Freeway Travel Time Prediction Using Data from Mobile Probes

by

Pedram Izadpanah

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2010

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

It is widely agreed that estimates of freeway segment travel times are more highly valued by motorists than other forms of traveller information. The provision of real-time estimates of travel times is becoming relatively common in many of the large urban centres in the US and overseas. Presently, most traveler information systems are operating based on estimated travel time rather than predicted travel time. However, traveler information systems are most beneficial when they are built upon predicted traffic information (*e.g.* predicted travel time). A number of researchers have proposed different models to predict travel time. One of these techniques is based on traffic flow theory and the concept of shockwaves. Most of the past efforts at identifying shockwaves have been focused on performing shockwave analysis based on fixed sensors such as loop detectors which are commonly used in many jurisdictions. However, latest advances in wireless communications have provided an opportunity to obtain vehicle trajectory data that potentially could be used to derive traffic conditions over a wide spatial area. This research proposes a new methodology to detect and analyze shockwaves based on vehicle trajectory data and will use this information to predict travel time for freeway sections.

The main idea behind this methodology is that average speed on a section of roadway is constant unless a shockwave is created due to change in flow rate or density of traffic. In the proposed methodology first the road section is discretized into a number of smaller road segments and the average speed of each segment is calculated based on the available information obtained from probe vehicles during the current time interval. If a new shockwave is detected, the average speed of the road segment is adjusted to account for the change in the traffic conditions. In order to detect shockwaves, first, a two phase piecewise linear regression is used to find the points at which a vehicle has changed its speed. Then, the points that correspond to the intersection of shockwaves and trajectories of probe vehicles are identified using a data filtering procedure and a linear clustering algorithm is employed to group different shockwaves. Finally, a linear regression model is applied to find propagation speed and spatial and temporal extent of each shockwave. The performance of this methodology was tested using one simulated signalized intersection, trajectories obtained from video processing of a section of freeway in California, and trajectories obtained from two freeway sections in Ontario. The results of this thesis show that the proposed methodology is able to detect shockwaves and predict travel time even with a small sample of vehicles. These results show that

traffic data acquisition systems which are based on anonymously tracking of vehicles are a viable substitution to the tradition traffic data collection systems especially in relatively rural areas.

Acknowledgements

I was honoured and privileged to work with my supervisors Professor Bruce Hellinga and Professor Liping Fu. I would like to express my sincere gratitude to them for their guidance, encouragement, motivation, mentorship, and financial support.

I appreciate the examining committee, Dr. Amer Shalaby, Dr. Jeff Casello, Dr. Carl Hass, and Dr. Fakhri Karray for their time and guidance towards improving this research.

Further, I am grateful to my friends and colleagues Soroush Salek Moghaddam, Reza Noroozi, Hossein Zarei, Akram Nour, Morteza Bagheri, Amir Hosein Ghods, Peter Park, Flavio Cunto, Behzad Hashemloo, and Samira Farahani for their friendship, advice, and time during the past 5 years.

I would like to thank my parents, Betty and Hamid Izadpanah, for their unconditional love, motivation, support, and guidance. I am indebted to them forever. Thank you mom and dad! I love you.

I want to thank my father-in-law and my mother-in-law, Mozaffar Hemmati and Nassrin Naghibi for their support and understanding.

My deepest gratitude goes to my sister, Parnam, my brother, Payam, and my sister-in-law Sahar for their support and friendship.

Finally and most importantly, I would like to thank my wife, Nasim, for her love, encouragement, and understanding during my PhD study. My PhD could neither start nor finish without her assistance and support.

Dedication

This thesis is dedicated to courageous people in the world who lost their lives or imprisoned for promotion of freedom and democracy.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Traffic congestion is a common problem in major cities around the world. Unfortunately, traffic congestion statistics shows an exacerbating trend over the past decades. Vehicles emissions contribute 15% of total green house gas emissions in North America, which is a significant factor in global warming (Oliver, 2005). According to a recent report (Schrank *et. al.*, 2005) Americans wasted a total of 3.7 billion hours and 2.3 billion gallons of fuel in traffic congestion annually. The report states that "…the current pace of transportation improvement, however, is not sufficient to keep pace with even a slow growth in travel demands in most major urban areas".

There are a number of techniques available to alleviate traffic congestion. Enhancing network capacity is one of the most common techniques employed to decrease the total delay of the network in the past. However, construction of new highways and expansion of existing highways have become cost prohibitive, especially when they are located at urban centres. Consequently, recent efforts have been focused on efficient utilization of existing capacity and effective travel demand management as a way of tackling traffic congestion. One of the most significant challenges to using the existing transportation infrastructure more effectively is the lack of accurate, up-to-date traffic conditions data for the entire road network. If wide area, up-to-date traffic conditions data would benefit two groups of users namely, travelers and operators of transportation networks. Travelers can use the information for pre-trip and enroute planning purposes which both would result in lowering the congestion level in the network. On the other hand, operators of the network would likely respond to incidents more quickly and effectively. They can build up a rich historical traffic information database for the entire network which has tremendous value to many applications in traffic engineering as well as other areas.

In large urban areas dedicated fixed sensors such as CCTV cameras and loop detectors have been used for many years as the main techniques to collect near real time traffic data. However, extensive deployment of these sensors is cost prohibitive and consequently they are typically deployed on only heavily travelled freeways within large urban centres. For example, Highway 401 is the only freeway that traverses the province of Ontario - running 820 km from Winsor in the West to Cornwall in the East - and is a vital transportation corridor for the movement of goods and people. Despite its importance, only approximately 59 km (7%) of the highway is instrumented.

New developments within the wireless communication field such as mobile phones provide the opportunity to obtain real time traffic information over a wide spatial area without the deployment of dedicated traffic sensors. Obtaining traffic information through mobile phones has become possible as a result of the development of the ability to determine the location of a mobile phone. The locationing ability was originally developed in response to the CRTC (Canadian Radio and Telecommunications Commission) and FCC (Federal Communications Commission) requirements that wireless carriers must provide an estimate of the location of a mobile phone in the event of an emergency call to 911. Currently, the accuracy of the estimated locations in mobile phone based systems is sufficient to support the estimation of speeds and travel times on roadway segments. A limited number of commercial systems have emerged in the market and several evaluation studies in North America are currently underway or have been recently completed (Izadpanah and Hellinga, 2007).

In mobile phone based systems, the positions of a sample of mobile phones within a specified geographical area are tracked anonymously over time. This process is called location referencing. The location referencing process is usually carried out by the wireless carrier and the resulting data consists of a randomly assigned probe vehicle identification number, time stamp, and position. The data are then transmitted to a processing centre for deriving information on traffic conditions such as link travel times and speed, incidents, and queues.

In the past few years wireless communication technologies have been evolving very quickly. Most new mobile phones and smart phones have embedded GPS and are capable to access the Internet. The positions obtained from GPS is generally more accurate than positions obtained from cell phone system except in areas where sight of satellites is blocked by high-rise towers (*e.g.* downtown Toronto). Coincidently, the cell phones based positioning systems are relatively more accurate in these areas due to small size of each cell. Therefore, the two techniques can complement each other to provide continuous accurate positions of vehicles over time.

Availability of accurate and frequent positions of mobile probes over time for a large spatial region provides new opportunities to acquire traffic information that is not readily available through traditional traffic surveillance systems. One of these opportunities is travel time prediction. Many Intelligent Transportation Systems (ITS) applications such as in-vehicle route guidance systems, and advanced traffic management systems rely on future state of traffic rather than current traffic conditions only (Chien *et. al.*, 2003). Predicted traffic information is an integral part of all travelers' information systems because travelers and managers of transportation networks are much less

interested in past traffic conditions (even if very recently) but are much more interested in the traffic conditions that will be experienced by vehicles that are entering a road segment or within currently a short period of time.

Since the advent of ITS, many researchers have worked on prediction of travel time in transportation networks. Most of them have used dedicated sensors, dedicated probe vehicle data, or fusion of these data with historical data to predict near future travel time. In this research, however, the main goal is to develop a reliable model to predict near future (less than one hour) travel time on freeways using travel time data obtained from mobile phone probes.

1.1 Background

1.1.1 Definition of Travel Time

Figure 1.1 is a space-time diagram illustrating the trajectory of several vehicles as they traverse a hypothetical section of roadway. The roadway section is denoted as section *i* and the individual vehicles are denoted by *j*. The time taken for a vehicle *j* to traverse the roadway section *i* is denoted as t_j.

Since many variables such as traffic volume, weather conditions, incidents, behaviour of the driver, vehicle characteristics, *etc.* affect travel time experienced by an individual traveler, travel time of a road section is a stochastic process which is impossible to estimate or predict exactly. Consequently, all the efforts in the literature focus on estimation or prediction of an expected (average) value of the travel time.

The expected travel time is computed as the average of the travel times associated with all vehicles that enter the section *i* in some time period. So, if travel time of vehicle *j* on section *i* is defined by $\tau_j^i(t)$ where *t* is the time instance that vehicle *j* entered the section *i*, the average travel time over time period Δ is defined as follows:

$$\tau_\Delta^i = \frac{1}{n}\sum_{j=1}^{n} \tau_j^i(t), t \in \Delta \tag{1.1}$$

Where, *n* is the number of vehicles entering road section *i* in time period Δ. For example in Figure 1.1 *n* is equal to 4 (*i.e.* vehicles *j*+1, …, *j*+4).

Note that this definition implies that even if the travel time of all vehicles can be measured (*i.e.* $\tau_j^i(t),\ t \in \Delta, j = 1, \cdots, n$) τ_Δ^i cannot be computed until all *n* vehicles have traversed the section.

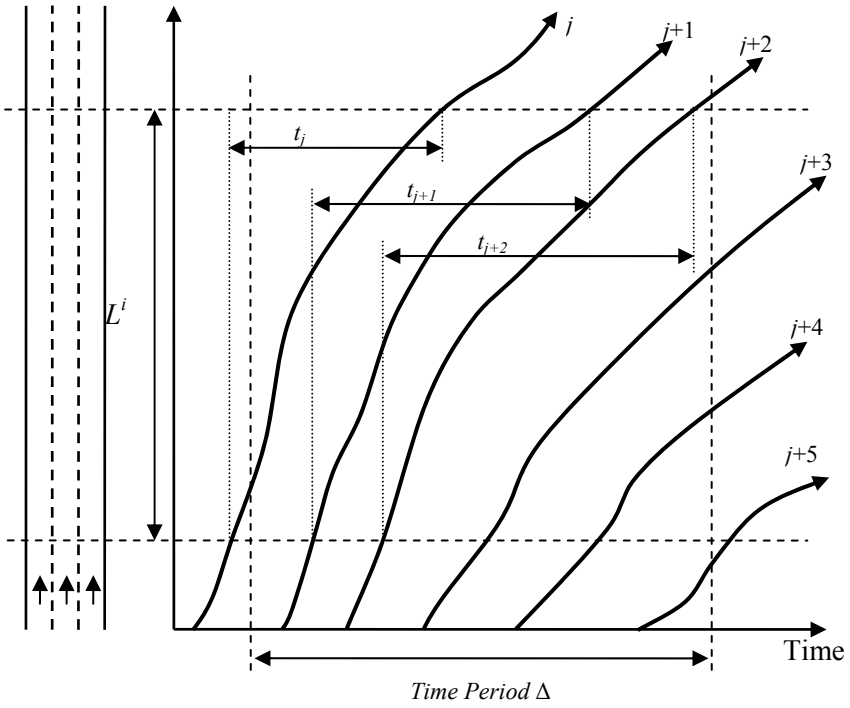


Figure 1.1: Space - time diagram for definition of travel time

1.2 Travel Time Estimation and Prediction

Travel time estimation is the process of calculating (average) travel time for vehicles that have finished traversing a road section and for which traffic conditions are known. In other words, travel time estimation entails calculation of travel time experienced by the vehicles through other traffic characteristics such as speed, flow, and density which can be measured in the field using traffic sensors such as loop detectors. However, the sensors are generally not able to directly measure travel time and consequently, traffic flow theory is usually employed to transform measurable traffic stream characteristics into estimates of travel time.

Conversely, travel time prediction is the process of determining travel time for future traffic conditions. The length of time in the future for which travel times are to be predicted is called the "prediction horizon". The length of the prediction horizon plays an important role in the choice of the approach to tackle this problem.

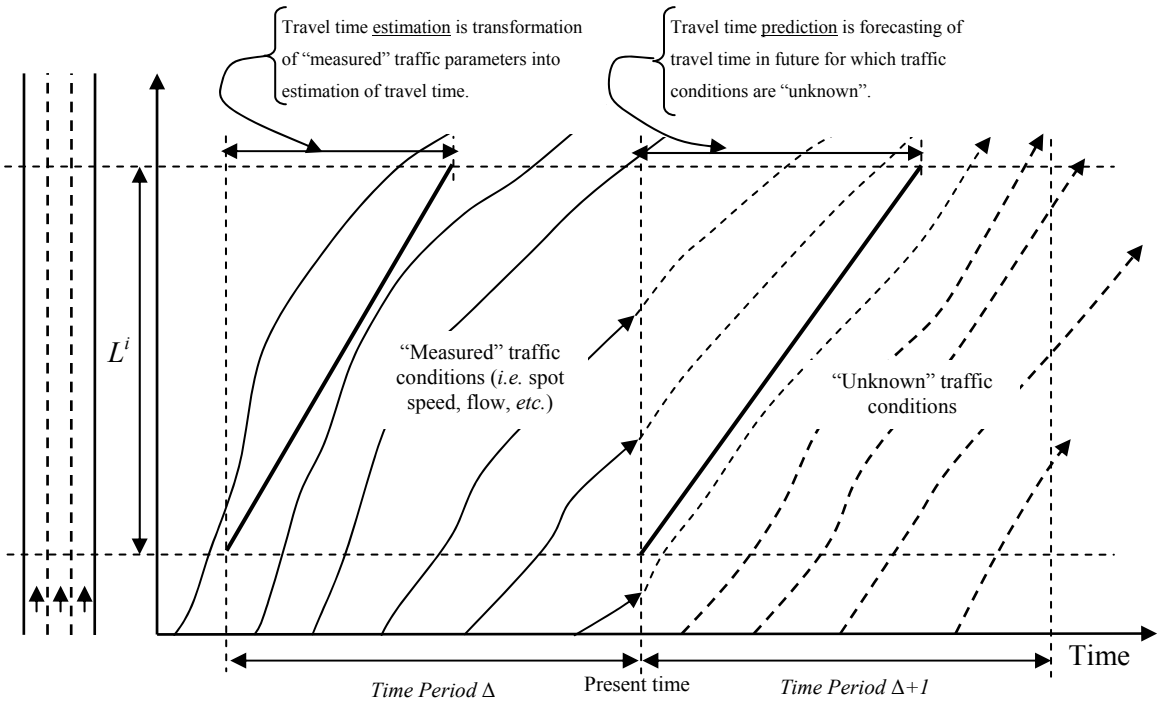


Figure 1.2: Travel time estimation and travel time prediction

Figure 1.2 illustrates the differences between travel time prediction and travel time estimation. Similar to Figure 1.1, the y-axis represents positions of vehicles along the roadway section and the x-axis represents time. Present time in this figure is the end of time period Δ. At present time, there is some information available about traffic in the past time intervals. Consequently, models can be used to estimate average travel time of vehicles that entered the road section in time period Δ. These models transform measured traffic conditions into estimation of travel time. On the other hand, for predicting travel time in the future, traffic conditions are not available. Consequently, travel time prediction models need to make use of previously available data and dynamics of traffic flow to forecast travel time in the future. Travel time prediction models might first predict other traffic metrics and then use a travel time estimation model to transform the predicted traffic stream characteristics into predicted travel times.

Presently, most traveler information systems are operating based on estimated travel time rather than predicted travel time due to lack of reliable prediction models. However, traveler information systems are most beneficial when they are built upon predicted traffic information (*e.g.* predicted travel time). The latest developments in the field of wireless communications such as mobile phones

provide the opportunity for researchers to develop prediction models based on wide area real time data.

For example, Ministry of Transportation Ontario (MTO) has developed a freeway travel time estimation/prediction system that determines travel time based on the average speed obtained from loop detectors. The underlying algorithm in this system is retrospective in nature and can yield accurate prediction results only if the state of traffic remains unchanged. However, in reality, the state of traffic is constantly changing due to changes in demand and capacity. This issue has been widely recognized by the transportation agencies such as MTO and they have initiated a number of efforts to address this deficiency. It should also be noted that the current algorithm used by MTO is applicable only on freeways with full FTMS instrumentation. One of the objectives of this research is to develop a framework which will enable transportation agencies such as MTO to make use of emerging technologies to estimate travel times for freeways which are not covered by FTMS.

1.1.1 Prediction Horizon

One of the critical questions arising in travel time prediction is the length of prediction horizon. For instance, one might ask "how long would it take to go from Toronto to Montreal next Monday?" or "how long would it take to travel from Highway 427 to Don Valley Parkway in Toronto along the Gardiner Expressway if someone enters the Gardiner in the next 5 minutes?" To answer these two questions, different approaches might be adopted by the analyst. For the first question historical data would suffice. However, answering the second question calls for a dynamic model that requires traffic conditions in immediate previous time steps.

To define the prediction horizon, let Δ^* be the current time period and Δ be the time period for which travel time ought to be predicted. Then, the prediction horizon is defined by:

$$T = \max\left[\Delta - \Delta^*, 0\right]. \qquad (1.2)$$

In the literature there are different categories of travel time prediction models based on the length of the prediction horizon (Van Lint, 2004):

- *Real-time Travel Time Prediction*: In these models, average travel time of vehicles entering the road segment at time period $\Delta = \Delta^*$ is desired. In other words, prediction horizon, T, in these models is 0. Real-time travel time prediction should be differentiated from travel time estimation. In travel time estimation, travel time of vehicles that in previous time periods entered the roadway are estimated. However, the goal in real time travel time prediction is to forecast travel time of vehicles that are entering the roadway section at current time interval which the traffic conditions are unknown.
- *Short Term Travel Time Prediction*: This category which is also referred to as near future travel time prediction models is aimed at prediction horizons that are in the order of 0 to 60 minutes (*i.e.* $0 \le T \le 60$ minutes). Obviously real-time travel time prediction models are a subset of short term travel time prediction models. This research focuses on this type of models.
- *Long Term Travel Time Prediction*: The objective of these models is to predict travel time of vehicles that will be entering the road segment in future time period which is longer than 60 minutes (*i.e.* $T > 60$ minutes).

In this research, the focus will be on real-time prediction of travel time.

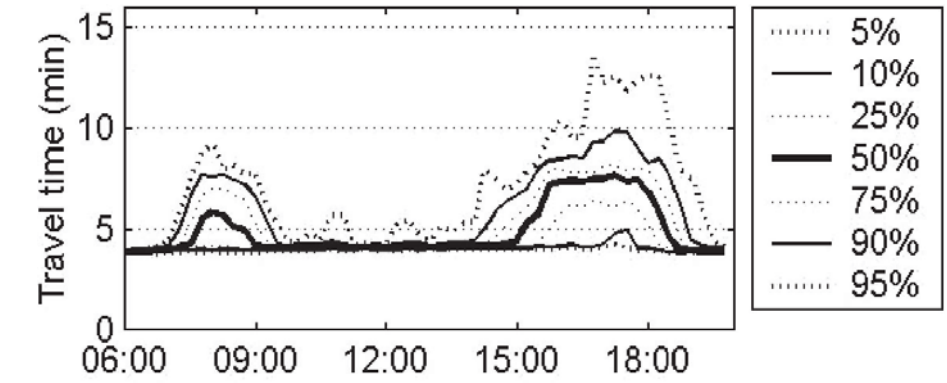
1.2.1 Travel Time Variability

Travel time of a certain section of roadway is subject to temporal variations. The temporal variations of travel time are not only seasonal, monthly, weekly but also daily which makes the prediction and estimation of the travel time difficult. The variability of travel time has been discussed by many researchers such as Meyer and Miller (2001).

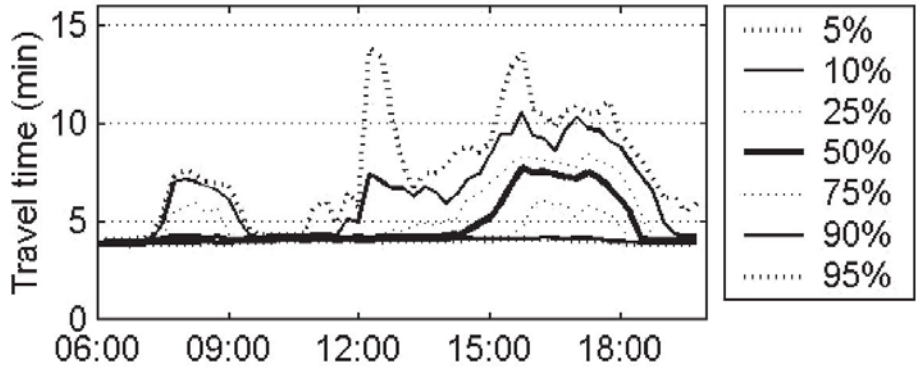
Figure 1.3 shows variability of travel time for different days of weeks in 2002 from 6 a.m. to 8 p.m along eastbound direction on a 6.5 km freeway in the Netherlands (Van Lint and Van Zuylen, 2005). This figure illustrates different percentiles of observed travel time on this section of the freeway. Figure 1.3(a) represents a typical weekday with morning and afternoon peaks. On the other hand, Figure 1.3(b) illustrates variations of travel time on Fridays which the afternoon peak starts earlier and lasts longer than weekdays. On Saturdays, which are shown in Figure 1.3(c), morning and

afternoon peaks are much lighter than weekdays. In fact the morning peak literally doesn't exist on Saturdays. It should be noted that this figure is generated for normal traffic conditions in absence of incidents.

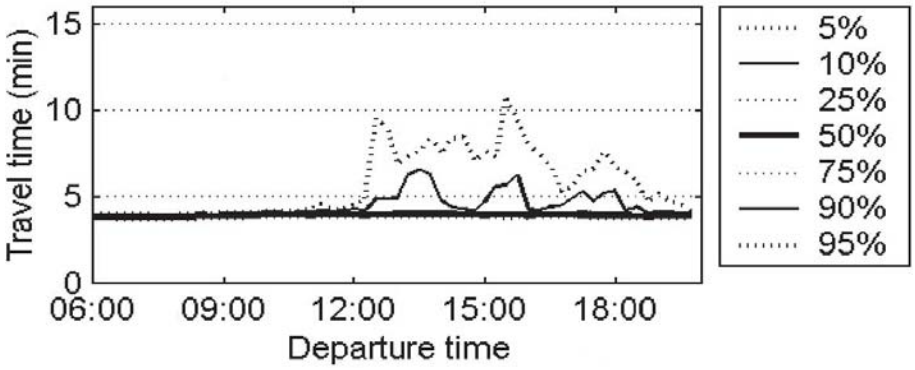
Changes demand and supply (due to weather conditions, incidents, *etc.*) create the temporal variations of travel time which in turn make the prediction of travel time difficult.



a) Thursdays



b) Fridays



c) Saturdays

Figure 1.3: Variability of travel times in 2002 for 15 minute time intervals between 6am and 8pm for different days of weeks (Van Lint and Van Zuylen, 2005)

1.3 Methods of Acquiring Travel Time Data

1.3.1 Fixed Point Technologies

Fixed point traffic data acquisition technologies collect various traffic stream measurements using sensors installed at predefined fixed points. A number of these technologies are used in practice:

1. Inductive Loop Detectors

Loop detectors have been used in traffic engineering as early as 1920's. Inductive loop detectors consist of insulated copper wire installed into the road pavement. When vehicles pass over the loop, the metal within the vehicle disturbs the magnetic field created by the loop and induces an electric current which is measured by the loop detector. The common practice is to either place a single loop detector or two loop detectors (one upstream of the other) in each lane. The single loop detectors are able to determine the number of vehicles passing over them as well as occupancy (a measure of the fraction of time the loop detects a vehicle). Double loop detectors, which are more common, have the ability to determine time mean speed in addition to vehicle counts and occupancy. Normally, loop detectors aggregate the data every 20 seconds and send them to a traffic control center through a data communication system. It should be noted that loop detectors are incapable of measuring travel time directly. However, due to their wide spread used for traffic surveillance, many models have been proposed to estimate travel time from loop detector data.

Loop detectors are relatively accurate and algorithms to acquire traffic information from loop detector data are available and widely used throughout the world. However, loop detectors are costly to install and maintain particularly because installation and maintenance require closure of the lanes. The other disadvantage of loop detectors is that loop detectors become especially unreliable during construction due to lane realignment or damage to the detection and communication hardware. Ironically these are the conditions for which there is a high demand to obtain traffic information.

2. CCTV Cameras

Cameras are mainly installed in transportation facilities to observe the traffic stream and support traffic control decisions such as confirmation and response to incidents. Recently, developments in image processing algorithms have made it possible to use these cameras as a source of travel time measurement. In this case the system operates similar to the traditional method of manual license plate recording. However, wide deployment of cameras is cost prohibitive. Furthermore, often cameras do not perform satisfactorily in poor weather conditions or in low light conditions.

3. Automatic Vehicle Identification

In this technique, vehicles are detected at fixed points along the subject route and vehicle identification and the time when the vehicle was detected are recorded. Consequently, the travel time associated with each detected vehicle between two consecutive fixed points can be computed. There are a number of technologies which can be used to detect vehicles. In most of these technologies a device with a unique ID is installed in vehicles and also readers are installed at certain locations in the network. A number of technologies can be used under this technique: radio frequency identification (RFID) tags, transponders, and Bluetooth devices. Examples of such systems are Highway 407 in Ontario and TrafficWatch system in Oakland California both of which have been deployed for electronic toll collection purposes.

1.3.2 Probe Vehicle Technologies

Probe vehicles are instrumented vehicles in the traffic stream that are able to disseminate position and time data to a traffic centre or can record position and time data which can be downloaded offline. Then, the data can be synthesized to obtain travel time or other traffic measures. Probe vehicles might be dispatched to the traffic stream for the purpose of data collection or they might be vehicles that are already in the network for other purposes but can be used as a data collection tool. Probe vehicles are categorized into dedicated probe vehicles and mobile probe vehicles.

1. Dedicated Probe Vehicles

Instrumented vehicles can be used as dedicated probes to obtain traffic conditions data. Vehicles are typically equipped with a GPS receiver, an electronic map database and a wireless communication link. The vehicles periodically send position data to a central data processing facility.

Dedicated probe vehicles have been used extensively for the purposes of obtaining traffic information; mostly for short term data collection (e.g. travel time or speed and delay studies). Special purpose vehicles such as trucks, taxis, public transit buses, or winter maintenance vehicle fleets have also been used for this purpose.

The most evident disadvantage of using commercial fleets as probe vehicles is that they may produce biased estimates because the probe vehicles may not be a random sample of the population of vehicles. A second disadvantage is the limited number of vehicle probes from which data may be obtained. Naturally this second limitation becomes less problematic when the fleet is sufficiently large.

2. Mobile Probe Vehicles

Mobile probe vehicles are a sample of vehicles in which the vehicles, drivers or passengers carry mobile devices which can be traced. For example, wireless carriers are able to track and locate mobile phones. There are a number of techniques to obtain the location of a mobile phone. A brief description of these techniques and the mobile phone technology can be found in Appendix C. Mobile probes provide an opportunity to derive traffic data at wide area scale without extensive instrumentation and at a low cost. However, the mobile probes are only able to produce positions and time stamps. Other characteristics of the traffic stream (*e.g.* volume, density, etc.) cannot be directly obtained from mobile probes.

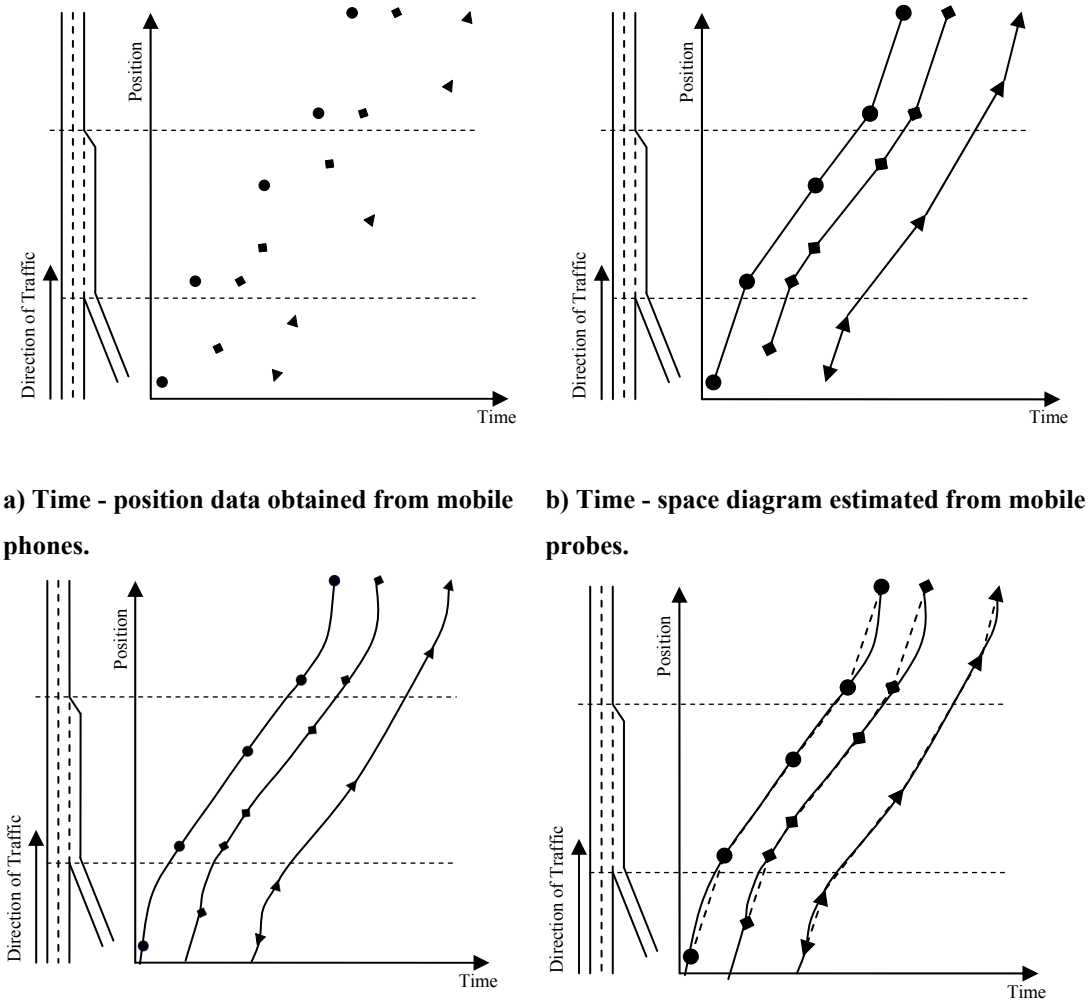
Figure 1.4 illustrates typical data that can be obtained through tracking mobile phones in a network. This figure shows a stretch of a freeway with an on-ramp and the associated time - space diagram for the freeway section. Wireless carriers that cover this section of the freeway are able to provide time stamped positions for every individual mobile phone user along this freeway. Figure 1.4(a) shows a discrete set of points representing the trajectory of different vehicles. Obviously the location data (i.e. estimated positions of mobile phones) are not completely accurate. To correct these data some map matching techniques should be used (Takada, 2006). The piecewise linear graph resulting from connecting all points of an individual mobile probe yields an approximation of the trajectory of the mobile probe (as depicted in Figure 1.4(b). The slope of the line between every pair of locations represents the average speed of the mobile probe between these two locations. Projection of the line segment on the x-axis yields the travel time of the probe between the two locations.

Figure 1.4(c), however, illustrates the true trajectories of these probe vehicles. Based on this figure, two different traffic regimes exist upstream of the on-ramp in the freeway section. Vehicles travel at near free flow speed and then join the slower moving queue upstream of the on ramp. Beyond the on ramp to some point beyond where the acceleration lane ends, the congested traffic regime prevails on the freeway section. Drivers experience the free flow speed at the end of the freeway section.

Figure 1.4(d) compares the real and approximate trajectories of probe vehicles. As can be seen in this figure, the exact position and time that a mobile probe transitions from an uncongested regime to a congested regime is not readily available from the approximate time-space diagram. Consequently, there are two main issues associated with the approximate trajectories regardless of measurement errors in locations obtained from mobile phones. First, data for a sample of vehicles traversing the

network is available rather than data for all vehicles. Second, the approximate trajectories of probe vehicles are available rather than the real trajectories.

The other problem associated with mobile phones is that no one can be certain that the mobile phone is in a moving vehicle. There are some algorithms in the literature that address this issue (Takada, 2006). In this research, it is assumed that all mobile phones are in moving vehicles.



a) Time - position data obtained from mobile phones.

b) Time - space diagram estimated from mobile probes.

c) Real time - space diagram.

d) Comparison of estimated and real time - space diagrams.

Figure 1.4: Time - space diagram depicting mobile probes on a freeway section with an on ramp

1.4 Problem Statement

As discussed previously, the most useful form of travel time information is predicted travel time rather than estimated travel time. The estimated travel time is useful for transportation planning and management purposes, but is not accurate enough for pre-trip or en-route planning purposes. Hoogendoorn (1997) shows that if current traffic information is used in future instead of predicted travel time, it may lead to oscillatory behaviour causing deterioration of traffic conditions instead of improvement.

Travel time prediction is intrinsically a difficult problem due to the fact that many factors influence travel time experienced by drivers. For example travel time in a particular road section is affected by drivers' behaviour, traffic demand, weather conditions, and incidents. Many of these factors vary over time and are difficult to predict. Furthermore, travel time itself is not easily measured by "traditional" data sources such as loop detectors and some models based on traffic flow theory are required to estimate travel time. To resolve this deficiency, some researchers have used probe vehicles to develop travel time prediction models using data from probe vehicles as input. However, wide area deployment of dedicated probe systems is not economically feasible and also such systems cannot guarantee dissemination of real time data.

Over the last three decades, a variety of models has been developed for travel time prediction. Some researchers have used time series analyses in the form of ARMA and ARIMA models (Yang, 2005b; D'Angelo *et*. *al*. 1999; and Ishak *et*. *al*., 2002). However, the traditional time series models are unable to capture rapid fluctuations in traffic stream (Head, 1995 and Abdulhai, *et.al.*, 1999). Kalman filter has been used to tackle travel time prediction problem (Chen, *et. al*., 2005; Chu, *et. al*., 2005; Chen and Chien, 2001; and Yang, 2005a). Some of these researchers have used probe vehicle data and the rest loop detector data to predict travel time. Researchers have found satisfactory results through applying Kalman filter especially due to the fact that Kalman filter can grasp quick fluctuations in travel time well. However, Kalman filter requires the dynamics of the process stated clearly which is a challenging task itself in the case of traffic flow due to many uncertainties. Some researchers have resorted to data-driven approaches such as Neural Networks due to complexity of the process in addition to imprecise and incomplete information about the process (Van Lint *et. al.*, 2005 and 2002; Rilette and Park, 1999). Furthermore, other techniques such as regression models have also been used by researchers (You and Kim, 2000, and Wu *et. al.*, 2004).

Advances in wireless communications have provided the opportunity of collecting wide area near real time traffic data. However, the data need to be interpreted and transformed into "classic" measures such as predicted travel time to be useful for travelers and transportation managers. There are three main challenges associated with the process of converting raw mobile phone data to traffic conditions: (1) Location data obtained from mobile phone is subject to measurement errors. The distribution and extent of the errors depend on the geographical area, the handset, and the localization technique being used. (2) Location referencing is not done very frequently and polling intervals between two consecutive locations affect accuracy of traffic information especially when two different regimes of traffic exist between these two consecutive locations. (3) Only a sample of vehicles contain an active mobile phone and only a sample of these phones is used to obtain traffic conditions in a network to minimize the cost for travel time prediction. The sampling process causes another type of error in the process of obtaining traffic conditions.

A number of researchers have endeavoured to obtain traffic conditions through mobile phones (Cayford *et. al.*, 2006; Smith, 2006; Takada, 2006; Fountain *et. al.*, 2004; Yim, 2003; Ygnace, 2001). However, the main focus of these researchers has been estimation of travel time or speed not travel time prediction. The available commercial packages that make use of mobile phone data only provide an estimation of travel time and/or speed despite the fact that mobile phone systems are potentially able to support many decision support tools in traffic engineering (Izadpanah and Hellinga, 2007).

In spite of the important efforts in the development of travel time prediction models and the advances in wireless traffic monitoring systems, some important issues remain to be addressed in order to develop a reliable model to predict travel time using data from mobile phones:

- There is no model available to make use of mobile phone data to predict travel time.
- There is no general consensus in the literature regarding superiority of a specific type of model for prediction of travel time (regardless of data sources such as loop detectors, dedicated probes, etc.).
- Three main sources of errors deteriorate accuracy of data collected via mobile phones: measurement errors, errors due to infrequent location inquiries, and errors due to sampling. The process of travel time prediction itself creates another source of error because of uncertainties about future. It is not clear how the quality of travel time prediction models is affected by the first three error terms and the size of fourth error term.

This research attempts to address the first two knowledge gaps and the last is left for future research.

1.5 Research Objectives

The main goal of this research is to develop the required tools for *real-time* prediction of travel time on freeways using data primarily obtained through anonymously tracking of vehicles. To achieve this goal the following objective should be met:

- Develop an algorithm to predict travel time of a freeway section using data obtained from vehicle trajectory data;
- Evaluate the proposed algorithm against the trajectory data from real freeways.
- Compare the performance of the proposed algorithm with traditional sources of travel time information.

1.6 Thesis Outline

The rest of the document is organized as follows: **Chapter 2** provides a synopsis of the state of the art to estimate travel time. This chapter categorizes travel time estimation models based on the logic behind the models. **Chapter 3** summarizes the efforts of researchers to predict travel time. Travel time prediction models are categorized into two broad groups: data driven models and traffic flow models. Furthermore, each group itself divided into different modeling strategies.

Chapter 4 describes the proposed methodology to predict travel time using trajectories of a sample of vehicles. Also, this chapter outlines two methodologies to obtain travel time of a route using data obtained from loop detectors. The loop detector data will be used as a bench mark to evaluate the performance of the proposed methodology. **Chapter 5** of this document describes the datasets which were used to show the performance of the proposed methodology. Also, this chapter describes the data collection efforts which were conducted in this research to compare travel time obtained from different sources of data on two freeway sections in Ontario. **Chapter 6** shows the results obtained from the application of the proposed travel time prediction model to different freeway sections. **Chapter 7** summarizes the conclusions and contributions of this research as well as provides directions for further research in this area.

Chapter 2
Travel Time Estimation

This chapter reviews current state of the art in travel time estimation. Although the main goal of this research is the prediction of travel time, travel time estimation models are often used as a part of prediction process in many travel time prediction models. Furthermore, the study of travel time estimation models provides insight into fundamentals of travel time variations over time that are essential for prediction of travel time.

Travel time estimation is the process of calculating travel times that have been experienced by road users over a road section. In other words, these models make use of traffic flow theory to determine travel time based on other known characteristics of traffic flow such as flow, density, speed, etc.

Travel time estimation models can be divided into four main categories namely, spot speed algorithms, stochastic queuing algorithms, section density algorithms, and multi-regime algorithms.

2.1 Spot Speed Algorithms

Spot-speed algorithms are a family of travel time estimation algorithms that rely on the use of speed measurements obtained at spot location (e.g. from loop detectors).

2.1.1 Average Speed Algorithm

The simplest form of loop detector based travel time estimation algorithms makes use of measured average vehicle speeds at a spot location and assumes vehicles travel at this speed over a fixed segment of the roadway (Lindveld *et al*., 2000). The length of this fixed segment of the roadway is often assumed to be half the distance between two consecutive detector stations, but can be defined as some other length. This scheme can be illustrated using the notation in Figure 2.1 and Figure 2.2.

.

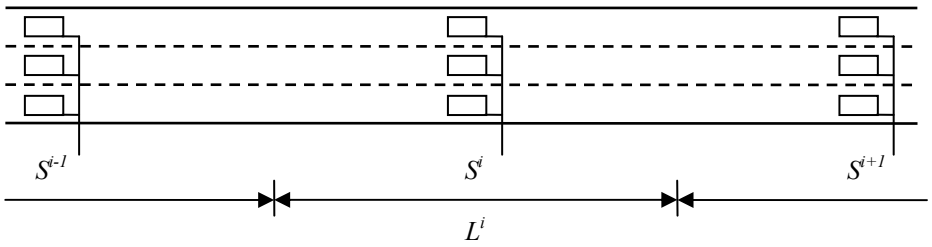


Figure 2.1: Roadway schematic for spot-speed algorithms

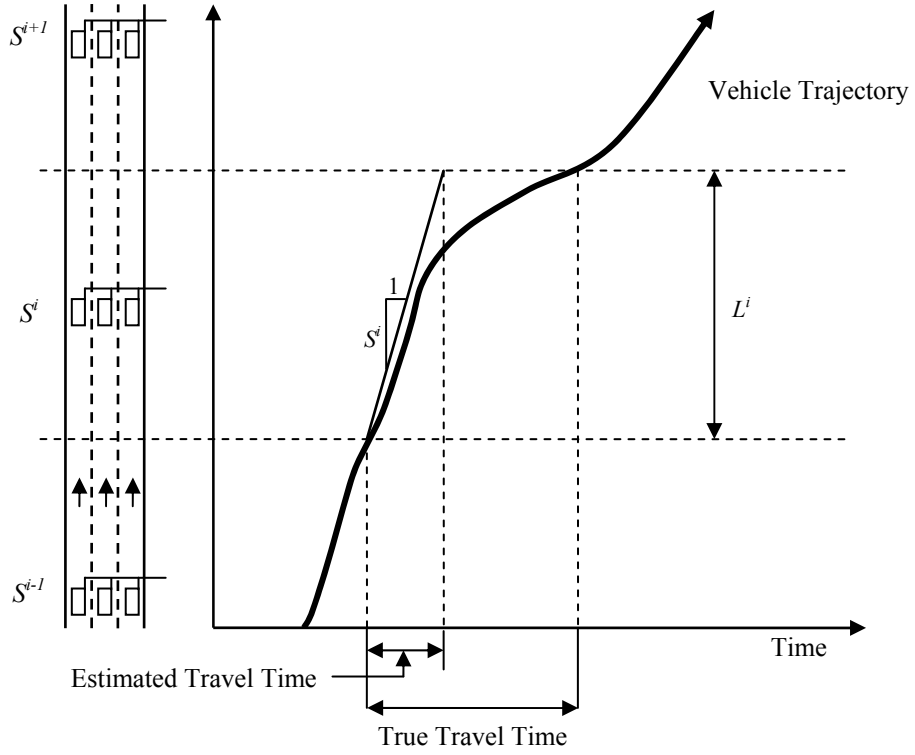


Figure 2.2: Space - time diagram

As illustrated in Figure 2.2, the average speed S^i represents the speeds of vehicles as they pass a specific location (i.e. the loop detector). Spot speed based methods assume that the vehicles travel at this speed over the entire section length (i.e. L^i). This assumption is generally valid when the road section is uncongested. However, as illustrated in Figure 2.2, when a portion of the section experiences congestion, then this assumption is no longer valid as the vehicle's speed changes significantly over the length of section L^i and consequently the estimated travel time will not be

accurate. If the section length (L^i) is short, then errors between estimated and actual travel time will be acceptably small. However, installing loop detectors more closely spaced than approximately 500 – 600m is generally not economically feasible.

It is important to note that, in general, point-speed algorithms tend to underestimate travel times when congestion is forming. The magnitude of the underestimation is a function of the severity of the congestion, the length of the road section (L^i), the portion of the segment congested, and the rate at which the queue is growing or dissipating.

2.1.2 Vehicle Trajectory Algorithm

The Average Speed Algorithm estimates a trajectory on the basis of the average of all the individual vehicle speeds measured at a point location. Coifman (2002) has proposed a method that develops a trajectory on the basis of speeds and arrival times of individual vehicles and assumptions about shockwave speeds.

The use of shockwaves (or more formally, the kinematics theory of traffic flow first proposed by Lighthill and Whitham, 1955 and Richards, 1956) within the method proposed by Coifman is illustrated in Figure 2.3. The steady state flow-density relationship is approximated by a triangular relationship. Shockwave speeds are computed as the slope of the flow-density curve, and therefore the speed is U_f and U_c for shockwaves in the uncongested and congested regimes respectively.

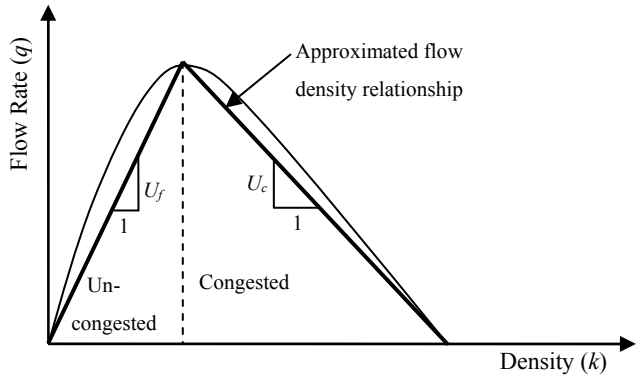


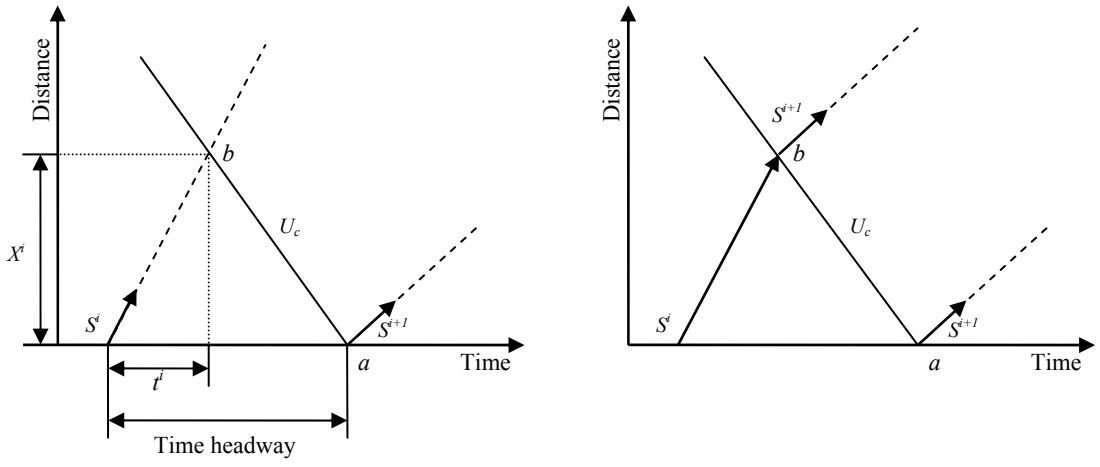
Figure 2.3: Shockwave speeds

Coifman assumes the use of advanced loop detectors that are able to measure and report the speeds and arrival times of individual vehicles. Consequently, the information assumed available about the road section is depicted in Figure 2.4(a). The bold arrows (S_i and S_{i+1}) represent the speeds of

individual vehicles as measured by a loop detector. The dashed lines represent the projected trajectory of these vehicles as if they were to continue travelling at the measured speed. Coifman refers to these projected trajectories as "chords". The time headway between the vehicles i and i+1 is just the time elapsed between the measurement of S_i and S_{i+1} by the loop detector.

If it is assumed that vehicle i is travelling in one flow state and vehicle i+1 is travelling in a different flow state, then the boundary between these two flow states (i.e. the shockwave) propagates at a speed equal to U_c (or U_f) depending on prevailed density. Consequently, it is possible to determine the length of time (t^i) and length of roadway (X^i) that vehicle i travels at speed S^i (corresponding to flow state i) before transitioning to speed S^{i+1} (corresponding to flow state i+1) by positioning a vector with speed U_c to intersect with the chord of vehicle i+1 at the time it is measured to pass the loop detector (point a). The intersection of this shockwave vector and the chord from vehicle i provides point b.

A trajectory is built by connecting the truncated chords from individual vehicles as illustrated in Figure 2.4(b). The chord from vehicle i+1 is connected to the truncated chord from vehicle i.



a) Data obtained from loop

b) Trajectory development procedure.

Figure 2.4: Coifman's trajectory estimation method

This method of incrementally summing the truncated chords is continued until the trajectory reaches the end of the link (Figure 2.5). The arrows that appear along the time axis are vectors representing individual vehicle speeds measured by the upstream loop detector (i.e. S^i, S^{i+1}, …). The dotted lines represent the shockwave (propagation) vector.

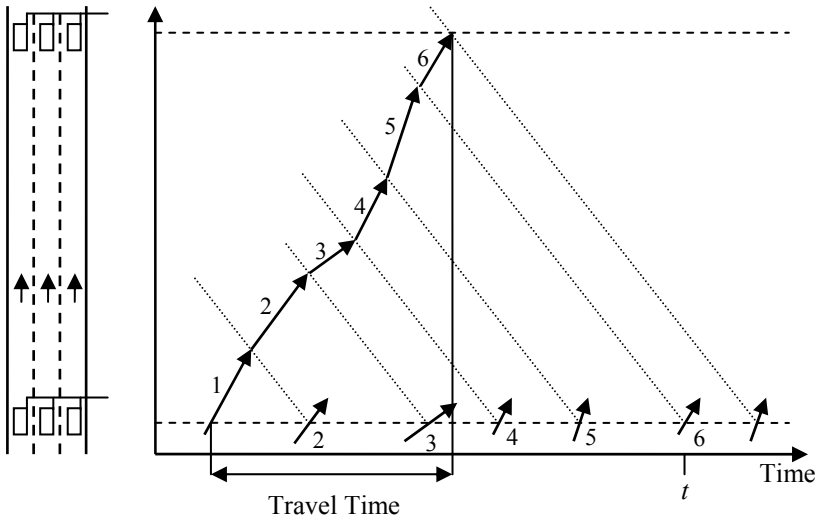


Figure 2.5: Coifman's trajectory estimation method

Coifman applied the proposed method to data from I-880 in California. Actual travel times were obtained from dedicated probe vehicles. The performance of the proposed method is relatively good as long as the road segment is not partially covered by a queue (because then the shockwave speed is neither U_f nor U_c). Consequently, the method is not reliable when recurrent or non-recurrent queues are growing or dissipating – limiting its practical application. Furthermore, Coifman's method makes use of individual vehicle information. Such data may not be available through commonly used traffic management systems in large urban areas such as COMPASS in Ontario.

2.1.3 Iterative Travel Time Algorithm

Cortés *et. al.* (2001) have proposed a method of estimating travel times that combines aspects of both the simple Average Speed Algorithm and Vehicle Trajectory Algorithm. The concept of the method is explained with respect to Figure 2.6.

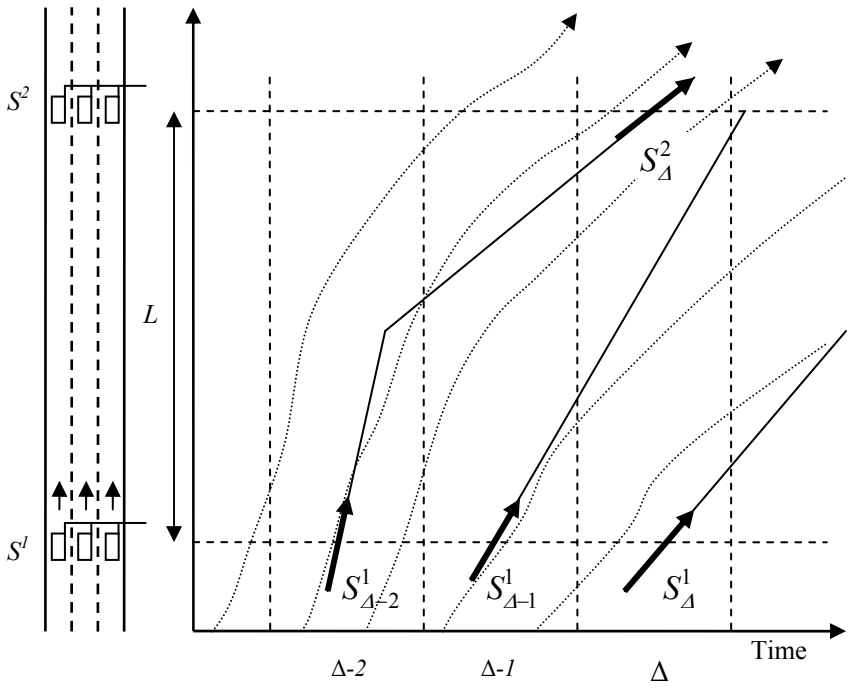


Figure 2.6: Iterative Travel Time Algorithm space-time diagram

Figure 2.6 illustrates a space-time diagram for a typical section of freeway. The section is bounded at the upstream and downstream ends by loop detector stations (S^1 and S^2 respectively) and has length L. As vehicles travel along the section, their speeds are measured by the loop detectors. Most loop detectors aggregate data over a defined polling period, typically 20 or 30 seconds. Consequently, the speeds of individual vehicles are not known – rather the average speed of vehicles passing each loop station is provided for each polling period (depicted in Figure 2.6 as the bold arrows on the space-time diagram). In Figure 2.6, the trajectory of 5 individual vehicles is depicted. The time scale spans 3 polling intervals (Δ, Δ-1, and Δ-2). The average speeds provided by the loop detectors are given by S^1_{Δ}, $S^1_{\Delta-1}$, $S^1_{\Delta-2}$, S^2_{Δ}, $S^2_{\Delta-1}$, and $S^2_{\Delta-2}$. The method proposed by Cortés *et. al.* is based on the assumption that the travel time experienced by vehicles on the section at time interval Δ is a linear combination of the speeds measured by loop stations 1 and 2 as per the following equation:

$$\tau_{\Delta}^{1,2} = \frac{L}{\alpha S^1_{\Delta-\tau_{\Delta}^{1,2}} + (1-\alpha) S^2_{\Delta}} \tag{2.1}$$

Where, the term alpha (α) is a weighting factor that must be calibrated; $\tau_{\Delta}^{1,2}$ is the travel time from loop station 1 to 2 during polling interval Δ; $S^{1}_{\Delta-\tau_{\Delta}^{1,2}}$ is the average speed reported by loop station 1 during interval $\Delta-\tau_{\Delta}^{1,2}$; and S_{Δ}^{2} is the average speed reported by loop station 2 during polling interval Δ. It is important to note that the unknown travel time appears on both sides of the equation meaning that it can only be solved iteratively.

The implication of solving this equation is illustrated graphically in Figure 2.6. The projection of the arrows depicting the average speeds during each polling interval indicates the trajectory of a vehicle travelling at this average speed. The iterative solution technique essentially searches for a solution that results in the two speed projection lines intersecting within section *L* and within the time period ($\Delta-\tau_{\Delta}^{1,2}$) to Δ.

The Average Speed Algorithm assumes that vehicles travel at a speed equal to the speed measured at loop station *i* from a distance halfway between the station *i* and station *i*-1 to a distance half way between station *i* and station *i*+1 (as in Figure 2.2). This is equivalent to computing the travel time from one detector station to the next as the distance between the two stations ($L^{i,i+1}$) divided by a weighted average of the speeds measured at the upstream and downstream stations in which the weighting is 0.5.

$$\tau^{i,i+1} = L^{i,i+1}/(0.5S^{i} + 0.5S^{i+1}) \qquad (2.2)$$

The Iterative Travel Time Algorithm is very similar to that used in the Average Speed Algorithm, when α = 0.5 and the term $S^{1}_{\Delta-\tau_{\Delta}^{1,2}}$ is replaced with S^{1}_{Δ}.

Cortés *et. al*. (2001) evaluated the performance of their proposed algorithm using only simulated data. They report a mean absolute percent error in the estimate of link travel times in the range of 7% and suggest that the corresponding mean absolute percentage error for the Average Speed Algorithm is in the range of 20 – 25%.

2.1.4 Piecewise Linear Trajectory Algorithm

Van Lint and Van der Zijpp (2003) proposed an algorithm to estimate travel time along freeway sections. The main idea of this algorithm is very similar to the Iterative Travel Time Algorithm and Average Speed Algorithm introduced before. However, the difference lies in the way speed is

estimated in the road section between the upstream and downstream loop detectors. In this algorithm, speed is estimated by a convex combination of the speeds measured by upstream and downstream loop detectors.

The concept of this algorithm is explained with respect to Figure 2.7. In this figure downstream and upstream loop detectors are located at x^i and x^{i-1} respectively. Average speeds at these two locations during time period Δ are designated by S_Δ^i and S_Δ^{i-1} respectively. The speed of vehicle *j* that enters section *p* of the freeway during time period Δ is estimated by:

$$S^j(t) = S_\Delta^{i-1} + \frac{x^j(t) - x^{i-1}}{x^i - x^{i-1}}\left(S_\Delta^i - S_\Delta^{i-1}\right) \qquad (2.3)$$

Where, $S^j(t)$ is speed of vehicle *j* at time *t* at location $x^j(t)$. In this equation $t_0 \le t \le t_1$ and $x^{i-1} \le x^j(t) \le x^i$. As can be seen in Figure 2.7, t_0 and t_1 are lower bound and upper bound of time period Δ respectively.

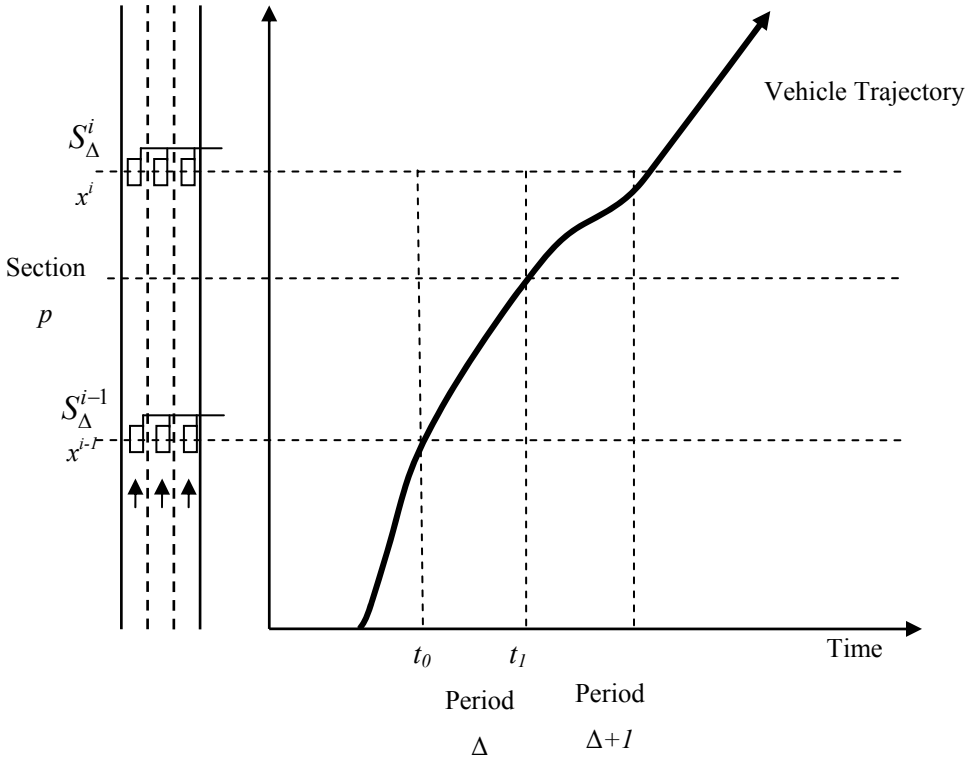


Figure 2.7: Piecewise linear trajectory algorithm - space-time diagram

In equation (2.3), speed of vehicles at any time is estimated by both loop detectors but the weight of closer loop detector is heavier than the further one.

Van Lint and Van der Zijpp (2003) tested the above algorithm against Average Speed Algorithm where it is assumed that speed of vehicles in the first half of the section is equal to speed of upstream loop detector and speed in the second half is equal to speed at downstream loop detector. They used FOSIM simulation model which was developed and calibrated at the Delft University, the Netherlands. They simulated a 7.3 km section of a freeway between Delft and Rotterdam in the Netherlands. Five different simulation runs were performed with different seeds and different traffic demand patterns. The loop detectors are assumed to be spaced 400 to 800 meters apart.

The researchers defined four measures to evaluate performance of the proposed method namely, root mean square error (*RMSE*), bias from mean (*BIAS*), root residual error (*RRE*), and mean relative error (*MRE*):

$$RMSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\hat{\tau}_k - \tau_k\right)^2} \tag{2.4}$$

$$BIAS = \hat{\mu} - \mu \tag{2.5}$$

$$RRE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left[\left(\hat{\tau}_k - \hat{\mu}\right) - \left(\tau_k - \mu\right)\right]^2} \tag{2.6}$$

$$MRE = \frac{100}{K}\sum_{k=1}^{K}\frac{\hat{\tau}_k - \tau_k}{\tau_k} \tag{2.7}$$

Where, $\hat{\mu} = \frac{1}{K}\sum_{k=1}^{K}\hat{\tau}_k$, $\mu = \frac{1}{K}\sum_{k=1}^{K}\tau_k$, $\hat{\tau}_k$ denotes estimated travel time at time step k, τ_k represents reference travel time obtained from simulation model, and K is total number of time steps. *RMSE* shows the overall error of the estimation models. Furthermore, having *BIAS* and *RRE*, *RMSE* can be calculated using $RMSE^2=BIAS^2+RRE^2$. Using the above measure of performance, the authors found that the Piecewise Linear Speed Based (PLSB) trajectory method outperforms the Average Speed Algorithm (ASA). Table 2.1 summarizes the results of the study.

Table 2.1: Comparison of Average Speed Algorithm and Piecewise Linear Speed Based trajectory method (Source: Van Lint and Van der Zijpp, 2003)

	RMSE (s)		*BIAS* (s)		*RRE* (s)		*MRE* (%)	
	ASA	PLSB	ASA	PLSB	ASA	PLSB	ASA	PLSB
Run 1	58.7	29.8	25.8	-0.96	52.7	29.8	5.96	0.96
Run 2	54.9	33.3	27.1	0.43	47.7	33.3	5.83	1.51
Run 3	69.7	33.1	30.4	-3.79	62.7	32.9	6.26	-0.12
Run 4	63.2	32.7	31.7	-4.92	54.7	32.3	6.28	0.32
Run 5	56.5	32.7	26.4	-3.44	50.0	32.5	5.35	1.12
Mean	60.6	32.3	28.3	-2.54	53.6	32.2	5.94	0.76

2.2 Stochastic Queuing Methods

Nam and Drew (1996) have proposed a stochastic queuing approach to estimating freeway travel times. This method uses only vehicle counts from loop detectors and does not require estimates of speed. If it is possible to obtain counts of vehicles entering and exiting the road segment as a function of time, then average travel time can be estimated from the cumulative arrivals and cumulative departures curves. This is illustrated in Figure 2.8.

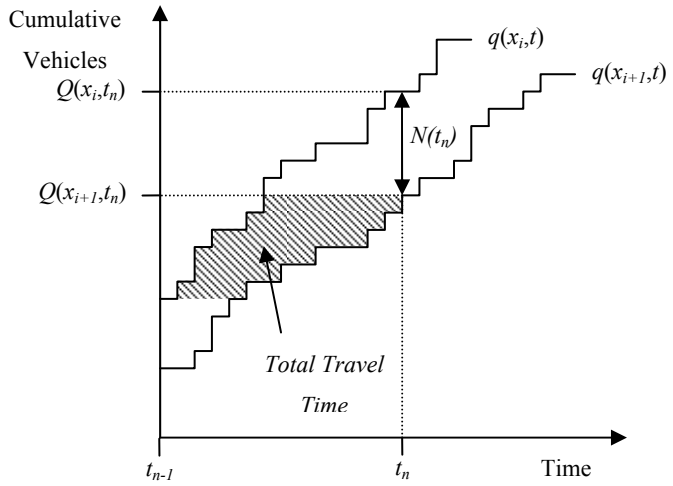


Figure 2.8: Estimation of travel times using cumulative vehicle diagram

In Figure 2.8 $q(x_i,t)$ and $q(x_{i+1},t)$ represent the time series of cumulative number of vehicles measured at the upstream (i.e. flow entering) and downstream (i.e. flow departing) ends of roadway

section i. The number of vehicles estimated to be within the section at time t_n, $N(t_n)$, is the difference between the cumulative number of vehicles entering the section and the cumulative number of vehicle departing the section. The total travel time of all vehicles passing through the section during the period from t_n-1 to t_n is shown as the shaded region. Then the average travel time during that period is calculated as the total travel time divided by the number of vehicles traversing the section.

There are several significant challenges to using this method in practice to obtain accurate travel time estimates.

First, the method requires that the road segment be a "closed" system, meaning that vehicles cannot enter or exit the road segment without being measured. In theory, this is not a problem as traffic management systems in major urban areas in North America (*e.g.* COMPASS system in Ontario) are generally configured to have loop stations on all on and off ramps and at regular spacing on the mainline lanes. However, if a loop station is inoperative due to a hardware failure or lane re-alignment has occurred due to construction/maintenance activities, then this method cannot be used.

Second, vehicle counts from loop detectors contain errors. In a method that relies on cumulative counts, measurement errors accumulate and can create significant errors in the estimates of average travel time.

Third, it is necessary to know how many vehicles are in the road segment at time zero. In theory, we can begin the process when no vehicles have arrived and therefore no vehicles are on the road segment. In practice this is not possible, yet the accuracy of the estimation process is sensitive to the accuracy of this initial number of vehicles.

Nam and Drew demonstrate the application of their method using data from a section of the QEW (Queen Elizabeth Way in Ontario) between Cawthra Road and Dixie Road. For this section, they observed an average error of approximately 3% between the total volume measured at the upstream end of the section and the total volume measured at the downstream section. The authors estimate travel times for this section of the QEW but do not have any independent travel time measurements and are therefore unable to quantify the accuracy of their estimates.

2.3 Section Density Algorithms

Oh *et. al.* (2002) have proposed an approach that is similar in many ways to the queuing method suggested by Nam and Drew as described in the previous section. Using the notation from Figure 2.9, the method by Oh *et. al.* estimates travel time (τ_Δ) on a road section during time period Δ as section

length (L) divided by average travel speed (S). This is essentially the same equation as used in the Average Speed Algorithm. However, unlike the Average Speed Algorithm in which a measured speed is used, the section density algorithm estimates speed using measured traffic volumes and average section density estimated on the basis of cumulative vehicle counts (Equation (2.8)).

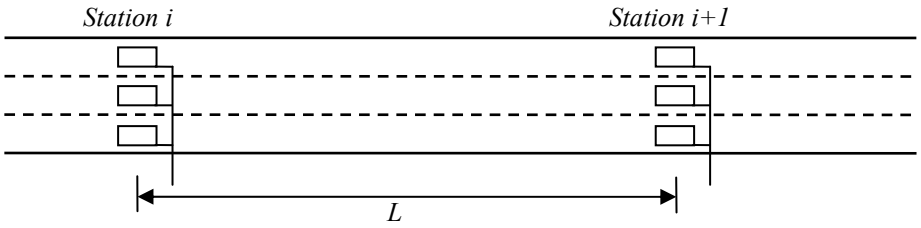


Figure 2.9: Roadway schematic for section density algorithm

$$T_\Delta = L / S_\Delta = L \times D_\Delta / q_\Delta \quad (2.8)$$

Where D_k is the density of traffic on the road segment during time period Δ and q_Δ is the average flow rate on the road segment at time period Δ. The average flow rate q_Δ is computed as the average of the measured flow rates at the upstream (q_Δ^i) and downstream (q_Δ^{i+1}) detector stations (*i.e.* $q_\Delta = \frac{q_\Delta^i + q_\Delta^{i+1}}{2}$).

This method faces the same challenges as the queuing method proposed by Nam and Drew. However, in this case, Oh *et. al*. suggest some means of addressing these challenges. For example, they suggest that the initial density (i.e. initial number of vehicles on the section) can be determined by calculating density on the basis of detector occupancy and an assumed average vehicle length ($D = o \times L/g$; o = detector occupancy; g = average vehicle length) when occupancy is quite low. They also suggest that systematic vehicle count measurement errors by detectors can be accommodated by calculating an average adjustment factor (similar to the 3% error reported by Nam and Drew).

Oh *et. al.* evaluated their proposed algorithm using simulated data (so that the true travel time can be known) and data from dual loop stations on I-880 in California. The results from the evaluation using the simulated data indicate good correspondence between the estimated travel times and the true travel times. However, these results can only be viewed as the best possible performance of the proposed algorithm as the simulated loop data on which the estimates were made, contained no measurement errors. The results from the evaluation using I-880 data do not provide conclusive

results. The proposed algorithm is compared to the Average Speed Algorithm. The results show that the proposed algorithm provides longer travel times than does the Average Speed Algorithm, however, because the true travel times are not known, it is not possible to conclude that these longer travel times are more or less accurate than the estimates from the Average Speed Algorithm.

2.4 Multi-Regime Algorithms

Dhulipala (2002) proposed a multi-regime approach in which separate travel time estimation methods are used for each of three traffic states; namely (a) lane closures, (b) incident conditions, and (c) normal operations. The methods assume volume and occupancy data from loop detectors are available.

Under lane closure conditions an estimate is made of the capacity of the section in the lane closure area (the method assumes a loop detector is present in the lane closure area). If the upstream volume is greater than the estimated capacity, then recurrent congestion is expected, and the algorithm for incident conditions is used. If the upstream volume is less than the estimated capacity then no congestion is expected to result from the lane closure and the algorithm for normal conditions is used. Consequently, the proposed method consists of two algorithms – one for congested conditions and one for un-congested conditions.

The algorithm for congested conditions estimates the length of the queue (congestion) and the number of vehicles in the queue. These estimates are made on the basis of shockwave analysis and cumulative vehicle counts. Travel times are estimated separately for the portion of the link upstream of the queue and the portion covered by the queue. The travel time for the uncongested portion of the link is based on speeds measured by a detector located upstream of the queue. The travel time in the queue is estimated using deterministic queuing theory (Equation(2.9)).

$$t_c = N/q_b \qquad (2.9)$$

where t_c = time spent in the queue; N = number of vehicles in the queue; and q_b = capacity flow of the bottleneck.

For normal conditions, the travel time is estimated by dividing the section length by the measured (or estimated) spot speed. This is essentially the Average Speed Algorithm discussed earlier.

For incident conditions, the proposed method attempts to make use of incident characteristics (e.g. number of lanes blocked, crash versus incident, severity of crash such as property damage only,

injury or fatality, hazardous material involvement, number of cars involved, number of trucks involved, etc.) to estimate incident clearance times. This is accomplished through the use of decision trees calibrated to an incident database from Fairfax County, Virginia (Ozbay, 1996). The result of the calibration is an estimated clearance time for each category of incident (branch in the decision tree). For the estimated duration of the incident, the travel times are estimated as for a lane closure condition.

The proposed algorithms were evaluated using simulated data (CORSIM) for a number of scenarios on a hypothetical 10-km long freeway segment with no on or off ramps. The performance has not been established using field conditions. The estimated number of vehicles in the queue and queue length are highly susceptible to inaccuracies in loop detector volume counts, but the evaluations performed with CORSIM did not consider this source of errors.

Chapter 3
Travel Time Prediction

Travel time prediction has been the focus of research for many years. Travel time prediction models can be categorized into two groups: data driven models and traffic flow models. Data driven models are models that try to find a relationship between given inputs and outputs and use this relationship to predict travel time given a set of inputs. Conversely, traffic flow models utilize traffic flow theory to capture the physical relation between inputs and outputs. This chapter provides a description of these two classes of models.

3.1 Data Driven Models

Data driven models or inductive models consider the traffic processes that generate variations in travel time as a black box. In other words, these techniques try to find a relationship between given inputs (speed, flow, travel time of previous time steps, weather information, *etc*.) and output (travel time) solely based on the data. The relationship is then used to predict travel time in the future given a new set of inputs. Data driven models possess either a statistical basis (such as regression models or time series models) or a machine learning origin such as neural networks, use heuristic approaches to find the relationship.

This research reviews the current state of the art in travel time prediction models. Regression models, time series models, Kalman filter, and neural networks are briefly explained and exemplified by some previous research.

3.1.1 Regression Models

The general purpose of a regression model is to find a closed form function that relates several independent or predictor variables to a dependent or criterion variable. Regression models can be separated into two categories namely, parametric regression models and nonparametric regression models.

The general form of a parametric regression model is shown in Equation (3.1). In this equation $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$ is a vector of parameters to be estimated, and $\mathbf{X}_i = (x_0, \ldots, x_n)$ is the vector of independent variables for the ith of n observations; the errors ε_i are assumed to be normally and independently distributed with mean 0 and constant variance σ^2. The function $f(\cdot)$ relates the expected value of

dependent variable y to the independent variables. In parametric regression models function f has a pre-specified form. Consequently, the challenge is to estimate the vector $\boldsymbol{\beta}$. There are a number of techniques to estimate parameters in a parametric regression models such as Least Square and Weighted Least Square.

$$y_i = f(\boldsymbol{\beta}, \mathbf{X}_i) + \varepsilon_i \tag{3.1}$$

Equation (3.2) shows the general form of nonparametric regression models. In this type of regression model, unlike the parametric regression models, the form of function f is not pre-specified. Most methods of nonparametric regression implicitly assume that f is a smooth continuous function. Again, there are several approaches to estimating nonparametric regression models such as local polynomial regression and smoothing splines.

$$y_i = f(\mathbf{X}_i) + \varepsilon_i \tag{3.2}$$

You and Kim (2000) developed a nonparametric regression model to forecast travel time along a freeway and an arterial in South Korea. They developed a hybrid framework consisting of a GIS application and the forecasting model. In this framework, users are able to have a prediction of travel time for the next 15 minutes between any OD pair along the freeway or the arterial. To predict travel time on the freeway, loop detector data (volume, occupancy rate, and spot speed) were used. The data in this study came from 8 loop detectors that cover 114.3 km which means that on average every loop detector covers 14.3 km of the freeway. This is far greater spacing than is typical for North American freeways which usually have loop detectors spaced approximately every 500-600 meters. However, probe vehicle data (distance traveled, travel time, and stopped time for each link) were used for travel time prediction on the arterial. On the arterial, 6 roadside beacons were installed with an average spacing of 800 meters in the network to receive and transmit probe vehicle data to the operation centre.

To evaluate performance of the framework a simulation model was calibrated using data from loop detectors and probe vehicles. Three performance measures were defined namely, root mean square error (RMSE), mean absolute percent error (MAPE), and correlation coefficient (ρ). Each time interval was chosen to be 5 minutes in this study. The researchers developed different models for the arterial and the freeway. The results of the evaluations for 2 different days at the same locations are shown in Table 3.1. As can be seen in this table the travel time prediction model performs better on the freeway than on the arterials.

Table 3.1: Results of travel time prediction in the arterial and freeway in two different days at the same locations (You and Kim (2000))

		RMSE (Sec)	MAPE (%)	*P*	Avg. Observed Travel Time (Sec)
Arterial	Day 1	15.47	8.02	0.47	148.50
	Day 2	19.63	9.88	0.66	161.67
Freeway	Day 1	14.27	2.22	0.83	404.48
	Day 2	8.46	1.67	0.90	390.52

Rice and Van Zwet (2004) developed a linear regression model to forecast travel time on California freeways. This research was part of an effort to provide Californian commuters with up to date real time data. The algorithm was designed to be part of an Internet based system through which commuters would input their origin and destination and obtain the shortest path based on predictive travel time. To respond to any query in a reasonable time, the prediction algorithm should be fast enough to process the huge dataset available for California.

The researchers state that they can compute two proxies for travel time at time instant Δ on day *d*: one of them is $\tilde{\tau}_{\Delta}^{d}$ which is the travel time computed using Average Speed Algorithm and the other measure, $\left(\overline{\tau}_{\Delta}^{d}\right)$ is travel time calculated based on historical data (Equation (3.3)):

$$\overline{\tau}_{\Delta}^{d} = \frac{1}{|D|}\sum_{i \in D} \tilde{\tau}_{\Delta}^{i} \tag{3.3}$$

Where, *D* is the set of all days for which data are available and $|D|$ is dimension of the set. The goal of the model by Rice and Van Zwet is to predict travel time on day *d* at time interval $\Delta+\delta$ ($\tau_{\Delta+\delta}^{d}$) where $\delta > 0$. The authors state that on day *d* at time interval Δ, $\tilde{\tau}_{\Delta+\delta}^{d}$ is more accurate than $\overline{\tau}_{\Delta+\delta}^{d}$ for a small δ and vice versa. The researchers define $\tilde{\tau}_{\Delta}^{d}$ and $\overline{\tau}_{\Delta}^{d}$ as two naïve measures of travel time. However, they found, based on empirical data, that a linear relationship exists between $\tau_{\Delta+\delta}^{d}$ and $\tilde{\tau}_{\Delta}^{d}$. The authors state that "this observation has held up in all of numerous freeway segments in CA that we have examined". Consequently, they assumed the following relationship between $\tilde{\tau}_{\Delta}^{d}$ and $\tau_{\Delta+\delta}^{d}$:

$$\tau_{\Delta+\delta}^{d} = \alpha_{\Delta}^{\delta} + \beta_{\Delta}^{\delta}\tilde{\tau}_{\Delta}^{d} + \varepsilon \tag{3.4}$$

Where, ε is a zero-mean random variable. As can be seen in Equation (3.4), the parameters of the model are dependent on Δ and δ. Furthermore, the authors assume that $\alpha_{\Delta}^{\delta} = \alpha_{\Delta}^{\prime\delta}\bar{\tau}_{\Delta}^{d}$ in order to express the future travel time as a linear combination of the available naïve measures of travel time where $\alpha_{\Delta}^{\prime\delta}$ is another parameter that changes with δ and Δ.

The model in Equation (3.4) uses only one data point in the past. However, there are more data points in the past that might be very similar to the present day. Consequently, the authors used the "nearest neighbour" method to identify the closest days in the past that can be used to predict the travel time. The nearest neighbour technique tries to find some days in the past that are closest to the present day in some sense. Then it is assumed that travel time in the future would be equal to corresponding travel time observed on that particular day and time interval in the past. In this research, two distance measures were defined between two days d_1 and d_2:

$$m(d_1, d_2) = \sum_{i=1}^{N} \sum_{\Delta<\Delta^*} \left| S_{\Delta}^{i,d_1} - S_{\Delta}^{i,d_2} \right| \tag{3.5}$$

$$m(d_1, d_2) = \left(\sum_{\Delta<\Delta^*} \left(\tilde{\tau}_{\Delta}^{d_1} - \tilde{\tau}_{\Delta}^{d_2} \right)^2 \right)^{1/2} \tag{3.6}$$

Where,

$m(d_1, d_2)$: distance measure between day d_1 and d_2,

i : an index denotes number of loop detectors,

N : total number of loop detectors,

Δ : an index denotes time intervals,

Δ^* : current time interval,

$S_{\Delta}^{i,j}$: Speed measured by loop detector i, at time step Δ, on day j = 1, 2.

If day d' minimizes one of the above distances to day d among all previous days available in the historical database, d' is the closest day to the present day and can be used to predict travel time in future (i.e. $\tau_{\Delta+\delta}^{d} = \tau_{\Delta+\delta}^{d'}$). Furthermore, instead of considering a particular day, the average of n closest days can also be used. The other argument that can be made here is that not all time intervals before

Δ^* are significant in selecting the most similar days, but a time window prior to Δ^* could be used in equations (3.5) and (3.6).

Rice and Van Zwet (2004) gathered loop detector data from 116 loop detectors along 48 miles of I-10 East in Los Angeles. The data are associated with trips that started between 5 AM and 8 PM. Time intervals were chosen to be 5 minutes in duration. Two prediction horizons, δ, were chosen, 0 and 60 minutes. They used Equation (3.6) and a time window of 20 minutes for the nearest neighbourhood technique. Furthermore, the average of two closest days in the past was used ($n = 2$). Root mean squared error was used to compare the two prediction models. An estimation of the true travel time was made using detector flow and occupancy measurements:

$$velocity = g \times \frac{flow}{occupancy} \tag{3.7}$$

where in this formula g is average length of vehicles passing over the loop detector. The average length of vehicles, g, varies from one location to the other and also it varies over time of a day. Furthermore, it is a critical parameter that should be selected carefully in order to obtain correct speed values. Jia *et. al.* (2001) proposed a technique to calibrate factor g which was used by Rice and Kim in their research.

Figure 3.1 and Figure 3.2 show the root mean squared error in terms of time of day for prediction horizon 0 and 60 minutes respectively. As expected the mean squared error is smaller during off peak hours and is smaller for a prediction horizon of 0 minutes than 60 minutes. Furthermore, the linear regression technique performs better than the nearest neighbour method.

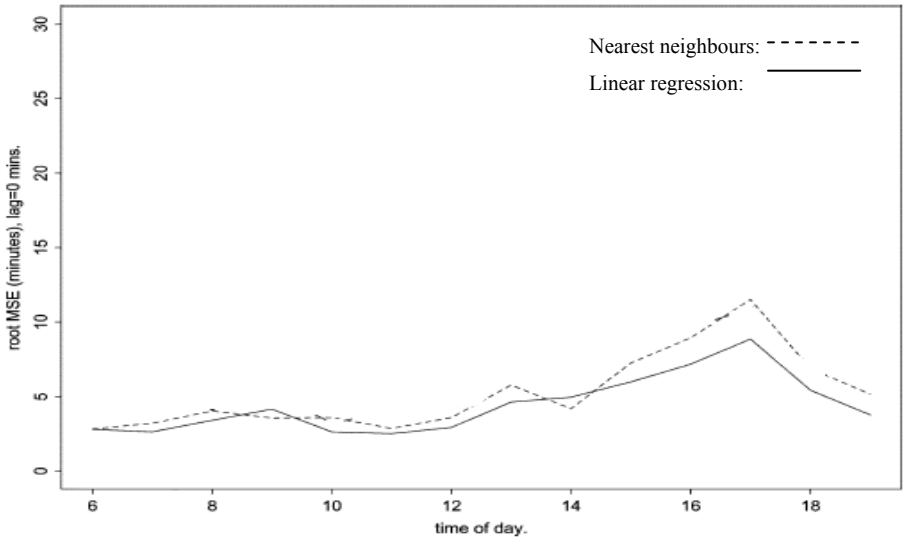


Figure 3.1: Comparison of nearest neighbour and linear regression for prediction horizon = 0 min (Source: Rice and Kim, 2001)

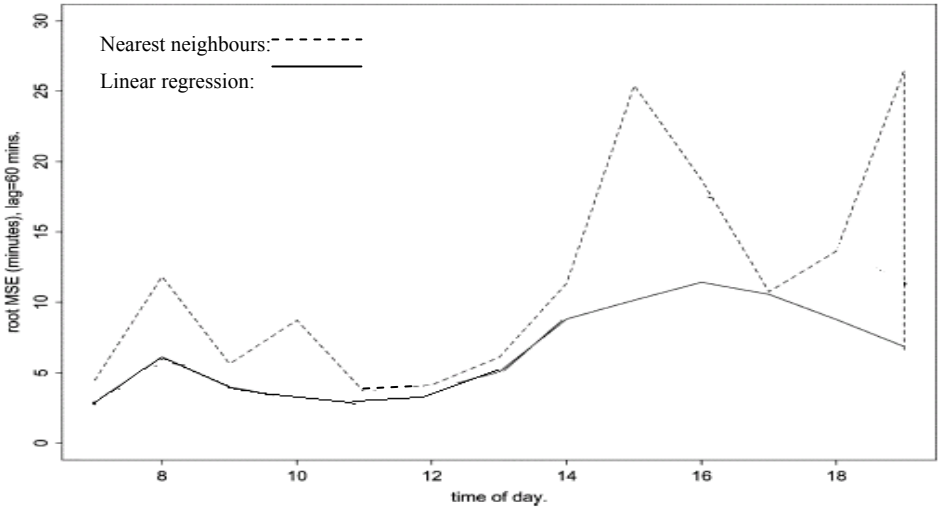


Figure 3.2: Comparison of nearest neighbour and linear regression for prediction horizon = 60 min (Source: Rice and Kim, 2001)

3.1.2 Time Series Models

Traffic data from either probe vehicles or loop detectors represent sequential observations over time of a stochastic process. Consequently, a number of researchers have used time series analysis to tackle the travel time prediction problem. The main goal of time series analysis is to understand the

mechanism of the process that generates those observations and also forecast the future values of the series.

Yang (2005b) used GPS equipped probe vehicles to predict travel time on an arterial in Minnesota. The ARIMA model was used to perform the prediction. However, in this paper no result was reported.

D'Angelo *et. al.* (1999) developed a time series model to predict travel time along a 18 km section of Interstate - 4 in Orlando, Florida. They used data from 25 loop detectors installed in this section of the freeway. Ishak and Al-Deek (2002) extended D'Angelo's model for a 62.5 km section of the same freeway using data from 70 loop detectors. The main objective of their study was to identify significant factors that affect accuracy of the travel time prediction model. In this study, spot speeds obtained from loop detectors are used as the only input to the prediction model. These researchers use the average speed algorithm described in the previous section to estimate average speed on the road section between the two consecutive loop detectors. Then, they utilize a nonlinear time series model to forecast the average speed in near future. The last step is to convert the average speed associated with each section to average travel time.

In this study, the researchers defined a term named "rolling horizon". The rolling horizon is the length of time in the past that is used in the time series model to detect fluctuations in the state of traffic. Furthermore, they defined "rolling step" which is the length of time associated with each time step in the rolling horizon.

The authors evaluated the model for a variety of traffic conditions from free flow conditions to heavily congested conditions. Moreover, they used different prediction horizons (i.e. 5, 10, and 15 min) and different rolling horizons (i.e. 15, 20, 25, and 30 min) with different rolling steps (i.e. 1, 3, and 5 min) to explore effects of these parameters and their combinations on accuracy of travel time prediction. They used relative travel time prediction error as the measure of evaluation.

The researchers found that the model was very sensitive to the level of congestion. The relative travel time prediction error varies between 5% and 30% in which the lower limit is associated with free flow conditions and the upper limit corresponds to extremely congested conditions. The prediction horizon was found to be the other important parameter in the model. The longer the prediction horizon, the less accurate the model was which is in line with expectations. Furthermore, the more congested road sections are more sensitive to the prediction horizon. The researchers

suggest that for free flow condictions, travel time can be predicted for longer periods of time without sacrificing the accuracy of the model.

3.1.3 Kalman Filter Based Models

The Kalman filter, which is called linear quadratic estimation (LQE) in control theory, was first proposed by R.E. Kalman in a famous paper in 1960 (Kalman, 1960). A Kalman filter is an optimal recursive data processing algorithm that is able to estimate the state of a stochastic discrete - time controlled process through a number of known, but noisy observations. The Kalman filter incorporates all available measurements, regardless of their accuracy, to estimate the current or future state of the system using (Maybeck, 1979):

a) knowledge about dynamics of the system and measurement devices,

b) statistical descriptions of the system noises, measurement errors, and uncertainty in dynamics models,

c) any available initial information about the state of the system.

For example assume that one attempts to predict travel time for a section of a freeway using loop detector data (spot speed, occupancy, and flow rate). Having knowledge of the relationship between these measured data and travel time and also errors in loop detector measurements and in the relationship between loop data and travel time, Kalman filter can be used to combine all of this data to generate an overall best estimation of travel time. Kalman filter is generally used to solve a set of equations which are over-determined. In other words, if a process can be described by a number of equations and also a number of measurements (which may contain errors), Kalman filter can be used to estimate or predict the process.

Yang (2005a) used GPS equipped vehicles and Kalman filtering to predict arterial travel time in event of a sudden traffic surge. In this research, three test vehicles with GPS receiver were used to collect and report travel time data for a pre-specified path for the University of Minnesota Duluth graduation ceremony. The test vehicles were running at 3 or 5 minute headway. Having returned to the original point, the test vehicles rejoined the traffic stream until the traffic condition around the place of graduation ceremony was back to normal. In this study the departure time interval of test vehicles is in fact the Kalman filter time step used for prediction of the next time step.

The state of process in the Kalman filter that should be predicted in this study was travel time. The state and measurement equations in the Kalman filter are shown as follows:

$$\tau_{\Delta} = \varphi_{\Delta}\tau_{\Delta-1} + w_{\Delta} \tag{3.8}$$

$$z_{\Delta} = \tau_{\Delta} + v_{\Delta} \tag{3.9}$$

Where, ϕ_{Δ} is a scalar that relates travel time at time step Δ to the travel time at previous time step. In these equations w_{Δ} and v_{Δ} are process and measurement noise terms that are assumed to be white and normally distributed with 0 mean and variance R and Q respectively.

Yang used the following procedure to predict travel time:

Step 1: Initialization

Set Δ = 0 and let $E\left[x_0\right] = \hat{\tau}_0$ and $E\left[e_0^2\right] = P_0$

Step 2: Prediction

State estimate prediction: $\hat{\tau}_{\Delta+1}^{-} = \varphi_{\Delta}\hat{\tau}_{\Delta}$

Error covariance prediction: $P_{\Delta+1}^{-} = \phi_{\Delta}P_{\Delta}\phi_{\Delta}^{T} + Q_{\Delta}$

Step 3: Kalman gain calculation

$$G_{\Delta+1} = P_{\Delta+1}^{-}\left(P_{\Delta+1}^{-} + R_{\Delta+1}\right)^{-1}$$

Step 4: Correction

State estimate correction: $\hat{\tau}_{\Delta+1} = \hat{\tau}_{\Delta+1}^{-} + G_{\Delta+1}\left(z_{\Delta+1} - \hat{\tau}_{\Delta+1}^{-}\right)$

Error covariance correction: $P_{\Delta} = \left(1 - G_{\Delta}\right)P_{\Delta}^{-}$

Step 5: Let Δ = Δ + 1 and go to Step 2 until the present time period ends.

Note that in this research P_{Δ} is scalar and Q_{Δ} is variance of travel time measurements at time step Δ and R_{Δ} is variance of process estimation error.

To quantify the error associated with travel time prediction, a mean absolute relative error (MARE) was used which is defined as:

$$MARE = \frac{\sum_{k=1}^{K}\left|\frac{\tau_k - \hat{\tau}_k}{\tau_k}\right|}{K} \times 100\% \tag{3.10}$$

Where, τ_k is the true travel time experienced by the test vehicles in time step k. The duration of data collection in this study was 45 minutes and the time steps were 3 minutes long. Thus, the total number of time steps, K, was 15.

They found that the prediction error, expressed in MARE, was 17.61% which the authors of the study claimed to be acceptable for the city's traffic operations activities. It was also found that as the process noise variance increases and the measurement noise variance decreases the MARE drops, which means that the *a priori* estimates are more accurate than measurements. In this study, they collected data for two similar days and used the first one as historical data.

3.1.4 Artificial Neural Networks

Artificial Neural Networks (ANN) has been extensively applied to different fields of engineering since late 80's. Traffic engineering and transportation planning have been in line with the other fields of engineering and have witnessed different applications of ANN. In general, ANN can be used for clustering, pattern recognition, function approximation, prediction of dynamic systems, etc.

ANN is designed to mimic the human brain. Humans are able to make decision based on incomplete and imprecise information and also in an uncertain and unknown environment. Similar to the human brain, an ANN consists of many interconnected and parallel processors that are called "neurons". Figure 3.3 shows a typical structure of ANN. As can be seen in this figure, ANN consists of some neurons or nodes[1] that are interconnected to each other through unidirectional or in some cases bidirectional weighted connections (Karray *et. al*., 2004). The ANN can be classified into different categories according to architecture, learning paradigm, and activation (transfer) functions.

In ANN, the nodes are laid in different layers. The way that layers and nodes are designed and connected to each other is termed "architecture or topology". The most commonly used architectures are "feedforward" and "recurrent" architecture.

The most popular topology is feedforward architecture which is illustrated in Figure 3.4(a). In this architecture the network starts with an input layer and ends with an output layer. The main computational capabilities of the network are concentrated in the layers which are termed hidden layers and are placed between the input and output layers. In feedforward architecture neurons are connected to each other through unidirectional connections.

In the recurrent architecture, the neurons are not necessarily connected to each other through unidirectional connection. In other words, the connections might be bidirectional or they can even provide feedback to the neuron itself (Figure 3.4(b)). This type of topology is of great importance especially in modeling and identifying dynamic systems (Karray *et. al*., 2004).

[1] Node and neuron are used interchangeably in this document.

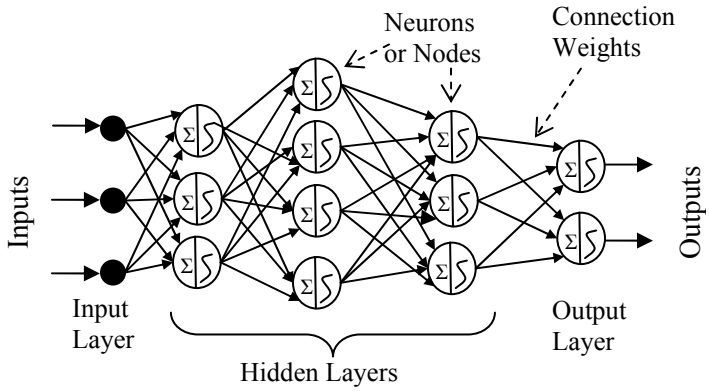
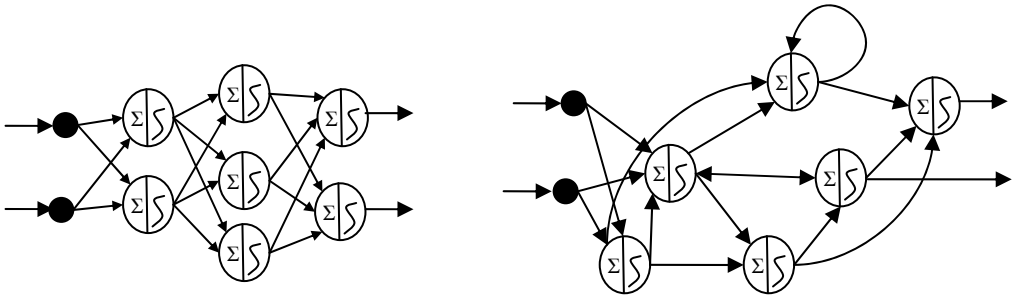


Figure 3.3: Typical structure of artificial neural networks



a) Feedforward architecture. b) Recurrent architecture.

Figure 3.4: Feedforward and recurrent architecture

One of the main characteristics of the ANN is "learning". Learning is the procedure of using available data to adjust weights of the ANN connections in order to infer as many right decisions as possible in future. Although there are a number of learning paradigms in the literature, the most commonly used techniques are "supervised" and "unsupervised" learning strategies.

In supervised learning, the ANN is exposed to *a priori* known data and the objective of the learning algorithm is to determine the weights associated with connections in order to minimize the distance between the model output and the known data. To solve the minimization problem different approaches such as genetic algorithms, simulated annealing algorithms, and gradient descent based algorithms can be used (Karray *et. al.*, 2004). In unsupervised learning, however, an external teacher is not involved in the process of learning. The network is provided with inputs but not with known outputs. The system itself must then decide based on some predefined guidelines what features it will use to group the input data.

The other important components of a neural network are “activation functions” which are the basic elements of the computational processor for the neural networks. As can be seen in Figure 3.5, any individual neuron consists of two components: the first component calculates the weighted summation of the inputs and the activation function performs a nonlinear or linear mapping before delivering the output to the other neurons. A variety of activation functions can be used, depending on the problem at hand and the location of the neuron in the layer. Figure 3.6 illustrates the four most commonly used activation functions.

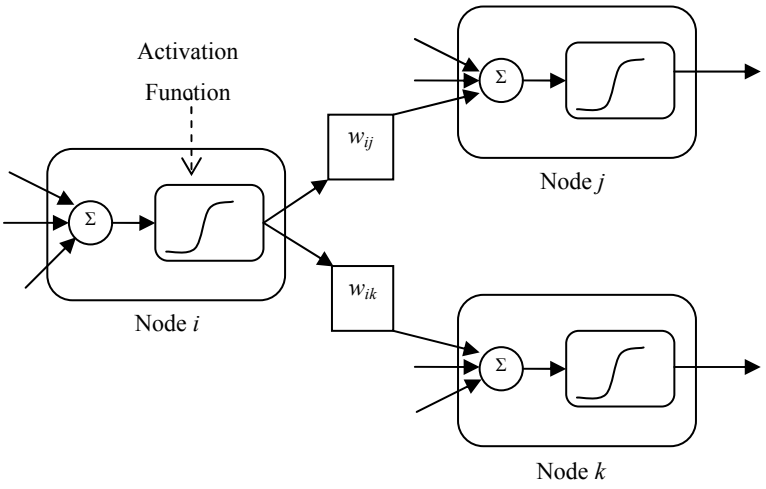


Figure 3.5: Interconnection between neurons (Karray *et. al*., 2004)

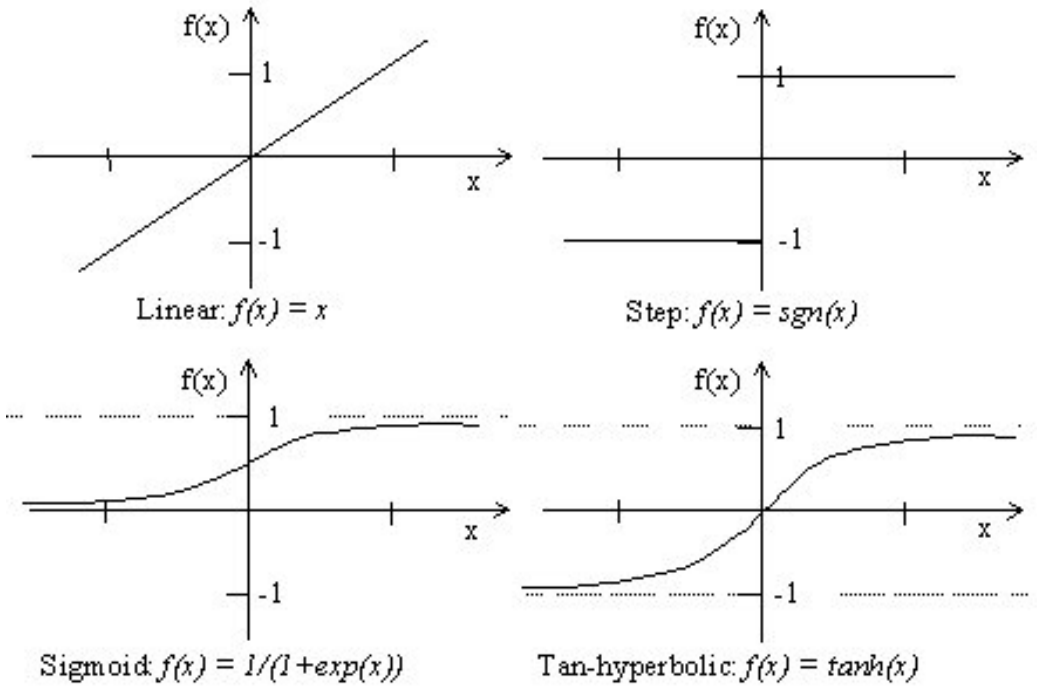


Figure 3.6: Typical profiles of four activation functions (http://www.stowa-nn.ihe.nl/ANN.htm)

Use of ANN models in travel time related problems began in the early 1990's (Nelson *et. al*., 1993 and Blue *et. al*., 1994). The first attempts to use neural networks in travel time prediction were done by Park and Rilett (1998, 1999).

Park and Rilett (1999) used a multilayer feedforward neural network to predict travel time on US-290 in Houston, Texas. In this study, they used probe vehicle data collected over a 27.6 km section of the freeway from April 1994 to July 1994. Data were collected 24 hours per day. The probe vehicles were equipped with transponders and were identified by an automatic vehicle identification (AVI) system. The freeway section contained 7 AVI tag readers, dividing the section into 6 links. The data were aggregated at 5-minute periods for each link. Three of the links had approximately 7 to 10 observations per link per time interval and the remaining 3 links had 3 to 6 observations per link per time interval. The neural network model was trained using data from the AM peak period.

The authors state that "… traffic flow patterns of the neighbouring links (upstream and downstream links) also should be considered for input into the neural network model. Intuitively, traffic flow-related parameters of the neighbouring links could provide useful information on the link of interest." The researchers used the correlation coefficient to investigate the relationship between travel time of

each individual link and traffic flow- related parameters of neighbouring links. They found that travel time observed on a link is highly correlated with travel time observed on neighbouring links. In some cases the coefficient of correlation was as high as 0.75. The authors found that the correlation between travel time of each link and the neighbouring links increases when the state of traffic is changing (*i.e.* the state of traffic is changing from uncongested to congested and vice versa).

The researchers used a fully connected multilayer neural network. The backpropagation algorithm was used for training the neural network. The backpropagation algorithm is a supervised learning method that is used in feedforward neural networks. The objective of this algorithm is to determine the weights between every pairs of neuron in order to minimize the total error between estimated values and target values.

Park and Rilett tested four different models for one of the middle links. The difference between the models was the number of input variables. Every input variable in this study is travel time of a time step which is associated with either the link of interest or neighbouring links. The first model used only travel time of the same link for prediction of travel time in the future. The rest of the models used different combinations of travel times associated with neighbouring links as well as travel time of the link of interest. Furthermore, they changed number of inputs for each model to find the best results.

Although there is no universal rule to determine the layout of layers in neural networks, the researchers state that "the appropriate number of hidden neurons and layers of the neural model depends on the pattern and complexity of the approximated function and the transfer function of the layers". Via trial and error, the authors found the best number of neurons for the hidden layer of each model to be between 4 to 6 neurons. Moreover, they realized that if more neurons are used, the accuracy of the model in this case decreases. In their proposed model (Figure 3.7), seven preceding travel times of the link of interest are used as input to the neural network. The model consists of three layers (one hidden layer) and the model is able to predict travel time in 5 time steps ahead.

Park and Rilett prepared a neural network model and tested it using 79 days of data for training and 1 day for validation. This procedure was used 80 times. In other words, every day of data was used as testing data once. To compare different models the average error was computed across all 80 trials. It was found that predicting one time step ahead incurs 7.4% to 10.5% error. As the prediction horizon increases, the error increases linearly.

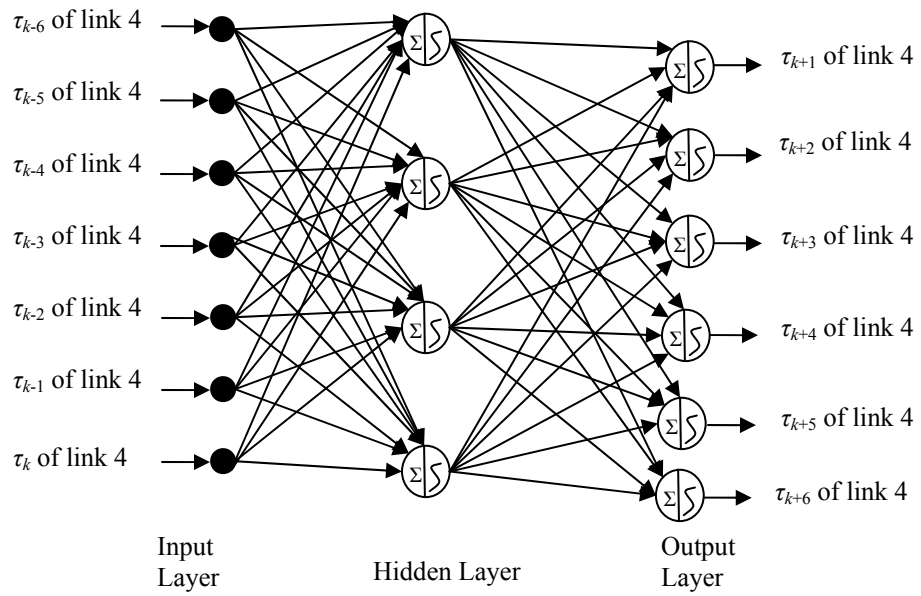


Figure 3.7: A sample of input-output structure of the artificial neural network

Van Lint *et. al*. (2002) used state-space neural networks to predict travel time in freeways. The state-space neural network is a recurrent neural network in which the state of a system at time k+1 is determined by its state at time step k and other inputs (speed and flow rate) at time step k, $\mathbf{u}_k$**.** In this study they introduced two state-space neural networks namely, fully connected and partially connected state-space neural networks. Figure 3.8 illustrates the fully connected model. This model consists of four layers: an input layer, one hidden layer, an output layer, and a context layer. The hidden layer includes N neurons each of which corresponds to a link of the route of interest. The context layer also consists of N neurons. The responsibility of this layer is to store the pervious state of each neuron in the hidden layer. The only difference between the fully connected network and partially connected network is the way the context layer is connected to the hidden layer. In the partially connected network only previous state of a particular neuron and previous states of the upstream and downstream neurons are connected to the corresponding neuron in the hidden layer while in the fully connected network previous states of all neurons are connected to every individual neuron in the hidden layer.

The authors used simulated loop detector data to train and evaluate the models. They simulated 7.3 km of the southbound direction of the A13 freeway in the Netherlands using FOSIM simulation

model. Three runs of the simulation were used to train the models and 2 simulation runs were used for verification of the models. They used a supervised learning algorithm to determine weights associated with each network.

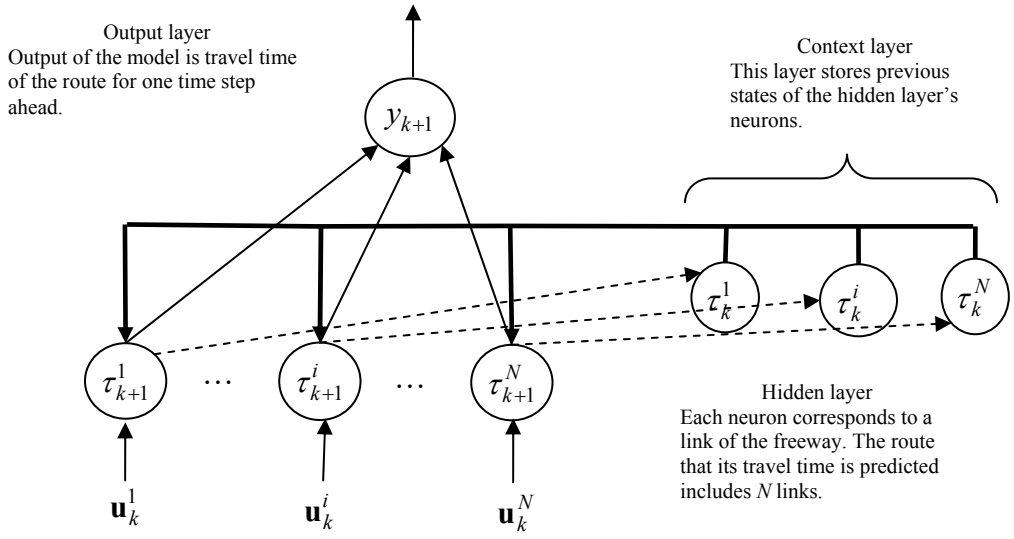


Figure 3.8: Fully connected state-space neural networks for travel time prediction along a route (Van Lint *et. al.*, 2002)

The researchers found that the outputs of the two models are not statistically different. Furthermore, they observed that the largest prediction error occurs when congestion is forming or dissipating. For other conditions, even severe congestion, both models predict travel time with errors less than 10% of the time travel time.

3.2 Traffic Flow Models

Traffic flow theory seeks to describe interactions of vehicles and infrastructure (highways and control devices) in mathematical forms. Traffic flow theories have been developed for and applied to supply-demand modelling, capacity and level of service analysis, traffic stream models, continuum modelling, shockwave analysis, queuing analysis, and simulation models (May, 1990), all of which can be used to estimate and predict travel time in a transportation network. However, the level of details and input requirements of these analysis methods differ. For example, traffic assignment models that fall into supply-demand models require supply characteristics (geometry of the network, cost functions of the links, *etc.*) and demand characteristics (*e.g.* OD matrix) and are likely not best

suited for short term travel time prediction. Microscopic simulation models can also be used to predict travel times but these models are challenging to calibrate and require future OD demands and network capacities –information which is generally not available.

3.2.1 Shockwave Analysis

Shockwaves are defined as boundaries in the time - space domain that represent discontinuities in flow and/or density (May, 1990). In other words, shockwaves are the boundary of an abrupt change in density. Figure 3.9 illustrates a 3 lane section of a roadway with no on ramp or off ramp. One lane is reduced in this roadway due to construction. If the traffic demand on this roadway section increases from a value equal to 2 times the capacity of a lane to 2.5 times the capacity of a lane, then a queue will begin to build upstream of the construction zone. Vehicles approaching the construction zone first travel in uncongested traffic (state *A*) and then join the queue (state *B*). In state *B* traffic is in the congested regime. The arrow that separates state *A* and state *B* is the shockwave. The shockwave in this case propagates backward (i.e. upstream). The speed of propagation is denoted by ω_{AB}.

The speed of the shockwave can be derived assuming conservation of vehicles. The shockwave speed is derived with respect to Figure 3.9. The number of vehicles leaving flow state *A*, N_A, must be exactly equal to the number of vehicles entering flow state *B*, N_B. Furthermore the speed of vehicles in state *A* and *B* relative to the shockwave respectively are $u_A - \omega_{AB}$ and $u_B - \omega_{AB}$ where u_A and u_B are the average speed of vehicles in state *A* and *B*. Consequently, N_A and N_B can be calculated as follows:

$$N_A = (u_A - \omega_{AB}) \times D_A \times T \tag{3.11}$$

$$N_B = (u_B - \omega_{AB}) \times D_B \times T \tag{3.12}$$

where, D_A and D_B are average densities corresponding to state *A* and *B* respectively and *T* is an arbitrary period of time. According to the above two equations and conservation of flow, the propagation speed of shockwaves can be calculated as follows:

$$\omega_{AB} = \frac{u_A D_A - u_B D_B}{D_A - D_B} = \frac{q_A - q_B}{D_A - D_B} = \frac{\Delta q}{\Delta D} \tag{3.13}$$

The above equation means that the speed of a shockwave is equal to the change in flow rate over the change in density.

Shockwave analysis is an important tool for analyzing freeways and intersections. However, as the size of the network increases it is very difficult to analyze the network using shockwaves. A shockwave based simulation model has been developed by Imada and May (1985) to analyze traffic in freeways. One of the reasons that the shockwave analysis and continuum modelling of traffic have not received enough attention is the fact that these models require a few parameters that are very difficult to estimate using traditional sources of traffic data. One of these parameters that these models are ironically very sensitive to is capacity of the bottleneck. The other requirement of these models is that they need flow rate of traffic at all entrances and exits. If these data are available or can be reliably estimated, then shockwave analysis can be used to estimate queue propagation over time-a factor having a significant influence on travel time. As described in the following chapter, shockwave propagation is one of the key ideas employed in the proposed travel time prediction model, in which existence and speed of a shockwave is determined using the estimated trajectories of the vehicles with mobile phone location data.

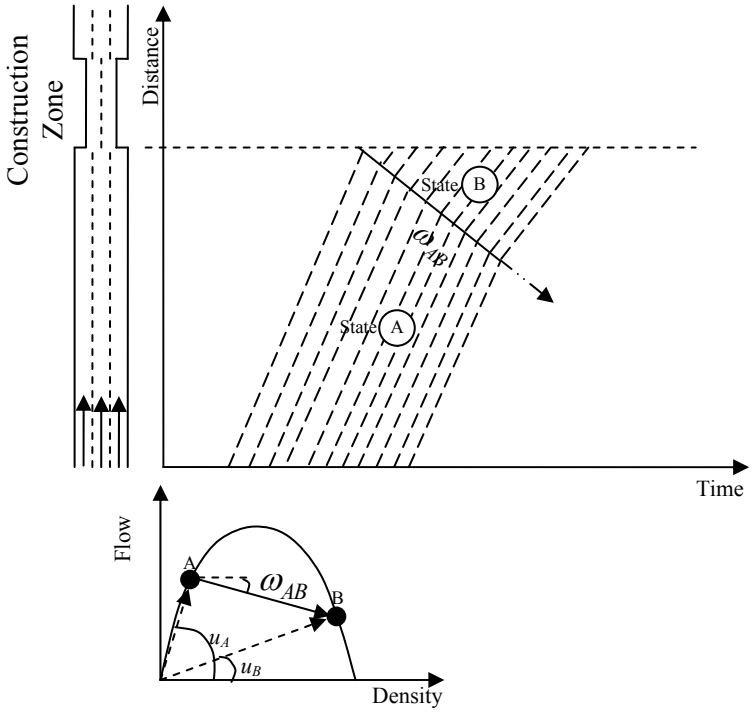


Figure 3.9: A typical shockwave in a construction zone

3.3 Summary

This chapter reviewed the-state-of-the-art in prediction of travel time. Travel time prediction models were divided into two groups: data driven models and traffic flow models. Regression models, time series models, Kalman filter, and neural networks were studied as data driven models. Traffic assignment models, microscopic simulation models, and shockwave analysis were briefly introduced under traffic flow models.

Data driven models have been used by researchers successfully. However, these models lack transferability. In other words, they are location specific which is not necessarily favourable. On the other hand, there are many ready-to-use applications available in the market which make application of these models easy. The other drawback of these models is that they are not flexible enough to account for new strategies that are adopted by managers of the network.

Traffic assignment models and microscopic simulation models are potentially powerful and reliable to predict travel time in a network. However, these models are very data intensive. They require OD matrix for future which are very difficult to predict. These models are unlikely to have promising results using data from mobile phones. Consequently, they are not considered in this study. The shockwave theory, however, can be used to determine the state of traffic in the network. This technique will be used as a part of the proposed framework which will be described in the next chapter.

Chapter 4
Methodology

This chapter comprised of two sections. The first section presents the proposed methodology for predicting travel time based on mobile phone location data. The second section presents two algorithms to predict travel time using loop detector data. The loop detector data will be used for comparison purposes.

4.1 Travel Time Prediction Using Probe Vehicle Trajectories

The main assumption behind the proposed methodology is that positions of a set of probe vehicles are available periodically. As discussed earlier, this kind of information is conceived to be available through dedicated probe vehicles (probe vehicles equipped with GPS) and mobile phone based traffic monitoring systems. The proposed methodology provides real time prediction of travel time (i.e. prediction horizon in this research is assumed to be 0).

Figure 4.1 shows a time-space diagram for a route. The route is discretized into a number of road sections denoted by i whose length is denoted by L^i. Time is divided into time intervals denoted by Δ. This figure also illustrates four different probe vehicles which show 3 different cases that may occur. At this time assume that it is desirable to predict travel time for time interval $\Delta = 2$. Probe 2 shows a case where a probe vehicle enters the route during a time interval and finishes the route in the same time interval. Therefore, travel time of this probe vehicle can be measured directly using its trajectory. Had all vehicles entered the route and finished their trip in the same time interval, travel time could have been directly measured. The other case is shown by probe 1 where this probe vehicle entered the road section during time interval 1 and finished the route in time interval 2. This probe vehicle is irrelevant to prediction of travel time associated with time interval 2. But, the data points associated with this probe vehicle in time interval 2 can be used to calculate average speed of each road section L^i. Probe 3 and probe 4 are examples of the case whose travel time is deemed to be predicted. The difference between these two probe vehicles is that for probe 3, only travel time associated with time interval $\Delta = 2$ has to be predicted and for probe 4 travel time of time intervals $\Delta = 3$ and $\Delta = 4$ need to be predicted.

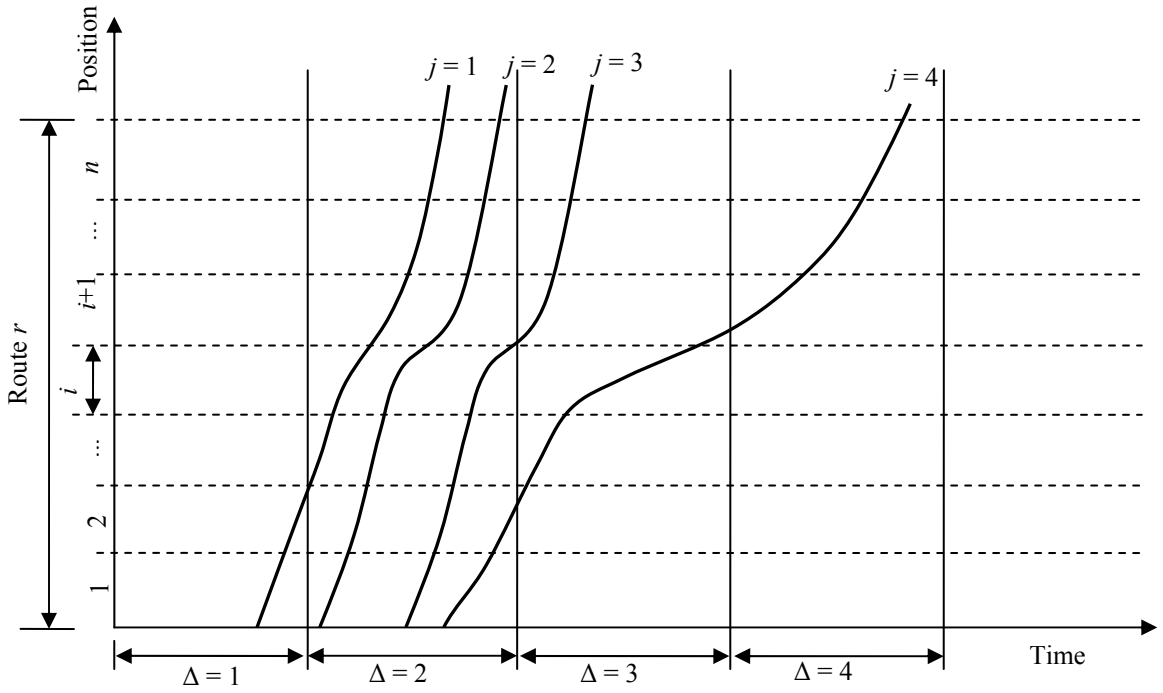


Figure 4.1: Discretization of time and space

Average speed of each road section over time period Δ can be computed using the following equation:

$$\overline{S}_{\Delta}^{i} = \frac{1}{J_{\Delta}^{i}} \sum_{j=1}^{J_{\Delta}^{i}} \frac{d_n^j - d_1^j}{t_n^j - t_1^j} \tag{4.1}$$

Where,

$\overline{S}_{\Delta}^{i}$: average speed of road section *i* over time interval Δ,

J_{Δ}^{i} : number of probe vehicles which have any observations during time interval Δ on road section *i*,

d_n^j : position component of the last observation on road section *i* associated with probe vehicle *j*,

d_1^j : position component of the first observation on road section *i* associated with probe vehicle *j*,

t_n^j : time component of the last observation on road section *i* associated with probe vehicle *j*,

t_1^j : time component of the first observation on road section *i* associated with probe vehicle *j*.

In order to visually explain the above equation, Figure 4.2 illustrates the time-space diagram associated with a hypothetical route. The route was discretized into two sections. It is desired to compute the average speed of section 2 during time interval 2 (i.e. $\overline{S}_2^2$). Four probe vehicles were observed during time interval 2. The first three probe vehicles have observations on section 2 of the route during time interval 2 (i.e. $J_2^2 = 3$). $t_1^j, t_n^j, d_1^j,$ and d_n^j are shown in Figure 4.2 using green circles. If these values are inserted into Equation(4.1), average speed of road section 2 during time interval 2 is calculated.

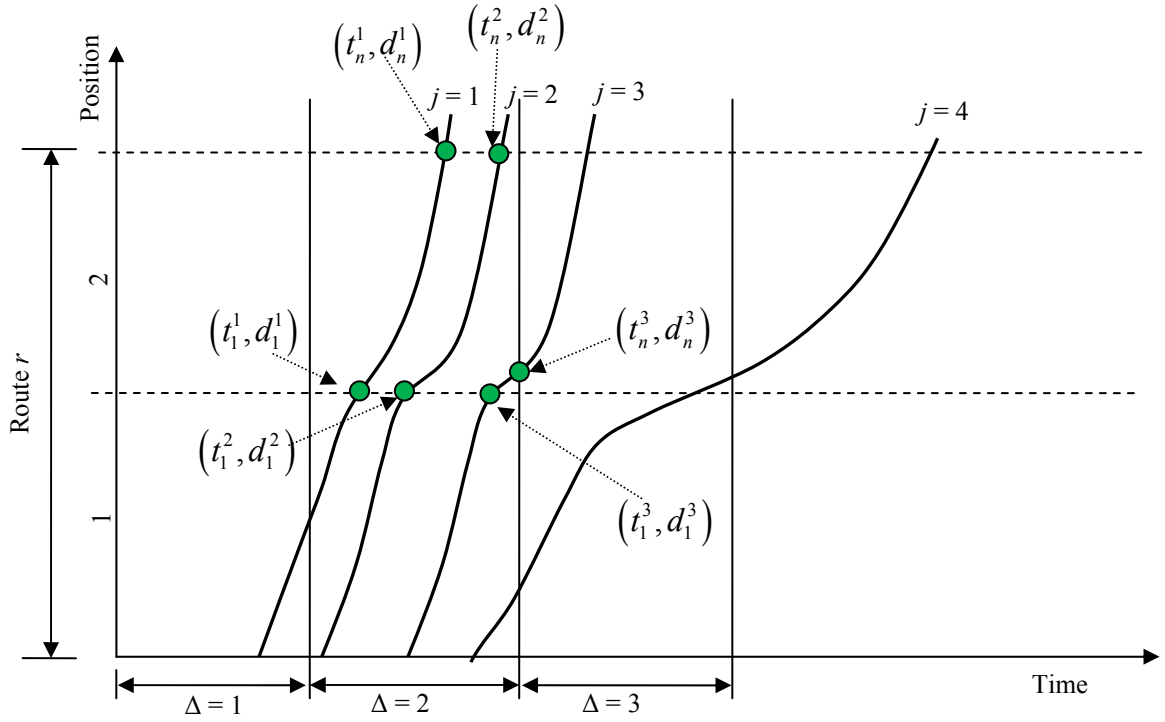


Figure 4.2: Calculation of average speed for each road section

Using the average speeds associated with each road section during the current time interval, travel time of route *r* can be predicted using the following equation:

$$\tau_\Delta^r = \sum_{i=1}^{n} \frac{L_i}{\overline{S}_\Delta^i} \qquad (4.2)$$

Where, τ_{Δ}^{r} denotes the average travel time of route *r* during time interval Δ and *n* is the number of road sections on route *r*. The assumption behind this equation is that speed of road section *i* is constant during all time periods. This equation is incapable of providing accurate travel time for the route if a substantial proportion of vehicles which enter the route finish the route in the next time intervals and the state of traffic associated with the road sections which are travelled during the next time intervals changes. In these cases, the average speed of those road sections in the future should be used in this equation. The following methodology which uses the shockwave information can be used to predict average speeds of those road sections in the future.

When the flow or density of traffic changes, a shockwave is generated along the boundary of the two states of traffic and the shockwave propagates along the route. Consequently, if shockwaves can be detected and their attributes be estimated, the trajectory of the shockwaves can be projected into the future and average travel time of each road section can be predicted using the projected trajectory of the shockwaves.

Figure 4.3 illustrates the same probe vehicles shown in Figure 4.1. However, this figure shows a shockwave which was detected during the current time interval (time interval 2). The detected shockwave is progressing backward throughout the subject route at speed ω. The projected trajectory of the shockwave is also shown in the figure with a dashed line. The projected trajectory shows that $\overline{S}_3^{i-1} \neq S_2^{i-1}$ and $\overline{S}_4^{i-1} \neq \overline{S}_3^{i-1} \neq S_2^{i-1}$ because the shockwave which was detected during time interval 2 will reach road section *i*-1 during time interval 3 and will traverse this road section during time interval 4. This shockwave is a boundary between the traffic condition which was prevailing on road section *i*-1 during time interval 2 and most of time interval 3 and the traffic condition which is downstream of the detected shockwave.

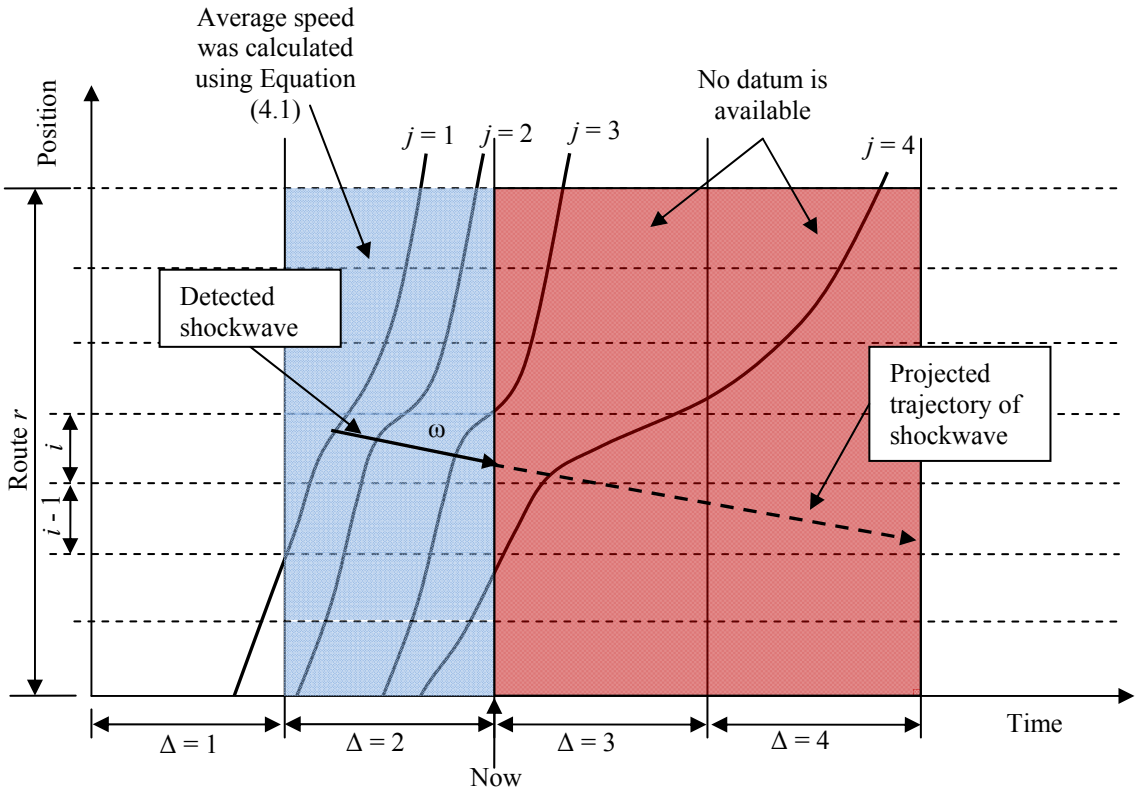


Figure 4.3: Projection of shockwave trajectories

A methodology to detect major shockwaves along the subject route is proposed later in this chapter. This methodology provides the following information about each shockwave:

1. Speed at which the shockwave is propagating along the route, ω
2. Average speed of vehicles before reaching the shockwave (upstream of the shockwave), S_u
3. Average speed of vehicles after impacting with the shockwave (downstream of the shockwave), S_d
4. Point in time and space where the shockwave was first detected during the current time interval, (t_0, x_0)
5. Point in time and space where the shockwave was last detected during the current time interval, (t_n, x_n)

Assuming that all shockwaves and information associated with them (which was described above) are known, Equation (4.3) can be used to modify the average speed associated with each road section for the future time intervals:

$$\tilde{S}^i_{\Delta+h} = \overline{S}^i_{\Delta} \times f \tag{4.3}$$

Where,

$\tilde{S}^i_{\Delta+h}$: denotes modified average speed of section *i* for time interval $\Delta+h$,

f : is the modification factor which is a number larger than 0.

The challenge here is to choose a precise modification factor in order to increase the accuracy of the prediction model. If the projected trajectories of shockwaves do not pass through road section *i*, the modification factor, *f*, is 1. If a shockwave with $S_d < S_u$ passes through road section *i*, *f* should be less than 1 and if a shockwave with $S_d > S_u$ passes through road section *i*, *f* should be greater than 1.

For the cases where n_s shockwaves pass through road section *i*, the following modification factor is proposed:

$$f = \begin{cases} \dfrac{1}{(n_s+1)}\left[\sum_{s=1}^{n_s} \dfrac{S^s_d}{S^s_u} + 1\right] & if\ n_s \geq 1 \\ 1 & if\ n_s = 0 \end{cases} \tag{4.4}$$

The basis for this equation is that when a shockwave passes through a road section, it changes the average speed of the road section. Figure 4.4 illustrates a case where a shockwave is detected in time interval $\Delta = 1$. Assuming the characteristics of the shockwave (i.e. propagation equation, average speed of vehicles upstream and downstream) are known based on the data received in time interval $\Delta = 1$, the projected trajectory of the shockwave passes through road section *i*. The implication of this projection is that the speed downstream of the shockwave changes from $\overline{S}^i_1$ to S_d. Now two assumptions are made:

1. A shockwave changes the average speed of a road section proportional to S_d / S_u .
2. Shockwaves divide a road section into equal part in time and space where the average speed of each section is influenced by the two boundary shockwaves.

In Figure 4.4 there is only one shockwave therefore $f = \left(\frac{1}{2}\right) \times \left(1 + \frac{S_d}{S_u}\right)$ and $\tilde{S}^i_2 = \overline{S}^i_1 \times \left(\frac{1}{2}\right) \times \left(1 + \frac{S_d}{S_u}\right)$. Assuming $\overline{S}^i_1 = S_u$ for illustrative purposes, then $\tilde{S}^i_2 = \left(\frac{S_u + S_d}{2}\right)$ which is intuitive.

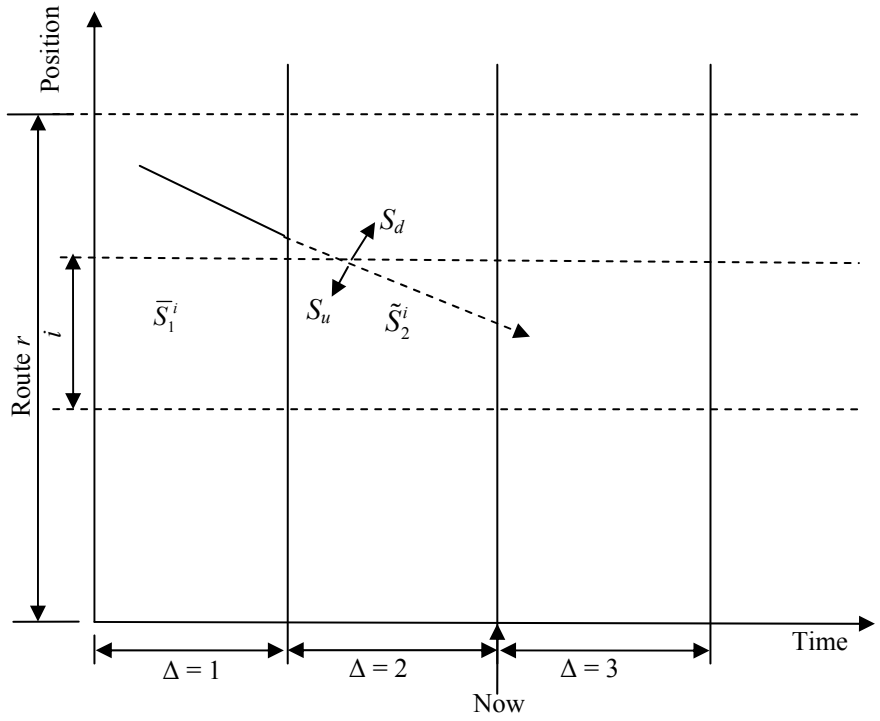


Figure 4.4: Speed change due to a shockwave

In order to predict travel time of route *r* using the average speed and modified average speed of each section, the flow chart shown in Figure 4.5 can be used. In this flow chart t_Δ denotes the duration of time interval Δ and S_m^i represents average speed for section *i* during time interval *m*. If average speed for section *i* during time interval *m* is not known yet, S_m^i represents modified average speed for section *i* during time interval *m*. According to this flow chart, travel times of the sections constituting route *r* are added one by one to $\hat{\tau}_\Delta$. After adding travel time of a section, $\hat{\tau}_\Delta$ is compared with the duration of the time interval, t_Δ. If it is larger than t_Δ, vehicles will likely finish the route during the next time intervals and therefore, for the remaining road sections the travel time obtained from the modified average speed will be used. In this flow chart, once travel time of all road sections are included into $\hat{\tau}_\Delta$, then $\hat{\tau}_\Delta$ will be equal to the predicted travel time for route *r* during time interval Δ (i.e. τ_Δ^r).

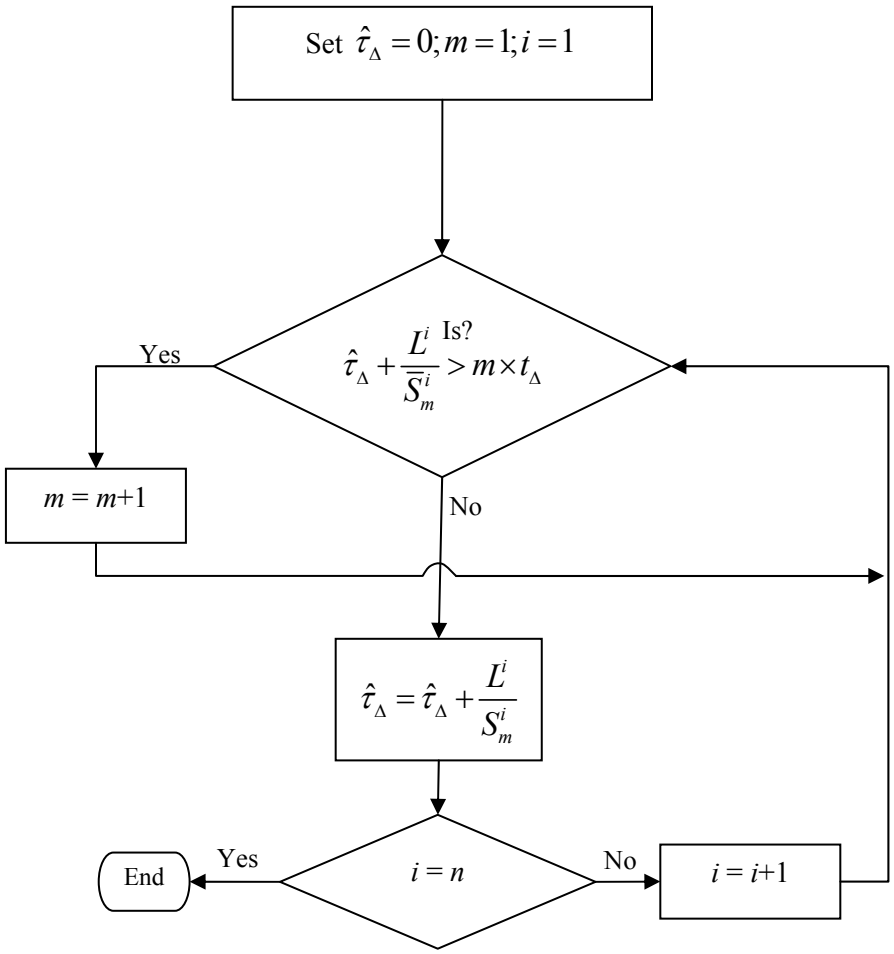


Figure 4.5: Flow chart for prediction of route travel time using average speed of road section

Equation (4.4) used in this methodology is an empirical equation which was used to simplify the problem and reduce the computation burden. This equation and the two assumptions behind the equation were shown to be intuitive. It is possible to calculate the modified speed of each road section more accurately using the shockwave diagrams overlaid on the time-space diagram associated with the freeway section. However, it is very difficult to develop a computer code for a general shockwave diagram.

As explained before, a methodology is proposed to detect major shockwaves. The information obtained from shockwaves is used to determine the modified average speed of each road section. The input data required for the proposed methodology are positions of a sample of vehicles over time. For

each given time interval, the algorithm assesses whether or not the trajectory of a probe vehicle has intersected with a major shockwave and finds the coordinates of the point of the intersection in space and time. Having detected all of the shockwave intersection points in a particular time interval, a linear clustering algorithm is employed to group all the shockwave intersection points associated with each shockwave. A linear regression model is fitted to the data points within each cluster to estimate attributes (i.e. speed, positions and times of start and end) of such shockwave.

In this research, it is assumed that (1) positions obtained from probe vehicles are accurate and (2) a linear model can be used to model the propagation of shockwaves. The first assumption would probably affect the accuracy of the output of the proposed algorithm. The extent of the degradation can be quantified but was not attempted in this research. The second assumption is only true when vehicle headways in the traffic flow are uniform which is rarely true. However, it is anticipated that errors resulting from violating this assumption are relatively small.

According to the above description, the proposed methodology is divided into three parts: (1) automatic shockwave detection, (2) data filtering, and (3) linear clustering algorithm. Details of each are provided in the following sections.

4.1.1.1 Automatic Shockwave Detection

An iterative two-phase piecewise or switching regression is used to detect shockwaves. This type of regression is used when two lines with different slopes fit the data and the "joint point" or the "change point" is not known a priori. In other words, suppose n pairs of data points (t_i, x_i), $i = 1, \ldots, n$ are available, where t_i is the time at which the probe vehicle is reported at position x_i along the roadway. Furthermore, it can be assumed that t_i are ordered in a way that $t_1 \le t_2 \le \cdots \le t_n$. Then, t and x can be related to each other according to the following set of equations:

$$x^1 = a_1 + b_1 t, \quad t \le t_0 \tag{4.5}$$

$$x^2 = a_2 + b_2 t, \quad t > t_0 \tag{4.6}$$

Where (t_0, x_0) is the joint point (Figure 4.6). This problem raises in different fields of science such as biology and econometrics (Vieth, 1989 and Worsley, 1983). Quandt (1958) was the first who addressed this problem. In the context of probe vehicles, the joint point is the point at which the speed of the vehicle has changed and as such likely represents the shockwave boundary between two flow states. These points are referred to as "inflection points". In this research, the two phase piecewise

linear regression model proposed by Vieth (1989) is applied to the time series of probe vehicle position data in order to find inflection points on the trajectories of these vehicles.

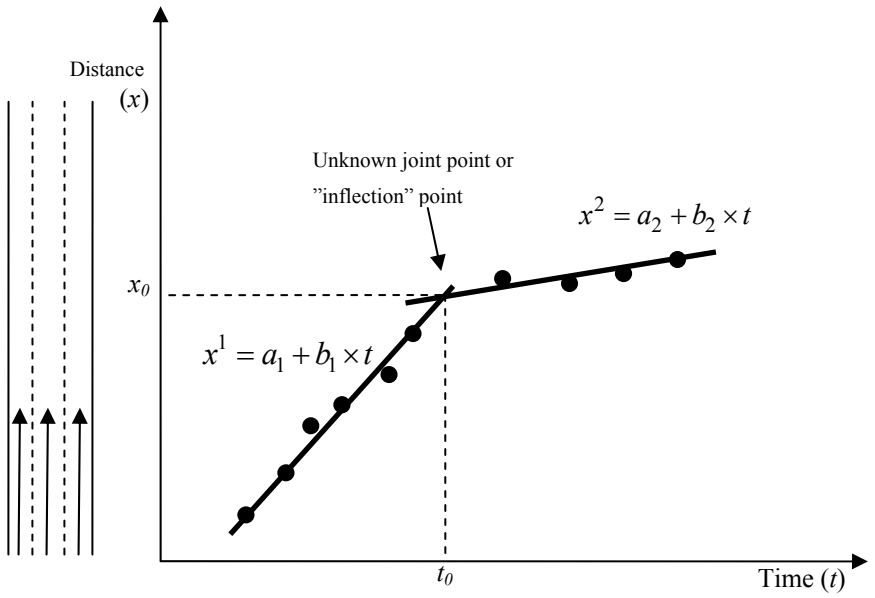


Figure 4.6: Two-phase piecewise linear model

Assume that a sample of P probe vehicles is available during time period Δ. For each probe vehicle p a set of points, $\Gamma^p = \left\{ \left(t_i, x_i \right) \middle| i = 1, \cdots, n^p \right\}$, in time and space is available during time interval Δ. Time interval Δ can also be discretized into a number of smaller time steps. The mathematical program to fit a two-phase linear regression to the points in set Λ^p can be formulated as follows where Λ^p is a subset of Γ^p and includes data points associated with one or more time steps depending on the number of the inflection points detected in previous time steps for probe p. Set Λ^p is defined formally later on with respect to Figure 4.8.

$$Min\, RRS = \sum_{t_i \le t_0} \left[x_i - \left(a_1 + b_1 t_i \right) \right]^2 + \sum_{t_0 > t_i} \left[x_i - \left(a_2 + b_2 t_i \right) \right]^2 \qquad (4.7)$$

s.t.

1) $(t_i, x_i) \in \Lambda^p \qquad \forall i = 1, \cdots, \left\| \Lambda^p \right\|$

2) $t_j \le t_0 < t_{j+1} \qquad \exists j = 1, \cdots, \left\| \Lambda^p \right\| - 3$

3) $\theta \times \left[x_0^l - \left(a_1 + b_1 t_0^l \right) \right] = 0$

4) $\theta = \begin{cases} 0 \text{ if no inflection point has been found so far,} \\ 1 \text{ if at least one inflection point has been found} \end{cases}$

where $\left\| \Lambda^p \right\|$ represents dimension of set Λ^p and $\left(t_0^l, x_0^l \right)$ is the last inflection point that has been found in previous time steps.

In the above mathematical program, the objective function is a minimization of the total residual sum of squares associated with both regression lines fitted to the data. The first constraint guarantees that the lines are fitted to the data in set Λ^p and the second constraint ensures that the inflection point exists within the time limits of the data in Λ^p. In other words, the above program finds the best piecewise linear regression, for which the inflection point falls within the available data. Figure 4.7 illustrates the effect of the second constraint. In this figure in the absence of the second constraint in the mathematical formulation (i.e. Equation (4.7)), piecewise lines AB and AC, are feasible. However, piecewise line AC is not a feasible vehicle trajectory. The second constraint prevents accepting piecewise line such as AC in Figure 4.7. The third and fourth constraints guarantee physical continuity in the trajectory of the probe vehicle. In other words, when an inflection point is detected, the next regression line in the next time step must pass through this point to maintain continuity.

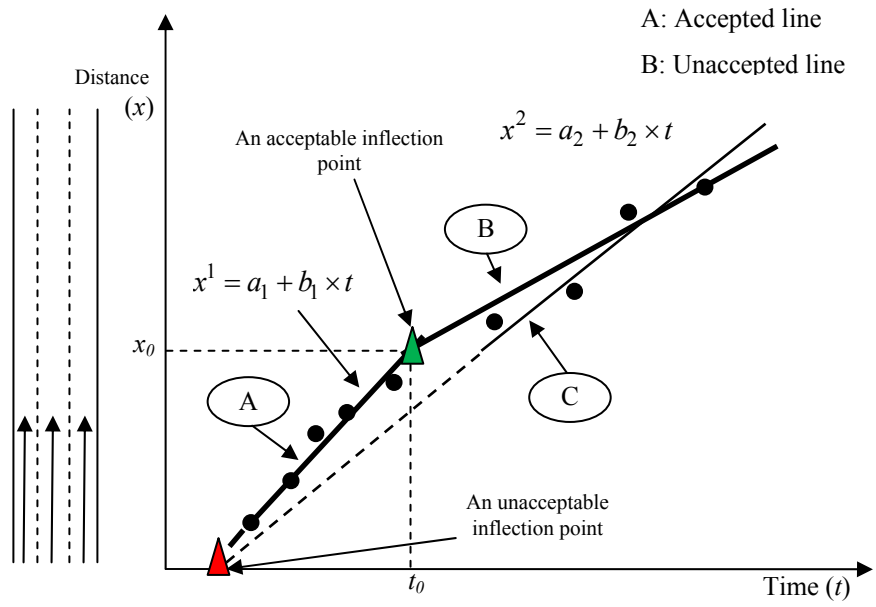


Figure 4.7: Accepted and unaccepted piecewise linear regressions

In order to solve the minimization problem defined by Equation (4.7), the following iterative solution algorithm can be used:

Step 0: Set j=3 and choose set Λ^p.

Step 1: while $j \le \left\|\Lambda^p\right\| - 3$

- Partition Λ^p into two mutually exclusive and collectively exhaustive sets: $\lambda_1 = \left\{(t_i, x_i) \middle| i = 1, \cdots, j\right\}$ and $\lambda_2 = \left\{(t_i, x_i) \middle| i = j+1, \cdots, \left\|\Lambda^p\right\|\right\}$.
- Perform two regular linear regressions in order to find the best lines fitted to λ_1 and λ_2. If an inflection point associated with probe p has been found, the piecewise regression line has to be constrained to pass through this inflection point.
- Calculate the objective function of Equation (4.7), *RSS*.
- Set $j = j+1$

Step 2: Choose the piecewise linear regression with the smallest *RSS*.

Λ^p used above is defined in Figure 4.8 which illustrates all the components of the proposed algorithm to automatically detect shockwaves.

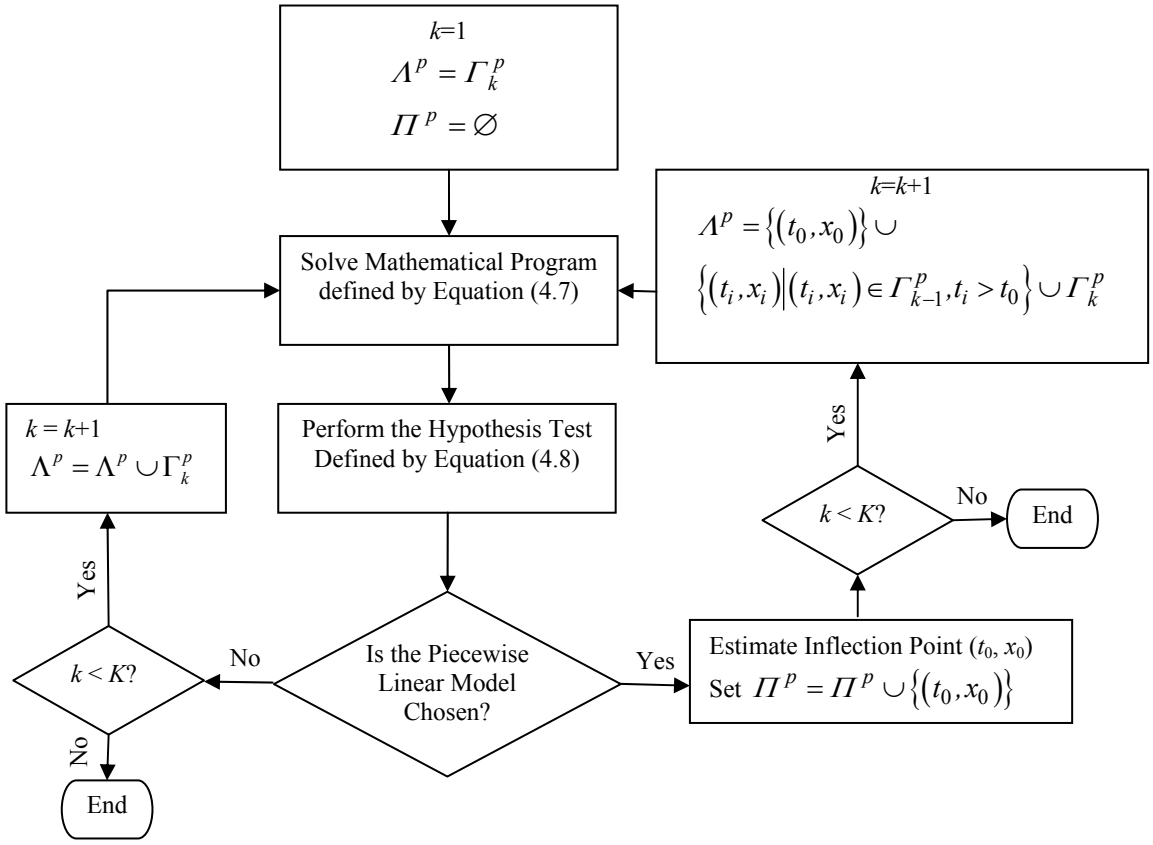


Figure 4.8: Proposed algorithm to estimate inflection points

Assume that in Figure 4.8 time interval Δ can be divided into *K* time steps. Set Γ_k^p in this figure includes all position measurements associated with probe vehicle *p* that are obtained during time step *k*. In this flow chart, Π^p is the set containing all inflection points for probe vehicle *p*. In other words, this set is the output of the above algorithm for probe vehicle *p* over time interval Δ. Set Λ^p includes data from the most recently found inflection point to the end of data points associated with the current time step. Figure 4.9 shows two cases to clarify the definition of set Λ^p. This figure shows trajectories of probe vehicles *A* and *B* during time interval Δ. The time interval is also divided into 4 time steps. Probe vehicle *A* has been traveling at approximately a constant speed. Consequently, the algorithm does not find any inflection point and at time step $k = 4$ set Λ^A includes all available data points ($\Lambda^A = \{(t_i, x_i) | i = 1, \cdots, 9\}$). In the case of probe vehicle *B* an inflection point was found in time step $k = 2$ and after that the probe vehicle travelled at approximately a constant speed resulting in no

other inflection point until $k = 4$. Therefore, set Λ^B contains all data point from the most recently found inflection point to the data point associated with the current time interval inclusive $\left(\Lambda^B = \left\{(t_0, x_0) \cup (t_i, x_i) \middle| i = 7, \cdots, 15\right\}\right)$.

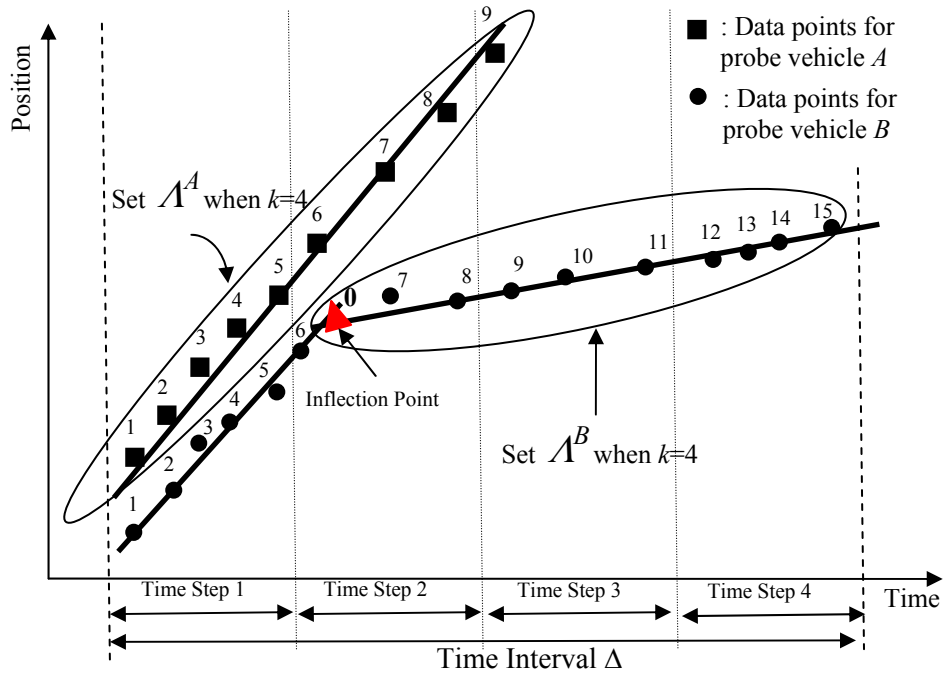


Figure 4.9: Two examples to show set Λ^p

The presence or absence of an inflection point is determined statistically. In Figure 4.10 the piecewise linear regression model defined by Equation (4.7) has been applied to both the trajectory data from two probe vehicles. However, in the case of probe vehicle *A* the piecewise linear regression is not statistically different from a single regime linear regression. Consequently, the one regime linear regression is chosen as the better model describing spatial and temporal movement of this probe vehicle and no point of inflection is defined. On the other hand, for probe vehicle *B* a piecewise linear regression model is statistically superior to the single regime model.

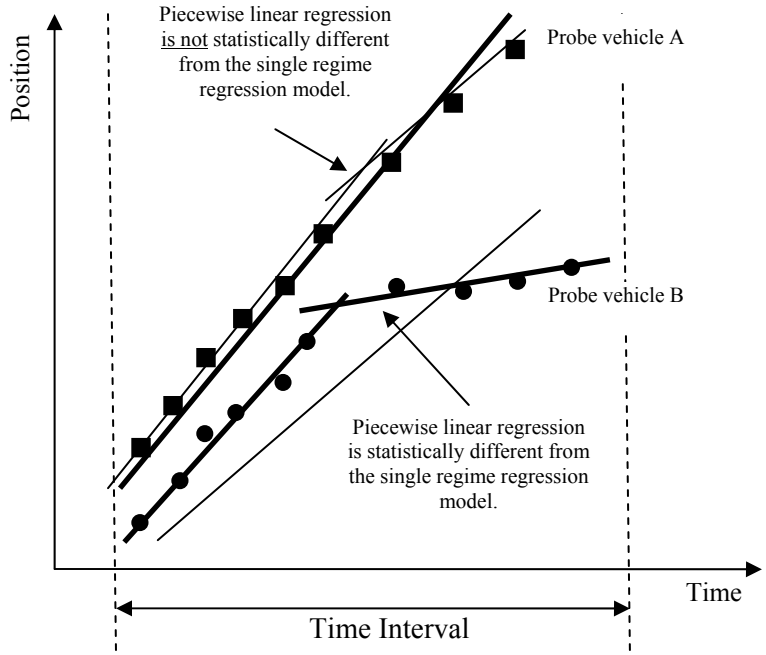


Figure 4.10: Statistically significant piecewise linear regression

The following hypothesis test can be applied to check whether or not a piecewise linear regression is statistically different from a single regime model (Vieth, 1989):

$$\begin{cases} H_0 : a_1 = a_2 \text{ and } b_1 = b_2 \\ H_1 : a_1 \neq a_2 \text{ or } b_1 \neq b_2 \end{cases} \tag{4.8}$$

The following statistic can be used to perform the above hypothesis test:

$$F = \frac{\left(RSSL - RSS\right)/3}{RSS/\left(\left\|\Lambda^p\right\| - 4\right)} \tag{4.9}$$

Where, *RSSL* in Equation (4.9) is the residual sum of squares of the single regime linear regression fitted to all point in Λ^p and *RSS* is the residual sum of squares associated with the piecewise linear regression which is defined by the objective function of Equation (4.7). The F-statistic defined by Equation (4.9) can be compared with an F-test table value with 3 and $\left\|\Lambda^p\right\| - 4$ degrees of freedom at a given level of confidence (e.g. 95%). If the F-statistic obtained using Equation (4.9) is larger than the critical value, the null hypothesis is rejected implying that the piecewise linear regression is

statistically different from the single regime linear regression and a point of inflection has not been found.

4.1.1.2 Data Filtering

At this stage, all inflection points associated with all probe vehicles in set *P* have been found. However, there are still two main challenges that need to be addressed. First, not every inflection point is part of a major shockwave. Second the points may be associated with multiple shockwaves. A data filtering procedure is proposed to address the first challenge in this section. A linear clustering algorithm is proposed to address the second challenge in the next section.

The main objective of the data filtering procedure is first to identify inflection points that are more likely to be part of a "major" shockwave and second to separate the resulting points into distinctive groups to facilitate the linear clustering algorithm which is described in the next section.

Our investigations of trajectories of vehicles in the real world show that in the case of a bottleneck or an incident, drivers do not decelerate and accelerate at a constant rate. Our observations show two stages of deceleration and two stages of acceleration as can be seen in Figure 4.11. In this figure, the black oval and the red circle show two stages of deceleration and the yellow circle and green oval show the two stages of acceleration for that particular probe vehicle. According to this observation, four different shockwave point groups can be identified among the detected inflection points (Figure 4.12). In order to expedite and facilitate the linear clustering procedure, the following rules are used to classify the points with similar characteristics into one group:

$$\begin{cases} \text{if } S_u^i > \Delta_u \text{ and } S_d^i < \Delta_u \Rightarrow g^i = 1 \\ \text{if } S_u^i > \Delta_c \text{ and } S_d^i < \Delta_c \Rightarrow g^i = 2 \\ \text{if } S_u^i < \Delta_c \text{ and } S_d^i > \Delta_c \Rightarrow g^i = 3 \\ \text{if } S_u^i < \Delta_u \text{ and } S_d^i > \Delta_u \Rightarrow g^i = 4 \end{cases}$$

Where,

i : denotes an inflection point, $i \in \Pi^p, p \in P$,
S_u^i : represents speed associated with the upstream of inflection point *i* (km/h),
S_u^i : represents speed associated with the downstream of inflection point *i* (km/h),
Δ_u : denotes the threshold speed associated with uncongested regime (km/h),
Δ_c : denotes the threshold speed associated with congested regime (km/h),
g^i : denotes the group associated with shockwave point *i* (km/h).

In Figure 4.12, group 1, group 2, group 3, and group 4 shockwave points are shown in black, red, yellow, and green respectively. The clustering algorithm, described in the next section, is applied to the four groups separately to improve computational efficiency and clustering accuracy.

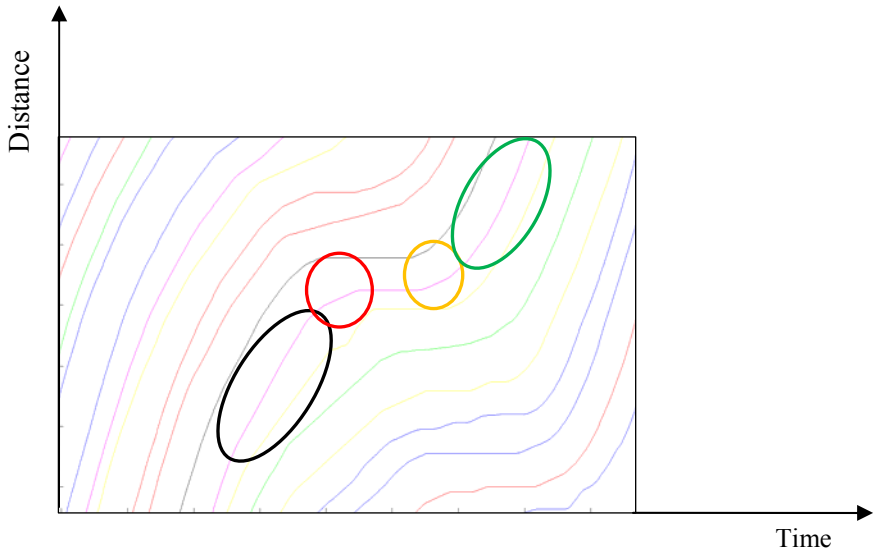


Figure 4.11: Acceleration and deceleration of real probe vehicles

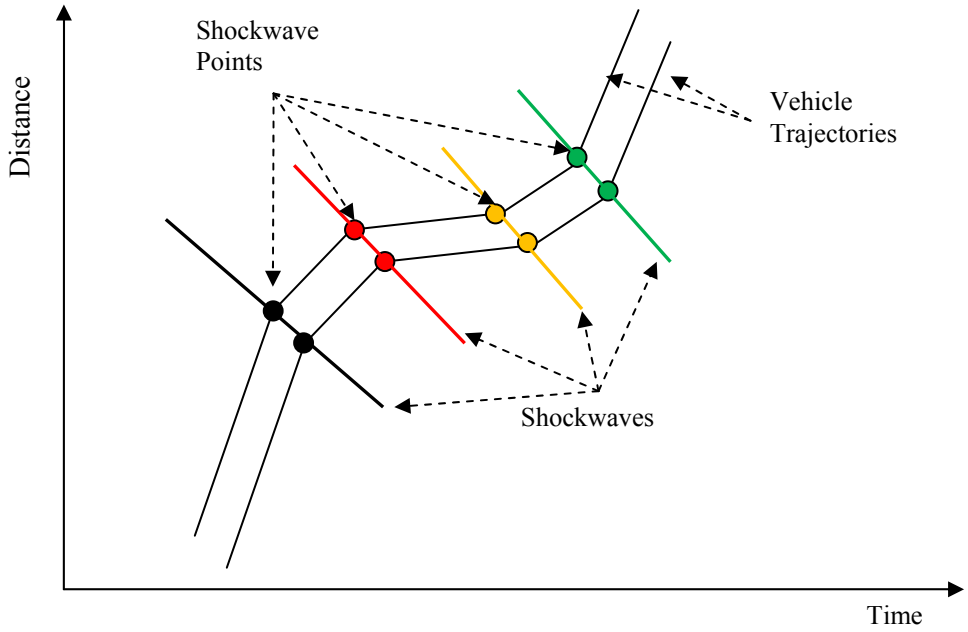


Figure 4.12: Four groups of shockwave points

4.1.1.3 Linear Clustering Algorithm

The data filtering procedure reduced the inflection points to the shockwave intersection points. Furthermore, the data has been categorized into four groups. The resulting shockwave intersection points need to be grouped into different shockwaves in order to find the characteristics of each shockwave. Under certain assumptions such as deterministic headway of vehicles and steady state of traffic, linear models can be used to model the propagation of shockwaves in a traffic stream. However, most clustering algorithms such as *k-means* algorithm (Hartigan, 1975) identify sparse and crowded datasets and are not appropriate for datasets with linear patterns (Van Aelst et. al., 2006). Consequently, a linear clustering algorithm is adopted to cluster shockwave intersection data points to distinctive shockwaves. A number of researchers have proposed different algorithms to cluster datasets with linear patterns (Van Aelst et. al., 2006; Murtaghi and Raftery, 1984, and Spath, 1982). In this research, a methodology proposed by Spath (1982) is used to cluster shockwave intersection points.

Assume that $\Omega_l = \left\{ \left(t_i, x_i\right) \middle| i = 1, \cdots, m_l \right\}$ is a set which contains all shockwave intersection points corresponding to shockwave group *l* where *l* is 1, 2, 3, and 4. The linear clustering problem can be formulated as follows:

$$Min \sum_{j=1}^{n} \sum_{i \in C_j} \left[x_i - \left(\alpha_{C_j} + \beta_{C_j} t_i \right) \right]^2 \tag{4.10}$$

s.t.

1) $C_j \subset \Omega_l \qquad \forall j = 1, \cdots, n$

2) $\left\| C_j \right\| \geq 3 \qquad \forall j = 1, \cdots, n$

3) $C_j \cap C_k = \varnothing \qquad \forall j, k = 1, \cdots, n \text{ and } j \neq k$

4) $\bigcup_{j=1}^{n} C_j = \Omega_l$

where, *n* denotes the number of clusters to be identified and C_j represents a cluster.

The objective function of the above mathematical problem is a minimization of the total sum of the squared errors of all clusters provided that $x = \alpha_{C_j} + \beta_{C_j} t_i$ is the regression line fitted to all data points in cluster C_j. Constraints 1 and 2 guarantee that data points of clusters are a subset of Ω_l and at

least 3 data points exist in each cluster respectively. Constraints 3 and 4 ensure that the clusters are mutually exclusive and collectively exhaustive.

The mathematical problem of Equation (4.10) is a nonlinear integer program which is intrinsically difficult to solve. However, this problem can be solved using the following algorithm proposed by Spath (1982):

Step 1: Choose an initial feasible solution.

Step 2: For each cluster C_j, perform a regression analysis to determine $\{a_{Cj}, b_{Cj}\}$.

Step 3: For a randomly selected $i \in C_j$ examine if there are clusters C_p with $p \neq j$ such that shifting i from C_j to C_p reduces the objective function. If so, choose C_r that maximizes the reduction. Then, set $C_j = C_j - \{i\}, C_r = C \cup \{i\}$. Continue this step until all data points are visited.

Step 4: Repeat Step 2 for a given number of times or until no reduction in objective function is achieved.

Set $\left\{\beta_{C_j} \mid j = 1, \cdots, n\right\}$ contains the estimated speed of the shockwaves associated with each cluster.

It should be noted that this clustering algorithm has two limitations: first, the performance of the linear clustering algorithm highly depends on the initial feasible solution. One method that can be used to create a "good" initial solution is to recognize the fact that shockwave points which are close together spatially and are far from each other temporally cannot be part of the same shockwave.

Second, the number of clusters is assumed to be known. However, in reality this is not the case. One approach to addressing this limitation is to apply the clustering algorithm for different values of *n* and then select the solution that maximizes the total marginal benefit. In other words, the optimal number of clusters, *n*, is selected according to the fact that increasing the number of clusters from *n*-1 to *n* results in a maximum reduction in the objective function defined in Equation (4.10).

4.2 Travel Time Prediction/Estimation Using Loop Detectors

4.2.1 Average Speed Algorithm (MTO Method)

The simplest form of loop detector based travel time estimation algorithms makes use of measured average vehicle speeds at a spot location and assumes vehicles travel at this speed over a fixed segment of the roadway (Lindveld *et al*., 2000). The length of this fixed segment of the roadway is often assumed to be half the distance between two consecutive detector stations, but can be defined as

some other length. This scheme can be illustrated using the notation in Figure 4.13 and the time-space diagram in Figure 4.14.

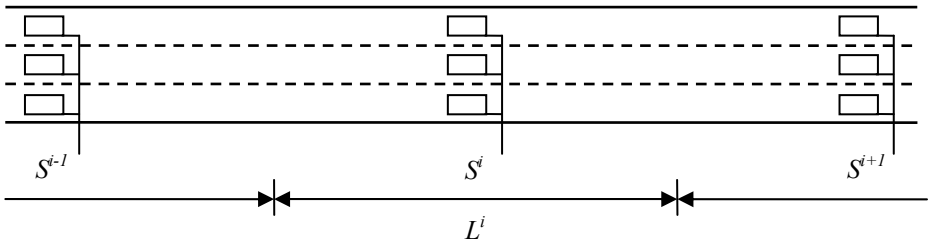


Figure 4.13: Roadway schematic for spot-speed algorithms

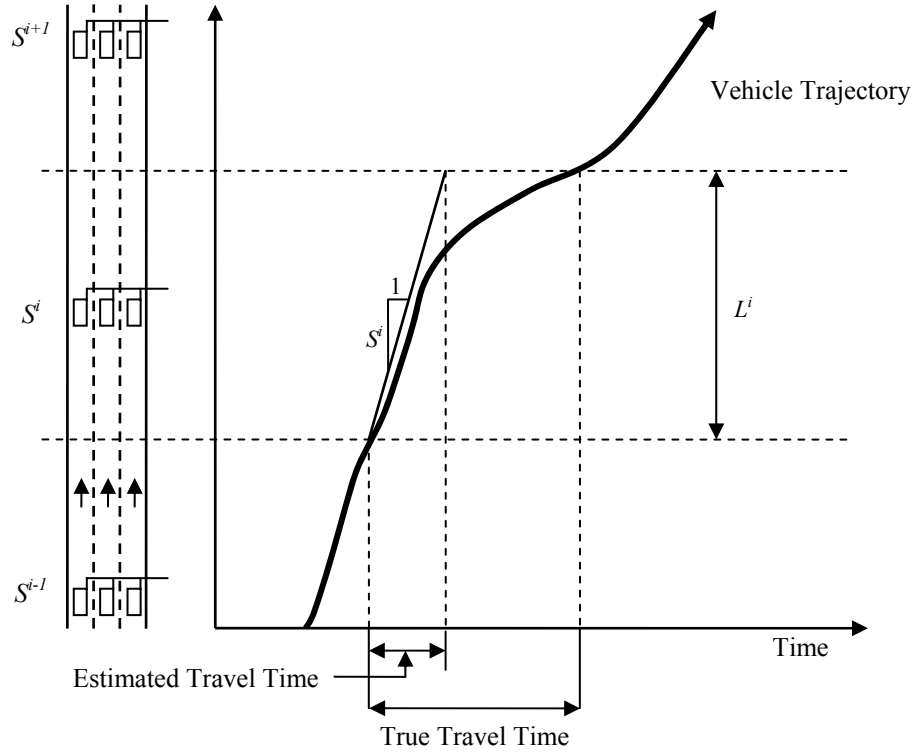


Figure 4.14: Space - time diagram

As illustrated in Figure 4.14, the average speed S^i represents the speeds of vehicles as they pass a specific location (i.e. the loop detector). Spot speed based methods assume that the vehicles travel at this speed over the entire section length (i.e. L^i). This assumption is generally acceptable when the road section is uncongested or short. However, as illustrated in Figure 4.14, when a portion of the section experiences congestion, then this assumption is no longer valid as the vehicle's speed changes

significantly over the length of section *i* and consequently the estimated travel time will not be accurate. If the section length (L^i) is short, then errors between estimated and actual travel time will be acceptably small. However, installing loop detectors more closely spaced than approximately 500 – 600m is generally not economically feasible.

It is important to note that, in general, point-speed algorithms tend to underestimate travel times when congestion is forming. The magnitude of the underestimation is a function of the severity of the congestion, the length of the road section, the portion of the segment congested, and the rate at which the queue is growing or dissipating.

The following equation can be used to predict travel time of a route which consists of *n* loop detectors:

$$\tau_M(t) = \sum_{i=1}^{n-1} \frac{L^{i,i+1}}{\left(S_i(t) + S_{i+1}(t)\right)/2} \tag{4.11}$$

Where,

$L^{i,i+1}$: Length of the highway section between loop *i* and loop *i*+1,

$S_i(t)$: Average speed of loop detector *i* during the subject time interval,

$\tau_M(t)$: Average travel time of the route for the subject time interval.

The assumption behind Equation (4.11) is that the route is bounded by two loop detectors.

MTO uses this methodology to estimate travel time of vehicles on 400 series highways and the QEW in Mississauga. However, MTO uses a number of adjustment factors to compensate for the shortcomings of this algorithm when congestion is forming or dissipating.

In order to determine the adjustment factors, a vehicle equipped with a GPS device travels the subject section of highway at different times of various days. This is performed to observe different states of traffic along the subject section of highway. Also loop detector data associated with the same section at the same time intervals are recorded. Furthermore, the subject section of the highway is divided into different congestion zones. The state of traffic on each congestion zone is determined in terms of moving well, moving slowly, or moving very slowly for every 20 second time period for which loop detector data are available. The following rules are used to determine the linguistic state of traffic on each congestion zone:

$$\begin{cases} \overline{S} \geq 75\text{km/h} \Rightarrow \text{Moving Well} \\ \overline{S} < 75\,\text{km/h or } \overline{S} \geq 40\,\text{km/h} \Rightarrow \text{Moving Slowly} \\ \overline{S} < 40\,\text{km/h} \Rightarrow \text{Moving Very Slowly} \end{cases}$$

Where $\overline{S}$ represents the average speed over a congestion zone calculated using the average speed algorithm described above. For each combination of linguistic traffic states, the ratio of travel time obtained from probe vehicles, τ_p, and travel time obtained from loop detectors using Equation (4.11) can be calculated. Finally, this ratio (adjustment factor) is applied in the future to travel time obtained from loop detectors whenever the same linguistic combination of traffic states is observed.

4.2.2 Trajectory Method

The trajectory method can be used in order to estimate travel time. This method is a modified version of the average speed algorithm but cannot be applied before all vehicles which entered the route during a time interval, complete the route. Equation (4.11) computes travel time as if the vehicles which passed over the first loop detector would pass the last loop detector at the same time. Obviously this assumption is not accurate. The trajectory method improves the accuracy of average speed algorithm by building the trajectory of an average vehicle. Then the travel time of the average vehicle (which is equal to the average travel time of vehicles which entered the route during the same 20 second time interval) is calculated by projecting the built trajectory on the time axis.

Figure 4.15 illustrates a route which includes three loop detectors and a time space diagram associated with the three loop detectors. Average speed of the first section can be calculated using the average speed of loop detector 1 and loop detector 2 associated with the time which the average vehicle entered the 20 second time interval. For the second section (i.e. the section between loop detector 2 and 3) the average speed is calculated using average speed of loop detector 2 and loop detector 3 associated with the time interval which the vehicles entered the route plus travel time of the first section. If the route consists of more than 3 loop detectors then the same approach can be used.

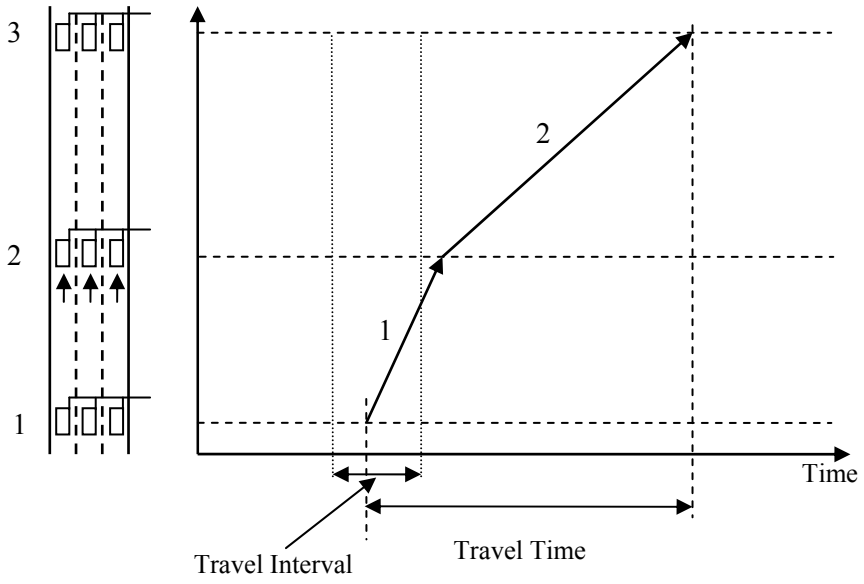


Figure 4.15: The trajectory method

The trajectory method explained above is an iterative procedure which can be shown by Equation (4.12) below:

$$\begin{cases} \tau_T^{1,2}(t) = \dfrac{L^{1,2}}{\left(S_1(t) + S_2(t)\right)/2} \\ \tau_T^{2,3}(t) = \dfrac{L^{2,3}}{\left(S_2\left(t+\tau_T^{1,2}\right) + S_3\left(t+\tau_T^{1,2}\right)\right)/2} \\ \vdots \\ \tau_T^{n-1,n}(t) = \dfrac{L^{n-1,n}}{\left(S_{n-1}\left(t+\sum_{i=1}^{n-2}\tau_T^{i,i+1}(t)\right) + S_n\left(t+\sum_{i=1}^{n-2}\tau_T^{i,i+1}(t)\right)\right)/2} \\ \tau_T(t) = \sum_{i=1}^{n-1}\tau_T^{i,i+1}(t) \end{cases} \quad (4.12)$$

Where,

$\tau_T^{i,i+1}(t)$: denotes travel time of the road section between loop detector *i* and loop detector *i*+1 when it is desired to estimate travel time for vehicles which entered the road section during time interval *t*,

$\tau_T(t)$: Travel time of the route associated with time interval *t* using the trajectory method.

It is expected that this method can provide very accurate travel time estimation even for long route. However, the major shortcoming of this method is that it can only estimate travel time for previous time intervals.

Chapter 5
Data Collection

Two Ontario freeways were selected as the study area. The first study area is a section of a suburban freeway in the Region of Waterloo, Ontario (Regional Highway 85) which is not covered by Freeway Traffic Management Systems (FTMS) and the second study area is a section of urban freeway (Highway 401 in the Greater Toronto Area) which is fully instrumented.

5.1 HWY 85: Non-instrumented Freeway

Highway 85 in the Regional Municipality of Waterloo is an example of a non-instrumented freeway in Ontario. A section of this freeway was chosen as the study area in this project. Highway 85 is a main roadway that connects Highway 401 to Waterloo through Regional Highway 7. The northern boundary of this freeway is Sawmill Rd. in St. Jacob (north of Waterloo) and southern boundary of Highway 85 is Highway 7 in Kitchener. Figure 5.1illustrates the study area which is bounded by Northfield Dr. in the north and Frederick St. in the south. The section is 8.370 km long and experiences recurrent congestion in both directions during AM and PM peak periods. The posted speed on the section is 90 km/h implying a free flow travel time of 5.58 minutes (actual free flow speeds are frequently higher than the posted speed limits). The data collection was performed on Thursday September 18, 2008 from 15:35 to 17:50. The rationale behind this time interval is that we were interested in collecting data before the start of congestion and after dissipation of congestion. The average temperature during the data collection was 19 Celsius and the weather condition was mainly clear (Environment Canada, 2008).

The data collection effort along Highway 85 included two tasks: (1) obtain the “Ground Truth” travel time and (2) probe vehicles runs.

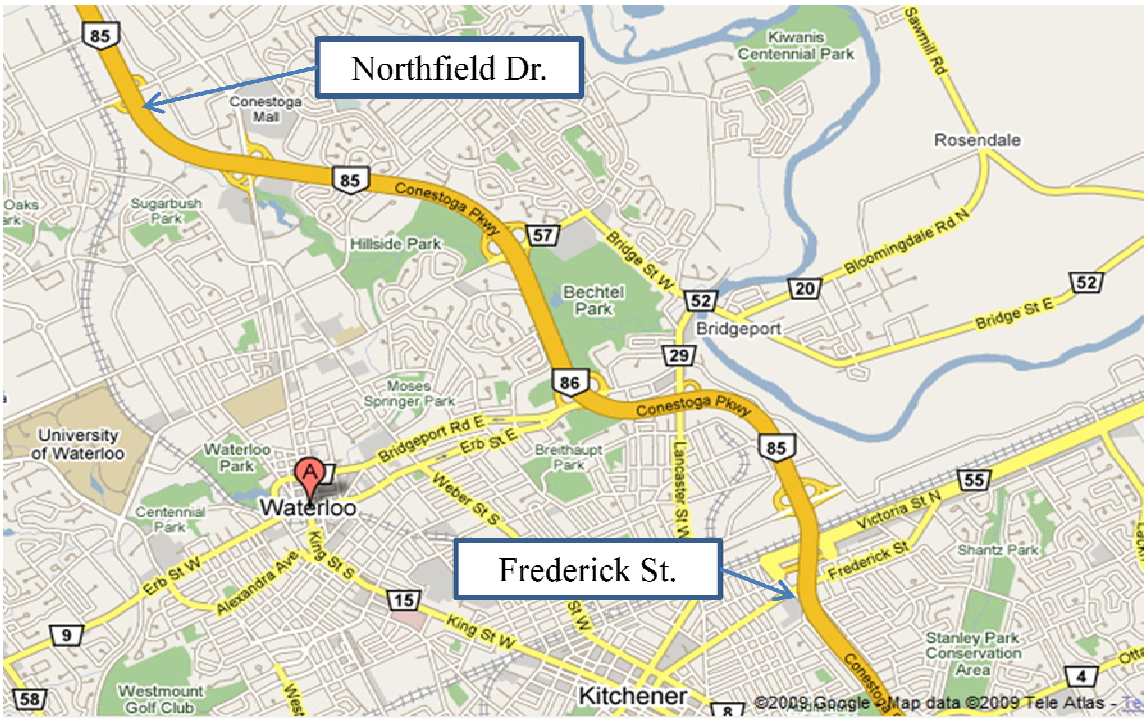


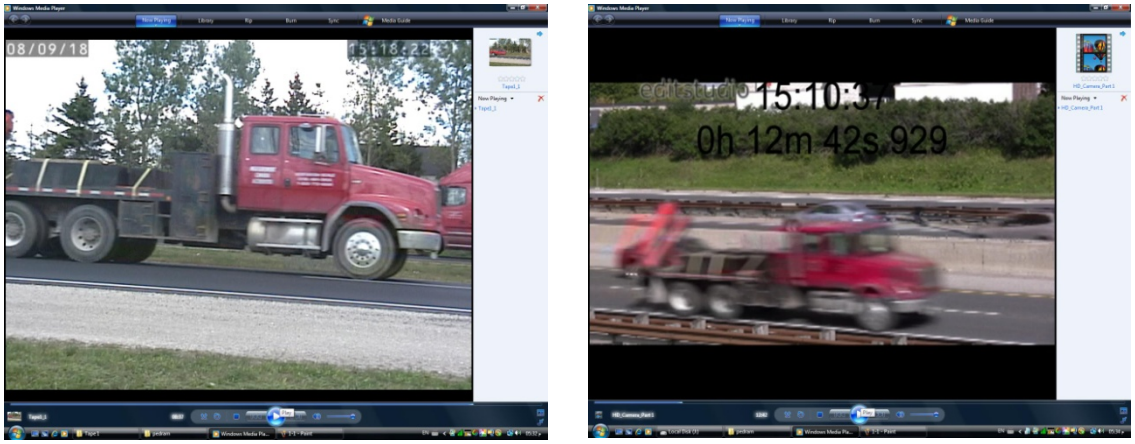
Figure 5.1: Highway 85 as study area in regional municipality of Waterloo (Source: www.maps.google.com)

5.1.1 Ground Truth Travel Time

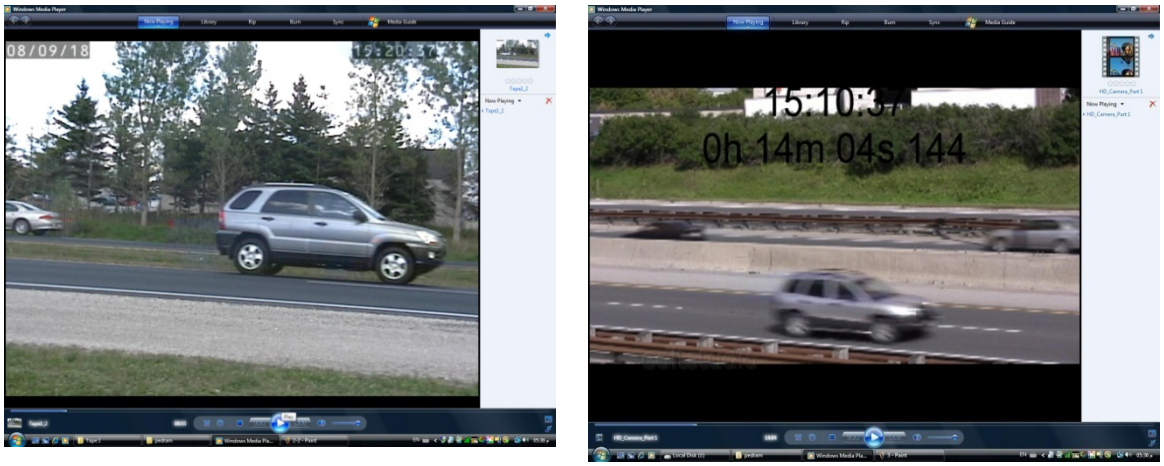
A method similar to licence plate number matching was used to capture true travel time of vehicles in the study area. Two video cameras were deployed, one at the northern boundary and the other at the southern boundary of the study area to film both directions of the freeway section. The cameras were deployed on the west side of the freeway outside of the MTO Right of Way. Each camera was monitored by an individual during the data collection time period. The internal clocks of the cameras were also synchronized to minimize the systematic error in travel time estimation. The videos were post-processed to obtain true travel time of vehicles. The first step was to overlay the timestamp on the video. It should be noted that timestamp of a video in most digital cameras are attached to the data but is not readily available to overlay on the video. Consequently, a commercial-off- the- shelf software was used to overlay the timestamp on the videos.

The next step to obtain travel time of individual vehicles is to match vehicles in two screen shots at upstream and downstream videos. In order to match vehicles, two computers were used simultaneously and a student manually found the matching vehicles in both screens. Figure 5.2 shows

two examples of matched vehicles. Travel time of each individual matched vehicle is simply the difference between the two timestamps of the screen shots of the vehicle. There are a number of off-ramps between the upstream and downstream cameras. Consequently, not every vehicle observed at the upstream video can be found in the downstream video. Also, it is likely that some vehicles which passed in front of both upstream and downstream cameras were not matched due to human error. In total, 352 vehicles were matched for the southbound direction. The results are presented later in this document.



a) A truck matched at the videos associated with the northern camera and southern camera (left image and right image respectively).



b) An SUV matched at the videos associated with the northern camera and southern camera (left image and right image respectively).

Figure 5.2. Matching process of vehicles to measure travel time

5.1.2 Probe Vehicles Runs

Five probe vehicles equipped with data logger GPS devices were continuously travelling on the study area. The data loggers are able to record time, position, and speed of the vehicle at a configurable rate. The rate (polling interval between two consecutive records) was chosen as 1 second in this project. The drivers of the probe vehicles were instructed to drive with the flow of traffic and try to be like an "average" vehicle.

Probe vehicles began the data collection effort near the north end of the study section and initiated their first trip at 3 minute headways. Each probe vehicle completed 6 runs per each direction.

The probe vehicle drivers were instructed to use King St. North interchange (north of Northfield Dr.) and Ottawa St. (south of Frederick Dr.) to turn around. This ensured that all vehicles covered the section between the two cameras. Consequently, in the post-processing of the GPS data, a methodology was designed to (1) identify and delete the data points north of the northern camera and south of the southern camera; and (2) identify the GPS data points associated with each run. Data points which were outside of the study area were visually identified in the Arcview GIS environment. Also, a spreadsheet was developed to separate data points associated with each run. The program assigns a four digit ID in the form of "*xyzz*" to the data points associated with each individual run. In this format *x* identifies the driver, *y* denotes the direction (where 1 is southbound and 2 is northbound), and *zz* is sequential. Therefore, 30 probe vehicle runs per each direction were collected.

5.1.3 Overview of the Data

The southbound direction of Highway 85 becomes more congested than the northbound direction during the PM peak period. Consequently, video matching was conducted on the southbound direction only. In total 352 vehicles could be matched at the upstream and downstream cameras which amounts to almost 16% of the total volume observed upstream of the upstream camera location.

Figure 5.3 illustrates the disaggregate travel times obtained from GPS equipped probe vehicles and travel time extracted from video matching. As can be seen in this figure, travel times obtained from probe vehicles closely match the travel time obtained from the video. Although some bunching of probe vehicles can be observed, probe vehicle observations have been spread across the data collection time period.

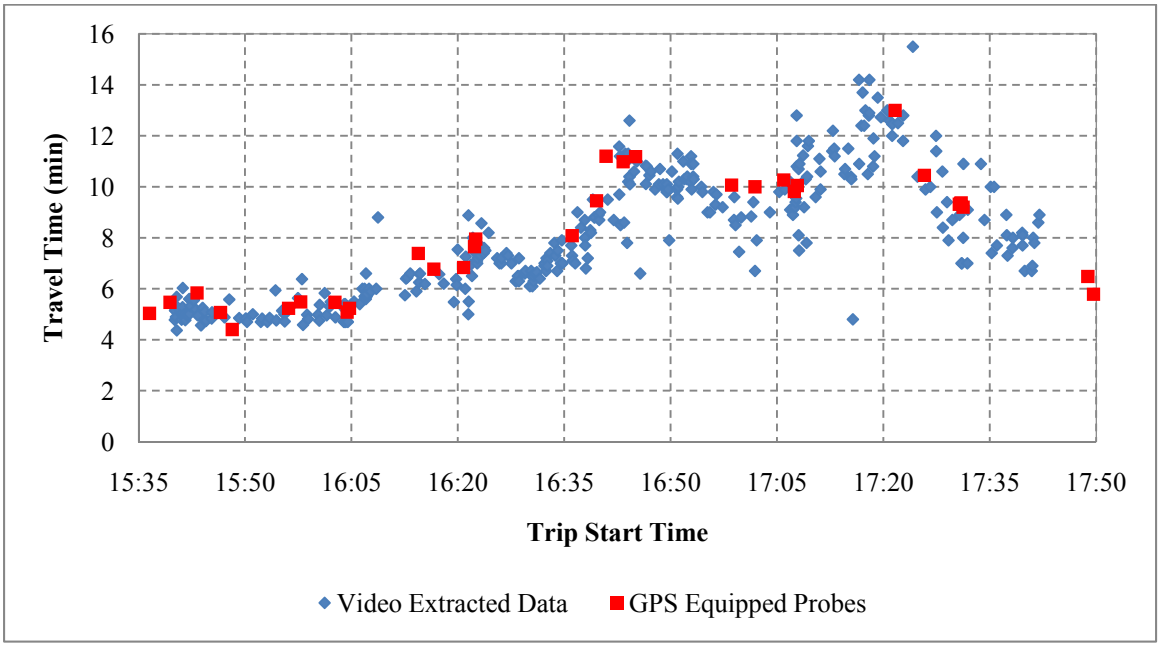


Figure 5.3. Disaggregate comparison of travel times of probe vehicles and video matching

In order to compare average travel time obtained from probe vehicles with average travel time obtained from video extraction during each 15 minute time interval, the following hypothesis test was conducted:

$$\begin{cases} H_0 : \overline{T}_{video} - \overline{T}_{probes} = 0 \\ H_1 : \overline{T}_{video} - \overline{T}_{probes} \neq 0 \end{cases} \quad (5.1)$$

Where,

$\overline{T}_{video}$: denotes average of all travel times obtained from video image matching during a 15 minute time interval,

$\overline{T}_{probes}$: denotes average travel times of probe vehicles that entered the road section during a 15 minute time interval.

Table 5.1 shows the result of the hypothesis test for 9 consecutive 15 minute time intervals starting at 15:35. As can be seen in this table the null hypothesis is rejected only for the last interval. For all other time intervals, there is insufficient evidence to conclude that the travel times obtained from video images and probe vehicles are different. It should be noted that, as can be seen in Figure 5.3, the video data were not available after sometime in the middle of the 9th time interval. Also, there are only two probe vehicle observations available coincidently at the very end of this time interval. Consequently, it is not unexpected that the null hypothesis for this time interval was rejected. It

should be also noted that the data associated with this time interval were removed from the data base and not used in the further analyses.

Table 5.1: Result of the test of hypothesis

Interval No	Video Extraction			Probe Vehicles			t-statistic	t-critical	Reject H_0
	No Observation	Average (min)	Variance (min^2)	No Observation	Average (min)	Variance (min^2)			
1	35	5.12	0.13	5	5.16	0.29	-0.230	2.024	No
2	33	5.07	0.18	5	5.30	0.03	-1.195	2.028	No
3	30	6.19	0.45	2	7.07	0.19	-1.832	2.042	No
4	64	6.99	0.42	3	7.48	0.33	-1.263	1.997	No
5	54	9.34	2.02	5	10.18	1.91	-1.265	2.002	No
6	37	9.66	1.01	2	10.03	0.002	-0.511	2.026	No
7	53	10.79	3.11	3	10.04	0.05	0.730	2.005	No
8	30	10.17	4.55	5	10.26	2.59	-0.091	2.034	No
9	17	7.95	0.72	4	5.59	0.63	5.069	2.093	Yes

Figure 5.4 can also be used to compare the travel times obtained from GPS equipped probe vehicles and the video matching technique. In this figure, the left vertical axis shows the travel time and the right vertical axis shows the difference between the travel time obtained from the GPS equipped vehicles and the travel time obtained from the video matching technique. The maximum difference is around 15% and on average the difference is 8.7%. Given the free flow travel time of the study section (i.e. 5.58 min) this figure shows that the travel time associated with the most congested state is twice as much as the free flow travel time. As explained earlier, the error associated with the last time interval was not calculated but the data points were shown to illustrate the fact that the congestion started to dissipate after 17:20.

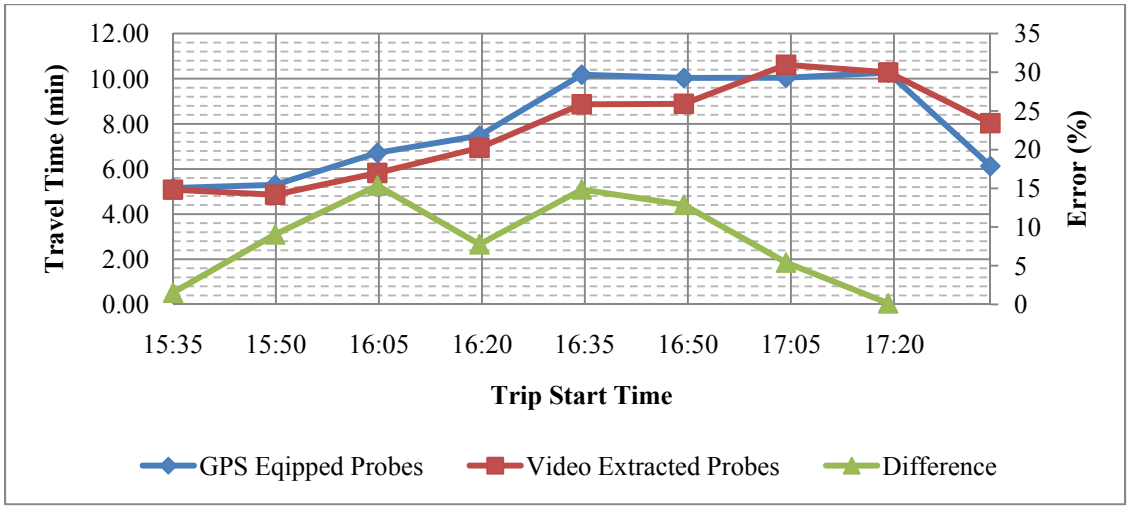


Figure 5.4. Comparison of Average Travel Time Estimated Using Data from GPS Equipped Vehicles and Video Extraction

5.2 HWY 401: Instrumented Freeway

A 21.35 km section of HWY 401 between Highway 400 and Kennedy Rd. was chosen for this study (Figure 5.5). This section of the 401 is covered by full FTMS infrastructure. This section experiences recurrent congestion during the AM and PM peak periods. The posted speed on Highway 401 is 100 km/h implying a free flow travel time of 12.81 min for each direction. The data collection was performed on Wednesday May 13, 2009 from 14:30 to 19:30. Similar to HWY 85, the rationale behind this time interval is to ensure that data can be collected before the start of congestion and after dissipation of congestion.

The average temperature during the data collection was 18 Celsius and the weather condition was mostly cloudy (Environment Canada, 2010). The main task during the data collection was to obtain GPS data associated with probe vehicles equipped with GPS devices. Data associated with loop detectors on the study area were also obtained from MTO. The following subsections summarize each of the two data types.

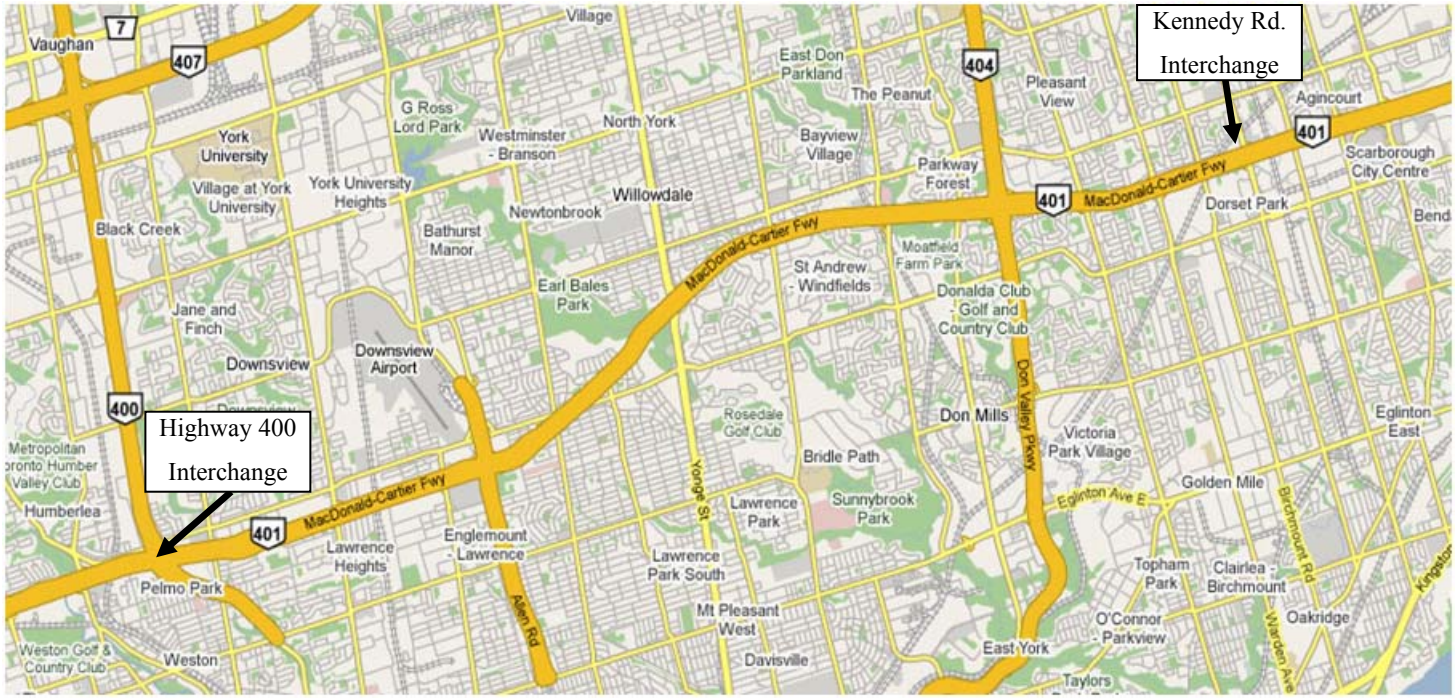


Figure 5.5: The Highway 401 study area in the Greater Toronto Area (Source: www.maps.google.com)

5.2.1 GPS Probe Vehicles Runs

Eight probe vehicles equipped with the same GPS data loggers as were used for the data collection effort on HWY 85 were continuously travelling along the study section over the study period. Each data logger was able to record time, position, and speed of the vehicle every second. Probe vehicles began the data collection effort near the west end of the study section and initiated their first trip at 5 minute headways. The probe vehicle drivers were instructed to drive with the flow of traffic and try to be like an "average" vehicle. They were also instructed to use Weston Rd. interchange (west of Highway 400) and McCowan Rd (East of Kennedy Rd) to turn around. Furthermore, in order to maximize the number of travel time observations given availability of limited number of probe vehicles, the drivers only drove on collector lanes. The same approach which was used for HWY85 was used to post-process the GPS data in order to identify the trajectory of each GPS equipped probe vehicle. Figure 5.6 shows travel time of probe vehicles on eastbound and westbound of HWY 401. Each point in this figure corresponds to one run of a probe vehicle which shows that each probe vehicle traveled on average 5 times each direction. Also, this figure shows that the eastbound direction is more congested than the westbound direction. A comparison between the travel time values at the beginning and end of the data collection effort with free flow travel time (12.36 min) shows that the forming and dissipating of congestion during that particular PM peak period were successfully captured.

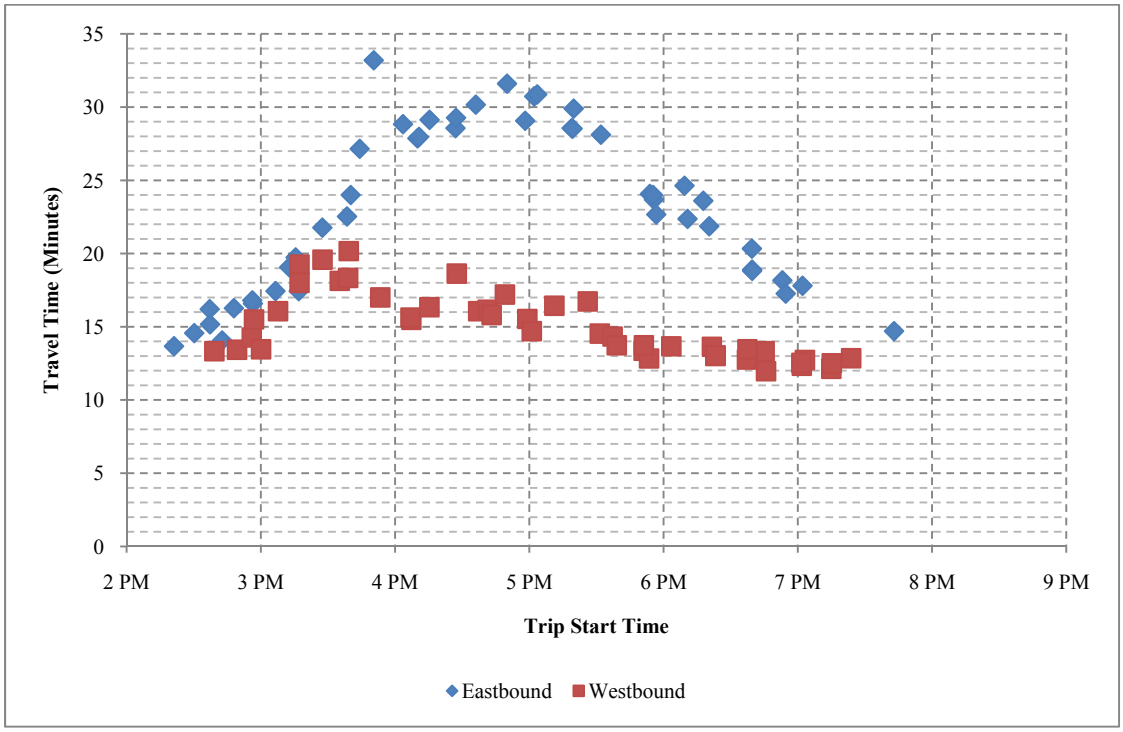


Figure 5.6: Travel time of probe vehicles equipped with GPS data loggers

5.2.2 Loop Detector Data

The Ontario Ministry of Transportation has developed the COMPASS system to monitor and manage 400 series highways in Ontario. The COMPASS system primarily relies on CCTV cameras and loop detector sensors which provide volume, speed, and occupancy. The data are transferred to the MTO traffic management centre where further analyses are performed in order to make various decisions. Loop detector data associated with the study section and for the duration of our data collection were obtained from MTO. The data include 20 second aggregated volume, speed, and occupancy for each lane of the collector and express lanes. However, only the data associated with the collector lanes were analyzed in this study because all of the probe vehicles drove on the collector lanes.

Not all of the loop detectors in the study area were operational during our data collection effort. Table 5.2 shows the detector ID and location of operational loop detectors in the eastbound and westbound directions of the collector lanes of HWY401 in the study area. According to this table, 6 loop detector stations (each of which consists of a double loop detector per each lane) are available in each direction. Consequently, the operational loop detectors have an average spacing of 4.27 km.

Table 5.2: Operational loop detectors within study section

Eastbound Collector Lanes		Westbound Collector Lanes	
Detector ID	Location	Detector ID	Location
401DW0060DEC	East of Weston	401DE0300DWC	At Midland
401DW0020DEC	East of Jane	401DE0210DWC	West of Hwy 404/DVP
401DE0030DEC	East of Keele	401DE0140DWC	East of Yonge
401DE0140DEC	East of Yonge	401DE0030DWC	East of Dufferin
401DE0170DEC	East of Bayview	401DW0020DWC	East of Jane
401DE0280DEC	East of Birchmount	401DW0060DWC	East of Weston

Data obtained from the loop detectors shown in Table 5.2 were processed and 15 minute average travel times were calculated using both the average speed (MTO) method and the trajectory method. Figure 5.7 and Figure 5.8 illustrate the travel times obtained from both methods for eastbound and westbound directions respectively.

Several observations can be made from these two figures:

1. The MTO method and the trajectory method closely match each other at the beginning and end of the data collection effort when traffic is approximately at free flow conditions. This observation is more evident in Figure 5.8 where the duration of this condition is longer at both ends of the data collection.

2. It appears that the travel times obtained from both methods are relatively similar at the peaks of the curves where traffic state is congested. This argument is applicable to both directions of the study section.

3. The other important observation is that the two curves in Figure 5.7 and Figure 5.8 are most different when congestion is forming and dissipating. These four instances are marked by green ovals in these two figures.

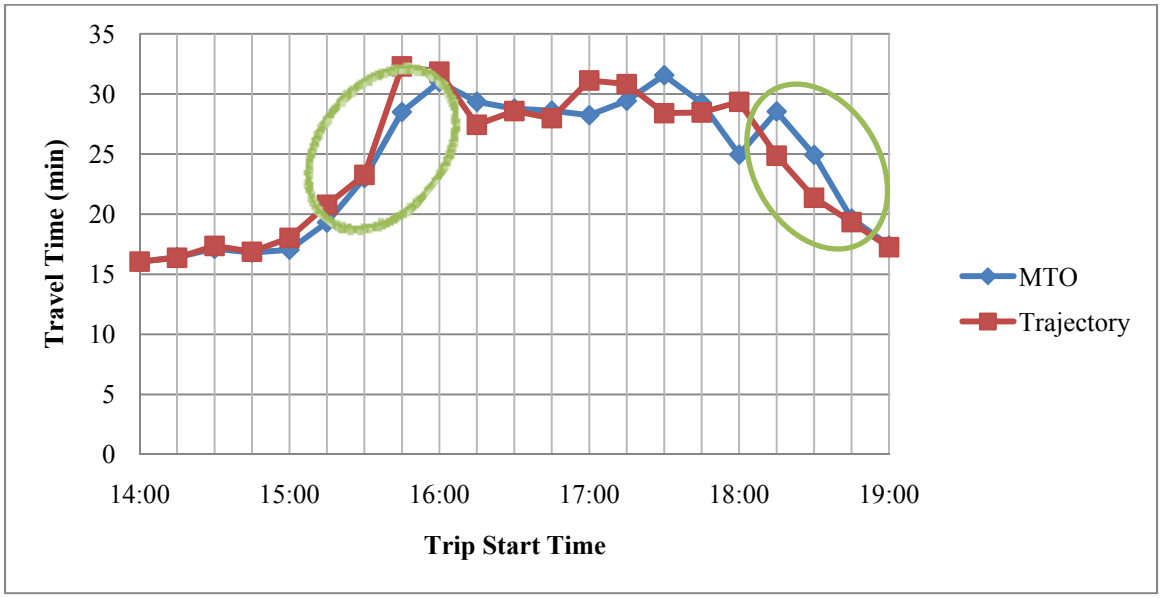


Figure 5.7: Comparison of the 15 min aggregated travel times estimated from loop detector data for EB HWY401

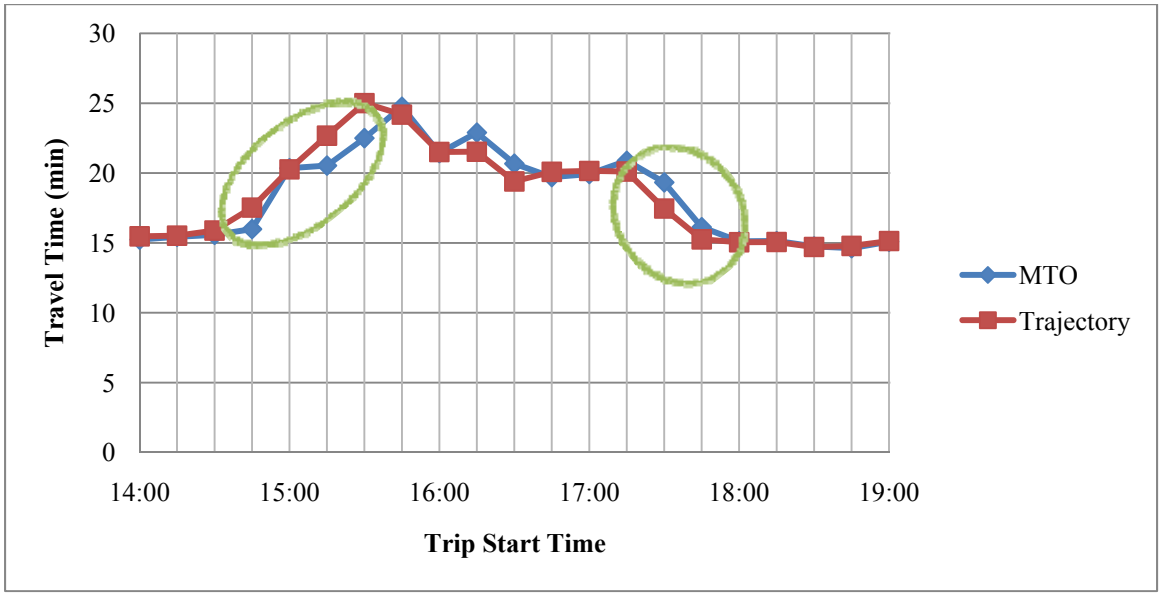


Figure 5.8: Comparison of the 15 min aggregated travel time estimated from loop detector data for WB HWY401

Figure 5.9 and Figure 5.10 compare the 15 minute aggregated travel time obtained from the MTO method and the trajectory method with the 15 minute aggregated travel time obtained from GPS equipped probe vehicles for the eastbound and westbound directions respectively. As can be seen in

Figure 5.9, travel times obtained from the trajectory method are more similar to travel times obtained from GPS equipped probe vehicles than are the travel times obtained from the MTO method. This observation can also be numerically verified. The Root Mean Square Error (RMSE) between travel times obtained from the trajectory method and obtained from GPS equipped vehicles is 2.38 minutes whereas the RMSE between travel times obtained from the MTO method and GPS equipped probe vehicles is 2.94 minutes.

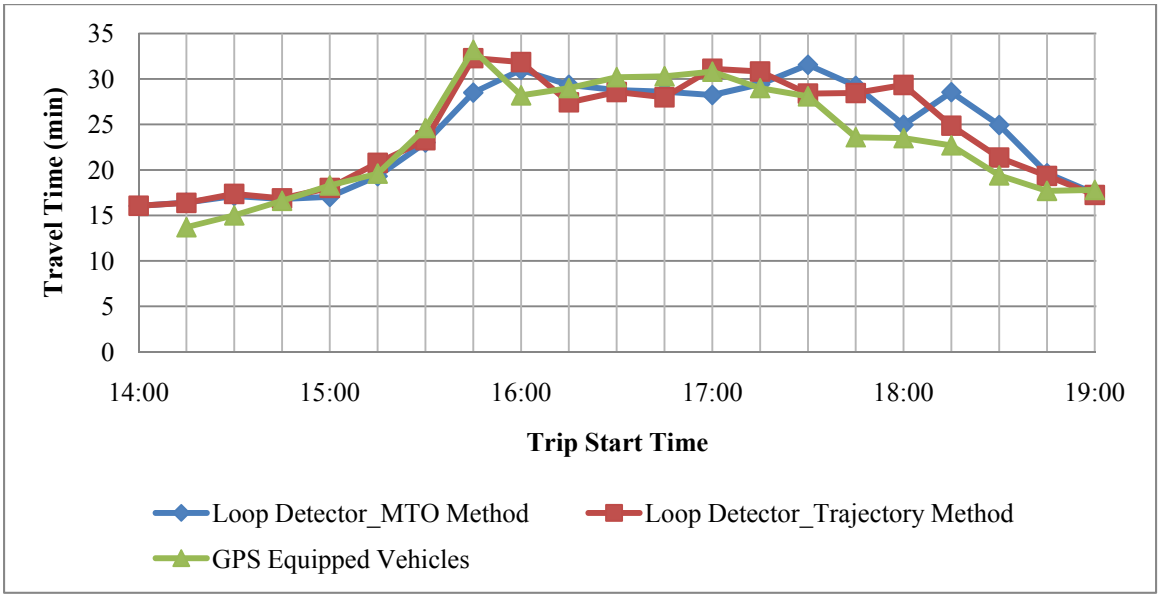


Figure 5.9: Comparison of travel time obtained from loop detector with travel time obtained from GPS equipped probes for EB HWY 401

As can be seen in Figure 5.10, the travel times obtained from the GPS equipped probes are consistently shorter than the travel times obtained via loop detector data. One of the reasons which can justify this observation is that the drivers of probe vehicle drove the westbound direction faster than an "average" vehicle despite the instructions. This was not possible for the eastbound due to heavier traffic congestion. The other reason might be poor performance of the loop detector based methods for the westbound direction of Highway 401 particularly in the light of the fact that he spacing of the operational loop detectors was rather excessive (4.27 km). However, the RMSE associated with the westbound is 3.92 min and 3.97 min for the trajectory method and the MTO method respectively.

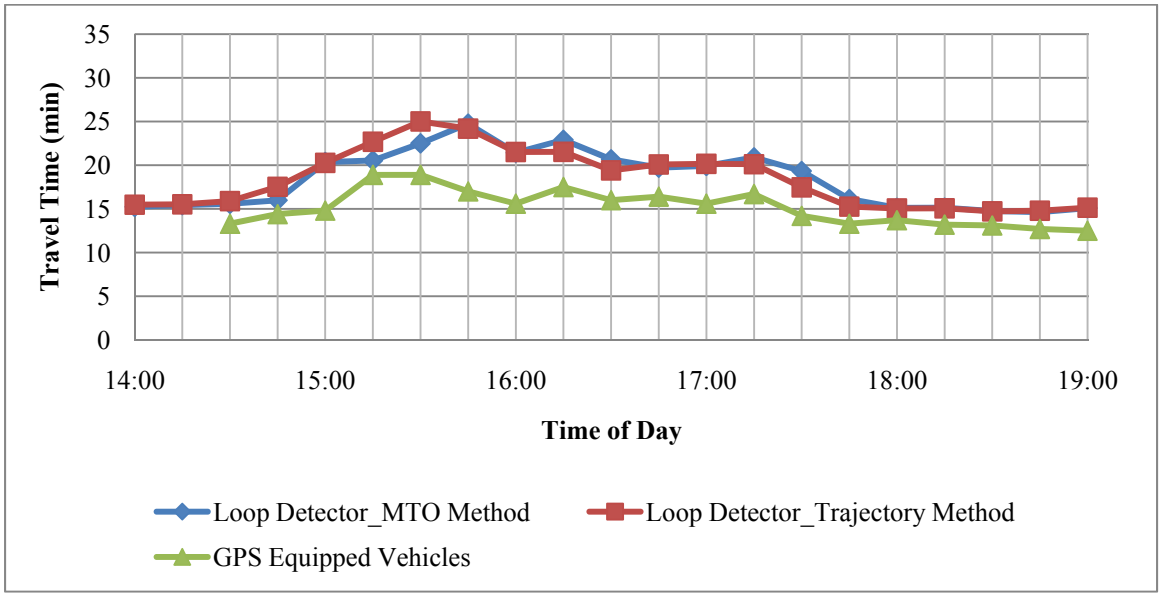


Figure 5.10: Comparison of Travel Time Obtained from Loop Detector with Travel Time Obtained from GPS Equipped probes for WB HWY 401

Chapter 6
Results

The objective of this chapter is to evaluate the performance of the proposed methodology. In order to achieve this objective the proposed methodology was applied to the data associated with one simulated intersection and 3 field datasets namely (1) NGSIM US-101; (2) HWY 85; and (3) HWY 401 were used. In this chapter, root mean square error (RMSE), Mean Absolute Percent Error (MAPE), and maximum absolute percent error (E_{max}) are used as measures of effectiveness for the propose travel time prediction model. The RMSE, MAPE, and E_{max} are calculated based on the following equations:

$$RMSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\hat{\tau}_k - \tau_k\right)^2} \tag{6.1}$$

$$MAPE = 100 \times \frac{1}{K}\sum_{k=1}^{K}\left(\frac{\left|\hat{\tau}_k - \tau_k\right|}{\tau_k}\right) \tag{6.2}$$

$$E_{\max} = \max\left\{abs\left(\hat{\tau}_k - \tau_k\right), \forall k = 1 \cdots K\right\} \tag{6.3}$$

where, $\hat{\tau}_k$ denotes predicted travel time at time interval k, τ_k represents reference travel time at time interval k, and K is total number of time intervals.

6.1 Application 1: Signalized Intersection

A single exclusive through lane of a signalized intersection (Figure 6.1) was simulated using the INTEGRATION simulation model (Van Aerde, M. & Assoc., Ltd, 2002a, 2002b). The signal was controlled by a fixed time two phase signal timing plan with a cycle length of 100 seconds; phase 1 and phase 2 have an effective green of 45 and 51 seconds respectively; total lost time of 4 seconds and offset of 52 seconds were also chosen. The approach under study had a saturation flow rate of 1900 vehicles per hour per lane. Traffic demand was assumed to be constant at 700 vph and consisted of only passenger cars. Furthermore, vehicles were generated with uniform headways. The approach link was 1 km in length and had a free flow speed of 60 km/h; a speed at capacity of 40 km/h and jam density of 125 veh/km/lane.

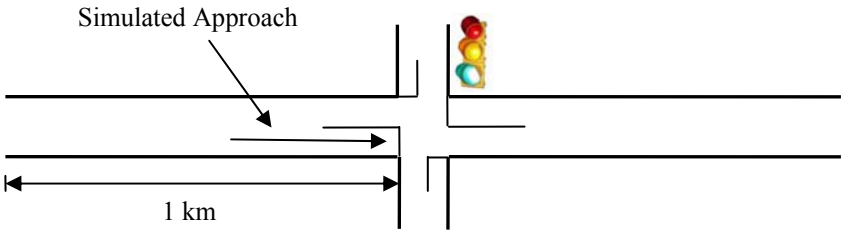


Figure 6.1: Simulated signalized intersection

The INTEGRATION simulation model uses a single regime four parameter macroscopic traffic flow model proposed by Van Aerde (Rakha and Crowther, 2002; Van Aerde, 1995; Van Aerde and Rakha, 1995). Van Aerde's model is defined by the following set of equations:

$$D = \frac{1}{C_1 + \frac{C_2}{\left(S_f - S\right)} + C_3 S} \tag{6.4}$$

$$C_1 = kC_2 \tag{6.5}$$

$$C_2 = \frac{1}{D_j\left(k + \frac{1}{S_f}\right)} \tag{6.6}$$

$$k = \frac{2S_c - S_f}{\left(S_f - S_c\right)^2} \tag{6.7}$$

$$C_3 = \frac{1}{S_c}\left(-C_1 + \frac{S_c}{V_c} - \frac{C_2}{S_f - S_c}\right) \tag{6.8}$$

where,

D = Density (veh/h/lane)

S = Speed (km/h)

S_f = Free flow speed (km/h)

S_c = Speed at capacity (km/h)

V_c = Capacity volume or saturation flow rate(vph)

D_j = Jam Density (veh/km/lane).

Using equations (6.4) through (6.8), the fundamental equation of traffic ($V=DS$), and the equation for calculating propagation speed of shockwaves, Equation (3.13), a shockwave diagram for this intersection approach can be constructed for one cycle of the simulated approach (Figure 6.2).

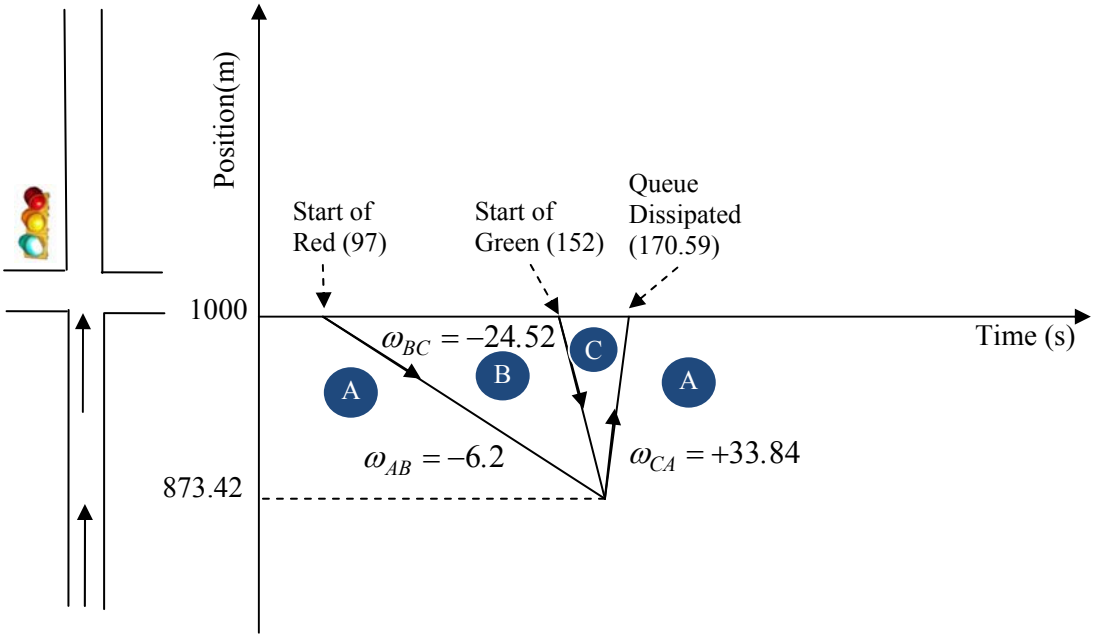


Figure 6.2 Shockwave diagram for one cycle of the simulated approach

In Figure 6.2 *A*, *B*, and *C* represent traffic states that exist upstream of the traffic signal. ω_{AB}, ω_{BC}, and ω_{CA} denote speed of propagation of shockwaves that are created between traffic states (*A*, *B*), (*B*, *C*) and (*C*, *A*) respectively. Table 6.1 quantifies the characteristics of traffic states *A*, *B*, and *C* and the shockwaves generated between them.

The proposed system was applied to the simulation output. All vehicles were treated as probe vehicles. The time intervals were chosen to be 5 minutes and each time step within the time intervals was 5 seconds. A 95% level of confidence was used to perform the hypothesis test explained by equations (4.8) and (4.9). Furthermore, the thresholds to identify shockwave intersection points among the inflection points were chosen to be 18 (km/h) and 14.4 (km/h) for shockwave group 1 (backward forming shockwave in this case) and shockwave group 2 (backward recovery shockwave in this case study) respectively. Because vehicles do not change their speed promptly when they meet a shockwave, more than one shockwave intersection point is expected to be obtained according to the above criteria. Consequently, for shockwave group 1 for each probe vehicle the last shockwave

intersection point is considered for clustering and for shockwave group 2 for each probe vehicle the first shockwave intersection point is taken into account.

Figure 6.3 shows 5 minutes of simulation data. Randomly selected colours have been used for better illustration of individual vehicles' trajectories.

Table 6.1: Attributes of traffic states in Figure 6.2 and shockwaves identified analytically

Traffic State	Flow Rate (vph)	Density (veh/km/lane)	Speed (km/h)	Shockwave Speed (km/h)
A	700	12.04	58.14	-6.20
B	0	125	0	-24.52
C	1900	47.5	40	+33.84
A	700	12.04	58.14	

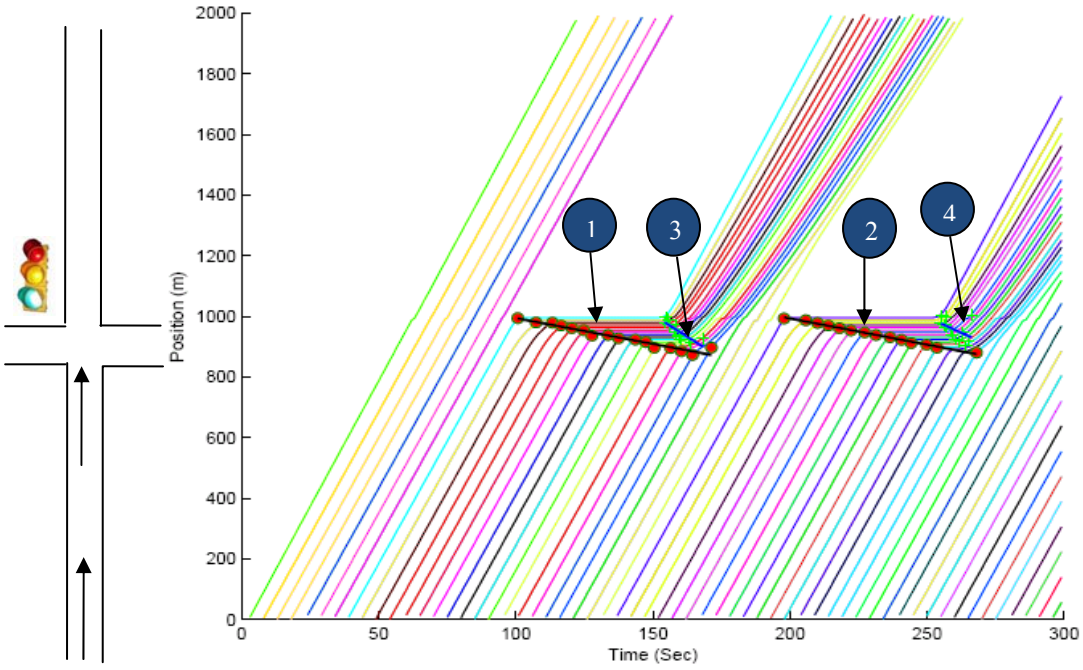


Figure 6.3: Trajectory of probe vehicles, shockwave intersection points and detected shockwaves in a signalized intersection

The red circles in this figure represent shockwave intersection points associated with backward forming shockwaves and the green crosses denote shockwave intersection points corresponding to recovery shockwaves. The lines passing through each group of inflection points, representing the identified shockwaves, are the result of the clustering module.

As can be seen in Figure 6.3, the framework could successfully identify both backward forming and backward recovery shockwaves but failed to identify the forward moving shockwaves. The main reasons for this might be (1) the propagation speed of the forward moving shockwave is higher than the other shockwaves; (2) the shockwave does not represent a boundary between two congested traffic states; and (3) this type of shockwave lasts for a shorter period of time.

Table 6.2 shows the final results obtained by applying the proposed framework to the signalized intersection. In this table columns are:

No: A sequential number representing shockwaves,

Type: Type of the shockwave (1 denotes backward forming shockwaves and 2 represents recovery shockwaves),

t_1: Time that the shockwave was first detected,

t_n: Time that the shockwave was last detected,

d_1: Location that the shockwave was first detected,

d_n: Location that the shockwave was last detected,

ω: Speed of the shockwave,

y_{int}: Y-intercept of the line which represents the shockwave,

S_u: Average speed of vehicles upstream of the shockwave,

S_d: Average speed of vehicles downstream of the shockwave,

n: Number of shockwave intersection points existing in each cluster and used to estimate attributes of the shockwave, n

R^2: Coefficient of determination associated with the regression line representing the shockwave.

Table 6.2: Output of the framework for the signalized intersection

No	Type	t_1 (s)	t_n (s)	d_1 (m)	d_n (m)	Speed (km/h)	y_{int} (m)	S_u (km/h)	S_d (km/h)	n	R^2
1	1	100.645	171.33	993.565	896.837	-6.138	1164.474	41.900	7.963	16	0.946
2	1	197.807	268.044	992.781	880.345	-6.221	1338.313	37.634	3.100	14	0.993
3	2	154.286	168.283	979.07	926.832	-20.765	1869.922	2.588	30.751	13	0.683
4	2	255.142	266.328	979.896	1004.406	-14.890	2031.226	5.044	32.904	13	0.215

A direct comparison of the shockwave speeds estimated by the proposed method and the true shockwave speeds is not possible as the INTEGRATION model does not explicitly output shockwave speeds and there does not exist a formal method of computing shockwave speeds on the basis of vehicle trajectory data. However, a comparison with the analytical estimates is possible (Table 6.3). As can be seen in this table the proposed framework was able to estimate the speed of the backward forming shockwaves accurately (<1 % difference). However, the relative difference between the analytical estimate and the results obtained from applying the framework is 27.3% in case of the recovery shockwave. It is expected that using different thresholds to differentiate between the inflection points and shockwave intersection points would improve the similarity between the proposed framework and the analytical estimates, however, the analytical estimates are themselves an approximation and subject to errors.

Figure 6.4 shows the shockwave diagram obtained from analytical estimates and applying the proposed framework. This figure can be used to temporally and spatially compare the analytical estimates and output of the proposed framework for one cycle of the traffic signal. In this figure the solid black and red lines represent the shockwave diagram based on the analytical framework and the proposed framework respectively. Furthermore, in this figure the red circles represent the beginning and end of the shockwave intersection points identified for both backward forming and recovery shockwaves. As can be seen the main difference is the fact that the proposed framework did not capture the forward moving shockwave.

Table 6.3: Comparison of analytical estimates of shockwave speed with the results obtained from applying the proposed methodology

Shockwave Type	Analytical Estimate of Propagation Speed (km/h)	Average Speed Obtained by the Proposed Framework (km/h)	Relative Difference (%)
Backward Forming	-6.20	-6.18	0.32
Recovery	-24.52	-17.83	27.3

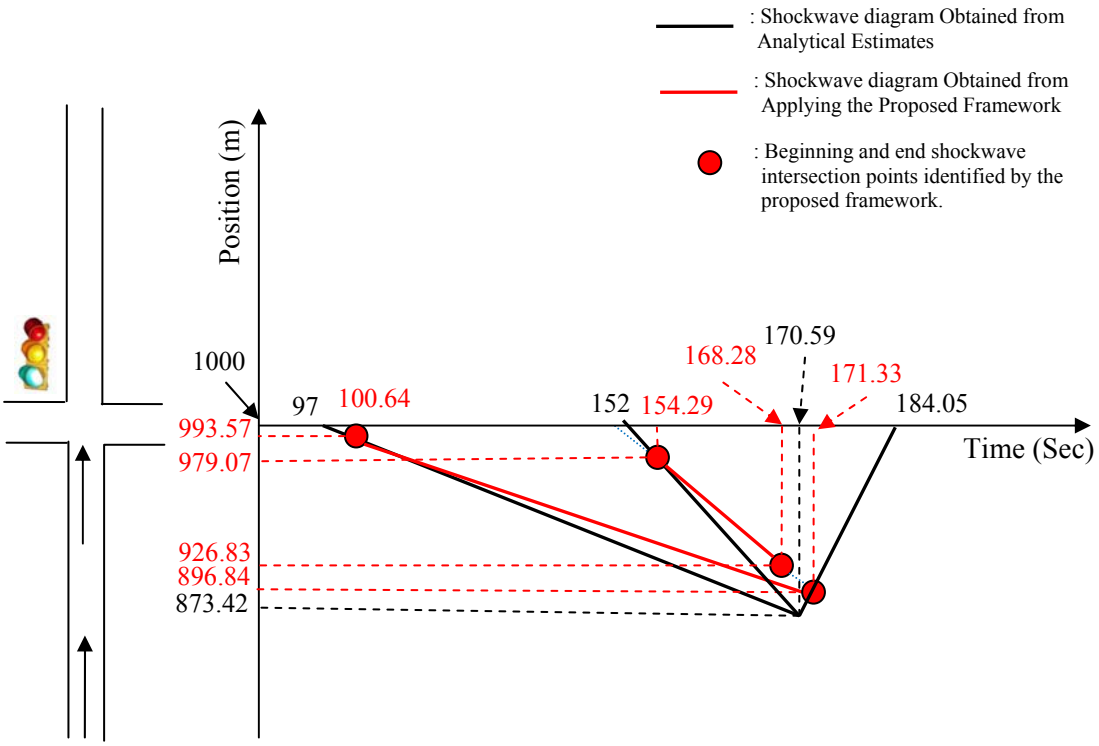


Figure 6.4 Temporal and spatial comparison of shockwave diagrams

6.2 Application 2: NGSIM US-101

The NGSIM data base is the result of Federal Highway Administration (FHWA) efforts to develop algorithms and datasets for calibration and validation of simulation models (FHWA, 2007). This data base includes trajectories of vehicles on freeway US101 in Los Angeles, California. The US101 test section is 0.5 km long. The US101 section consists of 5 lanes which includes a weaving section. Eight cameras were deployed along the study section which videotaped the traffic flow. Image processing software was used by FHWA to post-process the videos in order to obtain detailed information about each individual vehicle including trajectory of each vehicle. In the NGSIM database, positions of individual vehicles were available every tenth of second.

The advantage of the NGSIM database is that it includes the trajectories of all vehicles over a given time period at a very high resolution (i.e. the polling interval is one tenth of a second). However, the length of the section where the data collection was conducted on US-101 is only 0.5 km. Given the fact that average speed during the most congested time during the data collection was 40 km/h, the

average travel time along the section would be less than 1 minute. This means that most of the vehicles which enter the road section will exit the section during even a short (i.e. 5 minute) time interval. It should be noted that the time intervals can be chosen shorter than 5 minutes. However, it is then likely that not enough shockwave points can be found to detect shockwaves in the traffic stream. The short length of the freeway section poses challenges in applying the proposed methodology but it is used for illustration purposes.

The route was divided into 9 sections each with a length of 76.2 m (250 ft) except the last one which had a length of 44.5 m. Three types of probe vehicles are conceivable along this section similar to any other road sections:

1. Probe vehicles which enter and exit the road section during the current time interval,
2. Probe vehicles which start but don't exit the road section during the current time interval,
3. Probe vehicles which entered the route during any previous time interval and exit the road section in the current time interval.

At any time period the complete trajectory of the first group of probe vehicles and their travel times are known. Most of the probe vehicles in the NGSIM data set fall into this group because of the short length of the US101 section. The travel times of probe vehicles in the second group are not known and should be predicted in order to predict the travel time of the current time interval. Travel times of vehicles in the third group could be predicted in order to predict the travel time of vehicles during the time interval in which vehicles in this group entered the road section. It is noteworthy that the partial trajectories of the vehicles in the third group which are available during the current time interval can be utilized to estimate average travel time of each road segment more accurately.

The proposed methodology was applied to the first 15 minute of the data (7:50-8:05). In this study we only analyzed those probes which were travelling on the rightmost lane of the highway. Moreover, due to the fact that the road section is short and most vehicles traverse the road section during a single time interval in which they entered the road section (vehicles in the first group), we separate the vehicles in the first group and the second group to evaluate the performance of the proposed methodology on the vehicles in the second group.

Table 6.4 shows the average speed of each road segment during each time interval (i.e. time interval 1, time interval 2, and time interval 3) which were calculated using data from probe vehicles in the first group. In this table the first column from the left shows each road segment. Each row in

the rest of the columns represents average speed of the corresponding road segment during each time interval. The last row of the table shows average travel time during each time interval. The travel times on this row were calculated based on Equation (4.2). Table 6.5 shows standard deviations of speed associated with each road segment. The same tables can be developed using the data associated with the probe vehicles in the second group.

Table 6.4: Average speed over each road segment using data associated with probes in the first group

	Average Speed (km/h) for Each 5 min Time Interval		
Road Section	7:50-7:55	7:55-8:00	8:00-8:05
1	41.99	33.47	36.14
2	43.14	35.06	37.83
3	44.47	37.37	35.76
4	44.41	33.90	34.69
5	43.34	30.87	32.34
6	50.26	45.31	38.54
7	59.02	57.43	44.12
8	62.88	61.16	43.08
9	63.07	59.72	41.96
Travel Time (sec)	53.31	64.02	68.50
No of Observations	121	94	86

Table 6.5: Standard deviation of speed over each road segment using data associated with probes in the first group

	Standard Deviation of Speed (km/h) for Each Time Interval		
Road Section	7:50-7:55	7:55-8:00	8:00-8:05
1	11.51	10.86	14.12
2	11.22	11.51	9.27
3	13.04	10.75	12.62
4	14.16	8.43	12.71
5	13.01	5.82	11.22
6	9.06	6.64	10.06
7	6.54	8.16	12.54
8	7.16	9.56	14.83
9	7.23	12.69	15.66
No of Observations	121	94	86

Table 6.6 and Table 6.7show average and standard deviation speed of each road segment respectively. The probe vehicles used in the speed calculations shown in Table 6.6 fall into the second group. This group includes the vehicles which have entered the road section during the current interval but have not finished the section. In other words travel times of these vehicles should be predicted. It should be noted that if a vehicle doesn't have any observation for a road segment the average of speed obtained from vehicles in the first group and the third group for that segment was

used. Consequently, as expected the standard deviations in the last two segments are very small. The intent of the shockwave detection is to improve this preliminary prediction.

Table 6.6: Average speed over each road segment using data associated with probes in the second group

	Average Speed (km/h) for each 5 min time interval		
Road Section	7:50-7:55	7:55-8:00	8:00-8:05
1	30.79	39.63	27.61
2	28.34	38.35	22.65
3	34.50	33.57	29.11
4	39.66	28.85	33.16
5	43.94	31.36	33.66
6	50.53	44.55	38.36
7	58.77	56.40	42.49
8	62.99	61.02	42.64
9	63.07	60.94	43.15
Travel Time (sec)	62.20	64.36	78.62
No of Observations	32	32	31

Table 6.7: Standard deviation of speed over each road segment using data associated with probes in the second group

	Standard Deviation of Speed (km/h) for Each Time Interval		
Road Section	7:50-7:55	7:55-8:00	8:00-8:05
1	12.89	7.23	14.52
2	12.47	6.93	11.93
3	11.78	8.01	10.15
4	12.12	9.70	8.30
5	4.69	3.38	8.06
6	3.48	2.80	9.05
7	2.32	3.39	10.37
8	0.90	2.89	9.38
9	0.02	0.00	2.41
No of Observations	32	32	31

The proposed shockwave detection algorithm was applied to the NGSIM data. Figure 6.5 illustrates the shockwave diagram in the first time interval (i.e. the first five minutes). In this analysis each time step within the time intervals was 10 seconds. A 95% level of confidence was used to perform the hypothesis test explained by equations (4.8) and (4.9). Furthermore, the thresholds to identify shockwave intersection points among the inflection points were chosen to be 21.94 km/h (20 ft/sec) and 23.04 km/h (11 ft/sec). The road section was already congested and these values were selected because the shockwave points were visually appealing. As Figure 6.5 shows, all four groups of shockwaves which were introduced previously in the report were detected by the proposed algorithm.

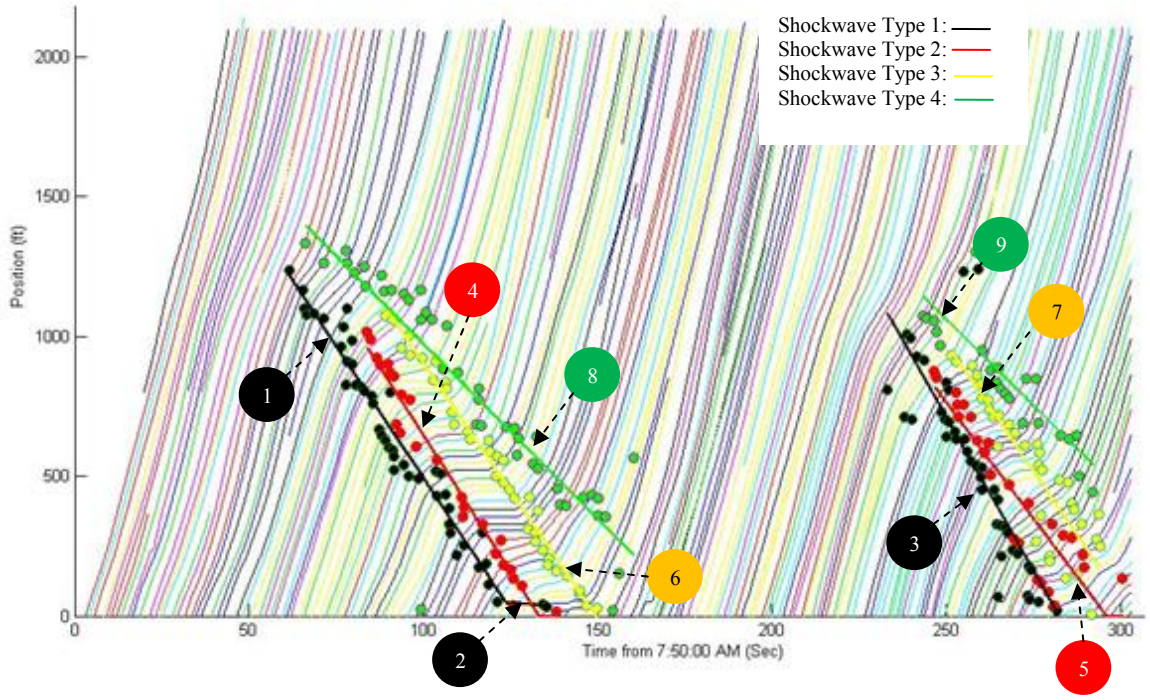


Figure 6.5: Trajectory of probe vehicles, shockwave intersection points and detected shockwaves on one lane of freeway US-101 (NGSIM data)

Table 6.8 shows the characteristics of shockwaves which were found.

Table 6.8: Output of the shockwave detection algorithm for NGSIM US-101 data

No	Type	t_1 (s)	t_n (s)	d_1 (m)	d_n (m)	Speed (km/h)	y_{int} (m)	Su (km/h)	S_d (km/h)	R^2
1	1	61.41	123.54	1219.70	48.69	-18.85	2377.16	33.98	31.17	0.95
2	1	123.54	135.40	48.69	36.89	-0.99	171.58	20.51	37.08	0.95
3	1	233.19	281.84	1081.48	10.56	-22.01	6214.42	34.13	27.88	0.69
4	2	83.81	138.17	961.23	-99.66	-19.51	2596.78	27.10	31.03	0.97
5	2	246.31	300.40	823.27	-75.30	-16.61	4915.13	25.05	26.07	0.82
6	3	89.06	149.69	1102.03	-4.31	-18.25	2727.06	31.49	42.66	0.98
7	3	251.43	294.03	948.33	168.99	-18.30	5548.84	26.75	33.97	0.79
8	4	66.00	160.28	1398.45	218.63	-12.51	2224.46	39.24	47.95	0.77
9	4	243.23	292.03	1143.26	541.34	-12.33	4143.22	37.26	42.90	0.68

No: A sequential number representing shockwaves (each shockwave in Figure 6.5 is marked with corresponding number),
Type: Type of the shockwave (1 and 2 denotes backward forming shockwaves whereas 3 and 4 represent recovery shockwaves),
t_1: Time that the shockwave was first detected (seconds after 7:50 AM),
t_n: Time that the shockwave was last detected (seconds after 7:50 AM),
d_1: Location that the shockwave was first detected (meters from upstream boundary of the study area),
d_n: Location that the shockwave was last detected (meters from upstream boundary of the study area),
ω: Speed of the shockwave (km/h),
y_{int}: Y-intercept of the line which represents the shockwave (meters),
S_u: Average speed of vehicles upstream of the shockwave (km/h),
S_d: Average speed of vehicles downstream of the shockwave (km/h),
R^2: Coefficient of determination associated with the regression line representing the shockwave.

The computational time for the shockwave detection algorithm depends on a number of factors including the number of probe vehicles, the length of the route, polling interval between consecutive positions, duration of time intervals, and duration of time steps. In this research, the shockwave detection algorithm and travel time prediction algorithms were coded in MATLAB. The shockwave detection algorithm was run on a computer with Intel Core 2 Duo 2.66 GHz processor and 2 GB RAM. The computation time for each time interval was 686.86 sec. It should be noted that for the NGSIM dataset, trajectories of all vehicles were used and the polling interval was 0.1 sec. Both factors contributed to longer computation time for the NGSIM set. On the other hand the length of the route associated with the NGSIM data was short.

The proposed travel time prediction method in this research utilizes shockwave information in order to improve the travel time prediction calculated using Equation (4.2). The travel time prediction obtained from Equation (4.2) which does not use the modification factor defined by Equation (4.4) is referred to as the benchmark method in the rest of the document.

In order to quantify the improvements achieved by using shockwave information in travel time prediction, the travel time prediction obtained from the proposed methodology and travel time prediction obtained from the benchmark method are compared with the true travel time of vehicles and the measures of effectiveness defined at the beginning of this chapter.

Figure 6.6 shows travel time prediction obtained from the proposed methodology, the benchmark method, and true travel time associated with each time interval for NGSIM data. The RMSEs associated with the travel time prediction obtained from the proposed and the benchmark methods are 7.71 sec and 29.84 sec respectively. The MAPEs associated with the travel time prediction obtained from the proposed and the benchmark methods are 11.44% and 36.00% respectively. The E_{max} for the proposed and benchmark method are 29.84% and 59.94% respectively.

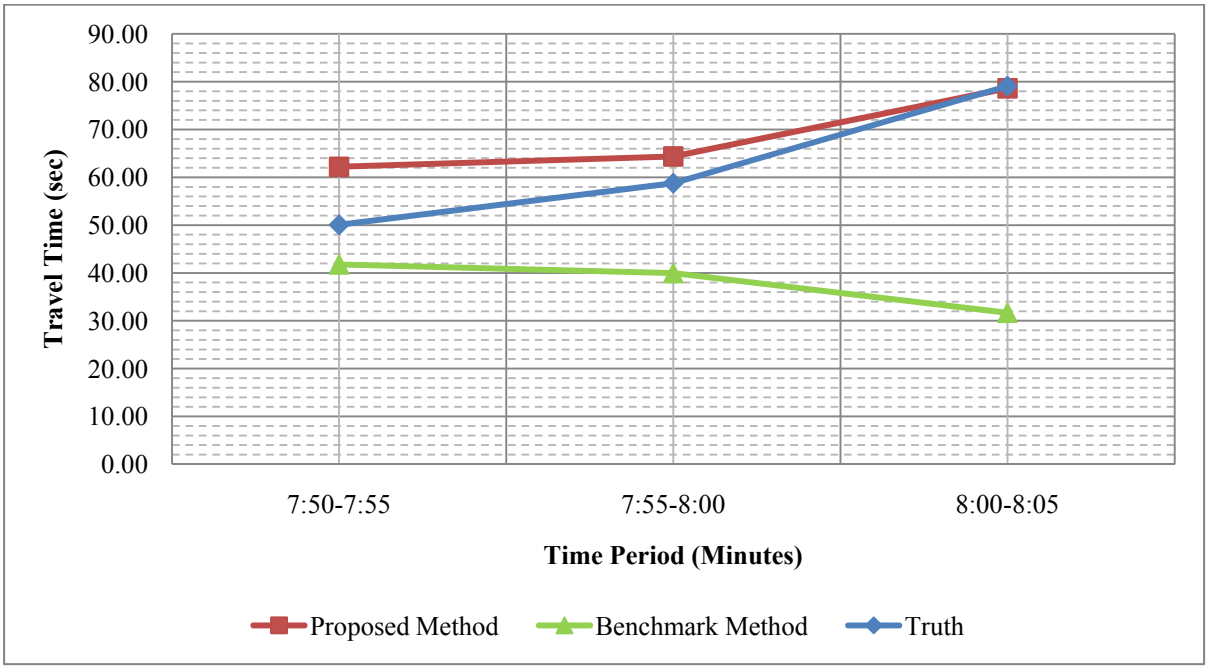


Figure 6.6: Comparison of travel time prediction from the proposed and benchmark methods with true travel time

The application of the proposed model to the NGSIM data (as described in the previous section) demonstrated the ability of the proposed method to identify shockwaves. However, the ability to generalize these findings are limited by the following:

1. The data included were only travelling on a known lane. In real world applications, the lane on which vehicles are travelling may not be identifiable.
2. The NGSIM data differ from the data likely to be available in particular from mobile phone probes because:
    a. NGSIM positions are available at a much higher polling frequency (i.e. every 0.1 seconds),
    b. Position accuracy is much better,
    c. Data are available from all vehicles (i.e. 100% sampling rate).
3. The NGSIM study section is much shorter than what would be considered in real world applications.

4. NGSIM data are available for only 45 minutes which significantly limits the range of traffic conditions captured in the database.

As an attempt to address these limitations, the proposed model is applied to the HWY 85 and HWY 401 study areas in the subsequent sections.

6.3 Application 3: HWY 85

The compilation of the HWY 85 data was described in Section 5.1. For the application of the proposed method, the 8.384 km freeway section was divided into sixteen 500 m segments and a 384 m segment. A time interval duration of 15 minutes was selected. Figure 6.8 illustrates the trajectories of probe vehicles during the study period as well as the positions of interchanges along the study section. Similar to the NGSIM dataset, three types of probe vehicles are expected to be observed:

1. Probe vehicles which enter and exit the 8.384 km road section during the current 15 minute time interval,
2. Probe vehicles which enter but don't finish travelling on the road section during the current time interval,
3. Probe vehicles which entered the route during any previous time interval and will complete the road section in the current time interval.

Table A1 and Table A2 in Appendix A show average speed associated with each road segment and standard deviation for each road segment respectively obtained from probe vehicles which have entered and exited the road section during the same time interval.

Table A3 and Table A4 in Appendix A show average speed associated with each road segment and standard deviation for each road segment respectively obtained from probe vehicles which started and not completed the route during the same time interval.

Figure 6.7 illustrate s the results of the shockwave detection technique for one time interval and Table 6.9 shows the attributes of each shockwave depicted in Figure 6.7. The comparison of values in this table and Table 6.8 shows that the R^2 values associated with the shockwaves identified on HWY 85 are not as large as those obtained through the NGSIM dataset. This finding can have two likely reasons: (1) in NGSIM case study, trajectories of vehicles travelling on one single lane were used but in the probe vehicles on HWY 85 were not constrained to a single lane; (2) US-101 in the NGSIM dataset was more congested than HWY 85. Therefore, vehicles were more constrained to follow each other than the HWY 85 case. It should be noted that the computation time associated with

the shockwave detection algorithm was 55.93 sec on a computer with Intel Core 2 Duo 2.66 GHz processor and 2 GB RAM.

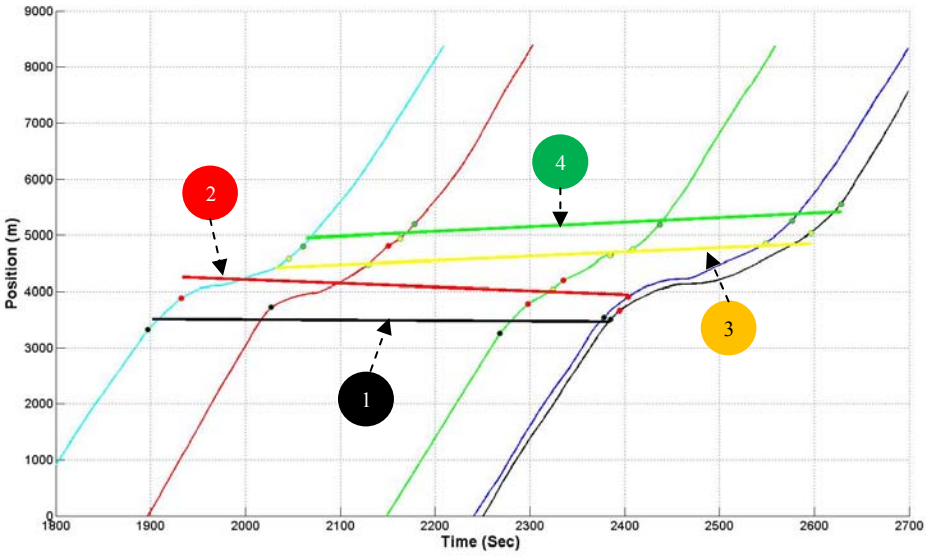


Figure 6.7: Shockwave Diagram for HWY 85 between 16:15:00 and 16:30:00

Table 6.9: Output of the shockwave detection algorithm for HWY 85 between 16:15:00 and 16:30:00

No	Type	t_1 (s)	t_n (s)	d_1 (m)	d_n (m)	Speed (km/h)	y_{int} (m)	Su (km/h)	S_d (km/h)	R^2
1	1	1896.78	2385.00	3511.08	3464.81	-0.34	9126.03	95.48	54.43	0.01
2	2	1932.15	2404.07	4254.41	3941.40	-2.40	43268.88	59.76	60.49	0.08
3	3	2045.51	2596.54	4421.34	4853.73	2.76	-40558.77	62.57	79.99	0.22
4	4	2060.71	2628.34	4954.96	5423.96	3.00	-43936.45	83.19	91.58	0.72

No: A sequential number representing shockwaves (each shockwave in Figure 6.7 is marked with corresponding number),
Type: Type of the shockwave (1 and 2 denotes backward forming shockwaves whereas 3 and 4 represent recovery shockwaves),
t_1: Time that the shockwave was first detected (seconds after 15:35),
t_n: Time that the shockwave was last detected (seconds after 15:35),
d_1: Location that the shockwave was first detected (meters from upstream boundary of the study area),
d_n: Location that the shockwave was last detected (meters from upstream boundary of the study area),
ω: Speed of the shockwave (km/h),
y_{int}: Y-intercept of the line which represents the shockwave (meters),
S_u: Average speed of vehicles upstream of the shockwave (km/h),
S_d: Average speed of vehicles downstream of the shockwave (km/h),
R^2: Coefficient of determination associated with the regression line representing the shockwave.

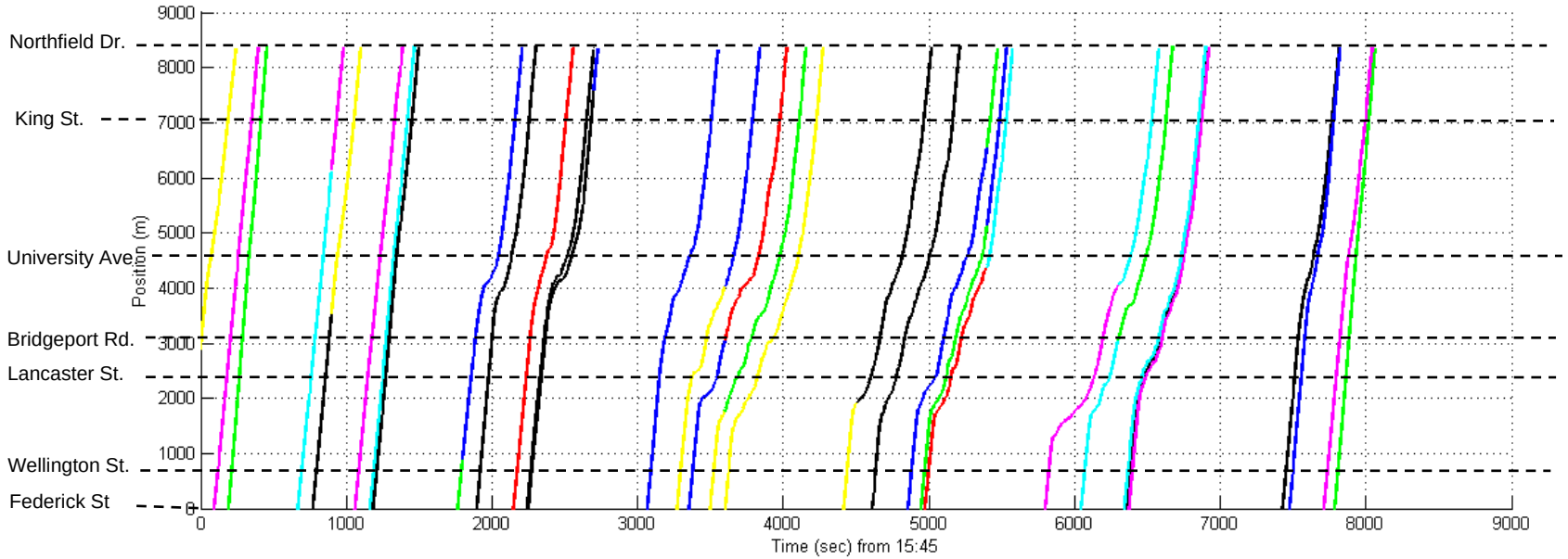


Figure 6.8: Trajectories of probe vehicles along SB of the HWY 85 study section from the start to end of the data collection effort

Figure 6.9 illustrates travel time prediction obtained from the proposed method, the benchmark method, and the truth over each time interval. True travel time in this figure was calculated based on travel time obtained from all probe vehicles equipped with GPS. As can be seen the travel time prediction curve is more consistent with the curve representing the truth. The RMSEs for the travel time prediction obtained from the proposed method and the benchmark method are 0.50 and 0.87 minute respectively. The MAPEs associated with the travel time prediction obtained from the proposed method and the benchmark method are 4.48% and 6.27% respectively. Based on this figure the proposed method is performing particularly better than the benchmark method where the congestion is forming and the congestion is dissipating. E_{max} associated with the proposed method and the benchmark method are 9.7% and 17.87%. Conclusions about the accuracy of the proposed prediction model are limited by the limited number of probe vehicle data points available in each time period.

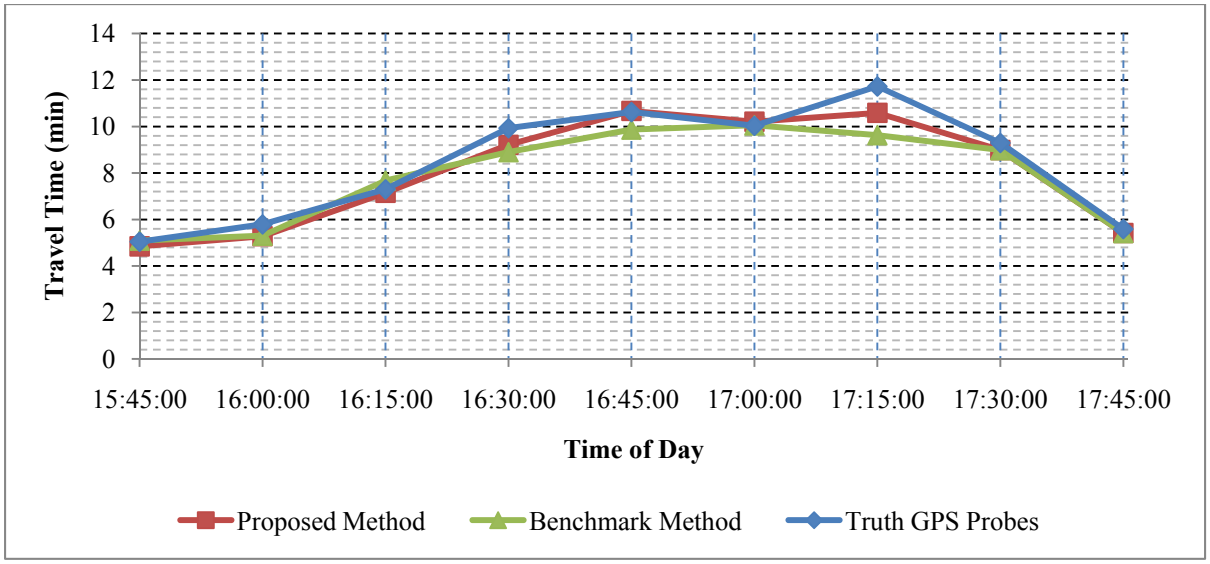


Figure 6.9: Comparison of travel time prediction obtained from the proposed method and the benchmark method with the truth

6.4 Application 4: HWY 401

The compilation of the HWY 401 data was described in Section 5.2. For the application of the proposed method, the 21.35 km freeway section was divided into forty two 500 m segments and a 351.8 m segment. A time interval duration of 15 minutes was selected. Unlike the NGSIM data and HWY 85 data only one probe vehicle entered and exited the study section during one time interval. The main reason was the length of the study section and also high level of congestion particularly in the eastbound direction. Figure 6.10 illustrates the results of the shockwave detection technique for

the period between 15:00:00 and 15:30:00. Table 6.10 shows the attributes of each shockwave depicted in Figure 6.10. In order to identify a shockwave a minimum of three shockwave points are required. Therefore, as can be seen in the shockwave diagram shown in Figure 6.10, the limited number of probes with respect to the length of the route has negatively affected the performance of the shockwave detection algorithm.

Figure 6.11 and Figure 6.12 show complete trajectories of probe vehicles along EB direction of the HWY 401 study section as well as the position of interchanges along the study section. Figure 6.13 shows three graphs associated with EB collector lanes of the HWY 401 study section: (1) measured travel time obtained from GPS equipped vehicles. It should be noted that these values are obtained from complete trajectories of each individual probe vehicle which has to be undertaken off-line; (2) prediction of travel time using the proposed method; and (3) travel time prediction obtained from the benchmark method. A consistent behaviour amongst the three graphs can be observed in this figure. The RMSEs of the proposed method and the benchmark method with respect to the measured GPS travel times which are referred to as truth in this figure are 1.88 and 3.86 minutes respectively. The MAPEs of the proposed method and the benchmark method with respect to the truth in this figure are 5.92% and 10.22% respectively. The E_{max} associated with travel times obtained from the proposed method and the benchmark method are 15.62% and 33.63% respectively.

Figure 6.14 illustrates the two prediction values (i.e. travel times obtained from the proposed method and the benchmark method) and the travel time obtained from MTO loop detectors calculated using the trajectory method. The RMSEs of the travel time prediction and travel time prediction without the modification factor with respect to the travel time obtained from loop detectors are 3.23 and 4.51 minutes respectively. The MAPEs of the two prediction values with respect to the trajectory method are 9.94% and 12.66%. E_{max} associated with the proposed method and the benchmark method are 19.70% and 37.63% respectively.

Figure 6.13 and Figure 6.14 clearly show the improvements of the shockwave information at the times when (1) congestion is forming and (2) congestion is dissipating. These time periods are exactly where most travel time prediction models in the literature have difficulty in providing accurate and robust predictions. It should be noted that the computation time of the shockwave detection algorithm for the duration of the study was 1040.21 sec or on average 54.74 sec per a 15 minute time interval using a computer with Intel Core 2 Duo 2.66 GHz processor and 2 GB RAM.

Table B1 and Table B2 in Appendix B show average speed associated with each road segment and standard deviation for each road segment respectively obtained from probe vehicles which have entered and not exited the road section during the same time interval.

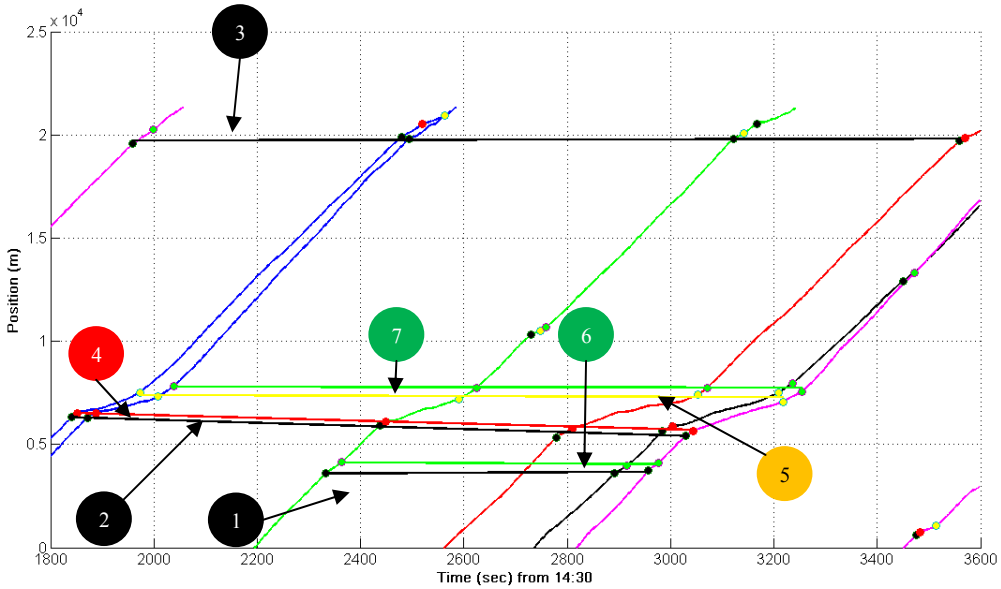


Figure 6.10: Shockwave Diagram for HWY 401 between 15:00:00 and 15:30:00

Table 6.10: Output of the shockwave detection algorithm for HWY 401 between 15:00:00 and 15:30:00

No	Type	t_1 (s)	t_n (s)	d_1 (m)	d_n (m)	Speed (km/h)	y_{int} (m)	Su (km/h)	S_d (km/h)	R^2
1	1	2332.66	2957.62	3583.78	3675.55	0.53	-4423.23	91.26	66.05	0.3257
2	1	1841.15	3029.75	6315.50	5412.36	-2.74	47377.66	80.11	66.68	0.8916
3	1	1958.68	3559.32	19723.73	19809.66	0.19	16816.17	85.05	51.51	0.0670
4	2	1850.92	3043.54	6520.88	5703.02	-2.47	43587.24	66.68	67.04	0.9473
5	3	1973.60	3217.62	7391.85	7284.47	-0.31	12067.82	67.04	84.53	0.0738
6	4	2364.12	2977.22	4129.72	4039.71	-0.53	12140.49	66.05	65.97	0.2635
7	4	2038.09	3254.57	7790.36	7751.48	-0.12	9523.81	85.20	88.84	0.0127

No: A sequential number representing shockwaves (each shockwave in Figure 6.10 is marked with corresponding number),
Type: Type of the shockwave (1 and 2 denotes backward forming shockwaves whereas 3 and 4 represent recovery shockwaves),
t_1: Time that the shockwave was first detected (seconds after 15:35),
t_n: Time that the shockwave was last detected (seconds after 15:35),
d_1: Location that the shockwave was first detected (meters from upstream boundary of the study area),
d_n: Location that the shockwave was last detected (meters from upstream boundary of the study area),
ω: Speed of the shockwave (km/h),
y_{int}: Y-intercept of the line which represents the shockwave (meters),
S_u: Average speed of vehicles upstream of the shockwave (km/h),
S_d: Average speed of vehicles downstream of the shockwave (km/h),
R^2: Coefficient of determination associated with the regression line representing the shockwave.

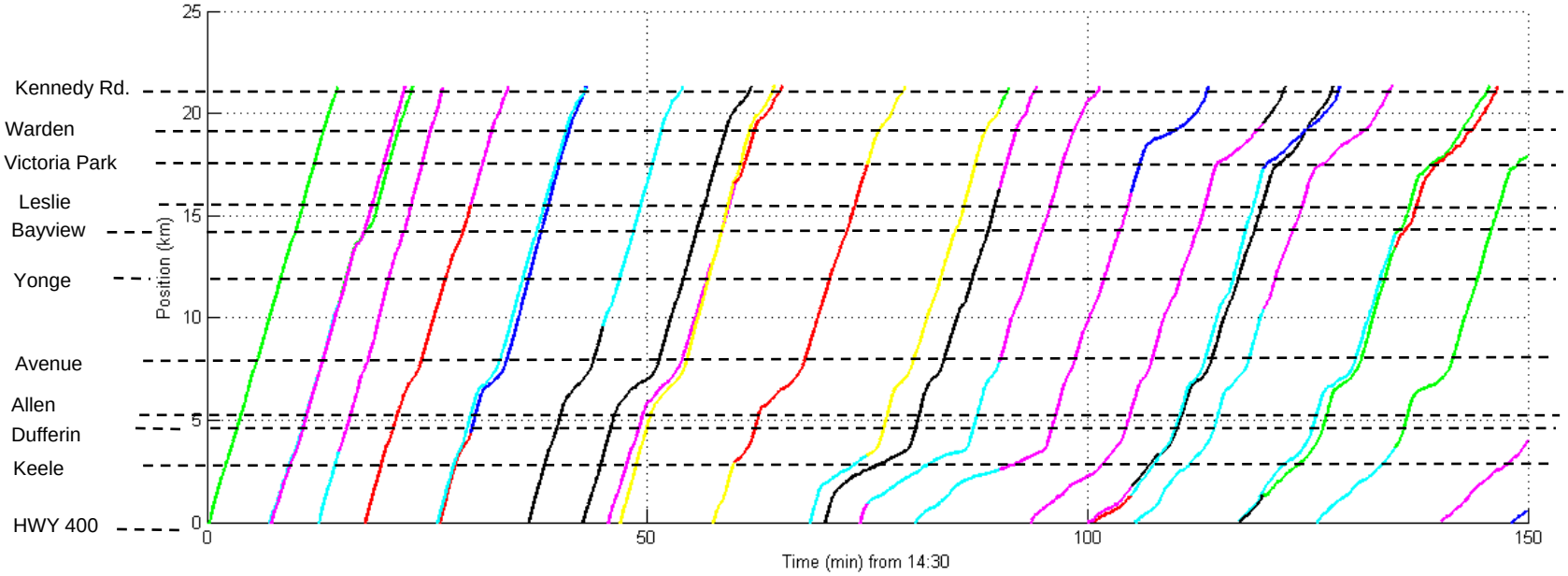


Figure 6.11: Trajectories of probe vehicles on EB of the HWY 401 study section from 14:30 to 17:00

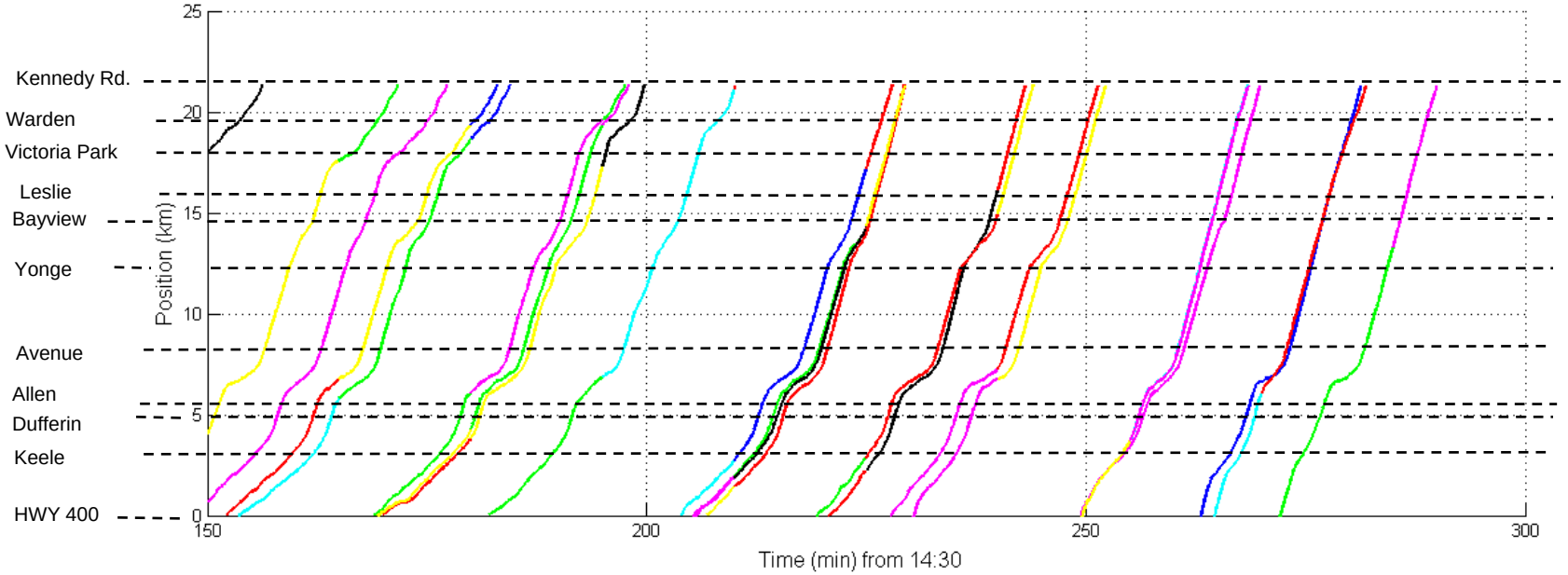


Figure 6.12: Trajectories of probe vehicles on EB of the HWY 401 study section from 17:00 to 19:30

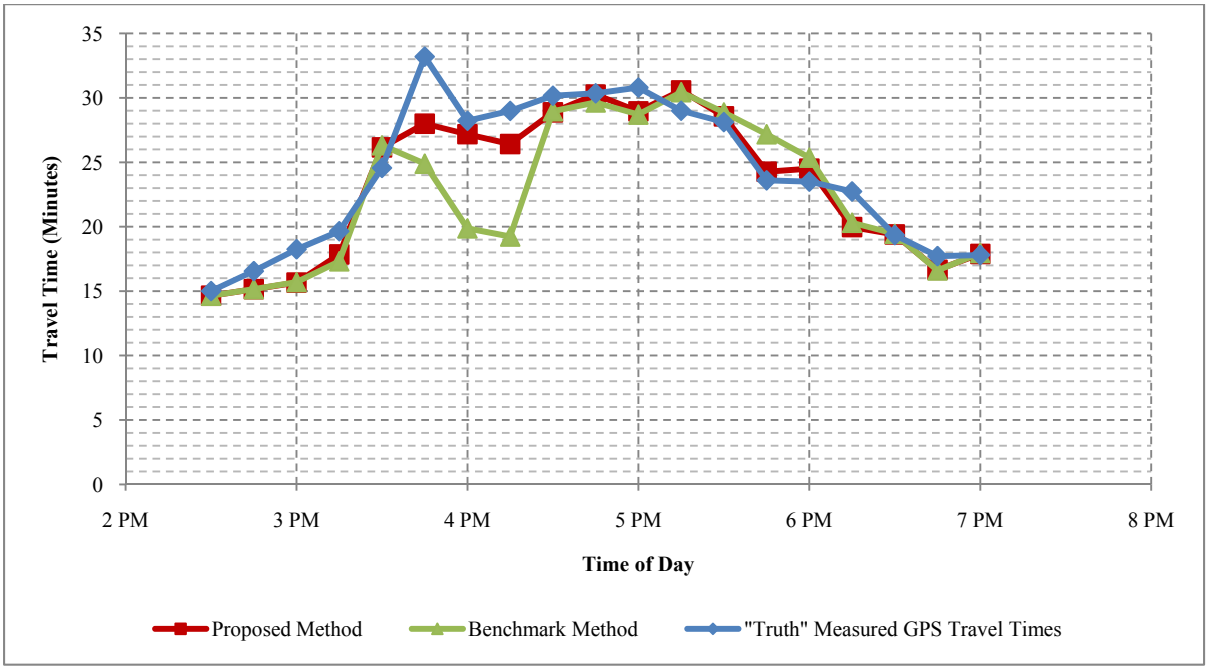


Figure 6.13: Comparison of travel time predictions obtained from the proposed method and the benchmark method with the measured GPS travel times for HWY 401 EB collector lanes

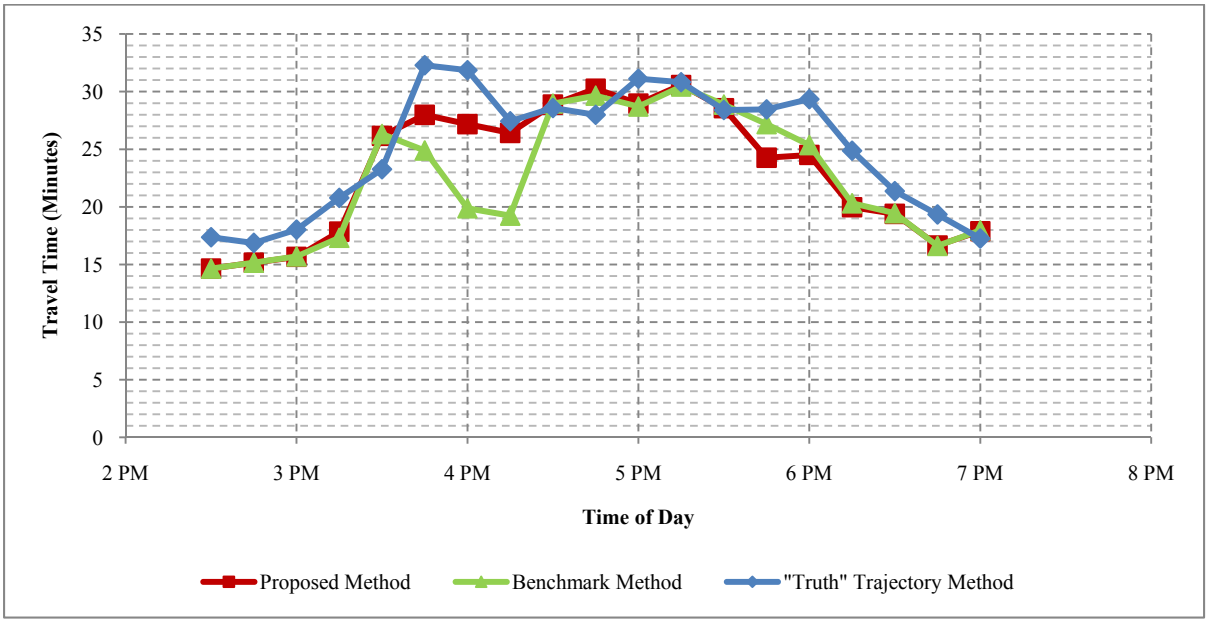


Figure 6.14: Comparison of travel time predictions obtained from the proposed method and the benchmark method with the trajectory method for HWY 401 collector lanes

In order to compare the results obtained from the proposed method and the benchmark method Figure 6.15 can be used. Figure 6.15 shows travel time obtained from loop detector data using the MTO method and the trajectory method as well as measured GPS travel times. This figure also was previously shown in Chapter 5 as Figure 5.9. RMSEs of MTO method with respect to the measured GPS travel times and the trajectory method are 3.23 min and 2.16 min respectively. The MAPEs of the MTO method with respect to the measured GPS travel times and the trajectory method are 9.77% and 6.11% respectively. E_{max} associated with the MTO method with respect to the measured GPS travel times and the trajectory method are 28.88% and 16.73% respectively.

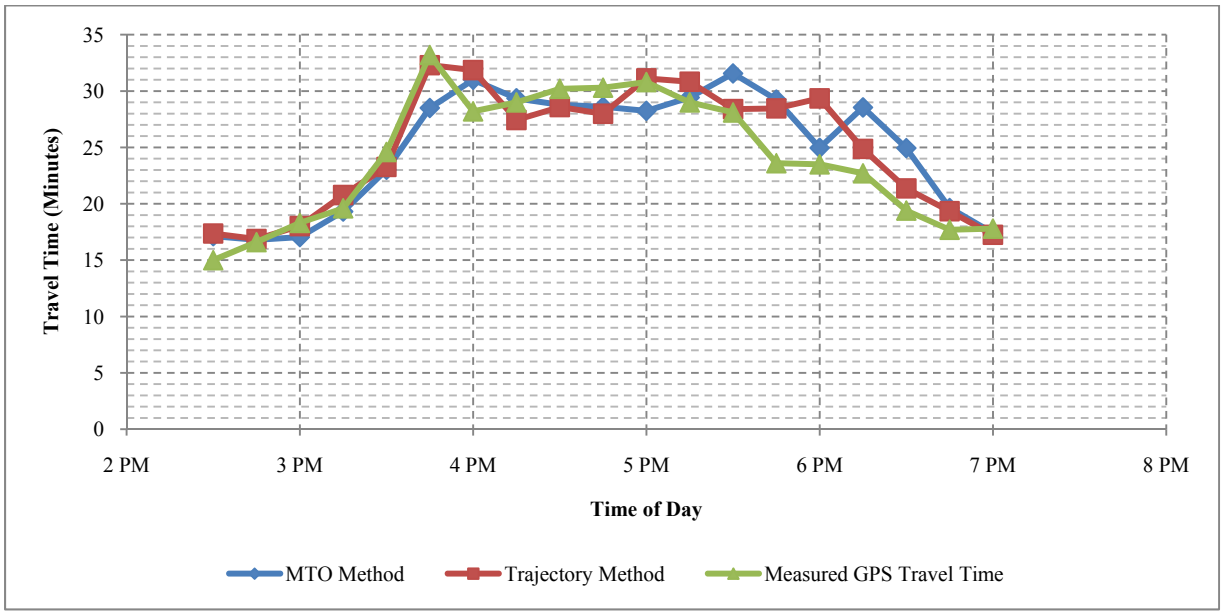


Figure 6.15: Comparison of travel time obtained from loop detector with travel time obtained from GPS equipped probes for EB HWY 401

6.5 Summary of Results

This chapter dealt with the application of the proposed methodology to four test networks, namely (1) a simulated hypothetical signalized intersection; (2) NGSIM data associated with a section of freeway US-101 in California; (3) a section of Highway 85 in the Region of Waterloo; and (4) a section of Highway 401 in the Greater Toronto Area.

The first test network was used to evaluate the performance of the proposed shockwave detection methodology to identify shockwaves at a signalized intersection. The proposed algorithm was able to successfully identify backward forming and backward recovery shockwaves but failed to identify the forward moving shockwave. The attributes of the identified shockwaves were compared with those

obtained from traffic flow theory and found that the relative difference between the propagation speeds of the backward forming shockwave obtained from the shockwave detection algorithm and traffic flow theory is 0.32%. The same value corresponding to the backward recovery shockwave is 27.3%. The maximum length of queue obtained from the shockwave detection algorithm was 103.16 m whereas the maximum length of queue obtained from traffic flow theory was 126.58 m which shows 18.5% difference. This highlights the capabilities of the proposed algorithm depicting the evolution of congestion in a traffic stream which per se has a number of applications such as end of queue warning systems and routing of emergency vehicles.

For the other three test networks, the proposed shockwave detection algorithm was able to successfully identify shockwaves in the traffic stream. Also, the proposed real-time travel time prediction was utilized to predict travel time associated with different time intervals of each test network. Table 6.11 shows RMSEs associated with the proposed method and benchmark method for each road section. As can be seen in this table the proposed method is consistently superior to the benchmark method. This table also shows the improvement achieved by the proposed method with respect to the benchmark method. Table 6.12 illustrates MAPEs and E_{max} associated with the proposed method and the benchmark method for each test network. This table is consistent with Table 6.11 and shows that the proposed method is superior to the benchmark method.

Table 6.11 and Table 6.12 also show the percent improvements achieved by the proposed algorithm with respect to the benchmark method. It should be noted that the main difference between the proposed method and the benchmark method is that the proposed method makes use of shockwave information and projects the shockwave trajectories into the future time intervals. Therefore, the improvements achieved by the proposed are attributable to the shockwave information.

In order to compare the proposed algorithm with another source which is used in practice by many jurisdiction, the travel times obtained from loop detector data using the MTO method was compared with the measured GPS travel times and the loop detector travel times obtained from the trajectory method. Table 6.11 and Table 6.12 show RMSEs, MAPE, and E_{max} associated with the MTO method with respect to the measured GPS travel times and the trajectory method. As can be seen in these tables the proposed method is superior to the MTO method when the measured GPS travel times are the reference. However, the MTO method is superior to the proposed method when the trajectory method is the reference. It is important to note that the spacing of loop detectors was on average 4.27 which is rather excessive which sheds doubts on the credibility of the travel times obtained from loop

detectors. One important conclusion can be however drawn from the comparison of the results obtained from the proposed method and the MTO method. The loop detector data are the results of significant capital investment in FTMS for freeways in major urban area. Such systems require regular maintenance due to damages caused by traffic, construction of the freeways, and adverse weather conditions. Moreover, the flow of data from loop detectors is jeopardized by most road works. The proposed algorithm with limited resources available in this research showed that the emerging technologies are potentially able to provide travel time predictions which are likely superior to the travel times obtained from loop detectors. The emerging technologies such as mobile phone probes make use of the existing infrastructure of the cellular network with minimal costs.

Table 6.11: RMSE associated with the proposed method and the benchmark method for each test network

Algorithm	US-101	HWY 85	HWY 401	
	Truth	Truth	"Truth": Measured GPS Travel Time	"Truth": Trajectory Method
Proposed Method	7.71 sec	0.5 min	1.88 min	3.23 min
Benchmark Method	29.84 sec	0.87 min	3.86 min	4.51 min
Improvement	74.2%	42.5%	51.3%	28.4%
MTO Method	-	-	3.23	2.16

Table 6.12: MAPE and E_{max} associated with the proposed method and the benchmark method for each test network

Algorithm	US-101		HWY 85		HWY 401			
	Truth		Truth		"Truth": Measured GPS Travel Time		"Truth": Trajectory Method	
	MAPE	E_{max}	MAPE	E_{max}	MAPE	E_{max}	MAPE	E_{max}
Proposed Method	11.44%	29.84%	4.48%	9.7%	5.92%	15.62%	9.94%	19.7%
Benchmark Method	36.00%	59.94%	6.27%	17.87%	10.22%	33.63%	12.66%	37.63%
Improvement	68.22%	50.22%	28.55%	45.72%	42.07%	53.55%	21.48%	47.65%
MTO Method	-	-	-	-	9.77%	28.88%	6.11%	16.73%

Chapter 7
Conclusions

This Chapter highlights the main conclusions and contributions of this thesis research and presents directions for future work. The main conclusion of this research, which is valuable from academic perspective, is that this research provides a systematic framework to study the formation and dissipation of congestion on freeways. This information is not only important for travel time prediction and traveler information systems which are the subject of this research but also is critical in a number of applications in the area of freeway traffic management systems and incident management systems.

In this research a dataset was collected and compiled for two freeway sections. This dataset provided an opportunity to gain insight into acceleration and deceleration behavior of probe vehicles along congested freeways. Moreover, the methodology presented in this research was applied to this dataset to assess the performance of the methodology in real applications.

Based on the findings of this research the following major conclusions can be made:

- The methodology presented in this research is capable of providing real-time travel time prediction. The results show that the average prediction error of the proposed algorithm is much less than the travel time predictions which are traditionally performed using loop detector data.
- This research demonstrated the possibility of identifying shockwaves in a traffic stream using a small sample of vehicles.
- It was shown in this research that the shockwave detection algorithm can estimate the shockwave attributes close to the values obtained from traffic flow theory.
- The study of trajectories of real probe vehicles in transient traffic conditions revealed more complex shockwave diagrams. It appears that drivers do not accelerate or decelerate at a constant rate. This pattern should be given full consideration in the future studies involving shockwaves.
- It was identified that for a small sample of vehicles equipped with GPS devices (0.23% in this research) the average travel times obtained from GPS equipped vehicles are not

statistically different from the true travel time. It should be noted that this conclusion has been drawn based on data from one freeway section. The number of probe vehicles required for a freeway section depends on the variance of speed among vehicles.

7.1 Major Contributions

Four major contributions are emphasized in this section: (1) the development of a real-time travel time prediction algorithm for freeway sections which involves two parallel components; (2) the development of a shockwave detection algorithm; and (3) The development of an algorithm to adjust travel time prediction.

1. Development of a Real-time Travel Time Prediction with Two Parallel Components: This research introduced a travel time prediction model for freeway sections with two parallel components. The first component provides a prediction for travel time using all available data points obtained during the current time interval. One of the novelties of this component is that any source of data can be used to carry out the travel time prediction. This provides an opportunity for data fusion which will be described in Section 7.2. Parallel to this component, the shockwave detection component identifies any shockwaves which may exist during the current time interval and estimates the attributes of each shockwave. Finally, the results of the two components are merged together in order to provide a more accurate and robust real-time travel time prediction using the proposed travel time prediction adjustment algorithm.
2. Development of a Shockwave Detection Algorithm: A new shockwave detection algorithm was developed in this research. The input to the shockwave detection algorithm is trajectories of a sample of vehicles regardless of the technology used to obtain the trajectory data. The main idea behind this algorithm is that when a vehicle and a major shockwave intersect, the speed of the vehicle changes abruptly. The shockwave detection model was built upon a strong theoretical basis of statistics and operations research. In this algorithm, first, a two phase piecewise linear regression is used to find the points at which a vehicle has changed its speed. Then, the points that correspond to the intersection of shockwaves and trajectories of probe vehicles are identified using a data filtering procedure and a linear clustering algorithm is employed to group different shockwaves. Finally, a linear regression model is applied to find propagation speed and spatial and temporal extent of each shockwave.

3. Development of an Algorithm to Adjust Travel Time Prediction: The travel time prediction adjustment algorithm proposed in this research projects the trajectories of the shockwaves detected in the traffic stream and identifies the road segments which each shockwave will likely traverse during the next time intervals. This algorithm adjusts the average speed of each road segment based on the average speed upstream and downstream of the shockwave. Finally, the algorithm predicts travel time of the route (which is composed of several road segments) based on the predicted travel time of each road segment.

7.2 Future Research

The following topics are recommended for future research. These topics improve and complement this research:

1. The prediction horizon in this research is zero. It is recommended that the methodology used in this research be extended to predict travel time for prediction horizons larger than zero. It is recommended that for the prediction horizons larger than zero historical data associated with the route be used for better prediction of travel time.
2. In the methodology presented in this research, first average speeds associated with each road segment for the current time interval are calculated. In this research only data obtained from trajectories of vehicles which traversed these road segments during the current time intervals were used. However, if other sources of data such as loop detector data are available, data fusion can likely improve the calculation of average speeds associated with each road segment.
3. The shockwave detection module in this research makes use the data available only in the current time interval. However, it is expected that if the data associated with previous time intervals are used, a better picture of shockwave diagrams could be obtained which could potentially improve the travel time prediction process.
4. The methodology presented in this research was only applied to freeway sections. However, it can potentially be applied to arterials. The main challenge of applying this methodology to arterials is that the shockwave diagrams are more complicated in arterials due to the interruption of traffic flow by traffic control devices.

5. One of the applications of the proposed shockwave detection algorithm is identification of the tail of the queue. It is recommended that the accuracy of the proposed algorithm in identification of the tail of the queues be investigated.

Appendix A

Table A1: Average speed for each road segment obtained from probe vehicles which entered and completed the HWY 85 road section during each time interval

	Average Speed (m/sec) for Each Time Interval								
Road Segment (m)	1	2	3	4	5	6	7	8	9
0-500	28.80	26.32	27.53	28.11	32.40	31.17	+	27.25	29.42
500-1000	28.09	26.76	28.10	28.13	30.42	31.34	+	27.25	29.38
1000-1500	29.65	28.11	28.53	28.63	18.30	29.76	+	27.90	29.09
1500-2000	30.04	27.77	28.26	28.63	3.38	8.61	+	25.57	29.27
2000-2500	30.74	28.24	26.69	28.63	8.41	7.50	+	7.33	30.23
2500-3000	30.86	28.34	27.09	19.17	9.62	11.61	+	7.45	29.63
3000-3500	30.86	27.75	26.28	11.12	5.97	10.03	+	12.41	27.84
3500-4000	30.96	28.65	12.33	7.52	7.16	8.83	+	6.86	22.83
4000-4500	31.01	28.06	8.71	8.12	9.31	8.01	+	12.21	20.27
4500-5000	30.63	24.59	11.64	7.94	13.63	9.70	+	17.49	16.81
5000-5500	30.34	22.92	18.78	18.88	19.30	18.25	+	15.07	22.06
5500-6000	30.58	25.10	22.50	21.94	21.76	18.68	+	24.30	24.69
6000-6500	29.60	25.20	26.24	21.94	21.54	11.32	+	26.35	26.62
6500-7000	30.33	27.11	28.00	24.65	25.37	24.93	+	26.01	26.84
7000-7500	31.58	28.05	28.00	25.72	28.45	29.59	+	25.73	28.25
7500-8000	31.58	25.91	26.24	27.64	29.36	28.23	+	24.79	28.43
8000-8384	31.58	25.82	26.70	27.64	28.56	27.66	+	24.92	28.34
Travel Time (sec)	275.97	314.45	405.33	493.81	702.14	603.9	+	539.18	325.14

+ No data is available

Table A2: Standard deviation of speed for each road segment obtained from probe vehicles which entered and completed the HWY 85 road section during each time interval

	Standard Deviation of Speed (m/sec) for Each Time Interval								
Road Segment (m)	1	2	3	4	5	6	7	8	9
0-500	4.43	1.10	1.81	-	-	-	+	0.45	1.99
500-1000	3.04	0.59	1.43	-	-	-	+	0.45	1.72
1000-1500	3.71	0.21	1.21	-	-	-	+	2.03	1.42
1500-2000	2.93	0.82	1.90	-	-	-	+	1.49	1.26
2000-2500	3.92	0.95	2.38	-	-	-	+	2.48	2.14
2500-3000	4.08	2.71	1.47	-	-	-	+	0.85	1.51
3000-3500	4.08	3.09	2.04	-	-	-	+	0.51	3.01
3500-4000	3.95	1.71	3.09	-	-	-	+	1.51	7.94
4000-4500	3.88	1.70	2.98	-	-	-	+	0.28	10.40
4500-5000	4.40	0.76	2.42	-	-	-	+	0.42	8.10
5000-5500	4.82	3.62	1.88	-	-	-	+	0.31	3.72
5500-6000	4.48	1.74	1.79	-	-	-	+	1.32	2.81
6000-6500	5.86	1.40	2.20	-	-	-	+	0.21	2.53
6500-7000	4.84	0.61	1.58	-	-	-	+	0.08	3.22
7000-7500	3.06	1.57	1.91	-	-	-	+	0.30	1.71
7500-8000	3.06	4.01	1.52	-	-	-	+	1.02	1.30
8000-8384	3.06	3.94	1.49	-	-	-	+	0.85	1.35

+ No data is available

- Only one observation existed

Table A3: Average speed for each road segment obtained from probe vehicles which entered and NOT completed the HWY 85 road section during each time interval

	Average Speed (m/sec) for Each Time Interval								
Road Segments (m)	1	2	3	4	5	6	7	8	9
0-500	26.80	26.37	29.40	28.09	29.17	26.62	26.85	+	+
500-1000	26.56	26.66	28.26	27.14	29.17	26.56	27.48	+	+
1000-1500	27.42	27.49	25.90	27.29	29.17	26.45	17.86	+	+
1500-2000	26.34	27.29	24.65	18.66	19.16	9.03	4.95	+	+
2000-2500	26.24	27.57	26.28	13.59	6.87	8.57	7.36	+	+
2500-3000	26.97	27.63	25.49	15.20	8.38	11.17	11.18	+	+
3000-3500	26.18	27.28	24.19	11.77	7.48	10.46	13.73	+	+
3500-4000	28.95	27.81	11.92	6.97	6.72	9.20	12.56	+	+
4000-4500	27.20	27.49	5.38	8.08	9.10	7.39	14.66	+	+
4500-5000	26.95	23.57	9.21	7.94	12.46	9.65	18.38	+	+
5000-5500	27.08	22.71	15.32	18.88	18.05	17.31	20.99	+	+
5500-6000	27.62	24.54	26.13	21.94	21.16	16.98	21.61	+	+
6000-6500	27.81	24.90	27.00	21.94	20.90	14.10	21.44	+	+
6500-7000	28.46	26.80	28.57	24.65	24.01	24.50	23.58	+	+
7000-7500	29.23	28.35	29.87	25.72	27.53	27.02	26.81	+	+
7500-8000	28.62	26.70	26.83	27.64	27.83	25.97	28.04	+	+
8000-8384	28.62	26.37	26.71	27.64	27.85	26.34	27.92	+	+
Travel Time (sec)	305.69	318.23	460.32	533.84	592.28	603.11	557.76	+	+

+ No data is available

Table A4: Standard deviation of speed for each road segment obtained from probe vehicles which entered and NOT completed the HWY 85 road section during each time interval

	Standard Deviation of Speed (m/sec) for Each Time Interval								
Road Segments (m)	1	2	3	4	5	6	7	8	9
0-500	0.14	-	-	2.21	-	1.16	0.41	+	+
500-1000	1.12	-	-	0.43	-	1.06	0.03	+	+
1000-1500	1.27	-	-	0.74	-	0.91	12.84	+	+
1500-2000	0.45	-	-	6.91	-	5.24	1.96	+	+
2000-2500	1.26	-	-	13.02	-	3.47	0.60	+	+
2500-3000	0.28	-	-	3.44	-	2.57	1.46	+	+
3000-3500	1.40	-	-	1.15	-	4.15	0.77	+	+
3500-4000	1.68	-	-	0.95	-	0.54	2.85	+	+
4000-4500	0.43	-	-	0.07	-	1.59	0.39	+	+
4500-5000	0.78	-	-	0.00	-	1.56	0.00	+	+
5000-5500	0.60	-	-	0.00	-	1.42	0.00	+	+
5500-6000	0.17	-	-	0.00	-	1.65	0.00	+	+
6000-6500	0.06	-	-	0.00	-	1.20	0.00	+	+
6500-7000	0.00	-	-	0.00	-	0.37	0.00	+	+
7000-7500	0.00	-	-	0.00	-	0.00	0.00	+	+
7500-8000	0.00	-	-	0.00	-	0.00	0.00	+	+
8000-8384	0.00	-	-	0.00	-	0.00	0.00	+	+

+ No data is available

- Only one observation existed

Appendix B

Table B1: Average speed for each road segment obtained from probe vehicles which entered and NOT completed EB of the HWY 401 road section during each time interval

Road Segments	Average Speed (m/sec) for Each Time Interval																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0-500	27.53	27.15	27.07	27.06	27.08	21.02	7.41	7.32	7.36	5.31	5.79	5.44	4.77	9.13	6.66	12.52	24.88	27.32	27.37
500-1000	23.48	25.73	25.05	21.77	23.19	14.53	5.25	8.20	7.73	9.73	7.39	5.76	4.37	7.73	7.99	11.12	11.37	22.68	27.07
1000-1500	22.48	25.68	24.50	21.89	16.80	12.15	8.86	15.69	6.56	10.35	7.51	6.99	10.09	8.89	8.29	7.88	13.66	21.67	24.82
1500-2000	22.36	25.03	23.40	23.47	7.51	5.92	11.96	14.58	9.06	8.84	7.04	7.24	10.34	10.42	9.66	9.16	7.51	12.63	18.54
2000-2500	21.96	24.14	23.16	24.46	6.43	3.66	13.23	14.41	6.53	9.86	6.88	6.36	6.43	11.17	8.94	8.90	9.25	9.93	15.01
2500-3000	22.82	24.09	25.83	25.19	9.07	6.31	13.18	14.32	6.93	8.89	8.96	7.45	10.20	11.92	8.27	10.55	7.50	10.80	11.95
3000-3500	21.04	20.37	25.46	27.33	9.84	6.40	14.86	14.21	8.33	10.43	7.47	8.34	9.50	12.08	9.41	12.26	12.01	14.65	14.77
3500-4000	20.85	18.67	21.73	24.51	11.74	12.11	17.62	16.28	13.07	11.39	11.37	10.34	13.49	12.08	13.14	14.81	15.04	14.74	15.23
4000-4500	23.27	21.46	23.01	24.80	13.22	18.08	21.08	21.27	16.73	14.33	16.93	13.61	18.58	12.08	20.35	19.50	15.44	21.12	21.60
4500-5000	24.93	25.24	24.83	26.09	15.47	19.48	22.39	22.61	17.75	16.58	18.06	13.62	23.06	12.08	21.11	22.95	15.44	24.55	22.18
5000-5500	23.75	23.68	24.70	24.31	15.47	19.46	22.40	24.17	18.23	17.55	19.66	15.09	22.27	12.08	20.89	23.31	15.44	24.50	23.42
5500-6000	24.19	23.39	24.30	11.52	10.55	20.12	22.69	24.75	18.78	16.92	11.60	12.62	7.50	12.08	17.71	19.24	15.44	24.23	24.10
6000-6500	24.25	22.85	16.41	11.55	10.34	18.21	22.99	21.57	16.18	10.30	8.23	10.09	8.67	12.08	15.16	7.29	15.44	11.93	16.99
6500-7000	24.49	20.18	10.39	9.26	10.28	9.09	21.92	11.94	10.29	9.96	6.96	11.01	6.07	12.08	14.28	7.71	15.73	7.78	5.36
7000-7500	23.46	20.46	14.17	13.57	11.54	12.96	19.71	14.07	12.18	11.66	9.25	12.85	9.94	13.41	16.24	13.06	18.12	10.76	9.86
7500-8000	22.67	21.59	21.24	22.63	13.66	20.33	21.16	21.25	15.86	15.03	13.86	17.00	14.94	15.40	21.73	20.03	21.64	16.97	19.70
8000-8500	23.91	23.13	22.96	25.53	14.21	22.21	22.82	25.27	18.14	16.32	14.91	20.11	17.98	16.14	23.14	21.36	23.61	18.87	24.74
8500-9000	25.37	23.82	25.09	26.40	14.69	22.43	24.38	25.98	18.75	16.92	16.27	20.49	19.67	16.70	23.26	22.47	24.44	20.28	25.86

Road Segments	Average Speed (m/sec) for Each Time Interval																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
9000-9500	26.31	24.67	26.48	26.77	14.80	23.47	24.55	26.02	19.33	17.15	16.92	20.38	20.22	17.17	23.40	22.99	25.31	21.49	26.93
9500-10000	25.65	25.25	27.03	25.17	14.57	23.34	22.02	25.35	18.91	16.78	16.42	20.83	19.75	17.03	23.65	23.35	25.39	21.60	27.07
10000-10500	23.76	24.04	26.40	23.77	14.30	22.60	19.73	22.86	17.93	14.97	15.74	20.17	17.97	15.35	23.54	23.52	24.65	20.94	25.64
10500-11000	21.83	23.48	26.53	24.28	14.58	21.54	19.07	18.66	16.01	15.74	15.87	18.61	16.55	14.23	24.13	23.53	24.43	20.92	27.02
11000-11500	22.98	24.84	25.80	25.02	14.53	22.63	21.19	20.51	18.37	15.21	16.14	18.62	15.44	15.22	23.85	23.41	24.50	21.05	27.62
11500-12000	23.86	27.19	25.99	26.23	14.74	23.45	23.14	25.22	20.00	16.27	17.02	21.03	17.62	16.31	24.29	23.74	25.27	22.20	27.46
12000-12500	23.37	25.92	26.13	25.90	14.42	22.69	23.52	25.26	19.89	16.85	16.36	19.65	17.43	16.43	23.97	20.59	21.89	22.48	27.23
12500-13000	22.73	25.61	27.03	23.98	13.74	20.59	23.09	24.16	18.01	17.53	15.09	19.47	11.72	15.55	15.93	17.37	16.38	22.02	26.93
13000-13500	22.51	21.90	25.84	23.92	13.74	20.84	21.42	23.93	17.55	17.55	15.40	15.32	10.14	13.72	11.84	16.07	16.35	20.63	25.76
13500-14000	23.82	14.35	24.62	26.09	13.31	20.25	22.44	24.38	17.72	16.09	13.01	7.61	10.84	13.28	11.32	16.24	19.24	21.80	26.09
14000-14500	24.17	19.74	24.58	25.68	13.99	19.83	21.17	23.91	16.84	13.50	13.46	11.51	13.19	14.47	14.77	21.25	21.76	22.18	24.48
14500-15000	27.57	22.78	25.03	26.57	14.09	20.68	21.82	23.66	16.64	19.45	15.55	15.40	15.07	15.48	17.45	25.30	23.76	25.92	25.98
15000-15500	27.61	25.04	25.78	26.56	14.86	24.69	22.55	24.24	17.24	20.72	16.78	17.97	17.62	16.32	19.25	25.36	24.35	26.19	27.50
15500-16000	27.70	25.66	26.23	26.85	14.86	23.91	24.16	25.82	18.84	23.78	16.73	18.47	18.93	16.65	19.99	25.42	24.99	26.76	27.52
16000-16500	27.76	27.24	27.23	26.59	14.86	24.15	24.86	26.51	20.01	25.75	16.71	19.00	19.10	17.11	20.70	25.62	24.51	27.38	28.17
16500-17000	26.98	25.81	24.78	24.58	15.04	23.80	24.65	25.88	19.68	18.70	15.05	16.93	19.75	17.17	20.93	24.79	26.84	27.64	27.65
17000-17500	26.78	25.47	25.75	26.11	17.89	22.80	24.77	24.62	12.20	10.45	11.88	9.41	19.63	17.03	20.61	25.61	26.14	27.33	26.39
17500-18000	27.13	27.14	26.35	26.22	19.33	23.61	24.51	19.32	8.33	7.56	11.60	6.27	19.57	17.68	21.14	26.59	26.47	28.57	27.38
18000-18500	26.97	26.99	23.68	24.87	18.58	24.35	23.38	17.52	9.83	6.59	11.60	11.68	17.61	17.13	20.42	26.81	27.75	28.07	26.23
18500-19000	26.50	26.62	23.98	25.20	18.03	23.01	23.38	15.21	10.02	9.75	12.45	11.90	14.83	9.21	14.21	26.22	29.00	28.26	25.98
19000-19500	24.17	24.05	25.38	25.47	17.15	12.45	21.04	15.69	10.12	7.81	11.84	11.97	13.93	6.83	11.51	26.14	28.59	27.85	25.86
19500-20000	24.36	22.73	21.61	19.97	14.02	13.37	18.15	16.77	10.67	8.62	12.11	10.76	13.64	8.45	12.44	26.01	25.74	26.18	24.97
20000-20500	25.44	23.84	16.21	13.63	14.22	19.42	21.59	17.39	11.60	11.02	12.47	15.22	15.58	18.57	15.98	26.20	25.07	23.13	23.05
20500-21000	24.64	25.40	15.41	9.57	15.37	12.50	16.69	19.88	14.58	8.76	12.74	11.65	15.28	19.43	20.09	26.51	25.24	25.49	25.96
21000-21351.8	23.72	23.37	16.45	9.51	15.11	15.33	16.54	20.96	15.41	7.58	12.91	13.53	17.09	17.38	18.86	26.19	25.05	24.72	26.29
No of Obs.	3	4	2	4	3	3	3	3	2	2	3	3	1	4	4	3	3	3	3
Travel Time (min)	14.64	15.13	15.70	17.09	26.32	23.73	19.87	19.24	27.78	29.65	30.21	30.45	28.87	27.18	23.58	20.83	19.45	18.31	17.00

Table B2: Standard deviation of speed for each road segment obtained from probe vehicles which entered and NOT completed EB of the HWY 401 road section during each

Road Segments	Standard Deviation of Speed (m/sec) for Each Time Interval																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0-500	2.87	0.46	2.45	2.05	0.94	11.03	1.70	1.22	1.13	0.24	0.96	1.48	-	2.83	0.87	9.71	0.46	3.17	3.12
500-1000	5.59	2.97	4.82	6.83	4.63	11.40	0.77	0.93	1.73	1.89	0.89	1.92	-	2.07	1.47	1.09	0.41	10.15	3.16
1000-1500	1.84	2.87	2.94	1.03	5.92	9.78	2.85	1.29	1.10	2.73	1.62	0.57	-	0.94	1.60	1.16	0.57	6.75	1.44
1500-2000	0.48	2.60	1.61	2.43	3.82	0.70	6.23	6.10	1.58	4.87	1.65	1.24	-	1.16	3.51	0.80	0.73	7.22	7.85
2000-2500	2.36	1.67	1.38	2.76	4.76	1.14	7.08	5.17	1.81	3.42	0.77	1.14	-	1.83	1.76	0.56	0.76	0.25	8.64
2500-3000	3.87	2.51	0.06	2.66	3.48	1.28	7.04	6.96	1.29	4.80	1.29	0.69	-	0.31	0.62	1.12	0.81	2.70	1.50
3000-3500	3.22	5.35	0.31	1.67	2.02	1.97	3.92	5.48	0.23	2.61	0.73	0.92	-	0.00	0.56	2.77	1.48	1.56	1.37
3500-4000	1.24	6.22	4.02	5.40	0.00	5.27	4.94	4.27	1.92	1.49	1.61	1.71	-	0.00	0.23	3.15	0.67	1.45	1.49
4000-4500	1.05	2.86	2.22	2.30	0.00	10.43	2.72	2.16	2.63	0.14	4.76	4.23	-	0.00	1.87	0.53	0.00	3.04	2.50
4500-5000	2.18	2.40	0.79	1.07	0.00	11.65	0.15	0.78	2.02	0.00	5.21	5.61	-	0.00	2.11	0.93	0.00	2.00	4.27
5000-5500	1.35	1.20	0.17	2.21	0.00	11.64	0.14	1.06	2.38	0.00	6.96	8.15	-	0.00	1.97	1.52	0.00	2.03	2.46
5500-6000	0.99	1.43	1.99	4.24	0.00	12.24	0.00	0.55	1.73	0.00	0.22	3.88	-	0.00	3.38	0.65	0.00	3.06	2.15
6000-6500	0.24	2.42	10.90	4.36	0.00	10.55	0.00	4.13	3.36	0.00	2.03	0.36	-	0.00	8.13	1.81	0.00	7.70	3.37
6500-7000	0.51	7.20	3.94	4.83	0.00	4.48	0.00	3.81	0.03	0.00	1.04	0.00	-	0.00	9.15	3.40	0.00	0.91	2.75
7000-7500	0.87	5.21	1.69	3.71	0.00	1.02	0.00	4.19	0.13	0.00	0.55	0.00	-	0.00	6.88	3.71	0.00	1.69	0.88
7500-8000	1.92	1.38	1.41	4.39	0.00	0.32	0.00	1.19	1.41	0.00	2.56	0.00	-	0.00	0.54	1.10	0.00	5.96	1.35
8000-8500	1.77	0.90	0.41	2.36	0.00	0.12	0.00	0.55	2.42	0.00	2.70	0.00	-	0.00	1.09	1.06	0.00	7.60	2.10
8500-9000	2.70	0.94	0.55	3.32	0.00	1.10	0.00	0.27	2.63	0.00	3.45	0.00	-	0.00	1.23	0.07	0.00	8.66	2.22
9000-9500	3.27	0.64	1.00	3.04	0.00	1.16	0.00	0.85	2.39	0.00	3.91	0.00	-	0.00	1.39	0.21	0.00	9.69	0.22
9500-10000	3.09	1.65	0.07	1.99	0.00	0.65	0.00	1.02	2.41	0.00	3.75	0.00	-	0.00	1.67	0.02	0.00	9.78	1.20
10000-10500	0.64	0.51	0.00	2.62	0.00	1.00	0.00	0.45	1.47	0.00	4.27	0.00	-	0.00	1.55	0.18	0.00	9.21	2.70
10500-11000	1.24	1.44	0.00	3.67	0.00	2.83	0.00	0.00	2.07	0.00	4.00	0.00	-	0.00	2.23	0.78	0.00	9.19	1.51
11000-11500	0.00	0.30	0.00	2.45	0.00	1.08	0.00	0.00	0.99	0.00	4.44	0.00	-	0.00	1.90	0.72	0.00	9.31	1.00

Road Segments	Standard Deviation of Speed (m/sec) for Each Time Interval																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
11500-12000	0.00	0.62	0.00	1.59	0.00	0.42	0.00	0.00	0.83	0.00	4.51	0.00	-	0.00	2.42	0.90	0.00	10.31	1.71
12000-12500	0.00	2.49	0.00	1.12	0.00	0.10	0.00	0.00	1.02	0.00	3.60	0.00	-	0.00	2.05	2.94	0.00	10.54	2.21
12500-13000	0.00	2.31	0.00	2.78	0.00	0.37	0.00	0.00	0.74	0.00	2.09	0.00	-	0.00	0.29	0.34	0.00	10.14	1.08
13000-13500	0.00	2.83	0.00	2.51	0.00	0.79	0.00	0.00	0.94	0.00	2.28	0.00	-	0.00	1.07	2.74	0.00	8.24	0.73
13500-14000	0.00	1.90	0.00	1.50	0.00	0.83	0.00	0.00	1.01	0.00	0.91	0.00	-	0.00	1.27	1.56	0.00	6.36	0.68
14000-14500	0.00	2.16	0.00	1.05	0.00	0.78	0.00	0.00	2.31	0.00	1.79	0.00	-	0.00	0.26	1.42	0.00	3.13	1.53
14500-15000	0.00	1.19	0.00	1.90	0.00	2.41	0.00	0.00	1.92	0.00	3.75	0.00	-	0.00	1.79	1.07	0.00	1.02	0.89
15000-15500	0.00	0.38	0.00	2.04	0.00	0.96	0.00	0.00	2.41	0.00	4.72	0.00	-	0.00	3.59	0.91	0.00	1.89	0.88
15500-16000	0.00	0.79	0.00	1.91	0.00	0.45	0.00	0.00	3.40	0.00	4.31	0.00	-	0.00	4.26	0.10	0.00	1.43	0.92
16000-16500	0.00	0.31	0.00	1.84	0.00	0.03	0.00	0.00	4.48	0.00	3.80	0.00	-	0.00	5.57	0.44	0.00	2.73	0.76
16500-17000	0.00	0.22	0.00	1.15	0.00	0.39	0.00	0.00	4.01	0.00	2.99	0.00	-	0.00	5.57	0.28	0.00	1.94	1.45
17000-17500	0.00	0.81	0.00	0.40	0.00	1.22	0.00	0.00	1.64	0.00	3.51	0.00	-	0.00	5.20	0.40	0.00	1.79	3.57
17500-18000	0.00	0.57	0.00	0.86	0.00	0.70	0.00	0.00	1.84	0.00	3.78	0.00	-	0.00	5.80	0.59	0.00	2.53	3.98
18000-18500	0.00	0.47	0.00	1.50	0.00	0.12	0.00	0.00	1.82	0.00	3.78	0.00	-	0.00	4.98	0.76	0.00	0.70	4.26
18500-19000	0.00	0.40	0.00	0.85	0.00	0.69	0.00	0.00	1.08	0.00	2.30	0.00	-	0.00	2.19	0.67	0.00	0.18	3.65
19000-19500	0.00	0.51	0.00	0.29	0.00	4.68	0.00	0.00	1.29	0.00	3.36	0.00	-	0.00	5.31	0.64	0.00	0.42	3.07
19500-20000	0.00	0.37	0.00	2.01	0.00	2.98	0.00	0.00	0.87	0.00	2.90	0.00	-	0.00	4.24	0.62	0.00	0.01	2.34
20000-20500	0.00	0.37	0.00	0.83	0.00	1.08	0.00	0.00	0.57	0.00	2.28	0.00	-	0.00	0.15	0.47	0.00	0.33	0.21
20500-21000	0.00	2.70	0.00	2.18	0.00	5.07	0.00	0.00	0.32	0.00	1.81	0.00	-	0.00	4.60	0.67	0.00	0.22	0.41
21000-21350	0.00	2.62	0.00	3.26	0.00	3.44	0.00	0.00	0.00	0.00	1.51	0.00	-	0.00	3.17	0.39	0.00	1.71	1.31
No of Obs.	3	4	2	4	3	3	3	3	2	2	3	3	1	4	4	3	3	3	3

Appendix C

Mobile phones are now used as a regular medium of communication around the world. A mobile phone system consists of a set of base stations, located on a cell grid typically depicted as a series of adjacent hexagons (Figure C1). One base station is associated with each cell. The base station consists of a tower and a small building containing the radio equipment to communicate with cell phones located within the cell and land-line equipment to communicate with a Mobile Telephone Switching Office (MTSO). The MTSO handles all of the phone connections to the normal land-based phone system for several base stations in a region. Consequently, the MTSO knows the cell in which each mobile phone currently connected to the network is located.

The cell ID information available at the MTSO represents one source of data for inferring traffic conditions. However, extracting meaningful information requires that:

1. It is possible to determine if the mobile phone is in a vehicle, and
2. The road segments a mobile phone has traversed can be identified from a time series of cell IDs and electronic road map database.

As the mobile phone is moving toward the edge of a cell, the base station associated with the cell can measure that the strength of the signal from the mobile phone is diminishing. At the same time the neighbouring cell notes that the mobile phone's signal is strengthening. Consequently, the two base stations communicate with each other through the MTSO and at some point, the mobile phone gets a message that tells the mobile phone to begin communicating with another base station. This process is termed "hand-off" and is depicted in Figure C2(Layton et al., 2006).

Knowledge of the time of hand-off and the geographical region of hand-off provides a second source of data for extracting road conditions data. This source provides more information than just the cell IDs obtained from the MTSO providing the potential for more accurate determination of the mobile phone's position.

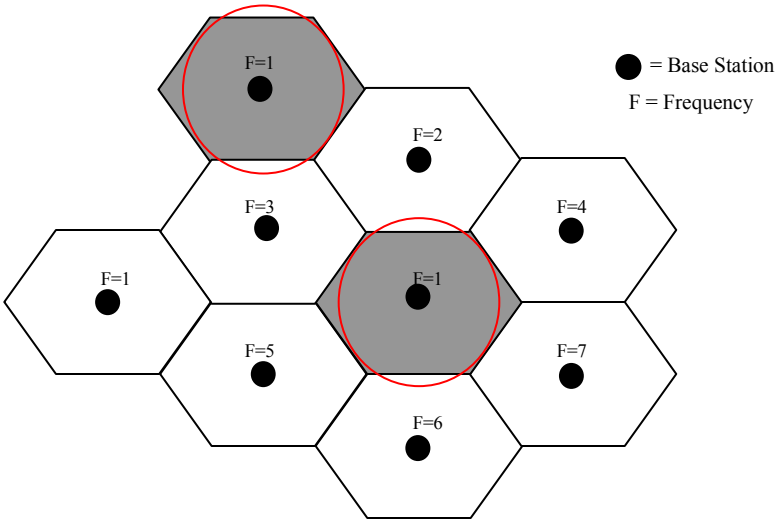


Figure C1: A typical cell grid and associated base stations

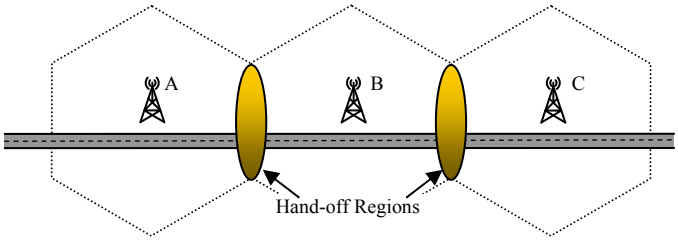


Figure C2: Hand-off mechanism of a moving mobile phone

Although, using locations and time associated with locations, traffic conditions along the route of the probe vehicles could be determined, it requires a few steps to be taken to convert the raw data into travel time and speed. (Hellinga *et. al.*, 2003) and (Hellinga *et. al.*, 2005) provide detailed information about these steps.

Mobile phone Location Identification Techniques

The cellular phone system was not designed originally to provide handset locations and therefore, carriers have had to develop new modules and deploy additional hardware to determine the locations with sufficient accuracy to satisfy the 911 requirements established by the CRTC and FCC.

Location identification techniques can generally be divided into one of 3 categories (Laitinen et al., 2001):

- Network-based
- Handset-based (GPS)
- Assisted- GPS

Each of these location identification techniques are described in the following sections.

Network-based Techniques

In network-based implementations, one or several base stations are involved in locating a mobile phone. Moreover, all required measurements are conducted at the base stations and the measurement results are sent to a location centre where the position is calculated. In this type of implementation, there is no requirement to make any changes to the current handsets. However, the mobile phone must be in active mode (i.e. in "talk" mode or sending a signal through the control channel) to enable location measurement. A number of network-based location identification techniques have been developed. Following are descriptions of the most common techniques.

Cell Identification Technique

Knowledge of the cell in which a handset is located is an intrinsic characteristic of a cellular phone system, and therefore there is no need for network hardware enhancements. In this technique, the location of the mobile phone is approximated by the location of the base station. Relatively minor software changes enable these cell IDs to be obtained continuously over time rather than only when a 911 call is initiated. Obviously, the accuracy of this technique is dependent on the size of the cell; the accuracy in rural areas, where the sizes of cells are substantially bigger than urban areas, is much lower.

Time of Arrival Technique

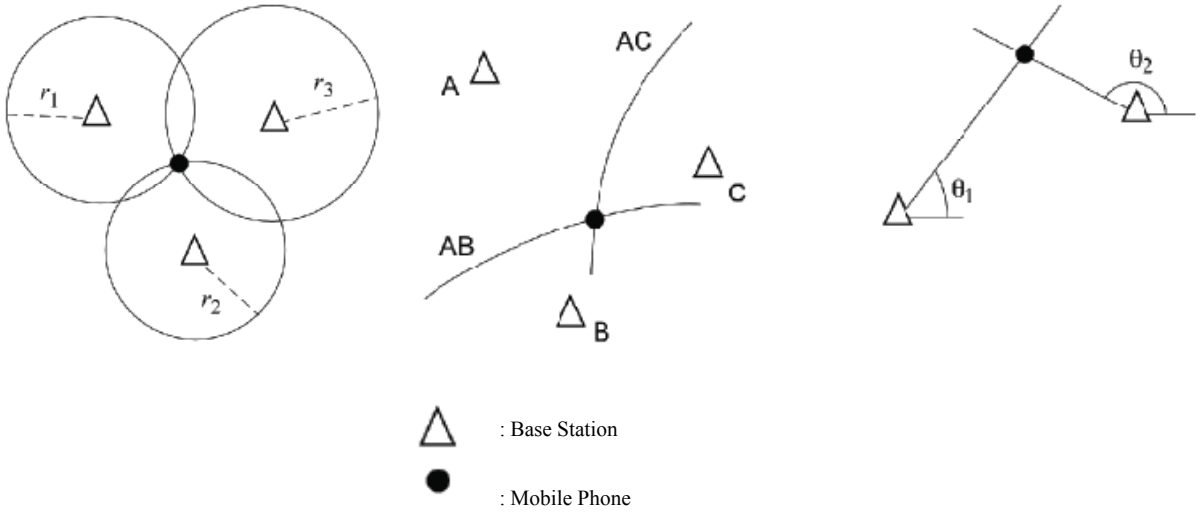
Since radio waves between base stations and mobile stations travel at a constant speed equal to the speed of light, the distance between a mobile phone and a base station is directly proportional to the time of arrival of the wave (Zhao, 2000). Consequently, if at least three base stations identify the time of arrival of a signal from a specific handset, then the location of the handset can be estimated as the intersection of the three circles centered on these base stations (Figure C3-a).

Time Difference of Arrival Technique

Another technique, termed “time difference of arrival” is based on the characteristic that the locus of time difference of arrival of a signal between two base stations and a mobile phone forms a hyperbola. Thus, the mobile phone’s location lies at intersection of two hyperbolas associated with two pairs of base stations (Figure C3-b).

Angle of Arrival Technique

The location of a mobile phone can also be determined by measuring the angle of arrival of the radio wave. In this case, the intersection of two directional lines of bearing defines a unique position (Figure C3-c) (Takada, 2006). This technique requires at least two base stations and also requires directional antennas or antenna arrays to be installed at the base stations to measure the angle of arrival. Since the angle of arrival technique requires line-of-sight propagation conditions to accurately estimate the location of a mobile phone, this technique is not appropriate in dense urban areas (Zhao, 2000).



a) Time of Arrival b) Time Difference of Arrival c) Angle of Arrival

Figure C3: Network-based methods of estimating handset locations.

Timing Advance

Timing Advance (TA) is a Time Division Multiple Access (TDMA) term used in Global System for Mobile communications (GSM) networks. GSM uses the TDMA technology for sharing one frequency between several users in order to avoid interference. The TA value is normally between 0 and 63 and each step represents an advance of one symbol period (approximately 3.69 microseconds). Since the radio waves travel at the speed of light (300,000,000 m/s), each TA step represents 550 m

from the base station (Eberspacher et al., 2001). In urban areas, the maximum TA step is usually 2 (i.e. cell diameter ≤ 2 km) while in rural areas it could be as large as 20 (i.e. cell diameter ≤ 20 km). Consequently, TA can be used to identify the location of a mobile phone with maximum error of approximately 550 m.

Handset-based Techniques

In handset-based implementations all the measurements and calculations are performed in the handset and the results (i.e. location) are transmitted to the base station. In this category of implementation, handsets must be able to measure their own locations, typically through the use of GPS.
GPS uses satellites orbiting the earth to determine position, speed, and time anywhere around the globe. The system is developed and maintained by the US Department of Defence. Civilian access is available through an agreement with the US Department of Transportation (Zhao, 2000). The GPS receiver determines the position of itself based on time of arrival technique.
The use of GPS in mobile phone as a locationing technique suffers from three main disadvantages (Zhao, 2000): First, the time required to obtain a GPS position is relatively long, ranging from 60 seconds to a few minutes due to the time required to acquire the satellite navigation message. Second, GPS signals are too weak to detect indoors and in urban canyons especially with small cellular sized antennas. Third, due to long signal acquisition time, GPS power dissipation is very high.

Assisted GPS Techniques

Assisted GPS (AGPS) is a technique devised to overcome the limitations associated with GPS based locationing. In an AGPS system, a network of fixed GPS receivers (often located at the base stations) is deployed. These receivers are located to have a clear view of the sky and can operate continuously. The reference network is also connected with the mobile phone network and continuously monitors the real time satellite constellation status and provides precise data. At the request of the mobile phone, data derived from the GPS reference network are transmitted to the mobile phone's GPS receiver to "bootstrap" the position acquisition process (Zhao, 2000). Through this technique, acquisition time and consequently power consumption is reduced due to the fact that the search space is limited by data from the reference network. Furthermore, sensitivity of the receiver is increased when the signals are weak (Zhao, 2000). Obliviously, legacy mobile phones cannot be used in this system.

Bibliography

Abbas, M.M. and D. Bullock (2003) On-Line Measure of Shockwaves for ITS Applications. American Society of Civil Engineers, Journal of Transportation Engineering, Vol. 129, No. 1, pp. 1-6.

Abdulhai, B., Porwal, H., and W. Recker (1999) Short Term Freeway Traffic Flow Prediction Using Genetically-Optimized Time-Delay-Based Neural Networks. Transportation Research Board, 78th Annual Meeting CD-ROM, Washington, D.C.

Algers, S., Bernauser, E., Boero, M., Breheret, L., Di Taranto, C., Dougherty, M., Fox, K., and J.F. Gabard (1997) Review of Micro-simulation Models. Review Report of the SMARTEST Project, deliverable 3, Accessed from http://www.its.leeds.ac.uk/smartest/deliv3f.html, on August 8, 2010.

Blue, V., List, G.F., and M.J. Embrechts (1994) Neural Network Freeway Travel Time Estimation. Proceedings of Intelligent Engineering Systems Through Artificial Neural Networks, American Society of Mechanical Engineers, New York, pp. 1135-1140.

Cayford, R. and Y. Yim (2006) A Field Operation Test Using Anonymous Cell Phone Tracking for Generating Traffic Information. Transportation Research Board CD-ROM, 85th Annual Meeting.

Chen, M., and S.I.J. Chien (2001) Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data: Link Based versus Path Based. Transportation Research Record, No. 1768, pp. 157-161.

Chien, S.I.J, and C.M. Kuchipudi (2003) Dynamic Travel Time Prediction with Real-Time and Historic Data. Journal of Transportation Engineering, ASCE, Vol 129, No.6, pp. 606-616.

Chu, L., Oh, J.S., and Will Recker (2005) Adaptive Kalman Filter Based Freeway Travel Time Estimation. Transportation Research Board, 84th Annual Meeting CD-ROM, Washington, D.C.

Coifman, B. (2000) Estimating Travel Times and Vehicle Trajectories on Freeways. Using Dual Loop Detectors. Transportation Research Part A: Policy and Practice, Vol. 36, No. 4, pp. 351-364.

Cortes, C.E., Lavanya, R., Oh, J.S., and R. Jayakrishnan (2001) A General Purpose Methodology for Link Travel Time Estimation Using Multiple Point Detection of Traffic. Institute of Transportation Studies, the University of California at Irvine, Accessed from http://www.its.uci.edu/its/publications/papers/CTSS/UCI-ITS-TS-WP-01-5.pdf, on August 8, 2010.

D'Angelo, M.P., Al-Deek, H., and M.C. Wang (1999) Travel Time Prediction for Freeway Corridors. Transportation Research Records, No. 1676, pp. 184-191.

Daganzo, C. F. (2003) A Variable Formulation of Kinematic Waves: Solution Methods. Institute of Transportation Studies, University of California at Berkeley, Research Report, UCB-ITS-RR-2003-7.

Dhulipala, S. (2002) A System for Travel Time Estimation on Urban Freeways. MASc Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Environment Canada. (2008) Accessed from http://www.climate.weatheroffice.ec.gc.ca/Welcome_e.htm, on January 15, 2009.

Environment Canada (2010) Accessed from http://www.climate.weatheroffice.gc.ca/climateData/hourlydata_e.html?Prov=XX&timeframe=1&StationID=5097&Month=5&Day=13&Year=2009&cmdB1=Go, August 8, 2010.

FHWA (2007) NGSim Homepage. Accessed from http://ngsim.fhwa.dot.gov in May 2007.

Fountain, M.D. and B.L. Smith (2004) Improving the Effectiveness of Wireless Location Technology-based Traffic Monitoring. Virginia Transportation Research Council, Report 05-17, Charlottesville, VA.

Gartner, N.H., Messer, C., and A.K. Rathi (1992) Traffic Flow Theory: A State of the Art Report. Chapter 1, Accessed from http://www.tfhrc.gov/its/tft/tft.htm, on August 8, 2010.

Grewal, M.S., and A.P. Andrews (1993) Kalman Filtering Theory and Practice. Prentice Hall, Englewood Cliffs, New Jersey.

Hartigan, J., 1975. Cluster AnalysisClustering Algorithms. New York, John Wiley.

Head, L.K. (1995) Event-based Short Term Traffic Flow Prediction Model. Transportation Research Record, No. 1510, pp. 45-52.

Hellinga, B., Fu, L., and H. Takada (2005) Traffic network condition monitoring via mobile phone location referencing – an emerging opportunity. Proceedings of the 6th Transportation Specialty Conference of the Canadian Society of Civil Engineers held in Toronto Canada June 2-4, 2005.

Hoogendoorn, S. P. (1997) Optimal Control of Dynamic Route Information Panels. In M. Papageorgiou & A. Pouliezos, eds, 'Preprints of the 1997 IFAC/IFIP/IFORS Symposium', Technical University of Crete, Chania, Greece, pp. 427–432.

Hoogendoorn, S.P., and P.H.L Bovy (2001) State of the Art of Vehicular Traffic Flow Modelling. Proceedings of Institution of Mechanical Engineers Vol. 215, No. 1, pp. 283–303.

I-95 Corridor Coalition (2007) Multi-State Vehicle-Probe Based Traffic Monitoring. Request for Information 82087N, Accessed from http://transport-communications.blogspot.com/2007/03/friday-march-30-2007.html, on August 8, 2010.

Imada, T., and A.D. May (1985) FREQ8PE: A freeway Corridor Simulation and Ramp Metering Optimization Model. UCB-ITS-RR-85-10, University of California, Berkeley, California.

Ishak, S., and H. Al-Deek (2002) Performance Evaluation of Short-Term Time-Series Traffic Prediction Model. ASCE Journal of Transportation Engineering, Vol. 128, No. 6, pp. 490-498.

Izadpanah, P. and B. Hellinga (2007) Wide-Area Wireless Traffic Conditions Monitoring: Reality or Wishful Thanking. Canadian Institute of Transportation Engineers, Annual Conference, 6-9 May 2007, Toronto, Ontario, Canada.

Jia, Z., Chen, C., Coifman, B., and P. Varaiya (2001) The PeMS algorithms for accurate, real- time estimates of g-factors and speeds from single loop detectors." Proceedings of the 4th International ITSC Conference, pp. 536–541.

Kalman, R. E. (1960) A New Approach to Linear Filtering and Prediction Problems. Transaction of the ASME—Journal of Basic Engineering, pp. 35-45.

Karray, F.O., and C. De Silva (2004) Soft Computing and Intelligent Systems Design: Theory, Tools and Applications. Addison Wesley Publications, ISBN 0-321-11617-8.

Khan, A, M, (2007) Intelligent Infrastructure-Based Queue-End Warning System for Avoiding Rear Impacts. IET Intelligent Transportation Systems, 2007, 1, (2), pp. 138–143.

Kindzerske, M.D., and D. Ni (2007) Composite Nearest Neighbor Nonparametric Regression to Improve Traffic Prediction. Transportation Research Records, No. 1993, pp. 30-35.

Kuhne, R. and P. Michalopoulos (1999) Continuum Flow Models. In H. Lieu, ed. Traffic Flow Theory. Turner-Fairbank Highway Research Center, FHWA, http://www.tfhrc.gov/its/tft/tft.htm, Chapter 5.

Lighthill, M. J. and G.B. Whitham (1955) On Kinematic Waves: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of Royal Society, London Series A 229, pp. 317-345.

Lighthill, M. J., and G.B.Whitham (1957) On Kinematic Waves, II: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society, London Series A 229, pp.317-345.

Lindveld, C.D.R., R. Thijs, P.H.L. Bovy, and N. J. Van der Zijpp (2000) Evaluation of Online Travel Time Estimators and Predictors. Transportation Research Record 1719, pp. 45 - 53.

Lu, X. Y. and A. Skabardonis, 2007. Freeway Traffic Shockwave Analysis: Exploring the NGSIM Trajectory Data. Transportation Research Board, 86th Annual Meeting CD-ROM, Washington D.C.

Malykhina, E. (2006) Nokia Wants Your Cell Phone to Tell You Where You Are. New York Times, October 9, 2006.

May, A.D. (1990) Traffic Flow Fundamentals. Prentice-Hall, ISBN 0-13-926072-2.

Maybeck, P.S. (1979) Stochastic Models, Estimation, and Control. Volume 1, Academic Press, New York.

Mehra, R.K., (1972) Approaches to Adaptive Filtering. IEEE Transactions on Automatic Control, Vol. 17, pp. 693-698.

Meyer, M. and E.J. Miller (2001) Urban Transportation Planning. McGraw-Hill, 2nd Edition.

Murtagh, F., and A.E. Raftery (1984) Fitting Straight Lines to Point Patterns. Pattern Recognition, Vol. 17, No. 5, pp. 479-483.

Nam, D.H., and D.R. Drew (1996) Traffic Dynamics: Method for Estimating Freeway Travel Times in Real Time from Flow Measurements. ASCE Journal of Transportation Engineering, Vol. 122, No. 3, pp. 185-191.

Nanthawichit, C., Nakatsuji, T., and H. Suzuki (2003) Application of Probe-Vehicle Data for Real-Time Traffic-State Estimation and Short-Term Travel-Time Prediction on a Freeway. Transportation Research Record, No. 1855, pp. 49-59.

Nelson, P., and P. Palacharla (1993) A Neural Network Model for Data Fusion in ADVANCE. Proceedings of Pacific Rim Conference, Seattle, Washington, pp. 273-243.

Newell, G.F. (1993a) A Simplified Theory of Kinematic Waves in Highway Traffic Part I: General Theory. Transportation Research, Part B, Vol. 27B, No. 4, pp. 281-287.

Newell, G.F. (1993b) A Simplified Theory of Kinematic Waves in Highway Traffic Part II: Queuing at Freeway Bottleneck. Transportation Research, Part B, Vol. 27B, No. 4, pp. 289-303.

Newell, G.F. (1993c) A Simplified Theory of Kinematic Waves in Highway Traffic Part III: Multi-Destination Flows. Transportation Research, Part B, Vol. 27B, No. 4, pp. 305-313.

Oh, J.S., Jayakrishnan, R. and W. Recker (2002) Section Travel Time Estimation from Point Detection Data. Institute of Transportation Studies, the University of California at Irvine, Accessed from http://www.its.uci.edu/its/publications/papers/CTSS/UCI-ITS-TS-WP-02-14.pdf, on August 8, 2010.

Oliver, B. (2005) Greenhouse Gas Emissions and Vehicle Fuel Efficiency Standards for Canada, Pollution Probe. Accessed from http://www.resourcesaver.org/file/toolmanager/CustomO16C45F62799.pdf in March 2007.

Ou, Q., Van Lint, J.W.C., and S.P. Hoogendoorn (2008) Piecewise Inverse Speed Correction by Using Individual Travel Times, Transportation Research Record, No. 2049, pp. 92-102.

Ozbay, K (1996) A Framework for Dynamic Traffic During Non-Recurrent Congestion: Models and Algorithms. PhD dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Park, D. and L.R. Rilett (1998) Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks. Transportation Research Record, No. 1617, pp. 163-170.

Park, D. and L.R. Rilett (1999) Forecasting Freeway Link Travel Times with a Multilayer Feedforward Neural Network. Computer Aided Civil and Infrastructure Engineering, Vol. 14, pp. 357-367.

Quandt, R.E., (1958) The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. Journal of the American Statistical Association, Vol. 53, pp. 873-880.

Rakha, H., and B. Crowther (2002) Comparison of Greenshields, Pipes, and Van Aerde Car-Following and Traffic Stream Models. Transportation Research Record, No. 1802, pp. 248-262.

Rice, J., and E. van Zwet (2004) A Simple and Effective Method for Predicting Travel Times on Freeways. IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 3, pp. 200-207.

Richards, P. I. (1956) Shock Waves on the Highway. Operations Research, Vol. 4, No. 1, pp. 42-51.

Rilette, L., and D. Park (1999) Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks. Transportation Research Board, 78th Annual Meeting CD-ROM, Washington D.C.

Roess, R.P., Prassas, E.S., and W.R. McShane (2004) Traffic Engineering. 3rd Edition, Prentice-Hall.

Roorda, M. J., Warkentin, C., Masters, P., and B. Sherman (2009) A System for Real-time Monitoring of Truck GPS, Truck Engine and Bluetooth Device Data on an Urban Freeway. Proceedings of the 44th Canadian Transportation Research Forum Annual General Meeting, Victoria, British Columbia, Canada.

Sanwal, K.K., and J. Walrand (1995) Vehicles as Probes, California PATH Working Paper UCB-ITS-PWP-95-11, the University of California at Berkeley.

Schrank, D., and T. Lomax (2005) The 2005 Urban Mobility Report. Texas Transportation Institute, The Texas A&M University System, Accessed from http://mobility.tamu.edu in April 2007.

Smith, B. (2006) Wireless Location Technology-Based Traffic Monitoring Demonstration and Evaluation Project - Final Evaluation Report, Smart Travel Laboratory Center for Transportation Studies University of Virginia.

Son, B. (1996) A Study of G.F. Newell's "Simplified Theory of Kinematic Waves in Highway Traffic. PhD Thesis, Department of Civil Engineering, University of Toronto, Ontario, Canada.

Spath, H. (1982) A Fast Algorithm for Clusterwise Linear Regression. Computing, Vol. 29, pp. 175-181.

Takada, H. (2006) Road Traffic Condition Acquisition via Mobile Phone Location Referencing. Ph.D. Thesis, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Tudor, L., H., Meadors, Alan, and R. Plant (2003) Deployment of Smart Work Zone Technology in Arkansas. Transportation Research Record, No. 1824, pp. 3-14.

US Department of Transportation (2008) Vehicle Infrastructure Integration Concepts. Accessed from http://www.its.dot.gov/vii/ in November 2008.

Van Aelst, S., Wang, X.S., Zammar, H., and R. Zhu (2006) Linear Grouping Using Orthogonal Regression. Computational Statistics & Data Analysis, Vol. 50, pp. 1287-1312.

Van Aerde & Assoc., Ltd (2002) INTEGRATION© RELEASE 2.30 FOR WINDOWS: User's Guide – Volume II: Advanced Model Features.

Van Aerde & Assoc., Ltd, (2002) INTEGRATION© RELEASE 2.30 FOR WINDOWS: User's Guide – Volume I: Fundamental Model Features.

Van Aerde, M. and H. Rakha (1995) Multivariate calibration of single regime speed-flow-density relationships. Proceedings of the Vehicle Navigation and Information Systems (VNIS) Conference, Seattle, WA.

Van Aerde, M., (1995) Single regime speed-flow-density relationship for congested and uncongested highways. Transportation Research Board, 74th Annual Meeting CD-ROM, Washington D.C.

Van Lint, J. W. C. and N.J. Van der Zijpp (2003) Improving a travel time estimation algorithm by using dual loop detectors. Transportation Research Record, No. 1855, pp. 41–48.

Van Lint, J. W. C., and S.P. Hoogendoorn (2007) The Technical and Economic Benefits of Data Fusion for Real-Time Monitoring of Freeway Traffic. In World Congress of Transportation Research, Berkeley, California.

Van Lint, J.W.C. (2004) Reliable Travel Time Prediction for Freeways. PhD thesis, Faculty of Civil Engineering and Geosciences, Transportation and Planning Section, Delft University of Technology.

Van Lint, J.W.C. (2006) Reliable Real-Time Framework for Short-Term Freeway Travel Time Prediction. ASCE Journal of Transportation Engineering, Vol. 132, No. 12, pp. 921-932.

Van Lint, J.W.C., Hoogendoorn, S.P., and H.J. Van Zuylen (2002) Freeway Travel Time Prediction with State-Space Neural Networks. Transportation Research Records, No. 1811, pp. 30-39.

Van Lint, J.W.C., Hoogendoorn, S.P., and H.J. Van Zuylen (2005) Accurate Freeway Travel Time Prediction with State-Space Neural Networks under Missing Data. Transportation Research Part C: Emerging Technologies, Vol. 13, pp. 347-369.

Vieth, E., (1989) Fitting Piecewise Linear Regression Functions to Biological Responses. Journal of Applied Physiology, Vol 67, No. 1, pp. 390-396.

Wang, Y., Papageorgiou, M., and A. Messmer (2006) RENAISSANCE- A Unified Macroscopic Model-Based Approach to Real-Time Freeway Network Traffic Surveillance. Transportation Research Part C, No. 14, pp. 190-212.

Worsley, K.J. (1983) Testing for a Two-Phase Multiple Regression. Technometric, Vol 25, No.1, pp. 35-42.

Wu, C.H., Wei, C.C., Su, D.C., Chang M.H. and J.M. Ho (2004) Travel Time Prediction with Support Vector Regression. IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No 4, pp. 276-281.

Yang, J.S (2005a) Travel Time Prediction Using the GPS Test Vehicle and Kalman Filtering Techniques. Proceedings of American Control Conference, 9-10 June, Portland, Oregon, USA.

Yang, J.S. (2005b) A Study of Travel Time Modeling via Time Series Analysis. Proceedings of the 2005 IEEE Conference on Control Applications, Toronto, Canada, August 28-31, 2005.

Yang, S., Hamedi, M., and A. Haghani (2005) Online Dispatching and Routing Model for emergency vechicles with area coverage contraints. Transportation Research Record, No 1923, pp. 1-8.

Yeon, J., and B. Ko (2007) Comparison of Travel Time Estimation Using Shockwave Analysis and Queuing Theory to Field Data along Freeways. Proceedings of IEEE International Conference on Multimedia and Ubiquitous Engineering, Seoul, South Korea.

Ygnace, J.L, and C. Drane (2001) Cellular Telecommunication and Transportation Convergence. Proceedings of 4th International IEEE Conference on Intelligent Transportation Systems.

Yim, Y., (2003) The State of Cellular Phones. California PATH Program, Inistitute of Transportation Studies, University of California, Berkeley. Accessed from http://www.path.berkeley.edu/PATH/publications/pdf/prr/2003/prr-2003-25.pdf in September 2007.

You, J., and T.J. Kim (2000) Development and Evaluation of a Hybrid Travel Time Forecasting Model. Transportation Research Part C: Emerging Technologies, Vol. 8, pp. 231-256.

Zhang, H.M., and W.H. Lint (2001) Some Recent Developments in Traffic Flow Theory. Proceedings of IEEE Intelligent Transportation System Conference, Oakland, California.