Real Time Prediction of Traffic Speed and Travel Time Characteristics on Freeways

by

Reza Noroozisanani

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, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner
Dr. Tony Z Qiu
Associate Professor, Dept. of Civil and Environmental Engineering
University of Alberta

Supervisor(s)
Dr. Bruce Hellinga
Professor, Dept. of Civil and Environmental Engineering

Internal Member
Dr. Liping Fu
Professor, Dept. of Civil and Environmental Engineering

Internal Member
Dr. James Craig
Associate Professor, Dept. of Civil and Environmental Engineering

Internal-external Member
Dr. Kumaraswamy Ponnambalam
Professor, Dept. of Systems Design Engineering

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

Travel time is one of the important transportation performance measures, and represents the quality of service for most of the facilities. In other words, one of the essential goals of any traffic treatment is to reduce the average travel time. Therefore, extensive work has been done to measure, estimate, and predict travel time. Using historical observations, traditional traffic analysis methods try to calibrate empirical models to estimate the average travel time of different transportation facilities. However, real-time traffic responsive management strategies require that estimates of travel time also be available in real-time. As a result, real time estimation and prediction of travel time has attracted increasing attention. Various factors influence the travel time of a road segment including: road geometry, traffic demand, traffic control devices, weather conditions, driving behaviors, and incidents. Consequently, the travel time of a road segment varies as a result of the variation of the influencing factors.

Predicting near-future freeway traffic conditions is challenging, especially when traffic conditions are in transition from one state to another (e.g. changing from free flow conditions to congestion and vice versa). This research aims to develop a method to perform real-time prediction of near-future freeway traffic stream characteristics (namely speed) and that relies on spot speed, volume, and occupancy measurements commonly available from loop detectors or other similar traffic sensors.

The framework of this study consists of a set of individual modules. The first module is called the *Base Predictor*. This module provides prediction for traffic variables while state of the traffic remains constant i.e free flow or congested. The *Congestion Detection Module* monitors the traffic state at each detector station of the study route to identify whether traffic conditions are congested or uncongested. When a congestion condition is detected, the *Traffic Propagation Module* is activated to update the prediction results of the *Steady-State Module*. The *Traffic Propagation Module* consists of two separate components: *Congestion Spillback* activates when traffic enters a congested state; *Congestion Dissipation* is activated when a congested state enters a recovery phase.

The proposed framework of this study is calibrated and evaluated using data from an urban expressway in the City of Toronto, Canada. Data were obtained from the westbound direction of the Gardiner Expressway which has a fixed posted speed limit of 90 km/hr. This expressway is equipped with mainline dual loop detector stations. Traffic volume, occupancy and speed are collected every 20 seconds for each lane at all the stations. The data set used in this study was collected over the period from January 2009 to December 2011.

For the *Steady-State Module*, a model based on Kalman filter was developed to predict the near future traffic conditions (speed, flow, and occupancy) at the location of fixed point detectors (i.e. loop detector in this study). Traffic propagation was proposed to be predicted based on either a static or dynamic traffic pattern. In the static pattern it was assumed that traffic conditions can be attributed based on the observed conditions in the same time of day; however, in the dynamic pattern, expected traffic conditions are estimated based on the current measurements of upstream and downstream detectors. The prediction results were compared to a naïve method, and it was shown that the average prediction error during the "change period" when traffic conditions are changing from free flow to congestion and vice versa is significantly lower when compared to the naïve method for the sample locations (approximately 25% improvement)

For the *Traffic Propagation Module*, a model has been developed to predict the speed of backward forming and forward recovery shockwaves. Unlike classic shockwave theory which is deterministic, the proposed model expresses the spillback and recovery as a stochastic process. The transition probability matrix is defined as a function of traffic occupancy on upstream and downstream stations in a Markov framework. Then, the probability of spillback and recovery was computed given the traffic conditions. An evaluation using field data has shown that the proposed stochastic model performs better than a classical shockwave model in term of correctly predicting the occurrence of backward forming and forward recovery shockwaves on the field data from the urban expressway. A procedure has been proposed to improve the prediction error of a time series model (*Steady-State Module*) by using the results of the proposed Markov model. It has been shown that the combined procedure significantly reduces the prediction error of the time series model.

For the real-time application of the proposed shockwave model, a module (*Congestion Detection Module*) is required to simultaneously work with the shockwave model, and identify the state of the traffic based on the available measurements. A model based on Support Vector Machin (SVM) was developed to estimate the current traffic state based on the available information from a fixed point detector. A binary model for the traffic state was considered i.e. free follow versus congested conditions. The model was shown to perform better compared to a Naïve model. The classification model was utilized to inform the *Traffic Propagation Module*. The combined model showed significant improvement in prediction error of traffic speed during the "Change Period" when traffic conditions are changing from free flow to congestion and vice versa.

Variability of travel speed in the near future was also investigated in this research. A continuous-time Markov model has been developed to predict the state of the traffic for the near future. Four traffic states were considered to characterize the state of traffic: two free flow states, one transition state, and

one congested state. Using the proposed model, we are able to predict the probability of the traffic being in each of the possible states in the near future based on the current traffic conditions. The predicted probabilities then were utilized to characterize the expected distribution of traffic speed. Based on historical observations, the distribution of traffic speed was characterized for each traffic state separately. Given these empirical distributions and the predicted probabilities, distribution of traffic speed was predicted for the near future. The distribution of traffic speed then was used to predict a confidence interval for the near future. The confidence interval can be used to identify the expected range of future speeds at a given confidence level.

Acknowledgements

I would like to thank my supervisor, Dr. Bruce Hellinga, for his insightful guidance, scholarly input, encouragement, financial support, and remarkable patience. It was my privilege and honor to work with Dr. Hellinga, and have him as my supervisor.

I appreciate the examining committee, Dr. Liping Fu, Dr. James Craig, Dr. Kumaraswamy Ponnambalam, and Dr. Tony Qiu for their helpful suggestions and insightful comments.

Further, I am grateful to my friends and colleagues at the University of Waterloo, Soroush Salek Moghaddam, Pedram Izadpanah, Hossein Zarei, Amir Hosein Ghods, Akram Nour, Wenfu Wang, Tae J. Kwon, Ehsan Bagheri, Babak Mehran, Shahin Karimi, Amir Zarinbal, Reza Golshan, Shahram Heydary, and Rita Hu for their friendship and advice.

I would like to extend my deepest gratitude to my parents, Farah Mollapour and Firooz Noroozi, for their unconditional love and support. Specially, I am extremely grateful for my mother's endless love and sacrifices. I would not be who I am today without her.

I want to thank my brother, Ali Noroozi, for his support and friendship throughout my life.

Finally, I would like to extend my loving gratitude to my wife, Narges Nazeri Rad, for her constant love and support during my PhD studies. This research would not have been possible without her encouragement, limitless patience, and sacrifices.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Traffic congestion is one of the major issues in large cities. Increasing population and population densities, urban sprawl, car ownership trends, and trends of increasing vehicle km travelled per capita suggest that traffic congestion and its consequences will challenge the quality of life more intensely in the future. For years, the development of transportation infrastructure was considered as the most appropriate response to satisfy the increasing demand. However, with the increasing cost of expanding roadway capacity greater emphasis has been placed on using various traffic and demand management strategies.

Congestion is a result of the interaction between transportation demand and infrastructure. Therefore, different strategies have been introduced which target at least one side of this interaction. Demand management strategies try to restructure the existing transportation demand. Some provide incentives to change trip scheduling, route, mode or destination, and some reduce the need for physical travel through more efficient land use, or transportation substitutes.

On the other hand, traffic management strategies focus on the existing infrastructure, and try to use it more efficiently. "Traffic management involves the allocation of infrastructure according to strategic operational and policy goals. These include efficiency, safety, environmental, economic and equity objectives." (European Commission 2009).

Traffic management strategies have evolved over the past years especially in terms of how they respond to the traffic demand; they include offline, real-time, and proactive approaches. Offline approaches are constructed on the basis of historical observations of the traffic situation; however, using traffic sensors, real-time strategies capture the existing traffic situation, and respond in real-time to optimize the usage of the facilities. In proactive traffic management approaches, the future traffic situation is predicted, and suitable strategies are implemented in order to avoid or delay the onset of congestion. In the course of their evolution, traffic management strategies rely on a variety of input parameters, and need to accurately estimate/predict them.

Travel time is one of the important transportation performance measures, and represents the quality of service for most of the facilities. In other words, one of the essential goals of any traffic treatment is to reduce the average travel time. Therefore, extensive work has been done to measure, estimate, and predict travel time. Using historical observations, traditional traffic analysis methods try to calibrate empirical models to estimate the average travel time of different transportation facilities. However, real-time traffic responsive management strategies require that estimates of travel time also be available in real-time. As a result, real time estimation and prediction of travel time has attracted increasing attention.

Various factors influence the travel time of a road segment including: road geometry, traffic demand, traffic control devices, weather conditions, driving behaviors, and incidents. Consequently, the travel time of a road segment varies as a result of the variation of the influencing factors.

1.1 Travel Time Definition

Travel time of an individual vehicle is defined over a certain segment of a roadway, and it is measured as the difference between the times that the vehicle enters and exits the segment. This concept can be illustrated on a space-time diagram. Figure 1-1 shows a schematic view of a space-time diagram along a segment of a freeway with length L_i. The trajectories of six different vehicles are illustrated in this figure. Vehicle v is the first vehicle that enters the segment, and the time taken to traverse the segment is denoted as τ_v.

The mean travel time is the arithmetic average of the travel time of the individual vehicles which traversed the segment in a certain time interval, and provides a measure of performance to characterize traffic condition of the road segment.

$$\tau_i^n = \frac{\sum_v \tau_{vi} \times \delta_v^n}{\sum_v \delta_v^n} \tag{1-1}$$

Where,

τ_i^n = Average travel time of segment i over time interval n,

τ_{vi} = Travel time of vehicle v along segment i of the roadway, and

δ_v^n = 1 if vehicle v enters in time interval n, and 0 otherwise.

It should be noted that the travel time of all individual vehicles which entered the road segment in time interval n is required to compute the mean travel time of the segment in that time interval (τ_v^n). This raises several issues:

- Most traffic surveillance devices capture the travel time of a sample of the vehicles. In this case, the computed average travel time is a sample mean of travel time, and should be considered as an estimate of the population mean travel time.
- There may be some vehicles which have entered the road segment during time interval n, but have not reached the end of the road segment during this interval. Therefore, the mean travel time cannot be computed until all vehicles have traversed the segment.

2
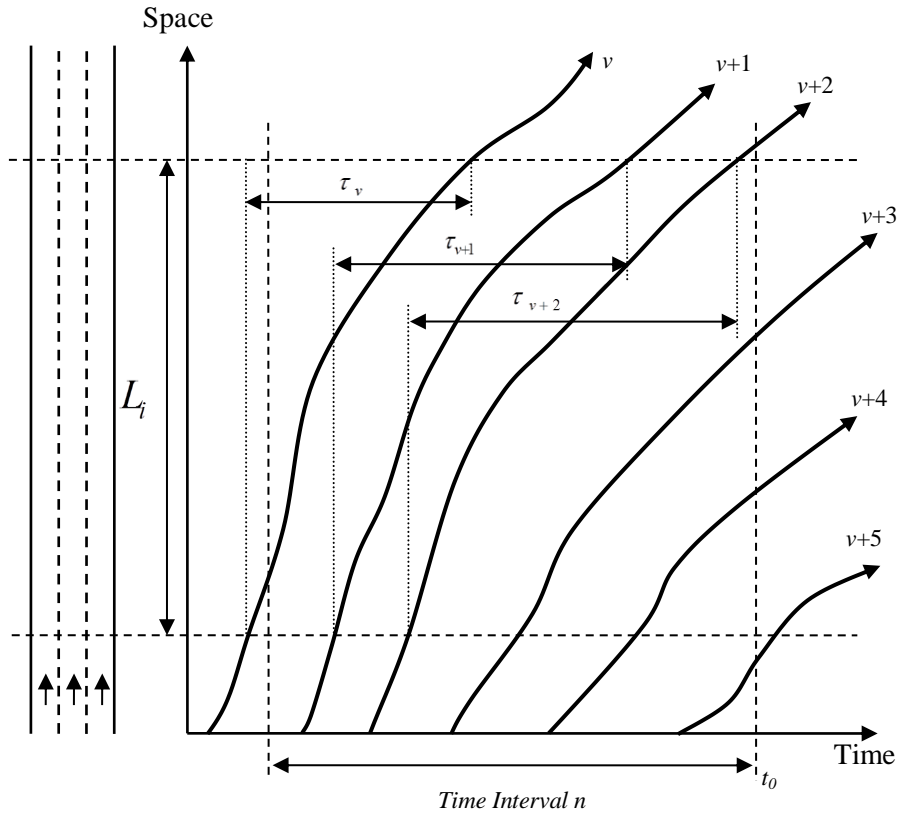
Figure 1-1: Space – time diagram for the definition of travel time (Izadpanah 2010)**.**

1.2 Travel Time Estimation vs. Prediction

Figure 1-2 illustrates a schematic view of a space-time diagram over a segment of a roadway. Time (the x-axis) has been discretized into time intervals with duration of Δ. This figure shows the trajectories of vehicles during two time intervals (time interval n and $n+1$) and the present time is denoted as t_{o}. The solid lines represent the part of the vehicle trajectories that has occurred. The dashed lines represent the future trajectories of the vehicles.

Using traffic sensors, we are able to measure traffic conditions during time interval n. Therefore, based on the sensors capabilities, some information about traffic characteristics in time interval $n+1$ would be available at the present time (time t_{o}). Estimation of mean travel time is the procedure which transforms the measured traffic data into an estimation of mean travel time for previous time intervals (e.g. time interval n). This procedure includes both of the following possible approaches:

- At time t_{o}: Using the measured traffic data, estimate the mean travel time of time interval n. The measured traffic data at this time may include the spot flow characteristics, travel time of vehicles which have traversed the entire length of the segment (first two vehicles in Figure 1-2), or the trajectory of all vehicles in time interval n (from time $t_{\mathrm{o}} - \Delta$ to t_{o}). It should be

3

noted that there are some vehicles which have not completed their travel to the end of the segment (vehicles 3 to 6 in Figure 1-2)

- At time t: At this time all of the vehicles which entered the road segment during time interval n have reached the end of the segment. Using the measured travel time or trajectory of all the vehicles (vehicles 1 to 6 in Figure 1-2), we can more accurately estimate the mean travel time of time interval n. However, this approach experiences a time delay $(t - t_0)$ to achieve the improved accuracy.

Prediction of mean travel time is the procedure that uses the measured traffic data to forecast the mean travel time for one or more future time intervals. In other words, in the prediction procedure, future traffic conditions must be modeled because they have not yet occurred.

Prediction horizon is one of the key elements in travel time prediction. Longer prediction horizons usually lead to less accurate results. Prediction horizon can be defined as follows:

$$T = \max \left[0, \left(n - n^* \right) \times \Delta \right] \tag{1-2}$$

Where:

$T =$ Prediction horizon,

$n^* =$ Current time interval,

$n =$ Time interval for which travel time is predicted, and

$\Delta =$ Length of the time interval.

Based on the length of prediction horizon, prediction approaches fall into different categories:

- Real time travel time prediction: the time horizon for this category is equal to zero. In other words, we are interested to predict the mean travel time of time interval $n = n^*$ at time t_0 (see Figure 1-2). This procedure includes the prediction of travel time of the vehicles which entered the segment in time interval n, but have not reached to the end of the segment at present time (vehicles 3 to 6 in Figure 1-2).

- Short term Travel time prediction: this category of approaches aims to predict mean travel time for the near future. Prediction horizons between 0 and 60 minutes are typically considered as short term prediction. The main focus of this research is on short term prediction.

- Long term travel time prediction: any prediction horizon longer than 60 minutes is considered as long term travel time prediction.
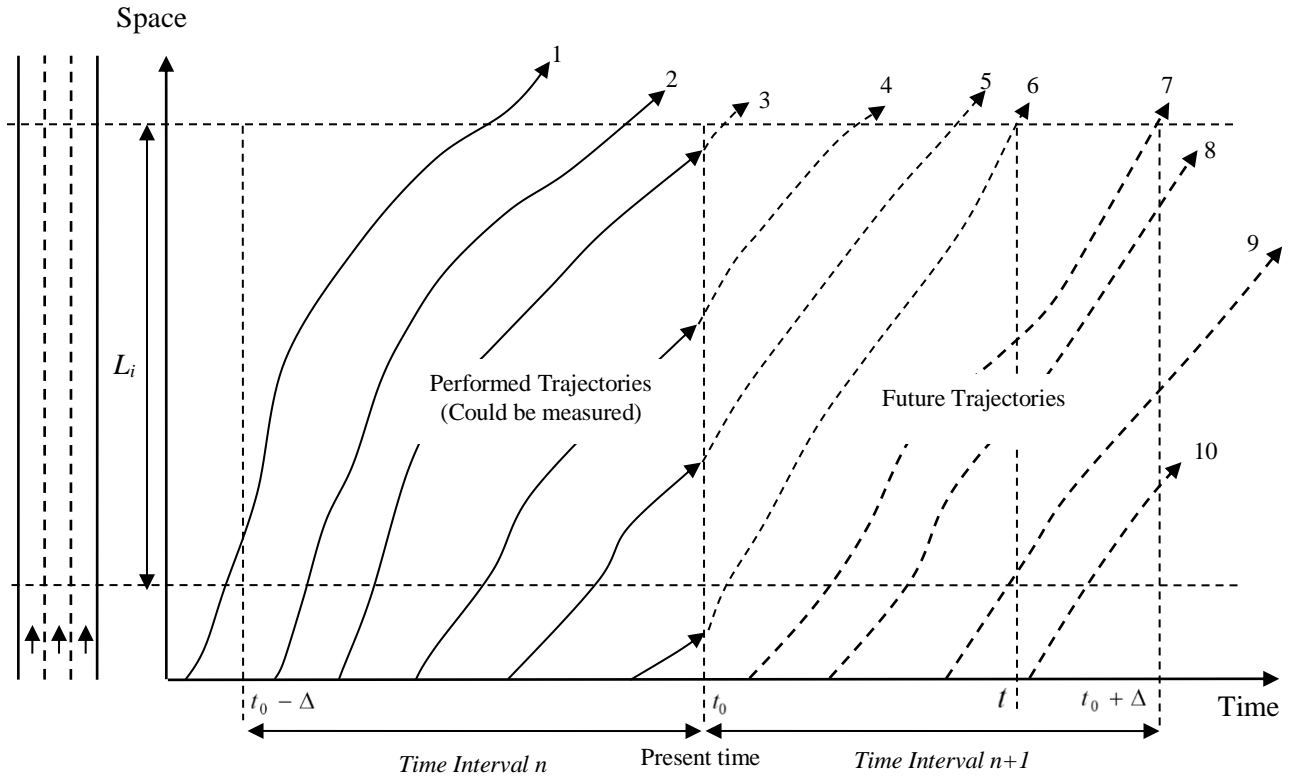
Figure 1-2: Travel time estimation vs. travel time prediction (Izadpanah 2010**).**

1.3 Traffic Sensors

1.3.1 Fixed Point Detectors

The sensors in this category provide data for a section of a roadway. Inductive loop detectors, microwave radars, and video image processors fall into this category. The components of these sensors are installed on the roadside, and no in-vehicle device is required.

Using this group of sensors, we can acquire macroscopic characteristics of traffic stream for a section of a roadway. Information such as volume, spot speed, occupancy, and vehicle classification would be provided by these detectors.

1.3.2 Automatic Vehicle Identification (AVI) Detectors

Sensors in this category uniquely identify vehicles or in-vehicle devices at a specific section of a roadway, and then re-identify them at some other downstream location. There are different technologies that can be implemented to identify vehicles. These technologies include automatic license plate recognition (ALPR), radio frequency identification (RFID), and Bluetooth detectors. For AVI detectors, roadside devices are required on both upstream and downstream locations. ALPR detectors do not need

5

any in-vehicle devices, and vehicle identification is on the basis of license plate number. However, RFID and Bluetooth detectors identify vehicles using an in-vehicle device; transponders for RFID and Bluetooth enabled devices for Bluetooth detectors are required.

Since these sensors uniquely identify vehicles, they are able to directly measure the travel time of individual vehicles along a segment of a roadway. Among AVI detectors, automatic license plate recognition sensors provide more data. Volume and occupancy are considered as the spot information which can be measured using ALPR sensors.

1.3.3 Floating Car Detectors

The sensors in this category are able to track vehicles at periodic intervals along the entire / part of their trip. GPS and cell phone tracking technologies are considered as floating car detectors. Although, in-vehicle devices are necessary in this category, there is no need for roadside equipment, and this attribute facilitates the deployment of this technology for wider geographic regions.

Since these sensors provide the trajectory of vehicles, they are able to directly measure not only the travel time of individual vehicles, but also the speed/acceleration profile of them.

1.4 Traffic Speed

Fixed point detectors are widely used by road agencies to collect historical characteristics of freeways. Most urban freeways in Canada and elsewhere are equipped with loop detectors, and record traffic volume, occupancy, and speed. Loop detectors provide valuable information regarding traffic conditions on a single section (spot location) of freeways.

Figure 1-3 shows a schematic view of a space-time diagram along a segment of a freeway. The segment is equipped with two loop detectors. The trajectories of seven different vehicles are illustrated in this figure. The speed of vehicle v measured at the location of detector d is denoted as u_{vd}.

The space mean speed at the location of each detector can be defined based on the speed of vehicles which pass over the detector during the time interval of the study, and provides a measure of performance to characterize traffic conditions of the road segment.

$$u_d^{(n)} = \frac{\sum_v \delta_{vd}^n}{\sum_v \frac{\delta_v^n}{u_{vd}}}$$

(1-3)

Where,

$u_d^{(n)}$ = Space mean speed at the location of detector d over time interval n,

u_{vd} = speed of vehicle v over detector d , and

δ_{vd}^{n} = 1 if vehicle v passes over detector d in time interval n , and 0 otherwise.

The space mean speed at the location of Detector 1 during time interval n is estimated based on the speed of vehicle $v+1$ to $v+4$. While the space mean speed at the location of Detector 2 is estimated based on the speed of vehicle $v-1$ to $v+2$ during the same time interval.
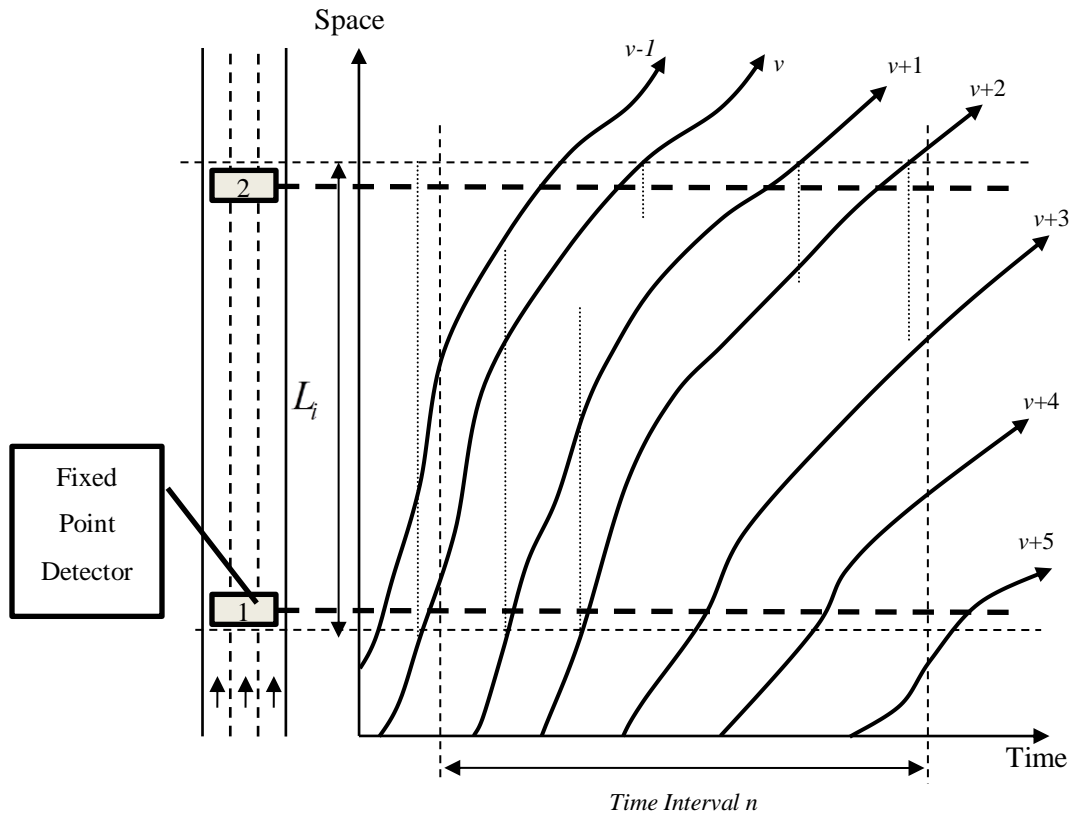
Figure 1-3: Space – time diagram for the definition of traffic speed

1.5 Travel Time Variability

The variability of travel time can be stated in terms of individual or mean travel time. Individual travel time variability reflects the travel time of different vehicles which enter a road segment in a certain time interval varies. However, the variability of mean travel time shows how a time aggregated index of travel time varies over the different time periods. Figure 1-4 shows the travel time variability for an expressway in Japan. The solid line illustrates the variation of hourly mean travel time for each hour of day. Furthermore, the range of individual travel time variation for each hour of day is also shown on this figure.

7
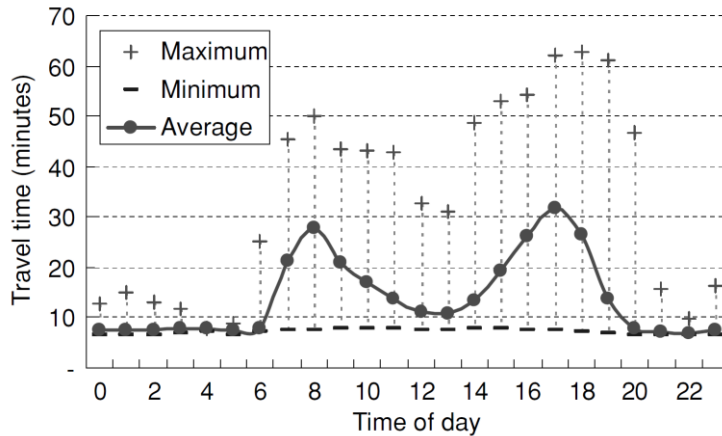
Figure 1-4: Individual and mean travel time variability for each hour of day in Hanshin expressway in Japan (Higatani et al. 2009)

Temporal variations of travel time are not limited to time of day variation. The mean travel time of a certain road segment may vary daily, monthly, or seasonally. Figure 1-5 illustrates the variability of travel time in 15-minute time of day periods on a freeway in the Netherlands. In this figure, the 5th, 10th, 25th, 50th (median), 75th, 90th, and 95th percentiles of travel time are shown. It is clear that the variability of travel time differs for different time of a day or different day of a week. Also, for a certain day of week e.g. Thursday and time of day e.g. 18:00, we may observe different values for 15-min period mean travel time.

A pattern can be recognized for the variation of travel time in Figure 1-5, and it suggests that some variations in travel time are predictable. In fact, mean travel time varies systematically over time. Variations between different times of day, different days of week, or different seasons of year are considered as systematic variation of travel time, and can be predicted using modeling procedures. However, we can observe variability in travel time even for a specific time period which is called random variability of travel time.

8

a) Thursdays

b) Fridays
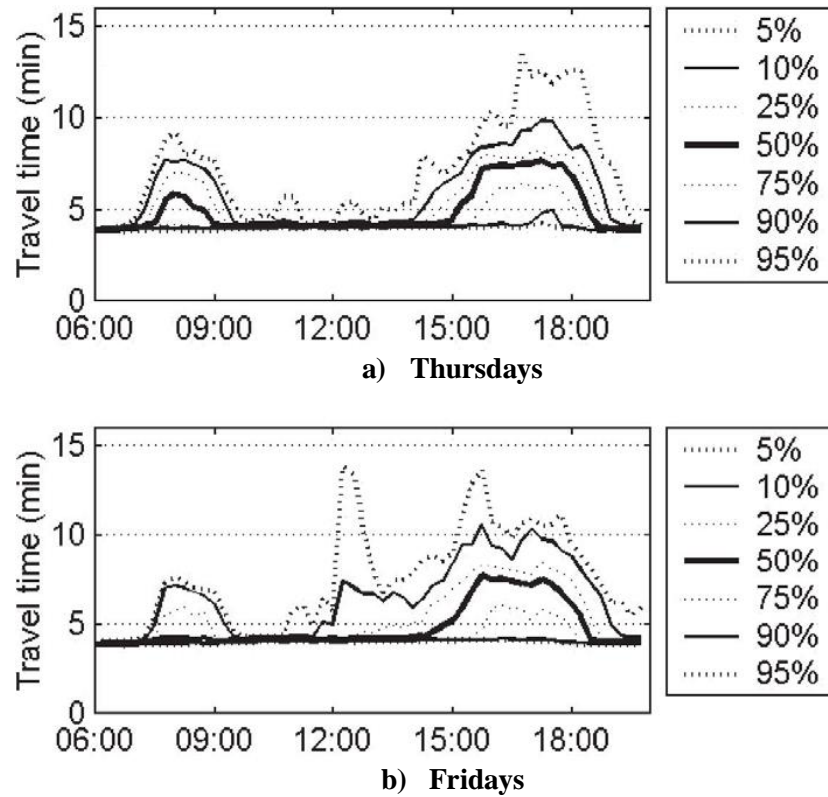
Figure 1-5: Variability of travel times in 2002 for 15-min time of day periods on northern Rotterdam Beltway A20 (van Lint and van Zuylen 2005)

In order to study the variability of travel time, we need to identify factors which influence travel time of a road segment. Travel time is a result of the interaction between demand and supply. Therefore, any factor which changes either the demand or the supply of a transportation network may change the travel time. However, even if we assume that demand and supply on a road segment are completely constant, we may observe variation in travel time as a result of variable driving behavior. Figure 1-6 illustrates important factors which influence the travel time of a road segment. These factors have been divided into three categories: factors which cause the fluctuation in demand, factors which cause the fluctuation in supply, and factors which cause the diversity of human behavior. Some factors such as weather conditions influence both demand and supply of a road segment. Some factors such as incidents result in the fluctuations of supply. However, these factors may indirectly influence the demand of a facility by changing the supply of the alternative or upstream facilities. Factors in the category of human behavior usually contribute to travel time variability by affecting the supply (capacity) of transportation facilities.

9
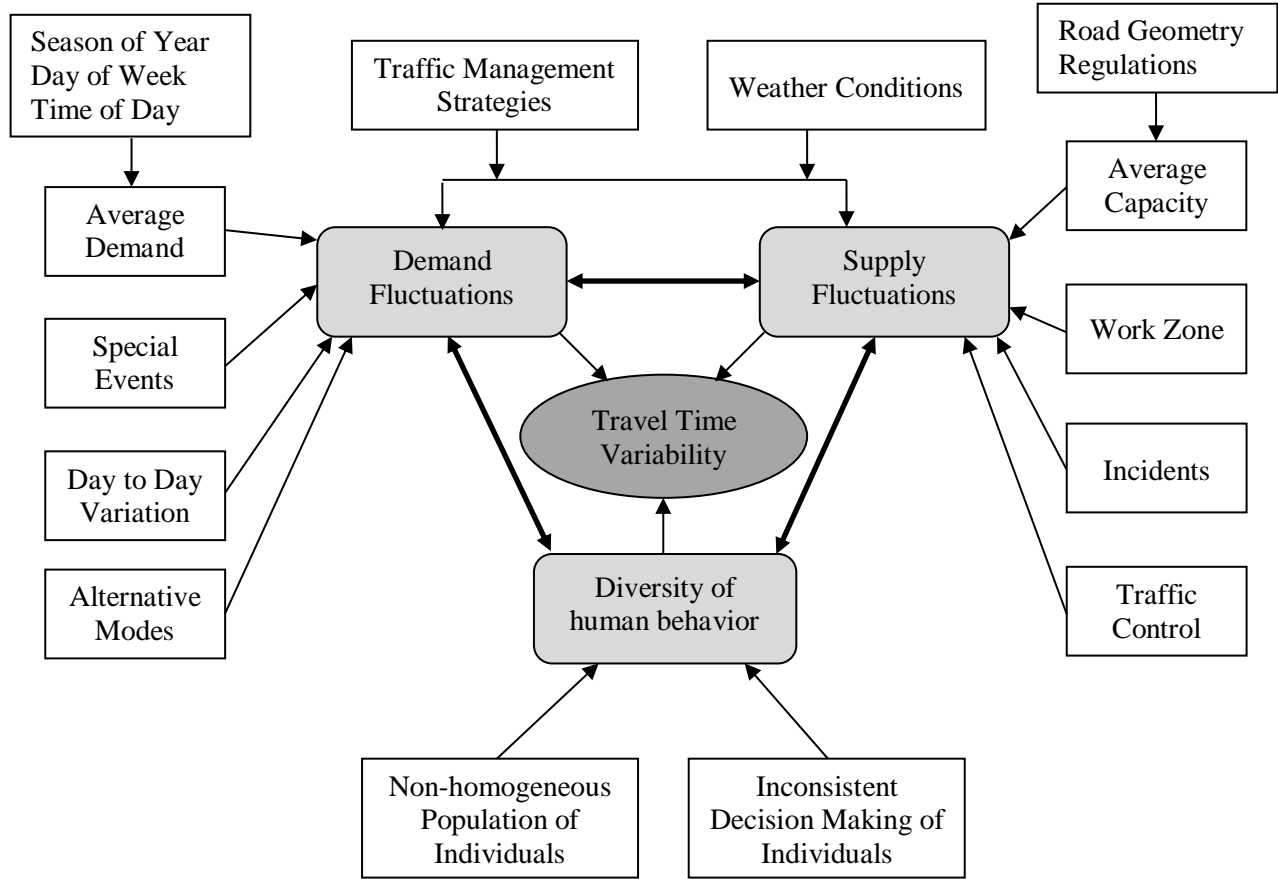
Figure 1-6: Factors influencing travel time variability

1.5.1 Segment vs. Route

Figure 1-7 illustrates a route consisting of S segments. Assuming travel time characteristics of all road segments are known, we are interested to study the travel time along route in this figure. It is clear that the mean travel time of the path can be simply obtained by addition of the mean travel time of all segments. However, the variation of travel time along the path cannot be simply explained based on the variation of each segment. In other words, the travel time over the path follows a distribution which is dependent on the correlation between travel times of different road segments. Therefore, the relationship between variability measures becomes more complicated compared to travel time itself. The following equations explain the characteristics of travel time along the route shown in Figure 1-7.

$$\tau_r = \sum_{i=1}^{S} \tau_i \tag{1-4}$$
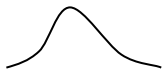
$$(\sigma_r)^2 = \sum_{i=1}^{S} (\sigma_i)^2 + \sum_{i=1}^{S} \sum_{\substack{j=1 \\ j \neq i}}^{S} \sigma_{i,j} \tag{1-5}$$

Where:

$\tau_r =$ Mean travel time along route r from origin 0 to destination d during analysis period,

$\sigma_r =$ Standard deviation of mean travel time along route r from origin 0 to destination d during analysis period, and

$\sigma_{i,j} =$ Covariance between mean travel time of segment i and segment j

Distribution			....		
Mean	τ_1	τ_2	....	τ_{S-1}	τ_S
Standard Deviation	σ_1	σ_2	....	σ_{S-1}	σ_S

o — Segment 1 — Segment 2 — Segment S-1 — Segment S — d

Figure 1-7: Schematic view of an OD path consisting of *S* road segments

1.6 Problem Statement

As mentioned earlier, travel time is one of the measures which characterizes the quality of service on freeways, and is a key input factor in most of the traffic management systems. Therefore, a rich body of research has been developed to explore the characteristics of travel time on freeways. Most of the previous research tried to investigate and predict mean travel time along segments of freeways.

Different methods have been developed to achieve reliable and accurate short-term predictions of mean travel time. These methods include but are not limited to traffic simulation models (Ben-Akiva et al. 2001; Muñoz et al. 2003; Tampère and Immers 2007), time series analysis (D'Angelo et al. 1999; Billings and Yang 2006), and neural network (Van Lint et al. 2005; van Hinsbergen et al. 2008). Some of these methods try to predict mean travel time using macroscopic or microscopic traffic behavior. However, some of them just seek to determine a trend in the data stream including historical and real time data (data derivative methods). In order to improve the performance of data derivative methods, some researchers tried to combine these two approaches and incorporate traffic behavior elements into data derivative methods. Among different methods to predict mean travel time, some seem to provide promising results, "but still not one of the methods can be considered the best method in any situation" (van Hinsbergen et al. 2007). Most of the methods perform well while the traffic condition is stationary. However, when the traffic state starts to change (from free-flow conditions to congestion or vice versa) they generally suffer from a lag to detect and predict this change and respond quickly. This suggests that the proposed algorithms should rely more on traffic elements, and try to predict traffic characteristics such as flow and density to come up with more reliable predictions of speed/travel time.

11

1.7 Goals and Objectives

This thesis focuses on the development of a method to predict the characteristics of travel time on freeways. The characteristics which are investigated in this study include mean and the variability of travel time. In this section, a list of questions which was investigated in the course of this thesis is provided, and each research question is followed by a corresponding research objective.

Data from fixed point sensors are available on freeways. This data has been mostly used to estimate traffic conditions on freeways. Many previous studies have been done to predict short-term traffic conditions and most of the proposed methods perform well while the traffic condition is stationary. However, when the traffic state starts to change (from free-flow conditions to congestion or vice versa) they generally suffer from a lag to detect and predict this change and respond quickly. This raises the following question:

Research Question 1: Using the available data on freeways (variables measured by fixed point sensors), how can we accurately predict traffic variables for the near future (future sensor measurements) while traffic condition is not stationary?

<u>*Research Objective 1:*</u> Develop a method to predict the average traffic speed during transitioning states (free flow to congested and vice versa) on freeways using data from fixed point sensors (e.g. loop detectors)

Another performance measure is the variability of traffic speed; variability will help to better characterize the uncertainty of prediction results as well as the reliability of traffic conditions.

Research Question 2: How can the variability of traffic speeds be predicted for the near future at the location of the fixed point detectors?

<u>*Research Objective 2:*</u> Develop and evaluate a method to predict the variability of traffic speed on segments of freeways using the results of the first objective.

Prediction of travel time along a route of a freeway which consists of some segments involves more challenges compared to a single segment. In general, the longer the length of the route on a freeway, the longer the prediction horizon tends to be, and consequently, there is more uncertainty associated with the prediction results.

Research Question 3: How can the segment level predictions of traffic speeds (resulting from Research Objectives 1 and 2) be extended to predict the travel time along a route of a freeway which consists of some smaller segments?

<u>*Research Objective 3:*</u> Develop and evaluate a method to extend the spot speed traffic predictions into travel time predictions along a route on freeways using the results of the first objective.

Chapter 2 provides a review of the existing literature. Chapter 3 describes the field data used for the calibration and validation of the models developed in this thesis and presents the overall structure for the proposed model.

Chapters 4 through 7 present and describe specific sub-models which are part of the overall proposed freeway travel time prediction model.

Chapter 8 describes the application of the proposed model to a freeway route in Toronto, Canada.

Chapter 9 provides conclusions, a summary of the thesis contributions, and recommendations for future work.

Chapter 2

Literature Review

This chapter documents the overall previous research on the prediction of traffic speed and travel time characteristics. General approaches and methodologies are discussed in this chapter, and specific models reviewed to develop the proposed methodology of this thesis are presented in the next chapters to better link to the proposed methodologies.

2.1 Short-term Traffic Prediction Models

2.1.1 Model-Based Approaches

Model-based approaches try to predict the characteristics of the traffic stream by implementing a traffic model which describes the interaction of vehicles and infrastructure in mathematical form. These approaches can be divided into different categories on the basis of the characteristics of the core model which is used to predict the future conditions.

2.1.1.1 Traffic Flow Theory Based Models

In this group of models, traffic is treated as fluid flow. Therefore, traffic is described in terms of flow, concentration and speed. The major assumptions of this theory are a) traffic is conserved and b) there is a one-to-one relationship between speed and density. These, assumptions lead to the following equations which are known as LWR equations (Gartner et al. 2001). The first equation is known as the conservation or continuity equation, and the second equation in conjunction with the equation $q = uk$ forms the fundamental relationship of traffic flow.

$$\frac{\partial q(x,t)}{\partial x} + \frac{\partial k(x,t)}{\partial t} = g(x,t)$$ (2-1)

$$u = f(k)$$ (2-2)

Where,

$q(x,t), \, k(x,t) =$ traffic flow and density at location x and time t, respectively,

$g(x,t) =$ the generation (dissipation) rate, and

$u =$ traffic speed.

Flow conservation equation can explain traffic propagation, queue formation, and queue dissipation on freeways. The boundary conditions of this differential equation determine the characteristics of the shockwaves.

Some research has used traffic flow theory to estimate the travel time on freeways (Nam and Drew 1999; Oh et al. 2003). Although this method is supported by a strong theoretical background, it requires extensive data including loop detector counts for each on-ramp or off-ramp. Furthermore, it needs complex side calculations to keep track of shockwaves (Daganzo 1994).

2.1.1.2 Cell transmission Models

The cell transmission model (CTM) has been proposed to eliminate the complex calculations of the classical traffic flow equations. CTM is a discrete approximation of the classical differential equations which explain the conservation of traffic flow (Daganzo 1994; Daganzo 1995). In CTM, each link is divided into some homogeneous "cells". The length of each cell is determined to be equal to the distance traveled in one simulation time step at free flow speed. CTM assumes that the relationship between flow (q) and density (k) can be expressed of the form illustrated in Figure 2-1, and approximates the LWR equations with a set of difference equations where the current conditions (state of the system) are updated for each simulation step (Daganzo 1995).

$$n_i(t+1) = n_i(t) + y_i(t) - y_{i+1}(t) \tag{2-3}$$

$$y_i(t) = \min\{n_{i-1}(t), Q_i(t), \delta[N_i(t) - n_i(t)]\} \tag{2-4}$$

Where,

$n_i(t) =$ the number of vehicles contained in cell i at time t,

$y_i(t) =$ the inflow to cell i in time interval $(t, t+1)$,

$Q_i(t) =$ the maximum number of vehicles that can flow into cell i when the clock advances from t to $t+1$.

$N_i(t) =$ the maximum number of vehicles that can be present in cell i at time t, and

$\delta = w/v$ (Refer to Figure 2-1)
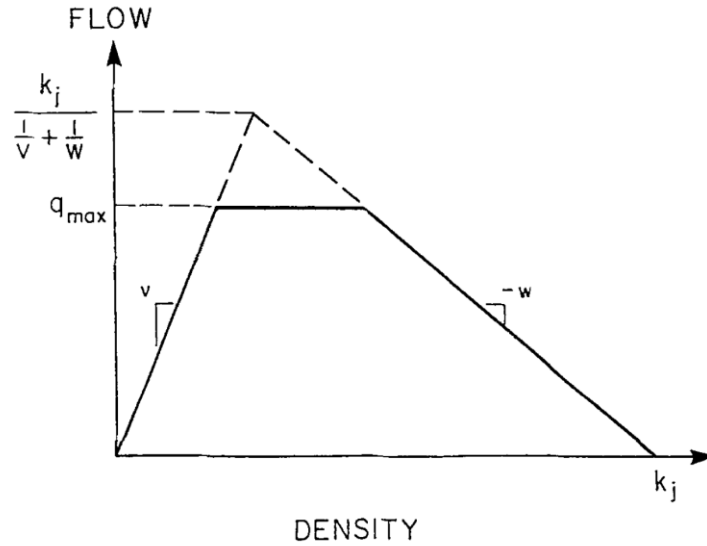
Figure 2-1: The equation of state of the cell transmission model (Daganzo 1995)**.**

CTM has been used as a simulation tool to estimate/predict traffic states by several researchers (Muñoz et al. 2003; Tampère and Immers 2007; Muñoz et al. 2004). Using CTM, Tampère and Immers (2007) proposed a method to predict travel time in real-time. Furthermore, CTM has been used to evaluate traffic management strategies on freeways (Lo 2001).

Estimating the traffic state using CTM involves less computational effort compared to the classical LWR equations. However, there is still a need to deploy traffic sensors for all on-ramps and off-ramps along the study route.

2.1.1.3 Traffic Simulation Models

Traffic simulation models are used to predict/estimate traffic state in the study area. Traffic simulators include a wide range of models, and are divided into macroscopic, mesoscopic, and microscopic approaches on the basis of the level of details that they consider.

In macroscopic traffic simulation models, only the global variables of the study network such as speed, density, and flow are considered. The LWR and CTM which were discussed earlier are considered as macroscopic approaches.

On the other side, microscopic approaches model individual vehicles in the network, and the interaction between them. They use driving behavior models to simulate the individual behavior of every vehicle in the system. Since this group of models consider extensive amount of details to estimate/predict traffic state, there are numerous parameters which need to be calibrated for these models. Furthermore, the relatively greater computational demands for microscopic models imposes challenges for using this group of models for real-time applications (Chrobok et al. 2004; Wahle and Schreckenberg 2001).

16

In mesoscopic approaches, the global characteristics of the network are estimated using macroscopic theory; then, the movement of the individual vehicles is simulated based on the macroscopic variables of the traffic stream. Mesoscopic models provide more information compared to macro-simulation models. However, they are more computationally efficient compared to micro-simulation models (Ben-Akiva et al. 2001).

2.1.1.4 Trajectory Method

Izadpanah (2010) proposed a method to estimate the travel time of a route of a freeway using the loop detector measurements. Based on the measurements of the spot speed, this method estimates a trajectory for the average vehicle traversing along the route. Then the travel time of the average vehicle is calculated by projecting the built trajectory on the time axis.

Suppose that a route consists of n segments which are separated by loop detectors. To estimate the average travel time along this route for vehicles entering the route at time n, the trajectory method estimates the time that this group of vehicles pass each of the loop detectors. Then the average speed of traveling along each segment is computed as the average spot speed of the loop detectors located on each side of this segment for the estimated passing time. This trajectory method, which is an iterative procedure, can be shown as follows (Izadpanah 2010):

$$\tau_T^{1,2}(t) = \frac{L^{1,2}}{\left(S_1(t) + S_2(t)\right)/2}$$

$$\tau_T^{2,3}(t) = \frac{L^{2,3}}{\left(S_2(t + \tau_T^{1,2}) + S_3(t + \tau_T^{1,2})\right)/2}$$

$$\vdots \hspace{6cm} \text{(2-5)}$$

$$\tau_T^{n-1,n}(t) = \frac{L^{n-1,n}}{\left(S_{n-1}(t + \sum_{i=1}^{n-2}\tau_T^{i,i+1}(t)) + S_n(t + \sum_{i=1}^{n-2}\tau_T^{i,i+1}(t))\right)/2}$$

$$\tau_T(t) = \sum_{i=1}^{n-1}\tau_T^{i,i+1}(t)$$

Where,

$\tau_T(t) = $ average travel time along the route associated with time interval t,

$\tau_T^{i,i+1}(t) = $ average travel time between loop detector t and t associated with time interval t.

2.1.2 Data Driven Approaches

Unlike model-based approaches which predict traffic state as a result of a traffic process, the focus of the data driven approaches is on establishing a relationship between the future traffic state and the present/past observations of traffic variables. This relationship between input and output variables is

solely on the basis of the observed correlations, trends, and patterns in the existing data set. In this section, some of the data driven approaches which are more common in prediction of travel time are reviewed.

2.1.2.1 Artificial Neural Network

Neural networks transform the prediction/estimation problem to a form of a network which consists of different nodes, and these nodes are linked by weighted connectors. Different architectures have been used in Neural Network approaches. However, the nodes can be usually divided into input, output, and intermediate nodes.

Models based on Neural Network are able to learn complex and non-linear relationships. Therefore, they have the potential to capture the trend of complex processes. Neural Network methods have been widely used to predict/estimate traffic variables (Van Lint et al. 2005; van Hinsbergen et al. 2008; van Hinsbergen et al. 2009; Park and Rilett 1999), and promising results have been reported. The main advantage of this approach is that models based on Neural Network are able to flexibly model the relationships between traffic variables, and they are easy to implement in a way that they do not require to consider extensive details in traffic flow theory.

Since Neural Network models may consider extremely non-linear structures to capture the trend of the process, they may suffer from over-fitting problem. In this case, despite the low calibration error for the study data set, the prediction error of the model for new observations may become very high. Since the model is sensitive to the training data set, Neural Network approaches may suffer from the lack of transferability and generalizability (Zheng 2011).

2.1.2.2 k Nearest Neighbors

The k Nearest Neighbors (kNN) is a classifying method which characterizes the outcome of a given condition by the outcomes of the similar conditions from the historical observations. In this method, the prediction variable for the given condition is assumed to be best represented by those of its k nearest neighbors in the available historical data set. Each observation is characterized by a feature vector which summarizes the prevailing conditions. The closeness of observation is measured by the distance between their feature vectors.

kNN has a simple yet solid theoretical foundation, and methods based on kNN are easy to implement. Therefore, k Nearest Neighbors has been widely used to predict the traffic variables (You and Kim 2000; Ul et al. 2003; Robinson and Polak 2005).

Although kNN is easy to implement, it requires a large set of historical data, and the computation time increases sharply when the size of the historical database increases. Models based on kNN strongly rely

on the historical observations, and do not consider the dynamic of the system. In other words, they do not seem to perform well for extreme or non-recurrent cases.

2.1.2.3 Time Series Analysis

Time series data usually refers to a set of observations collected sequentially in time, and time series analysis try to establish a relationship between different observations of a given data set. Unlike many statistical models which are built on the assumption of independency of the observations, time series analysis works on the basis of the fact that the observations of a random variable for different time intervals are correlated.

Time series methods have been widely used to predict traffic variables (D'Angelo et al. 1999; Billings and Yang 2006; Al-Deek et al. 1998). These models predict traffic state or travel time on the basis of the observations of the previous time intervals. The main advantage of this group of models is capturing the existing trend in the data set. However, they have difficulties in predicting non-recurrent traffic congestions.

Traditional time series models are statically calibrated on the basis of a set of historical time series data. Therefore, they tend to focus on the average trend of the process instead of the expected trend of the current conditions. Dynamic time series approaches tackle this problem, and try to calibrate the model parameters in real time.

2.1.2.4 Kalman Filter

Kalman filter is a recursive algorithm which operates on streams of noisy input data to produce a statistically optimal estimate of the system state variables for the underlying process. The underlying dynamic of the system is assumed to be linear. In other words, the current state of the system can be linearly explained on the basis of the previous state variables. Furthermore, the available measurements of the system states for the previous time steps are used to improve the prediction results for the future.

Methods based on Kalman filter have the potential to identify the noise in the data and capture the actual trend of the process. This approach has been widely used to predict traffic variables (Nanthawichit et al. 2003; Wang and Papageorgiou 2005; Steven et al. 2003; Chu et al. 2005; Suzuki et al. 2000), and the performance of Kalman filter is reported to be promising. However, since the traffic state is not stationary, methods based on Kalman filter are reported to suffer from lack of accuracy when the traffic condition transforms from free flow to congestion.

Kalman filter is defined by two equations: stochastic state model and measurement equation. Stochastic state model expresses the dynamic of the system, and can be written as follows:

$$\mathbf{x}^{(n)} = C\mathbf{x}^{(n-1)} + D\mathbf{u}^{(n-1)} + \boldsymbol{\varepsilon}^{(n)} \tag{2-6}$$

Where,

$\mathbf{x}^{(n)} =$ vector of state variables at time n,

$\mathbf{u}^{(n)} =$ vector of control variables at time n,

$\boldsymbol{\varepsilon}_i^{(n)} =$ multivariate process noise at time n,

$C, D =$ model parameters.

The measurement equation defines the relationship between the measurement variables and the state variables, and it can be written as follows:

$$\mathbf{z}_i^{(n)} = H\mathbf{x}_i^{(n)} + \boldsymbol{\delta}_i^{(n)} \tag{2-7}$$

Where $\mathbf{z}^{(n)}$ is the vector of measurements at time n, $\boldsymbol{\delta}_i^{(n)}$ is multivariate measurement error at time n, and H is the model parameter.

Kalman filter assumes that the process and measurement errors are normally distributed with mean of zero, and their covariance matrix could be dependent on time.

$$\boldsymbol{\varepsilon}^{(n)} \sim N(\mathbf{0}, A^{(n)}), \boldsymbol{\delta}^{(n)} \sim N(\mathbf{0}, B^{(n)})$$

Practical use of Kalman filter consists of two main steps. In the first step, *a priori* estimate of the both process state and prediction covariance for the future time interval is provided. This temporary projection is made on the basis of the stochastic state model as follows:

$$\begin{aligned} \tilde{\mathbf{x}}^{(n+1)} &= C\hat{\mathbf{x}}^{(n)} + D\mathbf{u}^{(n)} \\ \tilde{\mathbf{P}}^{(n+1)} &= C\hat{P}^{(n)}C^T + A^{(n)} \end{aligned} \tag{2-8}$$

Where:

$\tilde{\mathbf{x}}^{(n+1)} = $ *a priori* estimate of the state vector for time $n+1$,

$\hat{\mathbf{x}}^{(n)} = $ *a posteriori* estimate of state vector for time n,

$\tilde{\mathbf{P}}^{(n+1)}$, $\hat{\mathbf{P}}^{(n)} = $ *a priori* and *a posteriori* error covariance matrix, respectively.

The second step is to update the *a priori* estimate on the basis of the new available measurements. This results in *a posteriori* estimate of the both process state and prediction covariance. The measurement update equations can be written as follows:

$$\begin{aligned} K^{(n+1)} &= P^{(n+1)}H^T(H\,P^{(n+1)}H^T + B^{(n+1)})^{-1} \\ \hat{\mathbf{x}}^{(n+1)} &= \tilde{\mathbf{x}}^{(n+1)} + \mathbf{K}^{(n+1)}(\mathbf{z}^{(n+1)} - \tilde{\mathbf{x}}^{(n+1)}) \\ \hat{\mathbf{P}}^{(n+1)} &= (\mathbf{I} - \mathbf{K}^{(n+1)})\tilde{\mathbf{P}}^{(n+1)} \end{aligned} \tag{2-9}$$

2.2 Traffic State Prediction

Most of the short-term travel time prediction methods which have been discussed perform well while the traffic condition is stationary. However, when the traffic state starts to change (from free-flow condition to congestion or vice versa) they generally suffer from a lag to detect and predict this change and respond quickly. Therefore, in order to tackle this problem, it is required to be able to predict the state of the traffic. Then using this additional piece of information, we may be able to provide more accurate predictions for the near future travel time.

Various methods have been implemented to predict the state of the traffic for the near future. Almost all of the methods which have been introduced earlier for prediction of travel time can be utilized to predict the state of the traffic. However, one of the methods which has been widely used to predict the state of the stochastic processes is Markov chain. Markov chain has a strong theoretical foundation, and has provided promising results for predicting the behavior of stochastic processes including traffic conditions (Dong and Mahmassani 2009; Yeon et al. 2008; Evans et al. 2001; Bickel et al. 2003).

Methods based on Markov chain define a state variable which explains the condition of the process, and utilize historical time series observations to simulate the mechanism governing the change in the state variable. The theoretical foundation of Markov chain has been explained as follows.

Generally speaking, $\{X(t), t \in T\}$ is called a stochastic process if $X(t)$ is a random variable (or random vector) for any fixed $t \in T$. T is referred to as the "index set", and is often interpreted in the context of time. As such, $X(t)$ is often called the "state of the process at time t". The index set T can be a continuum of values or can be a set of discrete points. The resulting process is called continuous-time or discrete-time, respectively.

A stochastic process $\{X^{(n)}, n = 0, 1, 2, ...\}$ is said to be a discrete-time Markov chain (MC) if the following conditions hold true:

(1) For any $n = 0, 1, 2, ...$, $X^{(n)}$ takes on a finite (countable) set of possible values.

(2) For any $n = 0, 1, 2, ...$,

$$\Pr(X^{(n+1)} = x^{(n+1)} \mid X^{(n)} = x^{(n)}, ... , X^{(0)} = x^{(0)}) = \Pr(X^{(n+1)} = x^{(n+1)} \mid X^{(n)} = x^{(n)})$$

The second condition implies that the conditional distribution of any future state $X^{(n+1)}$ given the past states $X^{(0)}, X^{(1)}, ..., X^{(n-1)}$ and the present state $X^{(n)}$ is independent of the past states. This relation is referred as the Markov property.

For any pair of states i and j, the transition probability from state i at time n to state j at time $n+1$ is given by $P_{n,i,j} = \Pr(X^{(n+1)} = j \mid X^{(n)} = i)$. Let $P_n = [P_{n,i,j}]$ be the associated matrix containing the elements $P_{n,i,j}$, referred to as the one-step transition probability matrix (TPM) from time n to time $n+1$.

If $P_{n,i,j} = P_{i,j} \ \forall i,j$ and $n = 0, 1, 2, ...$, then the Markov chain is referred as stationary or homogeneous. In this case, the one-step TPM becomes $P_n = P = [P_{i,j}]$, or

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i,0} & P_{i,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

2.3 Travel Time Variability

As discussed in Chapter 1, the average travel time explains some of the characteristics of traveling along a route, and the variation of the travel time may add more insight to the prediction of the future conditions of the study roadway. Therefore, study of the travel time variability has attracted increased attention. The research, which has been done to investigate reliability of travel time, falls into three categories. In the first group, researchers try to define an index of travel time distribution which is able to more consistently explain the variability of travel time in different traffic conditions (Higatani et al. 2009; Bogers et al. 2008).The second group of research in this area focuses on the incorporation of travel time variability indices into the current traffic/transportation analysis (SHRP2 2009; Noland and Polak 2002). This group which mostly deals with offline applications of travel time variability tries to improve the current analysis through incorporation of variability indices.

The third group of research focuses on estimation / prediction of travel time variability; some tries to estimate/predict an index of travel time variability; however, some targets the distribution of travel time, and tries to find an analytical or empirical form for this distribution (Tu 2008; Zheng 2011; Kharoufeh and Gautam 2004). However, the limited previous work is mostly on the estimation of travel time variability rather than the prediction of the variability of travel time.

2.3.1 Estimation of Travel Time Variability

Unlike the mean travel time, little research has been done to predict/estimate the variability of travel time. Some research has tried to estimate an index of variability, and some others tried to model the distribution of travel time. Some of the attempts to model the variability of travel time are summarized as follows:

Using long term historical observations, van Lint and van Zuylen (2005) investigated the variability of travel time on a freeway in Netherlands. They characterized the variability of travel time by two factors including time of the day and day of the week, and the empirical observations showed that the distribution of travel time is dependent on these factors. Furthermore, they suggested that the shape of the distribution

is affected by the traffic flow conditions, and presented four conceptual shapes for the distribution of travel time for free flow, congestion onset, congestion, and congestion dissolve conditions. Finally, they implemented an Artificial Neural Network model to regenerate the long term distribution of travel time for the study freeway. The ANN model estimated the different percentile of travel time (10th, 50th, 90th) for different time of the day and day of the week (van Lint and van Zuylen 2005).

Hollander and Liu (2008) tried to estimate the distribution of day-to-day travel time by using multiple simulation runs. They assumed that the results of each single run of a micro-simulation model can be considered as estimates of traffic conditions on a single day. To support this assumption, they tried to establish an analogy between a single run and a single day by calibrating the micro-simulation model in respect to travel time distribution. The simulation model used in this research was DRACULA, and they chose a set of 21 parameters including traffic behavior and variable demand parameters to calibrate the simulation model to regenerate travel time distributions similar to the distributions observed in the real world. Therefore, distribution of travel time on the basis of the day-to-day observations was compared to that of the multiple simulation runs using Kolmogorov–Smirnov test for each location and time period. Finally, an optimization problem was implemented to estimate the optimal set of parameters which are able to regenerate the distribution of travel time (Hollander and Liu 2008).

Dong and Mahmassani (2009) investigated the distribution of travel time using historical observations. They suggested that the travel time and traffic flow follow a joint distribution function. However, the strong correlation between the two variables makes it very challenging to come up with a closed form for the joint probability density function. Therefore, they discretized the flow rate into a set of different bins, and then suggested an empirical distribution of travel time for each range of traffic flow rate. Since for a given traffic rate, there exist two different traffic conditions: free flow and congestion, they proposed a Gaussian mixture model to express the distribution of travel time. The proposed Gaussian mixture model is a convex combination of two Gaussian distributions which represent free flow and congested conditions. Finally, using historical observations of travel time for each road location, they calibrated the parameters of each individual Gaussian distribution and the combination coefficients (Dong and Mahmassani 2009).

In another research, Tu (2008) tried to model the reliability of travel time along a route on a freeway. In this research, first a measure to quantify the reliability of travel time was proposed, and then an empirical model was calibrated to estimate this measure for different values of traffic flow. Tu suggests that the reliability/variability of travel time along a segment of a freeway can be explained as a function of traffic flow rate on that segment. He used the distance between the 90th and 10th percentile of the travel time to express the variability of travel time. However, for a given flow rate, traffic conditions could be either free flow or congested. To capture the effect of both traffic conditions, he suggested that the overall

variability index should be computed as a linear combination of variability index for free flow and congested traffic conditions. Then, using regression analysis, an empirical model to estimate the percentile of the travel time as a function of traffic flow was presented (Tu 2008).

2.3.2 Estimation of Route Level Travel Time Variability

As discussed in Chapter 1, the variation of travel time along a route on a freeway cannot be simply explained based on the variation of each segment. In other words, the travel time along the route follows a distribution which is dependent to the correlation between travel times of different segments. Suppose that route r consists of S segments; then the variance of travel time along this route can be written as follows.

$$(\sigma_r)^2 = \sum_{i=1}^{S}(\sigma_i)^2 + \sum_{i=1}^{S}\sum_{\substack{j=1 \\ j \neq i}}^{S}\sigma_{i,j} \qquad \textbf{(2-10)}$$

Where:

$\sigma_r =$ Standard deviation of mean travel time along route r, and

$\sigma_{i,j} =$ Covariance between mean travel time of segment i and segment j

Therefore, it is required to estimate the covariance between travel times of all the segments along the route. In order to avoid to estimate the covariance between segment travel times, some researchers have proposed to directly estimate the variance of travel time for the entire route. This approach appears to be attractive as it can to provide the estimates for the route variability without the need to estimate the variability for the segments, and it avoids the need to estimate the covariance. However, this approach assumes that the entire route is a homogenous path along a freeway, and all the segments share the average characteristics of the entire route. Furthermore, this method independently computes the travel time variability for each route, even when these routes have a significant overlap (Tu 2008).

Some authors assume that the coefficient of variation of a route can be approximated by the average coefficient of variations over all the segments. Although this method may provide promising results in some conditions, there is no strong theoretical support for the method (Kaparias et al. 2008; Rakha et al. 2006).

$$\frac{\sigma_r}{\tau_r} = \frac{\sum_{i=1}^{S}\dfrac{\sigma_i}{\tau_i}}{S} \qquad \textbf{(2-11)}$$

The Strategic Highway Research (SHRP) 2 proposed a method to predict the standard deviation of travel time (SHRP2 2009). Suppose that route r consists of two segments with identical length, and travel time characteristics ($L_a = L_b = L; \tau_a = \tau_b = \tau; \sigma_a = \sigma_a = \sigma$)

If the travel time distributions for these two segments are completely independent, then $\sigma_r = \sigma \times 2^{0.5}$, and if the travel times on these two segments are completely correlated, then $\sigma_r = \sigma \times 2$. Therefore, they concluded that the real situation falls between these two extreme cases, and they proposed a general formula as follows:

$$\sigma_r = \sigma \times 2^{1-\eta} \qquad\qquad\qquad\qquad \text{(2-12)}$$

Where $0 < \eta < 0.5$ represents the level of correlation between travel times on the segments of the route. However, it assumes that the entire route is a homogenous path along a freeway, and the level of correlation is a constant attribute of the route.

Given the available literature, it appears that the following areas are not fully explored:

1- Methods for near future travel time prediction using real time data do not seem to accurately predict travel time/traffic speed when traffic conditions are changing. Real time prediction of traffic states may be helpful to inform prediction methods to adjust for the potential changes in the travel time/traffic speed where traffic states are predicted to change.

2- Little research has been done to estimate travel time/traffic speed variability in real time. Distribution of traffic speed in the near future may be used to inform proactive traffic management approaches.

Chapter 3

General Framework and Field Data

Predicting near-future freeway traffic conditions is challenging, especially when traffic conditions are in transition from one state to another (e.g. changing from free flow conditions to congestion and vice versa). This research aims to develop a robust method to perform real-time prediction of near-future freeway traffic stream characteristics (namely speed) and that relies on spot speed, volume, and occupancy measurements commonly available from loop detectors or other similar traffic sensors. This chapter describes the overall structure of the proposed model and describes the field data set that is used throughout this thesis for calibrating and validating the individual model components.

3.1 General Framework

Suppose that we are making predictions along route r which consists of S loop detector stations. Figure 3-1 illustrates a space-time diagram of such a freeway route. At time t_0, the objective of the proposed framework is to predict the traffic variables along the route for the average vehicle ($v^{(n+1)}$) which represents the vehicles entering the route during time interval $n+1$. In other words, we are interested in regenerating the downstream traffic conditions while the average vehicle traverses the route. Assuming that it takes m time intervals for vehicle $v^{(n+1)}$ to reach the end of the route, it is required to predict traffic variables for the next m time intervals.

The framework of this research is proposed to fulfill the following requirements:

- The proposed models can be calibrated automatically to enhance the transferability of the proposed framework for use in different locations

- The prediction method is computationally efficient to provide the opportunity to be used for real-time applications.

- The proposed framework only utilizes the traffic measurements along the main route, and measurements from on-ramps and off-ramps are not necessary.
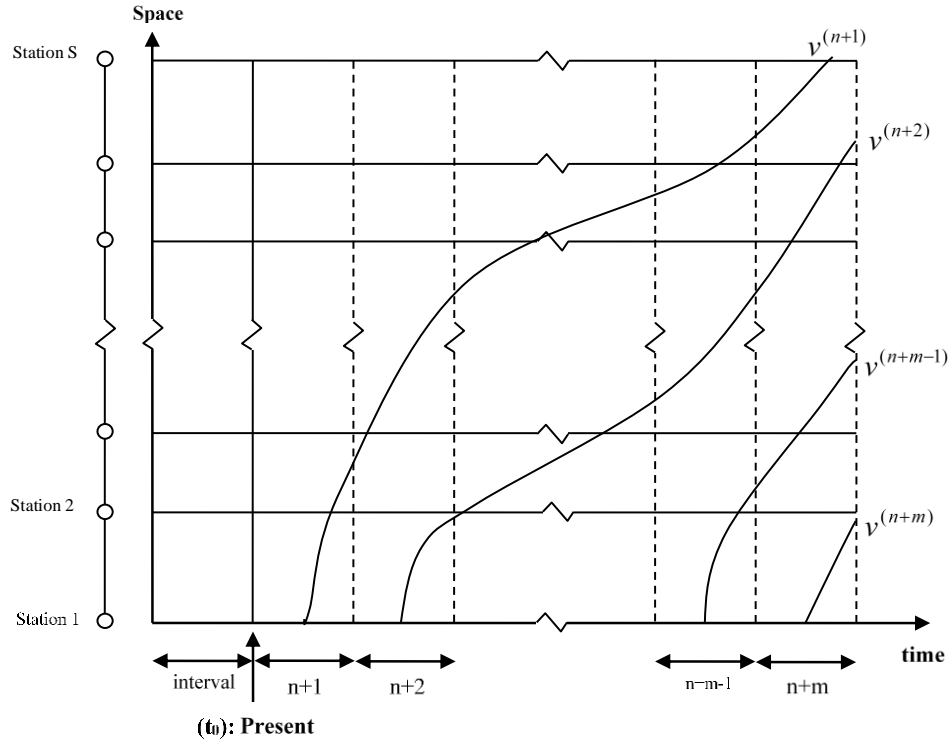
Figure 3-1: Schematic view of route _r_ consisting of _S_ detector stations

Figure 3-2 illustrates the proposed framework. The framework represents the application of a set of individual modules in each time interval. During the interval, a new set of traffic stream measurements are obtained from the traffic sensors (loop detectors). The suite of modules makes use of these new measurements, as well as historical data, to predict the near-future traffic conditions on each section of the freeway.

The first module is called the _Base Predictor_. This module is structured to provide near-future predictions of traffic conditions when traffic conditions remain stationary. This module provides prediction for traffic variables assuming traffic conditions remains stationary.

The _Congestion Detection Module_ monitors the traffic state at each detector station of the study route to identify whether traffic conditions are congested or uncongested. When a congestion condition is detected, the _Traffic Propagation Module_ is activated to update the prediction results of the _Base Predictor_. The Traffic Propagation Module consists of two separate components: _Congestion Spillback_ activates when traffic enters a congested state; _Congestion Dissipation_ is activated when a congested state enters a recovery phase.

Chapter 4 explains the details about the _Base Predictor_. Several time series models (Kalman filter) are proposed and evaluated using field data. The _Traffic Propagation Module_, discussed in Chapter 5,

27

describes a novel stochastic shockwave model which is used to update the predictions from Chapter 4. Chapter 6 describes the *Congestion Detection Module*, which is a classification model used to identify the state of the traffic in real-time. Finally, a stochastic model is developed to predict the empirical distribution of traffic speed for the near future; the details of this model is presented in Chapter 7.
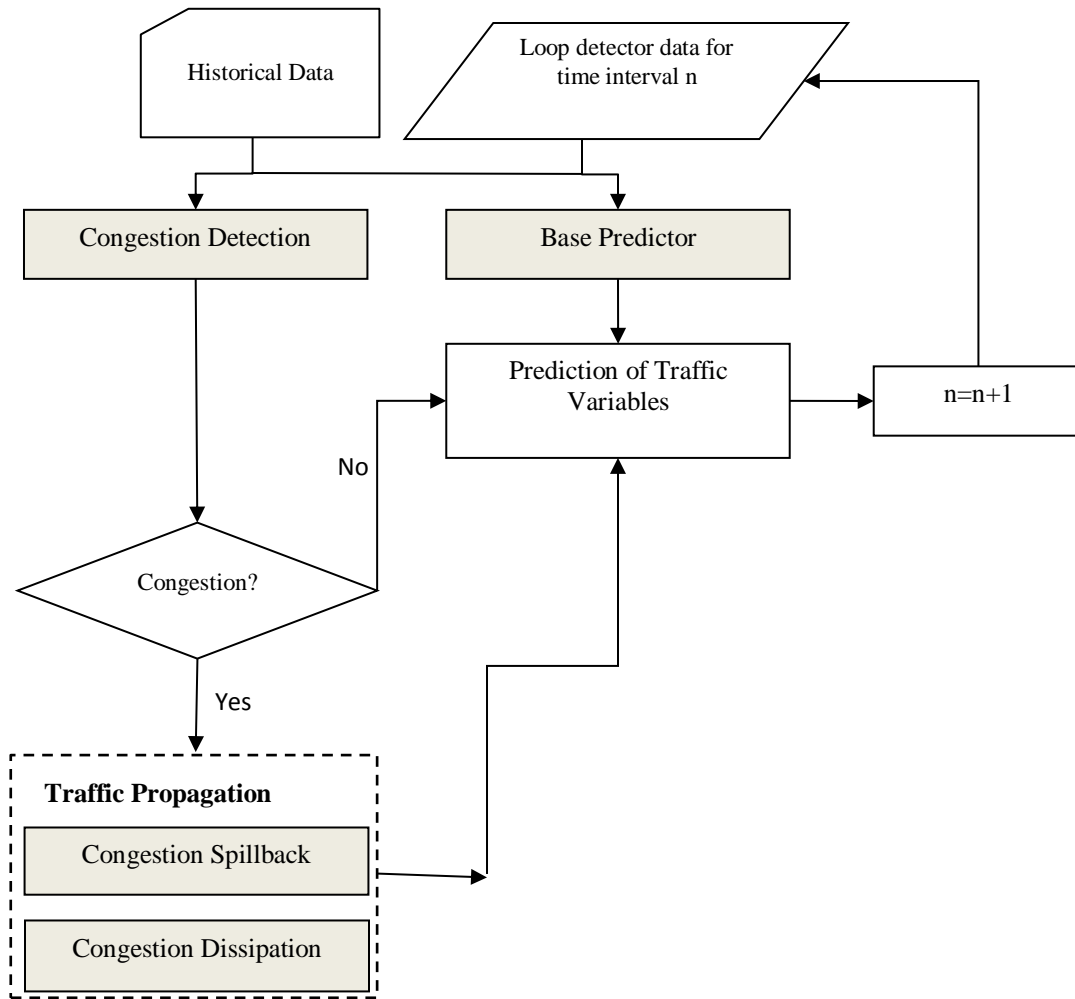
Figure 3-2: The proposed framework of the study

3.2 Field Data

The proposed framework of this study is calibrated and evaluated using data from an urban expressway in the City of Toronto, Canada. Data were obtained from the westbound direction of the Gardiner Expressway which has a fixed posted speed limit of 90 km/hr. This expressway is equipped with mainline dual loop detector stations. Traffic volume, occupancy and speed are collected every 20 seconds for each lane at all the stations. The data set used in this study was collected over the period from January 2009 to

December 2011. Figure 3-3 illustrates the location of the data collection detector stations in this study. The distances between the detectors are provided in Table 3-1. Loop detector stations are spaced approximately 500m apart. However, we selected data from only seven detector stations in order to create sections (freeway between two consecutive loop detector stations) of varying length. The loop detector stations are labeled from 1 to 7 starting from the upstream station. It should be noted that the study route contains sections with short length (550-590 m), medium length (1060-1120 m), and long length (3050 m).

The study route also contains sections with different access/egress characteristics; there are two access/egress points (on-ramp and off-ramp) between Stations 1 and 2, no access points between Stations 2 and 3, one on-ramp between Stations 3 and 4, no access point between Stations 4 and 5, one off-ramp between Stations 5 and 6, and one on-ramp between Station 6s and 7.

For the purpose of this study, the loop detector measurements are aggregated into longer time intervals (1 minute and 5 minute), and the lane traffic measurements are aggregated into station variables as follows; traffic speed is aggregated to provide the volume weighted space mean speed.

$$q^{(n)} = \frac{\sum_{l=1}^{L}\sum_{k=1}^{N_n} q_{lk}}{L \times N_n}$$

$$o^{(n)} = \frac{\sum_{l=1}^{L}\sum_{k=1}^{N_n} o_{lk}}{L \times N_n} \qquad \qquad \text{3-1)}$$

$$u^{(n)} = \frac{\sum_{l=1}^{L}\sum_{k=1}^{N_n} q_{lk}}{\sum_{l=1}^{L}\sum_{k=1}^{N_n} \frac{q_{lk}}{u_{ik}}}$$

Where,

$q^{(n)}$, $o^{(n)}$, $u^{(n)}$ = traffic flow, occupancy, and speed during aggregated time interval n,

q_{lk}, o_{lk}, u_{lk} = traffic flow, occupancy, and speed for lane l during 20 second interval k, and

N_n = number of 20 second intervals in aggregated time interval n,

L = number of lanes for the station of study.
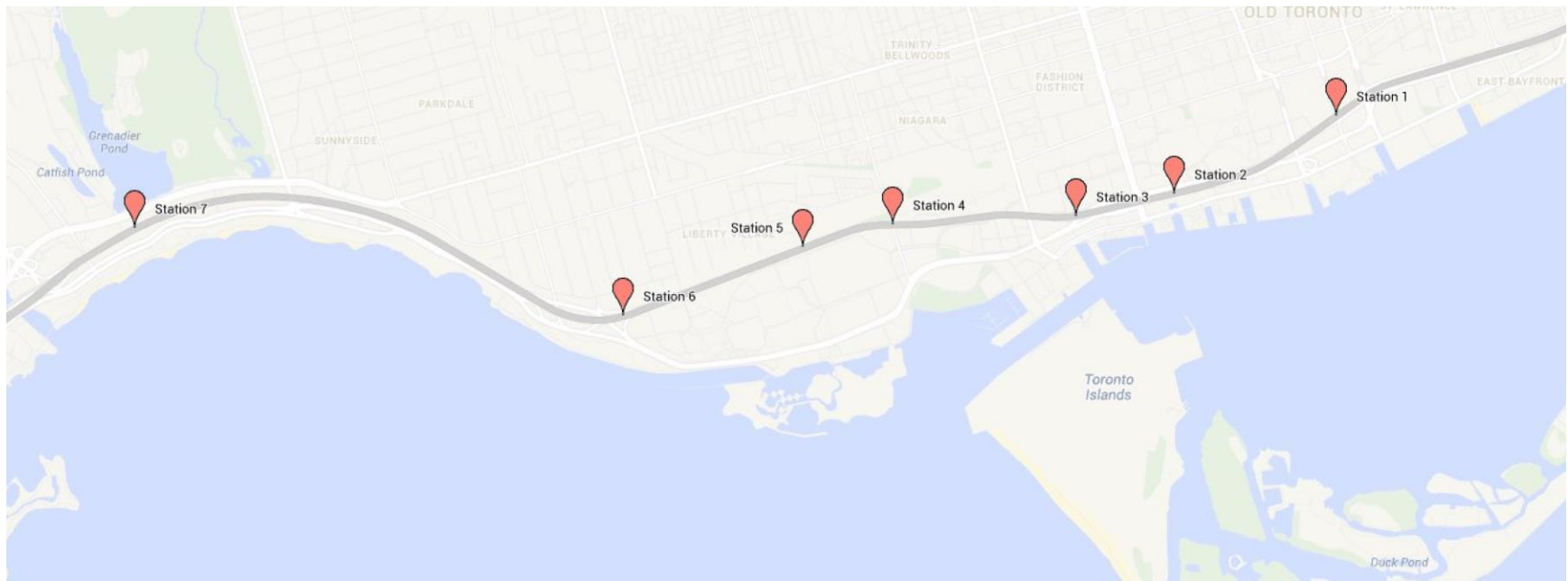
Figure 3-3: Extent of the study route and the location of loop detector stations (Map image source: Google)

Table 3-1: Distance between detector stations

Description	Station						
	1	2	3	4	5	6	7
Distance from the upstream detector (m)	0	1060	590	1070	550	1120	3050
Distance from Station 1 (m)	0	1060	1650	2720	3270	4390	7440

During the study days, there are cases that traffic data is missing or it is labeled as invalid by the measuring system; therefore, we used a simple imputation approach to treat the missing data for those cases. We assume that the occurrence of missing data is random meaning that the probability that an observation is missing is unrelated to the value of the traffic variables. Linear interpolation is used to estimate the value of missing traffic measurements. If traffic information on a given day is missing for either (i) more than 10% of the analysis period (5:00AM to 9:00PM) or (ii) one consecutive hour, the day was eliminated from the historical data pool. Traffic data for 567 days was utilized for this study.

In order to characterize the state of the traffic for the proposed modules of this research, we require a macroscopic model to define the fundamental relationship between traffic variables. We used the fundamental diagram proposed by Van Aerde (1995) in this research. The Van Aerde fundamental diagram defines the relationship between traffic speed and density as an S-shaped function, and involves the calibration of four parameters. The speed-density relationship for Van Aerde's model is presented as follows:

$$k = \frac{1}{c_1 + \dfrac{c_2}{u_f - u} + c_3 \times u} \tag{3-2}$$

Where, k and u are traffic density and speed, and c_1, c_2, and c_3 are model parameters which need to be calibrated as well as u_f (free flow speed).

Van Aerde (1995) proposed a multivariate optimization problem to automatically estimate the model's parameters based on loop detector data as follows:

$$Min \; E = \sum \left\{ \left(\frac{u_i - \hat{u}_i}{\tilde{u}} \right)^2 + \left(\frac{k_i - \hat{k}_i}{\tilde{k}} \right)^2 + \left(\frac{q_i - \hat{q}_i}{\tilde{q}} \right)^2 \right\}$$

$$S.T.:$$

$$\hat{k}_i = \frac{1}{c_1 + \dfrac{c_2}{u_f - \hat{u}_i} + c_3 \times \hat{u}_i} \quad \forall i \tag{3-3}$$

$$\hat{q}_i = \hat{u}_i \times \hat{k}_i \quad \forall i$$

$$\hat{q}_i, \hat{u}_i, \hat{k}_i \geq 0 \quad \forall i$$

Where, $\tilde{q}$, $\tilde{u}$, and $\tilde{k}$ are mean of flow, speed, and density of traffic.

Since traffic density is not measured by loop detectors, we have used measured occupancy values to approximate the traffic density. Suppose that N vehicles pass through a loop detector with effective detection length of L_d during a time interval of length T, and the length and speed of vehicle i are L_i

and u_i, respectively. Assuming $\bar{L} = L_d + L_i$ for each vehicle i, the relationship between traffic occupancy and density during the interval can be approximated as follows:

$$o = \frac{\sum_i \dfrac{L_d + L_i}{u_i}}{T}$$

$$o = \frac{\bar{L} \times \sum_i \dfrac{1}{u_i}}{T} = \bar{L}\frac{N/u}{T} = \bar{L}(q/u) = \bar{L} \times k$$

(3-4)

Figure 3-4 illustrates a calibrated fundamental diagram for data from a sample detector station along the study route. A separate curve is fitted for each station. Moreover, in order to consider the effect of seasonality in traffic condition, a separate curve is calibrated for each month of the year.
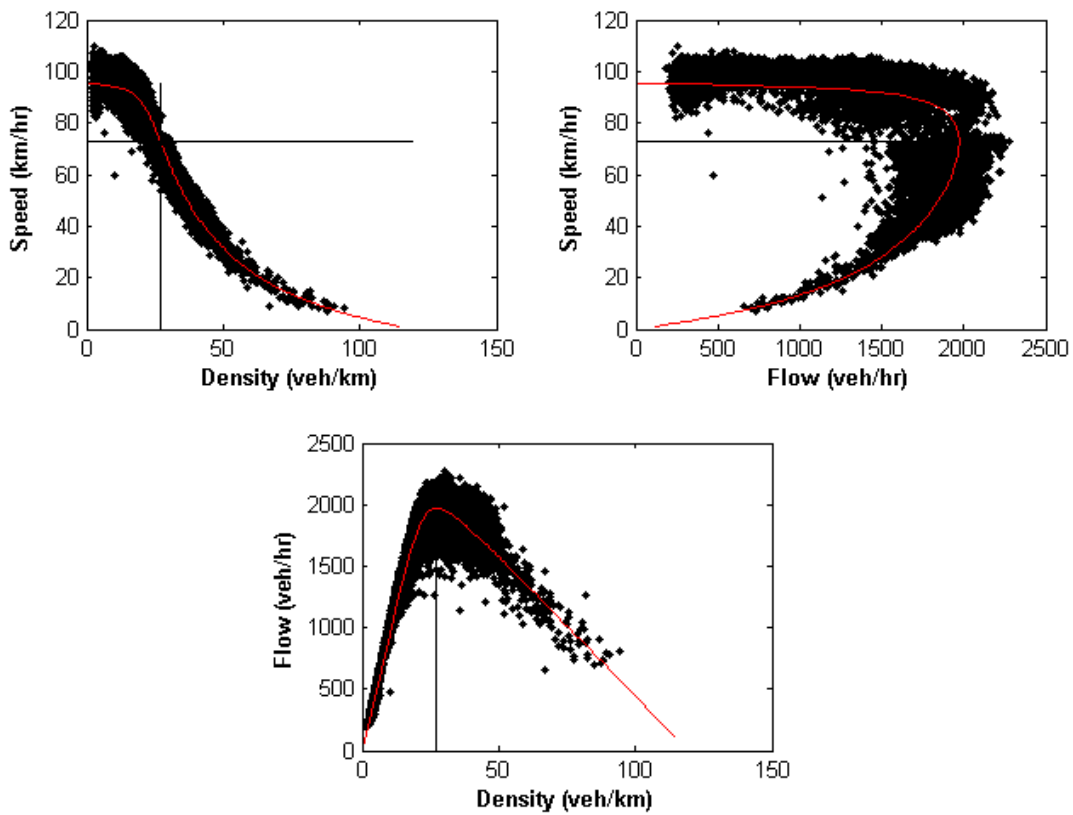
Figure 3-4: Fundamental traffic diagram for a sample detector station for month June

Traffic conditions along this section of Gardiner Expressway experience a wide range of traffic conditions including free flow and congested conditions. Figure 3-5 illustrates the observed 5-minute aggregated detector station speed as a function of location (detector station number) and time of day for a sample day. Darker blocks represent congested traffic conditions in this graph. On this particular day, there is very little severe congestion at Station 6 and the severe congestion observed at Station 7 seems to be local disturbances in traffic condition as the congestion did not propagate upstream. However, two distinctive congestion patterns can be observed; one in the morning and one in the afternoon. In the morning, traffic first breaks down at Station 5 at approximately 7 AM, and then it propagates to upstream stations (upstream stations break down later in time). The same pattern can be noted for the recovery phase. Traffic first recovers at the upstream station (Station 2) at shortly after 8 AM, and then downstream stations begin to recover one after another. This patterns of congestion forming and dissipation is typical of recurrent congestion, for which capacity remains constant, but demand varies (first increasing to exceed capacity and declining to below capacity).

Figure 3-6 and Figure 3-7 illustrates the space-time variations of detector occupancy and traffic volume for the same day. As expected, the congestion patterns exhibited by the speed data are also clearly identifiable in the detector occupancy data, but are not visible in the traffic volume data.
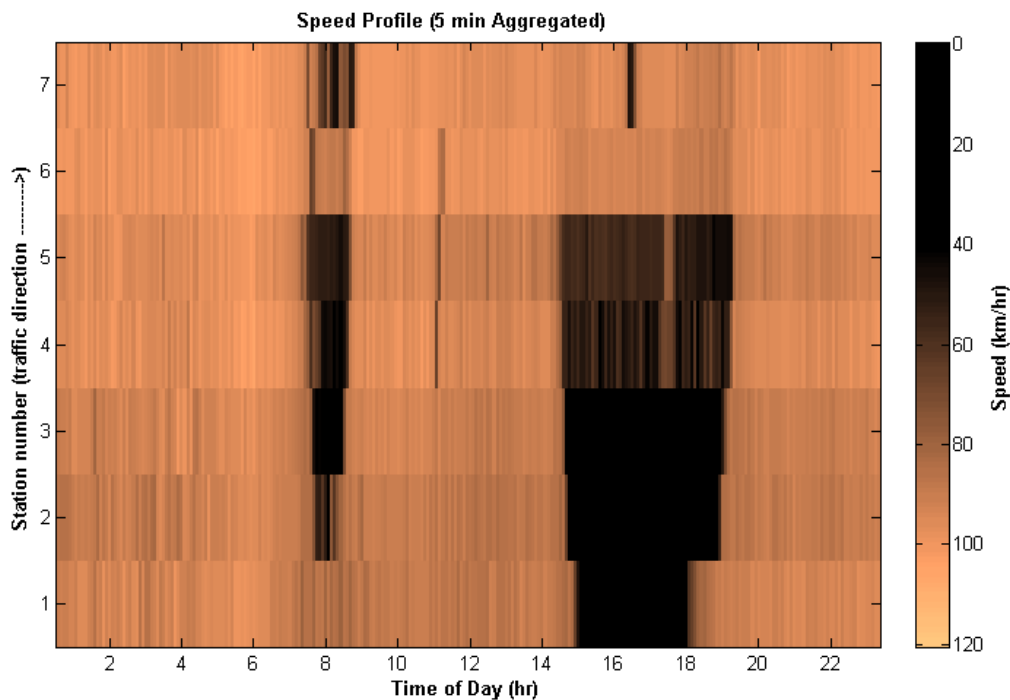
Figure 3-5: Speed profile along the study route for a sample day
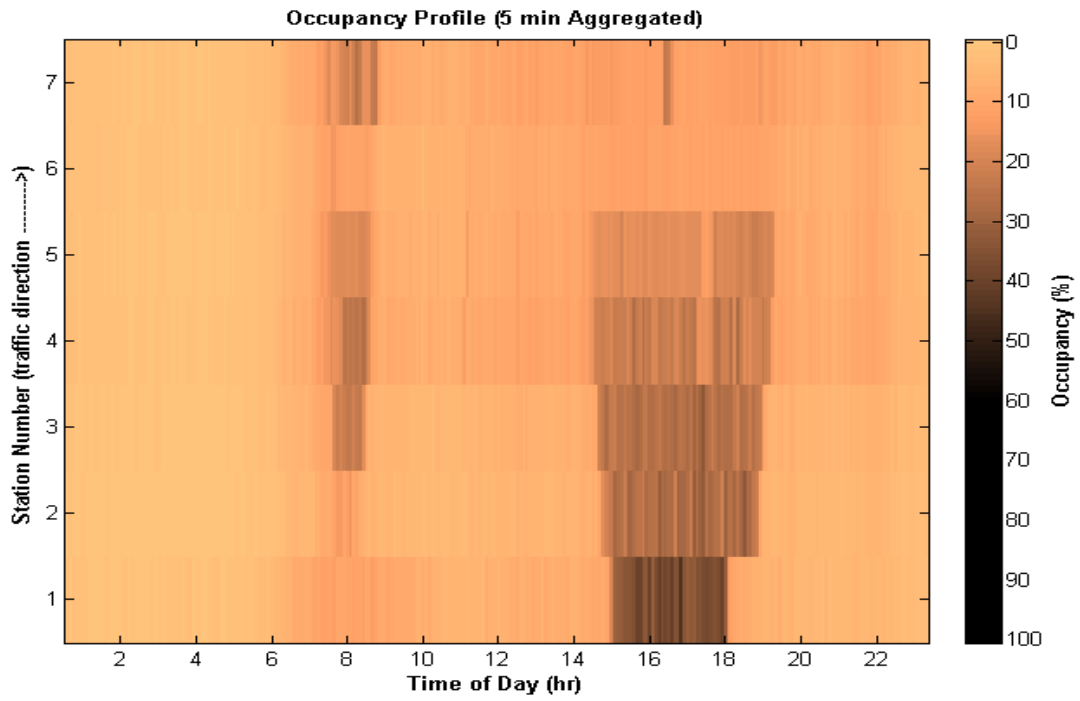
Figure 3-6: Occupancy profile along the study route for a sample day
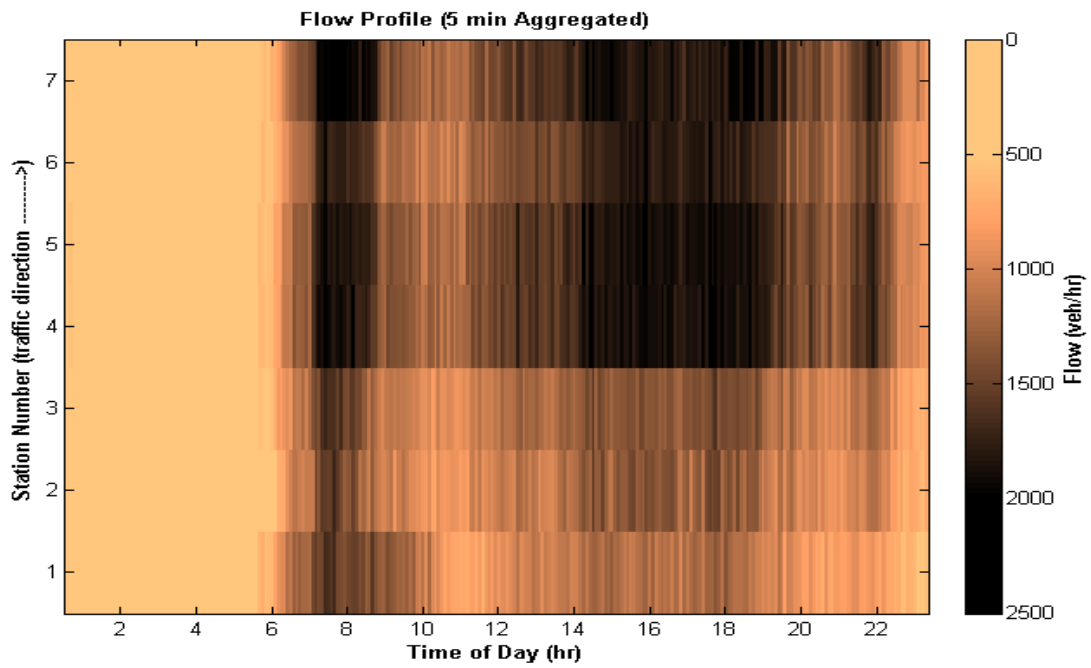
Figure 3-7: Traffic flow profile along the study route for a sample day

Chapter 4

Base Predictor Model

The state of the traffic is a stochastic variable which evolves over time for a given day. In other words, each day goes through a stochastic process. The traffic speed profile at any given location of a freeway may be affected by both the temporally varying pattern of the travel demand and the variation due to random events (e.g. weather, incidents, special events, etc.). Therefore, incorporating both the random and the historical patterns may lead to better predictions for traffic speed.

As discussed in Chapter 3, the main goal of the *Traffic Propagation Module* is to predict the speed of traffic when the traffic condition remains stationary (i.e. does not vary over time). During stationary traffic conditions, the traffic state on a given section of a freeway is not affected by traffic conditions on downstream or upstream segments. Therefore, for these conditions, the characteristics of the traffic stream on each segment may be explained by the recent observations of traffic stream on the same segment and the expected pattern of the travel demand for the near future.

As long as the state of the traffic remains stationary, the recent traffic stream measurements contain information implying that there is no evidence to believe that the traffic variables are going to change dramatically (e.g. change from speeds associated with the uncongested regime to speeds in the congested regime or vice versa) in the near future; in other words, the speed of traffic may deviates around the values which have been observed recently. Therefore, under a steady state assumption, traffic stream speed may follow a time series relationship. Therefore, time series methods have been widely used to predict traffic variables (D'Angelo et al. 1999; Billings and Yang 2006; Al-Deek et al. 1998). These models predict traffic state or travel time on the basis of the observations of the previous time intervals.

In this chapter, we have developed a model to predict the traffic stream speed for a given segment of a freeway based on recent measurements of the traffic stream characteristics of the same segment.

4.1 Background

Kalman filter is a recursive algorithm which operates on streams of noisy input data to produce a statistically optimal estimate of the system state variables for the underlying process. The underlying process of the system is assumed to be linear, and the available measurements of the system states for the previous time steps are used to improve the prediction results for the future.

Kalman filter is defined by two equations: stochastic state model and measurement equation. The stochastic state model expresses how the system evolves over time, and can be written as follows:

$$x^{(n)} = Cx^{(n-1)} + Du^{(n-1)} + \varepsilon^{(n)} \qquad\qquad \textbf{(4-1)}$$

where,

$x^{(n)} =$ state variable at time n,

$u^{(n)} =$ control variable at time n which is a deterministic term,

$\varepsilon_i^{(n)} =$ process noise at time n,

$C, D =$ model parameters.

The measurement equation defines the relationship between the measurement variable and the state variable, and can be written as follows:

$$z^{(n)} = Hx^{(n)} + \delta^{(n)} \tag{4-2}$$

Where $\mathbf{z}^{(n)}$ is the measurements of $x^{(n)}$ at time n, $\mathbf{\delta}^{(n)}$ is measurement error at time n, and H is the model parameter. Kalman filter assumes that the process and measurement errors are normally distributed with mean of zero, and their covariance matrix could be dependent on time.

$$\varepsilon^{(n)} \sim N(0, A^{(n)}), \ \delta^{(n)} \sim N(0, B^{(n)}) \tag{4-3}$$

4.1.1.1 Kalman Filter Steps

Kalman filter consists of two main steps: the prediction step and the estimation step. The prediction step tries to predict the future state of the system using the stochastic state model. The predicted state is called *a prior* estimate which has been predicted on the basis of the available information before the measurement of the system state.

$$\widetilde{x}^{(n+1)} = C \hat{x}^{(n)} + Du^{(n)}$$
$$\widetilde{P}^{(n+1)} = C \hat{P}^{(n)} C + A^{(n)} \tag{4-4}$$

Where $\widetilde{x}^{(n+1)}$ is *a priori* estimate of the state of the system for time $n+1$. This *a priori* estimate is actually our prediction for time $n+1$, and $\widetilde{P}^{(n+1)}$ is the *a priori* error covariance.

The estimation step tries to update the *a priori* estimate of the prediction step on the basis of the new measurement of the system state. The updated state of this step is called *a posteriori* estimate of the system state which uses the new measurement information.

$$\hat{x}^{(n+1)} = \widetilde{x}^{(n+1)} + K^{(n+1)}(z^{(n+1)} - H \widetilde{x}^{(n+1)}) \tag{4-5}$$

Where $\hat{x}^{(n+1)}$ is *a posteriori* estimate of the state of the system for time $n+1$, and $K^{(n+1)}$ is called Kalman gain that minimizes the *a posteriori* error covariance ($\hat{P}^{(n+1)}$), it is computed as follows

$$K^{(n+1)} = \widetilde{P}^{(n+1)} H^T (H \widetilde{P}^{(n+1)} H^T + B^{(n+1)})^{-1} \tag{4-6}$$

$$\hat{P}^{(n+1)} = (I - K^{(n+1)}H)\widetilde{P}^{(n+1)} \tag{4-7}$$

4.1.1.2 Stochastic State Model

Kalman filter has been extensively used to model stochastic dynamic systems. Models based on Kalman filter explain the evolution of the system through a stochastic state equation, and these models can be categorized on the basis of the type of this stochastic state equation.

The dynamics of the system can be modeled by differential equations. In this category of models, the stochastic state model is actually a difference equation which governs the evolution of the system.

Another category of models consider the stochastic system as an autoregressive process. In this category, the evolution of the system is modeled as a linear combination of the previous values meaning that a statistical relationship is utilized to project the future state of the system rather than a physical mechanism. The model we propose in the next section falls into this category. It should be noted that using autoregressive process as the stochastic state model assumes a smooth trend for the process.

4.2 Proposed Model

4.2.1 Field Data

The model is calibrated and validated using the field data explained in Chapter 3. The data is collected for 567 days during years of 2009 to 2011. An evaluation data set is constructed by randomly choosing 30 days, and the rest of the data is used for model calibration.

4.2.2 First Order Autoregressive Model

In this section, the stochastic state model is assumed to be a first order autoregressive model meaning that the state of the model for the near future can be solely explained by the current state of the system. Suppose we are interested in predicting variable x, then, the stochastic state model can be written as follows:

$$x^{(n)} = x^{(n-1)} + \varepsilon^{(n)} \tag{4-8}$$

Since $\varepsilon^{(n)}$ is normally distributed with mean equal to zero, the expected value of $\mathbf{x}^{(n)}$ would be equal to the expected value of $x^{(n-1)}$ meaning that the system is stationary with a constant expected value over time. In other words, the state of the system can be explained as a random walk around the expected value. Therefore, the stochastic state model for the traffic process is assumed as follows:

$$u^{(n+1)} = u^{(n)} + \varepsilon_1^{(n+1)} \tag{4-9}$$

$$q^{(n+1)} = q^{(n)} + \varepsilon_2^{(n+1)} \tag{4-10}$$

$$o^{(n+1)} = o^{(n)} + \varepsilon_3^{(n+1)} \tag{4-11}$$

Where u, q , and o are traffic speed, flow, and occupancy at a detector station, and ε_1 , ε_2 , and ε_3 are the associated error terms respectively.

One of the steps required to implement a Kalman filter model is to provide an estimate for the process and measurement error ($A^{(n)}$ and $B^{(n)}$ in Equation (4-3)). These filter parameters are difficult to estimate as we typically do not have the ability to directly observe the errors. However, often these parameters are obtained by conducting an off-line calibration and then the estimated error parameters are used for the online application of the Kalman filter (Madsen 2007).

A variance estimate for both measurement and process error of each traffic variable can be formulated separately. Suppose that the variance of process and measurement error for one of the traffic variables are constant over time and denoted as A and B respectively. The stochastic state models considered for the traffic process in Equation (4-9) to (4-11) allow us to simplify the general Kalman filter formulations; therefore, the *a priori* error variance, the Kalman gain and the *a posteriori* error variance for time interval *n+1* can be computed as follows:

$$\widetilde{P}^{(n+1)} = \hat{P}^{(n)} + A \tag{4-12}$$

$$K^{(n+1)} = \frac{\widetilde{P}^{(n+1)}}{\widetilde{P}^{(n+1)} + B} \tag{4-13}$$

$$\hat{P}^{(n+1)} = (\frac{B}{\widetilde{P}^{(n+1)} + B})\widetilde{P}^{(n+1)} \tag{4-14}$$

By replacing $\widetilde{P}^{(n+1)} = \hat{P}^{(n)} + A$ in the last two equations, the Kalman gain and the *a posteriori* error variance for time interval *n+1* can be written as a function of the *a posteriori* error variance for time interval *n* and the variance of process and measurement errors.

$$K^{(n+1)} = \frac{\hat{P}^{(n)} + A}{\hat{P}^{(n)} + A + B} \tag{4-15}$$

$$\hat{P}^{(n+1)} = (\frac{B}{\hat{P}^{(n)} + A + B})(\hat{P}^{(n)} + A) = \frac{\hat{P}^{(n)} + A}{\hat{P}^{(n)} + A + B} \times B = K^{(n)} \times B \tag{4-16}$$

Then, the equation for the Kalman gain can be simplified as the following recursive equation:

$$K^{(n+1)} = \frac{K^{(n)} \times B + A}{K^{(n)} \times B + A + B} = \frac{K^{(n)} + A/B}{K^{(n)} + A/B + 1} \tag{4-17}$$

Therefore, it can be concluded that the Kalman gain for each step is only dependent on the Kalman gain in the previous step $(K^{(n)})$ and the ratio of the variance of the process error (A) over the variance of the measurement error (B). Knowing that A and B are constant overtime, the value of Kalman

gain ($K^{(n)}$) approaches a limiting value when the number of steps (n) is sufficiently large. In other words, after some steps Kalman gain becomes constant.

Figure 4-1 illustrates the limiting value of the Kalman gain as a function of the ratio of the variance of the process error over the variance of the measurement error (A/B) and the initial value of the Kalman gain. It is clear the limiting value of the Kalman gain is not affected by its initial value assumed at the beginning of the modeling process and therefore, the ratio A/B is the value of interest and this needs to be calibrated.
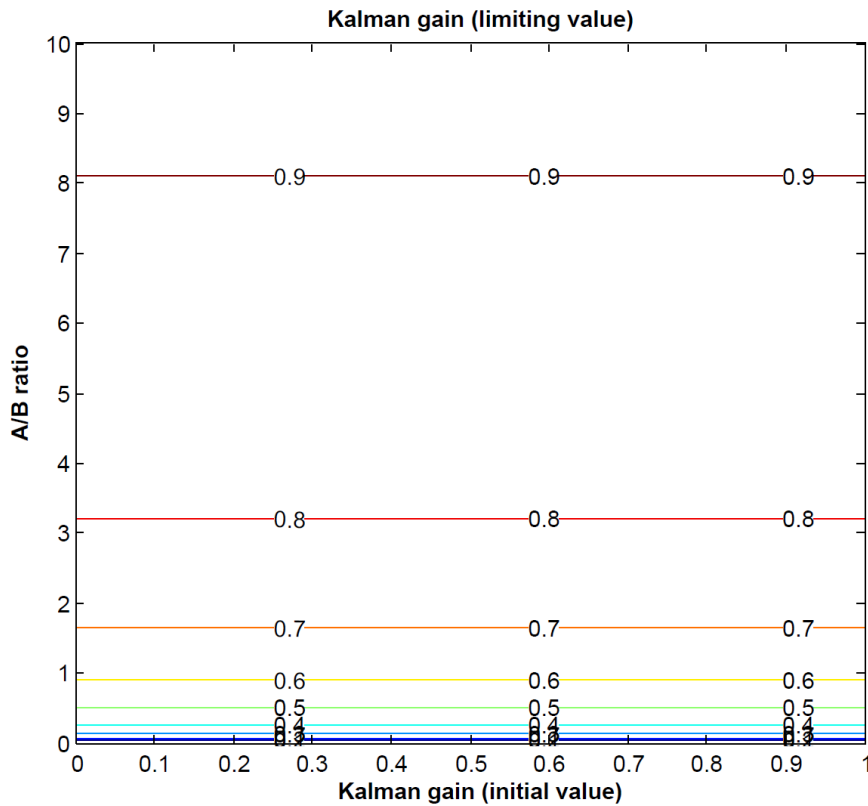
Figure 4-1: Variation of Kalman gain as a function of A/B and initial Kalman gain

In order to tune the error variance ratio (A/B), thirty days of Gardiner Expressway data were randomly selected, and the average prediction error was computed for different values of A/B. The prediction error is quantified as the mean absolute percentage error (MAPE) as follows:

$$MAPE = \frac{\sum_{n=1}^{N} \frac{v_p^n - v_0^n}{v_0^n}}{N}$$

(4-18)

Where,

39

$v_0^n =$ the actual value (ground truth) of a traffic variable for time interval n,

$v_p^n =$ the predicted value of a traffic variable for time interval n, and

$N =$ the number of time intervals.

Since the traffic measurements by the loop detectors are subject to error, we do not have access to ground truth to compute the error. Therefore, we assumed that the loop detector measurements can be used as the ground truth to evaluate the model performance. In other words, a model which predicts traffic variables as accurate as the sensor measurements is assumed to be perfect.

The prediction horizon in this study is assumed to be 15 minutes. Therefore, for each time interval, the measurements available 15 minutes prior to the interval of interest are used to predict traffic variables.

To accurately predict different traffic variables (speed, flow, and occupancy), a separate value of error variance ratio (*A/B*) is calibrated for each traffic variable. Then the appropriate value is used to predict speed, flow and occupancy.

Figure 4-2 illustrates the average prediction error for traffic stream speed as a function of error variance ratio (*A/B*). The results show that increasing the value of *A/B* reduces the average prediction error; however, the average error remains constant for values of *A/B* greater than 2. The same pattern can be observed for all 5 stations of the study route.

The average prediction error for traffic flow is depicted in Figure 4-3 as a function of error variance ratio (*A/B*). Unlike the results obtained for the traffic stream speed, the prediction error does not remain constant for greater values of *A/B*. Instead the prediction error is minimized for a value of *A/B* approximately equal to 1.

As shown in Figure 4-1, larger values of error variance ratio (*A/B*) result in larger values of Kalman gain ($K^{(n)}$); therefore, the *a posteriori* estimate relies more on the future measurements than the *a priori* estimate (Equation (4-5)). In an extreme case that measurements are not subject to any error (*B=0*), error variance ratio (*A/B*) approaches infinity, and the Kalman gain equals to 1 meaning that the *a posteriori* estimate for time interval n will be equal to the measurement for time interval n once it becomes available. Comparing the optimal value of *A/B* for traffic flow (1) and that of travel speed (2) suggests that the Kalman filter finds speed measurements more reliable than the flow measurements.

Figure 4-4 illustrates the average prediction error for occupancy as a function of error variance ratio (*A/B*). The results show a very similar pattern to traffic stream speed. The prediction error remains constant for greater values of *A/B*, and the optimal value for *A/B* is close to 2. Therefore, the *A/B* ratio is

considered to be equal to 2 for speed and occupancy predictions while it is considered to be equal to 1 for flow predictions.
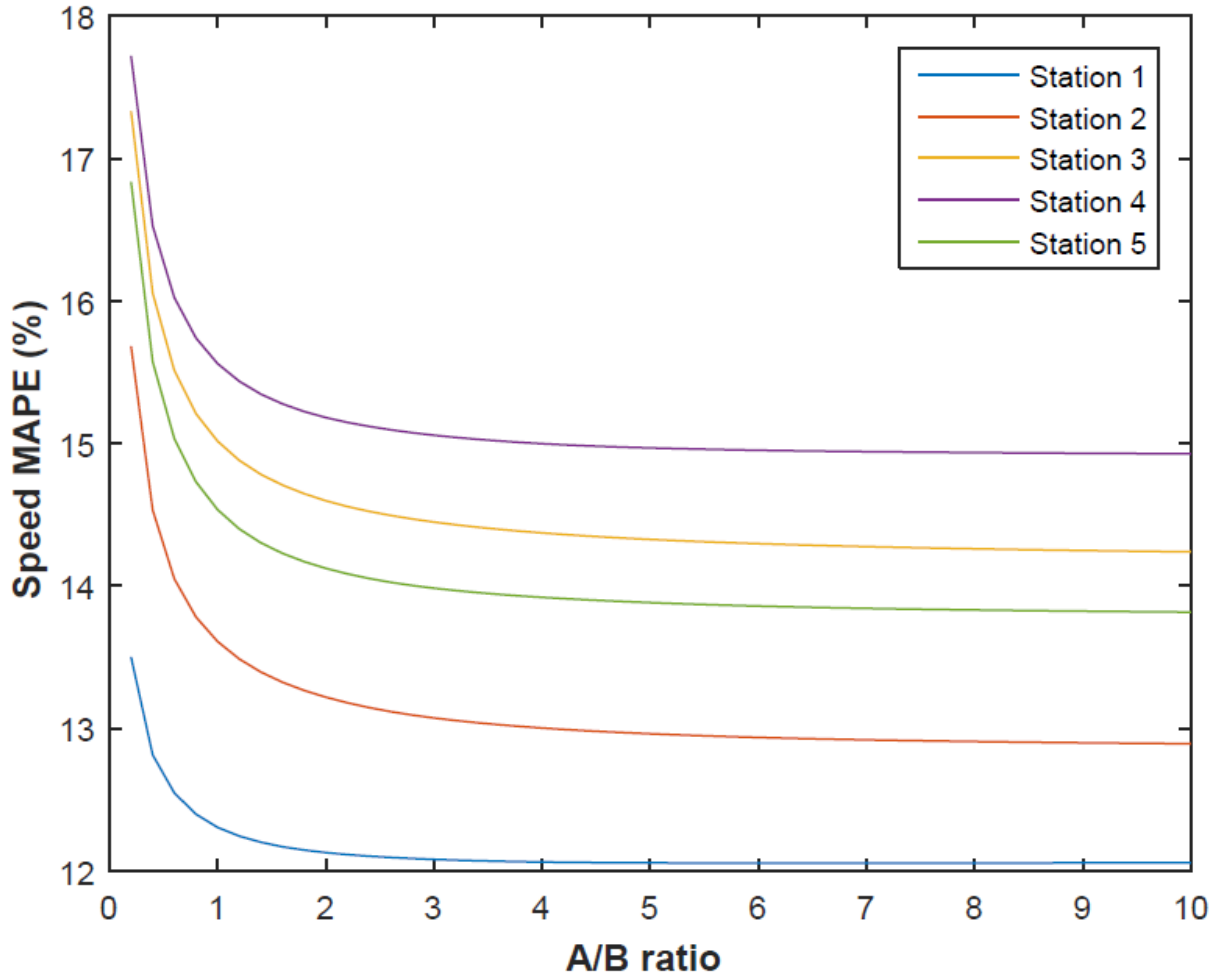
Figure 4-2: Prediction errors of traffic speed for different stations as a function of error variance ratio (*A/B*)
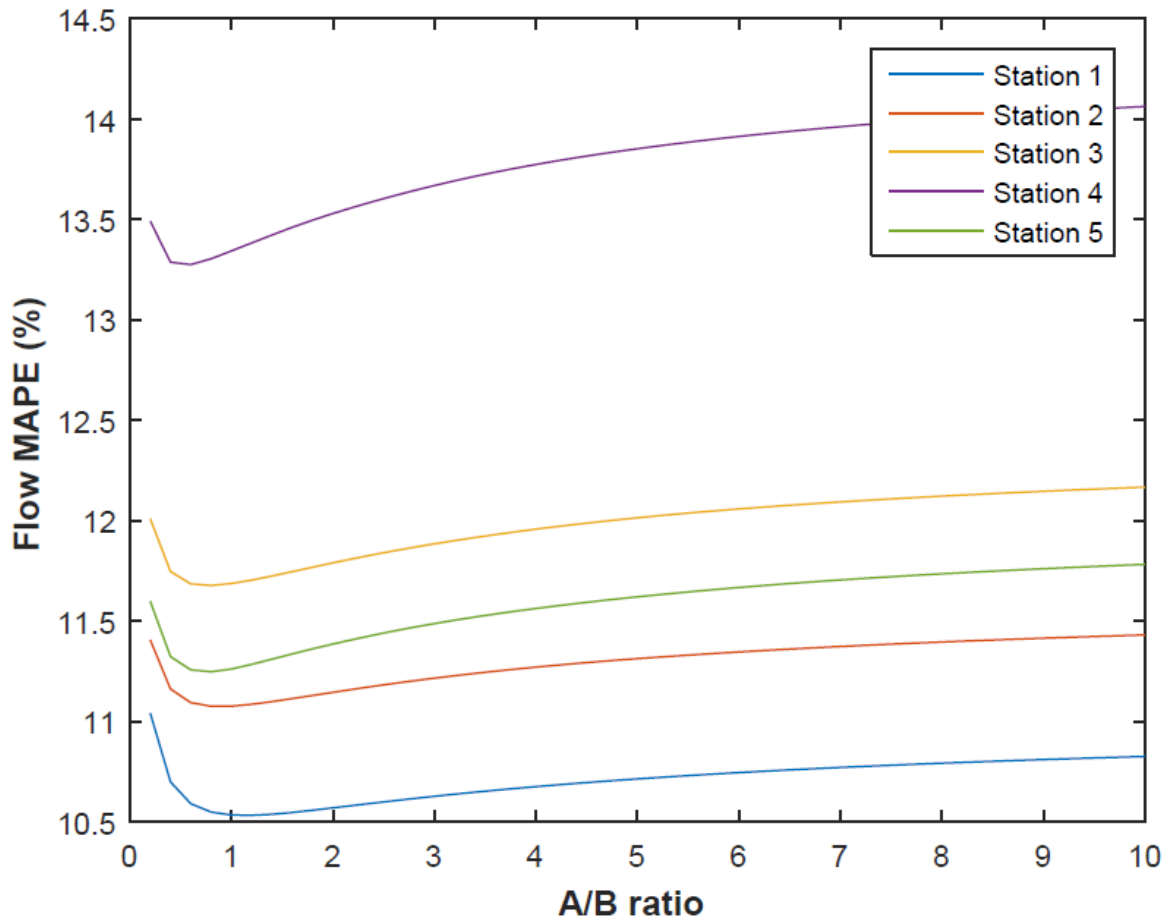
Figure 4-3: Prediction errors of traffic flow for different stations as a function of error variance ratio (*A/B*)
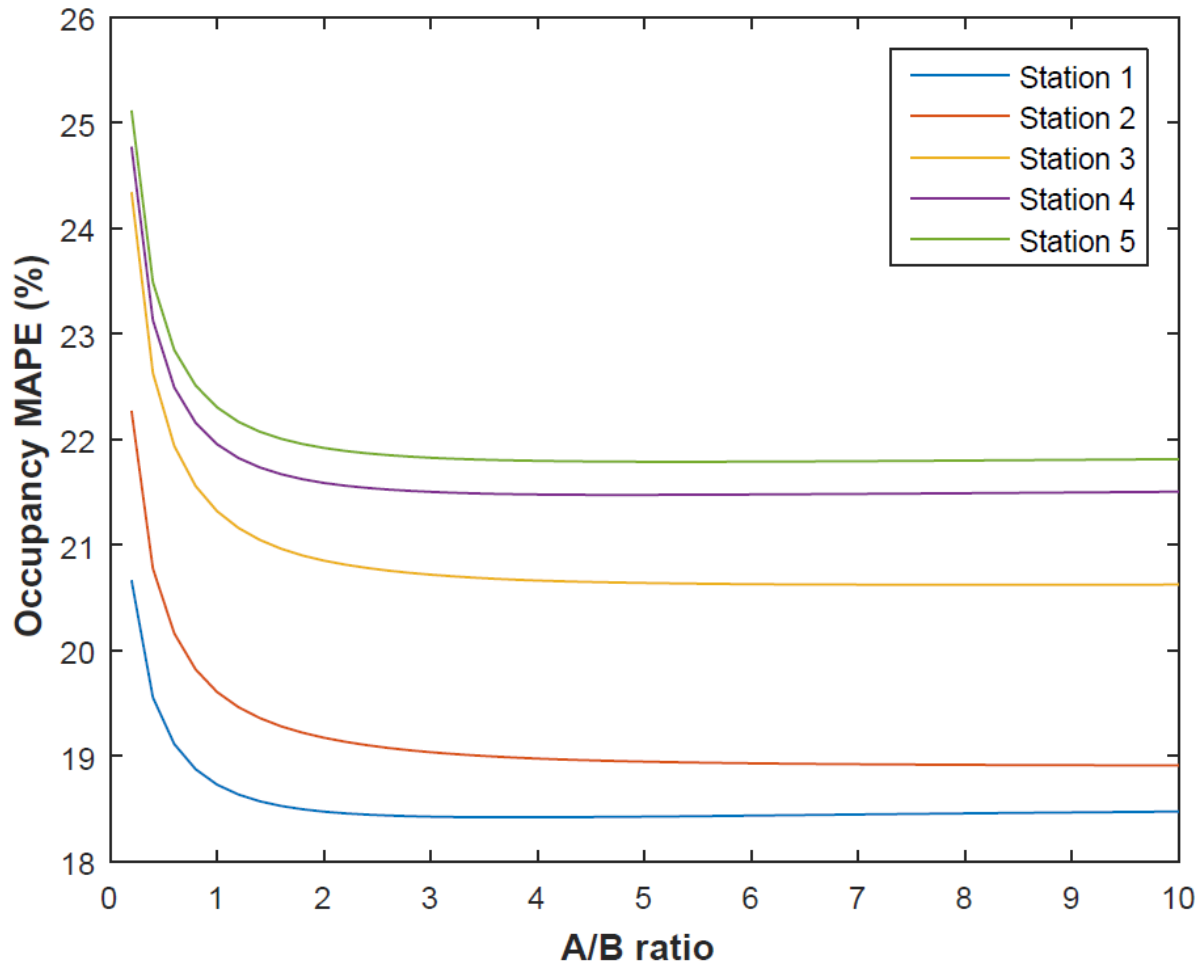
Figure 4-4: Prediction errors of traffic occupancy for different stations as a function of error variance ratio (*A/B*)

4.2.3 First Order Autoregressive Model with Constant Temporal Traffic Pattern

In this section, the stochastic state model of the Kalman model was developed to consider the expected changes in traffic pattern. It is assumed that traffic conditions follow a known pattern over the course of a day.

Figure 4-5 illustrates the relationship between different traffic variables at a detector station on a freeway. It should be noted that for uncongested traffic conditions the underlying mean speed of traffic seems to remain constant while traffic flow and density are changing. It is clear that the range of flow for "uncongested" conditions is quite large (say 0 to 1700 vphpl). So assuming a constant mean (i.e. stationary) conditions, is not valid. And because speed is almost constant over this range of low rates, the relationship between flow and density is almost linear, suggesting that density is also not stationary.
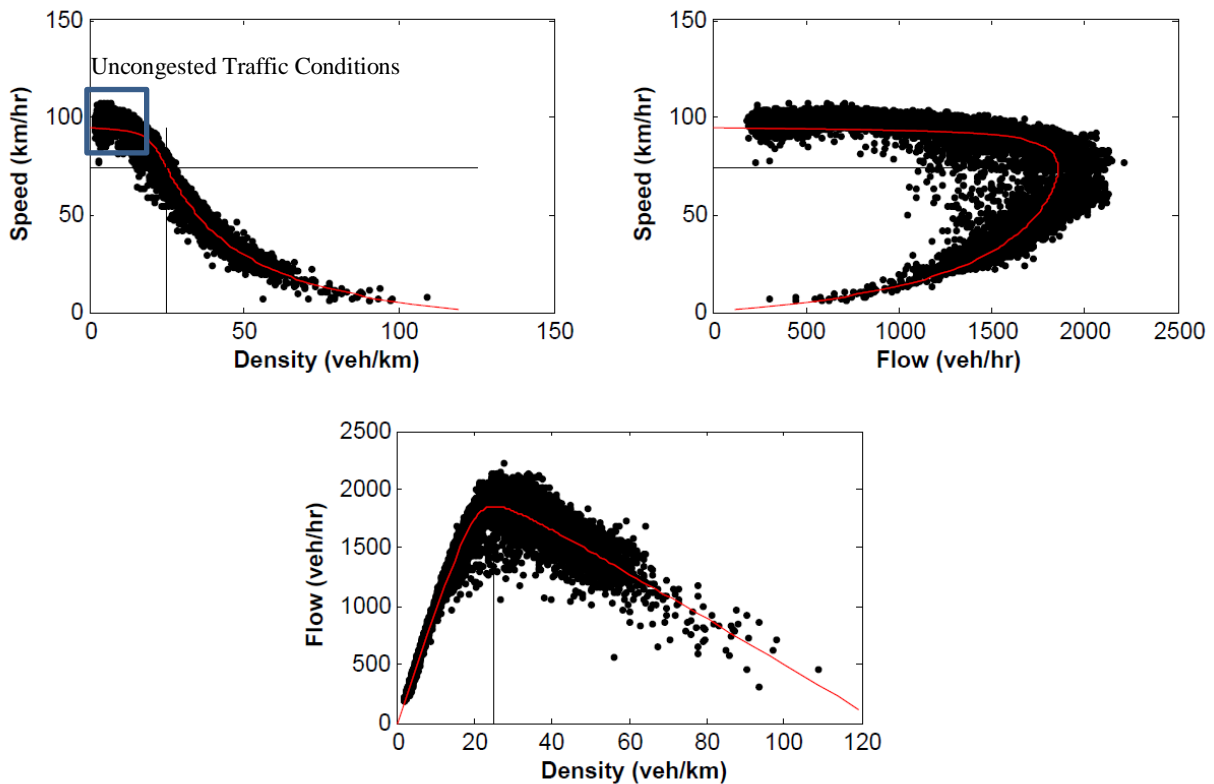
Figure 4-5: Relationship between traffic variables at a freeway station

If we cannot make the assumption of a stationary system for flow and density, then we need to include a term in the stochastic state model to consider the changes in traffic pattern during the uncongested traffic conditions. Therefore, the stochastic state model of the Kalman model may be written as follows:

44

$$u^{(n+1)} = u^{(n)} + \varepsilon_1^{(n+1)} \tag{4-19}$$

$$q^{(n+1)} = q^{(n)} + \Delta_2^{(n)} + \varepsilon_2^{(n+1)} \tag{4-20}$$

$$o^{(n+1)} = o^{(n)} + \Delta_3^{(n)} + \varepsilon_3^{(n+1)} \tag{4-21}$$

Where $\varepsilon_1^{(n+1)}$, $\varepsilon_2^{(n+1)}$, and $\varepsilon_3^{(n+1)}$ are normally distributed with mean equal to zero, and $\Delta_2^{(n)}$ and $\Delta_3^{(n)}$ are the changes in the expected value of traffic flow and occupancy for the next time interval. Kalman filter considers these terms as deterministic input to the model; therefore, they can be estimated based on the long term observations of traffic flow and occupancy. Taking expected values of both sides of the traffic flow stochastic state equation, we obtain:

$$E(q^{(n+1)}) = E(q^{(n)}) + \Delta_2^{(n)} \tag{4-22}$$

$$\Delta_2^{(n)} = E(q^{(n+1)}) - E(q^{(n)}) = q_m^{(n+1)} - q_m^{(n)} \tag{4-23}$$

Where $q_m^{(n)}$ is the historical mean flow rate at time n. The values of $q_m^{(n)}$ are proposed to be estimated using the historical observations of the traffic flow rate at the detector station. Therefore, the control term for the flow equation ($\Delta_2^{(n)}$) can be estimated using the historical pattern of observed traffic flow in the past.

The control term for the occupancy equation ($\Delta_3^{(n)}$) can be estimated based on the fundamental relationship between traffic variables. Assuming the expected value of traffic speed is constant during uncongested traffic conditions, the pattern of changes in occupancy can be estimated by the changes that we are expecting to happen to traffic flow.

Based on the detector measurement data, we have access to the occupancy of the detectors; however, the fundamental relationship between traffic variables is expressed in terms of density. Therefore, traffic density can be estimated on the basis of detector occupancy as follows:

$$k = \frac{o}{L_e} \tag{4-24}$$

Where o and k are occupancy and density, respectively, and L_e is the average effective length of the vehicles passing the detecting station. Average effective length of vehicles is equal to the summation of the length of the average vehicle and the length of the detection area. Then the following equation explains the relationship between traffic variables:

$$q = u \times \frac{o}{L_e} \tag{4-25}$$

$$q^{(n+1)} - q^{(n)} = u^{(n+1)} \times \frac{o^{(n+1)}}{L_e^{(n+1)}} - u^{(n)} \times \frac{o^{(n)}}{L_e^{(n)}} \tag{4-26}$$

Assuming that $L_e^{(n+1)} = L_e^{(n)} = L_e$ and $u^{(n+1)} = u^{(n)}$, it can be written:

$$o^{(n+1)} - o^{(n)} = \frac{L_e}{u^{(n)}} \times (q^{(n+1)} - q^{(n)}) \tag{4-27}$$

Then the one step change in the mean of occupancy can be written as follows:

$$\Delta_3^{(n)} = \frac{L_e}{u^{(n)}} \times \Delta_2^{(n)} \tag{4-28}$$

Since the model in this section considers a fixed pattern for the evolution of traffic flow for different times of day, it is called a first order autoregressive model with static pattern. In other words, the expected change in the traffic flow or occupancy at a given time of day is constant, and it is estimated based on the long term observations of traffic flow at the same location.

4.2.4 First Order Autoregressive with Dynamic Traffic Pattern

As explained for the static traffic pattern model, the future pattern of the traffic may be modeled using the historical observations of traffic variables; however, for the static pattern it is assumed that the changes in traffic can be modeled based on the time of day. In other words, the long term average speed, flow, and occupancy temporal profile is the representative of any day in the prediction dataset. However, historical traffic measurements show that the temporal profile of speed, occupancy and flow vary for different days. Therefore, we need to develop a model which is able to dynamically estimate the future pattern of the traffic.

Suppose that we are predicting traffic variables for a specific day; the idea of dynamic estimation of traffic pattern suggests that the future pattern of traffic for the prediction time interval may be estimated by finding a set of historical traffic conditions similar to the one that we are investigating. In other words, if we find time periods in the historical dataset that follow the same pattern as the prediction period, there is no reason to believe that the future pattern of the prediction period will be different than those of the similar historical periods. In order to dynamically estimate the future pattern of the traffic, a model based on k nearest neighbors (kNN) is developed in this section.

In the model proposed in this section, the stochastic state equations of the Kalman model are written as follows:

$$u^{(n+1)} = u^{(n)} + \Delta_1^{(n)} + \varepsilon_1^{(n+1)}$$
$$q^{(n+1)} = q^{(n)} + \Delta_2^{(n)} + \varepsilon_2^{(n+1)} \qquad \textbf{(4-29)}$$
$$o^{(n+1)} = o^{(n)} + \Delta_3^{(n)} + \varepsilon_3^{(n+1)}$$

Where $\varepsilon_1^{(n+1)}$, $\varepsilon_2^{(n+1)}$, and $\varepsilon_3^{(n+1)}$ are normally distributed with mean equal to zero, and the control terms ($\Delta_1^{(n)}, \Delta_2^{(n)}, \Delta_3^{(n)}$) are assumed to be estimated based on a kNN model. These control terms can be estimated as follows:

$$\Delta_1^{(n)} = E(u^{(n+1)}) - E(u^{(n)})$$
$$\Delta_2^{(n)} = E(q^{(n+1)}) - E(q^{(n)}) \qquad \textbf{(4-30)}$$
$$\Delta_3^{(n)} = E(o^{(n+1)}) - E(o^{(n)})$$

Where $E()$ is the expected value of a traffic variable for a given time interval. In order to estimate $\Delta_i^{(n)}$, we need to find days similar to the prediction day, and estimate the expected value of traffic variables for the appropriate time intervals.

4.2.4.1 Parameters Tuning

To implement the proposed model of this section, the following key parameters are required:

- Feature Vector: the attributes upon which the similarity is defined. These attributes are used to rank the available historical observations based on their similarity to the prediction period. In this research, we have considered the traffic variables for the past 15 minutes (three 5-minute intervals) for the station of interest as well as those for upstream and downstream stations as the feature vector. The time of day is not explicitly considered in the feature vector meaning that similar traffic conditions are considered in the analysis even if it occurs on a different time of day.

- Distance Metric: determines the closeness of the prediction period with any historical observation. In this research, Euclidean distance metric on a normalized variable space is implemented.

- Local estimation method: estimates the output based on the information from k nearest neighbors. Since we are interested in the expected value of a traffic variable for a given time interval, we have applied arithmetic mean in this research.

- Number of nearest neighbors (k): based on feature vector and distance metric, we can rank all the records available in the historical dataset. k determines the number of top ranked observations which are considered in the local estimation method. In this research,

we have investigated the performance of the model for different values of k; then an appropriate number for k is proposed.

In order to calibrate the number of neighbors, thirty days are randomly selected, and the average prediction error is computed for different values of k. The values of k that are evaluated in this research are 1, 5, 10, 15, 20, 30, 50, 75, and 100. The prediction error is stated based on mean absolute percentage error (MAPE).

Figure 4-6 illustrates the average prediction error for travel speed as a function of the number of neighbors. Since the feature vector includes information for both upstream and downstream stations, the algorithm can only be implemented on data for Stations 2, 3 and 4. Prediction error seems to be minimized for values of k between 10 and 20. For greater values of k, the prediction error increases; however, the rate of increase is not as great as for smaller values of k.

The average prediction error of traffic flow is depicted in Figure 4-7 as a function of the number of neighbors. The results show that prediction error remains constant for greater values of number of neighbors; however, the optimal number of neighbors seems to be very close to that of the traffic speed.

Figure 4-8 illustrates the average prediction error for occupancy as a function of the number of neighbors. The results show a very similar pattern to travel speed. The optimal number of neighbors is somewhere between 10 and 20.
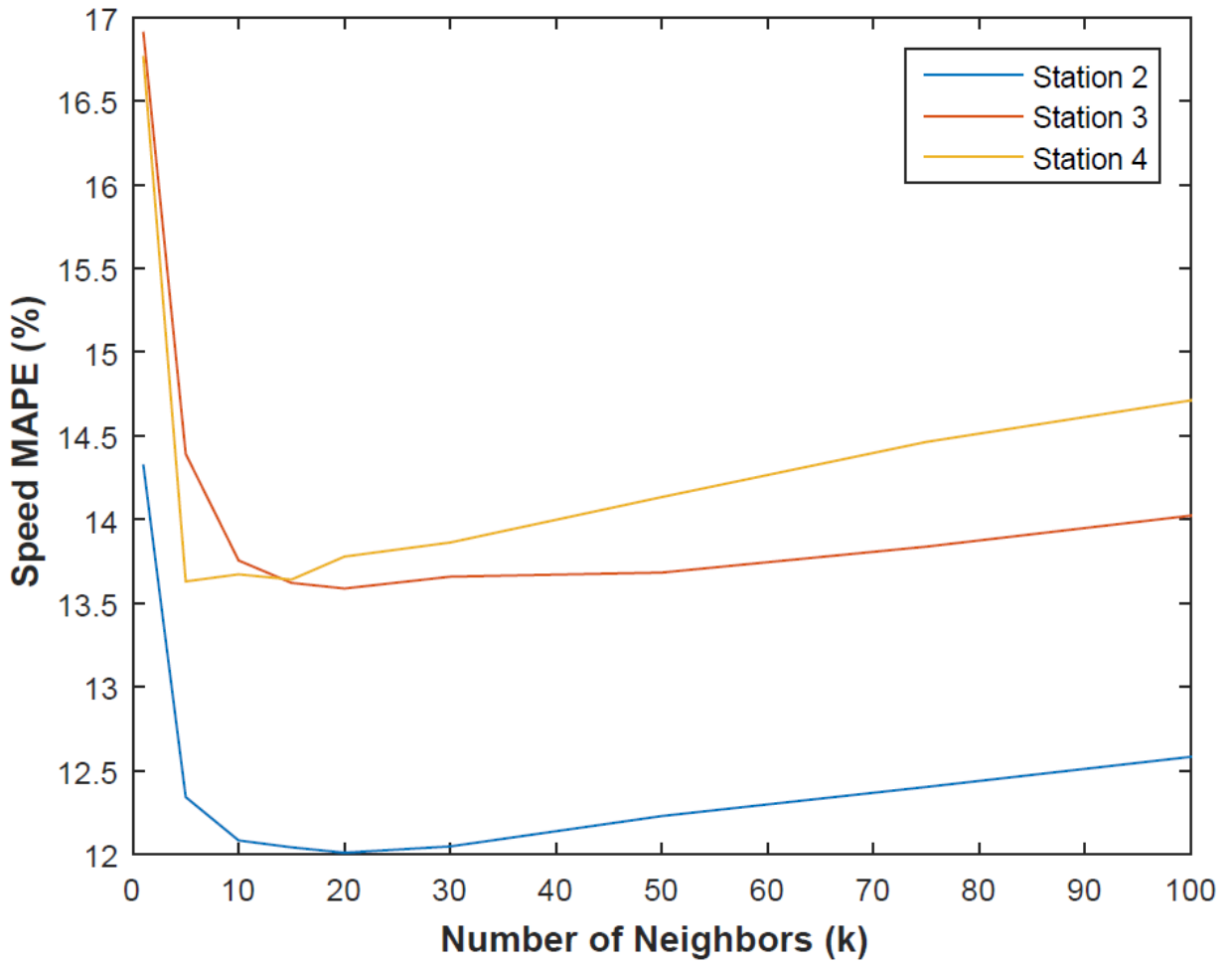
Figure 4-6: Prediction errors of traffic speed for different stations as a function of number of nearest neighbors
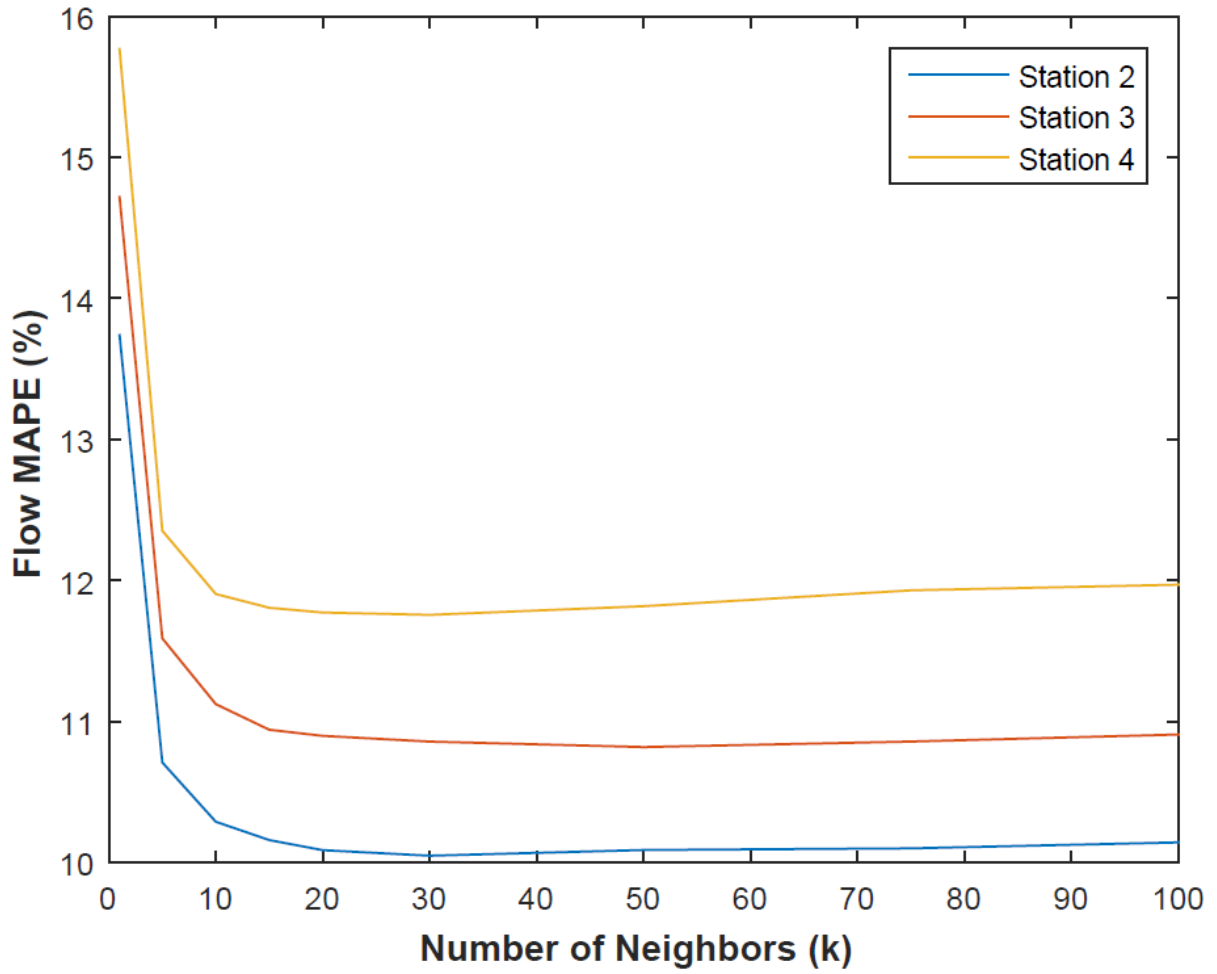
Figure 4-7: Prediction errors of traffic flow for different stations as a function of number of nearest neighbors
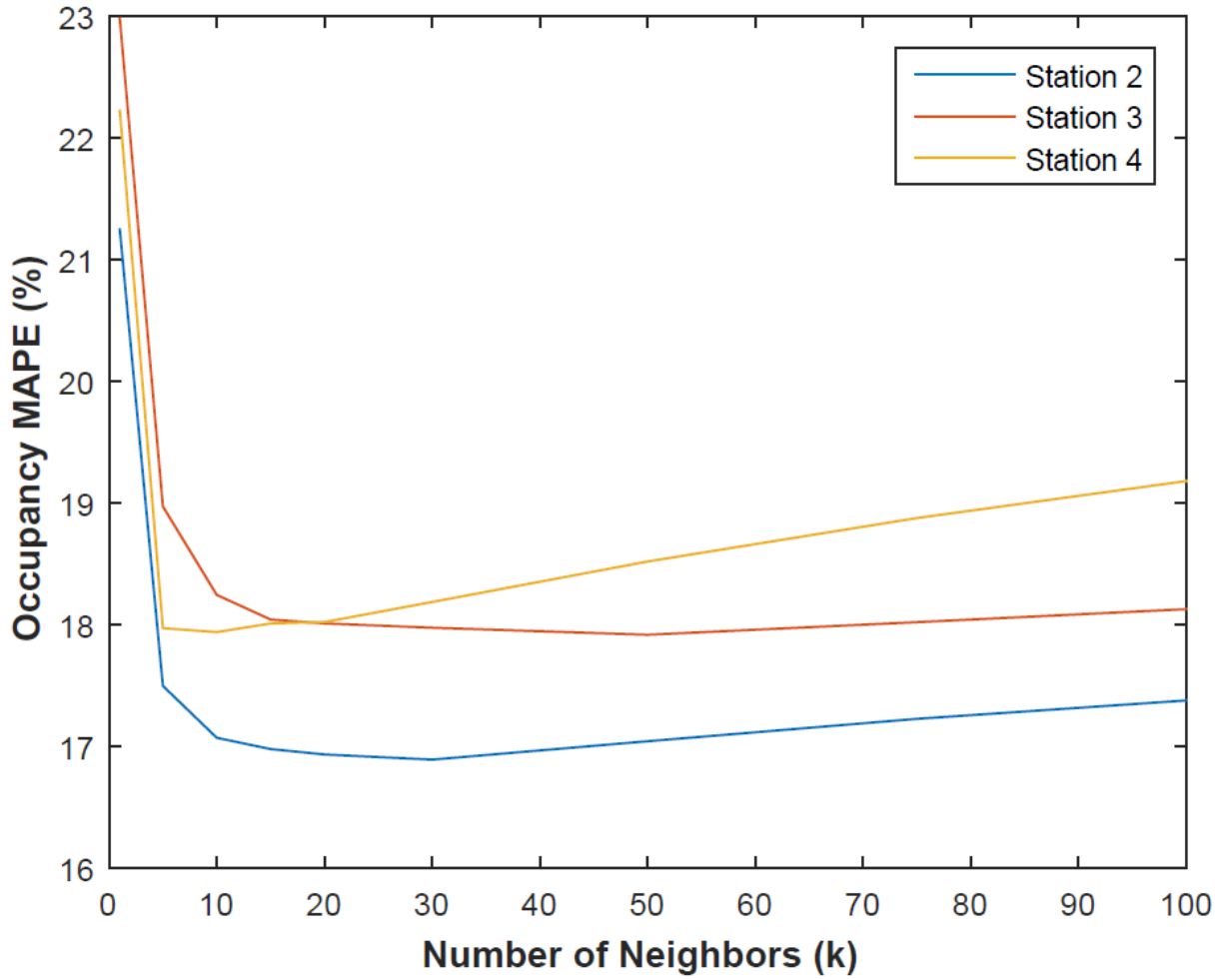
Figure 4-8: Prediction errors of traffic occupancy for different stations as a function of number of nearest neighbors

4.3 Prediction Results

In this section, the prediction results for 30 test days are presented. As mentioned earlier, the measure of performance is the mean absolute percentage error (MAPE), and the loop detector measurements are assumed to be the ground truth. Since Kalman model with dynamic pattern utilizes the information for the downstream and upstream stations, the prediction results for Stations 2, 3, and 4 are presented in this section.

Figure 4-9 illustrates travel speed prediction results for a sample day for one of the detectors. As expected the prediction errors are larger when traffic conditions were changing from free flow to congested and vice versa.

The average prediction error for the 30 test days are calculated to assess the overall performance of the model. Table 4-1 presents the average daily prediction error for travel speed. The prediction results for four different models are presented in this table. The models are denoted as follows:

- Naïve: a simple time series prediction model. The traffic variable for the prediction time interval is assumed to be equal to the latest available information. The Naïve model does not take into consideration any potential changes in traffic conditions between two consecutive time intervals. The state model is not stochastic, and can be formulated as $x^{(n+1)} = x^{(n)}$ where x is the traffic variable to be predicted.

- Kalman Base: a Kalman model based on first order autoregressive stochastic state model with no control term.

- Kalman Static: a Kalman model based on first order autoregressive stochastic state model with a static pattern as the control term.

- Kalman Dynamic: a Kalman model based on first order autoregressive stochastic state model with a dynamic pattern (kNN) as the control term

The first section of Table 4-1 presents the prediction errors for each model for the three different stations. The second section of this table presents the improvement of each model relative to the Naïve model. In this section, a p-value less than 0.05 indicates that the improvement is statistically significant. The results in this table demonstrate that the Kalman Base and Kalman Static models do not perform better than the Naïve model and that the Kalman Dynamic model does perform better than the Naïve model. Furthermore, these improvements in performance are statistically significant for Stations 2 and 4.

Table 4-2 presents the average daily prediction error for traffic flow. It should be noted that for traffic flow all three proposed models perform better than the Naïve model and these improvementsare statistically significant for all the stations.

Table 4-3 presents the average daily prediction error for detector occupancy. In this case, the Kalman Base model does not perform better than the Naïve model; however, both the Kalman Static and Kalman Dynamic model perform better than the Naïve model and these improvements are statistically significant for all stations.
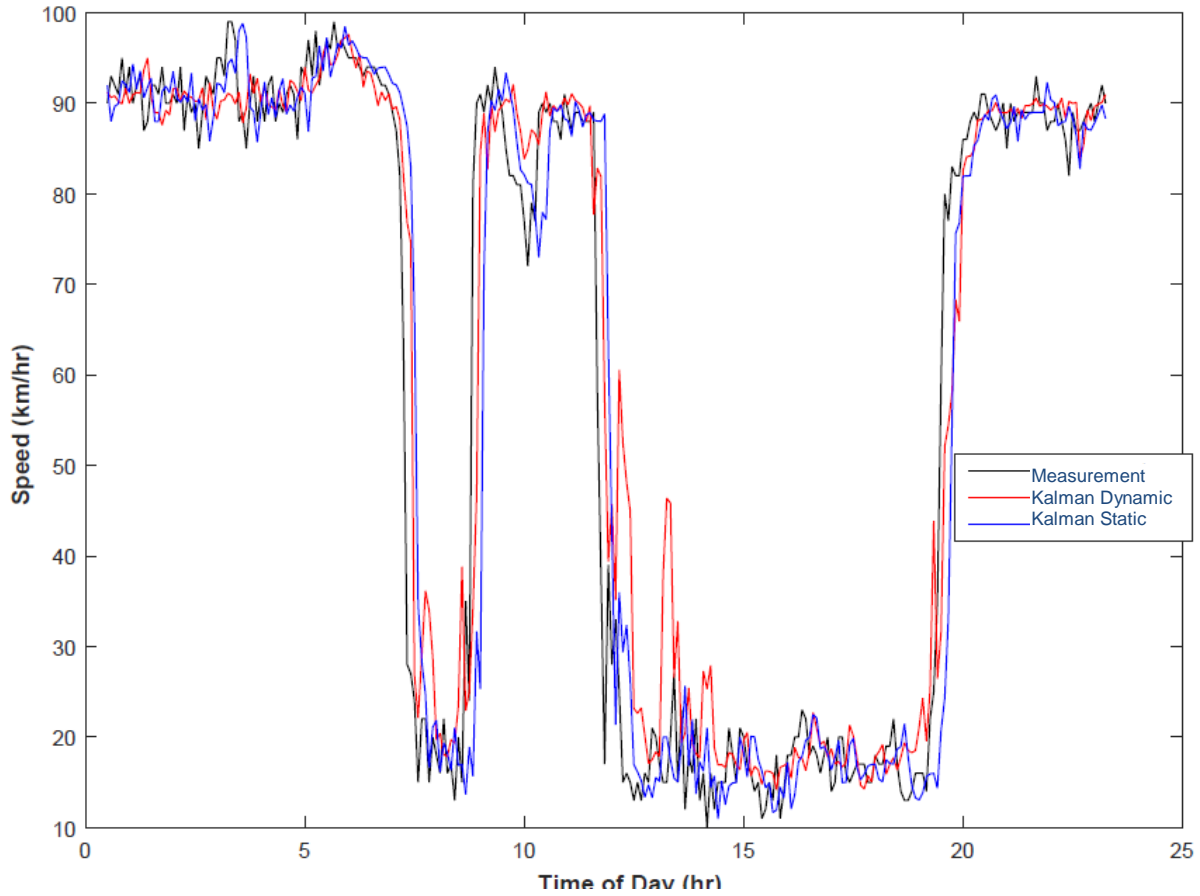
Figure 4-9: Prediction of Traffic Speed on Station 2 for a Sample Day

Table 4-1: Average prediction errors for travel speed over the entire day

Description	Model	Value	Station 2	Station 3	Station 4
Prediction Error	Naïve	MAPE (%)	14.9	14.2	12.8
	Kalman Base	MAPE (%)	15.0	14.4	13.0
	Kalman Static	MAPE (%)	15.0	14.4	13.0
	Kalman Dynamic	MAPE (%)	13.7	13.8	12.1
Improvement	Kalman Base	Relative (%)	-0.6	-1.5	-1.3
		Difference (%)	-0.1	-0.2	-0.2
		Standard Error (%)	0.2	0.3	0.2
		p-value	0.9860	1.0000	0.9999
	Kalman Static	Relative (%)	-0.6	-1.5	-1.3
		Difference (%)	-0.1	-0.2	-0.2
		Standard Error (%)	0.2	0.3	0.2
		p-value	0.9860	1.0000	0.9999
	Kalman Dynamic	Relative (%)	8.3	2.9	5.9
		Difference (%)	1.2	0.4	0.8
		Standard Error (%)	2.5	2.8	1.9
		p-value	0.0050	0.2191	0.0183

Table 4-2: Average prediction errors for traffic flow over the entire day

Description	Model	Value	Station 2	Station 3	Station 4
Prediction Error	Naïve	MAPE (%)	14.4	12.4	11.6
	Kalman Base	MAPE (%)	13.6	11.8	11.2
	Kalman Static	MAPE (%)	12.1	10.4	9.9
	Kalman Dynamic	MAPE (%)	11.9	11.1	10.3
Improvement	Kalman Base	Relative (%)	5.5	4.6	3.8
		Difference (%)	0.8	0.6	0.4
		Standard Error (%)	0.2	0.2	0.2
		p-value	0.0000	0.0000	0.0000
	Kalman Static	Relative (%)	15.7	16.1	14.9
		Difference (%)	2.3	2.0	1.7
		Standard Error (%)	0.4	0.4	0.3
		p-value	0.0000	0.0000	0.0000
	Kalman Dynamic	Relative (%)	17.3	10.4	11.5
		Difference (%)	2.5	1.3	1.3
		Standard Error (%)	1.1	1.0	0.8
		p-value	0.0000	0.0000	0.0000

Table 4-3: Average prediction errors for occupancy over the entire day

Description	Model	Value	Station 2	Station 3	Station 4
Prediction Error	Naïve	MAPE (%)	21.6	20.7	18.9
	Kalman Base	MAPE (%)	21.5	20.7	19.0
	Kalman Static	MAPE (%)	21.1	20.0	18.4
	Kalman Dynamic	MAPE (%)	17.9	18.2	17.1
Improvement	Kalman Base	Relative (%)	0.7	0.0	-0.2
		Difference (%)	0.1	0.0	0.0
		Standard Error (%)	0.3	0.3	0.3
		p-value	0.0123	0.9586	0.3581
	Kalman Static	Relative (%)	2.4	3.4	3.1
		Difference (%)	0.5	0.7	0.6
		Standard Error (%)	0.7	0.5	0.3
		p-value	0.0001	0.0000	0.0000
	Kalman Dynamic	Relative (%)	17.0	11.7	9.8
		Difference (%)	3.7	2.4	1.9
		Standard Error (%)	2.2	2.6	1.8
		p-value	0.0000	0.0000	0.0000

The previous tables presented the daily average errors for the stations of the study route; however, looking at the daily average errors may be misleading because prediction errors for free flow conditions are very low and there are many hours in a day that traffic condition is free flow. Moreover, the challenging part of prediction of traffic variables is when traffic conditions are changing from free flow to congested and vice versa. Therefore, we also examine the prediction errors during periods when traffic conditions change. We denote these periods as "change periods". Traffic conditions during a "change period" are not steady state and a change in the traffic state may happen. The "change period" is defined based on the shockwave model discussed in Chapter 5, and it includes all the instances in which one of the shockwave models was activated.

Table 4-4 presents the average prediction error for travel speed during change periods. Two observations can be made:

1. The improvement provided by the Kalman Dynamic model during this period is considerable (relative improvement between 30%-40%), and it is statistically significant.

2. The average prediction error during this period is much higher than the daily average (i.e. in Table 4.1).

Table 4-4: Average prediction errors for travel speed over the change period

Description	Model	Value	Station 2	Station 3	Station 4
Prediction Error	Naïve	MAPE (%)	76.8	66.7	53.4
	Kalman Base	MAPE (%)	80.3	71.6	56.9
	Kalman Static	MAPE (%)	80.3	71.6	56.9
	Kalman Dynamic	MAPE (%)	46.8	41.2	35.4
Improvement	Kalman Base	Relative (%)	-4.6	-7.4	-6.6
		Difference (%)	-3.6	-4.9	-3.5
		Standard Error (%)	2.2	2.9	3.0
		p-value	1.0000	1.0000	1.0000
	Kalman Static	Relative (%)	-4.6	-7.4	-6.6
		Difference (%)	-3.6	-4.9	-3.5
		Standard Error (%)	2.2	2.9	3.0
		p-value	1.0000	1.0000	1.0000
	Kalman Dynamic	Relative (%)	39.1	38.2	33.7
		Difference (%)	30.0	25.4	18.0
		Standard Error (%)	24.6	25.7	23.9
		p-value	0.0000	0.0000	0.0001

4.4 Conclusions

This chapter documents the findings of a prediction methodology based on Kalman filter as follows:

- The model is developed to predict the near future traffic conditions (speed, flow, and occupancy) at the location of fixed point detectors (i.e. loop detector in this study)

- The method requires live stream of traffic measurements (speed, flow, and occupancy) to be available for the near future predictions.

- Traffic propagation was proposed to be predicted based on either a static or dynamic traffic pattern. In the static pattern it was assumed that traffic conditions can be attributed based on the observed conditions in the same time of day; however, in the dynamic pattern, expected traffic conditions are estimated based on the current measurements of upstream and downstream detectors.

- Prediction results in this chapter are provided for a 15 minute prediction horizon.

- The prediction results were compared to a naïve method, and it was shown that the average daily prediction error of the proposed method is lower when compared to the naïve method for the sample locations (10-17% improvements)

- The prediction results were also compared during the "change period" when traffic conditions are changing from free flow to congestion and vice versa. It was noted that the model improvements are considerably greater during this period (approximately 25% improvement).

- Although travel speed predictions errors are significantly reduced when compared to the naïve method, the average prediction error during the periods when traffic conditions are changing rapidly (i.e. the "change periods"), remain large (MAPE = 35-45%). The focus of the next chapters is to propose and assess methodologies to reduce these prediction errors during these "change periods".

Chapter 5
Congestion Spill back and Dissipation[1]

In the previous Chapter, we proposed three Kalman Filter based prediction models. Though the Kalman Dynamic model was shown to be the best and to be statistically superior to the Naïve model, prediction errors were unacceptably large when traffic conditions were changing (i.e. from uncongested to congested or vice versa).

Suppose that we are predicting traffic speed along a freeway route. Traffic may breakdown at a section of the freeway as a result of recurrent or non-recurrent congestion. The fact that traffic has broken down on this section provides valuable information which can be used to update the predictions for all the upstream sections along the route. In other words, we need to develop a mechanism to predict how the downstream congestion propagates upstream, and determine if the queue spills back into the upstream section(s) during the prediction horizon.

Shockwave theory is a potential method for carrying out these predictions. With knowledge of the traffic states on a downstream and upstream section of a freeway, shockwave theory is able to estimate the speed of the shockwave, and the location of the boundary between the two states at different times. However, using the classical shockwave theory for this purpose faces two challenges: a) Shockwave theory requires traffic measurements not only at the main sections of a freeway, but also at all access (e.g. on-ramps) and egress points (e.g. off-ramps) along the study route; and b) our preliminary analysis shows that in many cases classical shockwave theory is unable to predict if a backward forming or forward recovery shockwave actually occurs. As is described later in this chapter, we applied classical shockwave theory to field data to predict the occurrence of shockwaves on a section of freeway (with no on or off-ramps). However, the model was only able to predict 30% of backward forming shockwaves and 59% of forward recovery waves that were observed in the field data. The main reason that shockwave theory does not perform well may lie in the fact that it works based on a deterministic framework and relies on the assumption that traffic conditions on the upstream and downstream of the shockwave are in steady state – an assumption typically violated especially when congestion is forming or dissipating.

The objective of this chapter is to develop a model to reconstruct the mechanism of backward forming and forward recovery shockwaves on freeways, and utilize the model in a traffic speed prediction framework. The proposed model follows a stochastic approach, and the probability of a shockwave

---

[1] The content of this chapter have been incorporated within an article presented by Noroozi and Hellinga in Transportation Research Board 94th Annual Meeting on January 2015: Noroozi, Reza, and Bruce Hellinga, 2015, "A Stochastic Model for Predicting Shockwaves on Freeways ", Presented in Transportation Research Board: 94th Annual Meeting, Washington DC.

occurring is predicted based on the latest available traffic measurements. To construct a more flexible structure, the model is calibrated to only use the measurements on the main sections of the freeway, and the data from on-ramp and off-ramp detectors are not required. Using the maximum likelihood technique, a multi-state stochastic model has been calibrated to estimate the probability of spillback or recovery.

5.1 Background

Traffic is usually defined in terms of flow, density, and speed; in theory, it can be treated as a continuum fluid flow passing through roadways. Consider a unidirectional continuous road section with an upstream and a downstream boundary. The conservation of traffic stream (conservation equation) can be written as follows (Kuhne, 1997)

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0 \tag{5-1}$$

Where q and k are traffic flow and density, and x and t are space and time, respectively.

Equation (5-1) expresses the relationship between two fundamental dependent variables, density and flow rate, and two independent ones (time and space), and it can be used to determine the traffic characteristics at any location and time. To solve this equation the fundamental relationship of traffic ($q = uk$; where u is the traffic speed) is considered as well as an additional assumption that speed is a function of density ($u = f(k)$; where f() can be any function). However, this assumption is only valid at steady state conditions (Kuhne, 1997).

The solution to Equation (1), first proposed by Lighthill and Whitham (Lighthill, 1955), results in a family of curves called "characteristics". They are straight lines emanating from the boundaries carrying the value of density and flow at those boundaries. The slope of the curves can be calculated as follows (Kuhne, 1997)

$$\frac{dx}{dt} = f(k) + k[f(k)] = \frac{dq}{dk} \tag{5-2}$$

When two characteristics intersect, the density and flow at the intersection point must assume values which are not possible in practice. This inconsistency is explained by the generation of shockwaves which are the boundaries between two different traffic states. The speed of a shockwave can be calculated as follows (Kuhne, 1997)

$$u_w = \frac{q_d - q_u}{k_d - k_u} \tag{5-3}$$

Where q_d and k_d represent downstream and q_u and k_u represent upstream flow and density.

Daganzo (1994) proposed the cell transmission model (CTM), a discrete approximation of the classical differential equations which explain the conservation of traffic flow. In CTM, each link is divided into some homogeneous "cells". The length of each cell is determined to be equal to the distance traveled in one simulation time step at free flow speed. CTM has been used as a simulation tool to estimate/predict traffic states by several researchers (Muñoz, 2003 and Tampère, 2007). Although CTM involves less computational effort compared to the classical differential equations, it is still computationally demanding and it also requires the deployment of traffic sensors for all on-ramps and off-ramps along the study route.

Izadpanah et al (2011) proposed a methodology for real-time travel time prediction utilizing the shockwave information. Their proposed method acquired the shockwave information from the trajectories of a sample of GPS equipped vehicles rather than fixed point detectors.

Liu et al (2009) proposed a method based on shockwave theory to estimate the queue length at congested signalized intersections in real-time. Since the nature of signals dictates the times that shockwave occur, they were more interested in estimating the speed of the shockwave rather than when the shockwave occurs.

There are some other studies that considered the effect of shockwave implicitly. Most of the pattern recognition algorithms to predict traffic speed such as neural networks, k nearest neighbors (kNN), and Kalman filters using traffic state models fall into this category. In this chapter, a method is proposed to explicitly predict the shockwaves on freeways, and update the prediction of traffic speed based on the shockwave information. Since the assumptions of shockwave theory (deterministic and steady state traffic conditions) are typically violated, a stochastic model is proposed to estimate if a shockwave occurs given the traffic conditions on the upstream and downstream stations.

5.2 Proposed Model

5.2.1 Field Data

The model is calibrated and validated using the field data explained in Chapter 3. The data is collected for 567 days during years of 2009 to 2011. An evaluation data set is constructed by randomly choosing 30 days, and the rest of the data is used for model calibration.

5.2.2 Stochastic Process

Generally speaking, $\{X(t), t \in T\}$ is called a stochastic process if $X(t)$ is a random variable (or random vector) for any fixed $t \in T$. T is referred to as the "index set", and is often interpreted in the context of time. As such, $X(t)$ is often called the "state of the process at time t". The index set T can be

a continuum of values or can be a set of discrete points; the resulting process is called continuous-time or discrete-time, respectively.

A stochastic process $\{X(n), \, n = 0, 1, 2, ...\}$ is said to be a discrete-time Markov chain (MC) if the following conditions hold true:

(1) For any $n = 0, 1, 2, ... $, $X(n)$ takes on a finite (countable) set of possible values which are called "states".

(2) For any $n = 0, 1, 2, ... $,

$$\Pr[X(n+1) = x(n+1) \,|\, X(n+1) = x(n), X(n-1) = x(n-1), \, ... \, , X(0) = x(0)] = \Pr[X(n+1) = x(n+1) \,|\, X(n) = x(n)]$$

The second condition implies that the conditional distribution of any future state $X(n+1)$ given the past states $X(0), X(1), ..., X(n-1)$ and the present state $X(n)$ is independent of the past states. This relation is referred to as the Markov property. In practical terms, this property implies that the probability of traffic conditions changing from uncongested to congested conditions in the next time interval is solely a function of the traffic characteristics at the current time and not on what the traffic conditions were in the previous time period. In general, a stochastic process may be explained by a Markov chain of order m where m is finite, and shows the memory of the model. For a Markov chain of order m, the future state $X(n+1)$ is dependent on the last m states of the process ($X(n), X(n-1), ..., X(n+1-m)$).

For any pair of states i and j, the transition probability from state i at time n to state j at time $n+1$ is given by $P_{n,i,j} = \Pr(X^{(n+1)} = j \,|\, X^{(n)} = i)$. Let $P_n = [P_{n,i,j}]$ be the associated matrix containing the elements $P_{n,i,j}$, referred to as the one-step transition probability matrix (TPM) from time n to time $n+1$

5.2.3 Traffic States

The first step to develop the proposed Markov model is to define the states of the traffic. When defining the states of the traffic the following points should be taken into consideration:

- Traffic states should be defined so that they differentiate between different circumstances; therefore, traffic conditions which lead to similar future conditions should be combined into one state.

- A traffic state should be defined to contain only similar traffic conditions.

- The greater the number of states that are defined, the more complicated the model calibration becomes because the number of parameters to be estimated is proportional to the square of the number of the states.

We define traffic states for the proposed Markov model on the basis of the macroscopic relationship of traffic variables. Figure 5-1 illustrates the traffic fundamental diagram for Station 4 from the Gardiner Expressway route. Each point is a 5-minute station average. The relationship between speed and occupancy and occupancy and traffic flow are illustrated in Part (a) and Part (b), respectively. Traffic conditions can be divided into three different categories. The points located on the left side of the graphs represent free flow conditions, and the points located on the right side of the graphs represent congested conditions. The points in the middle represent transition conditions. These conditions happen when the traffic state is changing from free flow to congestion or vice versa.

Measurements of traffic variables in Figure 5-1 show a wide range of variability. For example, for a given value of occupancy the speed may vary within a wide range; therefore, a macroscopic model is calibrated to explain the fundamental relationship between traffic variables. This fundamental model can be utilized to define the boundaries of traffic states. In this research, it is assumed that the fundamental traffic flow relationship can be explained by Van-Aerde's single regime model (19950).Van-Aerde et al (1995) proposed a three dimensional optimization problem to calibrate the four parameters of the model. Since the fundamental relationship may vary for different situations including weather conditions, a separate model has been fitted to the data associated with each month of the year. The calibration has been separately implemented for each detector station (i.e. each station along the route); consequently, this procedure has led to 12 different models for each station. Therefore, for a given day of the year, the appropriate model has been utilized to define the state of the traffic for each measurement interval.

Part (c) of Figure 5-1 shows the fitted Van-Aerde model on the speed-occupancy graph for the month of June. Three critical points are highlighted on this graph. Points A, B, and C represent free flow, capacity, and jam conditions, respectively. Capacity on this graph is estimated from a deterministic point of view; however, in the real world, the conditions which lead to a traffic breakdown (i.e. transition from uncongested conditions to congested conditions) may be observed on the left or right side of point B on the graph. Therefore, the transition state includes observations on both sides of point B. Observations on the very right side of the graph represent congested conditions, and observations on the very left side of the graph show free flow conditions. To more accurately explain the free flow conditions of traffic, the free flow observations can be divided into two different categories.

In this research, four traffic states are considered including two free flow states, one transition state, and one congested state. Points b1, b2, and b3 define the boundaries between the four traffic states. Points b1 and b2 divide the occupancy between free flow (A) and capacity (B) into three equal bins, and point b3 divides the speed between capacity (B) and jam density (C) into two equal bins.

Traffic states are labeled as State 1 to 4: State 1 is when traffic is completely free flow; State 2 is the second free flow condition; State 3 is the transition condition; and State 4 is congested.
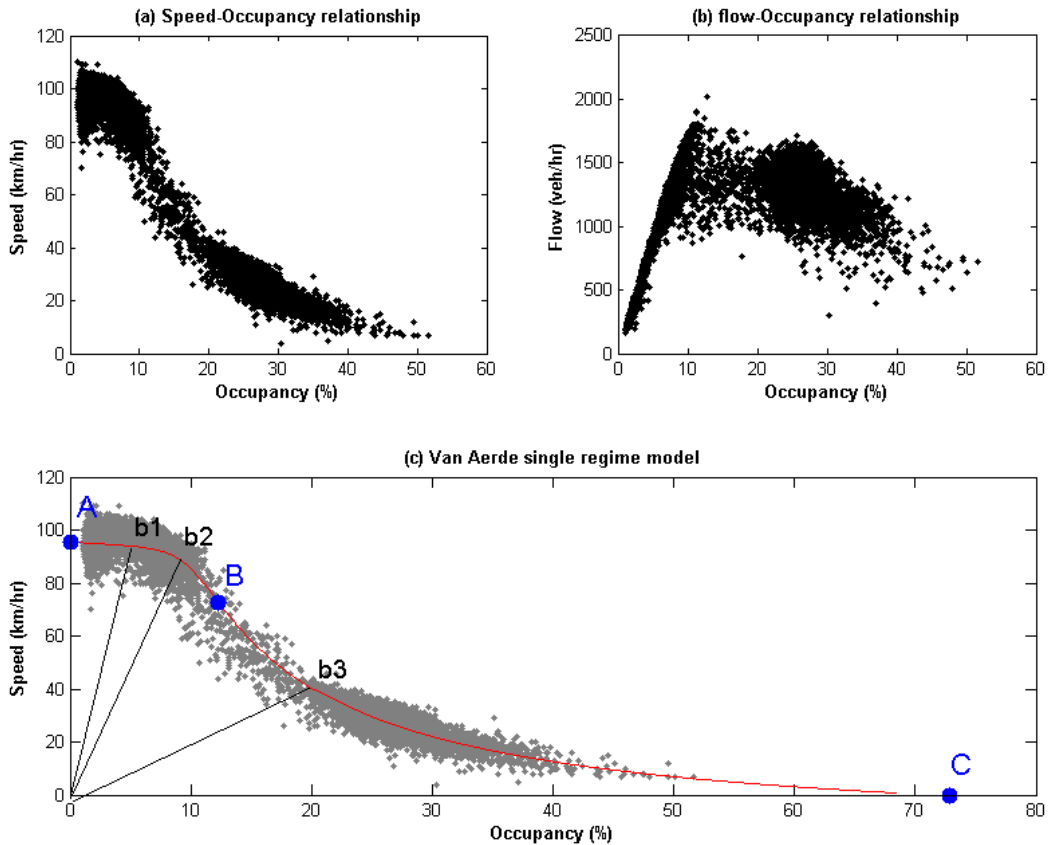
Figure 5-1: Fundamental diagram and traffic states on Station 4

For the purpose of congestion detection model, traffic conditions at the location of each detector station are characterized by two different states: a) congested, or b) uncongested. We chose these states because the objective of the proposed model is to predict how a congested state propagates upstream or dissipates. Traffic states in this chapter are defined on the basis explained in Figure 5-1. However, because we are only interested in congested versus uncongested states, the first three states from Figure 5-1 are combined into a single uncongested state. Furthermore, we are interested in traffic conditions that result in shockwaves that propagate upstream or downstream and not short term local traffic disturbances. Therefore, a congested traffic condition which lasts for more than one time interval (5 minute) is called a congested state in this chapter. A prediction framework based on all four states is proposed in Chapter 7.

An off-line method is used to determine the traffic state (i.e. label as uncongested or congested) for each 5-minute time interval in the field data set. An off-line method is used so that for each time interval both the past and the future measurements are assumed to be available. The state of the traffic is assumed to be congested if for two consecutive 5-minute time intervals the state of the traffic is not State 1 or State 2, and for at least one of the two time intervals the state of the traffic is State 4.

Since the fundamental relationship may vary for different situations including weather conditions, a separate Van Aerde model has been fitted to the data associated with each month of the year. The calibration has been implemented separately for each detector station.

5.2.4 Model Structure

The proposed model is developed to predict the future state of the traffic on a station of a freeway given the current traffic conditions on the upstream and the downstream stations. Two different mechanisms are considered in this study: a) backward forming shockwaves which occur when the upstream station is uncongested and the downstream station becomes congested, and b) forward recovery shockwaves which occur when the downstream station is congested and the upstream station becomes uncongested.

Suppose that the station of interest and the downstream station are currently uncongested. In this situation, we can assume that the traffic conditions on the station of interest are not affected by the downstream conditions. However, as soon as the downstream station becomes congested, there is the potential for the queue to grow and spillback into the station of interest (i.e. a backward forming shockwave to exist). Therefore, as soon as a congestion state is observed at the downstream station, the backward forming model is activated and we designate the traffic state as "Potential Spillback" (State A). As illustrated in Figure 5-2-a, there are three possible traffic states for the station for the next time interval:

- The process remains in the "Potential Spillback" mode (i.e. downstream station is congested and station of interest is uncongested).

- The congestion on the downstream station dissipates before affecting the station of interest (i.e. both the downstream and station of interest are uncongested). We refer to this as "Dissipation".

- The congestion on the downstream station spills back and causes the station of interest congested to become congested as well (i.e. downstream station and station of interest are congested).
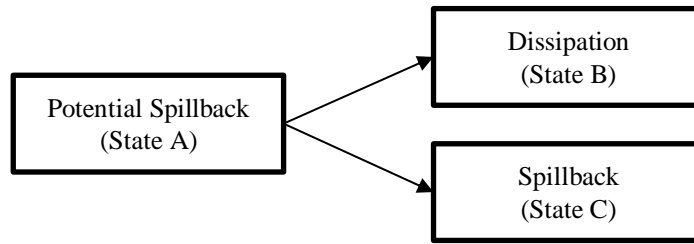
Part (a) Backward Forming Shockwave
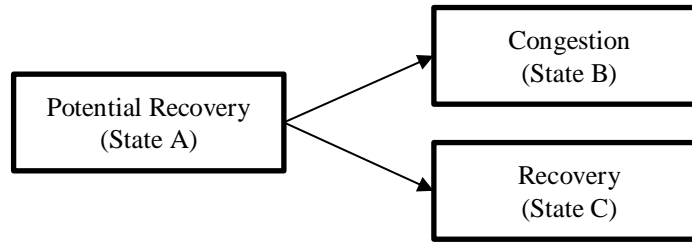
Part (b) Forward Recovery Shockwave

Figure 5-2: Stochastic model structure

The same structure can be defined for the forward recovery shockwave (Figure 5-2-b). Suppose that both the station of interest and the upstream stations are congested. A forward recovery shockwave occurs when the upstream station becomes uncongested. Therefore, as soon as the state of the upstream station is observed to be uncongested, the forward recovery model is activated. The state of the process upon activation is called "Potential Recovery", and there are three possible outcomes similar to backward forming shockwaves.

Stochastic models with this structure are called competing risk models meaning that when the model is activated (State A) there are two parallel ultimate outcomes (States B and C) competing with each other. When the process enters either of the ultimate states, there is no returning back to State A. Instead, the process stops here and waits for another activation.

5.3 Model Calibration

The proposed models need to be calibrated to estimate the transition probabilities between different states. The model parameters have been estimated using Maximum Likelihood estimation method and historical observations of the state transitions.

In this chapter, a continuous-time Markov model is developed to estimate the transition probability matrix. In continuous Markov models, the transition from one state to another may occur at

any real time. Therefore, the defining parameters of the model are transition intensities. The transition intensity from State i to State j (λ_{ij}) is defined in Equation 5-4. Then, the transition probabilities for a given time interval can be computed as a function of transition intensities and the length of the time interval as follows:

$$\lambda_{ij} = \lim_{\Delta t \to o} \frac{\Pr(X(t+\Delta t)=j \mid X(t)=i)}{\Delta t} \tag{5-4}$$

$$\Pr(\Delta t) = e^{Q \times \Delta t} \tag{5-5}$$

Where, $\Pr(\Delta t)$ is the matrix of transition probabilities for a time interval of Δt, Q is the matrix of transition intensities, and $e^{(\ )}$ is matrix exponential which is computed by the power series.

The transition intensities are assumed to be a function of traffic conditions, and a parametric model has been developed to estimate the transition intensities. To fit a model based on the maximum likelihood method, a statistical package has been used in R programming language called "msm" (Jackson, 2011). This package is capable of calibrating multi-state Markov and hidden Markov models in continuous time.

As discussed earlier, traditional shockwave models require measurement of the traffic variables at the location of each on-ramp and off-ramp. However, in this approach we are trying to implicitly consider the effect of mid-segment access points in the calibration process. The model only uses the measurements on the main section of the freeway; however, the parameter estimates for each segment reflect the very specific characteristics of that segment. Therefore, the model needs to be calibrated for each segment separately.

The simplest model is one for which the transition intensities are assumed to be constant meaning that the transition between two states is not affected by traffic conditions. Therefore, the probability of a spillback or a recovery is only affected by the time that the process is activated. This model is considered the "Base model" in this chapter, and is designated as Model 1. The transition intensity matrix for Model 1 has been estimated using maximum likelihood, and then the transition probability matrices are estimated for a 5 minute interval. The estimated transition probabilities in this table represent the historical averages that shockwaves impact the traffic state on the downstream or upstream section in the next time interval. For example, consider the result for the Backward Forming Shockwave model. If the station of interest is Station 1 and it is uncongested and the downstream station (Station 2) is congested, then the historical data indicate that the probability of transitioning from State A to State C (i.e. Station 1 will be congested in the next time interval) is 3.4%. The transition probabilities vary across the different stations as a result of differences in geometry. For example, if the station of interest is Station 3, then P_{AC} is 15.3%, much higher than for Station 1. This likely reflects the much shorter length of Section 3

(L3=280m) as traffic conditions are able to propagate across shorter segments more quickly than longer segments.

Table 5-1: Statistical results for the Base Markov model (Model 1)

Freeway Station	Backward Forming Shockwave P_{ij}(5 min)			Forward Recovery Shockwave P_{ij}(5 min)		
	1 to 1	1 to 2	1 to 3	1 to 1	1 to 2	1 to 3
1	92.1%	4.4%	3.4%	-	-	-
2	84.5%	7.9%	7.6%	92.4%	4.4%	3.2%
3	73.1%	11.5%	15.3%	84.1%	6.3%	9.6%
4	80.0%	10.9%	9.1%	75.0%	9.3%	15.7%
5	-	-	-	81.7%	6.7%	11.6%

There are different factors which may affect the formation of a shockwave on freeways. Based on the traditional deterministic shockwave theory, the speed of a shockwave on a segment of a freeway can be estimated as a function of the traffic flow and the density of the downstream and upstream stations. Suppose that we are studying a segment of a freeway on which the relationship between flow and density follows a triangular pattern with given parameters. Therefore, for any measured traffic density, the traffic flow can be estimated. If the downstream station becomes congested, the range of traffic conditions on the upstream which would lead to a queue spillback can be estimated.

Figure 5-3-a illustrates a hypothetical triangular fundamental diagram for a segment of a freeway. q_c, k_c and k_j represent the flow at capacity, density at capacity, and jam density, respectively.

Suppose that traffic flow and density measured at the downstream station are q_d and k_d, respectively, and the downstream station is congested ($k_d > k_c$) while the upstream station is uncongested ($k_u < k_c$). In order for a backward forming shockwave to propagate from the downstream station to the upstream station, the traffic variables on the upstream station (q_u and k_u) must satisfy the following inequality:

$$q_u > q_d \tag{5-6}$$

$$\frac{q_c}{k_c}k_u > q_c - (\frac{q_c}{k_j - k_c}) \times (k_d - k_c) \tag{5-7}$$

$$\frac{1}{k_c}k_u + (\frac{1}{k_j - k_c})k_d > 1 + (\frac{k_c}{k_j - k_c}) \tag{5-8}$$

Equation (5-8) states that the criteria for a shockwave to occur can be explained by the density on the upstream and the downstream stations, and Figure 3b illustrates all the combinations of upstream and downstream traffic densities that result in a backward forming shockwave given the triangular fundamental diagram in Figure 5-3-b.
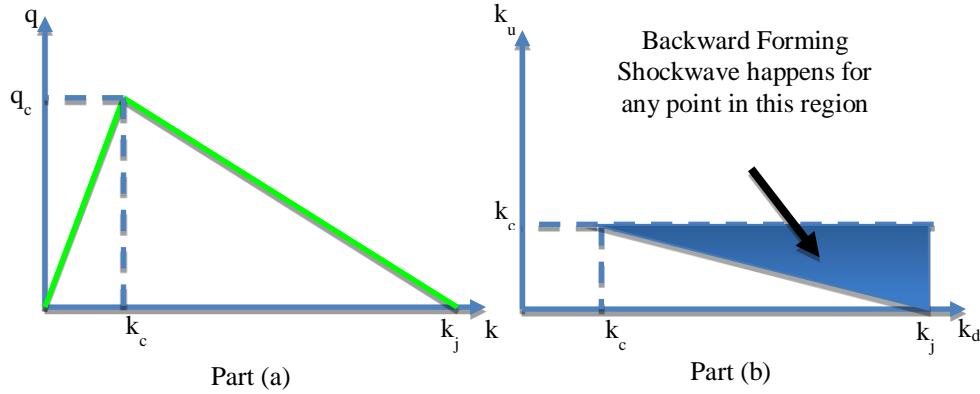
Figure 5-3: Hypothetic example of backward forming shockwaves

In practice we don't have measures of density, and therefore we assume that the measured loop detector occupancies on the upstream and the downstream stations affect the transition intensities of the proposed model. The transition intensities between different states in the proposed model of this study (Model 2) are considered as follows:

$$\lambda_{ij} = \begin{cases} \lambda_{0,ij} \times e^{(\alpha_{ij} \times ocp_d + \beta_{ij} \times ocp_u)} & \forall i \neq j \\ -\sum_{k \neq i} \lambda_{ik} & \forall i = j \end{cases}$$

(5-9)

Where,

λ_{ij} = transition intensity from state i to state j,

$\lambda_{0,ij}$ = base transition intensity from state i to state j (model parameter) ,

α_{ij}, β_{ij} = log-linear effects of covariates (model parameters), and

ocp_d, ocp_u = traffic occupancy on the downstream and upstream stations, respectively.

Table 5-2 presents the parameter estimates of Model 2. The values in parentheses indicate the lower and upper 95% confidence limits of the parameter estimates. The bold numbers show the statistically significant log-linear effects of covariates. It is clear that the effect of both downstream and upstream occupancy (α_{ij}, β_{ij}) is statistically significant for all the stations of the study. Since the model is formulated as a competing risk model, the relative values of parameter estimates are important. For

68

example, if traffic occupancy on the downstream station increases, it affects both the transition intensity to State B and C. For a backward forming shockwave, we expect an increase in the downstream occupancy to increase the chance of a traffic spillback (transition to State C). Therefore, the value of α_{AC} is expected to be greater than the value of α_{AB}. Therefore, a statistically valid model requires at least one of α_{AB} or α_{AC} to be statistically significant and $\alpha_{AC} > \alpha_{AB}$ for backward forming shockwaves. The same is true for β_{AB} and β_{AC}. The parameter estimates of backward forming shockwaves in Table 5-2 for all the stations hold the above conditions. For the forward recovery shockwaves, we expect that an increase in the downstream occupancy decreases the chance of recovery (transition to State C). Therefore, the value of α_{AC} is expected to be smaller than the value of α_{AB} which holds true for all the stations.

It can be shown that in a continuous time Markov model, the time that a process remains in a given state follows an exponential distribution (Kijima, 1997). Suppose that the process is in State A, there are two independent exponential processes with parameters λ_{AB} and λ_{AC} competing with each other. The time that the process remains in State A is exponentially distributed with parameter $\lambda_A = \lambda_{AB} + \lambda_{AC}$. Therefore, for the backward forming shockwave model the probability that the station of interest becomes congested (i.e. State C - Spillback) can be calculated as follows (t_{AB} and t_{AC} are time to transition from State A to B and C, respectively) (Kijima, 1997).

$$\Pr(Spillback) = \Pr(t_{AC} < t_{AB}) = \frac{\lambda_{AC}}{\lambda_{AB} + \lambda_{AC}} \tag{7}$$

Table 5-2: Statistical results for the proposed Markov model with covariates (Model 2)

Station	Parameter	State A to A	State A to B	State A to C
Backward Forming Shockwave				
1	$\lambda_{0,ij}$	-0.2441 (-0.3546,-0.1681)	0.2441 (0.1681,0.3546)	7.713e-06 (3.506e-06,1.697e-05)
	α_{ij}		**-0.2201** (-0.2396,-0.2007)	**0.1019** (0.08583,0.1179)
	β_{ij}		**0.1919** (0.1547,0.2291)	**0.6719** (0.6169,0.7269)
2	$\lambda_{0,ij}$	-0.8254 (-1.132,-0.6016)	0.8253 (0.6015,1.132)	6.984e-05 (3.231e-05,0.000151)
	α_{ij}		**-0.1893** (-0.2173,-0.1613)	**0.1488** (0.1309,0.1666)
	β_{ij}		0.03568 (-0.02271,0.09408)	**0.5592** (0.4881,0.6303)
3	$\lambda_{0,ij}$	-1.166 (-1.7,-0.8005)	1.16 (0.7941,1.694)	0.006715 (0.004263,0.01058)
	α_{ij}		**-0.08131** (-0.1086,-0.05401)	0.01419 (-0.01,0.03838)
	β_{ij}		**-0.1312** (-0.1767,-0.08571)	**0.3336** (0.2929,0.3742)
4	$\lambda_{0,ij}$	-0.353 (-0.4997,-0.2493)	0.3529 (0.2492,0.4996)	0.0001102 (6.488e-05,0.0001871)
	α_{ij}		**-0.1379** (-0.1537,-0.1222)	**0.1094** (0.09907,0.1198)
	β_{ij}		**0.1483** (0.1126,0.184)	**0.4955** (0.4553,0.5357)
Forward Recovery Shockwave				
2	$\lambda_{0,ij}$	-9.937 (-14.57,-6.776)	3.608e-05 (1.885e-05,6.905e-05)	9.937 (6.776,14.57)
	α_{ij}	0	**0.09869** (0.0803,0.1171)	**-0.32** (-0.3426,-0.2974)
	β_{ij}	0	**0.495** (0.4672,0.5229)	-0.01011 (-0.0553,0.03507)
3	$\lambda_{0,ij}$	-1.598 (-2.196,-1.163)	0.0007187 (0.0003898,0.001325)	1.598 (1.162,2.196)
	α_{ij}		**0.07247** (0.04774,0.0972)	**-0.1707** (-0.193,-0.1484)
	β_{ij}		**0.3316** (0.3028,0.3605)	0.03301 (-0.02527,0.09129)
4	$\lambda_{0,ij}$	-20.96 (-29.22,-15.03)	0.06676 (0.04025,0.1108)	20.89 (14.96,29.16)
	α_{ij}		**-0.1427** (-0.1745,-0.1109)	**-0.257** (-0.2821,-0.2318)
	β_{ij}		**0.2781** (0.2442,0.312)	**0.0592** (-0.09291,-0.02549)
5	$\lambda_{0,ij}$	-13.08 (-20.92,-8.182)	0.0004248 (0.0001958,0.0009219)	13.08 (8.182,20.92)
	α_{ij}		**0.05179** (0.02837,0.07521)	**-0.1632** (-0.1807,-0.1457)
	β_{ij}		**0.3585** (0.3244,0.3927)	**-0.118** (-0.1481,-0.08794)

Numbers in the brackets show the 95% confidence interval of the estimates. The bold numbers show the statistically significant log-linear effects of covariates.

Figure 5-4 illustrates the probability of spillback for all the stations of the study while a backward forming shockwave process is activated. The probability of spillback is presented as a function of the loop detector occupancy in the station of interest (upstream) and the downstream station. Each curve on the graph shows the combinations of upstream and downstream values of occupancy which result in the same probability of spillback. The curve of 0.5 shows the boundary beyond which the probability of a spillback (State C) is more than the probability of Dissipation (State B) The contour lines appear to be nearly linear, suggesting that a linear combination of the detector occupancies on the two stations results in a constant probability of queue spillback. The contours for Station 3 appear to be different from the other three stations. This is likely a reflection of the short length of Section 3 (280m) and indicates that traffic conditions on Station 3 are extremely correlated with those on Station 4. Therefore, the probability of spillback seems to be less dependent on the traffic condition of the downstream, and it is mostly defined based on the traffic condition of the upstream station.
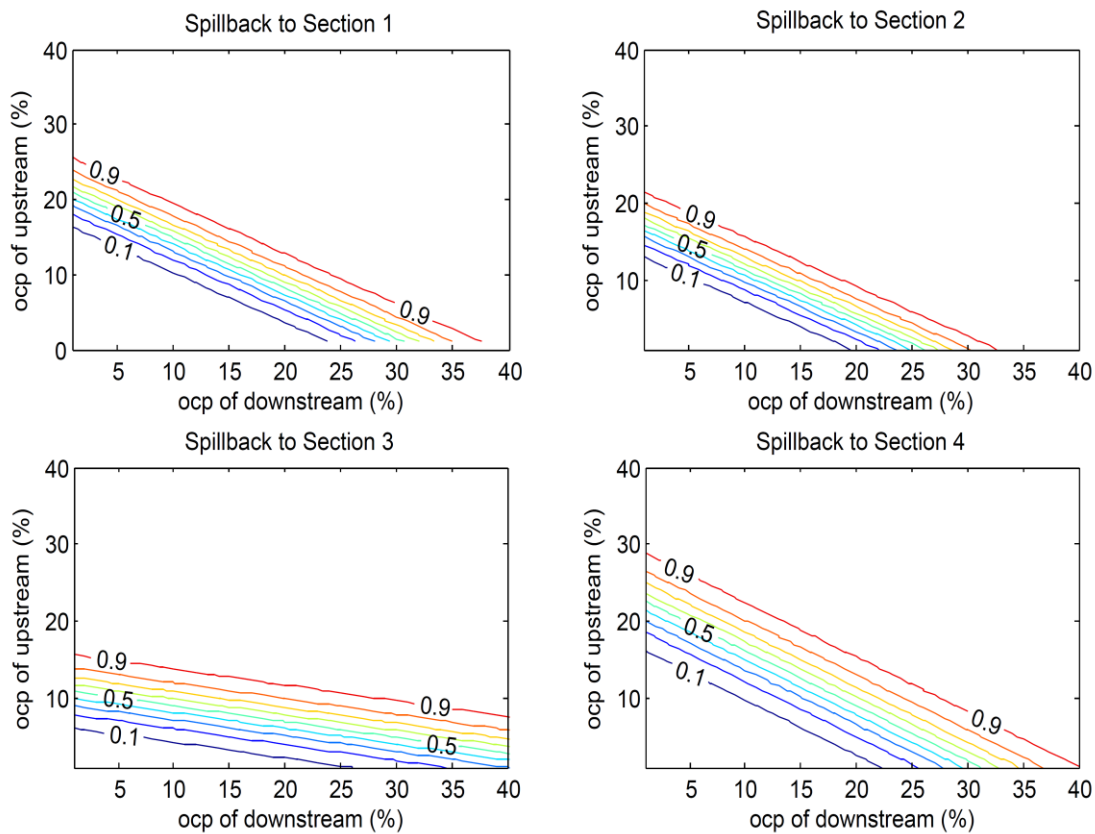
Figure 5-4: Probability of spillback based on the detector occupancy of the station of interest (upstream) and the downstream station (Model 2)

71

Figure 5-5 illustrates the probability of recovery for all the stations of the study while a forward recovery shockwave process is activated. The probability of recovery is presented as a function of the loop detector occupancy in the station of interest (downstream) and the upstream station. Each curve on the graph shows the combinations of upstream and downstream values of occupancy which result in the same probability of recovery. The curve of 0.5 shows the boundary beyond which the probability of a Recovery (State C) is more than the probability of Congestion (State B). The contour lines appear to be nearly linear, suggesting that a linear combination of the detector occupancies on the two stations results in a constant probability of queue dissipation.
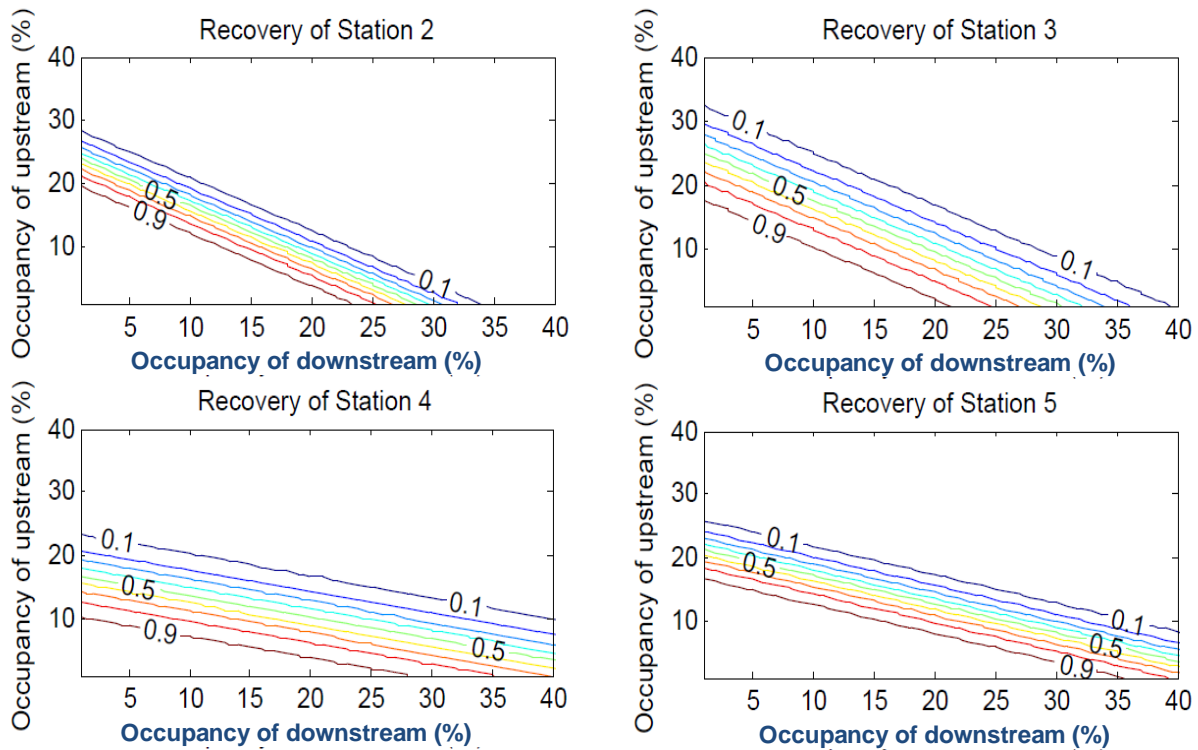
Figure 5-5: Probability of recovery based on the detector occupancy of the station of interest (downstream) and the upstream station (Model 2)

5.4 Model Evaluation

The proposed models were evaluated using loop detector data from 30 days randomly selected from the original data set. These data were not used for the model calibration.

5.4.1 Traffic Speed Prediction

In this section, we propose a framework to implement the proposed stochastic shockwave model (Model 2) to predict the speed of the traffic for the near future. The proposed framework of this study uses the stochastic shockwave model to predict the state of the traffic for the upstream segments of a freeway if traffic breaks down. As long as traffic state is uncongested for all the stations along the route of study (stationary conditions), traffic speed of a station can be predicted independently from the upstream and downstream stations. As soon as the traffic state is measured to be congested, the proposed shockwave model is activated, and the traffic state is predicted based on the measurements on the station of interest as well as the upstream and downstream stations. The proposed framework consists of two main modules: 1) a time-series predictor which is used while the traffic condition is stationary, and 2) a congestion detector.

5.4.1.1 Time-series Predictor

The framework can rely on any time-series model which uses the measurement of a fixed point detector. In this chapter, a simple time-series model is utilized to present the prediction framework and the capabilities of the proposed shockwave model. The time-series predictor in this chapter is called Naïve, and predicts the future traffic variables to be equal to the latest measurement (i.e. speed for time interval n+1 is predicted to be equal to the speed for time interval n).

5.4.1.2 Congestion Detector

For model calibration, traffic states were labeled in an off-line format meaning that for each time interval both the past and the future measurements were utilized. However, for a real-time application such as traffic speed prediction, a module is required to define the state of the traffic based on only past measurements. In this chapter, we consider two labeling methods which represent a lower and upper bound in practice. The first method is a naïve classifier in which we classify the state of the traffic on a station of a freeway as congested if it is in State 4 during the last 5 minutes and in State 3 or 4 during the 5 minutes before. We expect any reasonable classification model to perform at least as well as this method and therefore the performance using this model represents a lower bound on the performance expected in practice. We call this method "Shockwave 1". The second method uses the off-line labeling information and therefore represents the best case in terms of the performance of a real-time labelling method. We call this method "Shockwave 2".

In Chapter 6, a classification model is presented which works on a real-time basis, and it is shown to perform better than the naïve classifier.

5.4.1.3 Algorithm

1- Using the Congestion Detector, classify the state of the traffic for all the stations along the study route.

2- If there is at least one congested station, go to 4; otherwise go to 3.

3- Using the Time-series Predictor, predict the traffic speed for all the stations along the study route for the prediction horizon, and go to 1 for the next time interval.

4- Using the proposed shockwave model and the current measurements, predict the probability of spillback; if it is more than 50%, go to 5; otherwise go to 3.

5- Predict how far the queue will spillback during the prediction horizon using the transition intensities and the maximum wave speed on the fitted fundamental curve.

6- If congestion spills back to an upstream station, the speed at the upstream would be equal to the speed of the congested state. Update any upstream station which is affected, and go to 1 for the next time interval.

A similar algorithm is used for forward recovery shockwaves.

Table 5-3 presents the prediction results for the 30 evaluation days. Traffic speed is predicted at each freeway station for a 15 minute period. Since Station 1 does not have any upstream information and Station 5 does not have any downstream information, those two stations are not included in this table. The prediction error is stated based on mean absolute percentage error (MAPE). Table 5-3 is divided into two different time periods; "all day" shows the average prediction error for the entire day. Naïve shows the prediction results for the Time-series Predictor when the shockwave model is not implemented. The p-value reflects the statistical significance of the results. A p-value of less than 0.05 indicates a performance improvement (compared to the Naïve model) that is statistically significant.

From the results, it is clear that both the Shockwave 1 and Shockwave 2 models improved the prediction accuracy (when considering the All Day errors) and these improvements are statistically significant. However, looking at the daily average errors may be misleading because prediction errors for free flow conditions are very low and there are many hours in a day (especially early mornings and late nights) that the traffic condition is free flow.

The second part of the table shows the prediction results for the "change period" which is the period of time when traffic conditions are not steady state and a change in the traffic state may happen. This period includes all the instances in which one of the shockwave models was activated. It is clear that

the prediction errors and the improvements are much larger during these periods. Shockwave 1 reduced the prediction error by 13 to 16% and Shockwave 2 reduced prediction errors by 18 to 28%.

Table 5-3: Improvements in speed prediction accuracy resulting from the use of the proposed Markov model

Time Period	Description	Model	Value	Station 2	Station 3	Station 4
All day	Prediction Error	Naïve	MAPE (%)	14.91	14.16	12.84
		Shockwave 1	MAPE (%)	14.36	13.81	12.46
		Shockwave 2	MAPE (%)	13.78	13.66	12.14
	Improvement	Shockwave 1	Relative (%)	3.67	2.46	2.98
			Difference (%)	0.55	0.35	0.38
			Standard Error (%)	0.79	0.52	0.62
			p-value	0.001	0.001	0.002
		Shockwave 2	Relative (%)	7.56	3.54	5.40
			Difference (%)	1.13	0.50	0.69
			Standard Error (%)	1.10	0.70	0.91
			p-value	0.000	0.000	0.000
Change Period	Prediction Error	Naïve	MAPE (%)	76.77	66.68	53.37
		Shockwave 1	MAPE (%)	66.30	56.91	44.70
		Shockwave 2	MAPE (%)	57.81	54.89	38.31
	Improvement	Shockwave 1	Relative (%)	13.63	14.65	16.24
			Difference (%)	10.47	9.77	8.67
			Standard Error (%)	13.37	13.64	14.84
			p-value	0.000	0.000	0.003
		Shockwave 2	Relative (%)	24.69	17.68	28.22
			Difference (%)	18.95	11.79	15.06
			Standard Error (%)	20.58	16.94	21.83
			p-value	0.000	0.001	0.001

5.4.2 Prediction Rate

The proposed shockwave models are only activated when congestion is forming or dissipating. We identified each 5-minute time period during the 30 days of the test data for which the shockwave models were activated. Because we are applying the models to historical data, the actual end state is available for every observation. Therefore, we compared the states predicted by the proposed model with the actual observed state for each time period when the models were activated.

Table 5-4 presents the comparison between model predictions and observations. For example for backward forming shockwaves in Station 1, 68 cases of "Potential Spillback" were observed, and 44 of them eventually ended up in "Spillback" (State 3) and 24 in "Inactive Spillback". For the first group (44 Spillback cases), Shockwave Model 2 predicts 35 cases correctly, and for the second group (22 Inactive Spillback cases), the model predicts 17 cases correctly. The results for other stations and also forward

recovery shockwaves can be found in the table. The last column of Table 5-4 presents the accuracy of the classic (deterministic) shockwave theory, and can be compared with the accuracy of the proposed model. The proposed model has increased the average accuracy of classic shockwave theory from 42.8% to 72.5% for backward forming shockwaves and from 61.4% to 72.1% for forward recovery shockwaves.

Table 5-4: Comparison between observed states and model predictions

Backward Forming Shockwave							
Station	Number of activations	Observation		Prediction		Accuracy of Shockwave Model 2 (%)	Accuracy of shockwave theory (%)
		State	Number of Instances	Spillback (State 3)	Inactive (State 2)		
1	68	Spillback (3)	44	35	9	79.5	70.5
		Inactive (2)	24	7	17	70.8	25.0
2	61	Spillback (3)	48	33	15	68.8	22.9
		Inactive (2)	13	3	10	76.9	69.2
3	58	Spillback (3)	45	39	6	86.7	53.3
		Inactive (2)	13	6	7	53.8	38.5
4	49	Spillback (3)	23	17	6	73.9	0.0
		Inactive (2)	26	12	14	53.8	57.7
Forward Recovery Shockwave							
Station	Number of activations	Observation		Prediction		Accuracy of Shockwave Model 2	Accuracy of shockwave theory (%)
		State	Number of Instances	Recover (State 3)	Inactive (State 2)		
2	55	Recover (3)	44	23	21	52.3	47.7
		Inactive (2)	11	1	10	90.9	45.5
3	60	Recover (3)	50	37	13	74.0	66.0
		Inactive (2)	10	4	6	60.0	50.0
4	47	Recover (3)	41	33	8	80.5	75.6
		Inactive (2)	6	2	4	66.7	16.7
5	35	Recover (3)	32	28	4	87.5	78.1
		Inactive (2)	3	2	1	33.3	0.0

5.5 Conclusions

- A model has been developed to predict the speed of backward forming and forward recovery shockwaves. Unlike classic shockwave theory which is deterministic, the proposed model expresses the spillback and recovery as a stochastic process. The transition probability matrix is defined as a function of traffic occupancy on upstream and downstream stations. Then, the probability of spillback and recovery can be computed given the traffic conditions.

- An evaluation using field data has shown that the proposed stochastic model performs much better than a classical shockwave model in term of correctly predicting the occurrence of backward forming and forward recovery shockwaves on an urban expressway.

- A procedure has been suggested to improve the prediction error of a time series model by using the results of the proposed Markov model. It has been shown that the combined procedure significantly reduces the prediction error of a simple time series model. The proposed shockwave model can also be combined with the propagation model proposed in Chapter 4. The results of the combined model is presented in Chapter 8.

- For a real-time application of the proposed shockwave model, a module is required to simultaneously work with the shockwave model, and identify the state of the traffic based on the available measurements. In Chapter 6, a classification model is presented which works on a real-time basis, and it is feeds the estimated state of the traffic to the proposed shockwave model.

Chapter 6
Congestion Detection

As discussed in the previous chapters, the knowledge of a traffic breakdown on a section of a freeway may be utilized to increase the accuracy of traffic speed predictions on upstream sections. The proposed method for doing this was described in the previous Chapter. However, in order to be able to implement the proposed method, we need to have a mechanism which detects the state of the traffic in real-time. Using the real-time measurements of traffic variables, this mechanism needs to be able to determine if the traffic condition is becoming congested.

The goal of the model presented in this chapter is to identify the state of the traffic on each station of the freeway. This method forms the basis for the *congestion detection module*. If the *congestion detection module* determines that congestion is forming (or dissipating), then the *congestion spillback and dissipation module* is activated (i.e. the Markov model presented in Chapter 5) to update the prediction of traffic speed.

The congestion detection module tries to differentiate between congested and uncongested traffic conditions given the measurement of traffic variables. This may be easily formulated as a classification problem. The classes in this case are congested and uncongested traffic states, and the features are traffic speed, flow, and occupancy for the recent past.

6.1 Background

Figure 6-1 illustrates the measurements of speed and density at a loop detector station on a freeway; each point on the graph represent the traffic conditions for a 5 minute interval. As discussed in the previous chapters, the fundamental macroscopic relationship between traffic variables can be represented by the Van Aerde model (Van Aerde, 1995). This relationship relies on four parameters, namely free flow speed, speed at capacity, capacity, and jam density. Values for these four parameters can be found by calibrating the model to a set of empirical data. As described in Chapter 4, assuming an effectives average vehicle length, there exists a relationship between density and detector occupancy.

Speed at capacity can be used as a criterion to classify traffic conditions; observations with speed lower than speed at capacity are considered congested and those with higher values are considered uncongested. The same can be done using occupancy at capacity. It is clear that for some observations the two criteria are not consistent; there are some points on the graph which are considered congested by one criterion and uncongested by the other (see Figure 6-1). A method which considers both speed and occupancy may perform better in this case. In other words, the best classifier on this graph may not be a

simple vertical or a horizontal line. Rather, a method which considers all traffic variables may perform better and be more consistent with the theory of traffic flow.
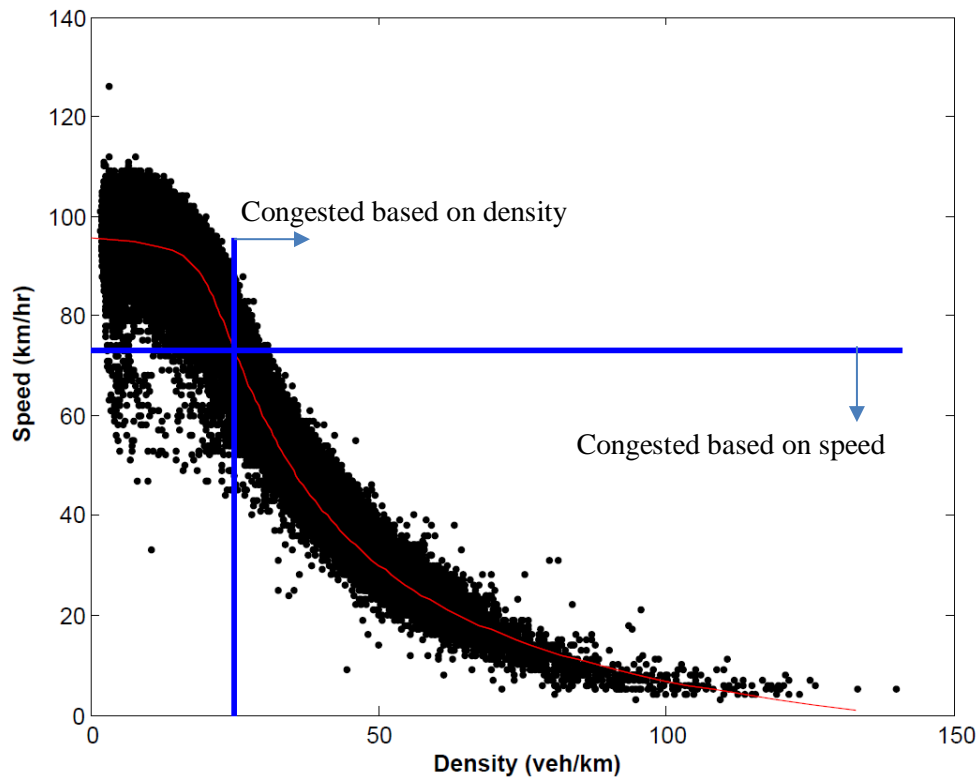
Figure 6-1: Speed-density measurement at a loop detector station on a freeway

Different methods have been presented to define and solve classification problems. They include Linear Discrimination Analysis (LDA), Naïve Bayes (NB), Logistic Regression (LR), Neural Network (NN), and Support Vector Machine (SVM); each of the methods are shown to provide promising results in some applications in the literature. In this research, we chose to use the SVM technique for the following reasons:

- In order to implement LDA, NB, or LR, the distribution of class features needs to be known; the first two approaches assume that the feature variables for each class are normally distributed, and the third approach assumes Weibull distribution. However, SVM does not require any knowledge about this distribution.

- The number of parameters to be estimated is lower in SVM compared to other parametric methods.

- Although NN is considered to be a flexible approach, it is vulnerable to overfitting, something that SVM is less susceptible to.

79

- SVM is a mathematical optimization problem and converges to a unique solution which is a discriminating hyperplane.

- The notion of a discriminating hyperplane is attractive as it is conceptually consistent with the use of a discriminating line as discussed in Figure 6-1.

6.2 Proposed Model

6.2.1 Traffic States

Since the proposed model of this chapter is going to be used as a real-time input into the *congestion spillback and dissipation module*, the traffic states in this chapter are exactly the same as those in Chapter 5. Therefore, traffic conditions at the location of each detector station are characterized by two different states: a) congested, or b) uncongested. Since we are interested in traffic conditions that result in shockwaves that propagate upstream or downstream and not short term local traffic disturbances, a congested traffic condition which last for more than one time interval (5 minute) is called a congested state in this module (Same as in Chapter 5).

6.2.2 Data Labeling

An off-line method is used to determine the traffic state (i.e. label as uncongested or congested) for each time interval in the field data set. An off-line method is used so that for each time interval both the past and the future measurements are assumed to be available.

As discussed in Chapter 7, traffic conditions are divided into 4 different states, and are labeled as State 1 to 4: State 1 is when traffic is completely free flow; State 2 is the second free flow condition; State 3 is the transition condition; and State 4 is congested. The congested state of this chapter relates to State 4 and the uncongested state relates to State 1 and State 2. The state of the traffic in this module is assumed to be congested if for two consecutive 5-minute time intervals the state of the traffic is not State 1 or State 2, and for at least one of the two time intervals the state of the traffic is State 4.

Traffic measurements in this module are aggregated into 1 minute intervals. For each interval, the state of traffic is labeled based on the measurements available for the 10 minutes in the future and the past. In other words, the future and the past information are aggregated into two 5 minute intervals as it is shown in Figure 6-2. The state of the traffic for time t is labeled as congested if the traffic state for time t is State 3 or State 4 and at least one of P1 or F1 intervals are in State 4, and the interval in State 4 is either followed or preceded by an interval in State 3 or State 4..

In order to determine the state of the traffic for each interval, the appropriate Van Aerde model (described in Chapter 3) has been used.

Time of interest

Interval P2   Interval P1   Interval F1   Interval F2

t-11   t-6   t-1   t   t+5   t+10   Time (min)
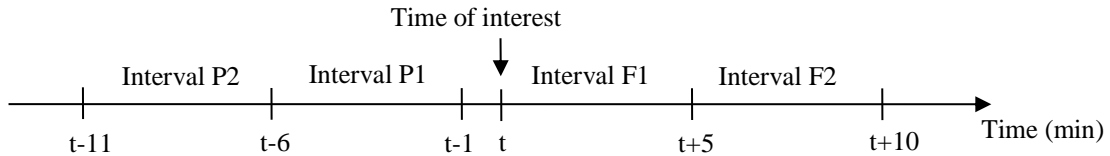
Figure 6-2: Schematic view of the time intervals used for data labeling

6.2.3 Feature Selection

The feature vector represents all the information available for the SVM model to classify the traffic state for a given time period. Since this module is required to be run in real-time, the feature vector can only include the information available in the past.

Traffic speed is considered as the most explanatory variable to define congestion; however, field observations show that 1 minute spot speed measurements experience a wide range of variability, and do not seem to be sufficient to characterize the state of the traffic (congested vs. uncongested). Therefore, the following points are considered to improve the chance of accurate detection:

- Using loop detector devices, we have access not only to the measurements of traffic speed, but also to the measurements of traffic flow and occupancy. A low speed observation on a freeway segment could be either a result of a local disturbance or a congested traffic condition; therefore, measured flow and occupancy provide more insight about the traffic state of the segment, and improve the probability of accurate classification. In this chapter, we have used two different feature vectors to calibrate the classification model: 1) a model based on traffic speeds only, and 2) a model based on speed, flow, and occupancy.

- Field observations show a wide range of fluctuations for 1 minute measurements. Therefore, the information contained within a 1 minute traffic measurement may not be stable enough to accurately classify the state of the traffic. Using a longer time period to aggregate data may reduce the chance of misclassification, but it may also result delay the detection of the congested conditions. Therefore, instead of using aggregated traffic measurements for a single, but longer, time period, we can use a vector of 1 minute measurements during a longer time interval.

In this chapter, the following alternatives are used as the feature vector for the classification model:

- Feature A: a vector of 1 minute measurements of traffic speed for the last 5 minutes. We call this model "SVM (speed)".

- Feature B: a vector of 1 minute measurements of traffic speed, flow and density for the last 5 minutes. We call this model "SVM (full)".

6.3 Model Training

The classification models proposed in this chapter are trained using MATLAB optimization. For each loop detector station a separate model has been trained. The traffic measurements at each station are aggregated into 1 minute intervals. Therefore, the state of traffic at each time is the variable of interest. Each one minute interval represents an observation in the training dataset. As mentioned in the previous section, each observation is labeled as either congested or uncongested based on the measurements from the future and past. The feature vector for each observation (1 minute interval) is constructed using the traffic measurements from only the past.

Since the fundamental relationship of traffic variables may be affected by seasonal variations, a separate model has been trained for each month of the year. Therefore, for each loop detector station 12 different models have been trained.

In order to evaluate the performance of the models, we need to define model misclassification rates to indicate the frequency of the cases for which the estimated class is not consistent with the assigned label. Suppose that d is an index representing the day and i is an index representing the 1-minute interval . Then the total and daily misclassification rates can be defined as follows:

$$TMCR^C = 100 \times \frac{\sum_d \sum_i L(c_{di}) \times M(c_{di})}{\sum_d \sum_i L(c_{di})} \tag{6-1}$$

$$TMCR^U = 100 \times \frac{\sum_d \sum_i (1 - L(c_{di})) \times (1 - M(c_{di}))}{\sum_d \sum_i (1 - L(c_{di}))} \tag{6-2}$$

$$MCR_d^U = 100 \times \frac{\sum_i L(c_{di}) \times M(c_{di})}{\sum_i L(c_{di})} \tag{6-3}$$

$$MCR_d^U = 100 \times \frac{\sum_i (1 - L(c_{di})) \times (1 - M(c_{di}))}{\sum_i (1 - L(c_{di}))} \tag{6-4}$$

Where,

$TMCR^{C}$, $TMCR^{U}$ = total misclassification rate for congested and uncongested conditions,

MCR^{C}_{d}, MCR^{U}_{d} = daily misclassification rate for day d for congested and uncongested conditions,

$L(c_{di})$ = 1 if traffic is labeled as congested for observation i of day d, and 0 otherwise,

$M(c_{di})$ =1 if the model classifies the traffic condition as congested for observation i of day d, and 0 otherwise.

The performance of the proposed models are compared to a naïve classification model in this study. The Naïve model uses the same logic as implemented in the labeling algorithm, but it only uses the traffic measurements from the past. Referring to Figure 6-2 , the state of the traffic for time t is congested if it is in State 3 or 4 and interval P1 is in State 4 and interval P2 is in State 3 or State 4.

Figure 6-3 illustrates the total misclassification rates for the proposed SVM classification models compared to the Naïve model. As mentioned earlier, a separate classification model has been trained for each month of the year to consider the seasonal effects. The results in Figure 6-3 show the performance of the models for the month of June. It should be noted that the SVM classification models do not perform better that the Naïve model when the state of the traffic is uncongested. However, for congested states, they outperform the Naïve model. Although the number of observations of uncongested states is much higher than the number of congested states, the overall performance of the SVM classification models seems to be better than the benchmark (Naïve model). This is true for all the stations of the study freeway route.

In order to investigate the performance of the classification models in more detail, the congested state of the study is divided into two different categories:

1) Congested observations which could be labeled as congested given only the information from the past. We call this category "Congested – Past Info", and

2) Congested observations which require the information from the future to be labeled as congested. We call this category "Congested – Future Info".

Figure 6-4 illustrates the total misclassification rates for each of these two categories of congested states. Although the overall misclassification rates of the Naïve model for congested states for June is less than 5 percent, the misclassification rate for "Congested – Future Info" is 30 to 50 percent. This situation is the part that we are most interested in. In other words, we are looking for a model which can detect the congested states as soon as possible. It should be noted that the misclassification rate of the proposed SVM models is less than 10% for this category.
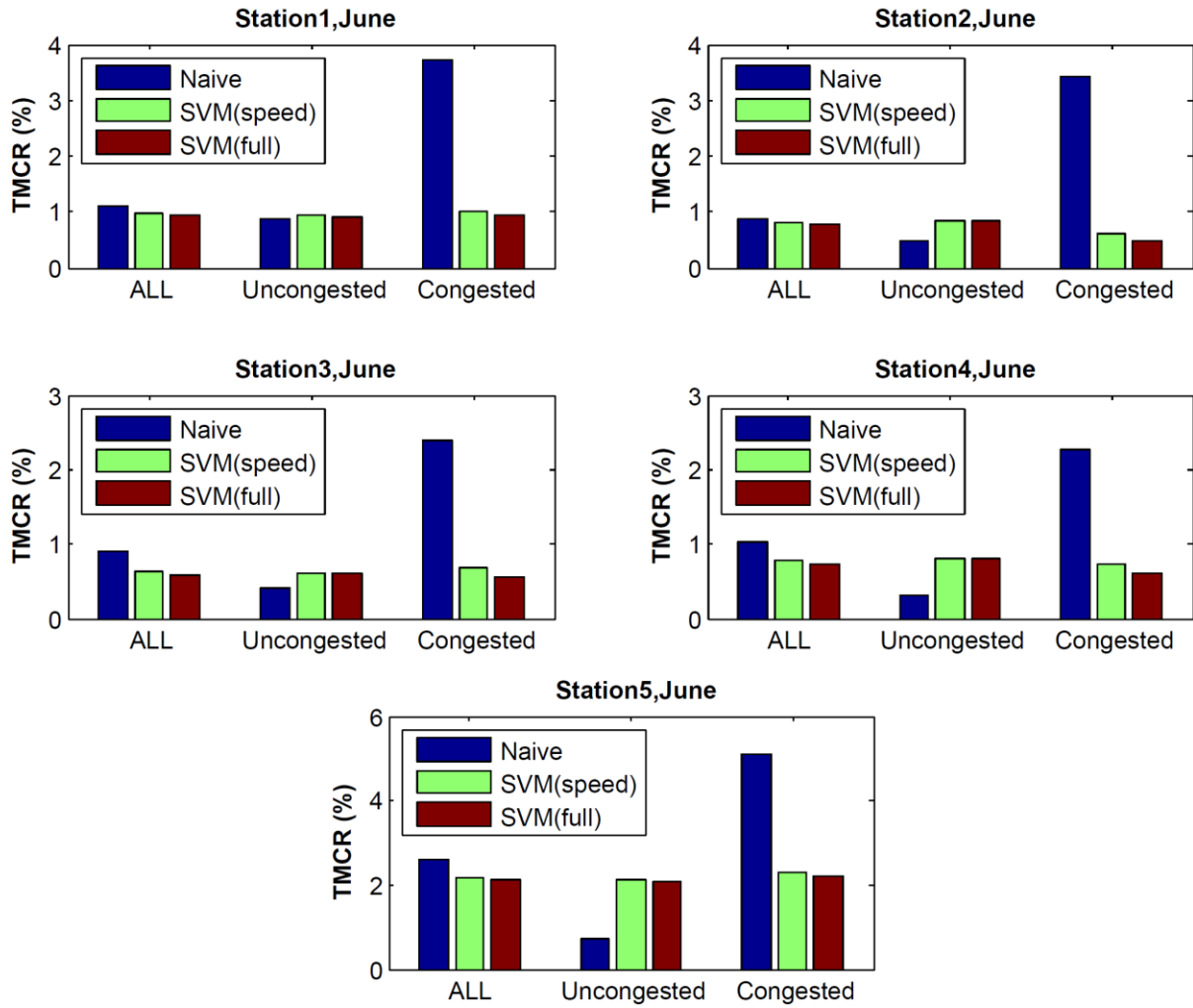
Figure 6-3: Misclassification rates of the SVM classification models for June
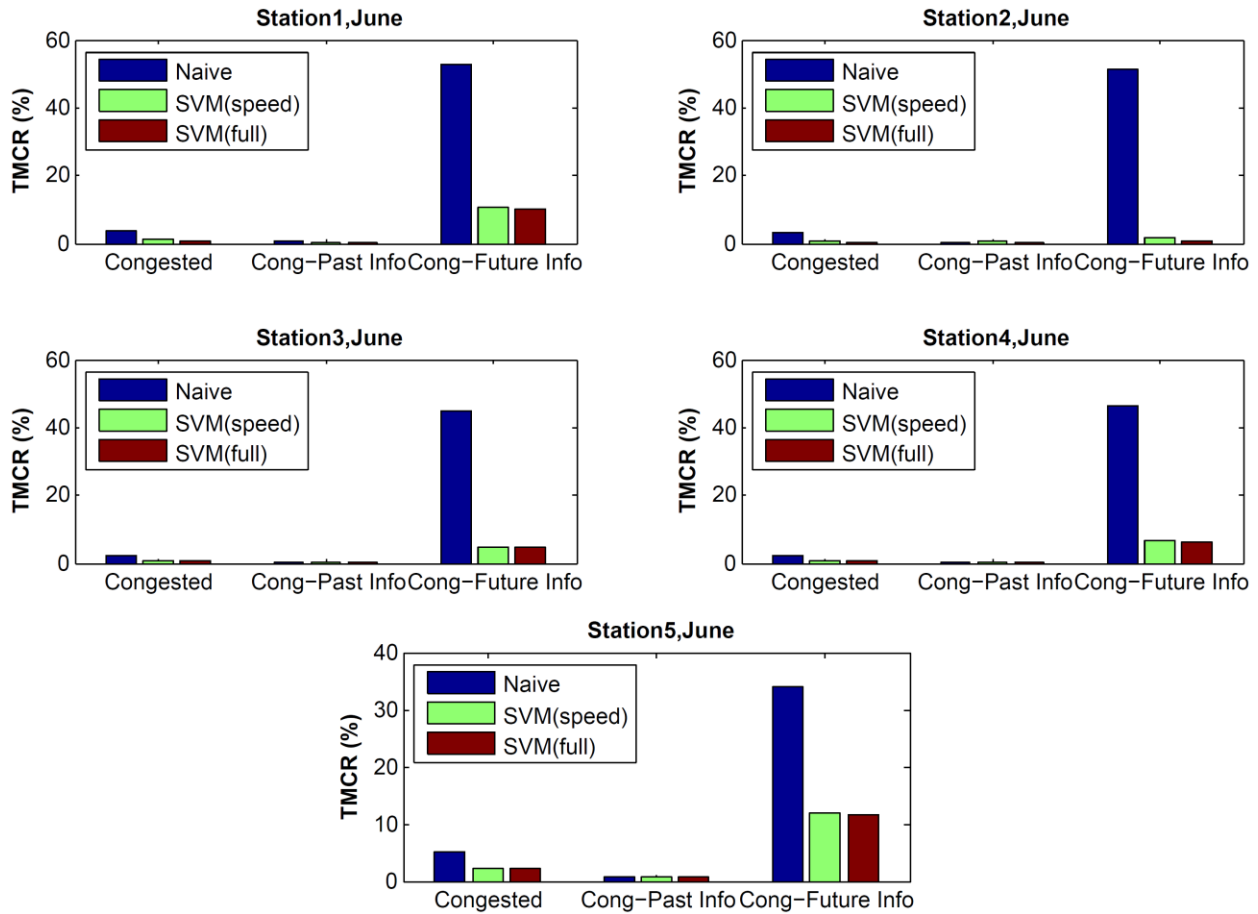
Figure 6-4: Misclassification rates of the SVM classification models for congested states for June

Table 6-1 presents the average daily misclassification rates for the training data set. It should be noted that the SVM (full) model reduces the misclassification rate of the Naïve model for both congested states and all observations.

Through the use of Student t test, it is also evaluated if this improvement is statistically significant.

Table 6-2 statistically compares the performance of the classification models. For all tests there are a total of 5760 observations (i.e. 30 days times 192 time intervals per day). The first column shows the hypothesis, and then the p-value for each category and detector station is provided. It is clear that the SVM (speed) model outperforms the Naïve model for congested states. However, this improved performance is not sufficient to make the SVM (speed) model statistically superior for the overall observations. In other words, we are 95 percent confident that the SVM (speed) model performs better than the Naïve model for stations 1, 4, and 5. However, the overall performance of the SVM (full) model is statistically better than both the Naïve model and the SVM (speed) model for all the stations of the study.

Table 6-1: Average daily misclassification rates (%)

Station	Model	All	Uncongested	Congested
1	Naïve	1.9	1.3	4.4
1	SVM (Speed)	1.9	2.3	0.8
1	SVM (full)	1.8	2.2	0.8
2	Naïve	0.8	0.3	2.7
2	SVM (Speed)	0.9	1.0	0.5
2	SVM (full)	0.7	0.8	0.5
3	Naïve	0.9	0.4	2.6
3	SVM (Speed)	0.9	1.0	0.5
3	SVM (full)	0.7	0.8	0.7
4	Naïve	1.0	0.5	2.5
4	SVM (Speed)	0.8	0.9	0.8
4	SVM (full)	0.7	0.7	0.7
5	Naïve	2.7	1.8	6.4
5	SVM (Speed)	2.5	2.6	2.3
5	SVM (full)	2.3	2.2	2.6

Table 6-2: statistical comparison between classification models

Hypothesis	Station	p-value of the test		
		All	Uncongested	Congested
Average daily misclassification rate for the SVM (speed) model is less than those of the Naïve model	1	0.0304	1.0000	0.0000
	2	0.9450	1.0000	0.0000
	3	0.2389	1.0000	0.0000
	4	0.0000	1.0000	0.0000
	5	0.0000	1.0000	0.0000
Average daily misclassification rate for the SVM (full) model is less than those of the Naïve model	1	0.0009	1.0000	0.0000
	2	0.0217	1.0000	0.0000
	3	0.0000	1.0000	0.0000
	4	0.0000	1.0000	0.0000
	5	0.0000	1.0000	0.0000
Average daily misclassification rate for the SVM (full) model is less than those of the SVM (speed) model	1	0.0000	0.0000	0.9283
	2	0.0059	0.0058	0.8794
	3	0.0002	0.0003	0.9647
	4	0.0000	0.0002	0.0127
	5	0.0000	0.0000	1.0000

6.4 Model Validation

6.4.1 Misclassification Rate

The proposed classification models were evaluated and compared to the performance of the Naïve model using loop detector data from 30 days randomly selected from the original data set. The selected days had not been used for the model calibration.

Figure 6-5 illustrates the average daily misclassification rates for the validation data set. The same pattern is observed here as it was for the calibration data set. The SVM (full) model is superior than the Naïve and the SVM (speed) models for all conditions and for only congested conditions. The range of the misclassification rates is not very different from the range observed for the calibration data set indicating that the model structure is not vulnerable to over-fitting issues.

Figure 6-6 depicts the misclassification rates of congested states for the validation data set in more detail. Similar to the results obtained for the calibration data set, these results indicate that the SVM (full) model outperforms both the Naïve and SVM (speed) models for the validation data set. The range of the errors is again similar to those obtained for the calibration data set.
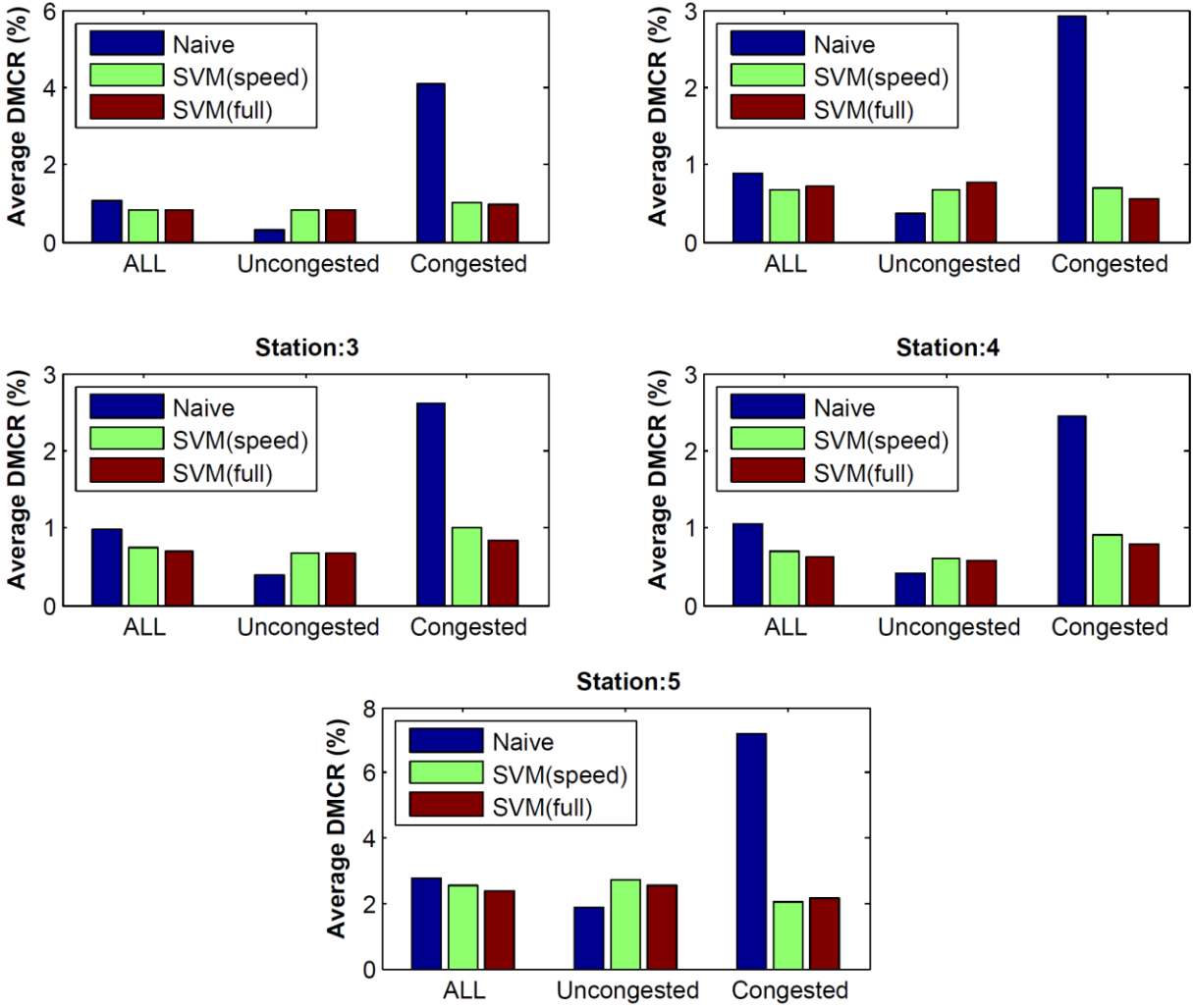
Figure 6-5: Misclassification rates for the SVM classification models for the validation data set
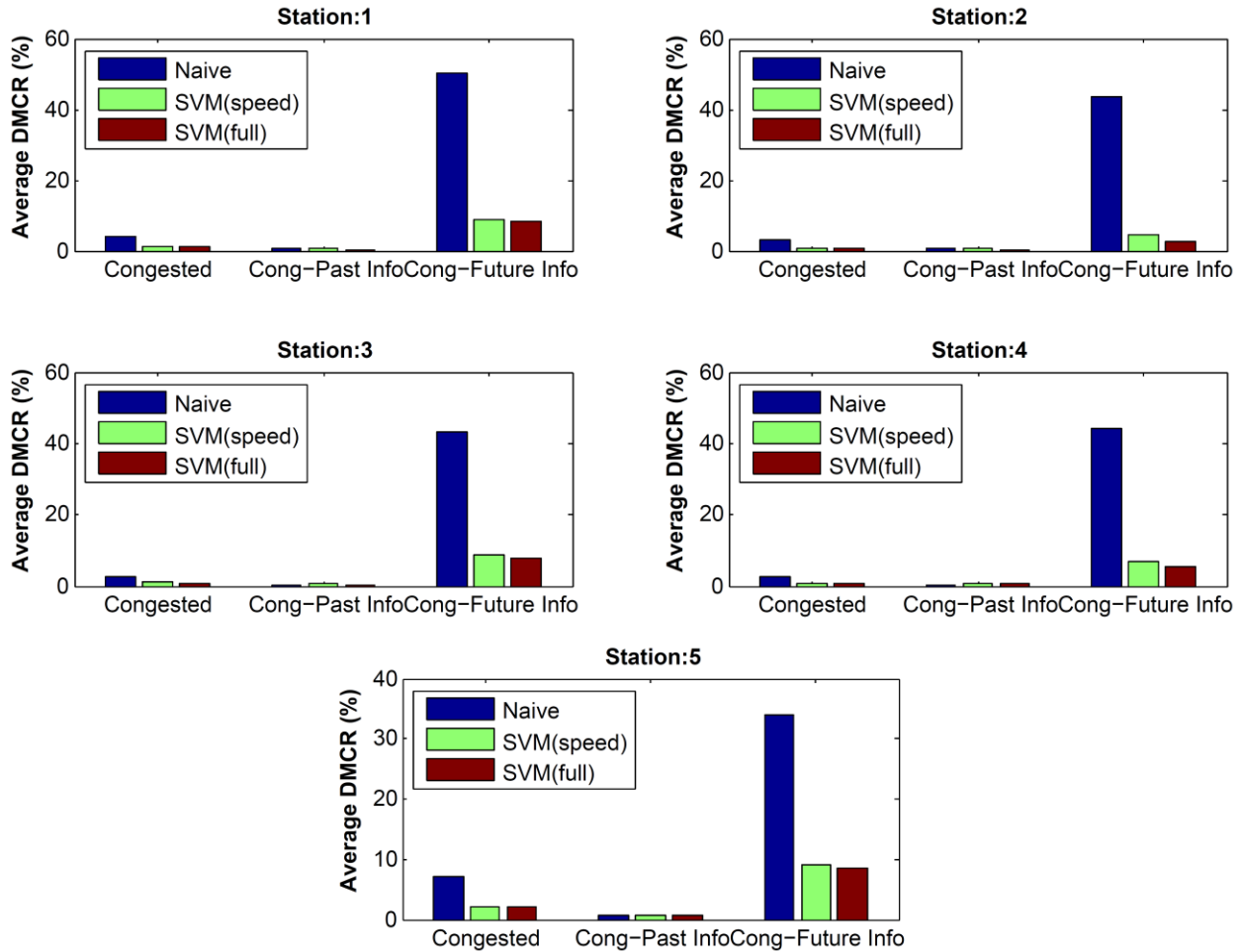
Figure 6-6: Misclassification rates for the SVM classification models for congested states for the validation data set

Table 6-3 presents the proportion of observations from the validation data set that fall into each category. Approximately 75% of each day, the roadway is uncongested and the remaining time, the roadway is congested. When the roadway is congested, approximately 93% of the time, the state can be correctly classified on the basis of past data only. For the remaining 7% of the time, past data alone is insufficient and it is for these conditions that the proposed classification models perform much better than the naïve model. It should also be noted that although these conditions do not occur frequency (only approximately 2% of the time), these are the times when travel time prediction is most valuable and most challenging.

89

Table 6-3: Proportion of the observed classes during the validation data set

Station	Portion of Each Class (%)			
	Uncongested	Congested	Congested - Past Info	Congested - Future Info
1	77.8	22.2	94.2	5.8
2	78.8	21.2	94.9	5.1
3	72.3	27.7	95.0	5.0
4	68.1	31.9	95.3	4.7
5	74.4	25.6	84.6	15.4
Average	74.3	25.7	92.8	7.2

6.4.2 Traffic speed prediction

Another approach to evaluate the proposed classification model is to investigate the contribution it may make in our proposed traffic stream speed prediction framework. In Chapter 5, a shockwave model has been proposed which requires a real-time state detection module. The performance of the shockwave model was evaluated using the Naïve state detection model. However, in this section the performance of the shockwave model using the proposed SVM (full) classification model is investigated.

The benchmark model in this section is the Chapter 5 Shockwave Model 2 using the Naïve classification model which we call "Shockwave – Base", and the proposed model is the Chapter 5 Shockwave Model 2 using the SVM (full) model which we call "Shockwave – SVM". Table 6-4 presents the traffic speed prediction results for the 30 validation [2]days. Traffic speed is predicted at each freeway station for a 15 minute period. Since Station 1 does not have any upstream information and Station 5 does not have any downstream information, those two stations are not included in this table. The prediction error is stated based on mean absolute percentage error (MAPE).

Table 6-4 is divided into two different time periods; "all day" shows the average prediction error computed using data from the entire day; and "Change Period" shows the average prediction errors for only those time periods when the traffic state changes.

The "All Day" results indicate that the Shockwave – SVM model provides lower prediction errors than the Shockwave – Base model. However, these improvements are small and are statistically significant only for station 4.

The "Change Period" results indicate that the prediction errors and the improvements provided by the proposed Shockwave- SVM model are much larger during these periods as compared with the All Day results. Furthermore, the improvements in performance are statistically significant for all three

---

[2] These are the same 30 days used for validating the SVM models in section 6.4.1.

stations except for Station 3. However, the p-value for this station is 0.0526 meaning that the improvement is statistically significant at 94 percent confidence level.

Table 6-4: Improvements in speed prediction accuracy resulting from the use of the proposed SVM classification model combined with the proposed shockwave model

Time Period	Description	Model	Value	Station 2	Station 3	Station 4
All day	Prediction error	Shockwave Base	MAPE (%)	14.36	13.81	12.46
		Shockwave SVM	MAPE (%)	14.22	13.75	12.31
	Improvement	Shockwave SVM	Relative (%)	1.0	0.5	1.2
			Difference (%)	0.14	0.06	0.14
			Standard Error (%)	0.87	0.30	0.41
			p-value	0.1876	0.1315	0.0319
Change Period	Prediction Error	Shockwave Base	MAPE (%)	66.30	56.91	44.70
		Shockwave SVM	MAPE (%)	61.67	54.89	38.89
	Improvement	Shockwave SVM	Relative (%)	7.0	3.6	13.0
			Difference (%)	4.6	2.0	5.8
			Standard Error (%)	14.71	6.64	15.0
			p-value	0.0475	0.0526	0.0213

6.5 Conclusions

This chapter documents the findings of a proposed congestion detection model as follows:

- The model is developed to estimate the current traffic state based on the available information from a fixed point detector. A binary model for the traffic state was considered i.e. free follow versus congested conditions.

- Two classification models were developed to identify traffic state; the first model only uses traffic speed to characterize the state of traffic (SVM –speed), and the second model uses speed. Flow, and occupancy for the classification (SVM-full). The model results were compared to a Naïve model. It was shown that the SVM-speed performs better compared the Naïve model, and SVM-full performs better when compared to SVM-speed.

- The SVM-full model was utilized as a state detection module to inform the Shockwave model developed in Chapter 5. The shockwave model of Chapter 5 used Naïve classifier to detect traffic state (Shockwave –Base), and the Naïve classifier was replaced by the SVM-full developed in this chapter (Shockwave-SVM).

- The combined Shockwave-SVM model showed little improvement in prediction error of traffic speed when averaged over the entire day; however, the model performed significantly better during the "Change Period. Therefore, it is concluded that the SVM model would provide improved prediction accuracy when traffic conditions are changing from free flow to congestions and vice versa.

Chapter 7
Traffic Speed Distribution[3]

In the previous chapters of this thesis, we proposed models to predict the mean traffic speed along a route on a freeway. In this chapter, a method to predict the variability of traffic speed in the near future is presented.

The objective of this chapter is to develop a method to characterize traffic conditions as a stochastic process, and utilize the results to predict the distribution of traffic speed for the near future. The methodology employs the concept of probabilistic breakdown and implements a Markov model to explain the stochastic changes in traffic conditions for each detector station. Using the maximum likelihood technique, a multi-state traffic model has been calibrated which defines the probability of transition among different traffic conditions as a function of traffic variables on the existing and downstream stations. Using the estimated transition probabilities, the characteristics of the near future traffic speed including mean, standard deviation, empirical distribution are predicted.

7.1 Background

A number of different methods have been proposed in the literature to provide accurate short-term predictions of traffic variables, and some methods seem to provide promising results. Nevertheless, according to van Hinsbergen et al. (2008) "… not one of the methods can be considered the best method in any situation". Most of the methods perform well when the traffic condition is stationary. However, when the traffic state starts to change (from free-flow conditions to congestion or vice versa) they generally suffer from a time lag to predict this change and therefore do not respond quickly. To tackle this problem, it is required to investigate the characteristics of traffic during congestion and the factors that cause congestion. Persaud et al. (1998) defined traffic breakdown as a "sudden and sharp drop in flow and speed" which is followed by congestion (i.e. formation of a queue of vehicles). Therefore, the breakdown phenomenon can be considered as a transition state between uncongested and congested conditions. However, field data for a given location shows that breakdown does not necessarily occur at the same level of flow rate (Elefteriadou et al 1995 and Persaud, 1998); and traffic conditions immediately prior to breakdown are considered as random variables. The probabilistic behavior of breakdown suggests that the traffic conditions of a given location follow a stochastic process.

---

[3] The content of this chapter has been incorporated within an article published by Noroozi and Hellinga: Noroozi, Reza, and Bruce Hellinga, 2014, "Real-time Prediction of Near-Future Traffic States on Freeways Using a Markov Model ", Transportation Research Record: Journal of the Transportation Research Board, Washington DC.

Evans et al. (2001) proposed a model to characterize the probability of breakdown at freeway merges using a Markov chain. They divided the merging area into different zones; and then the state of the traffic was defined as the possible combinations of whether a zone is occupied. The transition probability between states was calculated based on arrival rate of the traffic. Then they calculated the probability of breakdown for different arrival rates.

Kharoufeh and Gautam (2004) proposed an analytical procedure to construct the travel time distribution for a vehicle traversing a link of arbitrary length. They assumed that roadway, traffic, and environmental factors can be modeled as a random environmental process which modulates the speed of the vehicle. Assuming the process follows a continuous time Markov chain, they analytically derived the distribution of travel time for individual vehicles. However, they assumed that the infinitesimal generator matrix which explains the transition probabilities of the environmental process is available.

Yeon et al. (2008) proposed a method to estimate the travel time on a freeway route using a discrete Markov chain. First, they divided the study route into different links. The states of each link correspond to whether or not the link is congested, and the states of the route correspond to all possible combinations of links states. Then, they estimated the transition probability matrix based on the percentage of each transition in the historical observations. Using the steady-state probabilities and the average speed of each state, they estimated the expected value of the route travel time.

Dong and Mahmassani (2009) proposed a methodology to predict travel time reliability assuming a probabilistic nature for traffic breakdown. They utilized a Markov chain model to characterize the transition probabilities between the states, and the transition probability matrix was estimated to be constant for different time of day and different environmental conditions. Assuming travel time is normally distributed for a short time period, they proposed a bimodal distribution (uncongested/congested) to predict/estimate the distribution of travel time.

Given the available literature, it appears that the following issues remain to be addressed:

- Most previous studies in which the Markov model has been used have assumed that the transition probability matrix is constant for different environmental conditions. This assumption is frequently violated leading to inaccuracies in the model prediction. The few studies which have included environmental conditions present a separate transition matrix for each environmental condition rather than incorporating these conditions directly into the model. This suggests that the effect of different factors on transition probabilities should be investigated and if significant factors are found, that a Markov model with covariates (including both discrete and continuous) be calibrated to characterize traffic conditions as a stochastic process.

94

- The application of a Markov model for real time prediction of travel speed and its distribution still needs to be fully investigated.

7.2 Proposed Model

7.2.1 Traffic States

Traffic states in this chapter are defined on the same basis described in chapter 5. In this chapter, four traffic states are considered including two free flow states, one transition state, and one congested state. Traffic states are labeled as State 1 to 4: State 1 is when traffic is completely free flow; State 2 is the second free flow condition; State 3 is the transition condition; and State 4 is congested.

7.2.2 State Transition

The main objective of a Markov model is to estimate the probability that the process ends up in each state after a given time period. These probabilities are estimated on the basis of the transition frequencies in the historical traffic observations. Table 7-1 presents the relative transition frequencies for each pair of states for a five minute interval. The sum of the transition relative frequencies of each detector (4 by 4 matrix) is equal to 100%. There are some cells in Table 7-1 that show very low relative frequencies. Looking at the transition frequencies of Detector 1, the summation of relative frequencies from State 1 to 3, State 1 to 4, State 2 to 4, State 3 to 1, State 4 to 1, and State 4 to 2 is less than 0.5 percent. The same observation is true for Detector 2. Therefore, it can be concluded that it is unlikely (very rare) for traffic conditions to change from one state to a very different state in a single 5-minute interval. For example, if the current traffic state is 2 (Free Flow 2), it is highly unlikely that in the next 5-minute interval traffic conditions will change to State 4 (Congestion). In other words, traffic does not transition from State 2 directly to 4; rather it first transitions to State 3 and then to State 4. As a result, in the proposed Markov model the rare occurrences (negligible frequencies) of transitions between states are ignored. This is done by setting the intensity of transition (λ) for those states equal to zero meaning they do not communicate directly; however, there still exists a chance of transition through an intermediate state. Therefore, the proposed structure for the Markov model is as illustrated in Figure 7-1.

The term λ_{ij} in Figure 7-1 represents the transition intensity from State i to State j, and is defined as follows:

$$\lambda_{ij} = \lim_{\Delta t \to o} \frac{\Pr(X(t + \Delta t) = j \mid X(t) = i)}{\Delta t} \qquad \textbf{(7-1)}$$

Where $X(t)$ is the state of the traffic at time t.

The next section describes how the transition probability matrix is calculated on the basis of the transition intensity matrix defined in Equation 7-1.

Table 7-1: Relative state transition frequencies for a 5 minute interval during the study period (%)

Detector 1				
From	To			
	State 1	State 2	State 3	State 4
State 1	39.40	7.16	0.10	0.00
State 2	7.01	18.93	1.67	0.04
State 3	0.24	1.50	2.97	1.23
State 4	0.01	0.06	1.20	18.46

Station 2				
From	To			
	State 1	State 2	State 3	State 4
State 1	59.04	3.04	0.17	0.01
State 2	2.75	3.36	0.93	0.03
State 3	0.44	0.63	3.50	2.92
State 4	0.02	0.04	2.89	20.24

Station 3				
From	To			
	State 1	State 2	State 3	State 4
State 1	43.25	4.52	0.15	0.01
State 2	4.14	6.38	1.17	0.02
State 3	0.48	0.81	4.56	3.60
State 4	0.01	0.04	3.58	27.29

Station 4				
From	To			
	State 1	State 2	State 3	State 4
State 1	23.31	6.21	0.12	0.00
State 2	5.80	16.85	1.64	0.03
State 3	0.36	1.36	4.58	2.20
State 4	0.01	0.03	2.19	35.33
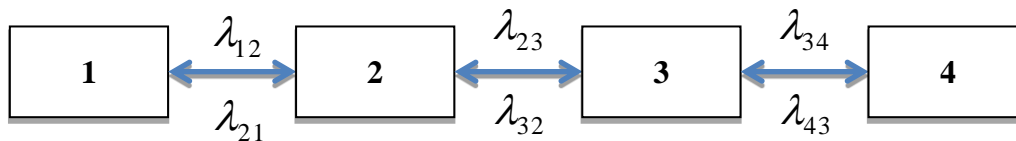
Figure 7-1: The proposed structure of the Markov model

7.3 Model Calibration

The same methodology of Chapter 5 was used to calibrate the model, and estimate the model parameters.

The simplest model to calibrate is a model which considers the state transition intensities to be constant for any given condition. This model assumes the intensity of transition from state i to state j is not affected by traffic conditions or any other factor. Therefore, the probability of transition is only affected by the length of the time interval. This model is called "Base model" in this chapter, and is

designated as Base Model. Using maximum likelihood estimator, the transition intensity matrix for the Base Model has been estimated. The statistical results of calibrating the Base Model are presented in Table 7-2. The parameters estimates (transition intensities), and their confidence intervals for all the stations are presented in this table, and the transition probability matrix can be calculated using Equation 7-2. For example for a 5 minute interval, the transition probability matrix for Station 4 is calculated as follows:

$$\text{Pr}(5\min) = \begin{bmatrix} 0.873 & 0.123 & 0.003 & 0.000 \\ 0.203 & 0.751 & 0.041 & 0.005 \\ 0.025 & 0.179 & 0.629 & 0.167 \\ 0.001 & 0.006 & 0.047 & 0.946 \end{bmatrix}$$

Table 7-2: Calibration results for the Base model

Station	Variable		State 1	State 2	State 3	State 4
1	Base intensities	State 1	-0.15377 (-0.1586,-0.14906)	0.15377 (0.1491, 0.1586)		
		State 2	0.25378 (0.2459, 0.26188)	-0.31437 (-0.32337,-0.3056)	0.06059 (0.0568, 0.06462)	
		State 3		0.26332 (0.24606, 0.2818)	-0.47903 (-0.5037,-0.45553)	0.21570 (0.20013, 0.23249)
		State 4			0.06085 (0.0564, 0.06566)	-0.06085 (-0.06566,-0.05640)
2	Base intensities	State 1	-0.04892 (-0.05134,-0.0466)	0.04892 (0.0466, 0.05134)		
		State 2	0.39029 (0.3710, 0.416)	-0.52283 (-0.5462,-0.5004)	0.13253 (0.1215, 0.14457)	
		State 3		0.08961 (0.0806, 0.09959)	-0.50352 (-0.5265,-0.48156)	0.41391 (0.39405, 0.43477)
		State 4			0.12480 (0.1188, 0.13113)	-0.12480 (-0.13113,-0.11878)
3	Base intensities	State 1	-0.09464 (-0.0975,-0.09187)	0.09464 (0.09187, 0.0975)		
		State 2	0.35408 (0.3433, 0.3653)	-0.45412 (-0.46674,-0.4418)	0.10004 (0.0943, 0.10606)	
		State 3		0.09015 (0.08404, 0.0967)	-0.49161 (-0.5066,-0.47705)	0.40146 (0.38832, 0.41504)
		State 4			0.11592 (0.1121, 0.11985)	-0.11592 (-0.11985,-0.11211)
4	Base intensities	State 1	-0.2104 (-0.2156,-0.2053)	0.2104 (0.2053,0.2156)		
		State 2	0.2388 (0.2328,0.2449)	-0.3062 (-0.3131,-0.2995)	0.06747 (0.06435,0.07076)	
		State 3		0.1669 (0.1585,0.1759)	-0.4369 (-0.4513,-0.4231)	0.27 (0.2591,0.2813)
		State 4			0.05826 (0.05591,0.0607)	-0.05826 (-0.0607,-0.05591)

Numbers are the calibrated transition intensities (λ_{ij})
Numbers in the brackets are the 95% confidence interval of the transition intensity estimates
Values of zero indicate a transition that is not possible (at least not in a single step) for the proposed Markov model

97

The estimates provided by the Base model represent the average probability of transitions for different possible situations. As stated in the previous sections, this way to calibrate a Markov model is the method most commonly found in traffic literature. However, the probability of transition may be affected by different factors, and the model should be able to capture the effect of those factors on the transition probabilities. Some factors affecting transition probabilities in this research are explained as follows:

- Markov models are usually assumed to be time homogenous meaning that the transition intensities are assumed to be the same throughout the entire period of study (a day in this case). However, observations of traffic conditions such as the average time spent in each state suggest that the transition intensities may vary for different times of day. Therefore, we define the binary variable *pm* to investigate the effect of time of day. *pm* is equal to 1 if time of day is between 12:00 noon and midnight, and 0 otherwise.

- In Markov models the future state of the process is characterized on the basis of the state of the process at the present time. However, we expect that the future state of the traffic on a given station of a freeway is affected not only by the current state of traffic at that location, but also the current state of the traffic at the next downstream station. For example, traffic is more likely to break down on a section when traffic on the next downstream section is congested (i.e. under these conditions it is more likely that the downstream congestion will spill back upstream and cause congestion on the upstream section). Therefore, we define a binary variable *ds* to investigate the effect of traffic conditions of the downstream station. *ds* is equal to 1 if the traffic state of the downstream station is congested (i.e. State 4), and 0 otherwise.

- Markov models are usually considered as memoryless meaning that the future state depends only on the current state and not on previous states. However, we expect that the previous traffic conditions may affect the future state of the traffic. Therefore, we define a continuous variable *ocp* to investigate the effect of traffic conditions during the previous time interval on transition intensities. *ocp* is equal to the measured station occupancy during the previous 5 minute interval for the study station. This new variable adds a short memory to the proposed Markov model. Therefore, the proposed model would not be classified as a pure Markovian process, and is categorized as a Markov process with a short memory.

We propose a Markov model that considers the effect of the abovementioned factors on the transition intensities. The statistical results of calibrating the Proposed Model are presented in Table 7-3.

The parameter estimates and their confidence intervals for all the stations are presented in this table. The transition intensity from State i to State j can be calculated as follows:

$$\lambda_{ij} = \begin{cases} \lambda_{0,ij} \times e^{(\alpha_{ij} \times pm + \beta_{ij} \times ds + \gamma_{ij} \times ocp)} & \forall i \neq j \\ -\sum_{k \neq i} \lambda_{ik} & \forall i = j \end{cases} \tag{7-2}$$

Where,

λ_{ij} = transition intensity from state i to state j,

$\lambda_{0,ij}$ = base transition intensity from state i to state j (from Table 7-3) ,

$\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ = log-linear effects of covariates (from Table 7-3), and

pm, ds, ocp = values of covariates.

The proposed model includes 6 independent transition intensities (i.e. λ_{ij}). The value of transition intensity from one state to itself is calculated as negative one times the summation of all the intensities in that row (i.e. the summation of all the elements on one row of the intensity matrix must be equal to zero). Therefore, the matrix of log-linear effects of each covariate includes 6 non-zero elements. The confidence interval of each element indicates whether the effect of that element is statistically significant (The estimate is considered to be statistically significant if confidence interval does not include the value of zero).

from The results in Table 7-3 indicate that the log-linear effects of covariate *pm* are statistically significant for all the stations (6 out of 6 for Station 1, 5 out of 6 for Station 2, 6 out of 6 for Station 3, and 5 out of 6 for Station 4). Therefore, it can be concluded that the resulting transition probability matrix for the PM period is statistically different from that for the AM period, and this result corroborates the speculation that a homogenous Markov model may not be sufficient to explain the transition probabilities for the entire day.

The results in Table 7-3 show that the log-linear effects of covariate *ocp* are statistically significant for all the stations (4 out of 6 for Station 1, and 6 out of 6 for all other stations). The log-linear effects of *ocp* on the upper diagonal (i.e. transitions from State 1 to 2, 2 to 3, and 3 to 4) are positive implying that the higher the occupancy is, the more likely the state of the traffic changes to a more congested state. On the other hand the effect of *ocp* on the lower diagonal (i.e. transitions from State 2 to 1, 3 to 2, and 4 to 3) are all negative meaning that the higher the occupancy is, the less likely the traffic recovers to a less congested state. Therefore, there is a statistically significant relationship between the transition probabilities and traffic occupancy during the previous time interval and this relationship is consistent with traffic engineering fundamentals.

The parameters estimated for the downstream station (ds) covariate confirm that there is a statistically significant relationship between the transition intensities and the traffic conditions on the downstream station. When the downstream station is congested, the traffic condition on the study station is more likely to transition to a more congested state (i.e. from State 2 to 3 or from 3 to 4), and it is less likely to transition to a less congested state (i.e. from State 4 to 3 or from 3 to 2). However, it seems that the condition of the downstream station does not have a statistically significant effect on the transitions between states 1 and 2.

Table 7-3: Calibration results for the proposed Markov model

Station	Variable		State 1	State 2	State 3	State 4
1	Base intensities	State 1	-0.04626 (-0.05031,-0.04254)	0.04626 (0.04254, 0.05031)		
		State 2	1.89377 (1.50823, 2.37787)	-1.94661 (-2.42946,-1.55972)	0.05283 (0.04018, 0.06947)	
		State 3		0.87574 (0.74584, 1.02827)	-1.00175 (-1.15573,-0.86828)	0.1260 (0.10144, 0.15651)
		State 4			0.24217 (0.15970, 0.36724)	-0.2422 (-0.36724,-0.1597)
	Effect of pm period	State 1		**-0.3795** (-0.4443,-0.31460)		
		State 2	**0.5161** (0.4479, 0.58427)		**-0.2296** (-0.3747,-0.08448)	
		State 3		**-0.4501** (-0.6143,-0.28597)		**0.8892** (0.6736, 1.10473)
		State 4			**-0.6557** (-0.9088,-0.40259)	
	Effect of occupancy of the previous time interval	State 1		**0.195714** (0.18715, 0.20427)		
		State 2	**-0.247186** (-0.27156,-0.22282)		0.022860 (-0.003969, 0.049690)	
		State 3		**-0.05353** (-0.06367,-0.04339)		**0.01577** (0.0249,0.0067)
		State 4			-0.006008 (-0.017192, 0.005177)	
	Effect of downstream station	State 1		0.10127 (-0.06259, 0.26513)		
		State 2	**-0.11912** (-0.23531,-0.00293)		**0.49408** (0.27013, 0.71802)	
		State 3		-0.06849 (-0.23533, 0.09835)		**0.2871** (0.09191, 0.48229)
		State 4			**-0.79592** (-0.97316,-0.61869)	
2	Base intensities	State 1	-0.01555 (-0.01711,-0.01414)	0.01555 (0.01414, 0.01711)		
		State 2	0.85372 (0.66421, 1.09731)	-1.12399 (-1.38538,-0.91191)	0.27026 (0.18911, 0.38624)	
		State 3		0.33876 (0.25316, 0.45331)	-0.57827 (-0.69606,-0.48040)	0.2395 (0.20104, 0.28533)
		State 4			2.23382 (1.69539, 2.94326)	-2.23382 (-2.9433,-1.6954)
	Effect of pm period	State 1		-0.07995 (-0.1776, 0.01773)		
		State 2	**0.22540** (0.1233, 0.32747)		**0.41730** (0.2431, 0.59149)	
		State 3		**-0.60166** (-0.8341,-0.36924)		**0.60735** (0.4662, 0.74856)
		State 4			**-0.84921** (-0.9909,-0.70748)	
	Effect of occupancy of the previous time interval	State 1		**0.20006** (0.18957, 0.210550)		
		State 2	**-0.11397** (-0.14549,-0.082438)		**0.01238** (0.007999, 0.01676)	
		State 3		**-0.06419** (-0.08212,-0.04625)		**0.01163** (0.00427, 0.019)
		State 4			**-0.04912** (-0.05971,-0.038524)	
	Effect of downstream station	State 1		**0.64281** (0.09512, 1.1905)		
		State 2	-0.03999 (-0.32853, 0.2485)		**1.43524** (1.17443, 1.6960)	
		State 3		0.11582 (-0.11661, 0.3483)		**0.48376** (0.37365, 0.5939)
		State 4			**-0.97710** (-1.09388,-0.8603)	

Numbers in the brackets show the 95% confidence interval of the estimates. The bold numbers show the statistically significant log-linear effects of covariates.
Effects of the variables are stated in log-linear scale

Continued: Calibration results for the proposed Markov model

Station	Variable		State 1	State 2	State 3	State 4
3	Base intensities	State 1	-0.02331 (-0.02537,-0.02142)	0.02331 (0.02142, 0.02537)		
		State 2	0.77560 (0.65221, 0.92233)	-0.86806 (-1.01506,-0.74235)	0.09246 (0.07483, 0.11424)	
		State 3		0.34959 (0.28837, 0.42381)	-0.51149 (-0.58690,-0.44578)	0.1619 (0.14269, 0.18369)
		State 4			3.55162 (2.86462, 4.40337)	-3.55162 (-4.4034,-2.8646)
	Effect of pm period	State 1		**0.3401** (0.2790, 0.40119)		
		State 2	**0.1617** (0.0972, 0.22621)		**-0.3160** (-0.4309,-0.20105)	
		State 3		**-0.2344** (-0.3803,-0.08854)		**0.4047** (0.3163, 0.49315)
		State 4			**-0.5496** (-0.6322,-0.46695)	
	Effect of occupancy of the previous time interval	State 1		**0.183271** (0.172895, 0.1937)		
		State 2	**-0.110382** (-0.130983,-0.0898)		**0.034139** (0.0113615, 0.0569)	
		State 3		**-0.06069** (-0.072718,-0.0487)		**0.00659** (0.0006, 0.01258)
		State 4			**-0.072693** (-0.080273,-0.0651)	
	Effect of downstream segment	State 1		0.56200 (-1.0485, 2.17245)		
		State 2	-0.06975 (-0.7841, 0.64462)		0.10539 (-0.9503, 1.16106)	
		State 3		**-0.26094** (-0.4346,-0.08733)		**0.73114** (0.6456, 0.81670)
		State 4			**-1.07899** (-1.1935,-0.96452)	
4	Base intensities	State 1	-0.04601 (-0.04966,-0.04263)	0.04601 (0.04263,0.04966)	0	0
		State 2	1.421 (1.176,1.716)	-1.444 (-1.739,-1.199)	0.02281 (0.01927,0.027)	0
		State 3	0	0.559 (0.4806,0.6502)	-0.6542 (-0.7454,-0.5741)	0.09515 (0.08341,0.1086)
		State 4	0	0	13.06 (10.4,16.4)	-13.06 (-16.4,-10.4)
	Effect of pm period	State 1	0	**0.2321** (0.1834,0.2807)	0	0
		State 2	**-0.2175** (-0.2694,-0.1655)	0	**-0.1117** (-0.2084,-0.01496)	0
		State 3	0	-0.0026 (-0.11,0.1048)	0	**0.2553** (0.1687,0.342)
		State 4	0	0	**-0.8293** (-0.9259,-0.7327)	0
	Effect of occupancy of the previous time interval	State 1	0	**0.2053** (0.1957,0.2148)	0	0
		State 2	**-0.2079** (-0.2314,-0.1843)	0	**0.1292** (0.112,0.1463)	0
		State 3	0	**-0.07785** (-0.08728,-0.06843)	0	**0.04934** (0.04319,0.05549)
		State 4	0	0	**-0.1777** (-0.1874,-0.1679)	0
	Effect of downstream segment	State 1	0	-0.8125 (-1.16,-0.4651)	0	0
		State 2	-0.1998 (-0.5613,0.1616)	0	**1.295** (1.037,1.553)	0
		State 3	0	**-0.4562** (-0.6691,-0.2432)	0	**0.7006** (0.5935,0.8078)
		State 4	0	0	**-1.197** (-1.326,-1.068)	0

Numbers in the brackets show the 95% confidence interval of the estimates
Effects of the variables are stated in log-linear scale
Bold values are statistically significant at the 95% confidence level

A model with five traffic states was also investigated. The five state model was constructed by dividing the transit state into two separate states. The model was calibrated using the same field data and the same method as the proposed four state model. A comparison between the two model indicated that there was no statistically significant difference in the transition probabilities from/to the transient states between the two models and therefore there is no evidence that the model with five states is superior to the proposed model for the investigated freeway stations. Consequently, the five state model was not pursued further and only the four state model was used.

Figure 7-2 illustrates the estimated transition probabilities for a 5 minute interval during the PM period for a Station 1. The graphs in Part (a) show the transition probabilities from a given state to all other states when the downstream station is not congested, and the graphs in Part (b) show the same probabilities when the downstream station is congested.
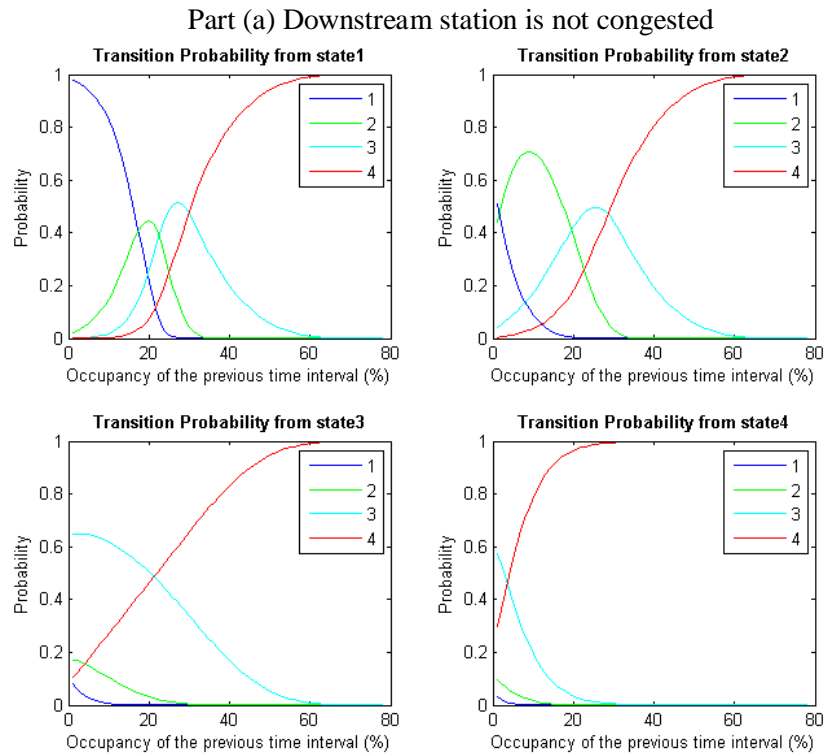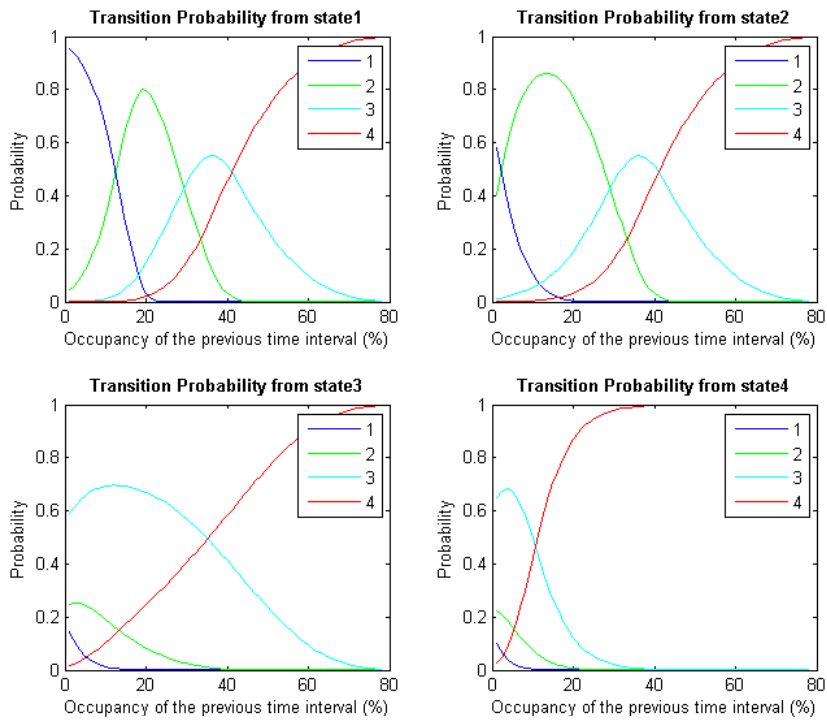
Part (a) Downstream station is not congested

Part (b) Downstream station is congested

Figure 7-2: Estimated 5 minute transition probabilities (PM period) for a sample station

104

7.4 Prediction of Traffic Speed Distribution

The Markov model proposed in the previous section provides the ability to predict, in real-time, the future state of traffic on a section of freeway. In this section, we use the proposed Markov model to predict the speed of the traffic stream on a section of freeway in a future time period.

As discussed in a previous section, traffic speed on a freeway station is considered as a random variable in the literature. Some previous research tried to investigate the distribution of traffic speed/travel time, and it has been shown that the distribution of speed tends not to be symmetrical. Li et al (2006) suggest a log-normal distribution for travel time. However, the empirical observations for travel speed show that the shape of the distribution highly depends on the traffic conditions; specifically it is more skewed for transition and congested states, and less skewed for free flow conditions..

Suppose that $U_d(t)$ is the random variable representing spot speed at time t on station d of a freeway. Based on historical measurements, the empirical distribution of speed, $f_d(u)$, can be constructed for each station of the freeway. Although $f_d(u)$ is the empirical probability density function of speed on station d, it does not provide an accurate explanation of the variability of speed for a specific time t because the traffic conditions prior to time t affect the range of the values that speed may possibly hold at time t ($U_d(t)$). Therefore, the random variable of interest would be $U_d(t)|C(t)$; where $C(t)$ represents traffic conditions at time t. In other words, the empirical distribution of speed at time t needs to be constructed based on the observations which share similar conditions.

It is clear that the range of the variation of speed for each traffic state should be different. Therefore, the first condition to which the distribution of speed at time t should be subjected is the state of the traffic. However, the state of the traffic for the near future is not known before it really happens, and traffic state itself is a random variable. Therefore, the empirical probability density function of speed for the near future can be predicted as a mixed distribution as follows:

$$g_d(U_d(t) \mid X(t_0) = i) = \sum_j p_{ij}(\Delta t, t_o) \times f_d(u \mid X = j) \tag{7-3}$$

Where,

$g_d(U_d(t) \mid X(t_0) = i) =$ empirical probability density function of speed given traffic is in state i at time t_0,

$p_{ij}(\Delta t, t_0) =$ transition probability from state i to j for a prediction horizon of Δt at time t_0 (these transition probabilities are obtained from the Markov model proposed in the previous section)

$f_d(u \mid X = j) =$ empirical probability density function of speed given traffic is in state j, and

$t_0 =$ present time,

It can be shown that the mean and the variance of the empirical distribution for time $t = t_o + \Delta t$ can be calculated as follows:

$$E[U_d(t) \mid (X(t_0) = i)] = \sum_j p_{ij}(\Delta t, t_0) \times E[U_d \mid X = j] \qquad \textbf{(7-4)}$$

$$Var[U_d(t) \mid (X(t_0) = i)] = \sum_j p_{ij}(\Delta t, t_0) \times (E[U_d \mid X = j]^2 + Var[U_d \mid X = j]) - E[U_d(t) \mid (X(t_0) = i)]^2$$

Figure 7-3 illustrates the empirical distribution of speed for each traffic state on Station 1 of the study area, $f_1(u \mid X = j)$. The empirical distributions have been constructed on the basis of historical observations of 5-minute station average speed on station 1. All 567 days of historical data were used in this analysis. The distribution of speed for States 1, 2, and 4 are single modal, and seem to vary within a shorter range compared to State 3. The speed of State 3 (transition state) exhibits a bimodal distribution that spans a wide range of speeds. Therefore, it seems that only using traffic state to characterize the distribution of speed might not be sufficient, and an unreasonably wide range of variation may be predicted for the distribution. Consequently, the inclusion of other conditions in $C(t)$ may result in a better explanation of the empirical distribution of speed.
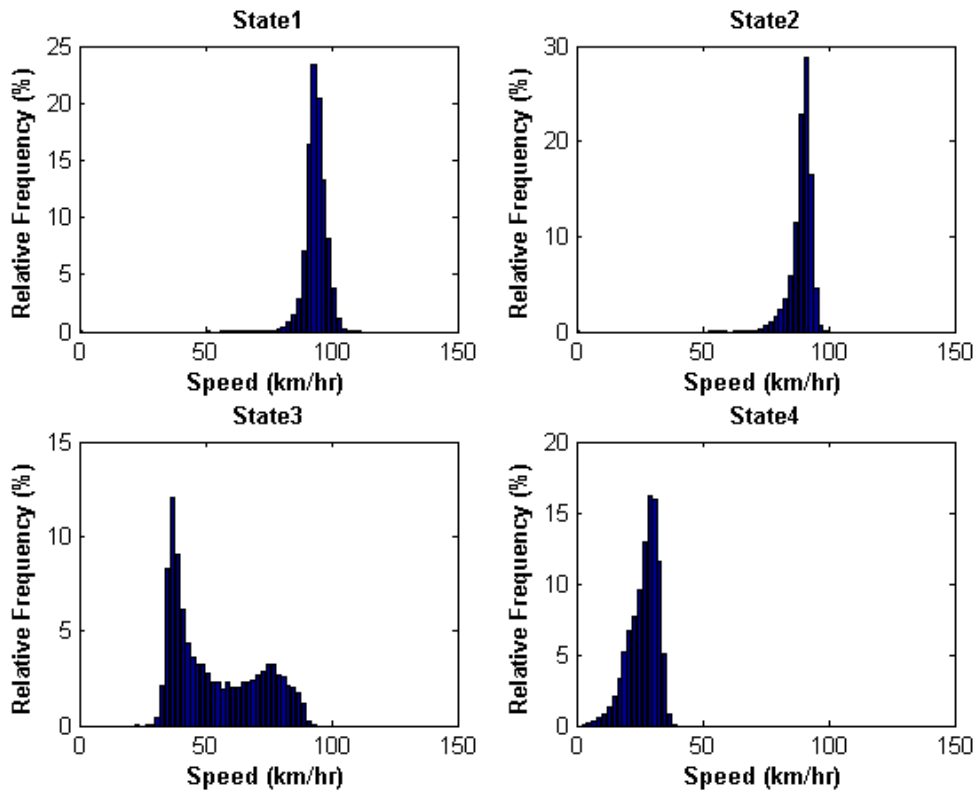
Figure 7-3:Empirical distribution of speed on Station 1 for different states

Consequently, we attempt to identify conditions which can improve the characterization of the distribution of speed. We examine the same conditions as were considered in the previous section (i.e. time of day (pm), traffic state of the downstream station (ds), and occupancy of the previous time interval (ocp)). We conduct a regression analysis to examine the effect of these three covariates on the mean of the distribution of speed for different states. The response variable in this analysis is mean traffic speed, and the independent factors are all categorical. Factors *pm* and *ds* are binary variables, and *ocp* is divided into four different categories. The regression model investigates the main effects and two level interactions of the factors. Table 7-4 summarizes the results of the regression analysis for State 3 for all the stations along the study route. It is clear that all the main effects and most of the interactions are statistically significant suggesting that the distribution of speed varies based on these factors. Therefore, all the factors can be effectively included in $C(t)$ to construct the empirical distribution of speed.

Table 7-4: The effect of Markov model covariates on the mean of speed for State 3

Station	Variable	Description	Coeff.	Lower	Upper
Station 1	μ	Mean	**69.8**	68.2	71.4
	ds	1 if downstream congested, 0 otherwise	**-15.2**	-18.3	-12.1
	pm	1 if after 12:00 pm, 0 otherwise	**-10.4**	-12.6	-8.2
	ocp_2	1 if 10<ocp<=20, 0 otherwise	-1.0	-2.9	0.8
	ocp_3	1 if 20<ocp<=30, 0 otherwise	**-16.3**	-18.7	-13.9
	ocp_4	1 if ocp>30, 0 otherwise	**-17.6**	-20.3	-15.0
	ds*pm	Interaction	**11.2**	8.7	13.6
	ds*ocp_2	Interaction	-0.4	-3.2	2.5
	ds*ocp_3	Interaction	**5.9**	2.7	9.1
	ds*ocp_4	Interaction	**5.9**	3.1	8.8
	pm*ocp_2	Interaction	**-4.3**	-7.1	-1.5
	pm*ocp_3	Interaction	**3.6**	0.3	7.0
	pm*ocp_4	Interaction	2.4	-0.9	5.7
Station 2	μ	Mean	**56.8**	55.8	57.8
	ds	1 if downstream congested, 0 otherwise	**-13.8**	-15.7	-11.9
	pm	1 if after 12:00 pm, 0 otherwise	**-4.3**	-5.6	-3.1
	ocp_2	1 if 10<ocp<=20, 0 otherwise	**-8.8**	-10.4	-7.3
	ocp_3	1 if 20<ocp<=30, 0 otherwise	**-14.4**	-16.2	-12.5
	ocp_4	1 if ocp>30, 0 otherwise	**-14.1**	-18.6	-9.6
	ds*pm	Interaction	**3.6**	2.0	5.2
	ds*ocp_2	Interaction	**6.8**	4.9	8.8
	ds*ocp_3	Interaction	**9.8**	8.1	11.6
	ds*ocp_4	Interaction	**7.7**	4.4	11.1
	pm*ocp_2	Interaction	**-1.8**	-3.7	0.0
	pm*ocp_3	Interaction	**-2.0**	-3.9	0.0
	pm*ocp_4	Interaction	-0.2	-4.5	4.1
Station 3	μ	Mean	**65.1**	64.4	65.8
	ds	1 if downstream congested, 0 otherwise	**-15.3**	-18.2	-12.5
	pm	1 if after 12:00 pm, 0 otherwise	**-8.7**	-9.7	-7.8
	ocp_2	1 if 10<ocp<=20, 0 otherwise	**-12.8**	-13.8	-11.8
	ocp_3	1 if 20<ocp<=30, 0 otherwise	**-19.1**	-20.2	-18.0
	ocp_4	1 if ocp>30, 0 otherwise	**-18.3**	-20.9	-15.7
	ds*pm	Interaction	**3.3**	2.1	4.5
	ds*ocp_2	Interaction	**7.0**	4.0	10.0
	ds*ocp_3	Interaction	**9.6**	6.7	12.4
	ds*ocp_4	Interaction	**8.1**	4.6	11.7
	pm*ocp_2	Interaction	**2.5**	1.1	3.8
	pm*ocp_3	Interaction	**3.5**	2.1	4.8
	pm*ocp_4	Interaction	**5.6**	3.1	8.1
Station 4	μ	Mean	**70.7**	69.9	71.6
	ds	1 if downstream congested, 0 otherwise	**-16.8**	-19.1	-14.4
	pm	1 if after 12:00 pm, 0 otherwise	**-3.8**	-4.8	-2.7
	ocp_2	1 if 10<ocp<=20, 0 otherwise	**-14.5**	-15.6	-13.3
	ocp_3	1 if 20<ocp<=30, 0 otherwise	**-25.6**	-26.8	-24.4
	ocp_4	1 if ocp>30, 0 otherwise	**-29.4**	-33.2	-25.6
	ds*pm	Interaction	1.0	-0.9	3.0
	ds*ocp_2	Interaction	**11.8**	9.2	14.4
	ds*ocp_3	Interaction	**17.0**	14.6	19.4
	ds*ocp_4	Interaction	**14.8**	9.9	19.7
	pm*ocp_2	Interaction	**1.6**	0.1	3.1
	pm*ocp_3	Interaction	**2.5**	1.1	4.0
	pm*ocp_4	Interaction	**5.9**	1.4	10.3

Bold = statistically significant at 95% level; Lower & Upper = 95% confidence limits

The distribution of speed for the near future can be predicted as follows:

$$g_d(U_d(t) \mid X(t_0) = i, C(t)) = \sum_j p_{ij}(\Delta t, t_o) \times f_d(u \mid X = j, C(t)) \qquad \textbf{(7-5)}$$

Where $f_d(u \mid X = j, C(t))$ is the empirical distribution of future speed while traffic is in state j and $C(t)$ explains the other conditions of traffic. $f_d(u \mid X = j, C(t))$ can be estimated based on the historical observations of traffic for each detector station. Only those observations which have been in state j and meet conditions $C(t)$ are considered to construct the empirical distribution of speed.

Using Equation 7-6, we are able to predict the mean and confidence interval for the traffic stream speed for the near-future. Figure 7-4 illustrates the constructed confidence interval for a typical day. It is clear that the width of the confidence interval is variable; it is wider for transient and congested states and it is narrower for free flow conditions.

In order to evaluate the confidence interval based on the constructed empirical distribution, we applied the proposed method to the 30 validation days. For each day, the number of future observations which do not fall into the predicted confidence interval is computed. For a confidence interval with significance level of α, it is expected that $100 \times \alpha$ percent of the speed observations will fall outside the boundaries. If the actual number of speed observations that fall outside the predicted confidence limits is less than $100 \times \alpha$, then the confidence interval is too conservative, and if it is more than $100 \times \alpha$, then the interval does not represent the level of significance of α.

Table 7-5 presents the average percentage of future speed observations which fall outside of the constructed confidence intervals for a 15 minute prediction horizon (three time intervals). Two different levels of significance (0.05 and 0.10) are used to evaluate the intervals for each station. Table 7-5 shows that the actual proportion of future observations outside of each interval is in agreement with the confidence levels upon which the confidence intervals are constructed.

Table 7-5: Average percentage of future observations which fall outside of the constructed confidence interval

Station	Confidence Interval ($\alpha = 0.05$)		Confidence Interval ($\alpha = 0.1$)	
	Mean of PCI_i (%)	Standard Deviation of PCI_i (%)	Mean of PCI_i (%)	Standard Deviation of PCI_i (%)
1	4.3	1.6	9.6	3.4
2	7.6	3.2	9.4	3.4
3	6.7	2.6	9.1	3.0
4	6.6	2.4	10.0	2.9

$PCI_i =$ percentage of future observations falling outside the empirical confidence interval for day i

109
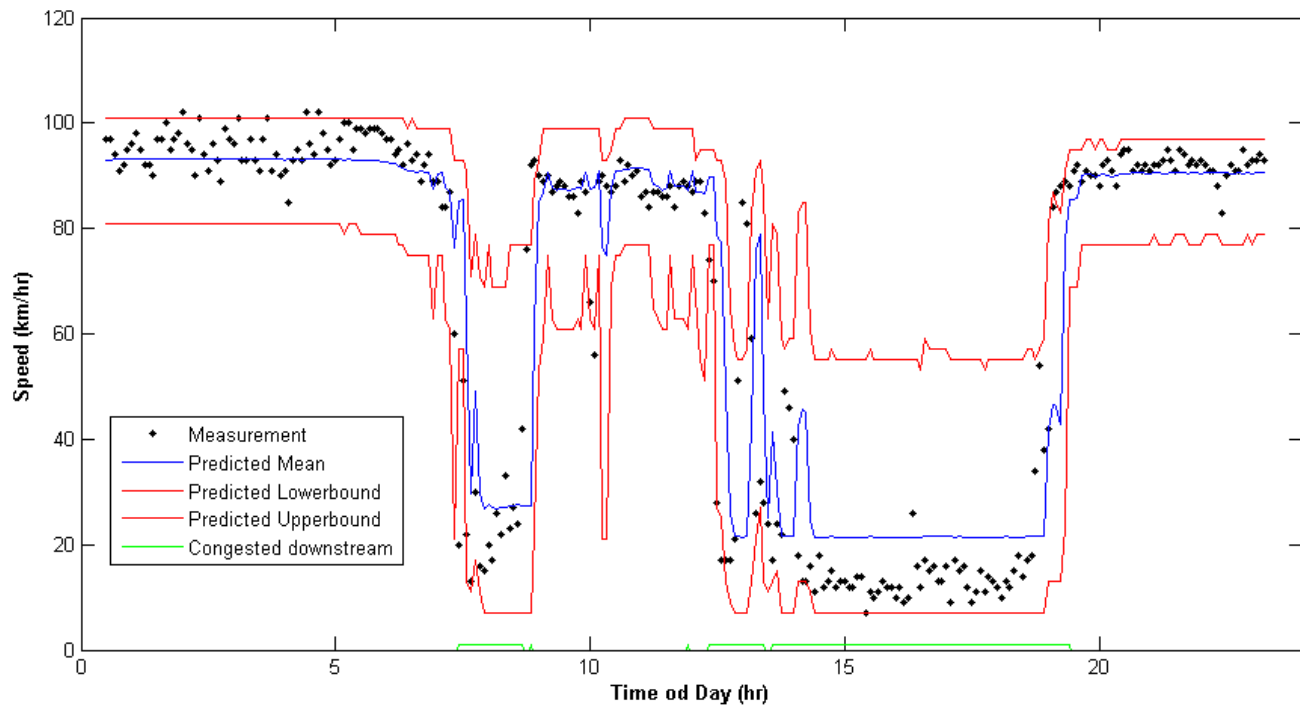
Figure 7-4: Prediction of the confidence interval of speed on Station 1 for a sample weekday

7.5 Conclusions

- A continuous-time Markov model has been developed to predict the state of the traffic for the near future. Four traffic states were considered to characterize the state of traffic: two free flow states, one transition state, and one congested state. Current traffic occupancy, state of traffic on the downstream segment, and time of day are found to have a statistically significant relationship with the near future traffic state.

- Using the proposed model, we are able to predict the probability of the traffic being in each of the possible states in the near future based on the current traffic conditions.

- The predicted probabilities then were utilized to characterize the expected distribution of traffic speed. Based on historical observations, the distribution of traffic speed was characterized for each traffic state separately. Given these empirical distributions and the predicted probabilities, distribution of traffic speed were predicted for the near future.

- The distribution of traffic speed then was used to predict a confidence interval for the near future. The confidence interval can be used to identify the expected range of future speeds at a given confidence level.

Chapter 8
Travel Time Prediction

The proposed models of this study have been presented in Chapters 4 through 7, and each model has been evaluated using a field dataset. The results demonstrated that the proposed models statistically improve the prediction of traffic speed; especially when traffic conditions are changing from uncongested to congested states.

In this chapter, the models are combined into the unified framework and the performance is evaluated on a longer route of a freeway.

8.1 Field Data – Extended Route

The evaluation described in this Chapter makes used of data from the Gardiner Expressway, the same urban freeway used to calibrate and validate the individual models in Chapters 4 through 7. However, in this chapter we apply the framework to a longer (7.44 km) section of the westbound direction of the Gardiner Expressway (see Chapter 3 for more details).

8.2 Prediction of Traffic Speed

Table 8-1 presents the prediction results for the 15 evaluation days. Traffic speed is predicted at each freeway station for a 15 minute period. Since Station 1 does not have any upstream information and Station 7 does not have any downstream information, those two stations are not included in this table. The prediction error is stated based on mean absolute percentage error (MAPE). Table 8-1 is divided into two different time periods; "all day" shows the average prediction error for the entire day, and "peak period" shows the average prediction error for the morning and evening peak period. The peak period in this figure is defined based on the typically conditions observed at the different stations during the 15 evaluation days. The Morning peak period is defined as 7:00 to 10:00 am and the Evening peak period is defined as 2:00 to 8:00 pm.

The prediction results of four different models are presented in Table 8-1. The Naïve model is a simple time-series prediction model which assumes that the traffic speed for the next time interval is equal to the latest available measurement from loop detectors. The Shockwave model in this chapter is equivalent to the Shockwave-SVM model from Chapter 6. The Shockwave model uses Naïve model as the time-series predictor; however, when the SVM model detects a congestion state at a station on the study route, the shockwave module is activated. Kalman-Dynamic model in this chapter is equivalent to the Kalman model with dynamic traffic pattern proposed in Chapter 4. Combined

model uses the Kalman-Dynamic as the time-series predictor and utilizes the shockwave module when a congestion state is detected.

The prediction results for Shockwave, Kalman-Dynamic, and Combined model show significant improvements for both entire day and peak period. The improvement of Combined model for the entire day ranges between 12.5 and 19.4 percent; the improvements for the peak period ranges between 10.2 and 20.5 percent. As shown in the previous chapters, both the Shockwave and Kalman-Dynamic models perform better than the Naïve model, and the results of this chapter show that the Combined model has lower prediction errors compared to both the Shockwave and Kalman-Dynamic models.

Table 8-1: Improvements in speed prediction accuracy for the proposed models of this study

Time Period	Station	Description	Naïve	Shockwave	Kalman-Dynamic	Combined Model
All day	2	Prediction Error (%)	12.3	11.5	10.5	9.9
	3		11.1	10.6	9.9	9.4
	4		10.1	9.8	9.6	9.0
	5		8.4	8.1	7.5	7.1
	6		9.1	8.2	8.3	8.0
	2	Improvement (%)	-	5.8	14.1	19.4
	3		-	4.8	10.7	15.8
	4		-	3.4	5.6	10.7
	5		-	3.2	10.0	14.8
	6		-	9.8	8.7	12.5
Peak Period	2	Prediction Error (%)	22.6	21.6	19.2	17.9
	3		19.6	19.0	16.7	15.8
	4		16.0	15.3	14.1	13.4
	5		14.3	14.0	13.2	12.5
	6		16.9	15.3	15.8	15.2
	2	Improvement (%)	-	4.3	14.8	20.5
	3		-	2.8	14.9	19.2
	4		-	4.4	12.2	16.8
	5		-	1.8	8.1	12.7
	6		-	9.5	6.6	10.2

Table 8-2 illustrates the prediction error improvement for each station; the distance to upstream and downstream stations are also provided in this table. Both the Shockwave model and the Kalman-Dynamic model use the traffic pattern at the station of interest along with the traffic pattern at upstream and downstream stations. Therefore, the length of the upstream and downstream sections may affect the performance of these prediction models. Station 2 has a medium range upstream

section followed by a short range downstream section; this is true for Stations 3, 4, and 5 (one medium and one short section). However, Station 6 has a medium range upstream section followed by a long range downstream section. The improvement in prediction accuracy for Station 6 is less than for the other stations and this may be the result of the longer distance between detectors.

Table 8-2: Improvements in speed prediction accuracy of Combined model compared to section length

Station	Distance To Station (m)		Access/Egress Points		Prediction
	Upstream	Downstream	Upstream	Downstream	Improvement (%)
2	1060	590	2	0	20.5
3	590	1070	0	1	19.2
4	1070	550	1	0	16.8
5	550	1120	0	1	12.7
6	1120	3050	1	1	10.2

Figure 8-1 illustrates the cumulative distribution function of the prediction error for Station 2. It is clear that the Combined model not only provides lower average traffic speed prediction errors, but also it performs better for every error percentile. The same pattern can be observed for Station 3 in Figure 8-2; the cumulative distribution function of the Combined model is higher than the other models for any given value of error.

Figure 8-3 illustrates the cumulative distribution function of the prediction error for Station 4. In this figure, for lower values of prediction error, the Kalman-Dynamic does not perform as well as the Naïve model or the Shockwave model, but it catches up slowly and for values of error greater than 11 percent it outperforms the other models. For example, the Naïve and the Shockwave models predicts the travel speed with less than 10% error about 50 percent of the time, while this is about 45% for the Combined and the Kalman-Dynamic models. However, the Naïve and the Shockwave models predict the travel speed with less than 30% error about 70 percent of the time, while this is about 85% for the Combined and the Kalman-Dynamic models. The same pattern can be observed for Stations 5 and 6.
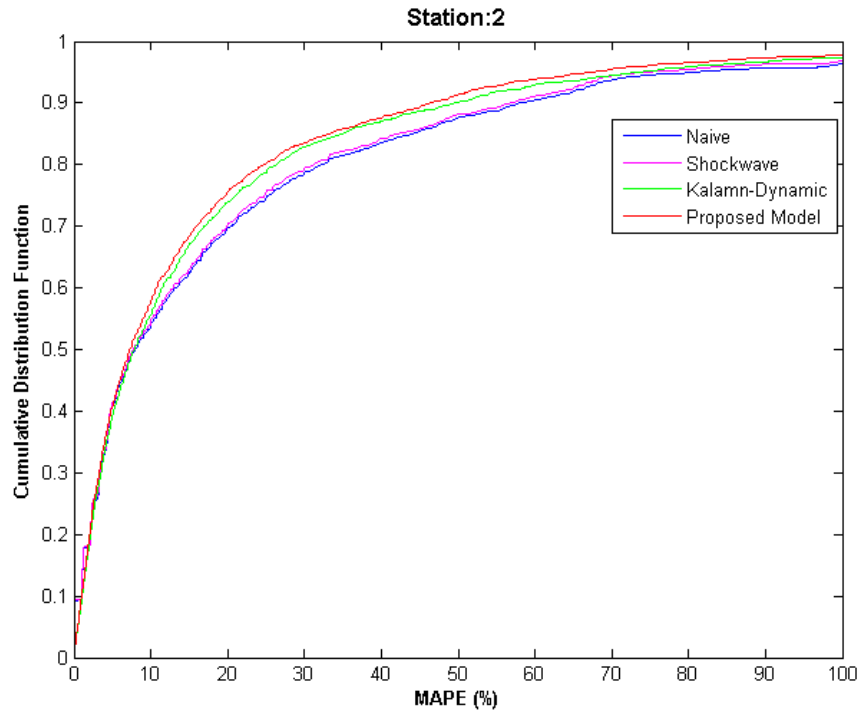
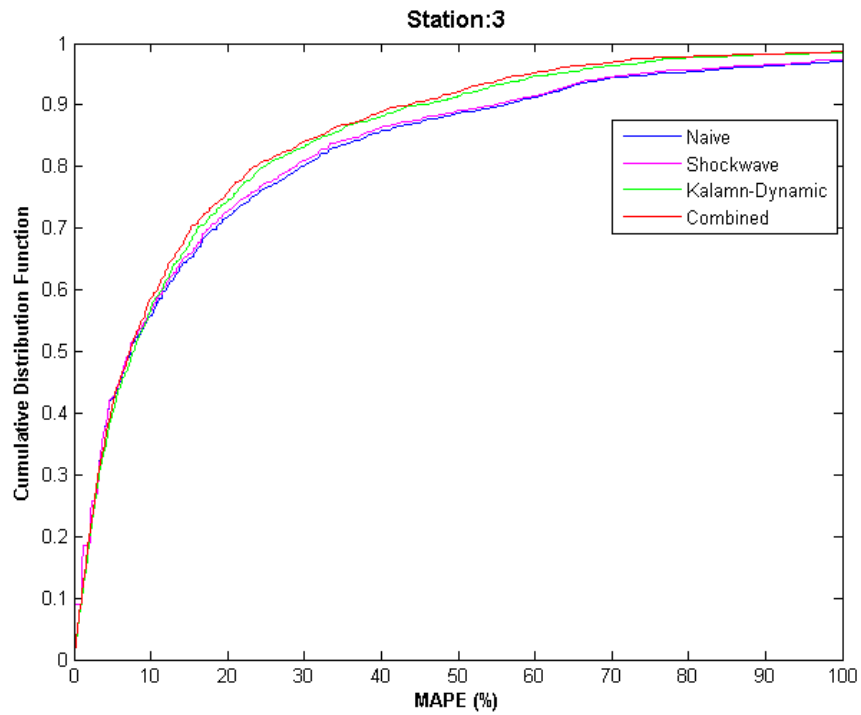**Figure 8-1: Cumulative distribution function of traffic speed prediction error for Station 2**

Figure 8-2: Cumulative distribution function of traffic speed prediction for Station 3

Figure 8-3: Cumulative distribution function of traffic speed prediction for Station 4
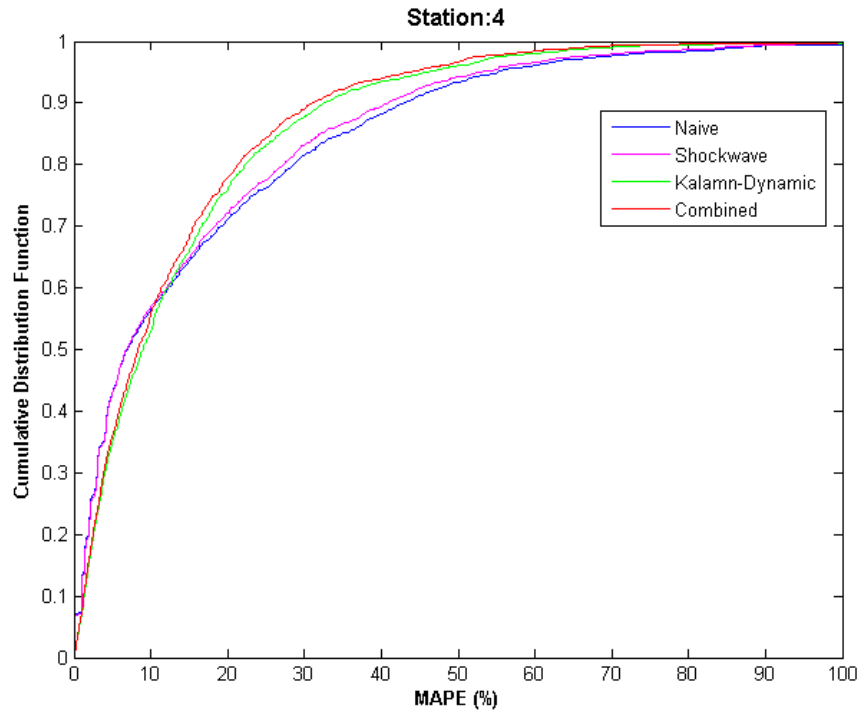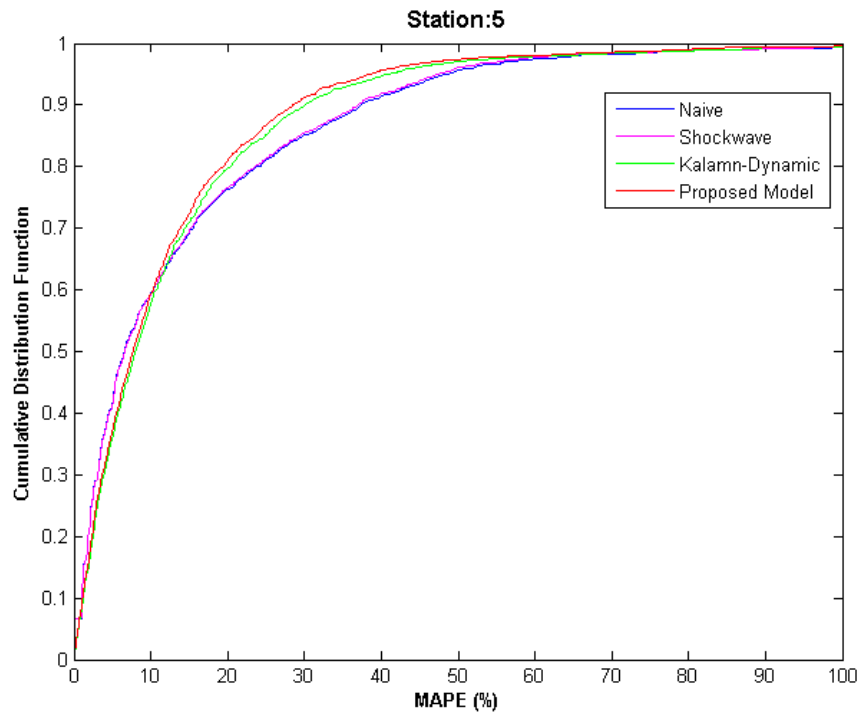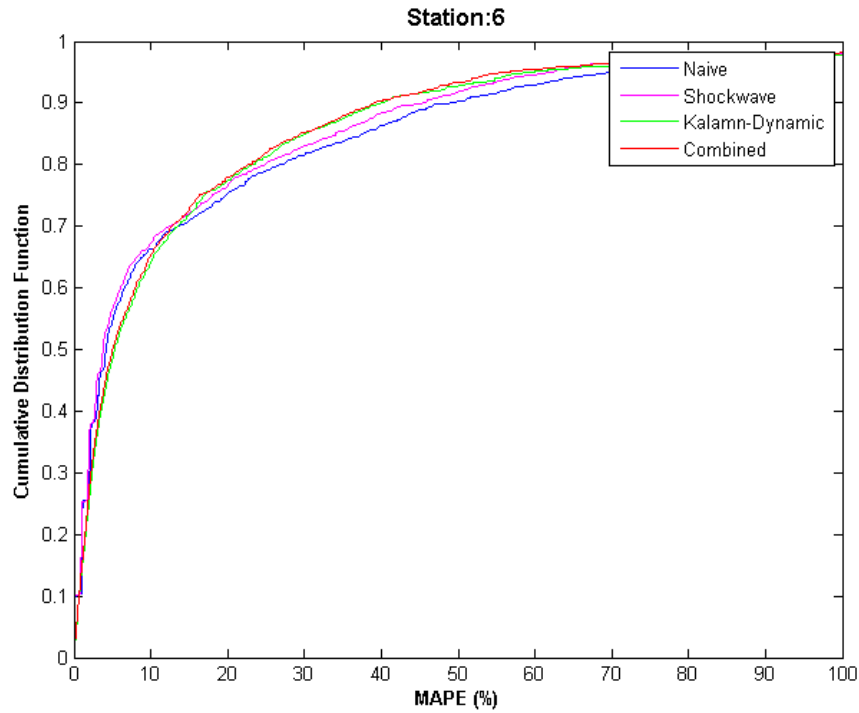

Figure 8-4: Cumulative distribution function of traffic speed prediction for Station 5

Figure 8-5: Cumulative distribution function of traffic speed prediction for Station 6

8.3 Prediction of Travel Time

In this section, the prediction results for traffic speed for each station are used to predict the travel time along the route of study. Suppose that a vehicle or group of vehicles are about to enter a freeway route at time t. Using the speed profile of the detectors, we are able to estimate the travel time for vehicles entering the route at time t. Izadpanah (2010) proposed a method to estimate the travel time of a route of a freeway using the loop detector measurements. Based on the measurements of the spot speed, this method builds a trajectory for the average vehicle traversing along the route. Then the travel time of the average vehicle is calculated by projecting the built trajectory on the time axis.

Suppose that a route consists of n segments which are separated by loop detectors. To estimate the average travel time along this route for vehicles entering the route at time t, the trajectory method estimates the time that this group of vehicles will pass each of the downstream loop detectors. Then the average speed of traveling along each segment is computed as the average spot speed of the loop detectors located at the upstream and downstream boundaries of this segment for the estimated passing time. The trajectory method is defined as follows (Izadpanah 2010):

$$\tau_T^{1,2}(t) = \frac{L^{1,2}}{\left(S_1(t) + S_2(t)\right)/2}$$

$$\tau_T^{2,3}(t) = \frac{L^{2,3}}{\left(S_2(t + \tau_T^{1,2}) + S_3(t + \tau_T^{1,2})\right)/2}$$

$$\vdots$$

$$\tau_T^{n-1,n}(t) = \frac{L^{n-1,n}}{\left(S_{n-1}(t + \sum_{i=1}^{n-2}\tau_T^{i,i+1}(t)) + S_n(t + \sum_{i=1}^{n-2}\tau_T^{i,i+1}(t))\right)/2}$$

$$\tau_T(t) = \sum_{i=1}^{n-1}\tau_T^{i,i+1}(t)$$

(8-1)

Where,

$\tau_T(t) =$ average travel time along the route associated with time interval t,

$\tau_T^{i,i+1}(t) =$ average travel time between loop detector i and $i+1$ associated with time interval t.

$S_n(t) =$ predicted average traffic speed at loop detector n for time interval t.

$L^{i,i+1} =$ length of the freeway segment between loop detector i and $i+1$

The travel time of the study route for vehicles entering the route at time t can be estimated using the trajectory method. Moreover, we are able to predict the travel time of the route using the values of traffic speed predicted for the future.

In this section, the speed predictions obtained from the Combined model are used to predict travel time for the study route. No independent measure of the true route travel time is available. Consequently, the true route travel times are estimated using the trajectory method described previously; however in this case the actual observed detector speeds are used. Obviously, it is not possible to do this in practice because this required future speeds to be available at the present time. However, since we are using archived data, we can apply this method to estimate the true route travel time.

Figure 8-6 illustrates the predicted travel time compared to the estimated travel time along the study route for a sample day. These results show that travel times varies from a low of approximately 4.5 minutes (equating to an average travel speed of approximately 98 km/h) to a high of approximately 10.5 minutes (equating to an average travel speed of approximately 43 km/h). The predicted route travel times appear to track the actual route travel times very closely.
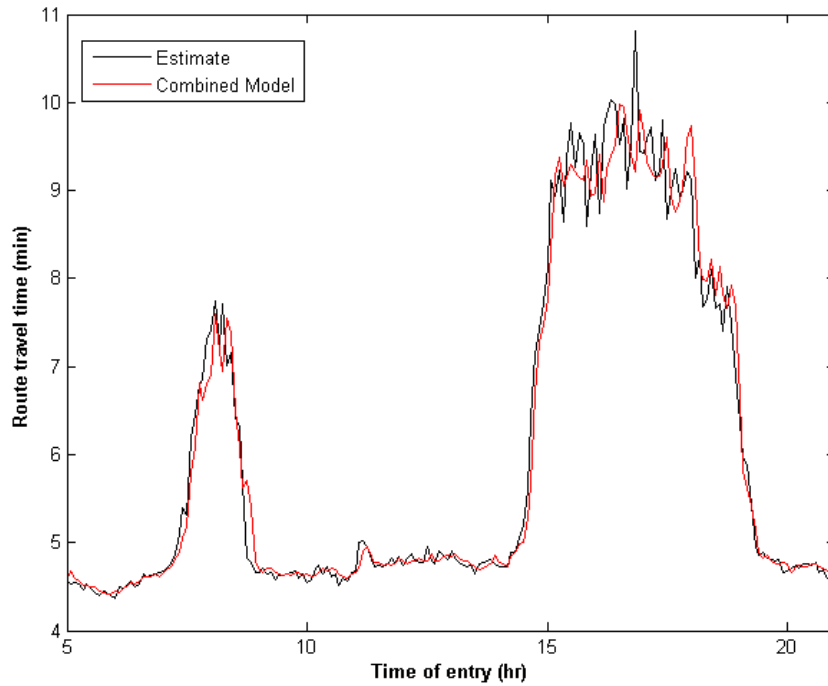
Figure 8-6: Estimation and prediction of travel time along the study route for a sample day

Table 8-3 presents the travel time prediction errors from the Naïve model and the proposed Combined model. The ground truth is assumed to be the estimated travel time using the trajectory method and the measured traffic speed. The proposed model provides improvements in travel time prediction accuracy of 13% when evaluating over the entire day and 23% when evaluating over just the peak periods. Figure 8-7 illustrates the cumulative distribution function of the travel time prediction errors for the Naïve and the Combined model. In addition to reducing the average error, the proposed Combined model performs better for every value of error percentile.

Table 8-3: Travel time prediction accuracy for the proposed model of this study

Time Period	Naïve	Combined Model	
	Prediction Error (%)	Prediction Error (%)	Relative Improvement (%)
All day	6.0	5.2	13.3
Peak period	9.0	6.9	23.3
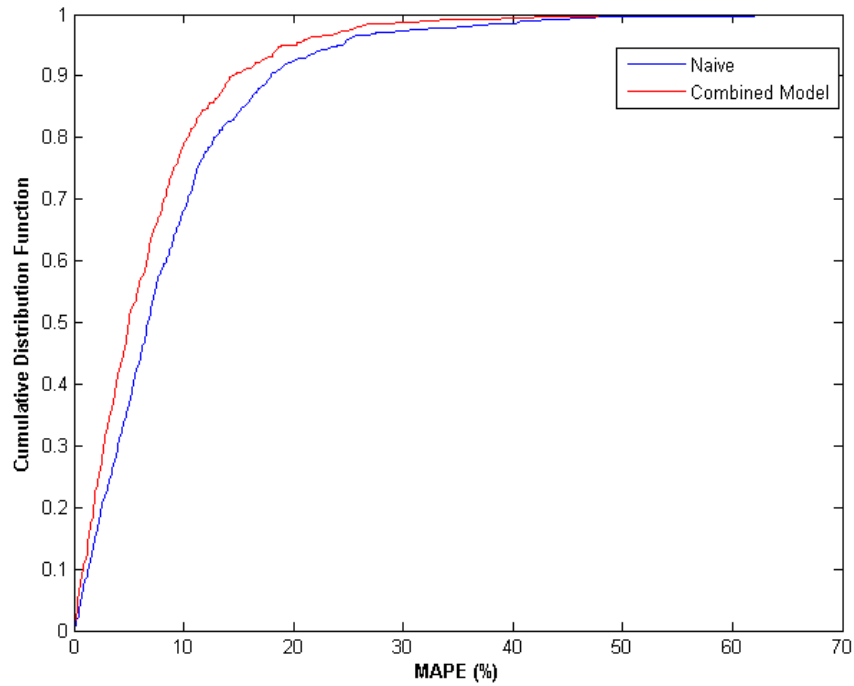
Figure 8-7: Cumulative distribution function of travel time prediction for the study route

8.4 Conclusions

This chapter documents a summary of the prediction results when the proposed modules of this thesis work together as follows:

- The proposed model of this chapter combines the models developed in Chapter 4, Chapter 5, and Chapter 6 and is called the "combined model". The performance of the proposed method was compared to that of the methods discussed in the previous chapters (Shockwave and Kalman-Dynamic). It was shown that the prediction errors for the Combined model are lower than each of those models separately.

- The proposed Combined model then was used to predict travel time along an extended freeway route (7.5 km). Travel time along the route was predicted by using the speed predictions at the location of each fixed point detector.

Chapter 9

Conclusions and Recommendations

This Chapter highlights the main conclusions and contributions of this thesis research and presents directions for future work.

9.1 Conclusions

A methodology was proposed to predict traffic conditions on freeways in real-time. The method can be summarized as follows:

1. Fixed point detector data (e.g. loop detectors) were used to acquire real time traffic data along freeways. The freeway route was divided into several segments each represented by a loop detector.

2. The methodology focuses on short-term prediction of traffic conditions. A fifteen minute prediction horizon was considered in this analysis.

3. The developed methodology consists of three main components. The first component which is called "Steady State Model" uses a Kalman filter model to predict general traffic conditions based on a data driven approach. The second component of the model adds another layer of analysis to the "Steady State Model" based on a model-based approach and is called "Congestion Spill back and Dissipation". This component implements the concept of traffic shockwave theory to inform the prediction model of changes in traffic conditions on the downstream and upstream segments. Conventional shockwave theory is deterministic; however, a methodology was developed to define the shockwave concept in a stochastic framework to tackle potential challenges raised by any uncertainty in data measurements and/or traffic conditions. A time-continuous Markov model was developed to define the stochastic shockwave model. The third component which is called "Congestion Detection" utilizes a classification approach to identify traffic state (Congestion vs Free flow) based on the available real-time traffic data. The predictions of this component are used in the "Congestion Spill back and Dissipation" module. A model based on Support Vector Machine was developed for the "Congestion Detection" component.

4. The prediction results of each individual component and the combined model wer compared to a naïve model. The results showed that the proposed "Steady State Model"

performs better than the naïve model. Moreover, adding the "Congestion Dissipation and Spill back" component to the "Steady State Model" improved the performance of the "Steady State Model" only, and finally, the combined model with all three components performs better than the combination of the first two components.

5. Prediction accuracy was most improved during the time periods when traffic conditions are changing from free flow to congestion and vice versa. These are the conditions when travel time predictions are of most value, but are also most difficult to predict.

6. The proposed combined model of this thesis was used to predict travel time along an extended freeway route (7.5 km). Travel time along the route was predicted by using the speed predictions at the location of each fixed point detector.

7. A stochastic framework was also developed to predict the empirical distribution of traffic speed. Four traffic states were considered to characterize the traffic conditions: two free flow states, one transition state, and one congested state. Using the proposed model, we were able to predict the probability of the traffic being in each of the possible states in the near future based on the current traffic conditions. As a result the empirical distribution of traffic speed was estimated for the traffic conditions in the near future. The distribution of traffic speed then was used to predict a confidence interval for traffic speed in the near future.

9.2 Contributions

The major contributions of this thesis can be summarized as follows:

1. Development of a Real-time Travel Time Prediction on Freeways: This research introduced a travel time prediction model for freeway routes with three components as described in Section 9.1. The model provides a prediction for travel time using traffic measurements from fixed point detectors (e.g. loop detectors). One of the characteristics of this model is that only traffic measurements on the freeway mainline are required – data from on and off ramps are not need. This is an attractive feature for practical applications because data from on and off ramps are less often available than data from the mainline. The model is capable of predicting traffic conditions for any prediction horizon in the context of short-term travel time prediction. The model was particularly aimed to predict travel time/traffic speed when traffic conditions are changing from free flow to congestion and vice versa. The proposed model was shown to significantly improve prediction accuracy when compared to a benchmark method. The improvement was shown to be the most during the

change period. The improvement in prediction of travel time was shown to be more than 20%.

2. Development of a Stochastic Shockwave Framework: A model has been developed to predict the speed of backward forming and forward recovery shockwaves. Unlike classic shockwave theory which is deterministic, the proposed model expresses the spillback and recovery as a stochastic process. This thesis introduced and formulated the concept of a stochastic shockwave approach to predict future traffic states. The transition probability matrix is defined as a function of traffic occupancy on upstream and downstream stations. Then, the probability of spillback and recovery can be computed given the traffic conditions. The proposed stochastic framework may better handle potential challenges raised by measurement error or variability in traffic conditions compared to the existing deterministic approach. It was shown in Chapter 5 that the accuracy of the prediction of future traffic state has been improved when the deterministic shockwave formulation was replaced by the stochastic shockwave approach developed in this thesis. The accuracy of traffic state prediction was improved by more than 20% when compared to the deterministic shockwave approach.

3. Development of a Real-time Method to Predict Empirical Distribution of Traffic Speed: This research introduced a traffic state prediction model based on stochastic process. The method provides a transition probability matrix to characterize the likelihood of the traffic states in the near future. The speed of traffic in the near future is considered as a random variable with an empirical distribution which was predicted based on the results of the developed model and the historical observed distribution of speed for each traffic state. The findings of this method can be used to develop a confidence interval for the average speed in the prediction horizon.

9.3 Future Work

1. This method was applied on a freeway route with a length of approximately 7.5 kilometers. It is recommended that the methodology used in this research be applied to predict travel times along a longer freeway route. The distance between two adjacent stations may affect the parameter estimates of the model. The proposed model implicitly considers the length of the segment via the parameter estimates; however, a model which explicitly considers the distance between detectors could be formulated, and the prediction results and transferability of the model should be investigated. The proposed method should also be

tested on a different section of freeway to confirm the transferability of the method to freeways with different geometry and traffic conditions.

2. There may be other factors which affect the transition probability between traffic states. Therefore, a further investigation to examine the effect of more covariates such as additional time periods and past traffic conditions could lead to more specific estimates for different traffic conditions.

3. Weather conditions may affect the stochastic process of the traffic in two different ways: a) it may affect the fundamental traffic diagram which leads to a different definition of traffic states; b) it may affect the probability of transition among states. Therefore, further investigations on the effect of different weather conditions may lead to better model estimates.

4. This method uses real-time traffic data provided by fixed point detectors. It is recommended that the methodology used in this research be extended to include additional data sources from floating vehicles (e.g. connected vehicles) to enhance the prediction accuracy especially during the change periods. Connected vehicle data may be used to inform traffic prediction models of traffic conditions on the downstream/upstream of the segment for which prediction is conducted.

References

Al-Deek, HM, M. P. D'Angelo, and MC Wang. 1998. Travel time prediction with non-linear time series. Paper presented at Fifth International Conference on Applications of Advanced Technologies in Transportation Engineering.

Ben-Akiva, M., M. Bierlaire, D. Burton, H. N. Koutsopoulos, and R. Mishalani. 2001. Network state estimation and prediction for real-time traffic management. *Networks and Spatial Economics* 1 (3): 293-318.

Bickel, P., C. Chen, J. Kwon, J. Rice, P. Varaiya, and E. van Zwet. 2003. Traffic flow on a freeway network. *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-*: 63-82.

Billings, D., and J. S. Yang. 2006. Application of the ARIMA models to urban roadway travel time prediction-a case study. Paper presented at IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC'06.

Bogers, E. A. I., H. W. C. Van Lint, and H. J. Van Zuylen. 2008. Reliability of travel time: Effective measures from a behavioral point of view. *Transportation Research Record: Journal of the Transportation Research Board* 2082 : 27-34.

Chrobok, R., A. Pottmeier, S. F. Hafstein, and M. Schreckenberg. 2004. Traffic forecast in large scale freeway networks. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering* 14 : 1995-2004.

Chu, L., S. Oh, and W. Recker. 2005. Adaptive kalman filter based freeway travel time estimation. Paper presented at 84th TRB Annual Meeting, Washington DC.

Daganzo, C. F. 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological 28 (4): 269-87.

Daganzo, C. F. 1995. The cell transmission model, part II: Network traffic. *Transportation Research Part B: Methodological* 29 (2): 79-93.

D'Angelo, M. P., H. M. Al-Deek, and M. C. Wang. 1999. Travel-time prediction for freeway corridors. *Transportation Research Record: Journal of the Transportation Research Board* 1676 : 184-91.

Dong, J., and H. S. Mahmassani. 2009. Flow breakdown and travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board* 2124 : 203-12.

Elefteriadou, Lily, RP Roess, and WR McShane. 1995. Probabilistic nature of breakdown at freeway merge junctions. Transportation Research Record(1484): 80-9.

European Commission. 2009. *TRAFFIC MANAGEMENT FOR LAND TRANSPORT: Research to increase the capacity,*

efficiency, sustainability and safety of road, rail and urban transport networks. Belgium: Directorate - General for Enrgy and Transport.

Evans, J. L., L. Elefteriadou, and N. Gautam. 2001. Probability of breakdown at freeway merges using markov chains. *Transportation Research Part B: Methodological* 35 (3): 237-54.

Gartner, N., C. J. Messer, and A. K. Rathi, eds. 2001. *Revised monograph of traffic flow theory: A state-of-the-art report*Transportation Research Board.

Greenshields, B. D. 1935. A study of traffic capacity. Paper presented at Highway Research Board Proceedings.

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

Hollander, Y., and R. Liu. 2008. Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C: Emerging Technologies* 16 (2): 212-31.

Izadpanah, P. 2010. Freeway travel time prediction using data from mobile probes. PhD., University of Waterloo.

Izadpanah, P., Hellinga, B., & Fu, L. 2011. Real-time freeway travel time prediction using vehicle trajectory data. Presented at Transportation Research Board 90th Annual Meeting (No. 11-2826).

Jackson, Christopher. 2011. Multi-state markov and hidden markov models in continuous time. Package'Msm', Version 1

Kaparias, I., M. G. H. Bell, and H. Belzner. 2008. A new measure of travel time reliability for in-vehicle navigation systems. *Journal of Intelligent Transportation Systems* 12 (4): 202-11.

Kharoufeh, J. P., and N. Gautam. 2004. Deriving link travel-time distributions via stochastic speed processes. *Transportation Science* 38 (1): 97-106.

Kijima, M. 1997. Markov processes for stochastic modeling (Vol. 6). CRC Press.

Kuhne, R., & Michalopoulos, P. 1997. Continuum flow models. Traffic flow theory: A state of the art report revised monograph on traffic flow theory.

Li, Ruimin, Geoffrey Rose, and Majid Sarvi. 2006. Using automatic vehicle identification data to gain insight into travel time variability and its causes. Transportation Research Record: Journal of the Transportation Research Board 1945 (1): 24-32.

Lighthill, M. J., & Whitham, G. B. 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 229(1178), 317-345.

Liu, H. X., Wu, X., Ma, W., & Hu, H. (2009). Real-time queue length estimation for congested signalized intersections. Transportation research part C: emerging technologies, 17(4), 412-427.

Lo, H. K. 2001. A cell-based traffic control formulation: Strategies and benefits of dynamic timing plans. *Transportation Science* 35 (2): 148-64.

Madsen, Henrik. Time series analysis. CRC Press, 2007.

Muñoz, L., X. Sun, D. Sun, G. Gomes, and R. Horowitz. 2004. Methodological calibration of the cell transmission model. Paper presented at Proceedings of the American Control Conference vol. 1.

Muñoz, L., X. Sun, R. Horowitz, and L. Alvarez. 2003. Traffic density estimation with the cell transmission model. Paper presented at American Control Conference, 2003. Proceedings of the 2003.

Nam, D. H., and D. R. Drew. 1999. Automatic measurement of traffic variables for intelligent transportation systems applications. *Transportation Research Part B: Methodological* 33 (6): 437-57.

Nanthawichit, C., T. Nakatsuji, 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: Journal of the Transportation Research Board* 1855 : 49-59.

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

Noroozi, R. and B. Hellinga, "Real-time Prediction of Near-Future Traffic States on Freeways Using a Markov Model", accepted for publication in the Transportation Research Record (TRR), Journal of the Transportation Research Board, 2014.

Oh, J. S., R. Jayakrishnan, and W. Recker. 2003. *Section travel time estimation from point detection data.* Center for Traffic Simulation Studies.

Park, D., and L. R. Rilett. 1999. Forecasting freeway link travel times with a multilayer feedforward neural network. *Computer-Aided Civil and Infrastructure Engineering* 14 (5): 357-67.

Persaud, Bhagwant, Sam Yagar, and Russel Brownlee. 1998. Exploration of the breakdown phenomenon in freeway traffic. Transportation Research Record: Journal of the Transportation Research Board 1634 (1): 64-9.

Persaud, Bhagwant, Sam Yagar, David Tsui, and Horace Look. 2001. Breakdown-related capacity for freeway with ramp metering. Transportation Research Record: Journal of the Transportation Research Board 1748 (1): 110-5.

Rakha, H., I. El-Shawarby, M. Arafeh, and F. Dion. 2006. Estimating path travel-time reliability. Paper presented at Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE.

Robinson, S., and J. W. Polak. 2005. Modeling urban link travel time with inductive loop detector data by using the k-NN method. *Transportation Research Record: Journal of the Transportation Research Board* 1935 : 47-56.

SHRP2. 2009. *Incorporating reliability performance measures in operations and planning modeling tools.* The Strategic Highway Research Program 2: Transportation Research Board Of The National Academies.

Steven, I., J. Chien, and C. M. Kuchipudi. 2003. Dynamic travel time prediction with real-time and historic data. *Journal of Transportation Engineering* 129 : 608.

Suzuki, H., T. Nakatsuji, Y. Tanaboriboon, and K. Takahashi. 2000. Dynamic estimation of origin-destination travel time and flow on a long freeway corridor: Neural kalman filter. *Transportation Research Record: Journal of the Transportation Research Board* 1739 (-1): 67-75.

Tampère, C. M. J., and LH Immers. 2007. An extended kalman filter application for traffic state estimation using CTM with implicit mode switching and dynamic parameters. Paper presented at Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE.

Tu, H. 2008. *Monitoring travel time reliability on freeways*Netherlands TRAIL Research School.

Ul, S., I. Bajwa, and M. Kuwahara. 2003. A travel time prediction method based on pattern matching technique. *Publication of: ARRB Transport Research, Limited*.

Van Aerde, M. 1995. Single regime speed-flow-density relationship for congested and uncongested highways. Paper presented at 74th Annual Meeting of the Transportation Research Board, Washington, DC.

Van Aerde, Michel, and Hesham Rakha. 1995. Multivariate calibration of single regime speed-flow-density relationships. Presented at 6th Vehicle Navigation and Information Systems Conference.

van Hinsbergen, C., H. van Lint, and H. J. van Zuylen. 2008. Neural network committee to predict travel times: Comparison of bayesian evidence approach to the use of a validation set. Paper presented at 11th International IEEE Conference on Intelligent Transportation Systems, 2008. ITSC 2008.

van Hinsbergen, CP, JW VAN LINT, and FM Sanders. 2007. Short term traffic prediction models. Paper presented at PROCEEDINGS OF THE 14TH WORLD CONGRESS ON INTELLIGENT TRANSPORT SYSTEMS (ITS), HELD BEIJING, OCTOBER 2007.

van Hinsbergen, CP, JW van Lint, and FM Sanders. 2007. Short term traffic prediction models. Presented at The 14th World Congress on Intelligent Transport Systems (ITS), held Beijing.

van Hinsbergen, CP, JWC Van Lint, and HJ Van Zuylen. 2009. Bayesian committee of neural networks to predict travel times with confidence intervals. *Transportation Research Part C: Emerging Technologies* 17 (5): 498-509.

van Lint, J. W. C., and H. J. van Zuylen. 2005. Monitoring and predicting freeway travel time reliability: Using width and skew of day-to-day travel time distribution. *Transportation Research Record: Journal of the Transportation Research Board* 1917 : 54-62.

Van Lint, JWC, SP Hoogendoorn, 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* 13 (5-6): 347-69.

Wahle, J., and M. Schreckenberg. 2001. A multi-agent system for on-line simulations based on real-world traffic data. Paper presented at Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001.

Wang, Y., and M. Papageorgiou. 2005. Real-time freeway traffic state estimation based on extended kalman filter: A general approach. *Transportation Research Part B: Methodological* 39 (2): 141-67.

Yeon, J., L. Elefteriadou, and S. Lawphongpanich. 2008. Travel time estimation on a freeway using discrete time markov chains. *Transportation Research Part B: Methodological* 42 (4): 325-38.

You, J., and T. J. Kim. 2000. Development and evaluation of a hybrid travel time forecasting model. *Transportation Research Part C: Emerging Technologies* 8 (1-6): 231-56.

Zheng, F. 2011. *Modelling urban travel times.*Netherlands TRAIL Research School.