levels of passenger probes. However, the significance of transit probe data diminishes with the increase of market penetration level of probe vehicles. Overall, the results of this thesis demonstrate the feasibility of using neighbour links data as an additional source of information that might not have been extensively explored.
TABLE OF CONTENTS

ABSTRACT ............................................................................................................................. ii
TABLE OF CONTENTS ...................................................................................................... iii
LIST OF TABLES ................................................................................................................. vi
LIST OF FIGURES .............................................................................................................. vii
ACKNOWLEDGEMENTS ................................................................................................. viii
DEDICATION........................................................................................................................ ix

CHAPTER 1: INTRODUCTION .......................................................................................... 1
  1.1 BACKGROUND .......................................................................................................... 1
  1.2 STATEMENT OF THE PROBLEM ............................................................................ 2
  1.3 RESEARCH OBJECTIVES ......................................................................................... 4
  1.4 THESIS STRUCTURE ................................................................................................. 5

CHAPTER 2: TRAVEL TIME DATA COLLECTION AND ESTIMATION METHODS: A LITERATURE REVIEW ................................................................. 6
  2.1 INTRODUCTION ........................................................................................................ 6
  2.2 TRAVEL TIME DATA COLLECTION METHODS ................................................. 6
    2.2.1 Test Vehicles Techniques ...................................................................................... 7
    2.2.2 License Plate Matching Techniques ..................................................................... 7
    2.2.3 Emerging and Non-traditional Techniques .......................................................... 8
    2.2.4 ITS Probe Vehicles Techniques ............................................................................ 8
  2.3 TRAVEL TIME ESTIMATION/PREDICTION MODELS ...................................... 13
    2.3.1 Travel Time Estimation/Prediction Using Statistical Models ............................ 13
    2.3.2 Travel Time Estimation/Prediction Using AI Models ......................................... 18
    2.3.3 Travel Time Estimation/Prediction Using Simulation Models ........................... 19
    2.3.4 Hybrid Travel Time Estimation/Prediction Models ............................................ 19
  2.4 TRAVEL TIME CORRELATION ............................................................................. 22
  2.5 GPS/GIS TRAVEL TIME STUDIES ......................................................................... 24
  2.6 VEHICLES AS PROBES ........................................................................................... 27
  2.7 CELLULAR PHONES AS PROBES ......................................................................... 30
  2.8 OPERATIONAL DEPLOYMENTS AND FIELD TESTS OF VEHICLES AS PROBES SYSTEMS .......................................................................................................... 33
    2.8.1 ADVANCE Project .............................................................................................. 33
    2.8.2 Taxi-Floating Car Data (Taxi-FCD) ................................................................. 33
    2.8.3 Floating Vehicle Data (FVD™) ......................................................................... 34
    2.8.4 OPTIS .................................................................................................................. 35
    2.8.5 IPCar System .................................................................................................. 36
    2.8.6 Singapore Taxis ................................................................................................. 37
CHAPTER 3: DATA DESCRIPTION .................................................................................. 40

3.1 THE STUDY AREA .................................................................................................. 40
3.2 REAL-LIFE DATA .................................................................................................. 41
   3.2.1 Survey Route and Method ......................................................................... 42
   3.2.2 Data Collection Program Design ............................................................. 43
   3.2.3 Sample Size ............................................................................................... 46
   3.2.4 Survey Implementation ............................................................................. 47
   3.2.5 Travel Time Data Analysis ....................................................................... 48
3.3 DOWNTOWN MICROSIMULATION MODEL ...................................................... 52
   3.3.1 VISSIM Background ................................................................................ 52
   3.3.2 Network Coding ....................................................................................... 52
   3.3.3 Model Update ........................................................................................... 55
   3.3.4 The Dynamic-based Simulation Model ..................................................... 58
   3.3.5 Calibration and Validation of the Microsimulation Model ......................... 59

CHAPTER 4: A PROOF OF CONECEPT USING REAL-LIFE DATA ....................... 78

4.1 INTRODUCTION ...................................................................................................... 78
4.2 A FRAMEWORK FOR NEIGHBOUR LINKS TRAVEL TIME ESTIMATION... 79
   4.2.1 Identification of Link Neighbours ............................................................. 80
   4.2.2 Choice of a Modelling Technique ............................................................. 81
   4.2.3 Application Entity .................................................................................... 91
   4.2.4 Source of Neighbour Links Travel Time Data ......................................... 92
4.3 CASE STUDY: DOWNTOWN VANCOUVER ....................................................... 92
   4.3.1 Data Description ....................................................................................... 92
   4.3.2 The Statistical Method ............................................................................. 93
   4.3.3 AI Methods ............................................................................................... 98
   4.3.4 Neighbour Links Travel Time Estimation: A Discussion ......................... 102
4.4 NEIGHBOUR CORRIDORS TRAVEL TIME ESTIMATION .............................. 103
   4.4.1 Analysis Corridors .................................................................................... 103
   4.4.2 Travel Time Distribution ......................................................................... 105
   4.4.3 Neighbour Corridor Travel Time Association ......................................... 107
   4.4.4 Description of the Used Models ............................................................... 108
   4.4.5 Error Measurements ............................................................................... 111
   4.4.6 Results ..................................................................................................... 112
   4.4.7 Models Comparison ................................................................................ 113
   4.4.8 Corridor Travel Time Estimation: A Discussion ..................................... 114

CHAPTER 5: USING SAMPLES OF PROBE VEHICLES FOR NEIGHBOUR LINKS TRAVEL TIME ESTIMATION ........................................................................... 116

5.1 INTRODUCTION ................................................................................................. 116
### LIST OF TABLES

Table 3.1 Required Sample Size at Different Confidence Intervals ......................................................... 47  
Table 3.2 Individual Route Travel Times (Seconds) .................................................................................. 49  
Table 3.3 Descriptive Statistics of Average Route Travel Times (Seconds) ........................................... 49  
Table 3.4 Descriptive Statistics of Average Link Travel Times (Seconds) ............................................... 51  
Table 3.5 Downtown HOV Lanes ........................................................................................................... 58  
Table 3.6 Summary of Route Choice Calibration Statistics ....................................................................... 73  
Table 3.7 Simulation and Real-Life Travel Times of the Calibration Corridors ......................................... 73  
Table 3.8 Results of Driver Behaviour Parameters Calibration .................................................................. 76  
Table 3.9 Results of Model Validation .................................................................................................... 76  
Table 4.1 Example Calculation of the Proposed Method for One Data Record ........................................ 97  
Table 4.2 Number of Neighbours Identified by Neuro-Fuzzy Models ...................................................... 100  
Table 4.3 Examples of the Rules Generated by the Neuro-Fuzzy Model for Link 2 ......................... 100  
Table 4.4 Results of Goodness of Fit Statistical Tests ............................................................................... 106  
Table 4.5 Fitted Statistical Distributions of Observed Travel Times ....................................................... 106  
Table 4.6 Results of ANOVA .................................................................................................................... 108  
Table 4.7 Parameters and Goodness of Fit Statistics of the Statistical Models ....................................... 112  
Table 4.8 Optimal $K$ Value and MSE of Each Weighting Method ......................................................... 113  
Table 6.1 Summary Statistics of Travel Times of the Analyzed Segments .............................................. 137  
Table 6.2 Defined Neighbours for Transit Sections .................................................................................. 140  
Table 7.1 Identified Neighbours for the Analyzed Segments ................................................................... 149  
Table 7.2 Bus/Passenger Probes Neighbourhood Exponential Models .................................................... 150
LIST OF FIGURES

Figure 3.1 The Proposed Study Area ................................................................. 41
Figure 3.2 Travel Time Survey Route ................................................................. 43
Figure 3.3 Logic behind the Developed Program ............................................... 45
Figure 3.4 Travel Time Sections ID (Real-Life Data) .......................................... 50
Figure 3.5 Dummy Links for Bus Routes Ending at Midblock ......................... 56
Figure 3.6 Movement Restrictions in Downtown Vancouver ............................ 57
Figure 3.7 The Proposed Calibration and Validation Procedure ....................... 62
Figure 3.8 Scatter Plots of Observed Volumes vs. Simulation Volumes .......... 72
Figure 4.1 B-Spline Fuzzy Membership Functions (Sayed and Razavi 2000) .... 88
Figure 4.2 Number of Neighbours for each Travel Time Section .................... 93
Figure 4.3 Error Measurements of Different Weighting Schemes ................. 95
Figure 4.4 Identified Number of Neighbours at Different Cut-off Values ....... 97
Figure 4.5 Estimation MAPE at Different Cut-off Values ............................... 98
Figure 4.6 Error Measurements of All Models ................................................. 101
Figure 4.7 Analysis Corridors ........................................................................... 104
Figure 4.8 Fitted Normal and Lognormal Corridor Travel Time Distributions .... 107
Figure 4.9 Mean Square Errors (MSE) of Different K Values and Weighting Methods 113
Figure 4.10 MSE of Different Models .............................................................. 114
Figure 4.11 MAPE and RRSE of Different Models .......................................... 114
Figure 5.1 Travel Time Sections ID (Simulation Data) ...................................... 122
Figure 5.2 MAPE of the Three Data Fusion Methods ..................................... 125
Figure 5.3 Probe-Neighbours Coverage at Different Market Penetration Levels 126
Figure 5.4 Calculated $\alpha$ for Link 28 ............................................................ 128
Figure 5.5 Estimated vs. True Link Travel Times (Variance Weighting) ......... 129
Figure 5.6 Analyzed Transit Sections ID .......................................................... 137
Figure 5.7 MAPE of the Three Methods of Data Fusion ............................... 141
Figure 5.8 Calculated $\alpha$ for the Four Sections ............................................. 142
Figure 5.9 Estimated vs. True Link Travel Times (Variance Weighting) ......... 144
Figure 7.1 Data Fusion Estimation Accuracy .................................................. 151
Figure 7.2 Fusion Accuracy of Using Passenger Probes only vs. Passenger Probes and Buses 152
Figure 7.3 Estimated vs. True Link Travel Times (Variance Weighting) ......... 154
Figure 8.1 Regional Congestion Indices in Metro Vancouver .......................... 160
ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my research advisor, Professor Tarek Sayed, from whom I have learned a lot during my five years of research at UBC. His continuous support, inspiration, and kindness have been undeniable. The constructive critique and constant availability of Professor Sayed helped me move smoothly in my research. His invaluable comments have contributed significantly to the improvement of my research and academic skills. I am also thankful for his constant consideration and support, especially when I was low in spirit. To me, he is not only a research advisor, but also a mentor and a big brother. Without the guidance of Professor Sayed, this dissertation would never have been completed. I will always feel fortunate that I had him as my Ph.D. advisor.

I offer my enduring gratefulness to my fellow students at UBC. I owe particular thanks to Dr. M. Wahba for his positive feedback and valuable suggestions on this thesis. My special thanks go to my dear friend, Karim Ismail, for his help and encouragement. I would also like to thank Clark Lim for his valuable discussions and priceless suggestions. Special thanks to Shweekar Ibrahim for her help, encouragement, and support.

I feel blessed to have a lot of close friends both in Egypt and Canada who supported and encouraged me during the years of my research. Thanks Ahmed Manzour, Sherif Ahmed, Sherif Essam and Sherif Moustafa for your remote support and true friendship.

I will never forget the joy, laughter, and happy moments that I had with Karim El-Basyouny, Mohamed Ammar, Amr Abo Elenein, Yehia Madkour, Ali Omran, Samer El Housseny, and Wael Ekeila.

I would also like to thank my wife, Heidi Gmeay, who has always been by my side. I want to thank my parents-in-law, who did their best to help me concentrate on my research.

Finally, the real appreciation and thanks are to the ones who maybe were far away from me, but yet, closest with their feelings and hearts. Special thanks are owed to my parents and my sisters, who have supported me throughout my years of education, both morally and financially.
DEDICATION

To Heidi, Zein, and Seif
CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Knowledge of the travel time over a transportation network is important for both transportation analysts and network users. From an analyst’s perspective, planners use travel time information as a basic input for many applications such as the calibration of travel demand forecasting models, the calibration of air quality emission models, and the ranking and prioritizing of major investments. As well, traffic professionals use travel time data to monitor network performance, provide route guidance, disseminate traveller information, observe congestion progression, and detect incident occurrence. From the road users’ viewpoint, travel time information is essential to better schedule their trips either by changing planned routes or by changing the trip time. Furthermore, commercial fleets benefit from accurate travel time information through better fleet management, and accordingly, improved service. Travel time is also a simple and robust Measure Of Effectiveness (MOE) that is perceived and well understood by both the public and politicians.

Travel time information is a major determinant for the success of any Intelligent Transportation Systems (ITS) application. ITS includes any system that improves the travel efficiency of people and goods by employing modern technology. Improving travel efficiency is meant to make the trip safer, shorter, cheaper, and more convenient. Travel time estimation on urban roads is more challenging than on freeways. Urban travel time estimation is not only complex because of the interrupted flow nature caused by control devices (e.g. signals), but also due to the presence of turning movements and shared lanes. Also, the existence of transit stops, on-street parking and commercial access points cause travel times in urban areas to fluctuate and hence more difficult to estimate/predict.

It is important to distinguish between two types of travel times: static and dynamic. Static travel time refers to average (e.g. historical) travel time during a specific period of the day. This is a simple concept and does not involve much effort for real-time data collection. It
may be adequate for road environments where traffic does not incur major or dynamic changes and is usually in the free flow state. In more dynamic environments, however, the approach may not be suitable due to considerable and rapid changes in traffic conditions. Dynamic travel time, on the other hand, considers the fact that travel times on a road segment can be influenced by many factors such as traffic flow, time of day, weather conditions, etc. Estimation of travel time dynamically involves the use of automated systems to continuously monitor the traffic and obtain new values for travel times. In fact, future predictions of traffic patterns and travel times are the most important determinants. Smith et al. (2002), suggests that “ITS must have a predictive capability.” There will always be a time lag between obtaining traffic information and trip time. Due to the rapid changes in traffic conditions in urban networks, traffic conditions do not usually remain constant over the following ten to twenty minute interval.

1.2 STATEMENT OF THE PROBLEM

Traffic professionals and engineers have been developing approaches to reduce congestion. One of these approaches is to provide travellers with traffic information, such as travel time, so that they can better manage their trips. Delivering updated real-time traffic information to travellers is the main objective of Advanced Traveller Information Systems (ATIS). Nevertheless, ATIS require data collection from one or more sources. Data collection costs can be very high, especially if the analysis area is extensive. Furthermore, costs to estimate travel times dynamically (i.e. during different periods of the day), are much higher. The challenge is to achieve an acceptable level of data collection accuracy within budget.

Network data collection coverage is a key determinant to the success of any ATIS. Transportation professionals are usually constrained by a limited number of network sensors that provide partial network coverage. Network coverage depends on sensor technology that is used for data collection. Static coverage is provided wherever fixed sensors such as Automatic Vehicle Identification (AVI) are used. If the traffic data are collected from moving sensors, such as Global Positioning System (GPS)-equipped probes, the coverage may change over time and hence is described as “dynamic.” In the
absence of coverage on some network links during some intervals (i.e. limited data), the common practice is to disseminate traffic information of covered locations only, or to use historical average data to compensate for missing data.

The focus of this research is on travel time estimation in urban networks with partial sensor coverage. More specifically, this research addresses a major issue associated with using Vehicles As Probes (VAP) for travel time data collection; that is, estimation of travel times on “untravelled” links during a measurement interval. This issue corresponds to the required sample size of probe vehicles. Probe vehicles can be commercial fleet vehicles, taxis, buses or any other type of vehicle not primarily used for traffic data collection. Rather, they move on a road network to serve a particular purpose and, hence, can be tracked in real-time to collect traffic information. Many studies have investigated the proper sample size of probe vehicles required for travel time estimation. It is difficult, however, in most cases to achieve the minimum sample size, especially if the road network is extensive and the data is aggregated at a high resolution (i.e. small data collection intervals).

This research proposes a solution to the problem of limited network sensor coverage caused by insufficient sample size of probe vehicles or an inadequate number of fixed sensors. The objective is to help fill in data gaps within a real-time traveller information system. In most existing traveller information systems, historical data are traditionally used in the absence of real-time information. The current research proposes another approach that can be used when there is an absence of real-time travel times on part of a road network because of limited sensor coverage. The approach makes use of travel time correlation between nearby (i.e. neighbour) links to estimate travel times on links with no data using neighbour links travel time data. Empirical and theoretical findings on the correlation between travel times of nearby links have been reported and discussed extensively in the literature (Hall 1986, Sen et al. 1997, Fu and Rilett 1998, Rilett and Park 1999, Chen and Chien 2001, Eisele and Rilett 2002, He et al. 2002, Gajewski and Rilett 2003). In general, travel time correlation between nearby links can be attributed to: (a) correlation in traffic demand, (b) similarity in traffic control and (c) queue spillback. The argument has always been that the correlation/covariance should be considered when
calculating route travel time from individual (i.e. link) travel times. Nevertheless, the potential for using travel time correlation for travel time estimation on segments with no data in urban networks has received little attention to date.

In this research, the concept of travel time “neighbourhood” is introduced. The term “neighbours” refers to nearby links that have similar characteristics and are subject to similar traffic conditions in a network. These nearby segments can be the preceding and succeeding segments on a route, parallel segments, or even intersecting segments.

Although this research focuses on the use of probe vehicles that can be tracked in real-time using GPS receivers or cellular phones, the proposed methodology is generic and can be applied to any other data collection method that uses fixed sensors (e.g. license plate matching, AVI, etc). It should be noted that the proposed method is not intended to replace other existing travel time estimation models. Rather, it is a supplementary and complementary model that can be integrated with almost all existing travel time estimation models through data fusion. The objective of the analysis is to demonstrate the feasibility of using neighbour links data as an additional source of traffic information that has not been extensively explored before.

1.3 RESEARCH OBJECTIVES

The primary objective of this research is to develop a general framework for the estimation of travel times on a road network using sparse travel time data. Particularly, to develop a framework to estimate travel times on links with no real-time data using travel time relationships between these links and their neighbour links. To achieve this primary research objective, the following secondary objectives need to be fulfilled:

− To validate the framework using real-life data,
− To investigate the use of various modelling techniques within the framework,
− To study the feasibility of using corridor travel time data to estimate travel times of a nearby corridor,
− To investigate the impact of probe vehicles’ sample size on the estimation accuracy of the proposed framework,
To explore the potential for using buses as probes for neighbour links travel time estimation, and,

To develop and compare data fusion methods that can be used to fuse historical data and real-time data of neighbour links.

1.4 THESIS STRUCTURE

The first chapter of this thesis provides an introduction that includes a background, a statement of the problem, and research objectives. Chapter two includes a literature review of some important topics that are related to the research area. In chapter three, the two sets of data which were used throughout this research are described. The proposed framework for neighbour links travel time estimation is presented in chapter four. In addition, chapter four describes the use of real-life data to validate the framework. The concept of neighbour corridors travel time estimation is also illustrated in chapter four. Application of the proposed framework using samples of simulated probes is described in chapter five. The potential for using buses as probes for neighbour links travel time estimation is demonstrated in chapter six. Chapter seven introduces data fusion methods that were used to combine link historical data, bus neighbour data, and passenger probes neighbour data to improve the accuracy of the estimation. Finally, chapter eight summarizes the research hypothesis, results and conclusions.
CHAPTER 2: TRAVEL TIME DATA COLLECTION AND ESTIMATION METHODS: A LITERATURE REVIEW

2.1 INTRODUCTION

The focus of this literature review is on the following topics: travel time data collection methods, travel time estimation/prediction models, travel time correlation, GPS/GIS travel time studies, and real-life attempts to use probes for data collection. First, a classification of different techniques to collect travel time data is presented. A review of different modelling approaches for travel time estimation/prediction is subsequently presented. The literature of travel time correlation is then discussed. Previous research studies which have integrated GPS and GIS to collect and estimate travel time data are introduced. Finally, a number of real-life deployments and field tests of “vehicle as probes” systems are presented.

2.2 TRAVEL TIME DATA COLLECTION METHODS

Many classifications can be used to categorize travel time data collection techniques. Quiroga and Bullock (1998), suggested that these techniques can be classified as roadside techniques and vehicle techniques. Roadside techniques include license plate matching and AVI, while vehicle techniques include the floating car method (May 1990), Automatic Vehicle Location (AVL), and cellular phone techniques. The Travel Time Data Collection Handbook (Turner et al. 1998) provides a comprehensive illustration of different methods used for travel time data collection. This handbook does not constitute an industry standard, but it was rather meant to provide guidance to transportation professionals and practitioners regarding travel time data collection, reduction, and reporting. The handbook categorized the methods used for travel time data collection into four groups; test vehicle techniques, license plate matching techniques, emerging and non-traditional techniques, and ITS probe vehicle techniques. The four groups are described in detail in the handbook, along with their advantages and disadvantages, guidelines for implementation, required
sample sizes, etc. The following is a brief background on each of the four methods with emphasis on ITS probe vehicle techniques.

### 2.2.1 Test Vehicles Techniques

Test Vehicles techniques have been used for many years to collect travel time data. Using these techniques, a vehicle is instrumented and dispatched into the traffic stream and the driver is asked to drive in one of three driving styles: average car, floating car, or maximum car. In average car, the test vehicle should be driven at a speed similar to the average traffic speed. In floating car, the driver is asked to safely pass as many vehicles as the test vehicle passed. In maximum car, the vehicle is driven at the posted speed limit unless impeded by the traffic stream. Collecting travel time data can be carried out by one of several methods based on the instrumentation level of the test vehicle. In one method, travel times are manually recorded using a stopwatch at predefined checkpoints by an observer. Another method utilizes an electronic Distance Measurement Instrument (DMI) which is connected to the transmission of the test vehicle to estimate vehicle speed and traversed distance. Travel time is computed by dividing the recorded distance by the speed. A third method employs the use of an on-board GPS receiver by which the location of the test vehicle can be determined at each time step, and hence, travel times can be estimated. This method of test vehicle is also known as the floating car technique or active vehicle technique. The term “active” refers to the fact that the main purpose of these vehicles is to collect travel time data. This method employs a number of test runs to estimate travel times at different times of the day.

### 2.2.2 License Plate Matching Techniques

In these techniques, the license plate characters of each vehicle are recorded at a number of checkpoints with their associated arrival times. The travel time is then computed as the difference between the arrival times of that license plate at two checkpoints. This method is straightforward and robust. A number of methods can be used to record the license plate characters. The simplest method is to record the data manually whether by using pen and paper, or a tape recorder. A portable computer can also be used by an observer to enter the
information, and hence, assist in the process of matching. A third method is to use video cameras to capture the vehicle’s license plate and then have an observer manually enter the data into a computer. Finally, the most advanced method is to use video cameras with computer vision algorithms. In this case, license plate characters are defined and automatically transferred to a computer and there is no need for an observer. These methods have many limitations which restrict their use. Most importantly, here, is the issue of limited coverage. The travel time data collection is limited to locations where observers or cameras can be positioned. Even if many cameras and observers are available, the accuracy expectations will still be low due to potential errors, especially when there are high volumes of traffic.

2.2.3 Emerging and Non-traditional Techniques

Unlike other methods, these techniques are described as “emerging” since they are still under research and development. Moreover, they are described as “non-traditional” since they do not measure travel time directly as in the case of traditional methods. Some of these techniques involve using point measurement equipment, such as inductive loop detectors, infrared sensors, ultrasonic detectors, magnetic sensors, or video cameras. Traffic parameters, such as spot speeds, occupancy, vehicle headway, and volumes are measured and used to estimate the unknown travel times. Pattern recognition is also utilized by developing image matching algorithms to match vehicle images captured at two consecutive observation points and to obtain corresponding travel times.

2.2.4 ITS Probe Vehicles Techniques

One recent approach for real-time travel time data collection is the use of vehicles as probes or “moving sensors” to collect traffic information directly from the traffic stream. Probe vehicles are not launched through the network for the purpose of traffic data collection, and therefore, are regarded as passive vehicles. These techniques employ the use of instrumented vehicles moving in the traffic stream and a communication mode between the vehicles and the Transportation Management Centre (TMC). The communication between the vehicle and the TMC can be either through wide area
communication coverage that connects vehicles directly to the centre, or by a dedicated short range communication link between vehicles and a roadside infrastructure which identifies the vehicle and sends the data to the centre. Vehicles floating in the traffic stream can provide useful information on links traversed, such as link speed and travel time. This information is then sent online to a TMC for further cleaning, reduction, and analysis. Specialized programs are usually designed and utilized for this purpose. Many technologies can be used to collect data from probe vehicles. The two most common techniques are Automatic Vehicle Identification (AVI) and Automatic Vehicle Location (AVL).

**AVI Tracking Technology**

The primary application of AVI systems is electronic toll collection on freeways and rural highways. However, their use can be extended to incident detection, traffic monitoring, and travel time data collection from probe vehicles. An AVI system utilizes an in-vehicle tag and a roadside beacon connected to a data reader via a coaxial cable. A transponder, or a tag, is an onboard device that transmits a signal with a unique vehicle code for each vehicle type to the roadside receiver (beacon). The roadside beacon transfers this information to the reader via the coaxial cable. Current advanced reader types integrate both the receiver and the reader in one unit with no need for cable connections. The reader stores the information with an assigned date and time stamp. The received information can be used for online applications such as transit signal priority, access control at gateways, emergency vehicle pre-emption at a signalized intersection, vehicle clearance at a toll plaza, or an inspection station on freeways. It can also be sent in real-time to a TMC where it can be processed. Additionally, the stored information can be used for offline analysis such as studying historical trends and travel time variability. In general, AVI systems are used for freeways where radio waves transfer freely through the air without obstructions. Noteworthy is that the use of this system in urban environments can be problematic. Most importantly, there is the potential problem of devoting a specific radio wave length for a system that must not interfere with other waves (i.e. radio channels). Also, there could be the potential for obstructed signals in areas with large buildings.
**AVL Tracking Technologies**

AVL systems are automated tracking systems. Due to the continuous enhancement in tracking technologies, coupled with the decline in hardware prices, these systems are gaining more popularity and are now often deployed. Tracking technologies used in AVL systems include dead reckoning and map-matching, signpost, ground-based radio navigation, Global Positioning Systems (GPS), and cellular phones. Current practice involves the use of GPS and cellular phones mature, widely-established, and well understood technologies. A brief description of each method is presented in the following subsections.

**Dead-Reckoning and Map-Matching**

This technique is categorized by a vehicle equipped with a digital compass and wheel sensors which are used to measure the vehicle’s heading and distance from a previous point. By using the previous position, the heading angle, and the distance travelled, the system is able to compute the vehicle’s new position. This system is prone to cumulative errors since positioning errors grow as the distance travelled increases. Accordingly, map matching is used to correct for these errors and vehicles are positioned to the road network on a digital map. Generally, these systems lack the appropriate positional accuracy and because of this, have become obsolete.

**Signpost**

This system is usually used by transit agencies to track and monitor their fleets. A number of signpost transmitters are installed along the transit route. Each signpost transmitter emits a unique code via a low power signal that can be captured by a receiver on an approaching bus. The equipped buses have on-board “vehicle location units” which assign a time stamp to the received information and send it periodically via a radio transmitter to a central computer. Identifying each pair of signpost IDs and time stamps makes it easy to obtain a vehicle’s speed and travel time. Another potential method of data collection is to include the odometer reading in the data sent to the central computer. This enables travel time and speed data collection at short time intervals rather than having to wait for the bus to approach a new signpost. This system has many disadvantages which include a low
coverage area, non-representative data, and costly maintenance. Many transit agencies are
moving toward upgrading their AVL systems by using GPS-equipped vehicles as opposed
to this outdated signpost technology.

Ground-Based Radio Navigation

This system is also known as terrestrial radio navigation or radio triangulation. In this
system, a number of antenna towers are set up in the data collection vicinity. Meanwhile,
vehicles are equipped with transponders, each with its own unique ID number. While
moving in the traffic stream, an equipped-vehicle broadcasts a radio frequency (RF) signal
that can be received by all nearby antennae. By measuring the time that the signal takes to
travel to the antenna, the distance between the vehicle and the antenna can be calculated. If
the signal sent by the vehicle is received by four or more antennae, the location of a vehicle
can be precisely estimated by triangulation. One disadvantage of radio-navigation is that
RF signals are usually obstructed by tall buildings and tunnels.

Global Positioning System (GPS)

A Global Positioning System (GPS), is an automated tool used to obtain positional
information anywhere in the world and at any time during the day. The GPS was originally
developed by the US Department of Defence (DoD) for military purposes and has been in
operation since 1995. The number of civilian users of GPS systems nowadays, however,
exceeds military users. The system is free of charge and only requires the availability of a
GPS receiver to receive satellite signals. The system uses a network of satellites, 21 active
and 3 spare, which are continuously orbiting the earth. It has the ability to locate any
subject on earth by triangulation with a reasonable accuracy which depends on many
factors, such as the number of visible satellites, the receiver’s accuracy, and location.
When the system was first introduced, the DoD degraded the accuracy of the GPS for fear
of being exploited by potential enemies attacking American targets. This action was a
denial policy known as the Selective Ability (SA) degradation. Accordingly, civilian users
were able to obtain their positions with an accuracy of no more than 100 m at a confidence
level of 95%. On May 1, 2000, US president, Bill Clinton ordered the SA degradation to
be removed at midnight with no additional cost to users. Removal of the SA was a positive
step towards enhancing the positional accuracy of the GPS. The accuracy of current GPS receivers ranges from 10 to 15 meters (Byon 2005). As the GPS depends, in principle, on the line of sight between the receiver and the satellites, signals can sometimes be lost underground or in urban canyons where tall buildings exist. In areas where this problem might exist, the GPS receiver is often integrated with an odometer and/or compass headings to extrapolate position information from the last GPS reading. A signpost can also be placed where signal loss is expected to take place.

A GPS-equipped vehicle can determine its position using signals received from four satellites. This information can then be sent to a central computer every short time segment (sometimes known as a “hit”). The length of a hit is dependent on several factors such as the accuracy required, network characteristics, and traffic conditions. By using position information, the trajectory of any vehicle can be tracked on a GIS map and used to obtain a vehicle’s speed profile and travel times.

**Cellular Phones**

The use of cellular phones has recently emerged as an attractive option for tracking purposes. This is due to the increasing number of cellular phone users around the world, coupled with recent technological enhancements. The approach offers a number of advantages as presented by Ygnace et al. (2000). First, the system has low implementation costs as it utilizes the existing infrastructure of the cellular phone network. Second, cellular systems have a built-in spectrum allocation (i.e. dedicated/licensed wave frequency range to operate within). Third, there are currently millions of cellular phone subscribers and active users (i.e. high market penetration levels). Finally, the ability of cellular systems to provide two-way communications links is key. This approach suffers from a number of problems which includes confidentiality and privacy issues and low positioning accuracy. The last concern is associated with three inherent problems. First, the difficulty in assigning a vehicle to an adjacent link, especially in compacted areas; second, determination of an incorrect travel direction, and finally, difficulties differentiating between cellular phones in vehicles moving on a road network versus other cellular phones (e.g. with pedestrians and/or inside buildings, etc). Another major issue is that using cellular phones affects driver’s behaviour. It cannot be assumed, therefore, that the
collected travel information is representative of the traffic stream. Nevertheless, in the past few years there have been many successful real-life deployments using cellular phones as probes for real-time traffic data collection.

2.3 TRAVEL TIME ESTIMATION/PREDICTION MODELS

Generally, available techniques for travel time estimation can be categorized according to their computational procedures as statistical/mathematical models, Artificial Intelligence (AI) models, and simulation models. Other hybrid approaches incorporate the use of two or more of the above techniques. Statistical models include, for example, Kalman Filtering (Chu et al. 2005), Bayes Analysis (Gajewski and Rilett 2003) and Non-Parametric models (Robinson and Polak 2005, 2007). Examples of Artificial Intelligence (AI) techniques include Artificial Neural Networks and Neuro-Fuzzy models (Palacharla and Nelson 1999). Online simulation has been recently researched for travel time estimation purposes (Walhe et al. 2001). Most of the previous methods have proven to provide satisfactory performance in linear environments, such as freeways. Only a few of them were shown to perform well in more complex environments, such as urban networks (ITS Orange book 2004). The following is a review of a number of studies that used some of these models for travel time estimation/prediction. These studies are classified according to the modelling approach (i.e. statistical, AI, simulation, and hybrid). For each of these studies, the type of data and input variables are also explained.

2.3.1 Travel Time Estimation/Prediction Using Statistical Models

Van Aerde et al. (1993) presented two models to compute link travel time: an analytical model and a simulation model. The analytical model used data from simulated probe vehicles, representing a market penetration of 20% of the total network volumes. This method was based on statistical analysis to compute mean link travel time and standard deviation and hence constitute confidence intervals for the obtained travel times. In the second method, mean travel time and standard deviation were computed using data from all vehicles in the simulation model. The analysis showed that both methods provided similar results.
Chen and Chien (2001) developed a Kalman filter model for travel time prediction. Kalman Filter was chosen because of its ability to update the state variable, travel time, whenever a new observation was available. It utilizes the use of two clues for travel time prediction, historical data and real-time data. Historical data can be average travel time for the same period during a previous day(s) or even for the same day during (a) previous week(s) or average travel time from the previous time period ($t-1$). The developed filter is computationally easy and robust. However, it has one unresolved issue - noise calibration. To test their model, CORSIM, a microsimulation software package, was used to simulate a freeway segment and generate travel time data from a sample of probe vehicles that represented 1% of all vehicles in the network. Two approaches were introduced for travel time prediction; a link-based and a path-based. In the path-based approach, path travel time is the recorded travel time for all vehicles completing the whole path. For the link-based approach, the authors presented two methods. The first was the simple addition of predicted travel times for all links during a time period, $t$, which may exceed the value of $t$ itself. The second method was a progressive one. With this method, predicted travel times are summed until the sum becomes larger than $t$ and then the addition procedure moves to the next time interval to obtain the predicted travel times for the remaining links, until the sums falls in the next time period and the destination is reached. Error measurements were used to compare the predictions of the three approaches: the path-based approach and the two link-based approaches. The results showed that the path-based approach slightly outperformed the link-based approach.

Chien and Kuchipudi (2003) combined both the link-based and the path-based approaches using a hybrid model. The new model refined the predicted travel times by choosing the more accurate prediction of the link-based and the path-based models. A measure of central tendency was used to filter the data. The mean and the variance of the recorded travel times during a certain time period were computed, and records representing the shortest and the longest 15% vehicle travel times were discarded. The root mean square error was used to compare the three models; the link-based, the path-based, and the hybrid. The model that showed the lowest prediction error during a specific time period was assigned to predict travel time for this period.
Chu et al. (2005) developed an adaptive Kalman filter to estimate travel times by fusing probe vehicles and loop detectors data. The algorithm used vehicle density as the state variable and travel time as the measurement variable. Noise terms were calibrated online using a simple empirical method that had the ability to handle both systematic and random errors in noise sequences. Both the measurement error and the state noise were assumed to be Gaussian. This method was compared to using only probe vehicles or only loop detectors data for travel time estimation. Simulated data were generated for two different scenarios; recurrent congestion and incident occurrence. The results showed that data fusion through the newly developed algorithm was capable of providing better travel time estimates than the other two methods. However, the significance of this method declined as the probe penetration increased.

Travel time prediction for transit vehicles has also been a topic of interest for many researchers. Although the nature of transit travel time differs from that of automobile travel time, the same prediction techniques can still be applied. For example, Shalaby and Farhan (2004) used Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) data to develop a bus arrival prediction model using Kalman Filtering. Their model comprised two separate simplified Kalman filters, one to predict bus running time and the other to predict transit vehicle dwelling time. For any Kalman filter, two sets of data are required: historical and real-time. The first algorithm, bus running time, used the travel time data of the last three days and the previous bus travel time data as historical and real-time data, respectively. The dwelling algorithm was similar to the bus running time algorithm, with the only difference being the use of passenger arrival rates rather than running times. A comparison was undertaken between the developed bus arrival prediction model and three previously developed models: historical average, regression, and Artificial Neural Network (ANN). The results showed that the Kalman filter model outperformed all the three models in terms of prediction accuracy.

Non-parametric (NP) statistical models have also been used broadly in the literature to estimate travel time. Robinson and Polak (2005) developed a K- Nearest Neighbours (KNN) approach to estimate urban link travel times using inductive loop detector data. The key assumptions of the approach were presented followed by the input parameters of the
KNN method. The authors defined four input parameters that needed to be optimized when using a KNN method: the number of attributes to be included in the feature vector (similarity attributes), the distance metric, which is a measure of closeness between the input and the historical observations, the number of nearest neighbours (K) to be used in determining the output value, and finally, the function that relates the KNN records to the output value. Observed real-life data of 37 days were used to optimize the values of the four parameters. The feature vector included the 15-minute flow and the occupancy measured at three detectors. For the distance metric, four distances were tested. Four estimation measures were also examined: mean, median, regression, and the Locally Weighted Scatter plot Smoothing (Lowess) model. The value of K was varied along 14 levels starting from 20 to 5000. After a detailed sensitivity analysis, the best local estimation method was found to be the Lowess model, and the optimum value of K was 2160. The authors compared the proposed optimized KNN approach and other approaches, such as the KNN median, linear regression, and the ANN, using median, mean, and historical data that belonged to the same day of the week; using historical data that belonged to the same 15-minute period; and using the historical data for the 15-minute period for the same day of the week. In terms of the Mean Absolute Percentage Error (MAPE), the KNN median model outperformed all other models followed by the KNN Lowess model. On the other hand, in terms of the Root Mean Square Error (RMSE), the KNN Lowess model outperformed all other models, followed by the regression, the ANN, and the KNN median models, respectively. The authors came to a number of important conclusions which included: 1) the KNN approach is not sensitive to the distance metric used, 2) a robust local estimation method is needed such as Lowess or median, and, 3) the optimal value of K depends on the size of the historical data set used.

Robinson and Polak (2007) extended their previous work by using the KNN approach to characterize Travel Time Variability (TTV). They showed that the TTV can be disaggregated into the sum of three components: day to day variability, period to period variability, and vehicle to vehicle variability. The three components were computed using both Automatic Number Plate Recognition (ANPR) and the KNN approaches. The values obtained from the ANPR were considered the true values. Using the ANPR data, it was shown that most of the TTV was caused by period to period variability (around 65%),
followed by vehicle to vehicle variability (around 23%) and finally, day to day variability. To see whether the KNN approach is capable of describing the TTV, the three terms were computed using the KNN approach and were compared to those obtained from the ANPR method. The results of the KNN method showed that the period to period variability represented about 80% of the total variability, while day to day variability represented 20% of the total variability. The KNN approach was unable to characterize the vehicle to vehicle variation as the model output was always a single value of travel time of one vehicle, using 15-minute volume and occupancy data. With respect to the total variability, the results of the KNN approach always underestimated the total variability when compared to the ANPR method. Another experiment carried out by the authors was an investigation of day to day travel time variability for the same period (travel time between 7:00 a.m. to 7:15 a.m. for different days of the week). In this experiment, an analysis was carried out to identify which days of the week were comparable. Cross correlation was used to identify similar days and it was found that all weekdays had correlation coefficients higher than 0.9 whether the travel times were computed using the ANPR or the KNN method. Also, Saturday and Sunday correlated with correlation coefficients above 0.85 using both methods. This indicated that the KNN approach was able to identify days with similar behaviour of the TTV.

Guo and Jin (2006) developed an approach to estimate link travel times using single loop detector data. This approach integrated both the cross correlation analysis and the probabilistic model of random average travel time. Comparisons between this approach and other previous approaches showed that this approach was indeed effective.

Liu and Ma (2007) aimed at developing an analytical travel time estimation model that utilized data from loop detectors as well as real-time signal status information. The model assumed that data from detectors and signal controllers could be transmitted through wires to a management centre where the model could be processed. Travel time of the analyzed urban corridor was divided into two components: free flow travel time and delay time. Free flow travel time was estimated by dividing the segment length by free flow speed. The delay time was computed as the summation of queuing delay and signal delay. The authors used a hypothetical urban corridor of 10 actuated-signalized intersections to test their
analytical model in a simulation environment. The RMSE statistic was used to compare the estimated travel times and the simulated travel times. Different demand levels were tested and the RMSE was always below 10% which was a result expected by the authors. The main contribution of this research was that it utilized not only loop detector data but also signal controller data to estimate corridor travel times.

2.3.2 Travel Time Estimation/Prediction Using AI Models

Palacharla and Nelson (1999) conducted a study estimating link travel times using data from loop detectors in urban areas. The typical data collected from loop detectors were vehicle counts and occupancy over a pre-set time interval. Fuzzy logic and neural networks were integrated and used in order to benefit from the strengths of both of them, rather than using any of them individually. The authors estimated the free flow component of the travel time by dividing the link length by the posted speed limit. Delay time was computed as the difference between link travel time and free flow travel time. First, fuzzy rules were employed to transform all input and output values into associated membership values. The input variables were traffic counts and occupancy, while the output variables were the measured travel times. A simple feed-forward neural network with three layers was then trained using the fuzzified input and output patterns. The standard back-propagation algorithm was employed in the learning process. The output of the neural network was a fuzzy pattern that needed to be defuzzified to the actual value. The authors concluded that the overall results of their models were satisfactory. In addition, they considered occupancy a better predictor for travel time in urban arterials than flow. Finally, the authors compared a fuzzy neural model, a fuzzy alone model, and regression models. The results showed that the fuzzy neural model significantly outperformed all other models.

Rilett and Park (1999) developed a direct one-step approach to forecast short-term corridor travel time. Predictions were made for 5, 10, 15, and 20 minutes ahead of the current analysis period. The secondary objective of this research was to investigate the appropriateness of the traditional two-step methods in estimating corridor travel time using empirical data instead of simulation or theoretical models. The authors compared three approaches for corridor travel time estimation. First, they tried a one-step approach that
used a Spectral Neural Network (SNN) model to directly estimate future corridor travel times. Next, they used a two-step model that started with predicting the travel times on links constituting a route using SNN and then combing them to forecast travel time on the corridor. Finally, they tested a two-step model that integrated both historical and real-time data for link travel time forecast. The authors tried three different approaches for the direct forecasting of corridor travel times. In the first approach, they used recent corridor travel times as an input for the SNN. In the second one, they used recent link travel times as an input. Finally, in the third approach, they used a combination of both link and corridor travel times. The authors used data of AVI-equipped vehicles on a freeway corridor. The results showed that the direct approach of travel time forecasting using SNN outperformed the traditional two-step models. In addition, a direct SNN prediction model that used a combination of both recent link and corridor travel times provided the best results among all other direct models.

2.3.3 Travel Time Estimation/Prediction Using Simulation Models

Wahle et al. (2001) used online simulation to estimate travel times. Traffic volume data were collected from 750 inductive loop detectors and online simulation was carried out. Travel time estimates were published on the web at one minute intervals. A route guidance system using fuzzy logic was developed using these estimates.

2.3.4 Hybrid Travel Time Estimation/Prediction Models

You and Kim (2000) integrated a non-parametric travel time estimation model within a GIS environment. The model had five modules: graphic user interface, real-time data collection, database (real-time, road network, and historical data), forecasting, and machine learning. The real-time data collection module collected data from loop detectors and probe vehicles every 30 seconds for highways and every 5 minutes for arterials. This module represented temporary storage for this data when forecasting was taking place. The forecasting module used a nonparametric KNN approach to predict travel times for the next 15-60 minutes. The forecasting process required a number of model parameters to be defined by the user before the prediction was commenced. These parameters included the
forecasting range, the search for data segment length, the day of week, the value of K, and the local estimation method. After a forecast had been made, the model compared the forecasted value with the observed value. In the event the error margin had exceeded a specific threshold, the Machine Learning (ML) module was activated. The role of the ML was to re-adjust the parameters of the estimation models so as to minimize prediction errors. The model was applied to two different road networks: arterials and highways. It was shown that the model performed much better with the highway network (MAPE<3%) than with the arterial network (MAPE<10%). In other words, the historical records from the highway network were more representative in the nonparametric analysis. The authors also compared the model with and without the ML module. The results showed that the ML significantly improved the prediction ability of the model.

Chien et al. (2002) developed a travel time prediction model for a real-time motorist information system in New Jersey. The developed model involved online simulation and a Kalman filter. Initially, a CORSIM simulation model of the study-bed was calibrated. Online traffic data were obtained from five acoustic sensors installed in designated locations within the study bed. Feeding the simulation model with real-time speed and flow data, travel time data could be generated. The Kalman filter was used to predict travel times for the next time period. Information was disseminated to motorists via VMS. To test their model, four hours of traffic operations were simulated from 6:00 a.m. to 10:00 a.m. Travel times of each link were recorded during a 5-minute period. Simulated travel times were used in the model as the real-time data and as the true data as well. Different error indices were used to compare the predicted and the true (simulated) values of travel time. The results were satisfactory with only minor prediction errors.

Liu et al. (2006a) aimed at developing an online travel time predictor using a simulation-based system. The lack of extensive historical data caused most of travel time prediction techniques to be unsuitable. They proposed a method that employed CORSIM to directly estimate travel times in real-time. The study area was covered with only ten sparse loop detectors. A number of issues, therefore, needed to be handled in order to perform an accurate online simulation. These issues included dealing with missing data, incident detection, estimating fractions of turning movements, predicting detector data at the
targeted time horizon, and calibrating the simulation model. A number of modules were
developed to tackle these issues. For their volume prediction module, two algorithms were
developed. The first was used when both historical (i.e. previous days) and online (i.e.
previous time intervals) data were available. A KNN model was developed to compute the
most similar traffic pattern of previous days to the current day and hence predict future
traffic volumes. The computed metric distance incorporated traffic volumes of the same
time intervals for previous days and the current day. The authors used the arithmetic mean
of the nearest neighbours as the predicted traffic volume at the targeted time interval.
Noteworthy is the fact that the authors did not define the size of K neighbours used in their
analysis. The second algorithm was used to obtain current volumes of malfunctioning
detectors. A decision tree based on three criteria: season, weekdays, and events, was used
to define the nearest neighbours to predict current volumes using historical data only.
Finally, an online travel time prediction system was developed on the web. Travel time
predictions were published for different time horizons. For quality assurance, the operator
could simply compare predicted travel times with actual travel times.

Liu et al. (2006b) proposed a hybrid model for urban travel time prediction. The model
integrated both the State Space Neural Network (SSNN) and the Extended Kalman Filter
(EKF). The EKF was used to train the data in the SSNN, rather than the conventional
approach. The authors used error measurements to compare their approach, denoted by
SSNNKF, to another two approaches namely the Kalman Filter method (KF), adapted
from the model used by Chien and Kuchipudi (2002), and the State Space Neural Network
trained by Levenberg-Marquardt (SSNNLM). The results showed that the proposed
method consistently outperformed the other two methods. Sensitivity analyses were carried
out for both the weighting parameters of the neural network and the Kalman filter noise
parameters. For the weighting parameters sensitivity analysis, results showed that, at the
beginning, there were great prediction errors due to the inappropriate allocation of
weighting parameters. However, as more observations became available, the prediction
errors declined sharply. As for the KF sensitivity analysis, three parameters were chosen
for the analysis: initial error covariance, state noise covariance, and observation noise
covariance. The results showed that initial error covariance did not play a significant role
in changing the results, whereas noise covariance, and observation noise covariance were responsible for major impacts on the prediction errors.

Zhang et al. (2007) introduced a new method to predict travel times in urban networks. The new method incorporated system recognition into a KNN approach. The hybrid model developed was compared to another two approaches to estimate urban travel time. The first was the KNN approach developed by Robinson and Polak (2005) and the second was the Kalman Filter method developed by Kuchipudi and Chien (2002). The error analysis showed that the hybrid model consistently outperformed the other two approaches.

### 2.4 TRAVEL TIME CORRELATION

Travel time correlation has been extensively researched from many perspectives. For example, the correlation or the relationship between traffic data obtained from test vehicles or loop detectors and the data obtained from transit vehicles has been investigated by Cathey and Dailey (2002, 2003), Tantiyanugulchai and Bertini (2003) and Chakroarty and Kikuchi (2004). Rakha and Van Aerde (1995), on the other hand, compared freeway travel time estimates from probe vehicles with those obtained from loop detectors. The loop detectors measured spot speeds, counts, and occupancy every 30 seconds. Speed estimates were aggregated for 5-minute intervals and travel times were estimated by dividing the segment length (1 mile) by the aggregated speeds. Statistical analysis was carried out using data comprising 6 days. The correlation coefficient was 0.83, which showed a general consistency between travel times estimated from probe vehicles and loop detectors. The authors also compared travel time estimates from GPS-equipped vehicles with those measured manually using a stopwatch operated by an observer. Nine test runs were carried out on two different networks: a major arterial network and a downtown grid network. The comparative results showed that reported errors were within 6 seconds.

Sanwal and Walrand (1995) found strong correlation between probe vehicle speeds and loop detector speeds. Sen et al. (1997) pointed out two types of travel time dependences: temporal and spatial. Temporal dependence refers to the correlation between travel time realization on the same link while spatial dependence refers to the correlation between travel time realizations on consecutive links that make up a route. The authors’ work
focused on temporal dependence, rather than spatial dependence. They concluded that vehicle travel times on the same link for the same measurement period are always correlated, except in the case of lightly travelled links. The problem of how to estimate travel time correlation between links on a corridor was also introduced by Sen et al. (1999). Theoretical analysis of this correlation was presented in Hall (1986), and Fu and Rilett (1998). Rilett and Park (1999) developed a one-step approach using ANN to directly predict corridor travel times and consider inter-correlation between link travel times. The authors suggested that using a separate model to predict the travel time on each link, without considering the covariance with other links, can lead to significant errors.

Eisele and Rilett (2002) presented empirical analysis using real-life data to estimate corridor travel time mean and variance while considering the dependence of travel times between links on the corridor. The mean and variance of corridor travel times were computed using first and second order approximation of Taylor’s series. Link travel time mean and variance were estimated once with the non-parametric Lowess model and then with the polynomial model for the two tested freeway corridors. The results showed that the Lowess approach, although promising in estimating link travel time mean and variance, appeared problematic in calculating corridor travel time parameters when systematic gaps in the data were found. Another finding of this research was that estimating corridor travel time mean and variance from loop detector measurements is problematic.

Gajewski and Rilett (2003) developed a Bayesian-based methodology to estimate the distribution of travel time correlation between links constituting a corridor. This developed approach did not seem sufficiently flexible in identifying the correlation under different traffic conditions which was noted as a major concern in this paper.

In summary, previous work on travel time correlation addressed the following: 1) the correlation between probe vehicle travel times and travel times estimated from loop detectors data for the same link; 2) the relationship between travel times and/or speeds obtained from transit vehicles and probe/test vehicles for the same link; and 3) the correlation (covariance) between travel times of consecutive links on a corridor.
Current travel time estimation methods always employ data from the link itself, whether these data are historical, real-time, or both. Studies that reported estimating travel times from neighbour links were sparse. Du and Hall (2006) used a geostatistical model to estimate travel times on local roads without GPS probes using sparse probe data from other links. The study made use of 16 months of household GPS travel survey data, where at most 12 GPS-equipped vehicles existed on the network at the same time. It was found that half of the road network was not travelled by any vehicle during the survey period. An index, referred to as the “speed ratio” was developed by the authors to facilitate the grouping of links. This index is a function of average speed per observation, speed limit, and the number of observations in each time period. The authors assumed that speed ratios would be spatially auto-correlated for links in the same group. A geostatistical Kriging model was developed to estimate travel times on links with no data. Validation results showed mean errors in speed ratios of about 19%. For further examination of their models, the authors compared actual trip travel times to alternative routes travel times. The hypothesis was that the difference in travel times should not be significant for an equilibrium state (all routes have the same travel time). The average difference between the routes actually travelled (recorded on the GPS) and the alternative routes (based on travel times from the model) was 3.8%, while the absolute average error was 16.8%.

Tam and Lam (2006, 2008) developed a real-time traveller information system which integrated historical and real-time travel time data. The system explicitly considered travel time covariance between two consecutive segments on a pre-defined route and used this covariance to estimate travel time on a segment using data from another segment. Chan et al. (2009) further extended the work of Tam and Lam (2006, 2008) by comparing three methods for travel time estimation: historical data only, historical and real-time data, and historical and K-Nearest Neighbours (KNN) real-life data. They showed that use of the KNN method significantly improved the estimation accuracy.

2.5 GPS/GIS TRAVEL TIME STUDIES

Czerniak et al. (2002) suggested that the integration of GPS and GIS has introduced a number of advantages that include cost reduction, reduced analysis time, and improved
map quality. Nevertheless, this integration has resulted in a number of problems, such as missing data points, erroneous data points, and the need for an accurate map matching algorithm to project GPS points on the GIS map. All of these issues have caused the integration of GPS and GIS to be an active research area.

Quiroga and Bullock (1998) presented a comprehensive GPS-GIS integrated methodology in order to perform travel time studies. First, they introduced two methods of building GIS vector maps; one based on current planning maps, and the other, based on GPS data. They showed that the second method is more robust in terms of simplicity, accuracy and potential for integration with GPS for travel time surveys. In other words, building a GIS vector map using the same GPS receivers used in the travel time survey will solve the problem of matching the accuracy of the GIS map and the GPS receiver. After developing the GIS map, all “discontinuities” such as intersections, on-ramps, and off-ramps were marked as checkpoints. Each link between two discontinuity points was further segmented by a number of intermediate checkpoints so as to have a checkpoint located every fixed distance. Accordingly, each segment between two checkpoints would have its own tag, geometric, and spatial characteristics. The authors investigated three issues related to this GPS-GIS integration methodology. The first was segment length, where the issue of interest was the appropriate segment length between each two consequent intermediate checkpoints. The second was the sampling rate (i.e. the impact of collecting GPS data at different sampling periods). The third issue was the central tendency analysis, which compared harmonic mean speeds and median speeds. The first analysis showed that relatively short segments (0.2-0.5 miles) are required to characterize local traffic conditions. The results of the second analysis showed that the sampling period should not exceed half the shortest travel time on the link. The authors suggested using a sampling period of (1-2 s) for urban highways with signalized intersections to account for the rapid changes in the vehicle’s speed profile. The last analysis showed that median speeds are more robust in the estimation of central tendency than harmonic mean speeds. This research was directed to static travel time estimation (i.e. offline travel time surveys).

In a research effort by Hunter et al. (2005), GPS-instrumented test vehicles were used to collect offline travel time data. The authors presented a method for initial data processing
and error checking. They consequently developed an algorithm to estimate travel time between each two consecutive intersections (nodes). The algorithm assumed hypothetical reference lines for each intersection. Each of these reference lines was associated with a specific movement direction. If a vehicle crossed a reference line, the movement direction (right, left, or through) could be identified. This methodology was used to establish an aggregated travel time matrix between different origins and destinations in the study area.

Byon et al. (2006) developed a GISTT (GPS-GIS Integrated System for Travel Time Surveys) software package to estimate travel time in both static and dynamic modes. The inputs of the static module included map data, route data, GPS trip data, time slot, and threshold speed. The output of the static module included a visual trip display, link travel time estimates, delays, quality of the data, and position/acceleration graphs. A map matching algorithm was developed, along with travel time and delay algorithms. Travel time was estimated by dividing the link length by the average link speed. The delay algorithm estimated the portion of time during which the vehicle was below a certain speed threshold. Data quality was judged using a number of parameters such as number of satellites within range, and length of lost signals. The dynamic travel time module received real-time GPS position information from probe vehicles, using wireless communication, and plotted it directly on the GIS map. To validate the static module, a field test was conducted using an equipped probe vehicle that travelled several routes in different network configurations. Travel time estimates obtained from the software were compared to simulated travel times of Paramics and also to travel times measured directly from the field using stopwatches. The stopwatch field measurements were supplied by highway operators and results were for a day other than the field test date. Accordingly, the authors stated that the errors in the estimates might be partially attributed to the difference in time periods of data collection. Generally, errors in the estimated travel times were of the order of 6% of the stopwatch travel times, except for CBD routes, where signal blockage resulted in higher error values. The dynamic module was also tested and found to be successful in tracking probe vehicles when the update was carried out every 3 seconds.

Pan et al. (2007) proposed a GPS-based method to obtain historical arterial travel times and intersection delays. A map matching algorithm was developed and a model of link
travel time and delay estimation was presented. The map matching algorithm was based on an offline analysis of all GPS points. The horizontal position error was assumed to be in the order of 20 m, and the error region was defined as a circle. The matching algorithm started with the search for a “good fitted point,” which is a point associated with only one link of the road network. The algorithm made use of three criteria to develop a likelihood measurement that could shrink the candidate link set. The three criteria used were: minimum distance between the point and the link, similarity of vehicle heading and roadway direction, and network topological characteristics. Once the map matching took place, link travel time could be estimated by subtracting the time stamp of the link starting point from the time stamp of the link ending point. A method to compute the delay components (acceleration-deceleration-queuing) was then presented. Two experiments were carried out using real-life data to test the efficiency of the methodology. The extracted travel times and delays from the GPS points were compared to those obtained from a Distance Measurement Instrument (DMI). The results showed that the proposed method provided good estimates of link travel time and overall delay time. However, the method could not accurately estimate delay components.

2.6 VEHICLES AS PROBES

An introductory report by Sanwal and Walrand (1995) illustrated the main elements of any operational Vehicles As Probes (VAP) system. The authors suggested that these vehicles are representative of the traffic stream with some variation, which is statistical in nature. Probe vehicles report their data at regular intervals to a central computer system for the purpose of data analysis. Using short intervals increases the communication costs, while using long intervals may impact the accuracy of the data. The authors defined five key elements for the implementation of a successful and economically feasible VAP system. These five elements were: system organization, system objectives/priorities, sample size or market penetration, polling rate (i.e. data acquisition period), and measured data. The authors discussed the appropriateness of GPS positioning in a VAP system. They suggested that using GPS could offer the best positional estimates in open areas where no obstruction to the GPS signals exists. Positional accuracy can be further increased using a Differential GPS (DGPS) receiver. Other approaches for position determination were also
introduced, such as using roadside beacons, or a combination of GPS and roadside beacons.

Yim and Cayford (2001, 2002) conducted a study of three phases to set up a system of vehicles as probes for travel time data collection, fusion, and dissemination. The three phases were: program development and initial field test, field operational tests, and a large-scale field implementation. The focus of this research was on the first phase mentioned: program development and initial field test. Two counties were simulated and probes with varying positional accuracies were generated. A positional accuracy of 10 m was sufficient to identify more than 99% of all links in the two counties. As the accuracy degraded, the percentage of unidentified links increased slightly, except for freeway segments, where the increase was dramatic. This research included a discussion of some behavioural issues related to the implementation of VAP systems. The authors stressed the importance of using probe vehicles that are representative of the traffic stream. Some fleets, such as service vans or patrol vehicles could have lower speeds and long stoppages which could deteriorate the accuracy of the collected data. Such behaviour issues should be identified and handled in a way that enhances the quality of the collected data. A software package was developed by California Partners for Advanced Transit and Highways (PATH) team to track probe vehicles on a GIS map using GPS information. The software contained a number of modules for link matching, path finding, and travel time estimation. A field test was conducted to investigate the potential for using GPS, DGPS and cellular phone technologies to track probe vehicles. The developed software was used to map the vehicle track, show the actual positional accuracy of the vehicle moving, and the success in terms of assigning the vehicle to the correct road. Data generated by the DGPS were accurate enough to identify about 93% of all followed paths. On the other hand, GPS data were less accurate such that only 86% of the traversed paths were identified. The cellular phone service provider was unable to provide the research team with any tracking data for the probe. Hence, the planned comparison was incomplete. The last part of Yim and Cayford (2001, 2002) research work studied the use of service patrols as probes. An equipped vehicle was driven on a freeway segment and data were sent to a central computer where the developed software processed the received information. Vehicle tracking was easy
even with a positional accuracy of 30 m. The estimated travel times were not compared to any ground truth travel times, and the accuracy of these estimates could not be assessed.

Demers et al. (2006) conducted what they called “the first attempt in history to share online traffic information between probe vehicles through wireless communication and change routes dynamically.” In this study, 200 vehicles were equipped with GPS receivers, PDAs and 3G wireless cards to obtain and share travel time information through a wireless network. Vehicles changed their routes dynamically in real-time based on the received travel times. The experiment procedure was as follows: first, the GPS on each vehicle was activated to obtain the vehicle’s location. Then, the 3G card was activated to establish a connection with the project server and the route guidance software, CoPilot. A number of virtual landmarks, called “monuments,” were superimposed on a network digital map. As a vehicle passed a monument, a monument to monument (M2M) message was sent to the server. These messages included the current monument ID, the last monument passed, the current time and the vehicle’s current position. The server used M2M to compute updated M2M travel times for the network. Every minute, the vehicles sent queries to the server to obtain the updated M2M travel times. CoPilot downloaded all recent travel times and checked the possibility of a better route based on the new information. If a better route (i.e. less expensive) was found, a message was relayed to the vehicle. Directions were given to the driver using aural route guidance. Finally, the authors presented examples showing the type of information that could be extracted based on the archived data.

Xu and Barth (2006) investigated a decentralized approach to share and transfer information between traveling vehicles without the need for a Traffic Management Centre (TMC). The authors stated that a decentralized system could be better than a centralized one which depends on a TMC, as the latter may suffer from single-point failures. The focus of this study was on evaluating different on-board travel time estimation algorithms using a decentralized system. The proposed system incorporated the following assumptions: 1) each vehicle had the ability to obtain its position with time, 2) each vehicle could communicate with other vehicles over a short range, and 3) each vehicle was equipped with an onboard unit capable of estimating its travel time based on position information and time stamps. To satisfy these conditions, each probe vehicle was assumed
to be equipped with a GPS receiver, a digital GIS map, a unit to estimate travel time and a Dedicated Short Range Communication (DSRC) wireless interface. The vehicle would receive traffic data from the surrounding vehicles, combine them with the data it had collected, and then send this information every 10 minutes to other vehicles. Travel time estimates could then be used as a basis for a Dynamic Route Guidance System (DRGS). Three different algorithms for travel time estimation were compared: blind averaging from all vehicles, use of a decay factor, and estimating only by direct-experienced vehicles. In the first scheme, when a vehicle received the travel time information, it used this information to fill in its empty cells of travel times. However, if the vehicle receiving data had records of travel time for the same links, these records were updated by averaging both received and existing records. A decay factor method simply assigned more weight to recent measurements of travel times than historical ones. Calibration of the decay factor was carried out using different simulation runs. The last algorithm used data solely from vehicles that had already traversed the link. Note that in the first two algorithms, travel time on link \( i \) could be transmitted many times by vehicles receiving their information from one vehicle only and, using this method, the records would be duplicated. The last algorithm solved this problem by updating travel time only from vehicles that traversed the link. A simulation experiment was carried out using Paramics and an NS-2 communication network simulation tool. The three travel time estimation algorithms were tested in the simulation environment. The third algorithm, estimating travel time from experiencing vehicles only, significantly outperformed the first two algorithms in terms of the MAPE.

2.7 CELLULAR PHONES AS PROBES

Published research by Ygnace et al. (2000) is considered one of the earliest attempts to investigate using cellular phones as probes. The main driving force for this research was the Federal Communications Commission E-911 mandate which necessitated that all cellular phones be located at an accuracy of about 125 m by October, 2001. The report presented an overview of cellular phone positioning techniques. The authors discussed three methods for cellular phone tracking: signal profiling, angle of arrival, and time measurements. A number of actual developments for cellular phone positioning were presented showing their potential for travel time data collection.
Yim and Cayford (2001, 2002) suggested that using cellular phones as probes showed some promise. However, this approach suffers from many technical problems. In comparison, using GPS as a mature technology for tracking showed higher success rates and more suitability for system deployment. The authors proposed a comparison between GPS and cellular technology. However, such a comparison was not carried out due to lack of necessary information at that time about cellular technology. Instead, two previous cellular data sets were analyzed using the developed tracking software. Analysis of these data showed that about 59% of the data points did not match any of the network links. The software used 44 hours of tracking data to match 107.4 km of the road network. In the absence of any information about the actual locations of these data points, the authors could not formulate a comparison to check the accuracy of tracking. No conclusive results were presented.

In a NCHRP project report (2005), sixteen real-life planned or completed deployments to monitor traffic using Wireless Location Technology (WLT) were presented. The authors stated that their review was limited to what could be found in the available literature. No interviews or surveys had been conducted with the parties actively involved in these deployments to collect more data. Benefits and opportunities offered by these systems were evaluated based on available results. As well, obstacles that could impede the evolution of these systems, were described. Each deployment was characterized in terms of system coverage, participants’ relationship with cellular service providers, technology, Department of Transportation (DOT) requirements, general results, and independent evaluators. Simulation-based experiments were also included in this report. According to the authors, although these systems may not prove or disprove any hypothesis, they do provide an additional source of information. General trends observed by the authors in the real-life deployments, were also summarized and illustrated.

Cayford and Yim (2006) conducted a complementary study to the research introduced in Yim and Cayford (2002). This paper introduced the second phase of their study: field operational tests. The Travel Information Probe System (TIPS) software, which was previously developed, was further enhanced and used in this study. Cell-probed vehicles were tracked for 24 hours in Tampa, Florida in 2005. Tracked vehicles were then assigned
to road segments and travel times were computed using the available records. Data were aggregated in 5-minute intervals and network coverage was computed. Network coverage is dependent on the spatial distribution of all cellular phones across the study area, and the temporal distribution of cell phones during 5-minute periods throughout the day on different links. Analysis showed that freeway coverage is significantly higher than surface street coverage. This was due to traffic volumes on freeways being much higher than volumes on surface roads. As well, coverage between 10 a.m. and 10 p.m. was much higher than coverage during other time periods, and was on average 76% for freeway segments. The data used were obtained from a single carrier only and represented only 15% of available phones in the area. In order to examine the impact of increasing market penetration (i.e. obtaining data from other carriers), the travel time information was aggregated over 15 minutes intervals, which was equivalent to tripling the number of cellular probes used. In this case, the coverage increased to 86% for all freeway segments between 10 a.m. and 10 p.m. Aggregating the data over 24 hours, the coverage was 99.7% for all freeway segments, and 98.7% for all surface streets. The authors investigated the issue of sample size required for reliable speed estimates. Using the statistical method presented in their previous work and different values for allowable errors and speed standard deviation, they computed the sample size required. Accordingly, they estimated the percentage of miles covered within a 95% confidence in the estimated travel speeds. The authors stressed the importance of comparing the generated data to other data sources obtained from loop detectors. They proposed this issue as a future research topic. Their final conclusion was that using cellular phones as probes shows great potential in the event broad network coverage is available.

Qiu et al. (2007) presented a comprehensive literature review of previous and current work dealing with cellular phones as probes in both academia and industry. The latest wireless location technologies were presented, followed by the results of previous projects. The authors also proposed a general architecture of a system that uses cellular technology to monitor traffic.
2.8 OPERATIONAL DEPLOYMENTS AND FIELD TESTS OF VEHICLES AS PROBES SYSTEMS

A large number of field experiments have been carried out to investigate using vehicles as probes for traffic data collection. The following is a description of some of these initiatives and their deployments.

2.8.1 ADVANCE Project

One of the first attempts to use vehicles as probes was the ADVANCE project which was a large-scale Advanced Traffic Management System (ATMS) initiative deployed in suburban Chicago (Sen et al. 1997). In this project, traffic data were collected first from a number of probe vehicles, then transmitted to a traffic management centre where they were processed and transmitted again to other vehicles. Data were transmitted in both directions using Radio Frequency (RF) every 5 minutes. Travel time predictions were made for 5, 10, and 15 minute intervals.

2.8.2 Taxi-Floating Car Data (Taxi-FCD)

A comprehensive taxis-as-probes system known as “Floating Car Data” (FCD) has been deployed in a number of European cities, namely: Berlin, Vienna, Munich, Nuremberg, Stuttgart, and Regensburg. Gühnemann et al. (2004) reported on using FCD collected from taxis in Berlin, Germany. Vehicles receive their locations through on-board GPS receivers with an accuracy of 10-20 m. Data messages are transmitted to a control centre through a radio channel. These messages include information about taxi ID, time stamp, position, and taxi status. The default polling rate is 1 message/minute and can sometimes be increased to as high as 4 messages/minute. Using the transmitted information, the central computer constructs the vehicle trajectory and matches it to a digital map. Link speed and travel time can then be computed. One application of the deployed FCD system is a dynamic route guidance (DRG) system. A web-based internet site was developed to enable real-time and pre-trip planning. Average taxi travel times for the last 30 minutes were used in the routing algorithm. If no taxi has passed a certain link during the past 30 minutes, historical data of the same time and the same day of the week are used. Speed profiles are
also developed using FCD data, and hence, speed-flow relationships are established fusing data from FCD and loop detectors.

### 2.8.3 Floating Vehicle Data (FVD™)

England Floating Vehicle Data (FVD™) is a system which collects, analyzes and predicts trip travel time using data collected from probe vehicles. Each probe vehicle is equipped with an on-board GPS receiver that transmits the vehicle’s current position and speed to a central computer via either the GSM network or radio waves. The system is operated by ITIS, which is a private company considered to be one of the leading providers of transportation information in the United Kingdom. ITIS claims to have the largest commercial fleet in the world capable of delivering traffic information via different platforms such as digital radio and automated telephone services (Simmons et al. 2002). The system was launched in February, 2000 with a limited number of vehicles. The company continued to expand the system for improved data collection. The system concentrates on the major UK network, covering over 30,000 miles of roads (Simmons et al. 2002). ITIS developed a new concept for the FVD unit. This concept is based on the fact that the average vehicle travels less than a for-business vehicle, which in turn travels less than a truck or bus. ITIS estimated that an average driver has a value of 1 FVD unit, a business driver has a value of 3 FVD units, and a truck or bus driver has a value of 30 FVD units. Three fleets are used in the ITIS FVD™ system, these are:

- **AA Patrol**: a roadside assistance patrol whose associated sample comprises 31500 FVD units,
- **National Fleet**: an accident management and full service maintenance and repair facility whose associated sample size is 16000 FVD units,
- **Stobart Group**: a haulage company whose associated sample size equals 22500 FVD units.

The FVD™ system does not provide a continuous stream of data, rather specific data for customers in need. An important characteristic of this system is an algorithm that defines the most probable vehicles on a specific road at a certain time period. Analysis of historical records is carried out to specify these vehicles and, hence, when a customer requires data
about a link at a certain time period, communication is activated between the TMC and the vehicle to obtain travel time information. Several commercial services have been developed utilizing a comprehensive database of historical and real-time travel information. These services include in-vehicle congestion avoidance advice, congestion analysis, and trip time prediction.

As this system utilizes commercial vehicles as probes, the obtained travel time data could, to a great extent, be biased simply because commercial vehicles are not representative of the average traffic stream. Many constraints such as speed limit, vehicle size, lane dedication, and motor horsepower cause the behaviour of commercial vehicles to be different from that of other vehicles. Accordingly, estimated travel times could be much higher than actual travel times. One major concern about this system is the unavailability of any of the developed algorithms. Development of such a large system must have necessitated the development of a number of complicated algorithms for vehicle tracking, removal of false alarms, data cleaning, online travel time estimation, and travel time prediction. Published descriptions of this system mentioned that it was validated using a number of pilot projects. Nevertheless, the validation methodology was not discussed and the results were never published.

2.8.4 OPTIS

Jenstav (2003), Bishop (2004), and Karlsson (2005) reported on a project called OPTIS (OPtimized Traffic In Sweden), jointly funded by the Swedish government and the Swedish automotive industry. The main goal of this project was to develop a cost-effective method to collect traffic data required for an accurate traveller information system. A field trial made use of a sample of probe vehicles and was carried out in 2002 in the City of Gothenburg. The sample size of probe vehicles is uncertain. Three different sample sizes were reported in three different publications. Jenstav (2003) mentioned that 250 vehicles were used in the field test, Bishop (2004) stated that the field test comprised 223 test vehicles, and Karlsson (2005) reported that the sample size included 220 vehicles. All probe vehicles in this field study were equipped with GPS receivers, GSM phones and combined GPS/GSM antennae to transmit information wirelessly. By tracking these
vehicles, reliable travel time estimates were obtained. Travel time estimates from probe vehicles were comparable to those obtained from cameras and queue warning measurements. The probe vehicle system proved to be more cost-effective in terms of providing wide network coverage. To investigate the issue of appropriate sample size, simulation was employed. Results showed that in a mid-sized city, with less than 1 million residents, the required market penetration was between 3-5%.

2.8.5 IPCar System

Ueda et al. (2000) reported on the initial development and field testing of the IPCar system in Japan. The letter “I” refers both to “Intelligent” and “Internet,” while “P” stands for “Probe.” The objective of this system was to gather travel information by employing floating cars as moving traffic sensors. The project was divided into three models or phases: the introduction model, the intermediate model, and the future model. Using the introduction model, a small number of vehicles would serve as probes. Commercial and service vehicles were candidates for this task as they are capable of large travel distances. In the intermediate model, the number of probe vehicles would be increased, and better quality and coverage could be obtained. In the future model, the system would be implemented at a large scale level, and good quality information could be obtained via networking between different services.

In the field test, ten GPS-probe vehicles were used to gather information over a period of three days. The vehicle’s speed, outside air temperature, wiper usage, switch settings, and GPS signals were detected every second and were transmitted to the traffic centre every 30 seconds using a packet communication system. Positioning information was found to be able to detect an object within a radius of 35 m with a probability of more than 95%. Travel times were measured and also obtained from averaging instantaneous speeds. Comparison of the two estimates revealed that the margin of error was around 40%, when a great dispersion in speed estimates existed, due to short averaging time (2 minutes). However, when averaging time was increased to 5 minutes, the error margin decreased to 10%. A final stage of this preliminary field test involved the fusion of voice information with sensor data. The test was conducted on three vehicles over a period of three days.
Drivers responded to periodic free-style questions posed by the system operator. The authors concluded that fusing both types of data provided valuable information.

2.8.6 Singapore Taxis

A large scale nationwide VAP system has been deployed in Singapore, making use of more than 10,000 taxis as probes. Real-time position and speed information was collected from these vehicles over a network area of 3000 km². Cheu et al. (2002) reported that the collected traffic information was disseminated to the public via the website “TrafficScan.” However, a search for this website, by the author, showed that it no longer exists.

2.9 LITERATURE REVIEW SUMMARY

Different techniques can be used for travel time data collection. The choice of a data collection technique depends on factors such as the road environment type, traffic volume levels, available budget, survey period, purpose of data collection, and permissible error. The Travel Time Data Collection Handbook (Turner et al. 1998) categorizes the methods used for travel time data collection into four groups; test vehicle techniques, license plate matching techniques, emerging and non-traditional techniques, and ITS probe vehicle techniques. In summary, license plate matching methods provide limited coverage, are costly, and are expected to be inaccurate in urban environments. Emerging and non-traditional techniques for travel time data collection are still under development and are relatively unreliable. Also, installation costs for these sensors are likely to be high, if deemed necessary for an entire network. The utilization of probe vehicles is a cost-efficient approach to collect real-time travel time data. Probe vehicles can be taxi fleets, buses, commercial vehicles or any other fleet type that exist on the road network to serve a particular purpose. Consequently, the approach does not incur any additional costs associated with equipment setup, hiring drivers, etc.

Travel time estimation refers to obtaining travel times using measurements of the current analysis period while in travel time prediction, travel time is projected into the future by various modelling means. Available techniques for travel time estimation can be categorized according to their input data, input variables, or computational procedures.
According to the input data, estimation models can use either simulation data, or real-life data. Real-life data can be historical data, real-time data, or a combination of the two. According to input variables, some models use travel time as both the input and output variables, while other models use traffic parameters, such as spot speeds, occupancy, vehicle headway, and volumes, which can be obtained from loop detectors. According to the computational procedures of travel time estimation models, three main groups can be identified. These include: statistical techniques, Artificial Intelligence (AI) techniques, and simulation models. A number of studies that estimated/predicted travel times were presented. These studies were classified according to the modelling approach, type of data, and input variables. Results of these studies were mixed and did not support one approach over another. However, a more sound conclusion would be that each problem or case study is, to some extent, unique, and the best modelling approach can only be determined by comparing a number of techniques.

Travel time correlation has been researched broadly from many perspectives. Previous work on travel time correlation addressed the following areas: 1) the correlation between probe vehicle travel times and travel times estimated from loop detectors data; 2) the relationship between travel times and/or speeds obtained from transit vehicles and probe/test vehicles; and 3) the correlation (covariance) between travel times of consecutive links on a corridor. The focus of this thesis will be on the last type of correlation listed here: the correlation between nearby links. The theoretical aspects of this correlation have been discussed in the literature. As well, empirical analyses using real-life data have been described in many studies. Problem formulation was always undertaken to account for correlation while calculating corridor travel time. Few studies were found in the literature which attempted to estimate travel times from neighbour links data. The current research supports this hypothesis, i.e. estimating travel times using relationships of nearby links, albeit viewed from a different perspective.

Several studies were carried out to integrate data collection using GPS and GIS analysis. The integration of GPS and GIS has its advantages, which include cost reduction, reduced analysis time, and improved map quality. On the other hand, this integration has resulted in a number of problems, such as missing data points, erroneous data points, and the need for
an accurate map matching algorithm to project GPS points on the GIS map. All of these problems have caused the integration of GPS and GIS systems to be an active research area. State-of-the-art research focuses on developing more accurate map matching algorithms, and extracting travel time and delay information from GPS data collected.

A number of field experiments have been carried out to investigate the use of vehicles as probes. Key issues of special interest to researchers are market penetration, accuracy of estimates, and methods for tracking vehicles. Examples of real-life initiatives that used vehicles to collect traffic data include: ADVANCE project in the USA, Taxi-Floating Car Data (Taxi-FCD) in Germany, Floating Vehicle Data (FVD®), developed and operated by ITIS in the United Kingdom, OPTIS Project in Sweden, IP-Car in Japan, and Singapore Taxis.
CHAPTER 3: DATA DESCRIPTION

In this chapter, data used throughout this research is described. Two different sets of data were used. The first data set is the real-life data that were collected through a travel time survey. The second set is synthetic data generated from a simulation model of the study area. The description of the two sets of data is given in the following sections.

3.1 THE STUDY AREA

Downtown Vancouver is a typical urban environment with traffic signals interrupting traffic operation. There is considerable congestion during the a.m. and p.m. peaks. The study area was selected to cover almost half of the downtown area as shown in Figure 3.1. The network is a grid network composed of 23 streets, 12 of which are in the east-west direction, and 11 of which are in the north-south direction. The 23 streets comprise 115 intersections, of which 108 are signalized. The average segment length of the east-west streets in downtown Vancouver is between 70-80 m, while for the north-south streets, it is between 160-190 m. The choice of this condensed network for the current analysis included spatially nearby links with similar land use characteristics. For example, all selected segments are classified as urban, are located in Central Business District (CBD), have approximately the same length, are signalized from both ends, and have similar cross-section elements, as per the number of lanes. This ensures that any high travel time correlation is not randomly found, but rather is attributed to the fact that all of these links are subject to similar traffic conditions. Pearson correlation is used to refine the selection to include only segments that have strong and statistically significant travel time correlation.
3.2 REAL-LIFE DATA

A travel time survey was conducted to simultaneously collect real-life average link travel times on a set of links of the analyzed study area, downtown Vancouver, during the peak periods of the day and for different days of the week. Real-life data collection served the following purposes:

- Proof of concept of neighbouring links travel time estimation
- Investigation of various modelling approaches that can be used in neighbouring links travel time estimation
- Calibration and validation of the microsimulation model of downtown Vancouver. This will enable the testing of other issues related to the research methodology that could not be studied using real-life data.
3.2.1 Survey Route and Method

The survey route included six heavily congested corridors in the north-south direction of the downtown area, namely: Homer, Richards, Seymour, Howe, Hornby, and Thurlow Streets. The selection of this route considered movement restrictions in the downtown area during peak hours. The route starts and ends at the same point on Homer Street at Drake Street, see Figure 3.2. The route length is about 9-km and its estimated driving time is approximately 28 minutes.

This survey was conducted using an Active Test Vehicles Technique which has been used for many years to collect travel time data. Details of this approach can be found in Turner et al. (1998). Several recent studies have advocated the merits of using GPS for travel time data collection (Quiroga and Bullock 1998, Czerniak et al. 2002, Hunter et al. 2005, Byon et al. 2006). However, the success of utilizing GPS receivers in this study was uncertain as they experience significant signal losses and signal drifting in urban canyons, such as downtown Vancouver. Accordingly, laptops equipped with a customized in-house program were used instead of GPS receivers to record the required travel times. The data collection program was created as a Microsoft Excel macro using Visual Basic for Applications (VBA). This approach is considered a variation of the manual method of active test vehicles. It offers the advantages of reducing errors and staff requirements. Every participant vehicle was equipped with a laptop placed on a driver-accessible spot with the program installed. Power inverters were provided for participating drivers to allow for continuous charging of the laptops from vehicle batteries.
3.2.2 Data Collection Program Design

The custom data collection program was designed to allow for preliminary information to be input, such as driver’s name, weather, and date. The designated route was preloaded in the program as a series of checkpoints with a checkpoint at each intersection. Each checkpoint was labelled with a unique ID number and accompanying text description. The first checkpoint was the route starting point. The driver approaches a checkpoint and then has to select an option (e.g. press a button on the laptop keypad), according to the arrival
scenario. If the driver arrives in a red-signal phase, he/she chooses “Stop.” When the driver safely crosses the intersection after stopping, or without stopping, he/she selects the option “Through.” To simplify this task for the participants and eliminate the need for them to look at the laptop monitor while driving, the program was designed to assign the option “Through” to any button on the laptop keypad, except for the four arrow buttons, which were assigned the option “Stop.” As human error was expected during the survey, another option was added to the program that allowed for the correction of incorrectly pressed buttons. This was done by pressing the “ESC” key which would undo the previous action and enable the user to choose another option. Another advantage of the developed program is its ability to automatically save the data file with every new entry. This eliminated the burden on the driver to execute data saves.

When a driver pressed any button, a time stamp was recorded in a hidden file with a description of the action taken (i.e. “Stop” or “Through”). The driver had no control over the hidden data file. The file could only be revealed by the program developer using a complex procedure and by inputting a password. This function was added to avoid any data manipulation by the drivers prior to data delivery. The hidden data file contained information about date, control point ID, stopping time stamp, through time stamp, as well as additional notes. The additional notes showed the starting and termination times of the program, and if any record had been corrected. Figure 3.3 shows the logic behind the developed program.

Pressing a keypad button was not expected to cause any disruption or significantly decrease the drivers’ concentration. The use of laptops in this survey had several advantages. First, there was no need for a second participant inside the car to manually record the time stamps. Hence, staff resources were kept to a minimum, which significantly reduced the total survey cost. Also, the creation of a module in the laptop program, to capture the delay, added another rich source of information to the collected data. Finally, travel times were directly computed at each checkpoint with no need for the extensive data cleaning, reduction, and analysis required for high-maintenance GPS data.
Figure 3.3 Logic behind the Developed Program
3.2.3 Sample Size

One of the important issues that had to be investigated while designing the survey was the number of runs to be considered in each aggregation period in order to obtain a reliable estimate of average travel time during this period. Turner et al. (1998) suggested using the following formula to compute the number of runs:

$$n = \left[ \frac{t \times COV}{\varepsilon_{max}} \right]^2$$  \hspace{1cm} (3.1)

Where:

- $n$ = the number of runs required for each measurement period
- $t$ = t-statistic from Student’s t distribution for a specified confidence level
- $COV$ = the coefficient of variation of link travel time
- $\varepsilon_{max}$ = the maximum allowed relative error (%)

The value of t-statistic was estimated according to the degree of freedom; $df= n-1$. This means that a sample size has to be assumed first, along with other variables, and verified against the previous formula. The travel time Coefficient of Variation (CoV) varies according to the road facility class, control type, and traffic conditions. Empirical studies showed that the CoV ranges between 9-17% from different findings. The maximum allowable relative error in travel time studies is commonly chosen to be $\pm (5-10\%)$ according to the study purpose. In this survey, the CoV was assumed to be 10%, and the maximum allowable error was set to 10%. Two confidence intervals were used to compute the required sample size: 90% and 95%. Table 3.1 shows the iterative process of computing the required sample size at the two confidence intervals. In order to achieve an acceptable accuracy level while also considering budget constraints, it was decided to run the survey at a 90% confidence level which required only five runs (records) for each aggregation period.
Table 3.1 Required Sample Size at Different Confidence Intervals

<table>
<thead>
<tr>
<th>A</th>
<th>n (assumed)</th>
<th>df = n-1</th>
<th>t</th>
<th>CoV</th>
<th>ε_max</th>
<th>n (computed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>3</td>
<td>2</td>
<td>4.30</td>
<td>10%</td>
<td>10%</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>3.18</td>
<td>10%</td>
<td>10%</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>2.78</td>
<td>10%</td>
<td>10%</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>2.57</td>
<td>10%</td>
<td>10%</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>2.45</td>
<td>10%</td>
<td>10%</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7</td>
<td>2.36</td>
<td>10%</td>
<td>10%</td>
<td>6</td>
</tr>
<tr>
<td>90%</td>
<td>3</td>
<td>2</td>
<td>2.92</td>
<td>10%</td>
<td>10%</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>2.35</td>
<td>10%</td>
<td>10%</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>2.13</td>
<td>10%</td>
<td>10%</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>2.02</td>
<td>10%</td>
<td>10%</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>1.94</td>
<td>10%</td>
<td>10%</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7</td>
<td>1.90</td>
<td>10%</td>
<td>10%</td>
<td>4</td>
</tr>
</tbody>
</table>

3.2.4 Survey Implementation

Two pilot runs were carried out to check for road construction zones, closed streets, and potential turn restrictions that were not considered while designing the route. Another objective was to gain insight on the round trip travel time to ensure that the required number of runs during each peak period could be achieved. Finally, it was important to test the developed program in real-life settings, while driving, to search for any coding errors and consequently modify the program accordingly. The two test runs showed some programming “bugs” that occurred when specific actions had taken place. These bugs were fixed prior to the survey. Travel time for the round trip (i.e. a complete cycle) was initially estimated as 30 minutes. After the pilot runs, it was found that the survey route could be heavily congested thereby increasing the travel time to 55 minutes, especially during the p.m. peak.

The survey was conducted over four days in December, 2008, during the a.m. peak period (7:00 a.m. to 10:00 a.m.) and over only two days during the p.m. peak period (3:00 p.m. to 6:00 p.m.). All days were regular business days with clear weather conditions. Five drivers were scheduled to run the survey. Each driver was asked to start the route between 7:00 a.m. and 7:15 a.m. for the morning peak and between 3:00 p.m. and 3:15 p.m. for the
evening peak. Drivers were instructed to continuously drive back and forth along the route for the three hours comprising the peak period.

### 3.2.5 Travel Time Data Analysis

Link travel times were computed by subtracting the time stamps of two consecutive check points. The data were checked for errors that may have occurred during the survey. Other outliers were found by searching for travel times that were less than 5 seconds or greater than 300 seconds. The average travel time on each segment was computed using the data collected from all drivers entering the survey route within a 25-minute window. A total of 139 runs were completed, of which 104 runs were done in the a.m. period, and 36 in the p.m. period.

**Route Travel Times**

Route travel time refers to the total time taken to finish a complete run (i.e. cycle) ending at the same starting point. In general, the total route travel time during the a.m. peak was much lower than that during the p.m. peak. This led to the completion of 5-6 runs in the a.m., and only 4 runs in the p.m. peak. Individual drivers’ route travel times are presented in Table 3.2. A descriptive analysis is shown in Table 3.3.
# Table 3.2 Individual Route Travel Times (Seconds)

<table>
<thead>
<tr>
<th>Day (1)</th>
<th>AM</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run (1)</td>
<td>1626</td>
<td>2275</td>
</tr>
<tr>
<td>Run (2)</td>
<td>2017</td>
<td>2084</td>
</tr>
<tr>
<td>Run (3)</td>
<td>2468</td>
<td>2725</td>
</tr>
<tr>
<td>Run (4)</td>
<td>1885</td>
<td>2275</td>
</tr>
<tr>
<td>Run (5)</td>
<td>1778</td>
<td>NA</td>
</tr>
<tr>
<td>Run (6)</td>
<td>NA*</td>
<td>2017</td>
</tr>
<tr>
<td>Average</td>
<td>1588</td>
<td>2045</td>
</tr>
<tr>
<td>Driver (1)</td>
<td>1430</td>
<td>2013</td>
</tr>
<tr>
<td>Driver (2)</td>
<td>1418</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (3)</td>
<td>1342</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (4)</td>
<td>1428</td>
<td>1860</td>
</tr>
<tr>
<td>Driver (5)</td>
<td>1351</td>
<td>1464</td>
</tr>
<tr>
<td>Driver (6)</td>
<td>1441</td>
<td>NA</td>
</tr>
<tr>
<td>Average</td>
<td>1394</td>
<td>1908</td>
</tr>
<tr>
<td>Day (2)</td>
<td>AM</td>
<td>PM</td>
</tr>
<tr>
<td>Run (1)</td>
<td>1645</td>
<td>2275</td>
</tr>
<tr>
<td>Run (2)</td>
<td>1887</td>
<td>2275</td>
</tr>
<tr>
<td>Run (3)</td>
<td>2150</td>
<td>2275</td>
</tr>
<tr>
<td>Run (4)</td>
<td>1945</td>
<td>1846</td>
</tr>
<tr>
<td>Run (5)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Run (6)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average</td>
<td>1460</td>
<td>1908</td>
</tr>
<tr>
<td>Driver (1)</td>
<td>1425</td>
<td>2013</td>
</tr>
<tr>
<td>Driver (2)</td>
<td>1420</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (3)</td>
<td>1389</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (4)</td>
<td>1306</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (5)</td>
<td>1510</td>
<td>1464</td>
</tr>
<tr>
<td>Average</td>
<td>1379</td>
<td>1908</td>
</tr>
<tr>
<td>Day (3)</td>
<td>AM</td>
<td>PM</td>
</tr>
<tr>
<td>Run (1)</td>
<td>1645</td>
<td>2275</td>
</tr>
<tr>
<td>Run (2)</td>
<td>1887</td>
<td>2275</td>
</tr>
<tr>
<td>Run (3)</td>
<td>2150</td>
<td>2275</td>
</tr>
<tr>
<td>Run (4)</td>
<td>1945</td>
<td>1846</td>
</tr>
<tr>
<td>Run (5)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Run (6)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average</td>
<td>1460</td>
<td>1908</td>
</tr>
<tr>
<td>Driver (1)</td>
<td>1425</td>
<td>2013</td>
</tr>
<tr>
<td>Driver (2)</td>
<td>1420</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (3)</td>
<td>1389</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (4)</td>
<td>1306</td>
<td>1864</td>
</tr>
<tr>
<td>Driver (5)</td>
<td>1510</td>
<td>1464</td>
</tr>
<tr>
<td>Average</td>
<td>1379</td>
<td>1908</td>
</tr>
<tr>
<td>Day (4)</td>
<td>AM</td>
<td>PM</td>
</tr>
<tr>
<td>Run (1)</td>
<td>1425</td>
<td>2013</td>
</tr>
<tr>
<td>Run (2)</td>
<td>1420</td>
<td>1864</td>
</tr>
<tr>
<td>Run (3)</td>
<td>1363</td>
<td>1864</td>
</tr>
<tr>
<td>Run (4)</td>
<td>1309</td>
<td>1464</td>
</tr>
<tr>
<td>Run (5)</td>
<td>1441</td>
<td>2013</td>
</tr>
<tr>
<td>Run (6)</td>
<td>1305</td>
<td>2023</td>
</tr>
<tr>
<td>Average</td>
<td>1379</td>
<td>1908</td>
</tr>
<tr>
<td>* Not Available</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Table 3.3 Descriptive Statistics of Average Route Travel Times (Seconds)

<table>
<thead>
<tr>
<th>Day (1) AM</th>
<th>Day (2) AM</th>
<th>Day (3) AM</th>
<th>Day (4) AM</th>
<th>Day (1) PM</th>
<th>Day (2) PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1945</td>
<td>1795</td>
<td>1862</td>
<td>1740</td>
<td>2707</td>
</tr>
<tr>
<td>Median</td>
<td>1889</td>
<td>1807</td>
<td>1858</td>
<td>1694</td>
<td>2696</td>
</tr>
<tr>
<td>Σ</td>
<td>323</td>
<td>350</td>
<td>293</td>
<td>262</td>
<td>538</td>
</tr>
<tr>
<td>CoV</td>
<td>17%</td>
<td>19%</td>
<td>16%</td>
<td>15%</td>
<td>20%</td>
</tr>
<tr>
<td>Minimum</td>
<td>1441</td>
<td>1305</td>
<td>1306</td>
<td>1309</td>
<td>2079</td>
</tr>
<tr>
<td>Maximum</td>
<td>2659</td>
<td>2482</td>
<td>2413</td>
<td>2153</td>
<td>3835</td>
</tr>
<tr>
<td>Count</td>
<td>27</td>
<td>26</td>
<td>26</td>
<td>24</td>
<td>20</td>
</tr>
</tbody>
</table>

* Not Available
The following observations can be inferred from Table 3.3:

1. Mean a.m. route travel times are always less than mean p.m. travel times
2. A.M. route travel times have less variability (CoV=17%-19%) than p.m. route travel times (CoV=20%-23%).
3. There are consistencies/similarities in route travel times of the a.m. peak for the four days involved. A similar conclusion can be reached for p.m. data.

*Average Link Travel Times*

The survey route included 55 travel time sections. According to Hellinga and Fu (1999), the statistical properties of travel time distribution is dependent on the direction from which vehicles enter/exit the segment. Therefore, and for consistency purposes, only 31 “through” travel time sections were considered for the purpose of this research (i.e. link entry and exit are “through” movements). Figure 3.4 shows the section ID of each travel time section.
The average travel time of each segment was computed using the data collected from all drivers entering the survey route within the same time interval. In total, 31 data samples were collected. Table 3.4 illustrates the statistical properties of the travel times of the sections analyzed.

Table 3.4 Descriptive Statistics of Average Link Travel Times (Seconds)

<table>
<thead>
<tr>
<th>TT_Sec_ID</th>
<th>Mean</th>
<th>Std Dev</th>
<th>CoV</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35.0</td>
<td>12.8</td>
<td>0.37</td>
<td>20.4</td>
<td>65.7</td>
</tr>
<tr>
<td>2</td>
<td>18.8</td>
<td>4.1</td>
<td>0.22</td>
<td>13.5</td>
<td>30.2</td>
</tr>
<tr>
<td>3</td>
<td>43.0</td>
<td>16.6</td>
<td>0.39</td>
<td>14.3</td>
<td>90.8</td>
</tr>
<tr>
<td>4</td>
<td>57.0</td>
<td>17.2</td>
<td>0.30</td>
<td>41.0</td>
<td>113.6</td>
</tr>
<tr>
<td>5</td>
<td>33.4</td>
<td>15.5</td>
<td>0.46</td>
<td>18.5</td>
<td>80.8</td>
</tr>
<tr>
<td>6</td>
<td>42.2</td>
<td>16.6</td>
<td>0.39</td>
<td>15.8</td>
<td>89.0</td>
</tr>
<tr>
<td>10</td>
<td>24.6</td>
<td>14.3</td>
<td>0.58</td>
<td>15.5</td>
<td>76.5</td>
</tr>
<tr>
<td>11</td>
<td>31.3</td>
<td>31.0</td>
<td>0.99</td>
<td>13.5</td>
<td>130.0</td>
</tr>
<tr>
<td>12</td>
<td>23.6</td>
<td>11.3</td>
<td>0.48</td>
<td>12.5</td>
<td>50.8</td>
</tr>
<tr>
<td>13</td>
<td>15.9</td>
<td>5.4</td>
<td>0.34</td>
<td>11.0</td>
<td>32.2</td>
</tr>
<tr>
<td>14</td>
<td>19.3</td>
<td>8.9</td>
<td>0.46</td>
<td>11.0</td>
<td>45.8</td>
</tr>
<tr>
<td>18</td>
<td>27.1</td>
<td>19.1</td>
<td>0.70</td>
<td>12.3</td>
<td>85.8</td>
</tr>
<tr>
<td>19</td>
<td>46.2</td>
<td>29.3</td>
<td>0.63</td>
<td>14.8</td>
<td>124.2</td>
</tr>
<tr>
<td>20</td>
<td>57.5</td>
<td>35.7</td>
<td>0.62</td>
<td>15.3</td>
<td>144.2</td>
</tr>
<tr>
<td>21</td>
<td>51.8</td>
<td>15.8</td>
<td>0.31</td>
<td>15.8</td>
<td>81.0</td>
</tr>
<tr>
<td>22</td>
<td>29.3</td>
<td>9.0</td>
<td>0.31</td>
<td>14.3</td>
<td>50.0</td>
</tr>
<tr>
<td>27</td>
<td>37.9</td>
<td>40.8</td>
<td>1.08</td>
<td>13.8</td>
<td>199.6</td>
</tr>
<tr>
<td>28</td>
<td>33.5</td>
<td>43.2</td>
<td>1.29</td>
<td>13.3</td>
<td>183.6</td>
</tr>
<tr>
<td>29</td>
<td>50.9</td>
<td>53.9</td>
<td>1.06</td>
<td>13.4</td>
<td>220.6</td>
</tr>
<tr>
<td>30</td>
<td>40.2</td>
<td>30.9</td>
<td>0.77</td>
<td>13.5</td>
<td>114.3</td>
</tr>
<tr>
<td>31</td>
<td>49.9</td>
<td>23.9</td>
<td>0.48</td>
<td>15.0</td>
<td>110.0</td>
</tr>
<tr>
<td>35</td>
<td>30.7</td>
<td>23.6</td>
<td>0.77</td>
<td>12.5</td>
<td>92.5</td>
</tr>
<tr>
<td>36</td>
<td>42.7</td>
<td>27.1</td>
<td>0.63</td>
<td>13.8</td>
<td>105.6</td>
</tr>
<tr>
<td>37</td>
<td>45.1</td>
<td>21.0</td>
<td>0.47</td>
<td>14.4</td>
<td>97.3</td>
</tr>
<tr>
<td>38</td>
<td>33.1</td>
<td>12.0</td>
<td>0.36</td>
<td>15.8</td>
<td>57.5</td>
</tr>
<tr>
<td>39</td>
<td>39.1</td>
<td>10.1</td>
<td>0.26</td>
<td>18.3</td>
<td>56.2</td>
</tr>
<tr>
<td>44</td>
<td>9.8</td>
<td>3.1</td>
<td>0.31</td>
<td>6.6</td>
<td>19.5</td>
</tr>
<tr>
<td>45</td>
<td>18.8</td>
<td>11.6</td>
<td>0.62</td>
<td>9.0</td>
<td>50.3</td>
</tr>
<tr>
<td>46</td>
<td>12.1</td>
<td>3.9</td>
<td>0.32</td>
<td>9.2</td>
<td>23.4</td>
</tr>
<tr>
<td>47</td>
<td>47.4</td>
<td>14.2</td>
<td>0.30</td>
<td>17.3</td>
<td>76.3</td>
</tr>
<tr>
<td>48</td>
<td>55.4</td>
<td>31.1</td>
<td>0.56</td>
<td>27.5</td>
<td>179.0</td>
</tr>
</tbody>
</table>
3.3 DOWNTOWN MICROSIMULATION MODEL

As mentioned earlier, two sources of data were used in this research. The first source, described in the previous section, is the real-life data. The second source of data is a VISSIM microsimulation model of the study area: downtown Vancouver. This model was developed in 2005 under the sponsorship of the City of Vancouver to test different scenarios associated with the implementation of the streetcar project in the downtown area (Ekeila 2005). The following is a brief description of the model development efforts and network coding of the model.

3.3.1 VISSIM Background

VISSIM is one of the most widely used simulation software packages and is gaining increasing interest among various parties in the transportation community. VISSIM is a stochastic time step microscopic simulation software package developed by PTV AG, Germany (VISSIM 5.1 User Manual). VISSIM is a behaviour-based simulation model that uses a Wiedemann psycho-physical car following logic to model traffic on urban streets and freeway environments. VISSIM is capable of simulating multi-modal traffic flows, including cars, trucks, buses, heavy rail, trams, LRT, bicycles, and pedestrians. It provides the flexibility to model any type of geometric configuration or unique driver behaviour encountered within the transportation system. VISSIM is versatile and provides the modeller with the ability to model a wide range of traffic operations in both the interrupted and uninterrupted traffic environment (VISSIM 5.1 User Manual). It has been used to analyze networks of varying sizes, ranging from individual intersections to entire metropolitan areas.

3.3.2 Network Coding

**Geometric Data**

Aerial photos were obtained from the City of Vancouver website, patched together, and used as the background for network development. The photos were scaled to have the correct coordinates of the roads and intersections within the network. For each link,
geometric data, including the number of lanes, width of each lane, and lane dedication, were collected. Using the scaled background and the information collected, the links were drawn to accurately match the real network. All of the defined links were classified as urban motorized links. Pedestrian paths were also modeled and classified as footpaths.

**Vehicle Volumes**

All vehicular traffic volume counts were collected from the Engineering Department of the City of Vancouver. Traffic counts were available for various years between 1992 and 2005. Many intersections had multiple counts in different years, while some had only single counts. Due to this variety in the traffic counts data, a systematic procedure was followed to summarize and prepare traffic volume data as model inputs. The procedure used was as follows:

1. If the configuration of the intersection had changed, all counts conducted before the configuration changes were eliminated.
2. Any counts conducted in 2001 were eliminated because of the transit strike that occurred that year.
3. If there was a significant difference between the old and new counts for the same intersection, then the old counts were not taken into consideration.
4. After following steps (1) to (3), the average of the remaining counts was computed.

**Pedestrian Volumes**

Significant pedestrian volumes exist in downtown Vancouver. Pedestrian volumes greatly affect the behaviour as well as the delays of left and right turning vehicles. They have a great influence, therefore, on the simulation output, and must be considered. Priority rules should be added to prevent any conflict between vehicular and pedestrian movements. Pedestrian counts were available for the years 1992 to 2004. The average of the available counts was computed for each intersection. Since pedestrian volumes accounted for the total number of pedestrians at each approach, they were split equally between the two opposing movements.
Traffic Signals

The simulation model incorporated both types of traffic signals: pre-timed signals, and actuated signals. Most of the traffic signals in the downtown area are pre-timed. Of 108 signalized intersections, 95 are pre-timed. All data on pretimed traffic signals in the network, cycle lengths, signal splits, phasing, lost times, and offsets were provided by the City of Vancouver Engineering Department. The data were used to model pre-timed signals in the VISSIM model.

Actuated traffic signals are signals with varying cycle lengths and phases. The signal controller is affected by current traffic demand on a cycle-by-cycle basis. Vehicle actuated programming was used to program the signal controller logic of actuated signals. Vehicle actuated programming is an optional add-on module of VISSIM for the simulation of programmable traffic actuated signal controls. In vehicle actuated programming, the control logic is described in a text file using a programming language (VISSIM 5.1 User Manual). The study area includes two types of actuated signals: semi-actuated signals (10 intersections) and pedestrian actuated signals (3 intersections). A comprehensive data set of the cycle lengths, phasing, vehicle green extension, maximum green extension, Flash Don’t Walk (FDW) timing, and other parameters associated with modelling actuated signals was provided by the City of Vancouver and was used to model both signal types.

Transit Lines

Forty-six transit lines were created according to 2005 routes and schedules. The South Coast British Columbia Transportation Authority (Translink) website was consulted to obtain route information for each transit line that runs through downtown, along with all corresponding bus stops. After defining transit routes and bus stops in the network, dwelling times were defined for each transit stop. Based on a study undertaken by the City of Vancouver on dwelling time distribution, it was found that dwelling times follow a normal distribution with a mean of 15 seconds and a standard deviation of 3 seconds. This distribution was used for all transit lines in the network.
3.3.3 Model Update

Careful revisions and updates were required to improve the earlier simulation model. The revision and update process included: analysis period, geometric design elements, transit lines, movement restrictions, priority rules, speed profiles, and HOV lanes.

Analysis Period

The a.m. simulation model was modified to deal with the peak period between 7:00 a.m. to 8:00 a.m. The a.m. peak period was previously defined as ranging between 6 a.m. and 9 a.m. However, due to varying traffic regulations during these three hours, it was decided to only analyze the period between 7:00 a.m. and 8:00 a.m., as the traffic regulations do not change within this time period.

Geometric Elements

All of the geometric elements of the model, including number of lanes, lane width, crosswalk locations, etc. were revised and modified according to 2008 network characteristics. Recent data were collected using aerial photos from Google Earth™ and a network survey that incorporated driving on all links of the analyzed network with manual recording of link geometries.

Transit Lines

The forty-six transit lines were updated according to the most recent routes and schedules as of December 2008. All new bus stops were identified including those that a bus could skip and those at which a bus had to stop. Many of the transit lines ended at midblock locations. In VISSIM, if the end of the transit route is left as it is in real-life, the transit vehicle will float on random routes until the end of the simulation period (VISSIM 5.1 User Manual). To account for this, dummy links were created and linked to the end point of the bus route to remove the buses from the network just after the destination station (See Figure 3.5).
Movement Restrictions

The Engineering Department of the City of Vancouver was consulted with respect to turning movement restrictions in the downtown core during the a.m. peak period. A detailed map was provided showing all movement restrictions in the downtown area during the analysis period. These restrictions were extracted from the map and applied to the model. Figure 3.6 shows a detailed map of movement restrictions in downtown Vancouver.

Priority Rules

Priority rules in VISSIM represent the logical safety rules that prevent any collision between two road users (i.e. two vehicles, a vehicle and a pedestrian). More than 2000 priority rules were created and added to the old model to account for different vehicular and pedestrian movement conflicts and to create “Keep Clear Zones” mid-intersection.
Speed Profiles

The desired speed distribution is one of the most influential parameters of a VISSIM simulation model. The speed limit in downtown Vancouver is 50 km/hr. Commuters usually drive at a speed slightly less than 50 km/hr because of the closely spaced signalized intersections which impede their ability to speed. Accordingly, a desired speed profile of 40-50 km/hr was used for cars in the simulation model, while a speed profile of 30-40 km/hr was used for HGV vehicles.

HOV Lanes

In Metro Vancouver, four types of HOV lanes exist: HOV 2⁺, HOV 3⁺, Bus/Bike & Authorized Vehicles and Bus/Bike Only. The first two types of HOVs do not exist in
downtown Vancouver. Table 3.5 shows the locations of the other two types of HOV lanes and their regulation periods (hours and days).

Table 3.5 Downtown HOV Lanes

<table>
<thead>
<tr>
<th>Street Section</th>
<th>Bus/Bike &amp; Authorized Vehicles</th>
<th>Bus/ Bike Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB Granville</td>
<td>24 HOURS</td>
<td></td>
</tr>
<tr>
<td>Robson to Georgia &amp; Pender to Hastings</td>
<td>(Closed for construction)</td>
<td></td>
</tr>
<tr>
<td>NB Granville</td>
<td>24 HOURS</td>
<td></td>
</tr>
<tr>
<td>Smithe to Hastings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB Howe</td>
<td>7AM– 7PM</td>
<td>Every Day</td>
</tr>
<tr>
<td>Robson to Drake</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB Seymour</td>
<td>7AM – 7PM</td>
<td>Every Day</td>
</tr>
<tr>
<td>Drake to Hastings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EB &amp; WB Pender</td>
<td>7-9:30AM, 3-6 PM</td>
<td>Mon-Fri</td>
</tr>
<tr>
<td>Cambie to Howe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB Burrard</td>
<td>7–9:30 AM</td>
<td>Mon-Fri</td>
</tr>
<tr>
<td>Pacific To Dunsmuir</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Emergency Vehicles, Commercial Vehicles With Permits, Taxis and Limousines

All lane segments defined in Table 3.5 were closed to vehicular traffic using the “lane closure” option of VISSIM, and open for use by transit vehicles only.

3.3.4 The Dynamic-based Simulation Model

One shortcoming of the earlier developed simulation model is that it involved using static assignment (routing decisions) based on the traffic counts and turning volumes of downtown intersections. This approach is not appropriate for medium/large networks where many routes between each Origin-Destination (OD) pair exist. Hence, it was necessary to convert the earlier static simulation model into a dynamic assignment model. The static simulation model was converted into a dynamic one by adding the necessary modelling elements such as zonal parking lots and network nodes. The calibration of the OD matrix and other parameters of the simulation model is presented in the following sections.
3.3.5 Calibration and Validation of the Microsimulation Model

Objective and Purpose

The objective of the calibration and validation of the microsimulation model of downtown Vancouver is to test several issues associated with the proposed neighbouring links travel time estimation. These issues include the impact of probe vehicles sample size on estimation accuracy, using transit as probes for neighbour links travel time estimation, and the fusion of passenger probes and transit vehicle data. There were inherent difficulties in using real-life data to test these issues. In this chapter, a general calibration and validation methodology for medium-size simulation models is presented. The methodology is applied to downtown Vancouver, the study area for this thesis.

Introduction

Microsimulation has been used by many practitioners and researchers for several years as a cost-efficient approach for testing transportation-related solutions and designs before implementation. Applications of these models include testing and evaluating new signal optimization strategies (Kesur 2009, Kim et al. 2009), analysis of route choice behaviour (Talaat et al. 2007), route divergence analysis (Quayle and Urbanik 2008), comparisons of unconventional intersection designs (El Esawy and Sayed 2007, 2009), and testing of new transit signal priority algorithms (Ekeila et al. 2009). Examples of commercially available and widely used microsimulation packages include VISSIM, CORSIM, AIMSUN, SimTraffic, Paramics, INTEGRATION, among others.

Acknowledging the importance of calibrating and validating microsimulation models, a considerable body of literature exists to address several issues related to the model’s development, calibration, and validation of different microsimulation models. In general, the current research can be classified into: small network models that deal with an isolated intersection or a corridor that includes a series of intersections (Park and Schneeberger 2003, Park and Qi 2005, Kim et al. 2005, Park et al. 2006, Choi et al. 2009), medium-sized networks of up to 100 intersections (Oketch and Carrick 2005, Balakrishna et al. 2007a) and large networks (Jha et al. 2004, Ambadipudi et al. 2006, Balakrishna et al. 2007b).
Small network models are usually easy to calibrate as their use involves static routing decisions and therefore their running times are relatively short. This removes the complexity and the necessity to calibrate OD matrices and route choice parameters, which are basic calibration requirements for medium/large networks. In the calibration efforts of small network models, the focus is usually on driver behaviour and lane change parameters. A set of calibration parameters is first identified, along with their ranges. An experiment is consequently designed to run different scenarios (i.e. combinations of calibration parameters). Some Measures of Effectiveness (MOE) are defined, such as average queue length, travel time, etc. and then used to compare the simulation measured values of each scenario to field observations. An optimization algorithm is used to determine the best set of parameters to make the model output match observations.

Calibration of medium/large scale networks on the other hand introduces difficulties and challenges. These models usually employ Dynamic Traffic Assignment (DTA) and therefore require the calibration of an OD matrix, and route choice parameters, in addition to other driver behaviour and lane changing parameters. In general, the existing calibration efforts of large scale networks focus only on OD matrix calibration and route choice (Jha et al. 2004, Oketch and Carrick 2005, Balakrishna et al. 2007a, 2007b). Manual adjustments are usually applied to other parameters (Ambadipudi et al. 2006). This was always justified by the long running time that might be needed to run the simulation model until convergence, which does not allow for an experimental design to be used to test different case scenarios.

The simulation of signalized networks involves several complexities compared to other roadway facilities, mainly because of the interrupted nature of the traffic flow. The complexity in model development can increase as the modeller remains keen on maintaining a high level of detail in the model. To illustrate, downtown areas are usually highly compacted with closely spaced intersections. This can cause a number of potential problems in vehicle routing, especially under high volume conditions. Other difficulties include modelling transit routes, stops, and exclusive/shared lanes. Pedestrian movements in downtown areas are usually significant and they have a considerable impact on the model output, and, hence, should not be neglected.
Although many calibration and validation procedures are available in the existing literature, a procedure to calibrate and validate medium-sized urban networks has not yet been well developed. In this chapter, a procedure for the calibration and validation of medium-sized simulation models is presented. A case study is applied to a VISSIM model of downtown Vancouver, British Columbia. The motivation behind this research is to apply a calibration effort that combines the two state-of-the-art approaches currently used for calibrating small and large network simulation models. This method considers the tradeoffs between time/cost constraints, calibration efforts, and the required accuracy.

Practitioners require a robust and efficient calibration procedure that suits their needs. Researchers need a well-developed simulation model upon which they can rely to test new methodologies and algorithms. The procedure presented in this chapter is intended to aid both researchers and practitioners interested in calibrating and validating medium-sized microsimulation models in general, and VISSIM models, in particular.

*Model Calibration Methodology*

The proper application of microsimulation models to derive a critical design decision depends on their accuracy and reliability. Accordingly, a proper procedure is required for the calibration and validation of these models. In the absence of such a procedure, the accuracy of the model will always be questionable. The calibration of medium/large scale microsimulation models involves OD matrix calibration, route choice parameter calibration, and driver behaviour parameters calibration. However, the last two stages are usually neglected and focus is only directed on OD matrix calibration. The proposed calibration procedure is a sequential procedure in which the modeller does not proceed to the following calibration stage unless necessary. First, the OD matrix is calibrated and the assigned traffic volumes of the model are compared to observed real-life volumes. If a close match is found between the two sets, the calibration of route choice can be skipped. Similarly, if the observed real-life MOEs such as travel times, match the simulation values, the calibration of driver behaviour and lane changing parameters can be also skipped. A general overview of the calibration and validation procedure is illustrated in Figure 3.7.
Figure 3.7 The Proposed Calibration and Validation Procedure
The following paragraphs describe the sequential method used for model calibration and validation.

**OD Matrix Calibration**

The purpose of OD matrix calibration is to obtain a travel demand matrix between each Origin-Destination pair in the study area during the analysis period. Obtaining such a matrix, using surveys, can prove very costly. Alternatively, OD matrices can be estimated from observed traffic counts and a seed/partial matrix. Simulation software packages are usually accompanied by a tool to estimate the OD demand matrix from static traffic counts and a seed matrix. For example, VISUM uses a matrix estimation module called “TFlow fuzzy,” while the equivalent module of Paramics is called “Estimator” (Oketch and Carrick 2005). In the absence of such modules, the analyst can use one of the well-developed mathematical methods which exist in the transportation engineering literature. See for example, methods of Cascetta and Nguyen (1988), Cascetta et al. (1993), Ashok and Ben-Akiva (2000), and Cascetta and Postorino (2001). For the scope of the current study, which uses a VISSIM model, the TFlow fuzzy module of VISUM was used for OD matrix estimation.

**Route Choice Parameters Calibration**

A subsequent step to OD matrix calibration is to verify the similarity between observed real-life traffic counts and simulation assignment volumes. If a close matching is found, the calibration of route choice parameters would not be needed and the analyst could skip this step. However, if such a close matching is not found, calibration of route choice parameters is mandatory. Calibration of route choice parameters is meant to change some attributes, such as link travel time, so as to change the results of traffic assignment. The objective is to reach a reasonable matching between observed and simulation traffic volumes. The similarity criteria between both volumes have to be defined along with their acceptable thresholds. A good starting point is to visually inspect the scatter plot of observed versus simulated volumes with respect to a 45° line representing a perfect fit (Jha et al. 2004, Ambadipudi et al. 2006, Balakrishna et al. 2007b). Other quantitative similarity measures include:
- Geoffrey E. Havers Statistic (GEH) which has been widely used to test observed versus simulated link volumes (FHWA guide 2004, Oketch, and Carrick 2005, Balakrishna et al. 2007b):

\[
GEH = \sqrt{\frac{2(Volume_{RL} - Volume_{Sim})^2}{Volume_{RL} + Volume_{Sim}}}
\]  
(3.2)

- The Mean Absolute Percent Error (MAPE), calculated as:

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Volume_{RL} - Volume_{Sim}}{Volume_{RL}} \right|
\]  
(3.3)

- % Error of Volumes Sum, calculated as:

\[
% \text{Error of Volumes Sum} = \left| \frac{\sum_{i=1}^{N} (Volume_{RL}) - \sum_{i=1}^{N} (Volume_{Sim})}{\sum_{i=1}^{N} (Volume_{RL})} \right|
\]  
(3.4)

- The Root Relative Squared Error (RRSE), calculated as:

\[
RRSE = \sqrt{\frac{1}{\sum_{i=1}^{N} Volume_{RL}} \cdot \sum_{i=1}^{N} \left( \frac{Volume_{RL} - Volume_{Sim}}{Volume_{RL}} \right)^2 \cdot Volume_{RL}}
\]  
(3.5)

- The Normalized Root Mean Square Error (RMSN), calculated as (Balakrishna et al. 2007b):

\[
RMSN = \sqrt{N \cdot \frac{\sum_{i=1}^{N} (Volume_{RL} - Volume_{Sim})^2}{\sum_{i=1}^{N} Volume_{RL}}}
\]  
(3.6)

- The Sample Correlation Coefficient (r), calculated as:

\[
r = \frac{N \cdot \sum_{i=1}^{N} (Volume_{RL} \cdot Volume_{Sim}) - \sum_{i=1}^{N} (Volume_{RL}) \cdot \sum_{i=1}^{N} (Volume_{Sim})}{\sqrt{N \cdot \sum_{i=1}^{N} (Volume_{RL})^2 - (\sum_{i=1}^{N} Volume_{RL})^2} \cdot \sqrt{N \cdot \sum_{i=1}^{N} (Volume_{Sim})^2 - (\sum_{i=1}^{N} Volume_{Sim})^2}}
\]  
(3.7)

Where:

\[Volume_{Sim} = \text{Simulation volume}\]

\[Volume_{RL} = \text{Observed real-life volume}\]

\[N = \text{Number of validation observations}\]
For a comprehensive literature review of different criteria used in the calibration of microsimulation models, readers are directed to consult Hollander and Liu (2008).

**Calibration Targets**

The FHWA Guide (2004) includes examples of calibration targets that were developed by the Wisconsin Department of Transportation (DOT). Although these targets were developed and used specifically for freeways, some of them have been widely applied in simulation models of urban areas (Oketch, and Carrick 2005, Balakrishna et al. 2007, Choi et al. 2009). The FHWA Guide (2004) suggests that calibration targets can vary according to the potential use for the model, the available resources of the analyst, and time constraints. Due to the extensive size of the modeled network and the challenges in achieving these criteria, the calibration targets used in this research were relaxed in order to reduce calibration efforts and time.

Noteworthy is that some of these criteria are already incorporated as convergence criteria in some simulation software packages, such as Paramics. However, in VISSIM, none of these are included as a modeller input. Therefore, with each change in route choice parameters, the model has to be run until convergence, and assigned volumes have to be compared to observed volumes.

**Driver Behaviour Parameters Calibration**

1. **Measures Of Effectiveness (MOE)**

The purpose of driver behaviour parameters calibration is to fine-tune a subset of driver behaviour parameters so that the model output matches field observed data (Hollander and Liu 2008). MOEs that can be collected from the field and generated from the simulation models have to be identified. These measures/variables (e.g. travel time) can be used for both driver behaviour parameters calibration and model validation. The extent (i.e. period and sample size) of data collection for model calibration is constrained by budget and time. The collected data set should be split into two sets: one for driver behaviour parameters calibration and one for model validation. This split can be either temporal or spatial. For example, if the data set is collected for two days, data from the first day can be used for the
calibration, while data from the second day can be used for the validation. Alternatively, if
the data were collected from many entities of the network (e.g. a number of corridors), data
from some entities (e.g. two corridors) can be used for the calibration, while data from the
other entities, can be used for the validation. A subsequent step, after defining the MOEs,
is to compare the observed real-life values to simulation values. If a close matching is
realized for both values, the calibration of driver behaviour parameters can be skipped.
However, if no such matching is found, it becomes necessary to calibrate some of the
driver behaviour parameters.

A small set of driver behaviour parameters should be carefully chosen and fine-tuned in
order to provide a close matching between observed and simulated values. The selection of
these parameters and their ranges is usually based upon engineering judgment, modeller’s
experience with the simulation software, and consideration of the previous literature.
Miller (2009) developed a procedure to identify the most influential calibration parameters
and eliminate other parameters. This procedure, however, is lengthy and may be inefficient
in terms of modelling large-scale networks.

2. Case Scenarios (Experimental Design)

Many experimental designs can be employed to obtain a sample that comprises different
combinations of calibration parameter values. Among these techniques, the Latin
Hypercube (LH) sampling technique (McKay and Beckman 1979), is considered an
efficient sampling method. It has been used in previous research to design similar
experiments (Park and Schneeberger 2003, Park and Qi 2005, Park et al. 2006). In this
sampling technique, the sample space of each input variable is randomly stratified into
partitions of equal marginal probabilities. In other words, the range of each variable $X_i$, is
divided into $N$ partitions of equal marginal probability $1/N$. Each variable is sampled once
from each of the $N$ strata. The variables are then randomly paired to form $N$ samples. This
technique, assuming that all variables are independent, provides a good sample, stratified
uniformly across the entire domain of all the variables (McKay and Beckman 1979). The
choice of an appropriate sample size is always a difficult question to answer. However, the
modeller can start with a small sample, of 50, for example, and test whether the sampled
scenarios cover the space of the observed MOE. If the sample does not include a scenario that closely matches the observed MOE, the ranges of the calibration parameters have to be altered and another sample design should be carried out to incorporate a larger sample.

Fine-tuning of driver behaviour parameters can, in some cases, be implemented manually (Merritt 2004, Shaaban and Radwan 2005). Hollander and Liu (2008) suggested that manual calibration should be limited to cases where the expected application of the simulation model is of a very limited scale. Alternatively, many calibration algorithms have been reported in the literature, including linear surface functions (Park and Schneeberger 2003), genetic algorithms (Kim et al. 2005, Park and Qi 2005, Park et al. 2006, Liu et al. 2006c), heuristic optimization methods (Ma et al. 2007), and simultaneous perturbation stochastic approximation (Balakishna et al. 2007a, 2007b), among others. In almost all of these algorithms, an optimization problem is formulated where the objective function is to minimize the difference between the observed and the measured values of the calibration MOE, subject to the simulation output of each case scenario. In this study, the objective is not to demonstrate a new calibration algorithm, but rather to present a comprehensive methodology for the calibration and validation. It is proposed to start with a simple algorithm and test its ability to calibrate the model before using other mathematically more expensive algorithms. An example is an algorithm that selects the best combination of calibration parameters which minimizes the Sum of The Squared Errors (SSE) between the observed and the simulated MOE such that:

\[
\text{Minimize } \sum (\text{MOE}_{RL} - \text{MOE}_{Sim})^2
\]

To run the \( N \) scenarios, the analyst can model only a small section of the network and run the case scenarios for the small model, then use the calibrated parameters of the small model with the large model. Although this approach reduces running time, it requires the building of a new model for the small network, and neglects the interactions between other elements within the network. It was not used, therefore, in this analysis. An alternative approach, proposed in this study, is to convert the dynamic-based model into a static one. VISSIM uses an option called “Create Static Routing” which enables the simple conversion of dynamic models into static-based models.
Model Validation

The purpose of model validation is to compare the output of the calibrated model to real-life measurements that were not used in the calibration. The validation data set can include data of other entities of the model (e.g. other travel time corridors) or measurements of the same entities for different periods and/or traffic conditions (Park and Schneeberger 2003), or even using other MOEs for the same entities (e.g. average queue length). The model is said to be “valid” if the chosen MOE from the unused real-life data set is close enough to the simulation value. Otherwise, driver behaviour parameters calibration has to be re-executed.

Application to Downtown Vancouver Network

Model Calibration

1. OD Matrix Calibration

In Dynamic Assignment (DA), traffic demand is modeled by means of origin-destination (OD) matrices rather than static routing decisions. Each cell in an OD matrix includes the number of trips between an OD pair during a given time period. Traditionally, OD matrices were obtained from travel surveys. In this study, and due to the absence of such a matrix, another approach was used to obtain the OD matrix. Initially, a VISSIM dynamic model was converted into a VISUM file. VISUM is a macroscopic planning package developed by PTV AG Germany that can be integrated with VISSIM. VISUM uses a module called TFlow Fuzzy to update an old (i.e. seed) OD matrix using recent traffic counts or aggregate origin and/or destination volumes.

To obtain the seed OD matrix, a simple method that employs the microsimulation model itself was used. First, all origins and destinations in the network were identified. The simulation model included 36 zones. Travel time detectors were placed at the start of each origin link and at the end of each destination link for all OD pairs to estimate OD travel times and traffic volumes. At this point, travel times were neglected and traffic volumes between each OD pair were used to build a seed (i.e. initial) OD matrix. The obtained trip matrix was then linked to the pre-defined traffic composition of downtown (2% trucks).
Consequently, all vehicular static routing decisions were deleted except those for pedestrians and transit. Recent directional traffic counts were obtained from the City of Vancouver VanMap, which is a web-based GIS database that enables users to extract a variety of traffic and non-traffic related information through a simple interactive interface.

The seed matrix obtained from the static VISSIM model was exported to VISUM and was run to obtain an initial traffic assignment. The TFlow Fuzzy module uses both the initial assignment results and the new traffic counts to obtain a new-updated OD matrix. In this module, the observed field traffic counts are modeled as fuzzy sets, instead of exact values. This is justified by the fact that traffic counts at the same location will vary from one day to the next. Hence, sampling errors are taken into account by allowing the traffic volumes to vary between the lower limit (membership of zero) and the upper limit (membership of zero) with the observed value, given a membership value of 1 (i.e. triangular membership function). An appropriate fuzzy bandwidth has to be defined in order for the module to work. Too tight bandwidths do not allow the module to reach a solution. A large bandwidth on the other hand can generate significant matrix estimation errors. One advantage of the TFlow Fuzzy module is its ability to update the seed matrix without the need for counts on all links in the network. In this calibration effort, hourly traffic volume data of 70 sections were used to update the old OD matrix. The output of the OD matrix estimation is a new OD matrix with 18,174 OD trips for the 36 zones in the morning peak hour.

2. Priority Rules Calibration

A VISSIM model without priority rules will cause vehicle collisions. Priority rules control the conflicts between two types of movements, such as left turn with opposing through. Other simulation packages such as SimTraffic do not require the definition of priority rules. The creation of a priority rule in VISSIM requires the identification of three basic parameters: minimum gap time, minimum headway distance, and maximum speed. Previous studies that dealt with calibrating priority rules parameters used fixed values for all sets of priority rules in the model (Park and Qi 2005, Park et al. 2006). The appropriateness of using fixed values for all rules is questioned due to the wide range of
intersection geometries that necessitates developing exclusive priority rules for each intersection. Accordingly, each priority rule was individually calibrated and its performance was inspected visually by running the model and ensuring that no collisions occurred. Although this process is time consuming, no other method was found to be suitable. Calibrating these rules should not be neglected as they significantly impact model outputs.

**Dynamic Assignment Control**

The DA process in VISSIM is based on iterated simulations. Hence, a modeled network is simulated repetitively and drivers choose their routes based on route choice criteria that they have previously experienced during preceding simulations. This behaviour is rational as it represents the driver’s growing experience of the network when driving repetitively on different days. In this study, the utility function, which describes the individual’s personal utility of each route, was based on route travel time. In VISSIM’s DA module, a number of paths are defined for each OD pair. The travel times on these paths are updated every small time interval, called the evaluation interval. The evaluation interval was selected to be 900 seconds as suggested by VISSIM’s user manual. To decrease convergence time and reduce the number of possible paths between each OD pair, paths with total costs of more than 50% higher than the best path, were rejected from the route search. Several iterations are required to achieve convergence or dynamic assignment equilibrium. The criterion selected to achieve convergence, or dynamic assignment equilibrium, was travel time on paths, and the tolerance value was set to 10%. The Method of Successive Averages (MSA) was used to compute path travel times of the past four iterations and compare it to path travel times of the current iteration. Hence, convergence is achieved, if the difference in travel times on every path, between the previous four consecutive iterations and the current iteration for all evaluation intervals, is less than 10%.

**Route Choice Calibration**

The calibrated OD matrix was imported to the dynamic VISSIM model. The model was run until convergence using the pre-defined convergence criteria and constraints. For the
model to converge, 23 iterations were needed, which consumed about 16 hours on an Intel Core 2 desktop with 2 GB of memory and a processor speed of 1.8 GHz.

As described earlier, a number of calibration targets have to be defined to correspond to the specified evaluation criteria. The calibration targets set out in this research were:

- \( GEH < 5 \) for the sum of all traffic volumes of all comparison sections,
- \( GEH < 10 \) for at least 85% of comparison sections,
- \( MAPE < 10\% \), and
- \% Error of Volumes Sum < 10\%.

These evaluation criteria and acceptable limits were used to compare the assigned link traffic volumes to the observed volumes. Again, if these statistics indicated a good fit, there would have been no need for local calibration of route choice parameters. However, the fit results were not satisfactory, and it was decided to perform local calibration. In VISSIM, local calibration can be carried out by adding link surcharges (e.g. cost components) to change assignment results. These surcharges have to be positive, hence the local calibration can only be undertaken by adding surcharges to links that attract more traffic than observed. These segments were defined using a GEH threshold of more than 10. In the absence of any guidance on the relationship between the amount of surcharge and traffic divergence according to this surcharge, an ad-hoc iterative approach was used to achieve the calibration targets. Four rounds of local calibration were needed to reach an acceptable calibration. In the first round, a constant surcharge of 3 was added to all links that exhibited high simulation volumes. The assignment results did not change significantly and therefore a constant surcharge of 8 was used in the second round. In the third and fourth rounds, surcharges between 10 to 30 were added to some links that still attracted high traffic volumes. It should be noted that the GEHs of all segments were checked after each round of local calibration and only links that maintained high traffic volumes were surcharged. The VISSIM model was re-run after each local calibration run until convergence was achieved under the new local calibration surcharges. The scatter plots of observed traffic volumes versus the base case (i.e. no local calibration) denoted by \((Ys)_0\) and the four calibration rounds \((Ys)_1, (Ys)_2, (Ys)_3,\) and \((Ys)_4\) are shown in Figure 3.8. As shown in the Figure, the scatter of the points improves with each new calibration run.
The best fit regression line of the fourth run, \((Y_s)_4\), is the closest to the 45° line. The regression line of real-life volumes and the best calibrated model was:

\[
Volume_{RL} = 0.9926 \cdot Volume_{Sim} \quad \quad R^2 = 0.8806
\]

![Figure 3.8 Scatter Plots of Observed Volumes vs. Simulation Volumes](image)

Table 3.6 provides a summary of the results of other calibration statistics, as described earlier. As shown in the Table, only the last calibration run could pass the condition of \(GEH < 10\) for at least 85% of the validation sections. The Table also shows that all calibration targets were achieved with significant improvements over the base case. Consequently, this stage of model calibration was terminated and the results deemed satisfactory.

**Driver Behaviour Parameters Calibration**

As previously mentioned, using travel time as an MOE has been extensively reported in the literature (Park and Schneeberger 2003, Jha et al. 2004, Dowling et al. 2004, Park and Qi 2005, Kim et al. 2005, Liu et al. 2006c, Ambadipudi et al. 2006, Park et al. 2006). In this study, corridor travel time was selected as an MOE for model calibration and
validation. The data collected from the travel time survey, as described previously, were used to obtain travel times on four corridors in downtown Vancouver: Richards, Seymour, Howe, and Hornby Streets, during the a.m. period. Average travel time on each corridor was computed using the data collected from all drivers on the same day. The average travel times of the two corridors: Richards and Seymour Streets were used for driver behaviour parameters calibration, while data derived from the other two corridors were used for model validation.

Table 3.6 Summary of Route Choice Calibration Statistics

<table>
<thead>
<tr>
<th></th>
<th>(Ys)_0 (Base Case)</th>
<th>(Ys)_1</th>
<th>(Ys)_2</th>
<th>(Ys)_3</th>
<th>(Ys)_4</th>
<th>% Improvement (Rel. to Base Case)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEH (Sum)</td>
<td>13.8</td>
<td>10.7</td>
<td>6.4</td>
<td>3.9</td>
<td>0.3</td>
<td>98%</td>
</tr>
<tr>
<td>GEH ≥10</td>
<td>20</td>
<td>20</td>
<td>16</td>
<td>13</td>
<td>8</td>
<td>60%</td>
</tr>
<tr>
<td>5&lt;GEH≤10</td>
<td>23</td>
<td>22</td>
<td>26</td>
<td>26</td>
<td>27</td>
<td>17%</td>
</tr>
<tr>
<td>GEH≤5</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>31</td>
<td>35</td>
<td>30%</td>
</tr>
<tr>
<td>MAPE</td>
<td>11%</td>
<td>9%</td>
<td>6%</td>
<td>7%</td>
<td>7%</td>
<td>34%</td>
</tr>
<tr>
<td>% Error of Volumes Sum</td>
<td>7%</td>
<td>5%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>98%</td>
</tr>
<tr>
<td>RRSE</td>
<td>39%</td>
<td>40%</td>
<td>37%</td>
<td>30%</td>
<td>27%</td>
<td>30%</td>
</tr>
<tr>
<td>RMSN</td>
<td>41%</td>
<td>41%</td>
<td>38%</td>
<td>28%</td>
<td>25%</td>
<td>39%</td>
</tr>
<tr>
<td>Correlation Coefficient r</td>
<td>0.88</td>
<td>0.87</td>
<td>0.88</td>
<td>0.93</td>
<td>0.94</td>
<td>7%</td>
</tr>
</tbody>
</table>

The simulation average travel times of the so-far calibrated model (i.e. default driver behaviour parameters) were compared to the observed travel times. The Sum of Squared Errors (SSE) and the MAPE were computed and the results are presented in Table 3.7.

Table 3.7 Simulation and Real-Life Travel Times of the Calibration Corridors

<table>
<thead>
<tr>
<th>Default Driver Behaviour Parameters</th>
<th>Richards</th>
<th>Seymour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean TT_{RL} (Seconds)</td>
<td>94.0</td>
<td>246.3</td>
</tr>
<tr>
<td>STD (σ) (Seconds)</td>
<td>26.6</td>
<td>122.4</td>
</tr>
<tr>
<td>μ+σ (Seconds)</td>
<td>120.7</td>
<td>368.7</td>
</tr>
<tr>
<td>μ-σ (Seconds)</td>
<td>67.4</td>
<td>123.9</td>
</tr>
<tr>
<td>Mean TT_{sim} (Seconds)</td>
<td>70.8</td>
<td>90.5</td>
</tr>
<tr>
<td>SSE (Seconds^2)</td>
<td>539</td>
<td>24271</td>
</tr>
<tr>
<td>Sum SSE</td>
<td></td>
<td>24810</td>
</tr>
<tr>
<td>MAPE%</td>
<td></td>
<td>44%</td>
</tr>
</tbody>
</table>

Mean \( TT_{RL} \): Real-Life mean travel time

Mean \( TT_{sim} \): Simulation mean travel time
As shown in Table 3.7, the simulation average travel times of the two downtown streets are by far less than the real-life travel times. Furthermore, the mean simulation travel time of Seymour Street is less than mean real-life travel time minus one of its standard deviation ($\mu - \sigma$). Therefore, it was decided to fine-tune a small number of driver behaviour parameters to output travel times that were close to what was observed in the field.

VISSIM allows several driver behaviour parameters to be calibrated according to user experience and available resources. These parameters can be categorized as: car following parameters, lane changing parameters, lateral behaviour parameters, and signal control parameters. A description of different calibration parameters is not included here, and interested readers are directed to read VISSIM 5.1 User Manual.

Only a few parameters were hypothesized to affect simulation average travel times, and, therefore, were chosen for the calibration based on best engineering judgment and the previous literature. Essentially, these parameters were expected to change the travel times on calibration corridors without changing speed profiles. The ranges of calibration parameters were based on previous literature findings (Park and Schneeberger 2003, Park and Qi 2005, Kim et al. 2005, Park et al. 2006). These parameters are:

- Number of observed vehicles: default 4, calibration values: 3, 4, 5
- Average standstill distance: default 2 m, calibration range: 1-4 m
- Additive part of desired safety distance: default 2 m, calibration range: 1-4 m
- Multiplicative part of desired safety distance: default 3 m, calibration range: 2-5 m
- Minimum headway: default 0.5 m, calibration range: 0.5-2 m

This experimental design was constructed using the Latin Hypercube (LH) sampling technique, as described earlier. The experimental design included 100 scenarios (i.e. cases) which represented different combinations of the calibration parameters values. The dynamic model was converted into a static model using the “Create Static Routing” option of VISSIM. Noteworthy is the fact that the model could have been run in a dynamic mode while disabling some of the dynamic assignment options that caused the model to continuously generate the last assignment (i.e. convergence) results. However, it was found that this approach consumes more time than using static routing, and accordingly, the latter
approach was used. For each scenario, the average travel times of the two calibration corridors were extracted and compared to real-life average travel times.

**Calibration Results**

The case scenario that minimized the SSE was Scenario (50) which used the following values for calibration parameters:

- Number of observed vehicles: 5
- Average standstill distance: 3.33 m
- Additive part of desired safety distance: 3.83 m
- Multiplicative part of desired safety distance: 4.21 m
- Minimum headway: 1.44 m

The values of all calibration parameters were higher than the default values. Intuitively, it was expected that increasing the safety distance components and the headway would increase travel times on the corridors. These values seemed reasonable, as driving in compacted urban networks, such as downtown, will always require the driver to adopt more generous safety distances and headways to avoid rear-end collisions. Changing car following parameters was adopted as an alternative, and proved to be an even better approach to changing the speed profiles to unrealistic values to decrease travel times on the calibration corridors. Moreover, changing the speed profiles required re-running the entire model for a new convergence and new local calibrations of route choice parameters.

In terms of error measurements, the SSE and the MAPE were used to evaluate the accuracy of the calibrated model. Results are presented in Table 3.8 and they show significant improvements over the results of the default model.
Table 3.8 Results of Driver Behaviour Parameters Calibration

<table>
<thead>
<tr>
<th>Calibrated Driver Behaviour Parameters</th>
<th>Richards St.</th>
<th>Seymour St.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $T_{TRL}$ (Seconds)</td>
<td>94.0</td>
<td>246.3</td>
</tr>
<tr>
<td>STD ($\sigma$) (Seconds)</td>
<td>26.6</td>
<td>122.4</td>
</tr>
<tr>
<td>$\mu+\sigma$ (Seconds)</td>
<td>120.7</td>
<td>368.7</td>
</tr>
<tr>
<td>$\mu-\sigma$ (Seconds)</td>
<td>67.4</td>
<td>123.9</td>
</tr>
<tr>
<td>Mean $T_{TSim}$ (Seconds)</td>
<td>68.1</td>
<td>243.9</td>
</tr>
<tr>
<td>SSE ($\text{Seconds}^2$)</td>
<td>671.7</td>
<td>5.7</td>
</tr>
<tr>
<td>Sum SSE</td>
<td>677</td>
<td></td>
</tr>
<tr>
<td>MAPE%</td>
<td>14.5%</td>
<td></td>
</tr>
</tbody>
</table>

The simulation mean travel time of both streets is bounded by one standard deviation of the real-life measured travel times which indicates a close proximity between observed and simulated values. At this stage, model calibration was considered successfully completed and the final step was model validation.

Model Validation

Average travel times on Howe and Hornby Streets were reserved for model validation. Note that this data set was not used in any stage of the model calibration. The calibrated model was re-run using the driver behaviour parameters of Scenario (50) to estimate the average travel times on the validation corridors. Table 3.9 illustrates the validation results of the model. As shown in the Table, all validation results were satisfactory with minimal MAPE errors. Also, the simulation travel times were still bounded by one standard deviation of the real-life travel times. Taking into account the scale of the network and the complexities in reaching higher accuracies within an appropriate time frame, it can be concluded that the model was successfully calibrated and validated.

Table 3.9 Results of Model Validation

<table>
<thead>
<tr>
<th>Calibrated Driver Behaviour Parameters</th>
<th>Howe St.</th>
<th>Hornby St.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $T_{TRL}$ (Seconds)</td>
<td>117.6</td>
<td>203.9</td>
</tr>
<tr>
<td>STD ($\sigma$) (Seconds)</td>
<td>35.2</td>
<td>86.2</td>
</tr>
<tr>
<td>$\mu+\sigma$ (Seconds)</td>
<td>152.7</td>
<td>290.1</td>
</tr>
<tr>
<td>$\mu-\sigma$ (Seconds)</td>
<td>82.4</td>
<td>117.7</td>
</tr>
<tr>
<td>Mean $T_{TSim}$ (Seconds)</td>
<td>95</td>
<td>203.9</td>
</tr>
<tr>
<td>SSE ($\text{Seconds}^2$)</td>
<td>509.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Sum SSE</td>
<td></td>
<td>510</td>
</tr>
<tr>
<td>MAPE%</td>
<td></td>
<td>9.5%</td>
</tr>
</tbody>
</table>
Summary, Conclusions, and Lessons Learned

A comprehensive procedure for the calibration and validation of medium-scale simulation models was presented. The methodology was applied to a case study of the VISSIM downtown Vancouver network. The chapter presented some practical and theoretical issues that are essential for the development and application of microsimulation models. These concepts are beneficial for both practitioners and researchers. The calibration procedure included OD matrix calibration using existing traffic volumes, route choice calibration by manipulating link surcharges, and driver behaviour parameters calibration using an experimental design and multiple runs. It is suggested that the analyst proceeds to a subsequent stage of the calibration only if needed. For example, if the assignment results of the model were satisfactory, there would be no need for route choice parameters calibration. Similarly, if the model output of different measures of effectiveness closely matches real-life observations, there would be no need for the experimental design, nor multiple runs. The model could be calibrated and validated with a reasonable tradeoffs between accuracy and modelling efforts. The most difficult and time consuming stage of model calibration was local calibration of route choice. Unfortunately, the VISSIM user’s manual does not provide information on the use of surcharges and their quantitative impact. Hence, an iterative manual approach to test the impact of different surcharge values had to be used. Another challenge was to select the proper method for driver behaviour parameters calibration. Available choices included modelling a small part of the network only and running the case scenarios for the small model. This approach reduces running time, but requires the building of a new model for the small network and also neglects the interactions between other elements of the network. It was therefore rejected. The alternative approach used in this study, involved converting the model into a static model which is straightforward in VISSIM. Running the 100 scenarios required using three desktop computers for about 30 hours each. Larger scale models can be more difficult to calibrate due to increased running time. The proposed procedure defines the main stages of medium/large scale simulation models calibration and validation. The methodology allows the analyst to control many stages of the calibration according to available time and resources.
CHAPTER 4: A PROOF OF CONCEPT USING REAL-LIFE DATA

4.1 INTRODUCTION

It is important first to distinguish between travel time estimation and travel time prediction. The former refers to obtaining travel times using measurements of the current analysis period, while for the latter, travel time is projected into the future by some means of modelling. In other words, considering the analysis period $t$, collecting traffic information and using this information to calculate travel time during $t$, the process is referred to as travel time estimation. On the other hand, if this data is used to estimate travel time in a subsequent time period $t+1$, $t+2$...$t+n$, then the process is known as prediction or forecast.

Generally, available techniques for travel time estimation can be categorized according to their input data, input variables, or computational procedures. According to the input data, estimation models can either use simulation data or real-life data. Real-life data can be historical data, real-time data or a combination of the two. According to input variables, some models use travel time as both the input and the output variables, while other models use traffic parameters such as spot speeds, occupancy, queue length, vehicle headway and volumes. Analytical models have been developed to estimate travel time using measurements from a single loop detector (Guo and Jin 2006) or multiple loop detectors (Rakha and Van Aerde 1995). In many other analytical models, data from loop detectors and signal controllers are integrated (Skabardonis and Geroliminis 2005, Bhaskar et al. 2009, Liu and Ma 2009). Noteworthy is the fact that the data requirements for these models are intensive.

The objectives of this chapter are:

- To present a general framework for neighbour links travel time estimation,
- To validate the proposed framework using real-life data,
- To study the potential benefits (i.e. improved accuracy) of using several modelling techniques within the proposed framework. Focus is given to statistical regression models, Artificial Intelligence (AI) methods, and Non-Parametric (NP) models, and
To study the use of corridor travel time in estimating the travel time of nearby corridors that represent alternative routes.

4.2 A FRAMEWORK FOR NEIGHBOUR LINKS TRAVEL TIME ESTIMATION

A framework is proposed to estimate real-time dynamic travel times on links with no sensors, using data from neighbour links. There are two sources for real-time travel time estimation. The first is historical link travel time data, and the second is real-time data from the link itself, or from neighbour links. In the absence of real-time data, the analyst may rely on historical records. However, if both types of data are available, a data fusion scheme can be applied to make use of the two sources. Let:

\[ x_{hl} = \text{Historical average travel time of link } l \text{ during a measurement period } t \]

\[ x_{rl} = \text{Real-time average travel time of link } l \text{ during } t \]

\[ x_{rn} = \text{Estimated real-time average travel time of link } l \text{ using data of neighbour links } n \text{ during } t \]

\[ \hat{x} = \text{The best estimate of real-time average travel time of link } l \text{ during } t \]

To estimate real-time travel time for link \( l \), the following cases can take place:

1. There are travel time sensors (e.g. a sufficient sample of probe vehicles) on link \( l \) during \( t \), then \( \hat{x} = x_{rl} \)
2. No sensors are available on link \( l \) during \( t \), while a set of neighbours \( n \) are equipped with sensors, then \( \hat{x} = f(x_{hl}, x_{rn}) \)
3. No sensors are available on link \( l \) or on its neighbours \( n \) during \( t \), then \( \hat{x} = x_{hl} \)

The focus of this research is on case (2), as cases (1) and (3) are straightforward. The research framework can be described by the following four modules:

1. Identification of link neighbours
2. Choice of a modelling technique
3. Application entity
4. Source of neighbour links travel time data

In the next subsections, a description of each of these modules is presented.
4.2.1 Identification of Link Neighbours

In this research, the term “neighbours” refers to nearby links that have similar characteristics and are subject to comparable traffic conditions within the network. In general, the choice of link neighbours should be according to many criteria which include: 1) area type and location, 2) road class, 3) traffic control level, and 4) travel time correlation. Appropriate choice of boundaries for the study (i.e. impact) area is essential to ensure that all links are subject to similar traffic conditions. Further segments not affected by demand changes within a specific area, should be excluded. Similarity in road class classification is important to ensure consistency in the speed profiles. For example, in many cases, a major arterial segment cannot be chosen as the neighbour for a local street. Traffic control level is also a major determinant in defining link neighbours. Travel times of segments with signalized intersections will be distinctly different from segments with stop signs. Finally, strong travel time correlation between nearby links of the same class and operational characteristics, facilitates developing accurate estimation models.

Pearson Correlation Coefficient \( r \) is used to quantify the travel time association between nearby links. Although Pearson’s correlation is a measure of linear dependency, it is still considered a good indicator of association even if the relationship is better represented by a curvilinear function. Moreover, for a clear nonlinear relationship, Pearson’s correlation understates the true association or correlation. This means that if a moderate correlation is found between two variables within an actual nonlinear relationship, then the true correlation is higher. Hence, the true association will always be greater than Pearson’s \( r \).

The \( p \)-value is used to test the null hypothesis of zero correlation, \( H_0: r = 0 \). If \( p < 0.05 \), then the probability of having a positive correlation is > 97.5%. Fisher transformation is used to normalize the distribution of the correlation coefficient and stabilize its variance. It is also used to construct confidence intervals for correlation coefficients. Consequently, an arbitrary correlation threshold is used to define link neighbours. This threshold should achieve a reasonable trade-off between estimation accuracy and number of neighbours. Relaxing this value will increase the number of defined neighbours, however it may impact the estimation accuracy, and vice versa. This concept can be viewed as a nonparametric approach, whereby the correlation represents the distance metric. If a correlation cut-off
value is used, then the problem is similar to a Kernel model. However, if a pre-set number of neighbours is defined, the problem becomes similar to a K-Nearest Neighbour (KNN) model. A sensitivity analysis of the impact of the correlation threshold on the estimation accuracy is presented later in this chapter.

Although it might be hypothesized that the use of spatial correlation could be more appropriate, it was decided to use the widely known Pearson correlation to identify link neighbours. Spatial correlation is used to describe the strength of the locational relationship between two entities. This does not necessarily mean a strong travel time relationship, although in many cases, it does. Also, spatial correlation is most commonly used with point measurements, rather than segment measurements. According to Eom et al. (2006), geostatistical methods, including spatial modelling, are applied to point-referenced data, where a variable of interest is measured at particular locations within a fixed spatial region.

Another approach to identify link neighbours is to use a neuro-fuzzy model that employs a dimension reduction technique. Data from all nearby links are used as input variables. The model starts from a simple architecture that incorporates only one variable and it builds upon as necessary. The B-Spline Associative Memory Networks (AMN) models adopted in this research use the Analysis Of Variance (ANOVA) (Brown and Harris 1995) as a reduction (i.e. decomposition) technique. The ANOVA representation allows the input/output mapping to be approximated using a limited number of “subnetworks” of much lower dimensions.

4.2.2 Choice of a Modelling Technique

Having defined link neighbours, models that relate target link travel times to travel times of neighbour links should be developed. A straightforward approach is to use a model that involves the unknown travel time as the dependent variable (i.e. response) and neighbour links travel times as the independent variables (i.e. predictors). This approach could be beneficial when the modeller obtains travel time data from the same neighbours at each time interval. This case corresponds to collecting data from fixed sensors such as AVI systems or moving sensors with fixed routes such as buses. However, if the system collects the data from a sample of probe vehicles that float randomly in the traffic stream, it would
be impractical to use a single model. Also, when using regression models, neighbour travel times correlation can cause inter-collinearity between explanatory variables, which in turn will weaken the overall predictive ability of the model.

Several modelling techniques are explored within the framework to select the technique that best suits the concept. The modelling techniques used in this study comprise a statistical method, a Multi-Layer Perception (MLP) Artificial Neural Network (ANN) model, a Neuro-fuzzy B-Spline model, and a Non Parametric (NP) K-Nearest Neighbours (KNN) model. These models were widely used in the literature of travel time estimation and hence were chosen for the current analysis. They represent different modelling families including parametric statistical models, Artificial Intelligence (AI) models, and non-parametric statistical models. Details of each of these methods/models are presented in the following sub-sections.

The Statistical Method

To overcome the problems of inter-collinearity and lack of data on some neighbours, a family of regression models can be developed to relate travel times of each pair of neighbour links independently (i.e. link to link models). Each model should incorporate one dependant and one explanatory variable only and both of these variables are travel times. This can be beneficial when limited travel time data are available, so that the travel times of only one neighbour link can be used. In this thesis, simple parametric statistical models are used. Regression models are developed using the least squares method. The number of models to be developed depends on the number of defined neighbours for each link in the study area.

Note that the developed models need to be dynamic to account for traffic condition variability. The strength of the travel time correlation between links is likely to change with time which necessitates the development of different models for different time periods. (e.g. morning peak, afternoon peak, etc.). A higher level of disaggregation can be attained by developing these models for each data collection interval (e.g. 15 minutes).
Developing link to link models can solve the problems of inter-collinearity and lack of data on some neighbours. Nevertheless, in many cases, travel time data might exist on more than one neighbour link and, therefore, more than one model can be used. This raises the question of how to combine estimates of different models to obtain a robust single estimate of the unknown travel time. More specifically, how should weights be assigned to each model estimate? Simple weighting schemes were applied to combine estimates of different models. The chosen schemes included straight average, weights by \( R^2 \), weights by travel time correlation \( r \), weights by the inverse of model’s variance \( (\sigma^2) \), weights by the exponent of travel time correlation \( r \), and weights by the exponent of the inverse of model’s variance \( (\sigma^2) \). The first scheme - equal weights, neglects statistics from individual models and assumes that they all result in the same level of accuracy. Applying weights using the model’s coefficient of determination \( (R^2) \) accounts only for the degree of variability explained by the model, while using the variance to assign weights accounts for the reliability of the travel time estimate for each model. Noteworthy is that using \( R^2 \) or model variance for weighting may overemphasize the significance of the statistical models. Using correlation to weight estimates of different models is hypothesized to be more appropriate as it accounts for the strength of the travel time relationship. The weighting schemes are expressed mathematically as:

**Straight Average:**

\[
    w_i = \frac{1}{n} \quad (4.1)
\]

**Coefficient of determination (\( R^2 \)):**

\[
    w_i = \frac{R_i^2}{\sum_{i=1}^{m} R_i^2} \quad (4.2)
\]

**Correlation coefficient (\( r \))**

\[
    w_i = \frac{r_i}{\sum_{i=1}^{m} r_i} \quad (4.3)
\]
Model variance ($\sigma^2$):

$$w_i = \frac{1}{\sum_{i=1}^{n} \frac{1}{e^{\sigma_i^2}}}$$  \hspace{1cm} (4.4)

Exponent of correlation ($e'$):

$$w_i = \frac{e'^i}{\sum_{i=1}^{n} e'^i}$$  \hspace{1cm} (4.5)

Exponent of model variance ($e^{\sigma^2}$):

$$w_i = \frac{1}{\sum_{i=1}^{n} e^{\sigma_i^2}}$$  \hspace{1cm} (4.6)

And

$$\sigma_i = \sqrt{\frac{SSE_i}{df}}$$  \hspace{1cm} (4.7)

Where:

$$\sum_{i=1}^{n} w_i = 1 \text{ and, } w_i = 0 \text{ if no travel time records exist on neighbour } i \text{ at time } t$$

$$SSE_i = \text{sum of squared errors of model relating unknown link travel time to neighbour } i$$

$$df = \text{degree of freedom} = \text{Historical data sample size} - \text{number of parameters to be estimated}$$

$$n = \text{Number of neighbour links}$$

A simple weighting formula can be used to calculate $\hat{x}$ as in case (2), such that:

$$\hat{x} = \alpha x_{rn} + (1 - \alpha) x_{hl}$$  \hspace{1cm} (4.8)

Where $\alpha, (1 - \alpha)$ are weights for $x_{rn}$ and $x_{hl}$, respectively, $0 \leq \alpha \leq 1$, and $\alpha = 1$ if an incident is detected on link $l$ during $t$.

An Empirical Bayes (EB) method is proposed to compute $\alpha$ based on the variance of neighbour links models.
In this method, $\alpha$ can be calculated as:

$$\alpha = 1 / [1 + \frac{\text{Var}(x_{rn})}{E(x_{rn})}] \quad (4.9)$$

Where:

$\text{Var}(x_{rn}) = \text{Variance of the estimated real-time average travel time of link } l \text{ using data of neighbour links } n \text{ during } t$

$E(x_{rn}) = \text{Expectation of the estimated real-time average travel time on link } l \text{ using data of neighbour links } n \text{ during } t$

To estimate the weight $\alpha$ in Equation (4.9), the expectation and the variance of the real-time average travel time of link $l$ using the data of neighbour links $n$ during $t$ denoted by $E(x_{rn})$ and $\text{Var}(x_{rn})$ need to be calculated. The method of statistical differentials is used to obtain $E(x_{rn})$ and $\text{Var}(x_{rn})$ as follows:

$$x_{rn} = \sum_{i=1}^{n} w_i \cdot x_{ri} \quad (4.10)$$

$$E(x_{rn}) = x_{rn} + \frac{\sum_{i=1}^{n} \frac{\partial^2 x_{rn}}{\partial x_{ri}^2} \text{Var}(x_{ri})}{2} \quad (4.11)$$

$$\text{Var}(x_{rn}) = \sum_{i=1}^{n} \left( \frac{\partial x_{rn}}{\partial x_{ri}} \right)^2 \cdot \text{Var}(x_{ri}) \quad (4.12)$$

Where:

$x_{ri} = \text{Estimated travel time on link } l \text{ using data of neighbour } i \text{ during } t$, where $i \in n$

$w_i = \text{Assigned weight to the estimated travel time of link } l \text{ from neighbour } i \text{ during } t$

$\text{Var}(x_{ri}) = \text{Variance of the estimated travel time of link } l \text{ from neighbour } i \text{ during } t$

Which leads to:

$$E(x_{rn}) = x_{rn} \quad (4.13)$$

$$\text{Var}(x_{rn}) = \sum_{i=1}^{n} (w_i)^2 \cdot \text{Var}(x_{ri}) \quad (4.14)$$
The following sequential points summarize the statistical method:

- Define a set of neighbours for each link based on preset correlation criteria,
- Use the historical data to develop regression models that relate link to link travel times,
- Prepare a weighting scheme to combine the outputs of different models to improve the estimate of link travel time,
- Receive travel time data from probe vehicles (or any other mode of data collection),
- Use the developed models to estimate travel times on links with no data using the available information of neighbour links,
- Apply the pre-prepared weighting scheme to fuse estimates of different models,
- Calculate the expectation and the variance of the real-time average travel time of link \( l \) using the data of neighbour links \( n \) during \( t \),
- Calculate the weight \( \alpha \), and
- Compute the best estimate of real-time average travel time of link \( l \) during \( t \) by fusing historical data of the link and neighbour links estimates.

**The ANN Model**

The Multi-Layer Perception (MLP) network structure trained with the Standard Back Propagation (SBP) algorithm (Rumelhart and McClelland 1986), is used to develop the ANN model. This network structure has been widely used and has proved satisfactory in many situations (Maier et al. 2000, Xie et al. 2004). The developed MLP network is composed of three layers, which is the minimum possible number of layers. The first layer is the input layer and it included a number of neurons equal to the number of input variables (in our case, the number of defined neighbours). The second layer is the output layer, and it included only one neuron representing the only output variable (i.e. travel time on the link with no data). The third layer is the hidden layer, which included \( n \) neurons. There is no general agreement on the value of \( n \). Xie et al. (2004) advocated using \( n = 2m - 1 \), where \( m \) is the number of neurons in the input layer. Rilett and Park (1999) suggested that increasing the number of neurons can result in long computation times and high probability of identifying local error minimums.
To optimize the value of $n$, values between 3 and $2m$ are tested until reaching the value which provided the lowest prediction error.

Other elements of the MLP model include the transfer functions, learning rate, and momentum. A linear function is used to connect the input layer to the hidden layer, and a sigmoid function is used to connect the hidden layer to the output layer. Using a very small learning rate can cause the net to take extensive time to converge. Too large values of the learning rate may result in the net becoming unstable during learning and never converge. In general, values between 0.1 and 0.5 for the learning rate are usually used. Similar to Sayed and Abdelwahab (1998), a learning rate of 0.2, and a momentum rate of 0.8, are used in this study. A training error of 3% was set as the convergence criteria for the ANN model.

*The Neuro-fuzzy Model*

ANNs have the ability to learn from a set of data, but the knowledge gained is hidden from the user (Sayed and Razavi 2000). The concept of neuro-fuzzy systems has emerged to combine the transparent, linguistic representation of a fuzzy system with the learning ability of an ANN (Brown and Harris 1994). A neuro-fuzzy system uses an ANN learning algorithm to determine its parameters (i.e. fuzzy sets and fuzzy rules) by processing data samples. Therefore, it can be trained to perform an input/output mapping, just as with an ANN, but with the additional benefit of being able to provide the set of rules on which the model is based. This gives further insight into the process being modeled (Sayed et al. 2003).

Associative memory networks (AMNs) have been shown to have the ability to approximate any continuous function, given sufficient degrees of freedom (Brown and Harris 1994). The input space of B-Spline AMNs is covered by a set of multidimensional overlapping basis (i.e. membership) functions. Model structure and complexity are determined by the size, shape, and overlap of the basis functions. The basis functions can take a number of forms, including B-Spline and Gaussian functions. B-Spline functions (Figure 4.1) are piecewise polynomials of order $k$ that have been widely used in surface
fitting applications. They can be used to represent fuzzy membership functions as described in Brown and Harris (1994).

![Figure 4.1 B-Spline Fuzzy Membership Functions (Sayed and Razavi 2000)](image)

The order of the B-Spline functions determines their smoothness. They can be used to implement crisp fuzzy sets ($k = 1$), the standard triangular fuzzy membership functions ($k = 2$), or other smoother representations. A univariate B-Spline function of order $k$ is non-zero only over $k$ intervals that are generated by a $(k + 1)$ knots. A multivariate B-Spline function can be formed by taking the tensor product of $n$ univariate basis functions (Brown and Harris 1994). The weighted sum of these basis functions produces a nonlinear adjustable input/output mapping. The network output can be represented by Equation (4.15):

$$f = \sum_{l=1}^{p} a_l w_l = a(x)^T w$$  \hspace{1cm} (4.15)

Where:

- $f$ = output
- $a(x) = [a_1(x), ..., a_p(x)] = $ vector of basic function outputs when presented by input $x = (x_1, ..., x_n)$
- $w = (w_1, ..., w_p) = $ vector of network weights

This type of network is equivalent to the union of a set of fuzzy rules in the form:

$$\text{IF } (x \text{ is } A_i) \text{ THEN } (f \text{ is } B_j) \ c_{ij} \hspace{1cm} (4.16)$$
Where \((x \text{ is } A_i)\) and \((f \text{ is } B_j)\) represent linguistic expressions for the input and output respectively; and \(c_{ij}\) is the rule confidence that relates the \(i^{th}\) fuzzy input set to the \(j^{th}\) fuzzy output set. This means that a weight can be fuzzified to produce a rule confidence vector \(c_i\) which can then be defuzzified to produce its original weight. This relationship allows the network to be trained in a weight space leading to considerable reduction in computational cost, while explaining the output with linguistic rules and the associated rule confidences.

**Curse of Dimensionality**

Many neuro-fuzzy systems suffer from what is referred to as the “curse of dimensionality.” The number of potential fuzzy rules in a fuzzy system is an exponential function of the dimension of the input space. This exponential growth of fuzzy rules, with a number of inputs, makes it impractical to use most existing neuro-fuzzy architectures for problems involving high dimensionality. To illustrate this “curse,” one may consider a fuzzy system with \(N\) input variables each of which have \(M\) membership functions. In such a system, as many as \(M^N\) combinations for potential fuzzy rules would exist.

The B-Spline AMN models adopted in this study use the Analysis Of Variance (ANOVA) (Brown and Harris 1995). In cases where terms of higher dimensions are zero or negligible, the ANOVA representation allows the input/output mapping to be approximated using a limited number of “subnetworks” of much lower dimensions.

It should be noted that each of these subnetworks represents a separate AMN, the outputs of which are summed to produce the overall model output. A fuzzy rule within each subnetwork is associated with a rule confidence \(c\). A rule confidence of zero indicates that the rule is not contributing to the output, while a rule confidence of one indicates 100% contribution. The number of fuzzy rules used in each subnetwork depends on the number of membership functions that are used to fuzzify the inputs of that subnetwork. The optimum structure of a B-Spline AMN is achieved through the selection of the smallest number of model inputs and the smallest number of basis functions for these inputs. Adaptive Spline Modelling Of Observation Data (ASMOD) (Kavli 1993) is an algorithm that uses the above decomposition to arrive at this optimal structure. In the ASMOD algorithm, training data are used to calculate the mean square error of the output.
algorithm starts from the simplest structure (e.g. only the first variable in one subnetwork with two triangular Splines) and iteratively refines its structure until some stop criteria are satisfied. In each step, among a number of potential changes to the structure, the one with the best performance is selected, and the process continues. The addition of a new input, adding an existing input to a subnetwork, splitting a subnetwork, and deleting an input, are all possible changes to the structure. Adding Splines in the middle of existing ones, deleting Splines, and changing the order of Splines, are examples of other changes that are considered by the algorithm. A more detailed explanation of the algorithm can be found in (Kavli 1993, Brown and Harris 1994). Some measure of statistical significance is used as stopping criteria. Among many such measures, Bossley et al. (1995) stated that, for noisy data, the Bayesian statistical significance measure appears to perform well. Hence, it is used in this study. The Bayesian statistical significance measure is given by (Brown and Harris 1994) as:

\[
K = L \ln(J) + p \ln(L)
\]  \hspace{1cm} (4.17)

Where:

\(K\) = performance measure,
\(p\) = size of current model,
\(J\) = mean square error, and
\(L\) = number of data pairs used to train the network.

The ASMOD algorithm uses training data to automatically determine the model inputs and the number of basis functions. However, the order of basis functions has to be determined \textit{apriori}. Higher-order functions result in smoother model outputs, but increase computational costs, and can lead to over-fitting of the data (Brown and Harris 1994). Therefore, it is desirable to use lower order functions, provided they are adequately able to model the relationship. The ASMOD algorithm appears to perform at least as well as the best of the other methods.
The Non-Parametric KNN Model

Nonparametric Regression (NPR) is an approach used to estimate the values of a dependent variable without a direct relationship (i.e., parametric model form) between the dependent and independent variables. This approach requires an adequate database in which similar previous cases exist. Where a case is characterized by the value of the dependent variable and the associated values of the independent variables. The approach then defines the “neighbourhood” of past, similar cases, and uses it to estimate the value of the dependent variable. The quality of the data set has a major influence on the effectiveness of any nonparametric modelling.

There are two main approaches to neighbourhood definition (Altman 1992): Nearest Neighbours (NN) and Kernel. In the Nearest Neighbours approach, the nearest (i.e., most similar) $K$ cases to the current case are selected and used by the modeller to estimate the value of the output variable, where $K$ is a predefined number. This method is known as the $K$-Nearest Neighbours approach and is denoted by $KNN$. The Kernel method defines a constant bandwidth in the state space and all neighbours within this bandwidth are used to estimate the value of the output variable. A drawback of the Kernel method is that the defined bandwidth may not include any data points and, hence, no estimate can be provided. In this research the KNN is applied for corridor travel time estimation.

4.2.3 Application Entity

The concept of neighbour links travel time estimation can be extended and applied at a corridor level. The two network entities: link and corridor, differ in terms of data needs, sensor coverage, strength of travel time covariance, and application. Link travel times in urban areas are highly fluctuating due to the existence of control devices (e.g., traffic signals), transit vehicles, and on-street parking, among other factors. Using corridor travel times, instead of link travel times, can normalize the impact of traffic signal delays, reduce high travel time variability, and lead to a stronger travel time relationship. Also, while it is impractical to install sensors to measure travel times of each link, it is feasible to add sensors (e.g., video cameras, AVI, etc) to measure corridor travel times. The end-use of
traffic data collection is information delivery to travellers, who indeed are more interested in acquiring route travel times than link travel times.

4.2.4 **Source of Neighbour Links Travel Time Data**

A broad range of sensor types can be used for travel time data collection as was shown in chapter two. These sensors can either be fixed or moving (i.e. probes). In this research, three cases are analyzed that correspond to the use of fixed sensors, samples of probe vehicles, and transit vehicles, as sources of neighbour links travel time data. The case of using fixed sensors is analyzed using corridors real-life travel time data. Although these data were collected using test vehicles, it was assumed that the time stamps associated with passing a checkpoint (i.e. intersection) are similar to the type of data that can be obtained using AVI systems. A sample of probe vehicles was used as the source of neighbour links data in chapter five, while transit vehicles data were used in chapter six. In chapter seven, different data fusion methods that can be used to integrate links historical data, and neighbour links real-time data, whether from vehicle probes or transit probes, are explored.

4.3 **CASE STUDY: DOWNTOWN VANCOUVER**

4.3.1 **Data Description**

As described in chapter three, a travel time survey was designed to simultaneously collect real-life automobile link travel times on a pre-designed route in downtown Vancouver. The selected analysis route covered six one-way arterials in the north-south direction of the downtown area. Average travel time on each segment was computed using the data collected from all drivers entering the survey route within the same interval. The aggregated data was split into two sets: one for the model’s development, and one for the model’s validation. In total, 31 data samples were collected. Of these, 21 records were used for the model’s development and 10 records were used for validation. The choice of the validation records was based on a stratified sampling method in which samples were randomly drawn from each day so as to yield a sample distribution similar to that of the calibration data set.
4.3.2 The Statistical Method

The travel time survey route included 55 travel time sections. Only 31 “through” travel time sections were considered in the current analysis (i.e. link entry and exit are “through” movements). The collected data were used to investigate the travel time correlation between the 31 travel time sections. Pearson Correlation was obtained for each pair of links and the $p$-value was used to test the null hypothesis of zero correlation, $H_0: r = 0$. A correlation threshold of 0.4 was used to define link neighbours. A sensitivity analysis of the impact of the correlation threshold on the estimation accuracy is presented later in this chapter. For all neighbour pairs, the null hypothesis of zero correlation was rejected with $p < 0.05$. At least one neighbour could be identified for each link with a maximum of 17 neighbours. The total number of defined neighbours for the 31 sections was 262, with an average of 8 neighbours per link. Figure 4.2 shows the number of neighbours for each travel time section using the pre-defined neighbourhood criteria.

![Figure 4.2 Number of Neighbours for each Travel Time Section](image)

Four travel time segments were selected randomly from the 31 segments as examples of links with unknown travel times, while the other 27 segments represented the neighbourhood population where travel time data exist. The section numbers of the four selected segments are 2, 13, 22, and 46 (Figure 3.5). Models that relate target link travel time to travel times of neighbour links were developed. The exponential model form showed a better fit with the data than did the linear and the power model forms. According
to the number of neighbours defined for the four segments, 25 exponential models were
developed after the exclusion of models that had at least one insignificant parameter. The
weighting schemes, as described earlier in this chapter, were used to combine estimates of
different models. The validation data set was used to assess the accuracy of the method.
Two error measurements were applied to compute the deviations of the estimated values
from the true values. These error measurements were:

- *The Mean Absolute Percent Error*, which indicates the expected error as a fraction
  of the measurement.

\[
MAPE = \frac{100}{N} \sum_{i=1}^{N} \frac{|x_i - x_{rn}|}{x_i}
\]  

(4.18)

- *The Root Relative Squared Error*, which captures large prediction errors and is
  calculated as follows:

\[
RRSE = \sqrt{\frac{1}{\sum_{i=1}^{N} x_i} \sum_{i=1}^{N} \left( \frac{x_i - x_{rn}}{x_i} \right)^2} 
\]  

(4.19)

Where:

- \(x_{rn}\) = the weighted estimates of travel times from neighbours exponential models
- \(x_i\) = the true link travel time
- \(N\) = number of validation observations

**Travel Time Estimation Results**

In general, all of the weighting methods showed satisfactory estimation accuracy with the
MAPE ranging between 10.7% and 19.7%, with an average of approximately 13.9% for the
four segments. The RRSE ranged from 16.0% to 24.1% with an average of 19.2%. Figure
4.3 shows the error measurement values of the applied weighting schemes.
It can be inferred from Figure 4.3 that the proposed method is insensitive to the weighting scheme and almost all of the schemes have the same accuracy. Noteworthy is that all of the proposed weighting schemes are based on historical data (as represented by the correlation, variance, or coefficient of determination). Other weighting schemes can be applied to incorporate real-time conditions.
An example of these schemes is a formula that incorporates the number of probe vehicles on each neighbour during a measurement interval \( (m) \) and the travel time variance of each neighbour during an interval \( (\sigma_{RT}^2) \) such that:

\[
w_l = \frac{\frac{m}{\sigma_{RT}^2}}{\sum_{i=1}^{m} e^{\frac{-\sigma_{RT}^2}{m}}} \quad (4.20)
\]

Unfortunately, this weighting scheme could not be applied due to the unavailability of an adequate data set.

The weight \( \alpha \) as in Equation (4.9) was computed for the four links and all the six weighting schemes. Noteworthy is that the weight \( \alpha \) is dynamic, and it changes from one measurement interval to another, according to the available neighbours, and the expectation/variance of the travel time estimate. As shown in Equations (4.9) and (4.10), the value of \( \alpha \) is mainly dependent on individual models’ variances. For models with high variances, (i.e. high \( Var(x_{ri}) \)), the variance of the estimated weighted travel time from neighbours \( n \), denoted by \( Var(x_{rn}) \) will be large. Consequently, the value of \( \alpha \) will be low.

The results showed that for Links 2 and 46, where neighbour links models had relatively low variances, \( \alpha \) ranged between 0.86 to 0.91, indicating that most of the weight in Equation (4.8) will be assigned to the estimates of neighbour links’ models. For Link 13, the weight \( \alpha \) ranged from 0.75 to 0.81, giving slightly lower weights to neighbours’ estimates, compared to Links 2 and 46. Finally, Link 22 showed the lowest weight \( \alpha \) among the four links and that ranged from 0.58 to 0.61. Again, the relatively low weight for Link 22 is attributed to the high variances of its models (72 to 82 sec\(^2\)). A sample calculation is presented in Table 4.1 for one data record of Link 22.
Table 4.1 Example Calculation of the Proposed Method for One Data Record

<table>
<thead>
<tr>
<th>Link with No Data</th>
<th>Link 22</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighbour ID</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>( \alpha^* )</td>
<td>26.36</td>
<td>26.22</td>
</tr>
<tr>
<td>( \beta^* )</td>
<td>0.00297</td>
<td>0.00294</td>
</tr>
<tr>
<td>Travel Times on Neighbours (TT)</td>
<td>14.60</td>
<td>17.20</td>
</tr>
<tr>
<td>( x_{ri}^* )</td>
<td>27.53</td>
<td>27.58</td>
</tr>
<tr>
<td>Model Variance (( \sigma^2 ) = SSE/df)</td>
<td>72.4</td>
<td>74.3</td>
</tr>
<tr>
<td>( 1/\sigma^2 )</td>
<td>0.0138</td>
<td>0.0135</td>
</tr>
<tr>
<td>Weight (( w_i ))</td>
<td>0.266</td>
<td>0.259</td>
</tr>
<tr>
<td>( E(x_{ri}) )</td>
<td>27.21</td>
<td></td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>True TT</td>
<td>32.20</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>4.99</td>
<td></td>
</tr>
<tr>
<td>Relative Error (RE)</td>
<td>15.5%</td>
<td></td>
</tr>
</tbody>
</table>

\( \alpha \) and \( \beta \) are the parameters of the exponential model: \( x_{ri} = \alpha \times e^{\beta_TT_i} \)

**Impact of Neighbourhood Cut-off on the Estimation Accuracy**

In order to investigate the impact of the correlation cut-off value on the estimation accuracy, a duplicate analysis was repeated for cut-off values of 0.4, 0.5, 0.6, and 0.7. Intuitively, increasing the cut-off value will decrease the number of neighbours to the point where no neighbours would be found at high cut-off values. Figure 4.4 shows the decrease in the number of neighbours for the four segments with different cut-off values, while Figure 4.5 presents the MAPE for the four links using different cut-off values.

![Figure 4.4 Identified Number of Neighbours at Different Cut-off Values](image-url)
As shown in Figures 4.4 and 4.5, increasing the cut-off value reduced the number of defined neighbours to the point where no neighbours were identified for some links at high cut-off values. Moreover, the estimation accuracy does not significantly drop with an increase in the cut-off value (except for Link 2). This indicates that the method is less sensitive to the used cut-off value and the modeller has the choice to select this value according to the strength of the travel time correlation in the study area. It should be noted that including neighbours with relatively weak positive correlation is also possible. However, estimates of their models will be assigned lower weights compared to other neighbours with stronger correlations.

### 4.3.3 AI Methods

The statistical-based methodology was applied in the previous section to define link neighbours and estimate travel times on links with no data, using the data from other links. In this section, another modelling approach was used for neighbour links travel time estimation using B-Spline neuro-fuzzy models. This approach was also compared to the statistical method and an Artificial Neural Network (ANN) model. The objectives of this part of the research were:

- To study the potential use of B-Spline neuro-fuzzy models to automatically select link neighbours, and,
− To compare the predictive ability of different modelling techniques in travel time estimation from neighbour links.

Models Development

The same four travel time segments that were previously selected were used once again to enable the comparison of the statistical method and the AI models (i.e. Links 2, 13, 22, and 46). A B-Spline neuro-fuzzy model was used to relate the travel time of each link to travel times of its neighbours. The method was compared to the statistical approach, as presented earlier in this chapter, in addition to an ANN model.

The B-spline AMN models were developed, and the ASMOD algorithm was used, to determine appropriate model inputs and the number and size of the basis functions. A training data set was used to build and train the neuro-fuzzy models. Four models were developed for the selected four segments. B-Spline AMN models were implemented using two types of input data. In the first, all links travel time data (i.e. 31 links: 30 inputs and 1 output) were used, while in the second, only data of the neighbours defined by correlation were included. The purpose of the first model was to examine the ability of the ASMOD algorithm to identify link neighbours and include them in the model. The second model was a two-step model where correlation was used first to define the model input variables (i.e. most influential variables). This model was developed to test whether model performance improves when reducing the number of input variables. The Bayesian statistical significance measure, as described earlier, was used as the stopping criteria. The maximum number of subnet inputs was set to 1 in order to explicitly identify the relationship between each pair of links.

Identification of Links Neighbours

Two types of neuro-fuzzy models were developed. The model that included travel times of all links as inputs is referred as to “Neuro-Fuzzy (All),” while the model that incorporated only neighbours links travel times is referred as to “Neuro-Fuzzy (Neighbours).” Table 4.2 shows the neighbours defined for each link using the correlation threshold and neuro-fuzzy models.
Table 4.2 Number of Neighbours Identified by Neuro-Fuzzy Models

<table>
<thead>
<tr>
<th>Method</th>
<th>Link</th>
<th>Neighbour IDs</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>5</td>
<td>35 36 20 6 37</td>
<td>6</td>
</tr>
<tr>
<td>Neuro-Fuzzy (All)</td>
<td>2</td>
<td>5  20 6 12 21</td>
<td>5</td>
</tr>
<tr>
<td>Neuro-Fuzzy (Neighbours)</td>
<td>5</td>
<td>20 6 37</td>
<td>4</td>
</tr>
<tr>
<td>Correlation</td>
<td>30</td>
<td>4 29 27 31 28</td>
<td>6</td>
</tr>
<tr>
<td>Neuro-Fuzzy (All)</td>
<td>13</td>
<td>29 31 39 2</td>
<td>4</td>
</tr>
<tr>
<td>Neuro-Fuzzy (Neighbours)</td>
<td>30</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>Correlation</td>
<td>28</td>
<td>27 5 29</td>
<td>4</td>
</tr>
<tr>
<td>Neuro-Fuzzy (All)</td>
<td>22</td>
<td>27 1 21</td>
<td>4</td>
</tr>
<tr>
<td>Neuro-Fuzzy (Neighbours)</td>
<td>28</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Correlation</td>
<td>27</td>
<td>29 48 28 36 5 1 39 45</td>
<td>9</td>
</tr>
<tr>
<td>Neuro-Fuzzy (All)</td>
<td>46</td>
<td>27 29 39 3 6 5</td>
<td>8</td>
</tr>
</tbody>
</table>

As shown in Table 4.2, many of the neighbours defined by the correlation threshold were identified by the “Neuro-Fuzzy (All)” model. Moreover, the optimal network structure of “Neuro-Fuzzy (Neighbours)” models always included only a subset or correlation neighbours.

All of the developed models used triangular membership functions \((k = 2)\) for all variable subsets. The output fuzzy rules were all revised and found to be logical. Examples of these rules are listed in Table 4.3. The numbers in parentheses in the consequent section represent the corresponding degree of belief.

Table 4.3 Examples of the Rules Generated by the Neuro-Fuzzy Model for Link 2

INTERNAL NETWORK 1 WITH 4 SUBNETWORKS

SubNetwork1 with 1 Input Variable and the 2 following Fuzzy Rules:
1: IF Link (5) TT is Low THEN Link (2) TT is Low (0.93) OR Link (2) is High (0.07)
2: IF Link (5) TT is High THEN Link (2) TT is Low (0.23) OR Link (2) is High (0.77)

SubNetwork2 with 1 Input Variable and the 2 following Fuzzy Rules:
3: IF Link (20) TT is Low THEN Link (2) TT is Low (0.99) OR Link (2) is High (0.01)
4: IF Link (20) TT is High THEN Link (2) TT is High (1.00)

SubNetwork3 with 1 Input Variable and the 3 following Fuzzy Rules:
5: IF Link (6) TT is Low THEN Link (2) TT is Low (0.80) OR Link (2) is High (0.20)
6: IF Link (6) TT is Medium THEN Link (2) TT is Low (0.80) OR Link (2) is High (0.20)
7: IF Link (6) TT is High THEN Link (2) TT is Low (0.15) OR Link (2) is High (0.85)

SubNetwork4 with 1 Input Variable and the 3 following Fuzzy Rules:
8: IF Link (37) TT is Low THEN Link (2) TT is Low (0.32) OR Link (2) is High (0.68)
9: IF Link (37) TT is Medium THEN Link (2) TT is Low (1.00)
10: IF Link (37) TT is High THEN Link (2) TT is Low (0.83) OR Link (2) is High (0.17)
**Estimation Accuracy**

In terms of the predictive ability of the model, Figure 4.6 shows the error measurements of all the models. As shown in the Figure, the statistical method and the “Neuro-Fuzzy (Neighbours)” models outperformed the ANN model and “Neuro-Fuzzy (All)” in all cases except for Link 2. The average MAPE for the four links was 19.3%, 16.4%, 13.9%, and 12.8% for the “Neuro-Fuzzy (All)”, ANN, statistical, and “Neuro-Fuzzy (Neighbours)” models, respectively.

![Figure 4.6 Error Measurements of All Models](image)

It was expected that the optimization algorithm of the B-Spline AMN model would be able to capture the links (i.e. inputs) that had high influence on the output (i.e. unknown link
However, due to the large number of input variables (i.e. 30 links), the model was unable to correctly find the most influential inputs. This may be attributed to the curse of dimensionality, as described earlier. Reducing the number of input variables using the correlation threshold, facilitated implementing the most optimal model, and significantly improved the model’s accuracy.

It should be noted that using ANN and neuro-fuzzy models can be problematic in some cases where data are not available for all neighbours during a specific time interval. Hence, developing a family of link to link travel time models could be beneficial as it enables travel times of one singular neighbour link to be used. However, ANN and neuro-fuzzy models can still be used in the following situations:

1. Modelling links to neighbours’ travel times where travel times on neighbour links are always available using fixed sensors (e.g. AVI, video cameras…etc), and,
2. Online training of the model according to the data available from specific neighbours.

4.3.4 Neighbour Links Travel Time Estimation: A Discussion

A general framework was presented for using neighbour links data for network-wide travel time estimation. This method is intended for real-time traveller information systems. It is anticipated that a large historical data set is required to initially define link neighbours. This historical database must be updated as new travel time information is obtained. The definition of neighbours in itself has to be dynamic. That is, the defined neighbours can differ from one time interval to another according to the strength of the correlation. Furthermore, the defined neighbours can change with any update to the historical database. The developed neighbours’ models should be revised and updated from one time to another (e.g. every month). Existing methods to estimate link travel times using loop detectors data, license plate matching, AVI systems, etc. can all be used to estimate travel times derived from the same entity (e.g. link). The proposed model can be integrated with all of these models to estimate travel times of neighbour links not equipped with sensors.
The main focus of this research is the estimation of link travel time using data from neighbour links, denoted by $x_{rn}$ and methods to combine real-time neighbour data with link historical data. Although most of the attention is focused on the use of probe vehicles that can be tracked in real-time, the methodology is generic and can be applied to any other data collection method that uses fixed sensors such as license plate matching and AVI systems. The potential for using a sample of probe vehicles for neighbour links travel time estimation is presented in chapter five, along with an investigation of the impact of probes sample size on the estimation accuracy. Using transit vehicles as probes instead of passenger vehicles for neighbour links travel time estimation is presented in chapter six. The fusion of both sources of data is illustrated in chapter seven.

The purpose of this research is not to establish a full operating system capable of estimating travel times using neighbour links data. Rather, the main objective is to present a prototype that shows the potential of using neighbour links relationships for travel time estimation. Hence, only a few segments are usually randomly chosen in any analysis as examples of links with unknown travel times, while the other analyzed segments represent the neighbourhood population where travel time data exist. It should be noted that the proposed method is not intended to replace other existing travel time estimation models. Nevertheless, it serves as a supplementary and complementary model that can be integrated with almost all existing travel time estimation models through data fusion. The objective of the analysis is to demonstrate the feasibility of using neighbour links data as an additional source of traffic information that might not have been explored extensively in the travel time estimation literature.

4.4 NEIGHBOUR CORRIDORS TRAVEL TIME ESTIMATION

4.4.1 Analysis Corridors

This section extends the concept of link-neighbourhood, that was presented in the previous sections, to corridor-neighbourhood. Two corridors were analyzed in this part of the research: Homer and Hornby Streets (Figure 4.7). The first corridor, Homer Street, is a
The two corridors represent alternative routes to the northbound traffic entering the downtown area. Each of the two corridors is composed of five short segments. They both start at Davie Street, in the south, to Georgia Street, in the north. Most of the traffic then diverges to Georgia Street, turning either right or left. The two corridors share the same length: approximately 834 m. The posted speed limit in downtown Vancouver is 50 km/hr. Hence, the free flow travel time on both corridors is approximately 60 seconds. However, they suffer severe congestions and delays during the peak periods, and actual travel times are much higher.

Corridor travel times were computed from the available data of the travel time survey. Average travel time on each corridor was computed using the data collected from all drivers entering a corridor within the same interval.
4.4.2 Travel Time Distribution

The travel time distribution of the two corridors was investigated. The most commonly used travel time distributions in the literature are normal and lognormal (Chen and Chien 2000, Gajewski and Rilett 2003, Rakha et al. 2006, Chen et al. 2007). It is always preferable to fit the travel times to lognormal distributions, considering the fact that travel times are non-negative and usually right-skewed (Rakha et al. 2006).

Several goodness of fit tests can be used to test the null hypothesis that the observed values of a variable represent a random sample from a specific theoretical distribution. The most widely used test is the Chi-squared test, which is used for testing against discrete and continuous distributions. This test requires the sample of observations to be large enough for the approximations to be valid. Other types of goodness of fit tests are based on the Empirical Distribution Function (EDF). The EDF tests offer many advantages over traditional Chi-squared goodness of fit tests including improved power and invariance with respect to the histogram midpoints. Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests are two examples of EDF tests. The null hypothesis of the two tests is that the observations are derived from the same statistical distribution. The null hypothesis is rejected if the test statistic is greater than the tabulated critical value. Some statisticians prefer to use the $p$-value associated with the goodness of fit statistic, rather than the statistic value. In this case, the null hypothesis is rejected if the $p$-value is less than the predetermined critical value ($\alpha$). A discussion on the EDF is provided in D'Agostino and Stephens (1986). In this study, both the K-S and A-D tests were used to assess the goodness of fit of the observed average travel times against normal and lognormal distributions.

The results of the goodness of fit tests on the travel times on the two corridors are shown in Table 4.4. The results support the null hypothesis that the observed sample of travel times on Homer Street can follow a normal or a lognormal distribution, with the statistics of the lognormal fit being more statistically significant than the normal fit. The K-S test supported the null hypothesis that travel times on Hornby Street can come from a normal distribution. However, the A-D statistic supported rejecting this hypothesis. Both the K-S
and the A-D tests supported the null hypothesis of the lognormal distribution of Hornby Street travel times.

Table 4.4 Results of Goodness of Fit Statistical Tests

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kolmogorov-Smirnov</td>
<td>Anderson-Darling</td>
</tr>
<tr>
<td></td>
<td>D* Statistic</td>
<td>p-value</td>
</tr>
<tr>
<td>Homer St.</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>Hornby St.</td>
<td>0.13</td>
<td>0.07</td>
</tr>
</tbody>
</table>

L = 0.201
AD = 0.757
α = 0.05

Table 4.5 shows the parameters of the fitted distributions of the travel times on both Hornby and Homer Streets. Figure 4.8 shows the fitted normal and lognormal distributions of the travel times on Hornby and Homer Streets, respectively.

Table 4.5 Fitted Statistical Distributions of Observed Travel Times

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Mu) (Seconds)</td>
<td>STD (Sigma)</td>
</tr>
<tr>
<td>Homer St.</td>
<td>197.9</td>
<td>56.6</td>
</tr>
<tr>
<td>Hornby St.</td>
<td>194.8</td>
<td>76.6</td>
</tr>
</tbody>
</table>
4.4.3 Neighbour Corridor Travel Time Association

The travel time data set was used to investigate the association (monotonic relationship) between travel times on the two corridors. The Pearson Correlation Coefficient ($r$) was used to test this association. The $p$-value was used to test the null hypothesis of zero correlation, $H_0: r = zero$. If $p < 0.05$ then the probability of having a positive correlation is $> 97.5\%$. The Pearson correlation of the travel times on the two corridors was found to be $0.69$ which shows a strong relationship between the travel times of the two corridors. The $95\%$ upper and lower limits of the correlation were $0.49$ and $0.82$, respectively. The $p$-value was less than $0.0001$, which rejects the null hypothesis of zero correlation at a confidence level of more than $99\%$. 

---

**Figure 4.8 Fitted Normal and Lognormal Corridor Travel Time Distributions**
A one-way ANOVA was further carried out to test whether the travel times on Homer Street have a significant impact on the travel times on Hornby Street. Results are shown in Table 4.6. They support the hypothesis of a strong relationship between travel times on the two corridors.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Mean</th>
<th>Variance</th>
<th>( F )</th>
<th>( F_{critical} )</th>
<th>( P-value )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homer St. TT</td>
<td>197.94</td>
<td>3203.43</td>
<td>0.062</td>
<td>3.96</td>
<td>0.805</td>
</tr>
<tr>
<td>Hornby St. TT</td>
<td>194.37</td>
<td>5467.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.4.4 Description of the Used Models

The collected average travel time data set was split into two sets, one for models’ development and one for models’ validation. For models’ development, travel times on Hornby Street represented the dependent variable (i.e. unknown) while travel times on Homer Street represented the independent variable. This scenario corresponds to having fixed sensors to obtain travel times on Homer Street, and having no sensors installed on Hornby Street. In the validation data set, the known travel times on Hornby Street were compared to the estimated travel times of different models, and the magnitudes of several error measurements were accordingly computed. Several modelling techniques were used and compared, these comprise: regression linear and non-linear models, and a Non Parametric (NP) K-Nearest Neighbours (KNN) model. In the following sub-sections, the development of the KNN model is presented, followed by the results of the analysis.

**The Non-Parametric KNN Model**

As mentioned earlier, the development of the KNN model requires the definition and optimization of a number of parameters that include: characterization of an appropriate state space, definition of a distance metric to determine the closeness between historical observations and the current case, choice of number of neighbours, and selection of an estimation method given a number of nearest neighbours.
**State Space (Feature Vector)**

The state space is simply the domain of independent variables that correlate to the dependent variable and hence can be used to search for similar cases. In this study, the state space includes only one variable (i.e. travel time on Homer Street) and each previous case is characterized by the values of travel times on Hornby and Homer Streets.

**Distance Metric**

When searching for neighbour cases, similarity between the current case and previous cases is determined using a distance metric. Similarities between current and previous cases are usually described by Minkowski distance metrics defined as (Smith et al. 2001):

\[ L_m = \left[ \sum_{i=1}^{N} |p_i - q_i|^m \right]^{\frac{1}{m}} \]  

(4.21)

Where:

\( N \) = the dimension of the state vector (i.e. number of independent variables)
\( p_i \) = the \( i^{th} \) variable in the state vector for the adjacent case
\( q_i \) = the \( i^{th} \) variable in the state vector for the current case

\( L_1, L_2, \) and \( L_\infty \) = the commonly known Manhattan, Euclidean, and max distances

In the current analysis, the input space incorporated only one variable and hence the distance metric could be expressed by:

\[ Dist_j = |TT_{Homer (j)} - TT_{Homer (new)}| \]  

(4.22)

Where:

\( Dist_j \) = the distance metric between the case \( j \) and the current case
\( TT_{Homer (new)} \) = travel time on Homer Street for the current case
\( TT_{Homer (j)} \) = travel time on Homer Street for case \( (j) \)

**Choice of Number of Neighbours (K)**

The number of Nearest Neighbours (NN), denoted by \( K \), is an important parameter that always needs to be optimized. The fact that a specific value of \( K \) has provided accurate
predictions in some modelling exercises does not mean that this value is transferable. In this research, values of $K$ between 1 and 16 were used, and the value that provided the lowest Mean Squared Error (MSE) was chosen as the optimal value.

**Forecast Generation**  
The most straightforward approach for generating a forecast of the output variable is to use the straight average approach, in which a simple average of the dependent variable values of the nearest neighbours is computed (Smith et al. 2002). This can be represented by Equation (4.23):

$$
\bar{T}_{Hornby} = \frac{1}{K} \sum_{j=1}^{K} T_{Hornby} (j)
$$

Where:

- $K$ = the number of neighbour cases found by the search procedure
- $\bar{T}_{Hornby}$ = the forecast of the output variable, travel time on Hornby Street
- $T_{Hornby} (j)$ = the value of the $j$th neighbour case of the historical data set

The weakness of this approach is that it neglects all information provided by the distance metric. It is more logical to assume that records that are closer to the current case (in terms of distance metric) can provide more information, and, hence, should have more impact on the forecast. A number of weighting schemes have been proposed for use within nonparametric regression as discussed by Smith et al. (2002). Generally, such weights are proportional to the distance between the neighbour and the current case. The most generic weighting scheme is to assign weights for neighbour cases by the inverse of distance metric. In this approach, neighbours are weighted according to their relative distance in relation to the current case. Hence, a case that is closer to the current case will have more impact on the forecast value. This can be expressed mathematically as:

$$
\bar{T}_{Hornby} = \frac{\sum_{j=1}^{K} \left( \frac{T_{Hornby} (j)}{\text{Dist}_j} \right)^C}{\sum_{j=1}^{K} \left( \frac{1}{\text{Dist}_j} \right)^C}
$$

Where: $C$ = power of the weight formula ($C = 1, 2, 3,...$)
The two weighting formulae in Equations (4.23) and (4.24) are used in the current study: average weighting and inverse of distance metric. In the second weighting formula, three values were used for $C$: 1, 2, and 3.

### 4.4.5 Error Measurements

The validation data set was used to assess the accuracy of all models. Three error measurements were applied to calculate the deviations of the predicted values from the true values. These error measurements are:

- **The Mean Square Error**, computed as:

  $$\text{MSE} = \frac{100}{M} \sum_{i=1}^{M} \left( TT_{\text{Hornby}} - \hat{TT}_{\text{Hornby}} \right)^2 $$  \hspace{1cm} \text{(4.25)}

- **The Mean Absolute Percent Error**, calculated as:

  $$\text{MAPE} = \frac{100}{M} \sum_{i=1}^{M} \left| \frac{TT_{\text{Hornby}} - \hat{TT}_{\text{Hornby}}}{TT_{\text{Hornby}}} \right| $$ \hspace{1cm} \text{(4.26)}

- **The Root Relative Squared Error**, calculated as:

  $$\text{RRSE} = \sqrt{\frac{1}{\sum_{i=1}^{M} TT_{\text{Hornby}}} \cdot \sum_{i=1}^{M} \left( \frac{TT_{\text{Hornby}} - \hat{TT}_{\text{Hornby}}}{TT_{\text{Hornby}}} \right)^2 \cdot TT_{\text{Hornby}}} $$ \hspace{1cm} \text{(4.27)}

Where:

$\hat{TT}_{\text{Hornby}}$ = the estimated travel time on Hornby Street, as obtained from the model

$TT_{\text{Hornby}}$ = the true travel time on Hornby Street

$M$ = number of validation observations
4.4.6 Results

Regression Models

The parameter “a” in the linear and power model forms was found to be insignificant at a 95% level (|t| < 1.96), therefore the two models were re-developed forcing a to be zero in the linear model and one in the power model. The Models’ coefficients and goodness of fit measures are shown in Table 4.7. The coefficients of three models appear with logical positive signs indicating monotonic travel time changes on the two corridors. All models’ parameters were significant at a 95% level. The models had similar coefficient of determination values of approximately 0.48.

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>68.39</td>
<td>4.54</td>
<td>0.00492</td>
<td>5.13</td>
<td>0.484</td>
</tr>
<tr>
<td>Linear</td>
<td>0.00</td>
<td>N/A</td>
<td>0.94863</td>
<td>19.18</td>
<td>0.474</td>
</tr>
<tr>
<td>Power</td>
<td>1.00</td>
<td>N/A</td>
<td>0.99025</td>
<td>102.14</td>
<td>0.474</td>
</tr>
</tbody>
</table>

The KNN Model

As mentioned earlier, the KNN model was optimized by varying the value of K and the forecast generation method. Values of K between 1 and 16 were used, with four possible weighting methods. Figure 4.9 shows the mean square error of all combinations of KNN models.

Figure 4.9 shows that the MSE decreases gradually with the increase of K until reaching the optimal value of K (i.e. minimum MSE), the MSE then starts to increase again. As is clear from the Figure, using the first order inverse (i.e. C = 1) of the distance metric to assign weights to each neighbour case significantly improves the predictive ability of the KNN model. The optimal KNN model was reached with K = 12 and C = 1. Table 4.8 shows the MSE of the optimal K value of each weighting method.
Figure 4.9 Mean Square Errors (MSE) of Different K Values and Weighting Methods

Table 4.8 Optimal K Value and MSE of Each Weighting Method

<table>
<thead>
<tr>
<th>Method</th>
<th>Optimal K</th>
<th>MSE (Seconds²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight Average</td>
<td>6</td>
<td>2990.6</td>
</tr>
<tr>
<td>Inverse Metric (C = 1)</td>
<td>12</td>
<td>2384.0</td>
</tr>
<tr>
<td>Inverse Metric (C = 2)</td>
<td>12</td>
<td>2806.0</td>
</tr>
<tr>
<td>Inverse Metric (C = 3)</td>
<td>13</td>
<td>3248.3</td>
</tr>
</tbody>
</table>

4.4.7 Models Comparison

In terms of the models’ predictive ability, Figure 4.10 shows the MSE error of all models, while Figure 4.11 presents the MAPE and the RRSE error measurements. Using the MSE for comparison, the optimized KNN model was found to be superior compared to all other models. The exponential regression model was found to be the second best, while the linear model performed the worst. The ranking of the models was the same when using the MAPE and the RRSE for the ranking. In general, all models showed satisfactory estimation accuracy with the MAPE ranging between 13.7% and 17.1% and the RRSE ranging from 20.3% to 22.3%.
4.4.8 Corridor Travel Time Estimation: A Discussion

Classical regression models have been used for many years in many fields. They are well understood, have a strong theoretical basis, and are easy to develop/evaluate. One assumption that must be checked prior to their deployment is the normality of the dependent variable. If the dependent variable follows another statistical distribution, rather than the normal or the lognormal, the modeller can use Generalized Linear Models (GLM), where the error structure is not constrained by the normality assumption. Another difficulty
associated with the development of regression models is the modeller’s inability to incorporate the statistical properties of the input variables. A common assumption is that input variables are known with high certainty, hence, their statistical properties are neglected.

KNN models, on the other hand, are distribution free which presents an advantage over statistical regression models where an implicit assumption in the transformation function is that all error terms are assumed to be independent and identically distributed with a normal distribution. KNN non-parametric models are easy to develop and use, but require the availability of an adequate (i.e. sufficiently large) data set that can be used to extract similar cases to the current ones. Using KNN models also incorporates many parameters that need to be optimized, as described in this chapter and in the previous literature.
CHAPTER 5: USING SAMPLES OF PROBE VEHICLES FOR NEIGHBOUR LINKS TRAVEL TIME ESTIMATION

5.1 INTRODUCTION

A “true” estimate of link travel time can only be obtained if travel times from all vehicles are sampled. As it is practically infeasible to sample the whole population, a small sample of probe vehicles are usually used for the estimation. The required sample size is dependent on many factors including: the type of data collected by the probes, the required level of accuracy, the required confidence in the accuracy achieved, the complexity of the area analyzed, the population size (i.e. in terms of traffic flow), time of day, and finally, the nature of the traffic state (e.g. steady or congested). Investigating the optimal probe vehicles sample size (also referred to as market penetration) has been extensively researched. Van Aerde et al. (1993) established a relationship between the travel time Coefficient of Variation (CoV) and the level of market penetration. The relationship had an inverse shape, showing a decrease of the CoV when the market penetration increases. The authors also distinguished the importance of considering both the temporal and the spatial variations of market penetration.

Sanwal and Walrand (1995) used a simulation model to investigate the accuracy of probe vehicles data at different penetration levels and different polling rates. Their results showed that a fraction as low as 4-6% of probe vehicles on a link could produce estimates with an error of 6-8% in travel time predictions.

Srinivasan and Jovanis (1996) developed a mathematical formulation of the minimum probe sample size required for reliable travel time estimation. The formulation was based on the assumption that the number of probes for a given measurement period would be sufficiently numerous (>25), for the central limit theory to be applicable. They showed that the minimum sample size required for a given period is dependent on the desired confidence level, maximum allowable error threshold, and coefficient of variation of travel time estimate. The authors stated that if the number of probes was fewer than 25 per interval, the central limit theorem would no longer be applicable and hence the previous
formulation would be invalid. Another important criterion when choosing the probes’ sample size is area coverage. A general heuristic algorithm was developed to reliably estimate the minimum number of probes required for specific network coverage. Simulation was employed to constitute the confidence levels for each probe sample size and network coverage. The authors concluded that probes can significantly cover a greater portion of freeways than arterials and other road classes.

Sen et al. (1997) investigated the relationship between travel time variance and the number of probe vehicles. They concluded that the variance of mean link travel time does not approach zero even when using a larger sample size of probes. In addition, at a certain sample size, more probes will not significantly decrease travel time variance. A final conclusion was that vehicles’ travel times on the same link at the same measurement period are always correlated, with the exception of lightly travelled links.

Hellinga and Fu (1999) examined the impact of sampling bias on the accuracy of probe vehicle estimates. An analytical model was developed based on queuing theory to prove that some specific sources of bias between vehicle population and probes can lead to a systematic bias in the sample estimate of the mean delay. The authors defined several sources of bias that can lead to these biased estimates of delays. The first source was the dissimilarity in link entry time distribution between the vehicle population and the probes. Link entry time was shown to be dependent on the turning movement required to access the link. Another source of bias was the type of movement at the downstream of the link. To support the results of their analytical model, the authors developed a microsimulation model for a signalized urban arterial using INTEGRATION. Three levels of analysis were carried out using one unbiased sample and then two biased samples. Results showed that biased samples can significantly underestimate/overestimate link travel times.

Chen and Chien (2000) suggested that for the formula of Srinivasan and Jovanis (1996) to be valid, one of two conditions had to be satisfied: a large sample size, or a normal vehicle travel time distribution on the travelled link. The authors used CORSIM to simulate a freeway segment which was divided into nine links. Individual vehicle travel times were collected every 5 minutes. The results showed that the normality assumption held for some
For links which had normal vehicle travel time distribution, the formula based on the central limit theorem was used to compute the required sample size. For other links on which vehicle travel time was not normally distributed, a heuristic model was developed to compute the required sample size. Under the stated traffic volumes, the authors found that a 3% market penetration of probe vehicles would be sufficient for a 95% significance level and an allowable error of 5% in travel time estimation. The authors showed that under low and high demand levels, the required sample size would be high, due to large travel time variance during free and congested flow conditions. During moderate demand levels, the sample size was lower.

Ygnace et al. (2000) developed two models to compute the sample size required to use cell phone-equipped vehicles as probes: analytical and simulation. The general results of the analytical model showed that a sample of 5% of probe vehicles would be sufficient to sample most links in the tested network in a 15-minute time interval. The simulation model compared travel time estimates of different penetration levels of probe vehicles: 5% to 50%, with increments of 5%. The authors tested 6 case studies in order to assess the impact of different network geography and traffic conditions on travel time estimation from probe vehicles. The results showed that with a low market penetration of 5% for cellular-probed vehicles, errors in travel time estimates can vary between 0-7% for the tested networks and traffic conditions.

Yim and Cayford (2001, 2002) presented three sampling methods to obtain the sample size required for reliable speed estimation. The first method was a statistical procedure that assumed a known standard error of speed before sampling. In the second sampling method, microsimulation was employed. The third method was based on the work of Ygnace et al. (2000), in which the authors related link length, density, and market penetration to network coverage. The main conclusion of this research was that a 4-5% market penetration level of probe vehicles can provide reliable travel time estimates. They also found that 150-160 observations per hour are sufficient for reliable speed estimation.

Chen and Chien (2001) studied the influence of increasing sample size on the accuracy of travel time prediction. Two samples of probe vehicles were used, representing 1% and 3%
of total network volume. They showed that as the sample size increased, the error measurements declined, however, the drops in the errors’ magnitudes were insignificant.

Cheu et al. (2002) explored the sample size required to accurately estimate average link speed. A microsimulation model was used to test different market penetration levels. It was shown that as the percentage of probe vehicles increased, the speed estimates from the probe vehicles approached their true values, as obtained from all vehicles. The authors noticed a “knee” at the 15% probe-level in the graph relating the percentage of probe vehicles and the standard error of average speed. This means that after a level of 15% of probe vehicles in the network is attained, the improvement in average speed accuracy will not be significant. Another finding of this research was that the required aggregate sample size over a network was in the order of 4% to 5%, or 10 probes per link within a sampling period of 700 seconds, and the maximum error in link average speed was not more than 5 km/hr for 95% of the measurement period.

Green et al. (2004) used the central limit theorem to estimate the probes’ sample size on a dynamic basis. The hypothesis was that travel time variance changes from one link to another, and changes on the same link occur from one time interval to another. Therefore, the sample size on each link has to change from time to time, according to travel time variance change. A comparison between the developed dynamic approach and the static one was carried out. In the static approach, the probes’ percentages on each link for each time interval was set at 5%. The results of this analysis showed that the static approach was unable to capture the dynamic nature of traffic demand. More specifically, the static method underestimated the sample size for higher levels of required accuracy and overestimated sample size for lower levels of required accuracy.

Cetin et al. (2005) investigated the appropriate sample size of probe vehicles required for reliable travel time estimation. A re-formulation for the sample size problem using a travel time estimation algorithm was presented. Paramics microsimulation was used because of its ability to simulate GPS-equipped vehicles. Average link travel times were obtained at 5 minute intervals. The true travel time for each 5-minute interval was estimated using data from all vehicles. Eight levels of sampling were tested, using 1 to 8 probes selected at
random, to compute the travel time for each interval. Generally, as the number of probes increased, the sum of squared residuals (error term) decreased. However, only marginal accuracy improvements were observed with the increase in number of probes. The authors formulated a travel time estimation algorithm using a simple exponential smoothing approach. The approach utilized travel time estimates of a previous time interval and incorporated new observations of travel times. The authors compared travel time estimates obtained directly from probe vehicles to those obtained from their model. The results showed that the latter approach outperformed the former in terms of error reduction.

Ishizaka et al. (2005) used a simulation model to develop a relationship between network coverage, measurement period, and percentage of taxis as probes. They used Paramics to simulate a part of the road network in the city of Bangkok, Thailand. Their analysis showed that using a measurement period of 15 minutes and 100% of taxis as probes, the network coverage would be slightly over 40%.

As presented previously, the existing literature suggests using a sample as low as 4-6% of the total network volume. In practice, this coverage can still be difficult to achieve. A proposed solution for this problem is to make use of travel time covariance between nearby (neighbour) links to estimate travel times on links with no data using neighbour links travel time data. The methodology of neighbour links travel time estimation was previously presented in chapter four. A proof of concept using field data was also introduced in chapter four. In this chapter, data generated from the microsimulation model of downtown Vancouver were used to investigate the impact of probe vehicles’ sample size on the proposed methodology. Also, fusion of neighbour links data and link historical data was presented.

5.2 DATA GENERATION

The microsimulation model of downtown Vancouver as described in chapter three was used to generate the data required for sample size analysis.
5.2.1 Historical Data

To generate the historical data required for the analysis, the network updated OD matrix was scaled by different factors starting from 60% and increasing to 100% in 10% increments to reflect different demand levels. The dynamic assignment model was run for each volume level until convergence (i.e. user equilibrium) was attained. The simulation period was one hour in duration. Average link travel times were measured on five analysis corridors in downtown Vancouver (Figure 5.1). The five corridors are comprised of 35 travel time sections. Noteworthy is that the section numbers of these links are different from the numbers used in real-life data analysis. Average link travel times were measured using virtual detectors placed directly after the stop line of each link to capture the delay at the downstream signal. The data aggregation interval used in this analysis was 10 minutes. Noteworthy is that in real-life, this value can be pre-defined by a Traffic Management Centre (TMC) and cannot be changed.

For a measurement period of 10 minutes and 5 volume levels, every link recorded 30 observations. Average link travel times were calculated using data from all vehicles. These data represented the historical data set and was used to investigate the travel time correlation between the 35 travel time sections. This data set was also used to define link neighbours and develop statistical models that relate average travel times on each link to its neighbours.

5.2.2 Real-time Data

To generate the real-time data set, the network volume OD matrix was again scaled by 75%, 85%, and 95%. Three different levels of probe vehicles were tested: 1%, 3%, and 5% of the total network volume under each demand level (i.e. 9 simulation scenarios in total). The same travel time virtual detectors were used to obtain average link travel times. The average travel time of each link, as obtained from all vehicles, was used as the true (i.e. validation) travel time. Neighbours’ travel times data, as obtained from probes of nearby links, were used as the data for online travel times. For the three volume levels, one hour of simulation, and a measurement interval of 10 minutes, the number of validation records totalled 18 observations.
5.3 RESULTS

5.3.1 Neighbour Links Travel Time Estimation Models

Link neighbours were identified using a correlation threshold of 0.3 with a positive Lower Limit (LL) of the correlation and \( p < 0.05 \). Four travel time segments were randomly selected from the 35 segments as an example of links with unknown travel times. The section numbers of the four selected segments are 14, 19, 23, and 28 (Figure 5.1). Each of these sections had at least one neighbour link. Similar to the analysis carried out in chapter four, regression exponential models that relate target link travel times to travel times of neighbour links, were developed. Each model involved only one dependant and one explanatory variable. According to the number of neighbours defined for the four segments, 25 exponential models were developed. A simple weighting scheme was applied to combine estimates of different neighbour models. The chosen scheme assigned weights
by model’s variance \((\sigma_i^2)\) as in Equations (4.4) and (4.7). Using this variance to assign weights accounts for the reliability of travel time estimate of each model. The weighting scheme is expressed mathematically as:

\[
w_i = \frac{1/\sigma_i^2}{\sum_{l=1}^{m} 1/\sigma_l^2}
\]

(4.4)

\[
\sigma_i = \sqrt{\frac{SSE_i}{df}}
\]

(4.7)

Where:

\(SSE_i\) = sum of squared errors of model relating unknown link travel time to neighbour \(i\)

\(df\) = degree of freedom = \(n\) - number of parameters to be estimated.

Recalling from chapter four:

\[
\hat{x} = \alpha \cdot x_{rn} + (1 - \alpha) \cdot x_{hl}
\]

(4.8)

Where \(\alpha\), \((1 - \alpha)\) represent weights for \(x_{rn}\) and \(x_{hl}\), respectively. In this analysis, \(x_{rn}\) was estimated using a sample of passenger probe vehicles that exist on neighbouring links. An Empirical Bayes (EB) method was proposed in chapter four to compute \(\alpha\) based on the variance of the neighbour links models. In this method, \(\alpha\) can be calculated as:

\[
\alpha = 1/[1 + \frac{\text{var}(x_{rn})}{\mu(x_{rn})}]
\]

(4.9)

In this chapter, two additional methods were compared to the EB method. In the first method, historical records were used to compensate for missing probes travel time records during certain time intervals. Nevertheless, historical records were not used when neighbour links probes data were available. This method will be referred to as “Historical” and can be expressed by the following rule:

\[
\text{IF } x_{rn} = \text{zero THEN } \hat{x} = x_{hl} \text{ OTHERWISE } \hat{x} = x_{rn}
\]

(5.1)
In the second method, the weight $\alpha$ is computed using the variance of the historical data and the variance of neighbour links models. This method will be referred to as “Variance Weights” and can be mathematically expressed as:

$$
\alpha = \frac{\frac{1}{\text{Var}(x_{rn})}}{\frac{1}{\text{Var}(x_{rn})} + \frac{1}{\text{Var}(x_{hl})}} = \frac{\text{Var}(x_{hl})}{\text{Var}(x_{hl}) + \text{Var}(x_{rn})}
$$

(5.2)

A similar weighting scheme to Equation (5.2) is found in Pu et al. (2009).

Where:

$\text{Var}(x_{rn}) = \text{Variance of the estimated real-time average travel time on link } l \text{ using data of neighbour links } n \text{ during } t$

$\text{Var}(x_{hl}) = \text{Variance of the historical average travel time on link } l \text{ during } t$

$E(x_{rn}) = \text{Expectation of the estimated real-time average travel time on link } l \text{ using data of neighbour links } n \text{ during } t$

5.3.2 Applying Weighting Schemes to the Developed Models

Thus far, historical data have been used to develop travel time relationships between neighbour links. Travel time data of neighbour links were obtained from probe vehicles. Neighbourhood models were used to estimate travel times on segments 14, 19, 23, and 28. For each measurement interval that did not have neighbours’ travel time records, the estimate of the neighbour link model was replaced by “zero” indicating that no data were available for the neighbour during this measurement interval. Moreover, in real-life, an outlier travel time record might be found on one of the neighbours due to a potential unusual event. To remove these outliers, a filtering scheme was applied to the estimates as follows:

$$
\begin{align*}
    x_{rl_{Filtered}} &= x_{rl_{Model}} & \text{If } & x_{rl_{Model}} \leq \mu + 2\sigma \\
    x_{rl_{Filtered}} &= N/A & \text{If } & x_{rl_{Model}} > \mu + 2\sigma
\end{align*}
$$

Where:

$x_{rl_{Model}} = \text{Model estimate of average travel time on link } l \text{ using data of neighbour } i \text{ during } t$
\( x_{ri_{Filtered}} \) = Filtered model estimate of average travel time on link \( l \) using data of neighbour \( i \) during \( t \)

\( \mu \) = Mean historical travel time on link \( l \) during \( t \)

\( \sigma \) = Standard deviation of the mean historical travel time on link \( l \) during \( t \)

In this filter, if the model estimate exceeds the \( \mu + 2\sigma \) value at a specific measurement interval, the links data were excluded from the set of neighbours during this interval. In real-life, the filter threshold value can be set as the maximum historical recorded travel time on this link during the same measurement time. The \( \mu + 2\sigma \) value theoretically corresponds to a 95% confidence level of the mean travel time for normally distributed travel times.

After refining the single model’s estimates, a weighting scheme using model variance was applied to combine estimates of all neighbours. The three data fusion methods were applied to add historical data to neighbour links travel time estimates. The true link travel times, as obtained from the simulation runs for the three validation scenarios, were compared against the fusion estimates using the MAPE. The MAPE was calculated as the average for the 18 intervals (i.e. 3 demand levels and 1 hour divided into 6 intervals), and the four segments. The average MAPE for the three methods is illustrated in Figure 5.2.

![Figure 5.2 MAPE of the Three Data Fusion Methods](image-url)
As shown in Figure 5.2, the estimation error declines with an increase in the probes’ sample size. The first and third data fusion methods are more sensitive to sample size changes. To illustrate, increasing the probes sample size caused significant decreases in the estimation error of the first and third data fusion methods only, while changes were noted to be insignificant when using the EB method. In general, the average MAPE was less than 18% for the three data fusion methods. This indicates that the methodology would provide acceptable accuracy even at very low market penetration rates close to 1%. Weighting by variance was shown to outperform the other two data fusion methods at the three market penetration levels.

In addition to the calculated MAPE, the coverage for each link and sample size were calculated as the percentage of time in which real-time neighbours data were available (e.g. number of intervals with real-time neighbours data/18). The results showed that, on average, the coverage attains about 80% at a 5% probes level. Detailed results are presented in Figure 5.3.

![Figure 5.3 Probe-Neighbours Coverage at Different Market Penetration Levels](image)

Plotting the detailed graphs of $\alpha$ and the three data fusion methods, it was found that increasing the sample size will always lead to higher weights being assigned to real-time neighbours data. Interestingly, this implies that the methodology is sensitive to sample size so that neighbour links travel times will be assigned higher weights as the probes sample size increases. The justification is that increasing the probes sample size leads to the
existence of probes on more neighbours. Consequently, the potential for having data on neighbours with stronger travel time correlation, increases. In turn, this adds more relevant (or accurate) information to travel time estimates of neighbouring links and improves the overall estimation accuracy. An example of $\alpha$ plots for Link 28 is presented in Figure 5.4 for the three market penetration levels. Note that the value of $\alpha$ in the first data fusion method will always be binary of 0 or 1. In the other two methods, $\alpha$ can vary between 0 and 1.

Using the EB method to compute $\alpha$ is only dependent on the statistical properties of the estimated travel time: $E(x_{rn})$ and $Var(x_{rn})$. Hence, it neglects the statistical properties of the historical data. This can lead to very low weights assigned to real-time neighbour travel time estimates, $E(x_{rn})$ in cases where they have high variances, although historical records might also have high variances, and vice versa. This case is clear in Figure 5.4, where the variance of neighbour links estimates was low, and therefore, the EB method assigned weights close to 1 to $E(x_{rn})$. However, because historical records also had low variances, weighting by variance accounted for the variability (i.e. variance) in both the historical data and the real-time neighbours’ estimates. The high weights that were assigned to neighbours’ estimates changed significantly to lay in the range of 0.3-0.9 when weighting by variance.

Figure 5.5 shows plots of estimated travel times using the variance weighting method versus the true travel times, as obtained from the validation runs for a sample size of 5%. The graphs show close matching between estimated travel times and true travel times. Indeed, the estimation accuracy is not perfect. Some intervals may have higher errors than others. Also, the accuracy varies from one link to another, due to the natural variations in link characteristics. The overall average MAPE was about 12.7% for all links at a market penetration level of 5%. This accuracy level was considered acceptable taking into account the high travel time fluctuations in the study area and the complex traffic patterns which included pedestrians, signalized intersections, shared lanes, etc.
Figure 5.4 Calculated $\alpha$ for Link 28
Figure 5.5 Estimated vs. True Link Travel Times (Variance Weighting)
Building upon the results obtained, it can be concluded that neighbour links travel times can be used to estimate travel times on nearby links with reasonable accuracy even in the existence of small samples of probe vehicles. Weighting by variance was found to be the most efficient data fusion method of the three tested methods. Finally, the estimation accuracy improves about 3.5% when the sample size is increased from 1% to 5%. Weights of neighbouring links travel times also increase with sample size increases.

### 5.4 SUMMARY AND CONCLUSIONS

Neighbour links are nearby links that have similar characteristics and are subject to similar traffic conditions. Strong travel time covariance exists between neighbour links for various reasons. This covariance can be useful in developing travel time relationships between neighbour links. These relationships, in turn, can be used to estimate travel times on links with no data, if the data are available for their neighbours. A general framework to integrate historical link travel time data with sparse neighbours travel time data for network travel time estimation was proposed. The methodology was applied to a case study in downtown Vancouver. Three different methods of data fusion were introduced to combine average historical travel times of the link itself with estimates of neighbouring links models. Three different market penetration levels of probe vehicles were also tested. It was shown that assigning weights to historical data and neighbour links data using their variances, outperforms the other two methods. In general, the travel time estimation error was around 17% for a sample size of 1%, while it was approximately 12.7% when using a sample size of 5%. This accuracy level was considered acceptable considering the high travel time fluctuations and the complex traffic patterns in the study area. It can be concluded that neighbour links travel time data can provide a useful source of travel time information for links that have similar traffic characteristics which are not covered by existing traffic sensors. The approach was proven to be useful even when using a small sample of probe vehicles in the network.
CHAPTER 6: ESTIMATION OF NEIGHBOUR LINKS TRAVEL TIMES USING BUSES AS PROBES

6.1  INTRODUCTION

Using buses as probes offers a number of advantages, as buses cover a large portion of urban networks, and the equipment required for data collection has usually already been installed by transit operators. A transit vehicle experiences more delays than automobiles due to acceleration/deceleration, stoppage at bus stops (dwelling time), and their relatively lower speeds. Despite the fact that transit vehicles and automobiles have different running behaviours, a relationship can be developed to estimate automobile travel times using transit data.

Transit vehicles run along predefined routes in the network, hence travel time estimation using buses as probes is usually limited to these routes. In this chapter, a method is proposed to estimate travel times on segments without real-time travel time information, using bus travel time data of nearby (neighbour) links. Unlike other probe vehicles, such as taxis, the routes and schedules for transit vehicles are predictable. This enables the identification of link neighbourhood \textit{apriori}. Furthermore, the number of transit vehicles that traverse each segment within a measurement interval can be predicted with reasonable accuracy (unlike the variability in coverage using passenger probe vehicles).

6.2  PREVIOUS WORK

Many researchers have investigated the potential use of transit vehicles as probes. A review of some recent research efforts that studied the use of transit vehicles as probes is presented.

Hall et al. (1999) investigated the potential for using data collected from transit vehicles to estimate automobile speeds and travel times. The authors suggested that bus tracking provided many potential benefits such as drivers’ adherence to schedules, dispatchers’ response to problems, schedulers’ allocation of adequate time between schedule checkpoints, and travellers having the benefit of access to real-time information on bus
arrivals. The proposed transit as probes system suffered a number of problems that included missing data from failed units, incomplete coverage on routes, and the inability to immediately update data at schedule changes. The authors concluded that the reliability expectations for an actual deployment of a transit as probe system were not met. In addition, they found little correlation between transit speeds and automobile speeds in normal conditions, while strong correlation existed between transit vehicle delays and automobile delays, when major incidents occurred. The analytical model that was used to estimate average link speed was:

\[
\text{Estimated Average Speed} = \frac{(N_1 \times SL)}{(ST - SDT - N_2)}
\]  

(6.1)

Where:

\( SL \) = physical length of segment  
\( ST \) = measured time to traverse the segment  
\( SDT \) = station dwell time  
\( N_1, N_2 \) = empirical coefficients to compensate for performance differences between automobiles and buses

Cathey and Dailey (2002) described a system that used transit vehicles as probes to collect travel time and speed data. Data from the King County Metro Transit AVL system in Washington State were used in their analysis. The developed system was composed of a tracking module with a Kalman Filter to smooth position and speed estimates, and a speed estimation algorithm. The authors developed a system of virtual probe sensors based on information obtained from the tracking and speed estimation modules. They found that the raw data had large variability and they therefore used a simple exponential smoothing equation to smooth their estimates. They showed that the smoothed speed estimates were similar to those obtained from loop detectors in terms of variability throughout the day.

Rather than creating a number of virtual sensors as was accomplished in their previous paper, Cathey and Dailey (2003) developed an approach to estimate corridor travel time and speed. A corridor(s) was defined and travel time and speeds were determined using information obtained from transit vehicles on this/these corridor(s). A comparison of
corridor speed profiles estimated from probes and inductive loop detectors showed some similarity, with transit probe speed estimates providing lower (i.e. more conservative) speeds.

In both of the previous two papers, the authors used transit vehicles to obtain travel times and compared them to the data obtained from inductive loop detectors. The studies implicitly assumed that transit vehicles are representative of the traffic stream. This assumption is not completely accurate. This is attributed to the fact that transit vehicles and automobiles have different running behaviours. Most of the results presented in the two previous papers showed that the spatial and temporal speed variability of transit vehicles data was similar to the variability of loop detectors data. The word “similar” suggests that the trends were similar but had different values. For example, in Cathey and Dailey (2002), the median speed profile from the detectors’ data was shifted upwards by 8 mph from the median speed profile of transit probes. The authors’ hypothesis could have led to more accurate results if a robust procedure had been adopted to isolate the impact of different factors which caused the behaviour to be different. Consequently, a relationship could have been developed to estimate automobile travel times using transit data.

Tantiyanugulchai and Bertini (2003a) compared speeds and travel times of probe buses to those obtained from GPS-instrumented vehicles. Each probe bus was equipped with a Differential GPS (DGPS) unit, an Automatic Passenger Counter (APC), and a wireless communication link to the transit centre. The recorded information included bus arrival and departure times at geo-coded bus stops, number of passengers alighting and boarding at each bus stop, and the maximum instantaneous speed achieved between two stops. The authors used bus trajectories to obtain the mean travel time and travel speed of buses. Consequently, they presented two imaginary scenarios involving what they called the “hypothetical” bus and the “pseudo” bus. The hypothetical bus, as defined by the authors, is a bus which does not stop for passengers. Hence, the travel time of a hypothetical bus is computed as the running time minus the total stopping time at all stops. A pseudo bus is a bus that runs between each two stops with the maximum recorded instantaneous speed between this pair of stops. Comparing test vehicles mean speed to the real bus, the pseudo bus, and the hypothetical bus mean speeds, the results showed that the average test vehicle
speed was about 1.66 times greater than real bus speeds. While the test vehicle speed was about 0.79 of the pseudo bus speed, the hypothetical bus speed was equivalent to 1.03 of the test vehicle speed.

Tantiyanugulchai and Bertini (2003b) extended their previous work by using bus data collected during a different day. The findings of this study were similar to their previous research findings. The ratios between the test vehicle speed and the actual bus speed, the pseudo bus speed, and the hypothetical bus speed were 1.63, 0.84, and 1.35, respectively.

Bertini and Tantiyanugulchai (2003) carried out the same analysis on data collected for the eastbound direction. The authors repeated the same work for the hypothetical and pseudo buses and added what they called “modified pseudo buses.” The concept of these imaginary buses is that they run with the maximum instantaneous speed recorded between each two stops, and dwelling time is added to their running times. Comparing the test vehicle travel time and the pseudo bus travel time, the ratio between the former and the latter was 1.36. It was also shown that the test vehicle speed was about 72% of the maximum instantaneous speed achieved by the buses.

Bertini and El-Geneidy (2004), used an analytical model to compute total transit travel time on a freeway. Their model was:

\[
Bus Travel Time (route) = T_0 + c_1 \cdot N_d + c_2 \cdot N_a + c_3 \cdot N_b
\]

Where:

\(T_0\) = average nonstop trip time of a bus

\(N_d\) = number of times a bus stops (dwells)

\(N_a\) = total number of passengers alighting a bus

\(N_b\) = total number of passengers boarding a bus

The model did not include components to account for delays caused by control devices and mid-block bus stops.

Chakroborty and Kikuchi (2004) investigated the use of transit vehicles as probes to provide automobile travel time information. Two major issues were investigated: data
stability and the required adjustments. Data were collected for bus travel times (BTT) and average automobile travel times (ATT) for five arterial segments during the a.m. peak, p.m. peak, and off-peak period. Both BTT and ATT data were collected using manual measurements. A linear regression model was developed to relate the BTT and the ATT. This model had the general form:

\[ ATT = A + b (BTT - TST) \]  

(6.3)

Where:

\[ TST = \text{total stopping time at all bus stops} \]

A model was developed for each of the analyzed arterial segments. The models performed well in terms of providing low estimation errors. Another modification of this general model was carried out such that:

\[ ATT = \frac{\text{Length of Travel Section}}{\text{Free Flow Speed}} + b(BTT - TST) \]  

(6.4)

Results improved when the new model was applied. The percentage of data points with an error of less than 10%, increased, which indicated better fit to the data. The authors noticed that the coefficient \( b \) had a narrow range, and, hence, they suggested using two general formulae for less frequently and more frequently congested roads. One major shortcoming of this research is that the same data set was used for both the calibration and the validation of the regression models. In linear regression, the coefficients are estimated by the least square method that minimizes the sum of squared residuals. In this context, it is expected that there will be minor estimation errors when using the same data for both models’ development and validation.

Pu et al. (2009) used a slightly different approach in relating bus and automobile travel times. Regression models that relate bus and car space mean speeds were developed using historical data. The mean and variance of the bus speed were calculated for each measurement interval, and consequently, confidence intervals of the bus speeds were obtained. Upon receiving new bus speed records in real-time, the mean of the new record is compared against the historical mean boundaries. If the mean of the new speed records
falls within the historical boundaries, the new records are discarded and the historical mean speed is used along with regression models to estimate average automobile speed on the link. In case the mean of the real-time speed record is outside the boundaries of the historical mean, it is used to update the historical mean using the variances of the historical mean and the real-time mean. The regression models developed in this research did not include either the dwelling time nor the acceleration/deceleration time. The estimation errors varied between -18% to 39% for the two segments and two peak periods tested.

Uno et al. (2009) used bus probes data to study travel time variability on urban corridors. Acceleration, deceleration, and stopping times were estimated and eliminated from bus travel time to estimate automobile travel times. The authors did not report whether the speed profiles of buses and vehicles were similar, nor how they accounted for any potential variability.

In summary, previous research efforts focused on using transit as probes to obtain travel times and speeds of the link itself. Most of the previous studies used regression models to estimate automobile travel time using bus travel times. In this chapter, the use of transit vehicles travel times for travel time estimation on neighbouring links, is proposed. Additionally, several fusion methods are compared to combine estimates from neighbour links transit data and links historical average data.

### 6.3 DATA GENERATION

The microsimulation model of downtown Vancouver, as described in chapter three was used to generate the data used in this chapter.

#### 6.3.1 Transit Routes in Downtown Vancouver

The simulation model included 46 transit lines which were updated according to recent routes and schedules. In general, bus frequency in downtown Vancouver ranges from 8 to 10 minutes. Most of the transit routes in downtown Vancouver run through four major streets: Richards, Howe, Seymour, and Burrard Streets (Figure 6.1). Seymour and Burrard Streets have HOV lanes dedicated only for bus traffic during the a.m. peak. The travel
behaviour, and consequently bus travel time, on HOV segments, is not indicative of that characterized by the general traffic stream. These segments were excluded, therefore, from the analysis. Furthermore, two individual segments on Howe Street did not have HOV lanes and were included in the analysis. In total, eight transit sections were analyzed, of which four segments had at least one bus stop (group 1) while the other four had none (group 2). Table 6.1 shows summary statistics of the analyzed segments, while Figure 6.1 shows the section ID for each transit travel time segment.

Table 6.1 Summary Statistics of Travel Times of the Analyzed Segments

<table>
<thead>
<tr>
<th>Street</th>
<th>Group</th>
<th>ID</th>
<th>Section ID</th>
<th>Number of Transit Lines</th>
<th>Length (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richards St.</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>180.1</td>
</tr>
<tr>
<td>Richards St.</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td>166.1</td>
</tr>
<tr>
<td>Richards St.</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>1</td>
<td>163.3</td>
</tr>
<tr>
<td>Howe St.</td>
<td>1</td>
<td>26</td>
<td>26</td>
<td>1</td>
<td>181.6</td>
</tr>
<tr>
<td>Richards St.</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>173.2</td>
</tr>
<tr>
<td>Richards St.</td>
<td>2</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>166.4</td>
</tr>
<tr>
<td>Richards St.</td>
<td>2</td>
<td>14</td>
<td>14</td>
<td>1</td>
<td>164.6</td>
</tr>
<tr>
<td>Howe St.</td>
<td>2</td>
<td>27</td>
<td>27</td>
<td>4</td>
<td>174.3</td>
</tr>
</tbody>
</table>

Figure 6.1 Analyzed Transit Sections ID
6.3.2 Historical Data

To generate the historical data required for this analysis, the network updated OD matrix was scaled by different factors starting from 60% and increasing to 100% at 10% increments, to reflect different demand levels. The dynamic assignment model was run for each volume level until convergence. Average link travel times were measured on 35 travel time sections, including the eight transit travel time sections. Link travel times were measured using virtual detectors placed directly after the stop line of each link. For each link, the generated information included average automobile travel time and average bus travel time, if any. The data aggregation interval used in this analysis was 10 minutes.

For a measurement period of 10 minutes and 5 volume levels, every link had 30 observations. This data represented the historical data set and was used to investigate the travel time correlation between the 35 travel time sections. This data set was also used to define link neighbours and to develop statistical models that relate automobile travel times on each link to bus travel times on its neighbours.

6.3.3 Real-time Data

To generate real-time data set, the network volume OD matrix was again scaled by 75%, 85%, and 95%. The same travel time virtual detectors were used to obtain the travel times of the transit probe vehicles and automobiles. The average automobile travel time for each section was used as the true (i.e. validation) travel time, while average bus travel times were used as the online neighbours’ travel times. For the three volume levels, one hour of simulation, and a measurement interval of 10 minutes, the number of validation records was 18 observations.

6.4 RESULTS

6.4.1 Neighbour Links Travel Time Estimation Models

In chapter five, link travel time as estimated from neighbour links models, denoted by $x_{nm}$, was computed using data from a sample of passenger probe vehicles that existed on
neighbouring links. In this chapter, bus travel times of neighbour links were used instead of vehicle probes. The three fusion methods, as described in chapter five, were used again here to fuse links historical data and neighbour links real-time bus data.

Pearson Correlation Coefficient ($r$) was used to define neighbour links. The $p$-value was used to test the null hypothesis of correlation. The correlation matrix was established for the 35 auto travel time sections and the 8 bus travel time sections (for a total of 35 sections). For the four transit sections with bus stops, the variable used was “average bus travel time – (average dwelling time × # of bus stops on the section).” Consequently, an arbitrary correlation threshold of 0.3 was used to define link neighbours for each transit segment. To further refine the selection of neighbours, records that had $p \geq 0.05$ were excluded from the selection of neighbours. Records with a negative Lower Limit (LL) of the correlation were excluded. Table 6.2 shows the defined neighbours for these transit sections.

As shown, some transit sections did not have any neighbour that satisfied the neighbourhood criteria. For all other neighbour pairs, the null hypothesis of zero correlation was rejected with $p < 0.05$.

The same four travel time segments, randomly selected in chapter five, were analyzed again in this chapter. This enables a comparison of the results of using samples of passenger probes versus transit probes. This also enables the fusion of both sources of information, as will be presented in the next chapter. The section numbers for the four selected segments were 14, 19, 23, and 28. As shown in Table 6.2, each of these sections had at least one neighbour link with bus data. Six regression models were developed according to the number of neighbours defined for the four travel time segments, as in Table 6.2. Estimates of different models were combined using the models’ variances.
Table 6.2 Defined Neighbours for Transit Sections

<table>
<thead>
<tr>
<th>Transit Segment ID</th>
<th>Neighbour Link ID</th>
<th>Correlation Estimate</th>
<th>95% Confidence Limit</th>
<th>p-value H₀: r = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>B12</td>
<td>C14</td>
<td>0.48</td>
<td>0.09</td>
<td>0.74</td>
</tr>
<tr>
<td>B14</td>
<td>C18</td>
<td>0.94</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>B9</td>
<td>C19</td>
<td>0.48</td>
<td>0.14</td>
<td>0.72</td>
</tr>
<tr>
<td>B14</td>
<td>C19</td>
<td>0.94</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>B9</td>
<td>C20</td>
<td>0.55</td>
<td>0.23</td>
<td>0.76</td>
</tr>
<tr>
<td>B14</td>
<td>C20</td>
<td>0.78</td>
<td>0.54</td>
<td>0.90</td>
</tr>
<tr>
<td>B14</td>
<td>C21</td>
<td>0.49</td>
<td>0.09</td>
<td>0.75</td>
</tr>
<tr>
<td>B9</td>
<td>C21</td>
<td>0.51</td>
<td>0.19</td>
<td>0.74</td>
</tr>
<tr>
<td>B9</td>
<td>C22</td>
<td>0.41</td>
<td>0.05</td>
<td>0.67</td>
</tr>
<tr>
<td>B14</td>
<td>C22</td>
<td>0.60</td>
<td>0.25</td>
<td>0.81</td>
</tr>
<tr>
<td>B9</td>
<td>C23</td>
<td>0.38</td>
<td>0.02</td>
<td>0.65</td>
</tr>
<tr>
<td>B14</td>
<td>C23</td>
<td>0.43</td>
<td>0.03</td>
<td>0.72</td>
</tr>
<tr>
<td>B9</td>
<td>C28</td>
<td>0.42</td>
<td>0.07</td>
<td>0.68</td>
</tr>
<tr>
<td>B12</td>
<td>C30</td>
<td>0.43</td>
<td>0.04</td>
<td>0.71</td>
</tr>
<tr>
<td>B9</td>
<td>C31</td>
<td>0.44</td>
<td>0.09</td>
<td>0.69</td>
</tr>
<tr>
<td>B13</td>
<td>C31</td>
<td>0.45</td>
<td>0.06</td>
<td>0.73</td>
</tr>
<tr>
<td>B9</td>
<td>C35</td>
<td>0.45</td>
<td>0.11</td>
<td>0.70</td>
</tr>
<tr>
<td>B14</td>
<td>C35</td>
<td>0.88</td>
<td>0.74</td>
<td>0.95</td>
</tr>
<tr>
<td>B9</td>
<td>C36</td>
<td>0.40</td>
<td>0.05</td>
<td>0.67</td>
</tr>
<tr>
<td>B14</td>
<td>C36</td>
<td>0.97</td>
<td>0.93</td>
<td>0.99</td>
</tr>
<tr>
<td>B9</td>
<td>C37</td>
<td>0.52</td>
<td>0.19</td>
<td>0.74</td>
</tr>
<tr>
<td>B14</td>
<td>C37</td>
<td>0.76</td>
<td>0.50</td>
<td>0.89</td>
</tr>
<tr>
<td>B14</td>
<td>C38</td>
<td>0.49</td>
<td>0.09</td>
<td>0.75</td>
</tr>
<tr>
<td>B9</td>
<td>C38</td>
<td>0.53</td>
<td>0.21</td>
<td>0.75</td>
</tr>
<tr>
<td>B14</td>
<td>C39</td>
<td>0.50</td>
<td>0.10</td>
<td>0.75</td>
</tr>
<tr>
<td>B13</td>
<td>C40</td>
<td>0.44</td>
<td>0.04</td>
<td>0.72</td>
</tr>
<tr>
<td>B9</td>
<td>C40</td>
<td>0.70</td>
<td>0.45</td>
<td>0.85</td>
</tr>
<tr>
<td>B9</td>
<td>C6</td>
<td>0.58</td>
<td>0.27</td>
<td>0.79</td>
</tr>
</tbody>
</table>

6.4.2 Applying Weighting Schemes to the Developed Models

Historical data were used to develop relationships between link automobile travel times and neighbours’ bus travel times. Consequently, bus neighbourhood models were used to estimate travel times on segments 14, 19, 23, and 28. For each measurement interval that did not have neighbours bus travel time records, the estimation of the neighbour links model was replaced by “zero,” indicating that no data were available for the neighbour
during this measurement interval. The same filter, proposed in chapter five, was applied to remove outliers of neighbour links travel time estimates.

The three data fusion methods (Equations 3.9, 6.1, and 6.2) were applied to add historical data to neighbour links travel time estimates. The true automobile travel times, as obtained from the simulation runs for the three validation scenarios, were compared against models’ estimates using the MAPE. The MAPE was calculated as the average for the 18 intervals and the four segments. The average MAPE of the three methods is illustrated in Figure 6.2.

![Figure 6.2 MAPE of the Three Methods of Data Fusion](image)

As shown in Figure 6.2, the estimation accuracy of neighbour links travel times is satisfactory with an average MAPE of less than 20% for the three data fusion methods. Weighting by variance outperformed the other two data fusion methods, with an average MAPE of about 15.4%.

The detailed plots of $\alpha$ for the 18 intervals and the four links are illustrated in Figure 6.3. Note that the value of $\alpha$ in the first data fusion method will always be binary 0 or 1. In the other two methods $\alpha$ may vary between 0 and 1. Using the EB method to compute $\alpha$ is only dependent on the statistical properties of the estimated travel time, $E(x_{rn})$ and $Var (x_{rn})$. Therefore, it neglects the statistical properties of the historical data. This can lead to very low weights assigned to the real-time neighbour travel time estimates, $E(x_{rn})$, although historical records might also have high variance, and vice versa.
Figure 6.3 Calculated $\alpha$ for the Four Sections
This case was clear when estimating the travel times of Links 19 and 23 which had a high degree of variability in both neighbours’ estimates and historical records. As shown in Figure 6.3, neglecting the variance of the historical data led to assigning very low weights to neighbour links travel time estimates. Weighting by variance accounted for the variability of both the historical data and the real-time neighbours’ estimates. The low weights that were assigned to neighbours’ estimates changed significantly with weighting by variance.

The variances of neighbour links estimates for Links 14 and 28 were high compared to the historical variance. Therefore, whether using the EB weighting or the variance weighting, the value of $\alpha$ was always low. This indicates that the approach using neighbouring links travel time estimation is more useful for links that experience high travel time variability, which are also the most challenging.

Figure 6.4 shows plots of the estimated travel times using the variance weighting method versus true travel times, as obtained from the validation runs. The graphs show similarity between the estimated travel times and the true travel time. Some intervals might have higher errors than others. Accuracy varies from one link to another due to the natural variations in link characteristics. The overall average MAPE was about 15.4% for all links and intervals. This accuracy level is considered comparable with similarly published results in the literature.

Building upon these results, it can be concluded that bus travel times can be used not only to estimate automobile travel time for the same link, but also for neighbouring links. It should be noted that estimation accuracy when using bus data to estimate neighbour links travel times is diminished compared to the accuracy of using bus data to estimate automobile travel time on the same link. It is still, however, within an acceptable limit.
Figure 6.4 Estimated vs. True Link Travel Times (Variance Weighting)
6.5 SUMMARY AND CONCLUSIONS

The potential for using transit vehicles as probes to estimate automobile (i.e. general) link travel time has been extensively investigated in the literature. However, using buses as probes for neighbour links travel time estimation has not been previously researched. Neighbour links are defined as nearby links that share similar characteristics and are subject to similar traffic conditions. This research proposes a general framework to integrate historical link travel time data and sparse bus travel time data for travel time estimation on a network. The methodology was applied to a case study in downtown Vancouver. Average link travel times were estimated using the bus travel times of neighbouring links. Three different methods of data fusion were introduced, to combine average historical travel times of the link itself, with estimates of neighbouring links models. It was shown that assigning weights to historical data and neighbour links data, based on their variances, slightly outperformed the other two methods. The MAPE of neighbouring links travel time estimation, using bus data, was about 15.4%. This accuracy level was considered acceptable in view of the considerable travel time fluctuations in the study area. It was concluded, therefore, that transit vehicles can be used as a useful source of travel time information for their travel links, as well as for nearby links with similar traffic characteristics.
CHAPTER 7: FUSION OF BUS AND VEHICLE PROBES DATA FOR NEIGHBOUR LINKS TRAVEL TIME ESTIMATION

7.1 INTRODUCTION

The use of vehicles as probes for real-time travel time estimation is usually limited to their travel routes. In the previous chapters, the potential use of probe vehicles travel times, to estimate travel times on their neighbouring links, was advocated. A general methodology for neighbour links travel time estimation was presented and validated using field travel time data in chapter four. A sample of probe vehicles was used as the source of neighbour links data in chapter five, while transit vehicles data were used in chapter six. In this chapter, different data fusion methods that can be used to integrate links historical data, and neighbour links real-time data, whether from vehicle probes or transit probes, are explored. Data fusion is meant to refine the identification, classification, or estimation of an object or a variable using the data of more than one source. Data fusion algorithms vary significantly in their degrees of complexity and procedure. Within the context of this chapter, the variable of interest is the mean link travel time, and the sources of information are historical data, bus travel times of a set of nearby links, and probe vehicles travel times of an additional set of nearby links.

7.2 PREVIOUS WORK

The fusion of Inductive Loop Detectors (ILD) data, and data of a sample of probe vehicles to improve travel time estimation has been investigated in the literature. Choi and Chung (2002) developed a fusion algorithm to estimate dynamic link travel time in urban road networks using probe vehicles data and loop detectors data. The algorithm was based on a voting technique, fuzzy regression, and a Bayesian pooling method. Real-life GPS probes and loop detector data were collected and used to validate the method. The results showed that the proposed fusion algorithm was more accurate than the mean travel time, as obtained from each individual source.
Nanthawichit et al. (2003) developed a Kalman Filtering (KF) model to predict traffic state variables using loop detector and probe vehicles data. The method was applied to synthetic data generated from an INTEGRATION microsimulation model of a freeway section. The method was further extended and applied to short term travel time prediction. The accuracy of the proposed method was shown to be comparable to other existing travel time prediction models.

Xie et al. (2004) compared travel time estimates of loop detectors data only, samples of probes only, and two fusion methods, to combine both sources of information. The two fusion methods were an ANN MLP model, and a linear regression model. The authors showed significant improvements in travel time estimation with the fusion of both data sources, especially when using the ANN model. The importance of loop detectors data was shown to diminish with an increase of probe vehicles sample size. A market penetration level of 4% to 5% of probe vehicles was suggested as the threshold above which loop detector data will not improve travel time estimation accuracy.

Berkow et al. (2009) presented the results of a case study in which data from traffic signal detectors and bus probes were combined and used to obtain improved arterial performance measures. Graphical techniques were developed that traced the spatial and temporal congestion boundaries along an arterial corridor.

El Faouzi et al. (2009) used a Dempster–Shafer (D-S) data fusion model to combine travel time estimates from inductive loop detectors and toll collection stations. The mean link travel time was defined as an interval or range according to the traffic state. Four discrete traffic states that represented different congestion levels were identified. Each of the four states was characterized by a particular travel time interval. For example, the first state, which represented low congestion, was defined by a travel time interval of the range between the free flow travel time and 1.1 of the free flow travel time. For each time aggregation period, the most probable travel time interval was selected as that with the largest belief. This technique did not provide an exact estimation of the travel time, but it identified the level of congestion using the interval of the travel time obtained.
As explained previously, the potential fusion of bus probes data and vehicle probes data has not been previously reported in the literature. In addition, combining neighbour links vehicle/bus probes data to link historical data has not been explored before.

7.3 DATA FUSION SCENARIOS

As previously described, two synthetic data sets were generated from the microsimulation model of downtown Vancouver. The first data set included average link travel times for each 10-minute interval for one hour of microsimulation time and five different network volume levels. This data set included both passenger probes and transit probes average link travel times. These data represented the historical data set and were used to establish a travel time correlation matrix for all analyzed links. Passenger and transit probe neighbours were identified using some pre-set correlation criteria. Regression models were developed to relate the unknown average link travel time of each link to its neighbour(s).

The second data set was obtained by simulating the network under three new demand levels. In chapter five, travel time data, as obtained from three samples of passenger probes, were used as the source of neighbour links travel times. In chapter six, transit vehicles travel times were used as neighbour links travel times. In this chapter, neighbour links travel times of a sample(s) of passenger vehicles and transit probes, in addition to historical data from the link itself, are combined by means of data fusion. Each link had two groups of neighbour links: transit neighbours and vehicle probe neighbours. During the measurement interval, if no real-time neighbour travel time data were available, whether from passenger probes or transit vehicles, historical data of the link itself were used as the best estimate of the average link travel time. If travel time data were available from one or more neighbour groups, the three fusion methods, as described in chapters five and six, were used to combine links historical data and neighbour links real-time data. In total, nine fusion scenarios were analyzed. These included different combinations of data sources, probe sample sizes, and fusion methods.
7.4 NEIGHBOUR LINK TRAVEL TIME ESTIMATION MODELS

The data used in chapters five and six were used in this chapter to fuse estimates of transit probes, as well as samples of passenger probes. The same four travel time segments analyzed in chapters five and six, were also analyzed in this chapter. The section numbers of the four selected segments were 14, 19, 23, and 28. Each of these sections has at least one neighbour link with bus data and one neighbour with passenger probes data. Table 7.1 shows the identified bus and vehicle neighbours for each of the four analyzed sections.

<table>
<thead>
<tr>
<th>Link Id</th>
<th>14</th>
<th>19</th>
<th>23</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighbour 1</td>
<td>12 Passenger</td>
<td>36 Passenger</td>
<td>22 Passenger</td>
<td>6 Passenger</td>
</tr>
<tr>
<td>Neighbour 2</td>
<td>12 Transit</td>
<td>37 Passenger</td>
<td>11 Passenger</td>
<td>11 Passenger</td>
</tr>
<tr>
<td>Neighbour 3</td>
<td>20 Passenger</td>
<td>39 Passenger</td>
<td>22 Passenger</td>
<td></td>
</tr>
<tr>
<td>Neighbour 4</td>
<td>18 Passenger</td>
<td>6 Passenger</td>
<td>13 Passenger</td>
<td></td>
</tr>
<tr>
<td>Neighbour 5</td>
<td>35 Passenger</td>
<td>18 Passenger</td>
<td>39 Passenger</td>
<td></td>
</tr>
<tr>
<td>Neighbour 6</td>
<td>21 Passenger</td>
<td>36 Passenger</td>
<td>35 Passenger</td>
<td></td>
</tr>
<tr>
<td>Neighbour 7</td>
<td>38 Passenger</td>
<td>9 Transit</td>
<td>7 Passenger</td>
<td></td>
</tr>
<tr>
<td>Neighbour 8</td>
<td>22 Passenger</td>
<td>14 Transit</td>
<td>9 Transit</td>
<td></td>
</tr>
<tr>
<td>Neighbour 9</td>
<td>9 Passenger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbour 10</td>
<td>39 Passenger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbour 11</td>
<td>5 Passenger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbour 12</td>
<td>9 Transit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbour 13</td>
<td>14 Transit</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two groups of regression models were developed using the least squares method. The first group included models that related target link travel time to average automobile travel time on neighbour links. The second group of models related target link travel time to the average bus travel time of its neighbours. According to the number of neighbours defined for the four segments, 31 models were developed as in Table 7.2.
As mentioned earlier, three sources of information were used to estimate the average travel times of each of the four links for each measurement interval. The first source was average link historical travel time, obtained from the historical data set. The second source was estimates of neighbour links models for transit (bus) neighbours. The third source was neighbour links travel time estimates, using data from a sample of passenger probes.

Bus neighbour links were added to the set of passenger probe neighbours as additional neighbours. In the calculation of $x_{rn}$, the weight assigned to the travel time estimate of each
neighbour, whether it was a passenger probe or a bus probe neighbour, was computed as the inverse of neighbour model’s variance. The three data fusion methods (Equations 3.9, 6.1, and 6.2) were applied to add historical data to neighbour links travel time estimates.

For each measurement interval and volume scenario, the true average link travel time was obtained from all vehicles travelling the link during the 10-minute measurement interval. The true average travel times were compared against those obtained using the fusion methodology in terms of the MAPE. The MAPE was calculated as the average for the 18 intervals. Figure 7.1 shows the errors associated with each data fusion method at different levels of market penetration of the passenger probes.

![Figure 7.1 Data Fusion Estimation Accuracy](image)

In general, the average MAPE was less than 18% for the three data fusion methods. Weighting by variance was shown to consistently outperform the other two data fusion methods, in terms of estimation errors. This indicates that historical data should be considered even in the event of available real-time travel time data from neighbour links. The importance of adding link historical data diminishes with the increase of the sample size of the probe vehicles. Intuitively, real-time travel time estimation accuracy improves when the probe vehicles’ sample size increases, and therefore the value added by utilizing the historical data becomes negligible. To illustrate, in the data fusion method denoted by “Historical,” link historical data would only be considered when no real-time neighbours’ data were available. As the market penetration level of probe vehicles increases, the
possibility of having measurement intervals with no neighbour data decreases. Accordingly, the method becomes almost equivalent to using neighbour links data only for the purpose of travel time estimation.

Next, the accuracy of fusing link historical data and a sample of probe vehicles was compared against the fusion of historical, transit, and passenger probes data. Weighting by variance was considered only because it was shown to outperform the other two weighting methods. Figure 7.2 shows the results of these comparisons.

![Figure 7.2 Fusion Accuracy of Using Passenger Probes only vs. Passenger Probes and Buses](image)

The average estimation error for all cases ranged between 13% and 17%, which is comparable to the results available in the current literature. An important conclusion that can be drawn from Figure 7.2 is that, as the market penetration level of probe vehicles decreases, the benefits of fusing transit neighbours travel time data and passenger probes data become more evident. This is expected to some extent. Higher market penetration levels of passenger probe vehicles increase the estimation accuracy, and, therefore, the addition of transit data is not expected to decrease the estimation error. However, using transit data, with a small sample of passenger probes, could lead to better estimation accuracy.

Figure 7.3 shows plots of the estimated travel times using the variance weighting method versus the true travel times, as obtained from the validation runs for a sample size of 5%.
The graphs show reasonable estimation accuracy. Taking into account the high travel time variability in the study area combined with the short length of the analyzed links, it can be concluded that the fusion methodology is successful in modelling the travel relationship between nearby links.

7.6 SUMMARY AND CONCLUSIONS

The potential fusion of probe vehicles and buses data for travel time estimation has not been previously studied. In this chapter, the fusion of a sample of probes and transit vehicles to estimate travel times on neighbour links was investigated. Three fusion methods were proposed to integrate historical data with neighbour links travel time data. A case study was presented using the microsimulation VISSIM model of downtown Vancouver. Three market penetration levels of probe vehicles were analyzed: 1%, 3%, and 5%. It was shown that the benefits of fusing transit neighbours’, travel time data, and passenger probes data, are more evident with smaller market penetration levels of passenger probes. The benefits of fusion decrease with an increase in the market penetration level of passenger probes.
Figure 7.3 Estimated vs. True Link Travel Times (Variance Weighting)
CHAPTER 8: SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

8.1 SUMMARY AND CONCLUSIONS

In this research, the concept of travel time neighbourhood was introduced. The term “neighbourhood” describes nearby segments that have similar characteristics and are subject to similar traffic conditions within a road network. These nearby segments can be preceding and succeeding segments on a route, parallel segments, or even intersecting segments. Strong travel time correlation exists between nearby links for various reasons. This correlation can be useful in developing travel time relationships between neighbour links. These relationships, in turn, can be used to estimate travel times on links with no data, if data are available for the links’ neighbours.

The main objective of this research was to develop a general framework for the estimation of real-time travel times on a road network using sparse travel time data. Specifically, the purpose was to estimate travel times on links not covered by existing sensors, by using their travel time relationships with neighbour links. Herein lies a potential solution to the problem of having partial network sensor coverage caused by a small sample of probe vehicles, or a limited number of detectors.

The proposed framework can be described by four modules: identification of link neighbours, choice of the modelling technique, application entity (e.g. link or corridor), and source of neighbour links travel time data.

Field travel time data were collected on major north-south street segments in downtown Vancouver, British Columbia, and used to validate the framework. The estimation accuracy of the proposed statistical method was assessed by the MAPE which had an average of less than 14% for four travel time validation segments. This level of accuracy was considered satisfactory in reference to the available literature. Estimation results, as obtained from the statistical method, were compared to similar results from B-Spline neuro-fuzzy and Multi-Layer Perception (MLP) Artificial Neural Network (ANN) models. Error measurements supported the superiority of the B-Spline neuro-fuzzy model and the
statistical method. Although the neuro-fuzzy model was shown to be efficient for neighbour links travel time estimation, it should only be used in limited situations where the data would always be available for all neighbours. In cases where floating sensors, such as probe vehicles, are used to collect travel time data, it becomes difficult to discern which links would be travelled in which time intervals. Therefore, using any model that includes a pre-set number of neighbours is deemed to be problematic.

The concept of neighbour links travel time estimation was extended and applied to estimate average corridor travel time using the available data of a nearby corridor. Real-life simultaneous travel time data were collected on two heavily congested corridors in downtown Vancouver. The travel time correlation of the two corridors was found to be significant. Models, consequently, were developed to estimate travel times on one corridor using data from the other corridor. The developed models included regression and K-Nearest Neighbours (KNN). The estimation accuracy of all models was assessed using several error measurements. The MAPE of all models ranged between 13.7% and 17.1% and, hence, the results were considered satisfactory. The KNN approach was found to be superior to the other models in terms of estimation accuracy.

To analyze the impact of the probe vehicles’ sample size on the proposed methodology, a case study was undertaken using a VISSIM microsimulation model of downtown Vancouver road network. The simulation model was calibrated and validated using real-life traffic volumes and travel time data. Three different levels of probe vehicles market penetration were tested. It was shown that the estimation errors of neighbour links travel times decline steadily when the sample size of the probe vehicles increases. This methodology provided reasonable estimation accuracy, even with small probe samples.

The potential for using bus travel time data to estimate automobile travel times of neighbour links was investigated. Data generated from the microsimulation model of downtown Vancouver were used in the analysis. Link travel time estimation accuracy using bus probes was assessed using the MAPE, the value of which was 15.4%.

Finally, the fusion of bus travel time data and passenger probes data for neighbour links travel time estimation was introduced. Three different fusion methods were applied to
combine link historical data and neighbour links travel times of passenger and bus probes. The fusion error of using buses and passenger probes always tended to be close to the error of the better source of information. Therefore, the fusion of both sources will not deteriorate the estimation accuracy, regardless of which source provides better estimates. It was shown that using transit data for neighbour links travel time estimation can improve the estimation accuracy at low market penetration levels. However, the significance of adding transit data diminishes with the increase of the market penetration level.

Assigning weights to historical data and neighbour links data using their variances outperformed the other two methods. In general, the travel time estimation error, after data fusion, ranged between 13% and 17%. This accuracy level was considered acceptable, considering the high travel time fluctuations in the study area and the complex traffic patterns that included pedestrians, signalized intersections, shared lanes, transit operation, etc. Hence, it was concluded that neighbour links travel times can be utilized as a useful source of travel time information for nearby links that have similar traffic characteristics, even when there is only a small sample of probe vehicles in the network.

It should be noted that the proposed framework is not intended to replace other existing travel time estimation methods. Rather, it is proposed as a complementary model. The objective of the analyses was to demonstrate the feasibility of using neighbour links data as an additional source of traffic information that may not have been explored extensively before. It was proven that this method was useful in estimating average travel times on links/corridors that did not have available real-time travel time data, while having strong travel time correlation with neighbouring links.

### 8.2 RESEARCH CONTRIBUTIONS

This research presented a number of contributions to the field of Intelligent Transportation Systems (ITS). It proposed a potential solution to the problem of limited network sensor coverage. The outcomes of this research can be summarized as:

1. Development of a general framework to expand travel time estimation to an entire urban road network using the data from only part of the network,
2. Proof of concept of the applicability/feasibility of using neighbours data for link/corridor travel time estimation, and,
3. Studying the potential use of different sources of neighbour links data (e.g. samples of passenger probes, transit vehicles) and the fusion of these sources.

8.3 FUTURE RESEARCH

Extensions of this research will focus on two areas: methodology improvement, and large scale deployment.

8.3.1 Methodology Improvement

Several modifications can be applied to improve different aspects of the methodology. Upon data availability, other criteria for neighbourhood definition will be examined. Statistical travel time correlation can be determined based on the statistical distributions of average link travel times. In this approach, correlation is defined by a statistical distribution, rather than by a single value. Another measure of relationship between travel times of nearby links is cross correlation. This type of correlation is used for time series data where the correlation is computed with a delay component. The use of classification regression trees as an alternate measure to define neighbours can also be investigated. Classification can be based on link characteristics, such as road class, number of lanes, existing transit, etc. Finally, the spatial covariance between links can be used as a criterion to define link neighbours. Another approach for neighbours’ definition would be to identify different classes of neighbours according to their connectivity to the link. First-degree neighbours are those connected directly to the link, second-degree neighbours are connected to the link through an intermediate first neighbour, etc.

Different statistical models, including parametric regression and non-parametric models, were used in this research to relate travel times of neighbour links. A proposed continuation would be to use Bayesian statistical models, where the first moments of the covariates (i.e. travel time mean and/or variance) are considered random variables in the model formulation. Another modelling technique that could be used to relate travel times of neighbour links is geostatistical modelling (e.g. Kriging models).
An important extension of this thesis is to develop other data fusion schemes to combine the data from different sources (e.g. probe vehicles, buses, etc). An Empirical Bayes method was used in this thesis to estimate travel times using neighbour links real-time data and historical data. The use of other methods, such as Kalman Filters can be explored.

8.3.2 Large Scale Deployment

The deployment of a large scale Vehicles As Probes (VAP) system in Metro Vancouver requires the identification of several important issues. The market first has to be surveyed to specify potential fleets that can be employed as probe vehicles. In Metro Vancouver, candidate fleets include taxis and Translink buses. According to the author’s recent discussions with Translink, cellular phone data will also be continuously collected and used to monitor network traffic conditions. The market penetration level of each probe type (i.e. taxis, buses, and cellular phones) needs to be obtained to determine whether a sufficient sample of probes would be available.

One of the main components of a large scale VAP deployment is the development of several algorithms. A map-matching algorithm needs to be developed to facilitate the association of each tracked probe to the correct link of travel in real-time on a GIS map. An additional algorithm is required to extract link travel time information from the raw position and time stamp data. The developed algorithms may be tested, validated and improved, if needed.

The choice of a data acquisition resolution (i.e. polling rate) should consider the tradeoffs between cost and accuracy. The data aggregation interval depends on the data acquisition resolution, dynamic changes of the traffic conditions, and the type of sensors used to collect data. The optimum data acquisition resolution and the measurement interval are two important parameters that must be specified during a large scale deployment of the VAP system.

Another extension of the current research includes the integration of other data sources to improve the travel time estimation/prediction accuracy. These sources include loop detectors, traffic signal controllers, video cameras, etc. Upon exploring all available
sources of traffic data, the feasibility of developing a data fusion scheme to merge data from different sources will be investigated.

As this research is application driven, a large scale Advanced Traveller Information System (ATIS) system can be deployed. This will incorporate the integration of several efforts of many stakeholders. For example, road-side Variable Message Signs (VMS) can be installed to inform commuters of potential delays, location of construction zones, or road incidents. Similar signage can be posted at bus stops to notify riders about predicated bus arrival times. Media companies can incorporate traffic notices and updates into the hourly news. Cellular phone operators can provide a service to their customers which would enable them to acquire traffic information by sending a message to a designated number. Internet users can benefit from a trip planning application that selects the best route for those drivers commuting during rush hours. An example of the output of the internet based application used to disseminate traffic information to road users is presented in Figure 8.1.

Figure 8.1 Regional Congestion Indices in Metro Vancouver
This Figure is extracted from the travel time survey report, carried out by the Greater Vancouver Transport Authority (TransLink) in Metro Vancouver, formerly the Greater Vancouver Regional District (GVRD), in 2003. It shows regional congestion indices values projected on a Metro Vancouver map. The regional congestion index is the measure of congestion calculated as the ratio between survey speed and posted speed. Finally, an incident detection and management system can be developed using traffic data collected from different sources.
REFERENCES


APPENDIX (I): PUBLICATIONS

