ABSTRACT

DEVELOPMENT OF DYNAMIC RECURSIVE MODELS FOR FREEWAY TRAVEL TIME PREDICTION

by
Xiaobo Liu

Traffic congestion has been a major problem in metropolitan areas, which is caused by either insufficient roadway capacity or unforeseeable incidents. In order to promote the efficiency of the existing roadway networks and mitigate the impact of traffic congestion, the development of a sound prediction model for travel times is desirable.

A comprehensive literature review about existing prediction models was conducted by investigating the advantages, disadvantages, and limitations of each model. Based on the features and properties of previous models, the base models including exponential smoothing model (ESM), moving average model (MAM), and Kalman filtering model (KFM) are developed to capture stochastic properties of traffic behavior for travel time prediction.

By incorporating KFM into ESM and MAM, three dynamic recursive prediction models including dynamic exponential smoothing model (DESM), improved dynamic exponential smoothing model (IDESM), and dynamic moving average model (DMAM) are developed, in which the time-varying weight parameters are optimized based on the most recent observation. Model evaluation has been conducted to analyze prediction accuracy under various traffic conditions (e.g., free-flow condition, recurrent and non-recurrent congested traffic conditions). Results show that the IDESM in general outperforms other models developed in this study in prediction accuracy and stability. In addition, the feature

and logic of the IDESM lead to its high transferability and adaptability, which could enable the prediction model to perform well at multiple locations and deal with complicated traffic conditions.

Besides the proficient capability, the IDESM is easy to implement in the real world transportation network. Thus, the IDESM is proven an appealing approach for short-time travel time prediction under various traffic conditions. The application scope of the IDESM is identified, while the optimal prediction intervals are also suggested in this study.

DEVELOPMENT OF DYNAMIC RECURSIVE MODELS FOR FREEWAY TRAVEL TIME PREDICTION

by

Xiaobo Liu

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation

Interdisciplinary Program in Transportation

May 2004

APPROVAL PAGE

DEVELOPMENT OF DYNAMIC RECURSIVE MODELS FOR FREEWAY TRAVEL TIME PREDICTION

Xiaobo Liu

Dr. Steven I-Jy Chien, Dissertation Advisor — Date
Associate Professor of Civil and Environmental Engineering, NJIT

Dr. Athanassios K. Bladikas, Committee Member — Date
Associate Professor of Industrial and Manufacturing Engineering, NJIT

Dr. Lazar Spasovic, Committee Member — Date
Professor of Civil and Environmental Engineering, NJIT

Dr. Janice R. Daniel, Committee Member — Date
Assistant Professor of Civil and Environmental Engineering, NJIT

Dr. Jian Yang, Committee Member — Date
Assistant Professor of Industrial and Manufacturing Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Xiaobo Liu

Degree: Doctor of Philosophy

Date: May 2004

Undergraduate and Graduate Education:

- Doctor of Philosophy in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 2004
- Master of Science in Transportation Management,
  Southwest Jiaotong University, Chengdu, P. R. China, 1999
- Bachelor of Science in Railway Transportation,
  Southwest Jiaotong University, Chengdu, P. R. China, 1996

Major: Transportation

Presentations and Publications:

Xiaobo Liu, Steven Chien and Kaan Ozbay,
"Predicting Travel Times for the South Jersey Real-Time Motorist Information System,"
Transportation Research Record, No. 1855, 2003.

Xiaobo Liu, Mei Chen and Steven Chien,
"A Dynamic Bus Arrival Time Prediction Model Based on APC Data,"
Journal of Computer- Aided Civil and Infrastructure Engineering, (forthcoming).

Xiaobo Liu, Mei Chen, Steven Chien, and Jeremy Brickey,
"Application of AVL/APC Data Support System,"
82nd Annual Meeting of Transportation Research Board, Washington, DC, 2003.

Xiaobo Liu, Mei Chen, Jason Yaw, Steven Chien,
"Using AVL and APC Data in Bus Arrival Time Prediction,"
82nd Annual Meeting of Transportation Research Board, Washington, DC, 2003.

Xiaobo Liu and Steven Chien,
"The Development of Dynamic Travel Time Prediction Models For South Jersey Real-Time Motorist Information,"
Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, Florida, 2002.

To my beloved wife and parents

ACKNOWLEDGEMENT

I am greatly indebted to my dissertation supervisor, Dr. Steven Chien, who constantly supported, inspired and encouraged me with his insight, dedication and understanding during my Ph.D. study. Special thanks are given to Dr. Athanassios Bladikas, Dr. Lazar Spasovic, Dr. Janice Daniel, and Dr. Jian Yang for serving as committee members and providing me with precious suggestions on research and industrial practice.

My fellow graduate students, Dr. Jiangtao Luo, Mr. Yimin Tang, Mr. Rajat Rajbhandari, Mr. Jiahua Song, Mr. Zhaowei Yang, Mr. Yongqiang Yang, and Ms. Fei Yang, are deserving of recognition for their support.

I also appreciate support from the New Jersey Department of Transportation, National Center for Transportation and Industrial Productivity, and New Jersey Institute of Technology associated with the research project "South Jersey Real-Time Motorist Information System."

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF SYMBOLS

Symbol **Page**

LIST OF SYMBOLS
(Continued)

Symbol **Page**

CHAPTER 1

INTRODUCTION

1.1 Background

Traffic congestion has been an escalating problem in many metropolitan areas for decades. A study (Spasovic, et al., 2001) conducted by the National Center for Transportation and Industrial Productivity (NCTIP) stated that the average annual cost of congestion for New Jersey was $880 per licensed driver in 2000. According to an urban mobility study conducted by Texas Transportation Institute (TTI, 2002), every driver wasted 51 hours sitting in traffic in 2001. The total cost based on excess travel time and fuel consumption was up to $68 billion for 75 major areas in the United States. General Motors (1991) estimated that highway congestion caused $96 billion in productivity losses each year and $70 billion in traffic accident costs. It was estimated that Americans spent over 1.6 billion hours in traffic jams each year. By the year 2005, the time lost to congestion would rise to 8.1 billion hours (Zhu, 2000).

Traffic congestion is caused by either insufficient roadway capacity or unforeseeable incidents, such as accidents. Transportation researchers have been investigating the relationship between traffic demand and roadway capacity. For example, a study conducted by the Center for Transportation Research Education (CTRE) predicted that the US population would grow 20 percent to 360 million by the year 2025, while the annual volume of weight will be doubled from 8 to 16.8 billion ton-miles. In addition, sixteen percent of the bridge inventory will be functionally obsolete or structurally deficient. Therefore, the financial, environmental, safety, and political

considerations made the policy-makers realize that expanding the size of the transportation system is almost impossible (Sussman, 1996).

With the enactment of the Surface Transportation Assistance Act of 1987 and the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, highway infrastructure began to utilize information technology to alleviate the impact of traffic congestion. The Federal Highway Administration (FHWA) placed a high priority on collaborative partnerships to mitigate the congestion impact by expanding congestion management systems and developing a regional Intelligent Transportation Systems (ITS) architecture. The application of information technology (IT) may be a viable way to improve the already congested transportation system and its safety through acquiring, storing and distributing accurate and timely traffic information. Golob and Regan (2001) explored the effects of IT applications in transportation and found that IT could greatly advance traveler information systems.

Advanced Traveler Information System (ATIS) is one of the most important components in ITS. Stathopoulos (2003) stated that "One of the most critical aspects of ITS success is the provision of accurate real-time information and short-term predictions of traffic parameters such as traffic volumes, travel speeds and occupancies." Based on predicted information, efficient traffic control could be applied to the transportation infrastructure, and then the negative impact of traffic congestion can be mitigated.

ATIS has been classified into three categories (Chestlow and Htcher, 1993): data collection, data consolidation, and data communication. Firstly, the traffic information should be collected by all technical means (e.g., loop detectors, probe vehicles and cameras). Secondly, the collected data should be consolidated as suitable and useful

information that can fit technical and practical requirements. Finally, the consolidated data should be transmitted to traffic control centers and/or road users.

In recent years, technologies have been developed to advance ITS applications and ease congestion, such as distributing real-time traffic information to travelers. Wahle et al. (2001) developed a dynamic route guidance system based on real time traffic data. Pattanamekar et al. (2003) employed predicted travel time in a dynamic shortest path problem. Accurately predicted travel times could bring huge benefits for most ITS applications, such as dynamic route-guidance systems (Fu and Rilett, 1998), congestion management and incident detection systems (Hellinga and Knapp, 1999), and in-vehicle information systems (Hellinga and Gudapati, 2000).

Since the availability of future traffic information is very important to the application of ITS, many prediction models have been used for that purpose, such as the Box-Jenkins models (Ahmed and Cook, 1979; Nihan and Holmesland, 1980), regression models (Frechette and Khan, 1998), nonparametric models (Smith and Demetsky, 1997), Kalman Filtering models (Okutani and Stephanedes, 1984), and artificial neural networks (McFadden and Durrans, 2001; Chien, et al., 2002). Though much progress has been made on traffic information prediction, the short-term forecasting of traffic conditions has had an active but somewhat unsatisfying research history.

1.2 Problem Statement

It is challenging to develop a reliable travel time prediction model because of two main reasons: (1) stochastic traffic conditions and (2) limited historical and real-time traffic information. Various factors affecting daily traffic operations were classified into internal

and external categories (Woodhull, 1987). Internal factors include time-varying traffic demand and vehicle composition (e.g., percentage of cars, buses and trucks), infrastructure (e.g., roadways, bridges, toll plaza), roadway geometric conditions (e.g., number of lanes, curvature, levels of terrains), and motorists' driving behavior. External factors include unexpected events (e.g., incidents), traffic control devices (e.g., signals, ramp meters), and weather conditions (e.g., snow, ice, fog and sun glare).

The stochastic factors make it difficult to formulate the relationship among traffic parameters (e.g., volume, speed, density) mathematically. Although the general relationship has been wildly explored in traffic flow theory, it might not be valid under saturated and very light traffic conditions (Hall and Persaud, 1989; Lin, 2002). Thus, the estimated traffic data may be biased, and inaccurate traffic information generated from the biased data will mislead motorists as well as worsen the already congested traffic condition. If such a condition persists, the deployed ITS would not be effective because of the lack of road users' confidence in it.

The complexity of traffic flow characteristics can also be observed in forms of various traffic conditions (e.g., free-flow and congested conditions), and prediction models may have different performance under different traffic conditions. For example, historic average models (Strobel, 1977) and the Box-Jenkins models (Ahmed and Cook, 1979) have generated satisfactory predictions of travel times under free-flow traffic conditions. However, those models cannot provide accurate results if the future traffic condition is different from that in the historic profile. Different prediction models have evolved to model various traffic conditions and locations (Kuchipudi and Chien, 2003).

Another challenge in developing prediction models comes from the difficulty to acquire a sufficient quantity and good quality of traffic data. Many models were developed based on historical and/or real time traffic data to predict future traffic information. For example, in Urban Transportation Control System (UTCS), Stephanedes el al. (1981) predicted traffic volume based on historical data. Later, Okutani and Stephanedes (1984) developed a Kalman Filter model to predict traffic volumes by employing real time traffic information. Ishak and Alecsandru (2003) employed both historical and real time traffic information to develop an Artificial Neural Network (ANN) based prediction model. The availability of traffic information was the major concern for the development of the prediction model.

Besides the availability of traffic data, the levels of detail of the data also affect the accuracy of the prediction model. Lan and Miou (1998) found that traffic flow distributions (e.g., normal, binomial distribution) under different aggregated levels (e.g., 5-minute average, 10-minute average) affected the prediction accuracy. Zietsman and Rilett (2000) investigated travel time estimation with different aggregated levels of data and found that modeling travel characteristics on a disaggregate level can improve the accuracy and result in better estimates for travel time. Kuchipudi and Chien (2003) analyzed the prediction interval impact on the accuracy of a prediction model by setting various interval lengths, and choosing as best intervals that minimized error and acquired a sufficient sample size. In addition, considering the stochastic traffic conditions and the availability of traffic data, different prediction models with different strengths are desired under different situations.

1.3 Objective and Scope of Work

The objective of this study is to develop dynamic short-term travel time prediction models by employing historical and real time data under various traffic conditions (e.g., free-flow and congested conditions).

To achieve the above objective, the advantage, disadvantage and limitation of previous and existing prediction models was thoroughly reviewed. Then, a number of prediction models were developed and tested under various traffic conditions. The properties and logic of those models were investigated, and the suitable situation for implementing each model was identified. The developed models were further enhanced to adapt to various traffic patterns (e.g., free-flow vs. congested conditions) to dynamically capture traffic changes and achieve the best prediction accuracy.

A microscopic simulation model (CORSIM) was used to simulate traffic operations under various conditions in a selected highway network in southern New Jersey. The simulated information was calibrated with data collected from acoustic sensors and used as actual traffic information for testing the developed prediction models. The performance of the developed models was evaluated based on various statistical measures (e.g., Mean Absolute Relative Error (MARE), Variance of Absolute Percentage Error (VAPE), Maximum Relative Error (MRE)) while considering the duration of the prediction interval.

1.4 Organization

This dissertation is organized into seven chapters. In this first chapter, the problem was identified and the objective and scope of the study were addressed. In Chapter 2, a thorough literature review summarizes findings from previous studies, including the applications of travel time information in ATIS, estimation of traffic information, and existing models developed for predicting travel times. A review of micro-, meso-, and macroscopic simulation models is also presented. Chapter 3 contains the development of the base prediction models. In Chapter 4, a study site is identified and a microscopic simulation model is developed with TSIS CORSIM, while the calibration and validation of the simulation model are addressed. Three traffic scenarios (e.g., free-flow condition, recurrent and non-recurrent congested traffic conditions) are designed and applied to test and evaluate the features and properties of the developed prediction models. In Chapter 5, three dynamic recursive prediction models are developed to dynamically optimize the weight parameters for improving prediction accuracy. In Chapter 6, the developed models are thoroughly evaluated in different periods under various traffic conditions. Results can be used to identify the application scope of each model and optimal prediction interval under complicated traffic conditions. Finally, conclusions and suggestions for future study are presented in Chapter 7.

CHAPTER 2

LITERATURE REVIEW

This chapter summarizes the review of previous models for short-term travel time prediction, while the importance of travel times for ATIS is introduced. In addition, the data applied for travel time prediction, including data collection and traffic parameters estimation, are reviewed. The reviewed prediction models are classified into time series models, regression models, and Artificial Neural Network (ANN) models and the limitations of each model type are discussed. Finally, simulation models (e.g., micro-, meso-, and macroscopic) are included, which can provide sufficient emulated traffic information for prediction model development and evaluation.

2.1 Traffic Information in ATIS

According to a study titled "Advanced Traveler Information Systems" conducted by the Intelligent Transportation Society of America (2002), congestion on the nation's streets and highways is getting worst. Forty-two states predicted they would be falling behind in their efforts to alleviate urban congestion on Interstates over the next decade. Advanced Traveler Information Systems (ATIS) can provide a good opportunity to understand traffic and transit conditions, present multi-modal options to travelers, and improve the operation of transportation networks. In other words, the ATIS is about providing citizens, businesses and commercial carriers with the right information at the right time to improve the quality and convenience of their trips and the overall performance of the transportation system. ATIS has played an important role with other ITS applications to improve traffic operations.

The core component within the ATIS is traffic information generated through a three-step procedure: collecting data from a surveillance system (e.g., loop detectors, probe vehicles, and cameras), consolidating the collected information (e.g., processing, estimating, and predicting traffic information), and disseminating the information to various agencies and the public. The ATIS can provide the location and severity of an incident, weather and road condition, delay and travel time. Among those items travel time is probably the most important because it is the most straightforward piece of information needed by travelers to plan their journey and for transportation agencies to optimize network performance.

Travel time information can be classified into three categories: historical, real-time, and future information. Historical information describes the state of the transportation system in previous periods (e.g., previous time interval, the same day of previous week, the same week of previous month, etc.). Real-time information contains the most up-to-date traffic conditions (e.g., the travel time at the current moment). Future information (e.g., what will happen after 5 minutes) can be predicted based on historical and real-time information.

Predicted travel time information can be classified into long-term and short-term. The predicted long-term information is mainly used for transportation planning. The short-term prediction often focuses on a short period in the future from a few minutes to a couple of hours. It is more suitable to use it for managing traffic operations and control, such as navigating vehicles, dynamic traffic assignment, and network wide traffic control.

Most of the existing traffic management systems provide real time traffic information and determine appropriate traffic control plans based on historical and real-

time traffic data, such as NAVIGATOR (Georgia DOT), TransGuide (Texas DOT), MONITOR (Wisconsin State DOT), SunGuide (Florida DOT), Compass (Ontario, Canada), GuideStar (Minnesota DOT), LISB (GERMANY), AUTOGUIDE (London), 511 system (Arizona DOT), Fastracs™, and the Advanced Driver and Vehicle Advisory Navigation Concept (ADVANCE) (Illinois DOT). The disseminated traffic information only contains current traffic information. Therefore, the performance of these systems is constrained because of the lack of a predictive feature. Some traffic management systems (e.g., Travel Time Prediction System (TIPS) in Ohio; TransGuide at San Antonio, TX; the AC Transit in Almeda County, California; the City-University-Energysaver (CUE) bus system in Fairfax, Virginia; the Vail Bus Service in Vail, Colorado; the Municipal Railway System in San Francisco, California; the Tri-Met Transit Tracker System in Portland, Oregon, etc.) were developed with predicted vehicle travel/arrival times. Thus, future traffic information can be provided to travelers.

With accurately predicted traffic information, travelers can make a smarter decision for selecting better departure times and routes to avoid traffic congestion and reduce the travel time. The impact of the ATIS information has been explored in numerous studies. Koutsopoulos et al. (1993) developed a simulation model to evaluate travelers' responses to the disseminated traffic information and explored the benefits travelers obtained. Daganzo and Lin (1994) found that the most useful information for route choice was predicted travel time. Motorists making decisions in absence of either predicted or real-time traffic information are implicitly based on historical information. If the travel choice is based on instantaneous information, it may lead to inferior solutions since conditions can damage while the traveler is on route. Therefore, accurate short-term

prediction of traffic conditions (e.g., future travel time ranges between 5 minutes to a couple of hours) is desirable for assisting trip planning.

Providing accurately predicted traffic information is critical for the successful implementation of ITS, especially for In-vehicle Route Guidance Systems (RGS). As soon as traffic conditions change, a more reliable routing plan can be dynamically generated considering future, rather than instantaneous, traffic conditions. Many studies have been conducted on RGS using real time and predicted information. Fu (2001) developed an adaptive routing algorithm for RGS, in which real-time travel time information was employed. The algorithm could predict the mean and variance of link travel times before a vehicle entered the link. Sung et al. (2000) employed an efficient labeling algorithm to determine the shortest path with time-dependent traffic speeds. Chabini (2001) developed a model with an enhanced feature for the shortest path problem with time-dependent link travel times under first-in-first-out (FIFO) conditions, which was proved a computationally efficient approach for the dynamic shortest path problem.

Hall (1996) investigated the potential impact of real time traffic information in route choice by increasing market penetration (e.g., percentage of drivers who subscribed to receive the traffic information). It was found that when drivers responded to accurately projected travel times, the system performance (e.g., delay time) could be improved with increased market penetration. It was also found that predicted travel time is the most useful information for traffic management rather than individual motorists.

2.2 Traffic Data for Travel Time Prediction

To predict future travel times, historical and real-time travel time information has to be collected, processed and analyzed in a systematic procedure. Though this study focuses on the development of prediction models, technologies and methods used for collecting data, estimating travel time and identifying traffic pattern are discussed in this section.

2.2.1 Data Collection

Travel time information, nowadays can be efficiently collected by using detectors (Dailey, 1993; Petty et al. 1998), probe vehicles (Chen and Chien, 1999; Dailey, 2003), and Global Positioning System (GPS) technology (You and Kim, 2000; Tsai, et al. 2003). Methods that are used to collect travel times can be directly classified into four categories (FHWA, 1998): test vehicle, license plate matching, ITS Probe Vehicle, and emerging and non-traditional techniques. "Test vehicles" are dispatched specifically to collect travel times along their trips. The "License Plate Matching" method compares license plates passing consecutive checkpoints, while the times at the checkpoints are manually or automatically recorded. Thus, the travel time between the checkpoints can be estimated. "Probe vehicles" require remote sensing devices (e.g., GPS) to record a variety of data used to calculate travel time. The sensing devices can collect the latitude and longitude coordinates with a time stamp. With this information, one can determine how long it took the vehicle to go from one place to another.

2.2.2 Methods for Estimating Traffic Measures

Most traditional surveillance systems (e.g., loop detectors) cannot collect travel time directly, but can collect point measurements such as flow, occupancy and spot speed. The methodology used to derive travel time from point detection measurements are presented in this section. It was important to realize that the variance of travel times (Sen, 1999; Eisele and Rilett, 2002) and traffic patterns (e.g., stable, non-recurrent congestion or recurrent congestion) are critical information for prediction model development (Akahame, 1986; Shbaklo, et al. 1992), and this will be also discussed in this section.

(1) Point Detection Estimation

The single loop detector is one of the most commonly used devices for collecting speeds. A number of models have been developed for single loop detectors (e.g., Hall and Persaud, 1989, Dailey, 1993, Nam and Drew, 1996), while most of them are based on the fundamental traffic flow theory to approximate the speed from traffic variables (e.g., volume and occupancy). The three fundamental traffic variables, including flow, density, and speed, are commonly used to describe temporal and spatial traffic characteristics.

The traffic volume can be directly detected with a single loop detector. However, the density is often substituted by the occupancy that is the product of density and a mean effective vehicle length (MEVL), which is roughly equivalent to the sum of vehicle length and the length of the single-loop detector. Then, speed can be estimated (McShane et al. 1998) using Equation 2.1:

$$\bar{s}(i) = \frac{n(i)}{T * \frac{o(i)}{\bar{l}(i)}} \tag{2.1}$$

where i is time interval index; $\bar{s}$(i) is space-mean speed; n (i) is number of vehicles per interval (volume); o (i) is occupancy; T is the interval length; and $\bar{l}$(i) is the MEVL for the interval.

Athol (1965) treated the MEVL as a constant in different time intervals to estimate speed for single loop data, as shown in Equation 2.2 where g is the reciprocal value of the MEVL. This method is commonly adopted in practice because it does not require complicated calibration.

$$\bar{s}(i) = \frac{n(i)}{T * o(i) * g} \tag{2.2}$$

Hall and Persaud (1989) tested the validity of Athol's model and found that biased estimates of speeds may be happening when two major assumptions at traffic flow theory are violated. The first one assumes that traffic flow is uniform, which is not true under congested and very light traffic conditions. The second assumption is that the occupancy and density of the traffic flow are linearly related (e.g., occupancy = c * density, c is a constant by assuming that vehicle lengths are identical), which is not true either.

Wang and Nihan (2000) developed a log-linear model to update "g" in Equation 2.2 periodically for different vehicle composition. The MEVL and "g" in different periods can be estimated with improved accuracy. Considering the vehicle length distribution in different traffic compositions, Wang and Nihan (2002) later employed a filtering algorithm to make MEVL consistent with the actual vehicle composition across intervals. They found that the estimation accuracy was significantly improved in

comparison with the unfiltered data case under both free-flow and moderately congested conditions.

Single-loop detector measurements are commonly used in practice for their simplicity and transferability. The estimated speed could be further used to estimate the link travel time. However, the calibration of the estimated travel speed requires data collected from other hardware such as from a video camera.

Dual-loop detectors have been used to monitor traffic volumes, occupancies and speeds with lesser calibration efforts. There are a number of models used to estimate travel times based on data collected from dual-loop detectors. Dailey (1993) used cross-correlation of traffic flows collected from upstream and downstream detectors to estimate the travel time between two detectors placed 0.5 miles apart. Later, Petty et al (1998) applied the same method to estimate travel times, while the accuracy of estimated data under different aggregated levels were explored. Nam and Drew (1999) developed a new dynamic traffic flow model that can automatically estimate spatial variables, such as travel times and space-mean speeds. The model was based on the conservation of flow-density-speed relationship, and it performed well for both normal and congested traffic conditions. With those models, the detected measures (e.g., speed, occupancy, etc.) could be applied to estimate the travel time.

(2) Variance Estimation

Variability of travel times on links is an important measure of service quality for travelers (Eisele and Rilett, 2002). The variance of travel times could affect the reliability of predicted travel time to determine the shortest path (Sen et al., 1999) and acts as an effective index to identify the traffic condition (Rilett et al., 1999). However, stochastic

traffic conditions and the lack of enough data collected by surveillance systems increases the difficulty to estimate travel time variance. For example, the estimated variance of travel time could be easily obsolete if the studied corridor length is long because the travel time changes dynamically (Eisele and Rilett, 2002). Further, the wide variation of travel times requires a large size of travel time data to determine the mean and variance of travel time subject to specified accuracy (Chen and Chien, 2001).

Chen et al. (2002) studied the average travel time and its variability based on data collected from a corridor in Los Angeles, CA. Results showed that both the travel time and its variability could be used as performance indices for determining the level of service. Rilett et al. (1999) developed a model to determine prediction intervals to forecast corridor travel times. With the mean and variance of link travel times, the mean and variance of the corridor travel time were approximated using a Bootstrap method and a Taylor series model.

Considering an estimated link travel time with high variance, Sen et al. (1999) developed a model to compute the minimum travel time between an origin and destination pair. Results showed that the variance of estimated travel times collected by probe vehicles would remain large irrespective of sample size increase. Thus the accuracy of predicted travel time was not necessarily improved by increasing the number of probe vehicles.

Eisele and Rilett (2002) developed a method to calculate the variance of corridor travel times based on link travel times. Based on the first- and second-order Taylor Series approximation, the variance and covariance of corridor travel times were computed considering the correlation among travel times of vehicles on the link. Gajewski and

Rilett (2003) applied the Bayesian Smoothing method to estimate the correlation of link travel times. A nonparametric regression model was first used to model the mean travel time of a link. Subsequently, a Bayesian-based method was developed for estimating the distribution of the correlation of travel times between links along a corridor in Houston, Texas, which was instrumented with Automatic Vehicle Identification (AVI) systems.

Traditionally, the estimation of travel time and its variability was based on aggregated data on different levels of detail (from a few minutes to a couple of weeks). Zietsman and Rilett (2000) analyzed travel times at a disaggregated level by using AVI data. They found that the increased quality of traffic data provided by AVI can improve the estimation accuracy.

Gerlough et al. (1971) found that the collected traffic data would exhibit different degrees of variability depending on traffic volume (volume-to-capacity ratio) and the duration of time interval for aggregating data. They selected Variance-to-Mean ratio (V/M) as an effective indicator for identifying the probabilistic distribution for model errors. For example, the observed traffic flow would be binomially distributed if V/M is less than 1. It was further suggested that typical situations associated with low V/M ratios were flows under congestion. Lan and Miaou (1998) explored the probabilistic distributions of traffic flows with different time interval duration and discussed the corresponding prediction boundaries for predicted traffic flow. The prediction boundaries were generated based on the estimated error distributions. The variance and probabilistic distribution of traffic data are important input parameters in many applications of ITS such as incident detection, traffic control, traffic route guidance, and traffic prediction.

2.2.3 Traffic Pattern Identification

Traffic patterns have been used to classify different traffic conditions while developing prediction models (Akahame, 1986, Shbaklo et al., 1992, Smith et al. 2002). Shbaklo et al. (1992) conducted a review of short-term travel time prediction models and found that it was an effective way to classify traffic conditions into different traffic patterns (e.g., free flow and congestion conditions). Traffic characteristics such as variation of travel time may be very different under various traffic patterns. Akahame (1986) stated that the traffic condition identification is a preliminary and important step for developing an appropriate and effective prediction model.

The traffic condition can be either free flow (un-congested) or congested conditions. Under a free flow condition, traffic moves smoothly without delay. Therefore, the traffic variables (e.g., speed, volume and occupancy) can be estimated with using traffic flow theory. Thus, the prediction of travel time under such a condition can be performed fairly well by many methods including the historical average method (Jeffrey, et al., 1987), ARIMA models (Okutani and Stephanedes, 1984), Kalman filter models (Suzuki et al., 2000), and artificial neural networks (Smith and Demetsky, 1994).

The traffic condition worsens and may become jammed when demand approaches road capacity. The congested traffic condition can be classified into recurrent congestion such as heavy traffic regularly happened during peak hours, and non-recurrent congestion due to reduced capacity caused by incidents (e.g., accidents, spilled truckloads, maintenance activities, and stalled vehicles). Recurrent congestion can be determined from historical data since peak demand occurs with predetermined period. The methods used to identify non-recurrent congestion or incidents are introduced below.

A number of incident detection methods were developed over the past several decades. Most of them used congestion as an index to identify whether the congestion was caused by an incident or a recurrent bottleneck condition (Gall and Hall, 1989). Persaud and Hall (1989) analyzed traffic data (e.g., volume, occupancy, and speed) collected near an accident on a freeway. The data were poorly explained by traffic flow theory, but fit very well by a catastrophe theory model considering the transition to and from the congested condition upstream of the incident. Thus, an outline of an alternative logic, based on the catastrophe theory model of the flow-occupancy-speed pattern, was suggested for freeway incident detection.

Gall and Hall (1989) developed a logic model to distinguish recurrent and non-recurrent congestion. The method was able to monitor detector data at regulated time intervals and to evaluate the characteristics of traffic operations under congested condition. Thus, the reduced capacity caused by incidents can be detected for non-recurrent congestion identification. Abbasi et al. (1999) developed a model to measure variance of speeds on freeways under a recurrent congested condition. The results were derived to measure congestion presence and severity as well as evaluate ATMS implementation.

To reduce false alarm rate of incident detection, Stephanedes and Chassiakos (1993) developed a filtering algorithm to detect freeway incidents. The algorithm considered smoothing spatial occupancy differences between upstream and downstream detectors. An incident would be detected when the difference increased significantly in a short time period. The method provided an alternative way for incident identification.

Artificial Neural Networks (ANNs) and fuzzy algorithms have been broadly used for incident detection. Cheu (1996) developed a new incident detection model by using an ANN to detect lane-blocking incidents on urban freeways. The study classified traffic surveillance data obtained from inductive loop detectors, and then used them to detect incidents. Dia and Rose (1997) developed a multi-layer feed-forward (MLF) ANN model to detect incidents, which was evaluated using field data. Results confirmed that the ANN model could provide fast and reliable incident detection on freeways.

Later, Hsiao (1997) developed a Fuzzy Logic Incident Patrol System (FLIPS) by integrating a fuzzy logic system and the learning capabilities of ANN to automatically detect freeway incidents. A Fuzzy Logic Adaptive Threshold System (FLATS) was developed to convert the output of FLIPS, an incident detection index into a binary value predicting the absence or presence of an incident. The evaluation results showed that the integrated learning process and adaptive threshold technique can significantly improve incident detection performance.

Jin et al. (2002) developed a constructive probabilistic neural network (CPNN) for freeway incident detection. The CPNN was structured based on a Gaussian model and trained by a dynamic decay adjustment algorithm. The results showed that CPNN has three main advantages over a basic probabilistic neural network (BPNN): (1) CPNN has clustering ability and therefore could achieve similarly good incident-detection performance within a much smaller network size; (2) each Gaussian component in CPNN has its own smoothing parameters that can be obtained by the dynamic decay adjustment algorithm with a few epochs of training; and (3) the CPNN adaptation methods have the ability to prune obsolete Gaussian components.

Sheu (2002) presented a model based on the fuzzy clustering theory to identify freeway incidents. The model was able to distinguish a congested traffic condition from incident-free traffic states and characterize temporal and spatial changes of traffic variables such as occupancy, density and speed related to the incident. Skabardonis et al. (2003) developed a method to measure total delay on urban freeways under recurrent and non-recurrent congested conditions, while the data collected by loop detectors were analyzed to approximate the average and the probability distribution of delay. The method could be used to evaluate congestion impact on travel time and its variations. Srinivasan et al. (2000) developed a hybrid ANN for detecting incidents on a transportation network. The comparison between the hybrid ANN and four other incident-detection algorithms (e.g., California algorithms, McMaster algorithms, Minnesota algorithms, Artificial Intelligence (AI) models) was conducted. Results demonstrated that the hybrid ANN had superior accuracy for incident detection.

2.3 Traffic Prediction Models

Many traffic prediction models have been developed based on available traffic information including historical and real time traffic data. Some models only used historical traffic data (Kreer, 1975 for UTCS-2; Ahmed and Cook, 1979; Hoffmann and Jando, 1990; Khattak et al. 1992; and Smith, 1997), some rely on real-time traffic information (Stephanedes et al. 1981 for UTCS-3; Lu, 1990; Suzuki, et al., 2000), and others (Williams, et al. 1999; Kuchipudi and Chien, 2003; Shalaby and Farhan, 2003) employed both historical and real-time data to predict future information. In this study, prediction models are classified into time series models, regression models, and ANN

models. The application, advantages, and disadvantages of each model class are discussed next.

2.3.1 Time Series Models

Time Series is defined as an ordered sequence of values of a variable at equally spaced time intervals (Croarkin, etc, 2001). Time series analysis is a statistical model (Smith and Demetsky, 1994) that can catch and explain the system behavior based on a set of time series data x(t) and thus use them to forecast future condition at time t+D, where D is the prediction interval. The model can predict the x(t+D) based on a time series x(t), x(t-D), x(t-2D), and so on. The time series technique has been wildly used in transportation research (e.g., Ahmed et al., 1979, Okutani and Stephanedes, 1984, Terry et al., 1986, Arem et al., 1997, Williams, 2001)

Time series models (Chatfield, 2002) can be classified into "univariate" and "multivariate" types based on the dimension of the observation variables (e.g., scalar or vector). However, multivariate time series models are not wildly used for traffic forecasting due to their complicated formulation requirements and no further details about them will be presented here. The two most popular and wildly used time series models are discussed below.

(1) Box-Jenkins Models

Enders (1996) defined the Box-Jenkins model as "a methodology for identifying, estimating, and forecasting" based on an autoregressive moving average (ARMA) that was made up by autoregressive and moving-average components. As a well-known Box-Jenkins model, the Autoregressive Integrated Moving Average model (ARIMA) was

wildly used to predict future traffic flow and travel time. ARIMA models can be simply defined (Williams, et al, 1999) as linear estimators regressed on past values of the modeled time series (the autoregressive terms) and/or past prediction errors (the moving average terms). ARIMA (p,d,q) represents a time series {Xt}, in which parameters p and q represent the order of autoregressive and moving average, respectively, and d represents the order of ordinary differencing (Smith, etc., 2002). It is assumed that travel time prediction is a point process and that the observed data can be identified by statistical techniques.

The fast computation times associated with the ARIMA model make it an efficient prediction method for real-time traffic management. One shortcoming of ARIMA is that the time series model relies on a consistent series of data. Once the data series is interrupted, the missing data will decrease the accuracy of the predicted results. For example, Okutani and Stephanedes (1984) applied the ARIMA model to UTCS. The results showed unsatisfactory goodness of fit and high errors due to missing data. Under such a condition, ARIMA models are not more accurate than a simple moving average method. Terry et al., (1986) stated that time series models should "be used with caution as the complexity of the situation increases". Later, Williams et al. (1999) used a seasonal ARIMA and an exponential smoothing method to predict traffic flow. The single exponential smoothing was a simplification of seasonal ARIMA assuming no seasonality, and the technique performed well in replacing the missed data and improving prediction accuracy.

The moving average model is one type of the Box-Jenkins models, and usually uses an average of historical traffic data in previous time intervals to forecast the future

condition. This method assumes that future traffic flow would follow the same pattern as the past one, relying mainly on the cyclical nature of traffic conditions. However, the lack of a dynamic reaction to the changing traffic will fail this method when the system experiences unexpected changes caused by incidents, such as adverse weather condition, accidents, and other special events.

Another type of smoothing method is called exponential smoothing. A single exponential smoothing method is actually a special case of ARIMA when (p,d,q) are set to (0,1,1) (Brown, 1963; Ross, 1982). The exponential smoothing model assigns time-varying weights to the previous observations when the prediction results are updated (Filliben, et al., 2002). Iwasaki and Saito (1996) developed a speed prediction model based on a single exponential smoothing method with historical data. The weight of the historical speed value was selected by a trial and error approach. Results generated with data collected from a traffic surveillance system on a rural motorway in Japan showed that the prediction model could overcome the time lag while predicting traffic flow and perform well in real practice.

Many researchers made various theoretical and empirical arguments for selecting appropriate smoothing parameters such as α discussed here. Bowerman and O'Connell (1979) stated that in practice smoothing constants between 0.01 to 0.3 usually worked quite well for prediction. However, in a study conducted by Makridakis et al. (1982), α values above 0.30 frequently yielded the best forecasts. The dependency of optimal α may be captured by time series methods that can analyze the autocorrelation of data. Gardner (1985) concluded that it was best to estimate an optimum α from the historical and real time observations, rather than a guessing value.

Newbold (1995) developed a method for selecting a smoothing constant. The recommendation was to select α that minimizes the sum of squared prediction errors. According to a study conducted by StatSoft Research Team (2004), α can be chosen by a grid search in the feasible region of the parameter space in practice; that is, different values of α should be tried, for example, starting with $\alpha = 0.1$ to $\alpha = 0.9$ by increment of 0.1. Then, the α which produces the smallest absolute mean residual error is chosen.

The Urban Transportation Control System (UTCS) was developed based on a Box-Jenkins model, whose parameters were determined off-line using a data set collected from the studied site (Kreer, 1975; Stephanedes et al., 1981) to predict traffic flow and travel time. The UTCS-2 predictor employed both historical and current-day measurements. Thus, it can reduce the prediction error resulting from the difference between deviations from the historical data and current conditiosn. The UTCS-3 predictor only employed real-time measurements to interpolate the most recent smoothed and un-smoothed measurements on the current day as the predicted values. Under free flow traffic conditions, the algorithm employing historical information (UTCS) as reference provided better prediction than those using current-day measurements.

The AUTOGUIDE system (Jeffrey et al., 1987) was developed in London. It utilized simply the storing at historical traffic data to predict travel time in different periods of a day. Though the model is not able to reflect dynamic traffic changes, it performed very well in practice when there is no available real-time traffic data. Kaysi et al. (1993) developed the LISB system by applying a historical average model to predict future traffic conditions based on historical and real-time data. A projection ratio of the historical travel time on a specific link to the current travel time as reported by equipped

vehicles was used to predict future link travel times. The weakness was to assume that the projection ratio remained a constant.

Coifman (1996) developed a basic signal processing theory to smooth freeway loop detector data. After introducing fixed-time averages, cumulative sums, and moving-time averages, a nontraditional smoothing technique was developed. Results indicated that good estimates could be achieved even under congested traffic conditions by employing historical average methods.

The Box-Jenkins model combined the moving average and the autoregressive method, which can increase the prediction accuracy. Ahmed et al. (1979) and Nihan et al. (1980) used the Box-Jenkins method to forecast traffic volume on freeways. Arem et al. (1997) developed a linear ARIMA model using data collected from inductive loop detectors to predict travel times. The predicted and actual travel times under both recurrent and non-recurrent congested conditions were compared. The difference was very small under recurrent congestion, but it was large under non-recurrent congested condition.

Vemuri and Pant (1998) developed a time series model for short-term forecasting of traffic delays for highways with construction zones. Travel times were estimated from two adjacent detectors and converted to a set of time series data. Thus, the problem of short-term traffic delay forecasting could be formulated as a time series evolution problem. A generic structure applied an on-line approximation model to predict travel time based on real time and historical data. Satisfactory performance was achieved.

Lee and Fambro (1999) developed a ARIMA model along with three time-series models to predict traffic volume. Results showed that all time-series models performed

well with reasonable accuracy for short-term forecasting tasks overall. On the average, the exponential smoothing model was better than time series models except for the ARIMA model. It was found that proper model identification couldn't be guaranteed when data have an ARMA structure. Williams (2001) developed a multivariate ARIMA model to predict downstream traffic flow by using data collected by upstream sensors. Results indicated that the multivariate ARIMA model outperformed a univariate model.

Kamarianakis and Prastacos (2003) developed two different univariate (historical average and ARIMA) and two multivariate (vector autoregressive moving average, or VARMA and single space-time ARIMA, or STARIMA) models to predict traffic flow. Results showed that the historical average model cannot predict well during the transition of changing traffic patterns. The STARIMA model could generate as accurate prediction results as other models (ARIMA, VARMA and STARIMA model) with relatively simple structure.

(2) Kalman Filtering Models

Static time series models, such as historic average and Ben-Jenkins models cannot capture dynamic changes in traffic operations (Chatfield, 2003). The prediction accuracy of these models depends on the similarity between the historical and future traffic conditions. The Kalman filtering model is a dynamic linear prediction model consisting of state and measurement equations. The state equation describes how the state variables change over time, while the measurement equation transfers the estimated state variables into observed variables. Then, the prediction error (difference of the predicted and observed variables) can be used to update the state variables in the next time interval.

The Kalman filtering model was originally developed for signal processing. Then it was applied in transportation research to predict traffic volume in an urban network (Okutani and Stephanedes, 1984). Unlike the methods using only historical data for prediction, the Kalman filter model uses time-varying parameters that can fully respond to dynamic conditions. The parameters are updated based on the most recent observation. Thus, the prediction error between the predicted results and future traffic conditions could be minimized.

Suzuki et al., (2000) applied a Kalman filtering model to simultaneously predict dynamic travel times and traffic flow on a freeway corridor in Bangkok, Thailand. The state and measurement equations in the Kalman filtering (KF) model were integrated with an ANN model, called Neural Kaman filter (NKF) model, to predict traffic states on the freeway corridor. Numerical analysis showed that the use of the NKF model was fairly accurate in predicting dynamic travel time and traffic flow.

Bhattacharjee et al., (2001) developed a Kalman filtering model to predict O-D demand, where the effect of travel time information on traffic diversion was formulated. The inherent dynamic nature of traffic characteristics was captured. Changes in drivers' perceptions as well as other randomness in the route diversion behavior have been modeled using an adaptive, aggregate, dynamic linear model where the model parameters are updated on-line using the Bayesian updating approach. Experimental results based on a freeway corridor in northwest Indiana indicated that the developed dynamic linear model (DLM) could significantly improve O-D demand prediction accounting for route diversion behaviors.

Jiang (2003) combined the Kalman filtering model and the ARIMA model for predicting freeway traffic. It was demonstrated that the integrated model predicted traffic flow in a more accurate manner than either one of the two models. The traffic flow on a freeway in the next time interval (e.g., 5 to 15 minutes later) can be predicted using traffic data collected in the current and past time intervals. Dynamic traffic predictions with the developed model can be performed for individual lanes as well as for all the lanes of each travel direction. The study also demonstrated that, if the traffic capacity was given, a dynamic prediction of traffic flow rate with this prediction model would also constitute a dynamic prediction of traffic congestion.

Sheu (2003) developed an advanced method for predicting queue lengths based on a stochastic system approach. A Kalman filtering model for predicting real-time queue overflows was developed, in which a discrete-time stochastic system was specified for modeling the queue length when traffic overflow occurred, while a random process was used to predict traffic arrivals. Results indicated that the proposed method held promise in real-time prediction of queue overflows.

Nanthawichit et al. (2003) developed a method to estimate and predict traffic variables (e.g., travel time) with data collected by probe vehicles and detectors. The data collected by probe vehicles were integrated into the measurement equation of the Kalman filtering model. Estimated travel times were updated with information from both stationary detectors and probe vehicles. The application of the developed method was extended to the estimation and short-term prediction of travel time.

Kuchipudi and Chien (2003) developed dynamic recursive models to predict travel time employing both real-time and historical data. Path-based and link-based

Kalman filtering models were developed and tested, while their performance under different traffic conditions was compared. The duration of the predicted time intervals was analyzed to evaluate its impact on the accuracy of the prediction model. The experimental results revealed that the predicted travel times with path-based model were better than those predicted with the link-based model during peak hours.

Chen et al. (2003) developed a neural/dynamic (N/D) model by integrating a Kalman filtering model and an ANN to predict bus arrival times under congested traffic conditions. The predicted bus travel times from the ANN model was taken as input to the Kalman filtering algorithm to recursively update the predicted data based on real time information collected by automatic passenger counters (APC). After a statistical analysis of all trips, the N/D model with a lower RMSE always performed better than the ANN model. Thus, it provided a promising way for predicting bus travel times.

(3) Non-linear Time Series Models

The linear models, such as the Box-Jenkins and Kalman filter models, are the main time series models used for traffic forecasting. However, Chatfield (2002) found that there was increasing interest in non-linear systems with exciting potential when the observed time series exhibited features that cannot be explained by a linear system.

Al-Deek et al. (1998) developed a non-linear model to predict travel times, which was tested with real world data collected under recurrent congestion. The maximum prediction errors of travel times were found when traffic was jammed. A minimum speed threshold was set as a boundary of congestion. The improved results were achieved by refining the prediction algorithm and smoothing the input data. Thus, the new approach produced reasonable errors for short-term (5-minute) travel time predictions.

Angelo et al. (1999) developed a nonlinear time series model for travel time prediction on a freeway network in Orlando, Florida. Both single-variable and multivariable types of prediction models were developed with real world traffic data. In single-variable prediction, speed time-series data are used to forecast travel times along the freeway corridor. Multivariable prediction schemes are developed using speed, occupancy, and volume data provided by inductive loop detectors on the study section. Results demonstrated that the performance of the single-variable model was superior to the multivariable model. It also proved that the developed nonlinear model has strong potential for on-line implementation during changing traffic conditions.

Nair et al. (2000) investigated the applicability of a non-linear system for traffic parameters (e.g., speed, volume, occupancy) modeling and prediction. They modeled a traffic time series as a deterministic system, implying existence of an underlying structure in the system, and applied the phase space technique to the non-linear time series analysis. The experiment was designed based on the data collected from loop detectors on the San Antonio freeway. The nonlinear time series technique was then used to predict average speed and demonstrated better prediction results than a neural network.

Ishak and Al-Deek (2003) investigated factors that had a significant impact on the accuracy of predicted travel times by a nonlinear time series model. The goal was to identify the operational settings and the anticipated predicting accuracy of the model before it can be fully implemented. The study was conducted using an extensive amount of real-time data collected from the 62.5 km corridor of I-4 in Orlando, Florida. Various scenarios were generated from a combination of model parameter settings and different traffic conditions to test the performance of the model. Relative travel time prediction

errors were used as measures of performance. Analysis was conducted to identify the impacts of parameters (e.g., prediction horizon, congestion index) on the model's performance.

2.3.2 Regression Models

Regression analysis is a statistical procedure commonly applied in solving management, business and engineering problems (Chatfield, 2002). Regression models could be classified into parametric and non-parametric models, by which a single dependent variable affected by one or more explanatory (independent) variables can be analyzed. Therefore, the dependent variable could be estimated based on a set of independent variables under different conditions.

(1) Parametric Regression Models

The parametric regression model includes linear least squares regression, nonlinear least squares regression, weighted least squares regression, etc. The linear regression model is developed with the "least squares" method to fit a line through a set of observations, and is by far the most widely used modeling method (Croarkin, et al., 2001). The applications of regression models in the area of traffic prediction are reviewed and discussed below.

Abdelfattah and Khan (1998) developed linear and a non-linear regression models to predict bus delays. The explanatory variables were chosen based on available data, while their significance on the dependent variable was explored. The developed bus delay prediction models were tested under normal and incident conditions. Results showed that

the predicted bus delays were close to the field data, and were used to assist transit operators updating bus schedules on a real time basis.

Frechette and Khan (1998) employed the Bayesian regression approach to predict travel times for urban streets. Two disparate sources of information were used: priori data (what is known before an experiment) and experimental data (information derived from an experiment). Results generated by NETSIM (a microscopic network simulation model developed by Federal Highway Administration) were used as priori information, whereas the experimental data were obtained from videotape records on the selected streets. A Bayesian regression software was used to develop the prediction model for travel times during peak periods in medium to large central business districts.

Zhang and Rice (2003) proposed a method to predict freeway travel times using a linear regression model, in which the coefficients of explanatory variables varied with the smooth weight functions of vehicle departure time. Thus, the travel time with closer departure time to the prediction time played a larger role in determining the coefficients of the regression model and obtained improved prediction accuracy. The effectiveness of the method has demonstrated and tested with loop detector data. The method was straightforward to implement, computationally efficient and applicable to widely available sensor data collected on freeways.

Sun et al. (2003) developed a local linear regression model for short-term traffic prediction. Local linear regression is one type of local weighted regression methods, which has been applied to artificial intelligence, dynamic system identification and data mining. The performance of the model was compared with nonparametric approaches (e.g., nearest neighborhood and kernel methods) by using traffic speed data collected on

the Houston US-290 Northwest Freeway. Results indicated that the local linear model outperformed the nonparametric models.

Cho and Rilett (2003) developed a modular ANN, a simple ANN model, and a standard linear regression model to predict train arrival times. The average speed, squared average speed, and estimated distance between the head of the train and the highway-railroad grade crossings (HRGC) were selected as explanatory variables to develop the regression model. It was found that a modular ANN architecture could produce superior results. When the updated speed was refilled and trained in the models, the accuracy of prediction was improved.

Shalaby and Farhan (2003) developed a linear regression model to predict bus travel times with AVL data, in which speed, distance, dwell time and delay were selected as explanatory variables. Historical average, Kalman filtering and ANN models were developed, and the prediction accuracy was compared. The results showed that the Kalman filtering model outperformed others methods for bus travel times in downtown Toronto.

(2) Nonparametric Regression Models

The nonparametric regression model is a regression model that does not require a functional relationship between the dependent and explanatory variables. Nonparametric regression prediction model is founded on chaotic system theory, which is defined by deterministic and non-linear state transitions. When there is no prior knowledge about the model that can represent the system with a finite number of parameters, nonparametric regression can be used to explore the data.

By applying historical observations in a specific area (neighborhood) close to the current variables, the nonparametric regression model assigns a weighted average value to predict the dependent variable. Smith etc., (2002) found that the quality of the database, particularly when storing historical data that represents all possible future conditions, greatly affects the accuracy of nonparametric regression models.

Altman (1992) classified the nonparametric model approaches into kernel and nearest neighbor based on the definition of past instances neighbor. A kernel neighborhood was defined as having a constant bandwidth on the independent variable space with the center on the current investigated state. A nearest neighborhood was defined as having a constant number of data points including those "nearest" to the current system state. Therefore, a prediction result was always available from the nearest neighbor approach, while it could be missing in a kernel neighborhood approach, because any past similar instances could possibly fall out of the pre-defined bandwidth.

Davis and Nihan (1991) investigated the non-parametric regression method for short-term traffic forecasting. They developed a K-nearest approach to overcome the inherent problems in parametric forecasting methods (e.g., nonlinear relationship between input and output). An empirical study using real world data collected on a freeway in Washington State was conducted, while a K-nearest model and a linear time-series model were tested. Results showed that though the K-nearest model did not perform better than the linear time-series model, the possible reasons were found, including a dataset that did not contain enough instances of traffic flow transitions. However, a more promising method was expected when more extensive data and K were provided to produce superior forecasting.

Smith and Demetsky (1996) developed a nonparametric regression model to predict traffic flow applying multiple prediction intervals. The model could expand the prediction horizon in intervals from 15 minutes up to 1-hour compared with the single-interval formulation of nonparametric regression models. The prediction performances were comparable for both models. Later, Smith (1997) developed four forecasting models (historical average, ARIMA, artificial neural network, and nonparametric regression) to predict traffic flow for two sites on the Capital Beltway in Northern Virginia. The comparison of results indicated that the nonparametric regression model outperformed the others with single interval prediction. Based on its success with the single interval forecasting, the nonparametric regression approach was used for multiple interval prediction and performed well in this application.

Smith et al., (2002) did a study to compare parametric and nonparametric models while forecasting traffic flow. Both ARIMA and nonparametric models were tested and the short-term traffic flow on a single point at London Orbital Motorway, M25 was predicted. They reported four fundamental challenges when applying nonparametric regression: definition of an appropriate state space, definition of a distance metric to determine nearness of historical observations to current conditions, selection of a forecast generation method given a collection of nearest neighbors, and management of the potential neighbors database. Results indicated that the heuristic method employing different historical data significantly improved the performance of the nonparametric regression model. It also demonstrated that using nonparametric regression was preferred when the implementation of seasonal ARIMA models was restricted by some requirements, such that missing data cannot be used in the model.

2.3.3 Artificial Neural Network Models

Artificial Neural Networks (ANNs) have been broadly used in transportation research, such as traffic flow estimation and prediction, traffic pattern identification, traffic control and transportation planning. Meldrum (1995) conducted a study and found two ANN inherent disadvantages inhibited in the ANN. One is that ANN sometimes requires long time for training data and may generalize new data outside the training set. Another was that a trial-and-error procedure was involved to find the optimum architecture and training techniques.

The ANN function can be taken as "a sophisticated form of regression" (Smith and Demetsky, 1994) though it doesn't require the formulation of related traffic measurements, which enables the ANN to handle nonlinear surfaces. The performance of the predictive ANNs substantially depends on the input-output specification and the quality of training samples. There is no robust theory to determine the best training procedures.

Smith and Demetsky (1994) developed a back-propagation ANN model and a time series model to forecast short-term traffic flow. The comparison of the prediction results showed that the back-propagation ANN model was more responsive to dynamic traffic conditions than the time series model, and held considerable potential application in real-time ITS applications.

Bae (1997) applied ANN to interpret auto travel time directly from bus travel time. The regression model was used to identify the correlation between bus and auto travel times on a link with dynamic traffic flow. Both dynamic and static field data, which affected the travel time of bus and auto were collected and used to validate the

travel time prediction model. The ANN results outperformed the regression model when dealing with a non-linear system.

McFadden and Durrans (2001) applied ANN to predict speeds on a two-lane rural highway. They developed two back-propagation ANNs for speed predictions and compared their prediction results with the regression models developed by Krammes et al. (1995). Comparision results indicated that the explanatory performances of the ANN models were comparable to the regression model, but ANNs were not limited by distributional constraints inherent to regression. It demonstrated that ANNs were more powerful for modeling different explanatory variables than regression models.

Chen and Grant-Muller (2001) integrated ANNs with a sequential learning process to develop a dynamic traffic flow forecasting system. An ANN was based on a Resource Allocating Network (RAN) that adapted to a sequential learning scheme under which data points were presented to the network in sequence. Results showed that the network performance could be improved by dynamically adding a new hidden unit. The number of hidden units should correspond to the complexity of the underlying function that reflected the observed data.

Lint and Hoogendoorn (2002) developed an ANN for freeway travel time prediction. The ANN was derived from a state-space formulation and was tested by data collected from a densely used highway in the Netherlands. Results indicated that the model could accurately predict travel time and produce approximately zero mean normally distributed residuals.

Kisgyorgy and Rilett (2002) applied an ANN with two different approaches to predict travel time for a freeway system in San Antonio, Texas. The first approach used

the ANN for speed prediction and then calculated the travel time, while the second one predicted the travel times directly with the ANN. The best results were obtained, while travel times were directly predicted by the ANN. However, this model cannot be applied in practice due to the lack of new data.

Yin et al., (2002) integrated a fuzzy method with an ANN, called Fuzzy Neural Method (FNM) to predict traffic flow for an urban network. The FNM consists of two modules: a gate network (GN) and an expert network (EN). The GN classified the input data into a number of clusters using a fuzzy approach, and the EN specified the input–output relationship using an ANN. While the GN grouped traffic with similar traffic patterns into clusters, the EN modeled the specific relationship within each cluster. Both simulation and real world data were used and satisfactory prediction results were obtained.

Ishak and Alecsandru (2003) used multiple ANN topologies for short-term traffic prediction on freeways under different network and traffic conditions. Using a mix of traditional and modern ANN topologies, the prediction performance was evaluated under different settings (e.g., traffic situations, type of networks) and prediction horizons (e.g., from 5 to 20 minutes). Results showed that the developed ANN resulted in better performance. The study implied that the identifying appropriate traffic situations, types of networks, and traffic settings could improve the prediction accuracy.

2.4 Simulation Models

2.4.1 CORSIM

Computer modeling and simulating transportation networks have been widely and increasingly implemented in traffic analysis and evaluation because of two reasons. One is that the performance of traffic operations under different conditions and the traffic impact of tentative modifications can be tested and evaluated before implementation in the field. The other is that some measures of effectiveness (MOEs), such as travel time on a particular OD pair, are difficult to measure in the field due to the lack of an appropriate surveillance system. For example, Xie, et al., (2002) pointed out the difficulty of collecting relevant field data to estimate certain performance measures (e.g., delay time) in reality for conducting signalized intersection-related studies, such as gap acceptance and capacity analysis.

Based on the representation of traffic flow or vehicle movement (Chu, 2002), traffic simulation models can be classified into microscopic (CORSIM, TRANSIM, MITSIM, PARAMICS, VISSIM, AIMSUN2, etc.), mesoscopic (FREFLO, METANET, AUTOS, VISUM, etc.), and macroscopic (DYNASMART, DYNAMIT, INTEGRATION, METROPOLIS, etc.).

Microscopic models can identify a vehicle by different type (e.g., auto, truck and bus), specify driver characteristics (e.g., aggressive or cautious), and emulate stochastic driving properties (e.g., lane-change maneuver, car following logic and driver decision process) of vehicles individually. Therefore, a microscopic model can describe dynamic interactions among vehicles in a complicated traffic environment and generate vehicle behavior with the highest degree of disaggregated level under various traffic

circumstances. However, the most detailed geometric and traffic data are required for the development and calibration of the simulated network.

Mesoscopic models can simulate vehicles at a lower level of detail than the microscopic models since mesoscopic models emulate vehicle behavior at a higher aggregated manner. For example, instead of detailed interactions (e.g., car following logic) between vehicles, mesoscopic models determine vehicle and driving properties (e.g., the lane change decision) based on their aggregated characteristics (e.g., speed, density). Macroscopic models simulate vehicle and driving behavior at the highest aggregated level, in which the traffic stream is represented by aggregated traffic flow, density and speed. Therefore, there is no real driving behavior of vehicles (e.g., lane change maneuver) emulated in macroscopic models.

In this study, short-term travel time will be predicted and evaluated for the developed prediction models. A microscopic simulation mode is desirable to generate the most detailed measures such as travel time in different time intervals. Thus, CORSIM is an ideal simulator that can meet the study needs.

CORSIM (microscopic CORridor SIMulation) is the powerful core engine in the TSIS (Traffic Software Integrated System) suite developed by the Federal Highway Administration (FHWA). It integrates two microscopic simulation models: the arterial network model NETSIM and the freeway model FRESIM and has been applied extensively to a wide variety of areas by both practitioners and researchers. Through simulating stochastic individual traffic vehicle operations, CORSIM can produce sophisticated MOEs based on car-following and lane-changing models.

2.4.2 Calibration and Validation

Simulation models should be properly calibrated before implementation. To fine tune the CORSIM simulation model to mimic real-world traffic conditions, its parameters need to be calibrated. These parameters are related to driver behavior, vehicle performance and roadway throughput capability, such as free flow speed, mean discharge headway, and mean startup delay) (Xie, et al., 2002), car-following sensitivity parameter, vehicle length (Cohen, 2002), yield value, time to complete the lane-changing maneuver, and lane changing aggressiveness (Skabardonis, 2002).

There are generally two different approaches to calibrate microscopic traffic simulation systems. The first approach is model calibration, which re-establishes the input-output relation to obtain the desired system accuracy by changing the basic modules that describe the complex relationship between the input and output at the simulation systems. Since this approach required a good understanding of the particular simulator modules and accessibility of the modules internal resources, only simulator developers can perform such kind of calibration. The second approach is referred to as parameter calibration, defined as re-establishing the input-output relation to obtain the desired system accuracy by changing only those parameters that govern the input-output relationship in the systems being modeled (Lee, et al., 2001). Parameter calibration in a simulation model may be regarded as an optimization problem in which a set of values for operating parameters that satisfy an objective function are to be searched (Cheu, el at. 1998).

Considering the interdependent influence among different parameters on the simulation results, the impact cannot be isolated when changing several parameters at a

time. Therefore, the other parameters should be fixed to test the impact of each parameter on the simulation results. To identify the key parameters influencing the simulation results, a sensitivity analysis for various parameters should be conducted, and the impacts of changing those parameters be observed.

The calibration procedure can be made during different comparisons including graphical, aggregate, and statistical. The graphical comparison is a subjective validation approach, which is especially useful for testing the results generated by the simulation model preliminarily. It makes the comparison easy and visible. Aggregate means and standard deviations give general indication of system performances in the real world and in simulation. However, they do not present an accurate trend or indication on how variables perform over time, what patterns are created, or how much individual measurements deviate. An aggregate comparison, along with the graphical comparisons of scattered plots, reveals the similarities and discrepancies of the magnitude and changing pattern of variables. A statistical analysis is crucial for validating the proposed model based on real world and simulation sample data. It can be used for assessing the accuracy of the model, testing various hypotheses, and determining degree of correlation between real world and simulation. The following indices are used for statistical comparisons.

The Mean Absolute Percent Error (MAPE) measures the percentage error between simulation results and field data, and can be given by Equation 2.3:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|S_i - O_i|}{O_i} * 100\% \tag{2.3}$$

where n, S_i and O_i are sample size, observation i of simulation output, and observation i of field measurement, respectively.

The Root Mean Square Error (RMSE) denotes the error between simulation results and field data, and can be computed by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_I)^2} \tag{2.4}$$

where n, S_i and O_i are the same as in Equation 2.3.

Once satisfactory values for the calibration parameters of the model have been obtained, the model should be checked (Jayakrishnan, et al., 2002) by comparing measures of effectiveness (MOEs) observed in the field against comparable measures of effectiveness calculated by the model (Cohoe, 2002). Verifying a calibrated model in this manner is commonly called validation. The validation process establishes the credibility of the model by demonstrating its ability to replicate actual network and traffic patterns (Jayakrishnan, et al., 2002). Validation should only be performed after the model has been properly calibrated.

2.5 Summary

Models developed for predicting traffic data (e.g., flow, speed, travel time) have been classified into time series, regression, and ANN models. Each model has demonstrated good performance for traffic prediction under situation suitable for it. Their advantages and disadvantages are summarized in Table 2.1. To fulfill the objective, creditable models should be developed to predict traffic information by employing real-time and historical travel times.

The regression method is not suitable for this study because of two reasons. The first reason is that the regression model predicts based on the relationship between the

Table 2.1 Summaries of Prediction Models.

Models		Advantage	Disadvantage
time series models	**Box-Jenkins models**	Fast computation time Provide correlation between variables in consecutive intervals	no standard procedure to identify proper ARIMA model Difficult to optimize weight value in dynamic environment
	Kalman filtering models	Fully dynamic respond to traffic change	inherent time lag fit only linear system with normal distributed state variable
	Non-linear models	Handle non-linear system	fit only non-linear system, no dynamic features
Regression models	**Parametric regression models**	Formulate the relationship between dependent and explanatory variables	Static coefficients of explanatory variables in aggregated data Cannot reflect correlation between value in consecutive intervals Dependent variable need be different from explanatory variables
	Non-parametric Regression models	not require functional relationship kernel and nearest neighbor approaches	require sufficient dataset to contain enough instance Cannot reflect correlation between value in consecutive intervals
ANN		No functional relationship required, stronger modeling ability that can handle non-linear system	Long time for training data, Lack of standard procedure to find the optimum architecture

dependent variable (e.g., travel time) and a number of explanatory variables (e.g., speed, departure time, link length, etc.). However, both input and output of the prediction model in this study is travel time, which does not fit the requirement of the regression model. The second reason is that the coefficients of the explanatory variables in the regression model are derived based on historical data. Even though the new data could be entered into model to update the coefficients, the correlation between travel times on consecutive intervals cannot be determined in this model. Thus, the dynamic changes in the traffic system cannot be accurately captured.

The ANN is a powerful modeling approach especially when the relationship between the predicted variable and explanatory variables is hard to formulate mathematically. This feature enables the ANN to handle nonlinear systems well. However, as "a sophisticated form of regression" (Smith and Demetsky, 1994), the ANN training process is still based on the historical data without considering the correlation between travel times on consecutive intervals. Besides that, the long time required for training the ANN and lack of a standard procedure to find the optimum architecture (Meldrum, 1995) further restricted its applicability to this study.

A stationary time series can be decomposed into a deterministic series and a stochastic series. Furthermore, the stochastic series can be represented as an infinite moving average time series (Smith, etc., 2002). By considering the stochastic nature of traffic flow and available travel times in successive intervals, it is suggested to propose a time series approach for travel time prediction in this study. The correlation between travel times in consecutive intervals could be formulated in a time series model when the

travel time measurements collected during the past intervals are taken as states variable in a time series problem.

One type of time series model is the moving average model, which can generate the future travel time by averaging the travel times collected in previous intervals. Different weights are allocated to travel times collected in different intervals. Conventionally, it is assumed that the previous observations closer to the current interval should have a higher weight and play a more important role for predicting traffic data in the next interval. Another type of a time series model is the exponential smoothing model, which can apply both historical and real-time traffic data to predict future information. It can formulate the deterministic and stochastic components by adding different weights to the historical and real-time data. This model is anticipated to perform well under complicated traffic conditions by allocating the appropriate weights to historical and real-time traffic data.

CHAPTER 3

DEVELOPMENT OF BASE MODELS

The development of three base travel time prediction models, including the exponential smoothing model (ESM), the moving average model (MAM), and the Kalman filtering model (KFM) are discussed in this chapter. The properties, logic and features of the base models are explored, tested, and evaluated through a simulated transportation network, which will be discussed in Chapter 4.

3.1 System Configuration

The configuration for developing the three base prediction models is shown in Figure 3.1, which is comprised by four major components, including network modeling, experiment design, model development, and system evaluation. Based on the field data collected by a number of acoustic sensors and data stations, a time-varying traffic distribution can be derived and utilized as input to a microscopic traffic simulation model. The developed simulation model is calibrated and validated to ensure that the simulated travel time is adequately substantiated by the field data. Therefore, the simulated traffic information can be treated as actual traffic data for developing, testing and evaluating the base models. Detailed information about simulation network modeling is presented in Chapter 4.

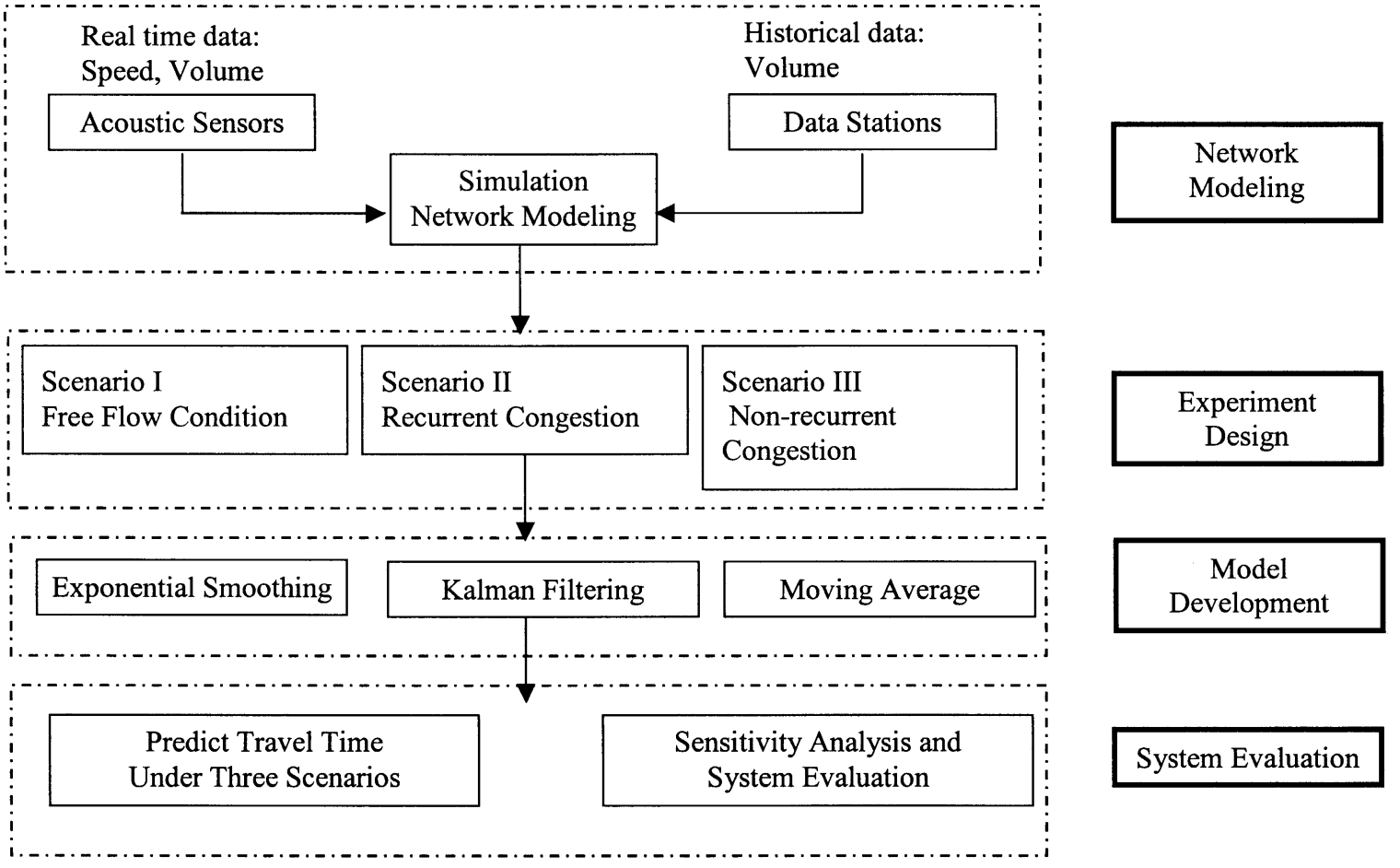

Figure 3.1 The Development of the Base Prediction Models.

3.2 Exponential Smoothing Model

The exponential smoothing model (ESM) has gained popularity for many years in predicting future transportation condition with significant work performed to study its theoretical property. Empirical studies conducted by Makridakis et al. (1982), as well as Gross and Craig, (1974) indicated that ESM could be the best choice for one-period-ahead forecasting. Thus, the exponential smoothing method could be a candidate approach utilized for short-term traffic prediction.

3.2.1 Single Exponential Smoothing Method

The proposed ESM discussed in this study is developed based on a single exponential smoothing method formulated in Equation 3.1:

$$S_t = \alpha x_{t-1} + (1-\alpha) S_{t-1} \tag{3.1}$$

where x_{t-1}: observation at time t-1,

S_t: predicted value at time t,

α: weight parameter of real time data, or smoothing constant, $0 \leq \alpha \leq 1$,

t: index of time interval, $t \geq 2$.

Based on Equation 3.1, the predicted value S_{t-1} can be derived as

$$S_{t-1} = \alpha x_{t-2} + (1-\alpha) S_{t-2} \tag{3.2}$$

By substituting S_{t-1} into Equation 3.1, S_t can be derived as

$$S_t = \alpha x_{t-1} + \alpha(1-\alpha) x_{t-2} + (1-\alpha)^2 S_{t-2} \tag{3.3}$$

Thus,

$$S_t = \alpha x_{t-1} + (1-\alpha)[\alpha x_{t-2} + (1-\alpha) S_{t-2}] \tag{3.4}$$

Similarly, by substituting S_{t-2}, S_{t-3}, and so forth until S_2 into Equation 3.4, S_t can be derived as

$$S_t = \alpha \sum_{i=1}^{t-2} (1-\alpha)^{i-1} x_{t-i} + (1-\alpha)^{t-2} S_2 \tag{3.5}$$

where $t \geq 2$ and $0 \leq \alpha \leq 1$. The weight $\alpha(1-\alpha)^{i-1}$ decreases exponentially as i increases, so does the influence of the real time observation x_{t-i} to S_t.

To demonstrate the impact of different weights on S_t, Equation 3.1 can be written as

$$S_t = S_{t-1} + \alpha[x_{t-1} - S_{t-1}] \quad \text{where } 0 \leq \alpha \leq 1 \tag{3.6}$$

From Equation 3.6, it is found that S_t is estimated by the sum of S_{t-1} and the product of the weight parameter α and the prediction error at time t-1 (e.g., $[x_{t-1} - S_{t-1}]$). An example is given to illustrate the principle of the single exponential smoothing method. By considering a data set consisting of 11 observations taken over time, the smoothed data with different weights (e.g., 0.1 and 0.5) are listed in Table 3.1 and shown in Figure 3.2.

Table 3.1 Example of Single Exponential Smoothing Method.

Time	X_t	$S_t(\alpha=0.1)$	$S_t(\alpha=0.5)$
1	71		
2	70	71	71
3	69	70.90	70.5
4	68	70.71	69.75
5	64	70.44	68.88
6	65	69.80	66.44
7	72	69.32	65.72
8	78	69.58	68.86
9	75	70.43	73.43
10	75	70.88	74.21
11	75	71.29	74.61

Note: X_t: original data observed at time t S_t: smoothed data at time t

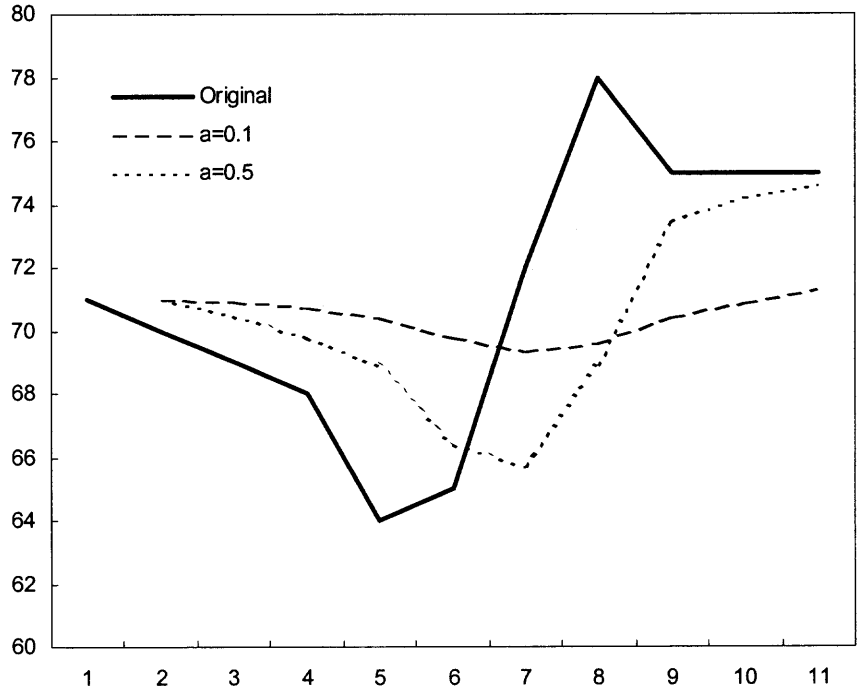

Figure 3.2 Graphical Example of Single Exponential Smoothing.

From Figure 3.2 indicates α dominates the closeness of the interpolated value S_t to the real time observation x_{t-1}. The relationship between the smoothing constant and prediction results can be summarized as follows:

- The speed of prediction adjustment to prediction error is determined by the smoothing constant α.
- The closer the value of α is to zero, the slower the forecast will be adjusted based on prediction errors.
- The closer the value of α is to 1.00, the greater impact will come from real time observation, while less influence will be made by the previous step smoothing data.
- Selecting a smoothing constant α that balances the benefits of real time observation and previous smoothing data can improve the prediction accuracy.

3.2.2 Double and Triple Exponential Smoothing Methods

Double exponential smoothing is another type of exponential smoothing method that adds a trend component to the single exponential smoothing concept. Double exponential smoothing is defined in Equations 3.7 and 3.8, which require two smoothing parameters α and γ.

$$S_t = \alpha x_{t-1} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1 \qquad (3.7)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \qquad 0 \le \gamma \le 1 \qquad (3.8)$$

S_t and x_t denote the smoothed estimate and observation, respectively, and b_t is the smoothed estimate of the trend of the time series at time *t*. In this approach, an observation X_{t-1} at t-1 is weighted with a previous estimate plus a trend factor, or estimated slope b_t. This slope is a smoothed estimate of the overall trend of the time series. By adding an estimate of the trend to the previous estimate of the value of a time series observation, double exponential smoothing is able to fit a time series that has a trend better than single exponential smoothing.

The third common type of exponential smoothing is triple exponential smoothing, which incorporates both a trend and a seasonal component. The seasonal component allows the smoothing estimate to better follow seasonal trends such as the oscillation of a wave or yearly traffic patterns. The basic equations for this method are defined as:

$$S_t = \alpha \frac{x_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1 \qquad \text{Overall Smoothing} \qquad (3.9)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \qquad 0 \le \gamma \le 1 \qquad \text{Trend Smoothing} \qquad (3.10)$$

$$I_t = \beta \frac{x_t}{S_t} + (1-\beta)I_{t-L} \qquad 0 \le \beta \le 1 \qquad \text{Seasonal Smoothing} \qquad (3.11)$$

where x_t is the observation,

S_t is the smoothed observation,

b is the trend factor,

I_t is the seasonal index,

t is an index denoting a time period,

L is the number of periods in each season,

Exponential smoothing can provide a good way to predict future events. The forecast of single exponential smoothing will always converge to a single value. Double exponential smoothing is suitable for prediction where trends exist. Likewise, triple exponential smoothing is appropriate to apply for forecast when the data show trend and seasonality.

3.2.3 ESM Development

It is anticipated that a good smoothing model will optimize the prediction accuracy with real-time traffic data under various traffic conditions. Thus, the proposed ESM in this study is developed based on the single exponential smoothing method in this study, where the travel time data x_{t-1} observed at time t-1 and historical travel time data $x_{t,h}$ at time t (e.g., the travel time of the same time interval observed the previous day or week) are fed into the single exponential smoothing method. Thus, the ESM can be formulated as Equation 3.12

$$x_{t,ESM} = \alpha x_{t,h} + (1-\alpha)x_{t-1} \qquad 0 \le \alpha \le 1 \qquad (3.12)$$

where α is the weight parameter of historical data at t,

$x_{t,ESM}$: Predicted travel time at t with ESM,

$x_{t,h}$: Historical travel time at t,

x_{t-1}: travel time observation at t-1.

In Equation 3.12, $x_{t,ESM}$ is computed as the weighted average of the observed travel time x_{t-1} at time t-1 and the historical data at $x_{t,h}$ at time t The value of α can significantly impact the prediction, especially when the difference between the observation and historical data is large. For example, x_{t-1} is completely ignored when α is equal to 1. On the contrary, $x_{t,h}$ is entirely ignored when α is equal to 0. The value of α between 0 and 1 will produce intermediate results, where a smaller value of α causes real time observations to have a stronger effects on the prediction, while a greater value of α puts more weight on historical observations. Therefore, optimizing the method of exponential smoothing becomes a matter of determining the best smoothing constant α.

3.3 Moving Average Model

Nau (2004) defined moving average as a series of arithmetic means of a series of data sets, in which short-term averaging has the effect of smoothing out the bump existing in the original data series. In this section, an equally weighted moving averaging model (MAM) is proposed and formulated in Equation 3.13

$$x_{t,MAM} = \frac{\sum_{i=1}^{N} x_{t-i}}{N} \tag{3.13}$$

where $x_{t,MAM}$: predicted information at time t with MAM,

N: total number of previous intervals' data considered by MAM,

i: the i^{th} previous interval data,

x_{t-i}: observed travel time at time t-i,

t: index of time interval.

Nau (2004) stated that the center of the simple moving average is located at period t-(N+1)/2. This implies that the estimate of the local mean tends to lag behind the true value of the local mean by about (N+1)/2 periods. When more previous interval data are used in MAM, the prediction curve of the moving average becomes smoother. The averaged data lags further behind the actual trend if data from more previous intervals are involved to affect the moving average value (Walker, 2004). Therefore, for short-term traffic prediction, N usually should be relatively small to decrease the time lag of the predicted data.

An example is given to illustrate the simple moving average method. Consider the following data set consisting of 11 observations taken over time, the moving average data are listed in Table 3.2 and shown in Figure 3.3.

Table 3.2 Original and Moving Average Value.

Time	X_t	X_{MAM} (N=2)
1	71	
2	70	
3	69	70.5
4	68	69.5
5	64	68.5
6	65	66
7	72	64.5
8	78	68.5
9	75	75
10	75	76.5
11	75	75

Note: x_t: original data at time t $x_{t,MAM}$: moving averaged data at time t

MAM sometimes can be treated as a special case of ESM when the smoothing constant α is equal to 0 and N=1. Under such a situation, the weight parameter for historical data is 0. Therefore, MAM only uses real-time data components for prediction instead of both historical and real-time components in ESM. Taft (2004) tested the

moving average and concluded that the MAM can perform well for traffic conditions that are very different from the historical trend. It will be appropriate to employ only real-time traffic data in MAM when the historical data may bring potential negative impacts to the predicting of future traffic.

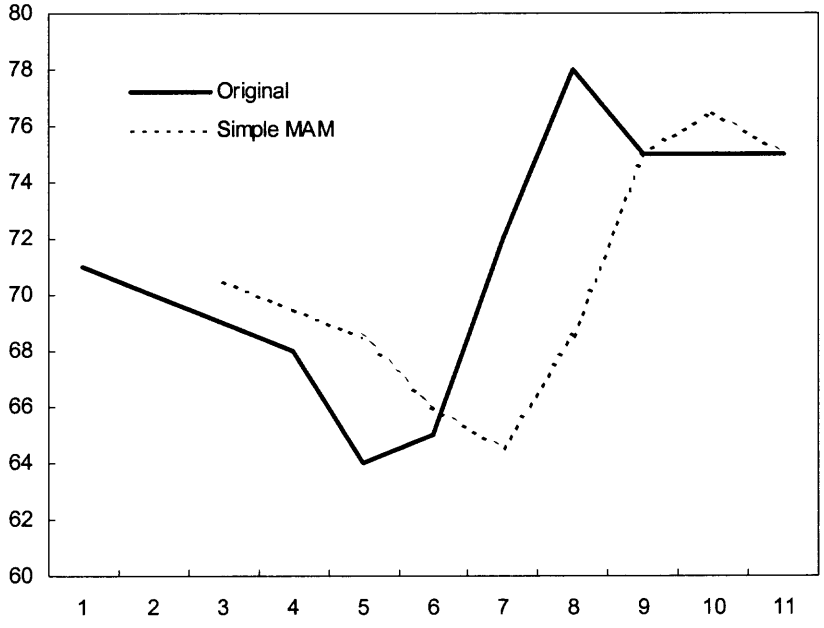

Figure 3.3 Example of a Moving Average Model.

3.4 Kalman Filtering Model

The Kalman Filtering Model (KFM) is a statistical time series modeling approach originating from state-space representations in linear control theory. The Kalman filtering model can employ both historical and real time data to predict future information based on the most recent prediction error.

Travel time may be affected by various factors, such as traffic volume, geometric condition, incidents, vehicle composition, and weather condition, whose relationship with

the factors is difficult to be captured. Travel time is thus treated as the state variable in the KFM. To develop a KFM for dynamic travel time prediction, the following assumptions are made:

(1) The state variable (e.g., travel time) to be predicted follows a linear relationship on successive time intervals based on Equation 3.14;

(2) The state variable is normally distributed in the prediction horizon.

Let x_t denote the state variable (e.g., travel time) to be predicted at time interval t, ϕ_t denote the transition parameter at time interval t which is externally determined to represent the linear relationship between successive state variables. ω_t denote a noise term of travel time that has a normal distribution with zero mean and a variance of Q_t, the covariance of travel time.

Let z_t denote the observation of travel time in time interval t, and v_t denote the measurement error at time t that has a normal distribution with zero mean and a variance of R_t, the covariance of the measurement variable. H is the measurement sensitivity, which represents the linear relationship between measurement variable z_t and state variable x_t.

Let P_t denote the covariance of the prediction error at time t. The notation (+) or (-) represents the priori or posteriori value for the studied variables, respectively. For example, P_t^- is the priori error covariance of the predicted travel time at time t, $\hat{x}_t^+$ denotes the posteriori (or updated) value of predicted travel time when the observation of measurement variable (e.g., travel time) is available at time t.

The KFM system model and measurement model can be formulated as

$$x_t = \phi_{t-1} x_{t-1} + \omega_{t-1} \tag{3.14}$$

$$z_t = H_t x_t + v_t \tag{3.15}$$

Assume that in a linear system $\forall$ i, j, $E[\omega_i * v_j]=0$, which means ω_i and v_j are uncorrelated. The derivation of equations in KFM is discussed in Appendix B. The basic steps of the computational procedure for KFM is shown below:

Step 0: Initialization

Set t=0 and let $E[x_0] = \hat{x}_0$ and $E[(x_0 - \hat{x}_0)^2] = P_0$.

Step 1: Extrapolation

State estimate extrapolation: $\hat{x}_t^- = \Phi_{t-1}\, \hat{x}_{t-1}^+$ ..

Error covariance extrapolation: $P_t^- = \Phi_{t-1}\, P_{t-1}^+\, \Phi_{t-1}^T + Q_{t-1}$

Step 2: Kalman Gain Calculation

$$K_t = P_{t,-}\left[P_{t,-} + R_t\right]^{-1}$$

Step 3: Update

State estimate update: $\hat{x}_t^+ = \hat{x}_t^- + \overline{K}_t [\, z_t - H_t\, \hat{x}_t^- \,]$

Error covariance update: $P_t^+ = [\mathrm{I} - \overline{K}_t\, H_t\,]\, P_t^-$

Step 4: Let t = t + 1 and go to Step 1 until the preset time period ends.

Measurement variable z_t could be obtained by averaging the travel times reported from probe vehicles or simulation results generated for time interval t. The state transition parameter ϕ_t represents the linear relationship between the state variable (travel times) of successive intervals in KFM. It is suggested to be the ratio of historical travel time data at successive time intervals. As for the historical travel time, it can be chosen as different measures (e.g., travel time in the previous time interval, travel time in the same time interval on the same day of the week before or average travel time in the same time

interval in previous week) depending on the traffic condition and available data. Since no traffic parameter other than travel time is involved in KFM, the measurement sensitivity H is set to be 1.

3.5 Summary

In this chapter, three base models (e.g., ESM, MAM, and KFM) are developed for short-term travel time prediction. ESM can employ both historical and real-time data for travel time prediction, which can predict the future traffic trend reasonably well when it is similar to the historical profile. Regarding MAM, it is an appealing approach to capture dynamic traffic information when non-recurrent congestion occurs. The KFM has been evaluated (e.g., Chien, etc, 2003) as a sound approach for prediction because it can dynamically improve the accuracy of prediction results based on the prediction error at the previous time interval. It optimizes prediction results and achieves the least square prediction error by assuming the state variable is normally distributed and has a linear relationship between successive intervals.

The three base models will be further tested with travel times from a real-world transportation network. Travel times under different traffic conditions will be incorporated into the developed base models to evaluate their performance for further analysis.

CHAPTER 4

CASE STUDY I

4.1 Background of the Studied Site

The site selected for testing and evaluating the developed base models for travel time prediction is a transportation network covering Routes 55, 42, 76 and 676 in southern New Jersey. The residents of southern New Jersey mainly use this corridor for destinations in the city of Philadelphia through the Walt Whitman and Ben Franklin Bridges. Congestion points are scattered over the studied roadways and at the toll plazas of both bridges during peak periods further increase the travel time variation.

It was known from historical data that the northbound Route 42 to the Walt Whitman Bridge (WWB) where it intersects the southbound Route 168 was congested during the morning peak period. Traffic conditions will be worse in the future due to the growing population in this region. Other congested locations in the study site during the morning peak were at the merging traffic streams from Routes 295 and 55 to Route 42 before entering the Ben Franklin Bridge (BFB) as shown in Figure 4.1. In addition, unexpected events such as incidents or construction maintenance activities within the study network are introduced by means of simulation.

4.2 Data Collection

Five acoustic sensors were installed in 2001 in the study network to collect real time traffic data that will be jointly used for calibrating the simulated model. The sensor locations were determined by considering the locations of potential congestion points and existing data stations. The links equipped with acoustic sensors are on Routes 76, 676, 42 and 295, where traffic volume and speed data were collected.

A list of sensor locations is shown in Table 4.1 and Figure 4.2:

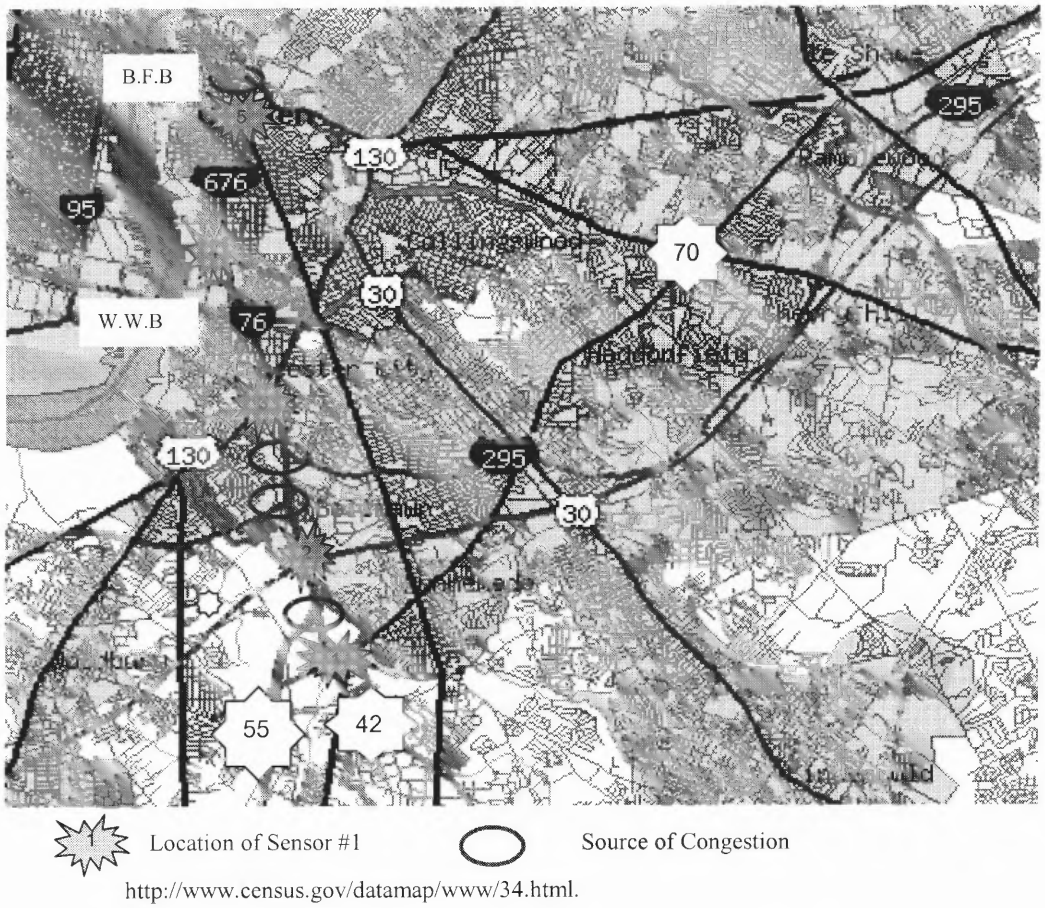

Figure 4.1 Map of the Study Site.

Table 4.1 Sensor Locations (On northbound of Routes 42, 76 and 676).

Sensor Number	Position	Mile Post
1	Northbound Route 42	13.5
2	Northbound I-295	26.92
3	Northbound I-76	2.5
4	Northbound I-676	1.0
5	Northbound I-676	4.0

Sensor #1 recorded the traffic at the starting point of the study network on northbound Rt. 42, which was located 50 feet before the ramp junction of Rt. 55 and Rt. 42, and is one of the congested places in the network. Sensor #2 collected data on northbound Rt. 76, where the traffic is fed by northbound Rt. 295. Sensor #3 monitored traffic entering the Toll Plaza on the Walt Whitman Bridge, where traffic from northbound Rt. 76 and westbound Rt. 130 merged. Sensor #4 recorded the traffic condition on northbound Rt. 676 between the entrances of the two bridges. Sensor #5 collected data on northbound Rt. 676, where traffic merged into the main stream to the Ben Franklin Bridge. Drivers destinated for Philadelphia can opt for the Walt Whitman Bridge if traffic on northbound Rt. 676 is jammed before the Ben Franklin Bridge. There are three different types of data collected by the acoustic sensors, namely vehicle counts, speed and occupancy. Due to power shortage for sensor operations, the frequency of data collection is set to 30 seconds at every 5 minutes and 15 minutes respectively for peak (6:00am to 9:00am and 3:00pm-6:00om) and off-peak (all of the other time) periods. An example of collected speed and volume data by an acoustic sensor on 10/26/20 is presented in Table 4.2.

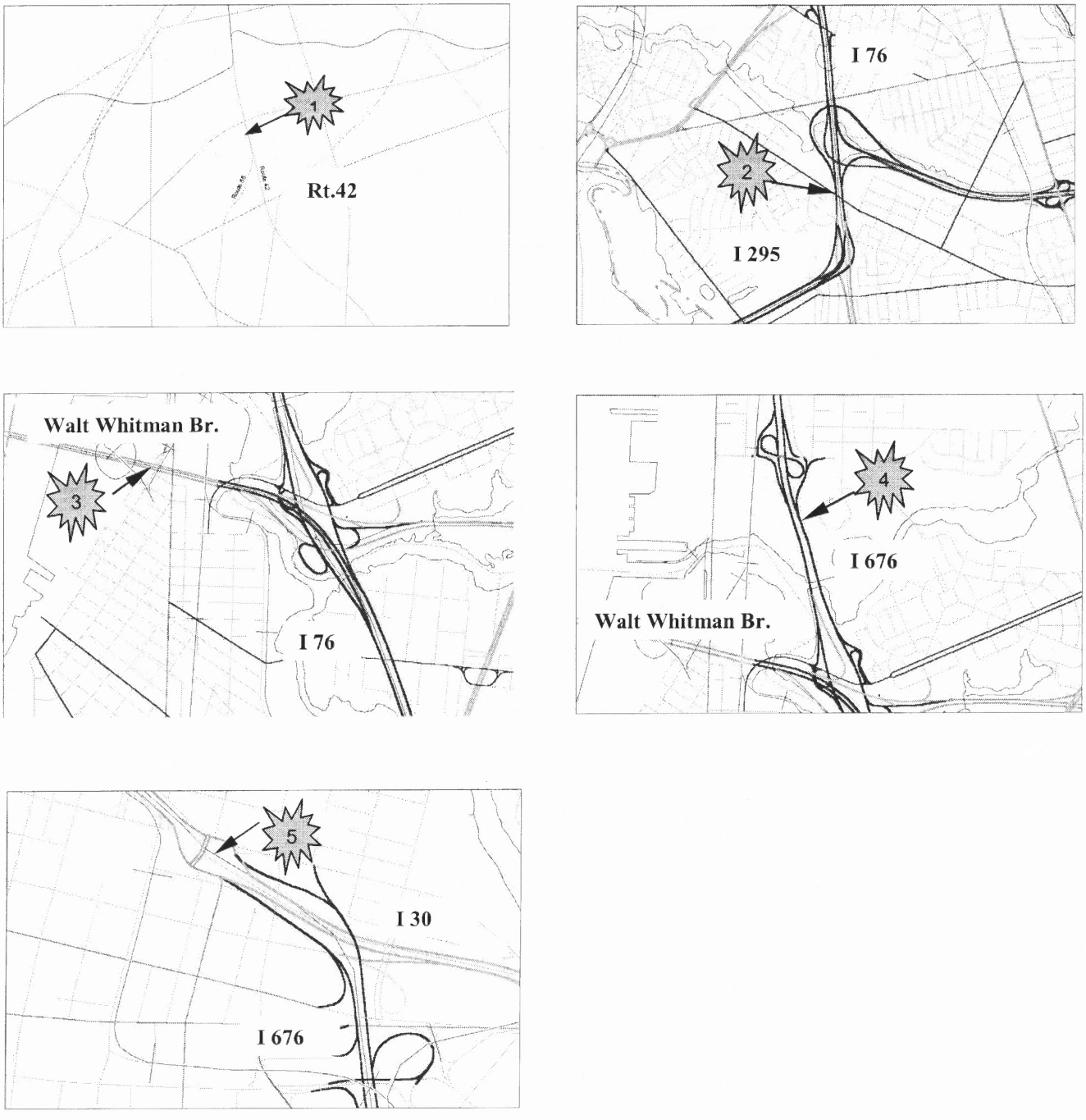

Figure 4.2 Sensor Locations.

Table 4.2 Sample Volume and Speed Data Collected by Acoustic Sensor 1.

Note: All traffic volumes are in vehicles per minute and speeds are in miles per hour, collection date is 10/26/2001.

Acoustic 001	Total Volume - Lane:					Speed - Lane:					Avg.	Acoustic 001	Total Volume - Lane:					Speed - Lane:					Avg.
Time	1	2	3	4	5	1	2	3	4	5	Speed	Time	1	2	3	4	5	1	2	3	4	5	Speed
6:00AM	2	4	10	2	4	59	46	59	49	32	49	7:55AM	0	6	6	6	4	--	44	40	45	31	40
6:05AM	0	14	6	6	6	--	47	61	54	30	48	8:00AM	0	16	12	14	0	--	48	35	37	--	40
6:10AM	0	18	16	8	6	--	44	52	57	39	48	8:05AM	0	14	10	10	6	--	39	46	37	30	38
6:15AM	2	12	12	6	6	43	46	51	47	53	48	8:10AM	0	8	6	14	6	--	47	38	36	31	38
6:20AM	0	14	22	8	6	--	47	72	51	46	54	8:15AM	0	16	12	4	8	--	46	34	40	32	38
6:25AM	0	26	12	18	6	--	46	66	46	58	54	8:20AM	0	14	20	12	4	--	57	57	40	42	49
6:30AM	0	18	8	14	10	--	58	63	44	51	54	8:25AM	2	12	14	14	4	59	47	51	43	45	49
6:35AM	0	16	16	14	12	--	49	37	50	32	42	8:30AM	0	14	14	4	18	--	57	45	50	44	49
6:40AM	0	12	10	14	2	--	45	48	39	36	42	8:35AM	0	18	8	10	4	--	41	41	43	47	43
6:45AM	2	12	12	20	4	52	44	43	41	30	42	8:40AM	0	10	8	10	4	--	45	39	49	39	43
6:50AM	2	20	14	16	4	47	43	38	36	36	40	8:45AM	0	10	18	18	8	--	48	38	44	42	43
6:55AM	0	18	10	12	12	--	37	37	38	48	40	8:50AM	0	6	18	14	6	--	42	32	45	37	39
7:00AM	0	12	8	12	4	--	46	39	30	45	40	8:55AM	0	10	4	6	2	--	58	43	36	55	48
7:05AM	2	20	16	12	6	60	50	51	44	40	49	9:00AM	0	8	10	14	2	--	59	51	43	39	48
7:10AM	0	12	12	10	6	--	51	62	44	39	49	9:15AM	2	10	12	10	2	59	57	59	53	12	48
7:15AM	0	20	12	8	2	--	63	54	36	43	49	9:30AM	0	12	8	6	4	--	57	44	28	51	45
7:20AM	0	18	14	18	6	--	67	53	49	39	52	9:45AM	0	10	10	16	2	--	49	62	42	51	51
7:25AM	0	18	12	8	6	--	58	53	57	40	52	10:00AM	2	4	6	8	0	46	48	40	42	--	44
7:30AM	0	20	14	10	10	--	64	48	60	36	52	10:15AM	0	16	10	14	6	--	56	53	45	46	50
7:35AM	0	16	6	10	10	--	55	52	49	32	47	10:30AM	0	6	12	16	2	--	56	45	43	48	48
7:40AM	0	14	8	16	2	--	58	45	40	45	47	10:45AM	0	18	8	8	0	--	40	59	51	--	50
7:45AM	0	4	4	8	8	--	50	58	43	37	47	11:00AM	0	12	10	14	0	--	46	52	46	--	48
7:50AM	0	18	10	12	0	--	42	33	45	--	40	11:15AM	2	2	0	8	0	53	37	--	39	--	43

4.3 Traffic Simulation Program

Computer simulation is an inexpensive approach to mimic traffic operations for all kinds of ITS applications. It would be cost-effective here to test and evaluate the performance of the developed base prediction models with simulated traffic operations in the studied transportation network.

A sound simulation model could generate all kinds of measures of effectiveness (MOEs) including speeds, volumes, travel times, occupancy, fuel consumption, pollutant emissions, etc. Collecting MOEs from simulation output can avoid the disruption to the traffic operations caused by field experiments. For example, it is impossible to design different traffic conditions such as an abrupt increase in traffic volume, incident-based congestion, and maintenance activities in a field study. In addition, the simulation can mimic the physical changes of the transportation facility required by some schemes, which are not acceptable for experimental purposes.

To develop a traffic simulation model for the study network, CORSIM was selected to generate travel time based on its capabilities. CORSIM can simulate traffic operations in a complex and large-scale roadway network, in which every vehicle is modeled as a distinct object and follows a stochastic lane-change, car-following, and gap acceptance logic. The movement of each vehicle, every traffic control device (e.g., traffic signals) and event are updated in every second to regenerate the status of vehicles. The stochastic factors such as driver behavior characteristics, vehicle characteristics, and traffic characteristics are considered to calculate the relationships among vehicles on the streets or links at every second. Therefore, CORSIM could generate various formats of

traffic data (e.g., real-time or historical, aggregated in different levels) for evaluating the developed prediction model.

4.4 Network Modeling

The study network is shown in Figure 4.3 and contains ten on-ramps and nine off-ramps. The fundamental elements in CORSIM network modeling are links and nodes. A node represents the start or end point of links. A link represents a directional freeway or ramp segment connected by two nodes. The node can be classified into entry/exit node, internal node, and dummy node. The points where traffic enters or exists of the network are defined as entry/exit nodes, while the nodes located inside the network are represented as internal nodes. Dummy nodes are another type of nodes placed on entry and exit links between the entry/exit node and the internal node for collection of traffic on the links. Therefore, the studied network was segmented along the mainline into different links by critical points such as interchanges, potential ramp meter locations, curvature/superelevation change, and entry or exit points. The developed simulation model contains 11 entry nodes, 10 exit nodes, 21 dummy nodes and 35 internal nodes connected by 11 entry links, 10 exit links and 54 internal links.

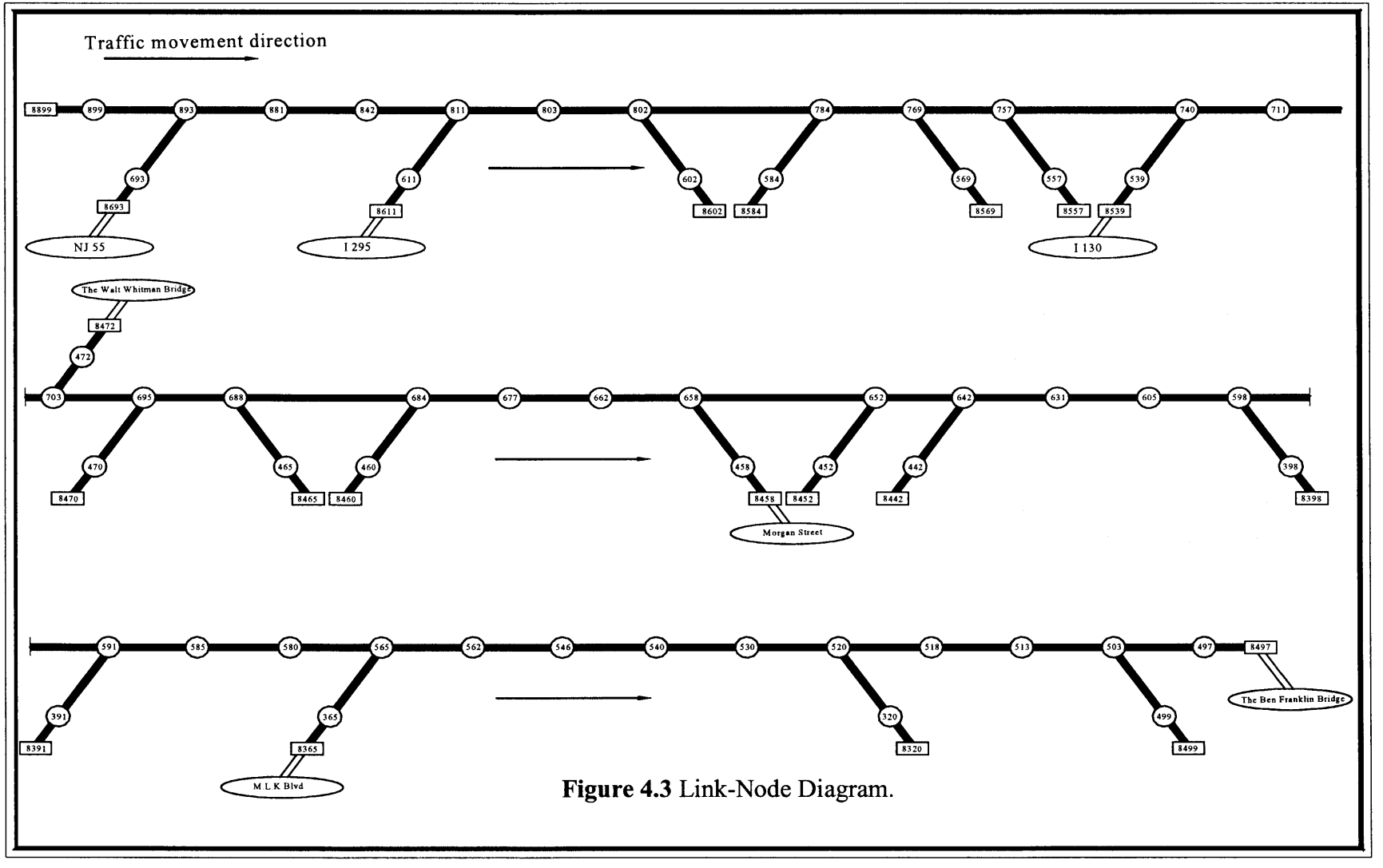

Figure 4.3 Link-Node Diagram.

4.4.1 Geometric Data

The geometric data were collected from construction plans provided by NJDOT, including the link length, number of lanes, radius of curvature, the grade percentages and super-elevation, etc. To get the geometric data, the straight-line diagram is an important source. Although the radius of the curvature is unavailable, names of the lanes, connecting ramps, Mile Post (MP), number of lanes, and traffic station ID, can be found.

Another source used to get geometric and traffic information was the NJDOT Geographic Information System (GIS) that contains roadway pavement information, which can fill the gap for accurate layout and related geometric information. In addition to that, the aerographic maps taken by satellite are available at http://terraserver-usa.com/. They reflect the real image of the study network in scale, and therefore provide the layout for identifying ramp junctions in the study area. The collected geometric data are summarized in Table 4.3.

Table 4.3 Geometric Characteristics of the Study Site.

				AUXILIARY LANE						
				ONE		TWO				
LINK NAME	TYPE	LNGTH (FT)	NO.THRU LANES	TYPE	LNGTH (FT)	TYPE	LNGTH (FT)	THRU DEST NODE	CURV RADIUS (FT)	GRADE
(8899,899	F	0	3	-	-	-	-	893	-	0
(899,893)	F	600	3	-	-	-	-	881	-	1
(893,881)	F	1200	3	A	300	-	-	842	-	0
(881,842)	F	3900	3	-	-	-	-	811	-	0
(842,811)	F	3100	3	-	-	-	-	803	-	0
(811,803)	F	800	3	W	800	-	-	802	-	0
(803,802)	F	500	4	D	400	-	-	784	-	0
(802,784)	F	1750	4	-	-	-	-	769	-	2
(784,769)	F	1500	5	W	1500	A	500	757	-	1
(769,757)	F	1200	5	D	700	D	400	740	-	-2
(757,740)	F	1700	5	-	-	-	-	711	-	0
(740,711)	F	3000	5	A	900	A	500	703	-	0
(711,703)	F	780	4	D	300	-	-	695	3396	0
(703,695)	F	850	3	-	-	-	-	688	-	-2
(695,688)	F	650	3	W	650	-	-	684	3459	0
(688,684)	F	400	3	-	-	-	-	677	6067	0
(684,677)	F	750	3	A	450	-	-	662	3428	0
(677,662)	F	1500	4	-	-	-	-	658	6067	0
(662,658)	F	400	4	D	350	-	-	652	3993	0
(658,652)	F	570	4	-	-	-	-	642	4000	0
(652,642)	F	960	4	A	400	-	-	631	-	0
(642,631)	F	1170	3	A	500	-	-	605	-	0
(631,605)	F	2600	3	-	-	-	-	598	-	0
(605,598)	F	550	3	D	400	-	-	591	-	0
(598,591)	F	700	3	-	-	-	-	585	700	0
(591,585)	F	600	3	A	300	-	-	580	1900	0
(585,580)	F	500	3	-	-	-	-	565	4000	0
(580,565)	F	1500	3		-	-	-	562	1920	0
(565,562)	F	300	3	A	250	-	-	546	1920	0
(562,546)	F	1600	3	-	-	-	-	540	-	0
(546,540)	F	600	3	-	-	-	-	530	4800	0
(540,530)	F	1000	3	-	-	-	-	520	-	0
(530,520)	F	1000	3	D	800	-	-	518	949	0
(520,518)	F	200	3	-	-	-	-	513	949	0
(518,513)	F	775	3	-	-	-	-	503	-	0
(513,503)	F	607	3	D	400	-	-	497	2200	0
(503,497)	F	918	3	-	-	-	-	8497	-	0
(497,8497	F	0	3	-	-	-	-	-	-	0

F: Freeway R: Ramp A: Acceleration lane D: Deceleration lane W: Full auxiliary lane

Table 4.3 Geometric Characteristics of the Study Site (continued).

				AUXILIARY LANE						
				ONE		TWO				
LINK NAME	TYPE	LNGTH (FT)	NO.THRU LANES	TYPE	LNGTH (FT)	TYPE	LNGTH (FT)	THRU DEST NODE	CURV RADIUS (FT)	GRADE
(497,8497	R	2289	1	A	-	-	-	471	2956	0
(703,472)	R	750	1	A	-	-	-	8471	-	0
(472,471)	R		1	A	300	-	-	695	-	0
(8470,470	R	600	1	A	-	-	-	688	700	0
(470,695)	R	200	1	A	-	-	-	8465	350	0
(688,465)	R		1	A	800	-	-	684	-	0
(8460,460	R	180	1	A	400	-	-	677	700	0
(460,684)	R		1	A	-	-	-	652	-	0
(8452,452	R	230	1	A	1500	A	500	642	2822	0
(452,652)	R		1	A	700	D	400	642	-	0
(8442,442	R	400	1	A	-	-		631	1578	0
(442,642)	R	300	1	A	900	A	500	8398	500	0
(598,398)	R	200	1	A	300	-	-	8458	-	0
(658,458)	R		1	A	-	-	-	591	450	0
(8391,391	R	300	1	A	650	-	-	585	1000	0
(391,591)	R	300	1	A	-	-	-	8602	-	0
(802,602)	R		1	D	450	-	-	784	2200	0
(8584,584	R	400	1	D	-	-	-	769	1100	0
(584,784)	R	300	1	A	350	-	-	8569	1000	0
(769,569)	R	300	1	D	-	-	-	8557	-	0
(757,557)	R		1	D	400	-	-	740	1100	0
(8539,539	R	460	1	D	500	-	-	711	-	0
(539,740)	R		1	A	-	-	-	565	5000	0
(8365,365	R	300	1	A	400	-	-	562	750	0
(365,565)	R	400	1	A	-	-	-	8320	-	0
(520,320)	R		1	A	300	-	-	893	1400	0
(8693,693	R	600	1	A	-	-	-	881	-	0
(693,893)	R		1	A	-	-	-	811	1300	0
(8611,611	R	600	1	A	250	-	-	803	350	0
(611,811)	R	400	1	A	-	-	-	8499	5000	0

F: Freeway R: Ramp A: Acceleration lane D: Deceleration lane W: Full auxiliary lane

4.4.2 Traffic Data

Five acoustic sensors were installed in the study network as shown in Figure 4.2 and introduced in Section 4.2, to collect traffic volumes and speeds at the designated locations. However, due to limited equipment, it was impossible to install enough acoustic sensors to cover the entire area. Thus, data stations are used to collect additional traffic counts. The average annual daily traffic (AADT) collected by data stations is the daily average of historical traffic observations. The acoustic sensor data are used as reference to normalize the AADT into a traffic volume distribution over time as shown in Figure 4.4. The seven data stations located in the studied region and collected AADT are shown in Table 4.4.

Table 4.4 Traffic Counts Lookup Results (Year 2000).

Station Number	Route Number	Milepost	Station Location	AADT
7-4-103	676	0.70	BETWEEN I-76 & MORGAN BLVD	69,252
7-9-355	676	2.50	JUST NORTH OF ATLANTIC AVE	61,047
7-4-104	676	2.95	AT HADDON AVE OVERPASS	58,065
7-5-001	76	0.50	JUST SOUTH OF MARKET ST	112,310
7-1-24	76	1.60	AT NICOLSON ROAD OVERPASS	136,310
7-2-11	76	2.40	WALT WHITMAN BR, TOLL	99,330
7-4-303	42	12.20	BETWEEN RT 544 & NJ 55	97,184

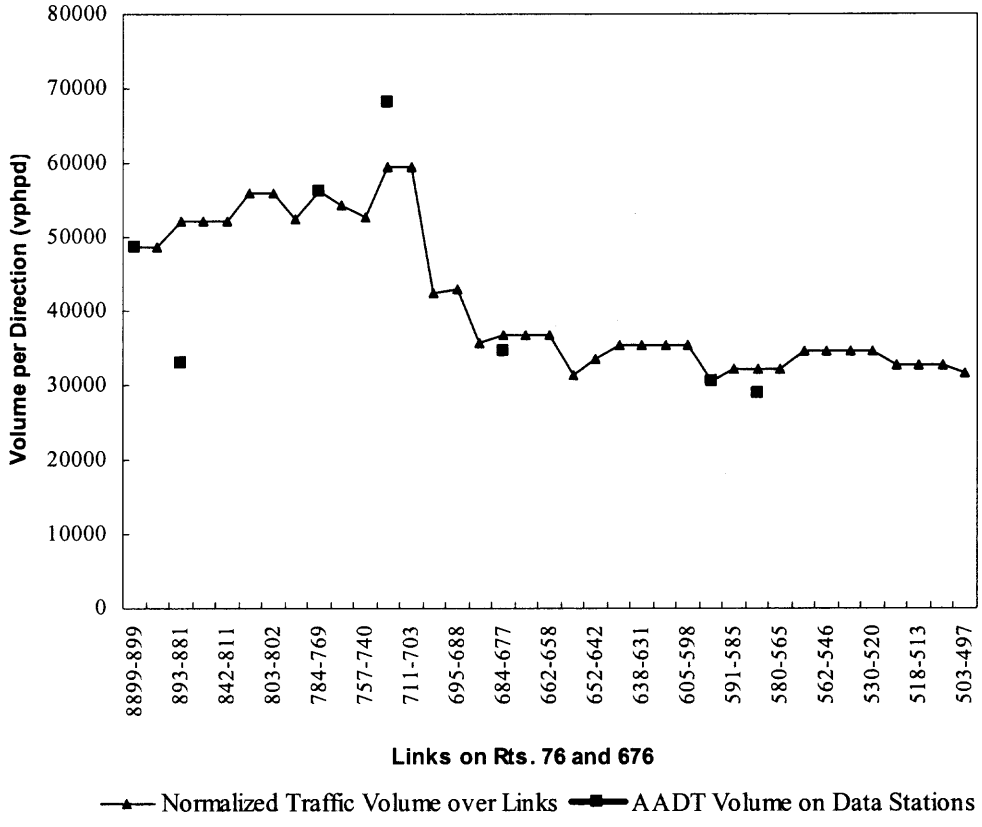

Figure 4.4 AADT Volumes.

The traffic volume distribution over links along the study network can be derived from the AADT, while the time-varying information is still missing. The traffic counts collected by the acoustic sensors are used for estimating an hourly traffic distribution for the purpose of simulation network modeling. The hourly traffic volume distribution, for example at Sensor #1, is shown in Figure 4.5. The free-flow speed in the simulation model is assumed 10 mph above the corresponding speed limit marked in the straight-line diagram.

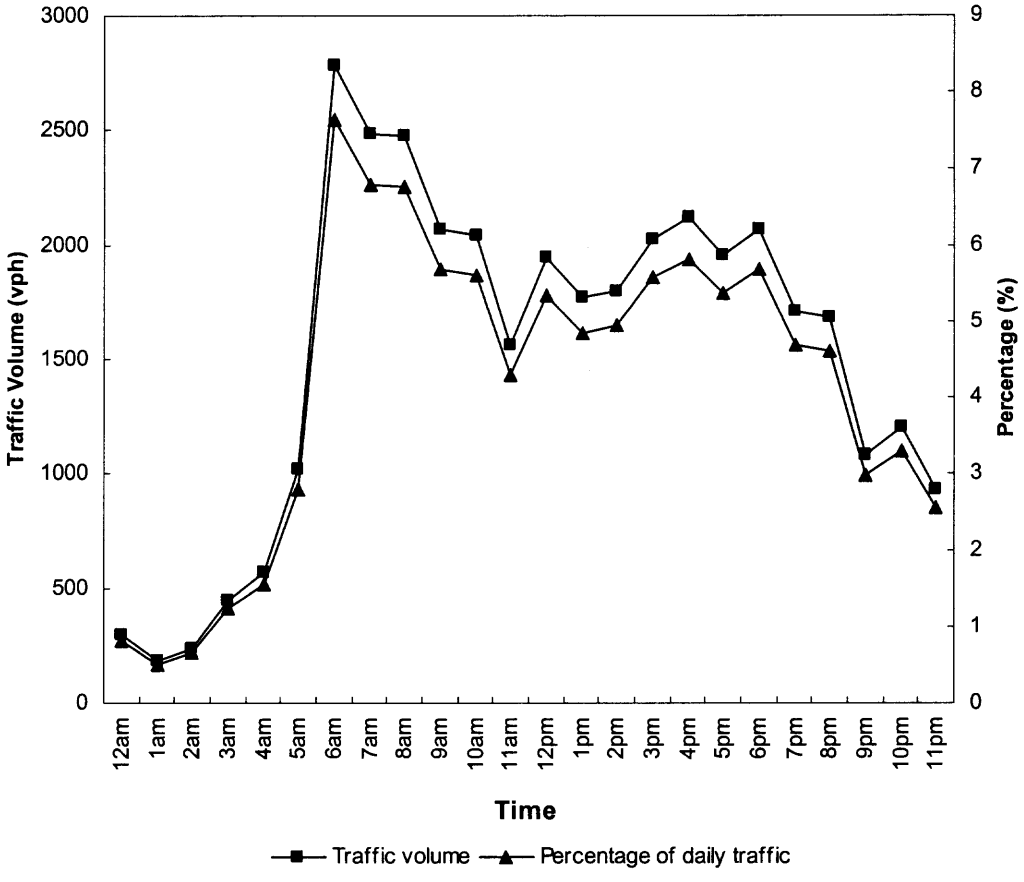

Figure 4.5 Traffic Distributions over Time at Sensor # 1.

4.5 Network Calibration and Validation

To generate credible simulation results testing the developed base prediction models, the developed simulation model should be proven to represent real world traffic operations reasonably well. Therefore, an analysis of simulation results considering comparable parameters, calibration and validation of the simulation model should be performed.

4.5.1 Model Calibration

With the input data of freeway geometry, traffic conditions (volumes, speeds and turning movement) and facility locations (e.g., locations of ramps and warning signs), 18-hours of traffic operations from 5:00 am to 11:00 pm on northbound Routes. 42, 76 and 676 were simulated. Two types of traffic parameters including traffic volumes at data stations and traffic speeds at the locations of acoustic sensors were selected as the reference to calibrate the simulated model. Since the acoustic sensor collected spot speed data at 5-minute interval only in two time periods (e.g., 6:00 am - 9:00 am and 3:00 pm to 6:00 pm), the speed data in time period 6:00 am to 9:00 am were selected to conduct the model calibration.

The discrepancies between simulated data and their field counterparts were identified. These discrepancies can be reduced by calibrating parameters in CORSIM, such as the car-following sensitivity factor (defined by headway between vehicles), lane change parameter, minimum separation for vehicle generation, the collision avoidance time period and the percentage of cooperative drivers. These parameters were adjusted to fine tune the simulated driving behavior. Decreasing the values of the car-following sensitivity factor could achieve larger volumes and higher speeds due to shorter gaps between vehicles. With more cooperative drivers, traffic volumes on ramps and their speeds increase because more vehicles are able to merge into the traffic stream on the mainline.

The impact of a particular parameter cannot be identified if several parameters are changed simultaneously in the simulation model. Thus, the impact analysis of each

parameter to the simulation results was conducted by varying a particular parameter while fixing others. The calibrated parameters are shown in Tables 4.5 and 4.6.

The simulated traffic volumes were compared with field data collected by the data stations. The results showed that the difference between simulated and actual data in the time period from 6:00 am to 8:55 am are less than 10% as shown in Table 4.7.

Table 4.5 Car-Following Sensitivity Factor (hundredth of a second).

Driver Type	1	2	3	4	5	6	7	8	9	10
Default Value	150	140	130	120	110	100	90	80	70	60
Calibrated Value	120	110	100	90	80	70	60	50	40	30

Table 4.6 Other Calibrated Parameters.

Parameter	Default Value	Calibrated Value
Pitt car following constant	10	5
Time to complete a lane-change maneuver	20	10
% of drivers desiring to yield right-of-way	20	30
Multiplier for desire to make a discretionary lane change	5	8

Table 4.7 Results of Calibrated Simulation and Actual Data (Traffic Volume).

DataStation	6:00 am to 6:55 am			7:00 am to 7:55 am			8:00 am to 8:55 am			Average
Locations	Actual	Simulated	Error %	Actual	Simulated	Error %	Actual	Simulated	Error %	Error %
8899-899	3887	3888	0.024	3401	3400	0.035	3401	3401	0.006	0.022
893-881	4180	4171	0.218	3658	3657	0.016	3658	3660	0.066	0.100
784-769	4492	4507	0.326	3931	4102	4.355	3931	4054	3.134	2.605
740-711	4767	4794	0.574	4171	4391	5.279	4171	4374	4.872	3.575
684-677	2938	3105	5.682	2571	2865	11.444	2571	2775	7.943	8.356
598-591	2442	2532	3.702	2136	2388	11.776	2136	2291	7.236	7.572
585-580	2579	2618	1.522	2256	2482	9.998	2256	2382	5.566	5.695

To compare the traffic speeds gathered by the acoustic sensor and generated by the simulation model, the mean and 95% confidence intervals were generated for the means of each data source as shown in Figure 4.8. It can be stated that the simulation readings are sufficiently close to the actual sensor readings based on a study conducted by a Rutgers University research team (Ozbay, etc., 2004). The reasons causing the bigger deviation on sensor 4 were attributed to (1) Aggregation of sensor data, (2) Normality assumption, (3) Ground truth data.

Table 4.8 95% Confidence Interval of Traffic Speeds (MPH).

Sensor	Simulated	Sensor Data
1	[44.32,45.32]	[41.30,50.74]
2	[47.90,49.56]	[45.74,61.54]
3	[46.21,54.59]	[51.15,60.11]
4	[49.15,67.45]	[49.60,72.14]
5	[54.97,59.31]	[55.56,66.18]

The mean absolute relative error (MARE) (Equation 4.1) is selected as another index to evaluate calibrated results.

$$MARE = \frac{1}{N}\sum_{t}\frac{\left|V_{actual}(t) - V_{simulated}(t)\right|}{V_{actual}(t)} \tag{4.1}$$

Note that N represents the sample size, while $V_{actual}(t)$ and $V_{simulated}(t)$ represent the observed and simulated travel speeds, respectively. With calibrated parameters, the MAREs during the period from 6:00 am to 9:00 am are 7.78%, 13.75%, 10.67%, 10.71%, and 10.46% at the locations of sensors 1 to 5, respectively. The results imply that the calibrated simulation model can reasonably well replicate traffic operations in the study network.

4.5.2 Model Validation

The simulation model is validated to ensure that the simulated results are reliable and able to mimic traffic operation in the study network. Graphical, aggregate, and statistical comparisons were considered for the simulation model validation.

A graphical comparison deals with data without numerical summaries or tabular representation. It is not suitable for this study because it is a subjective validation approach that cannot generate a quantitative index for comparison. An aggregate comparison could provide a general indication in terms of mean and standard deviation, but cannot present accurate details about how variables perform over time, how patterns changed, or how variables deviate over time. Since this study requires the validated data at a very detailed level (e.g., in 5-minutes time intervals), both graphic and aggregate approaches are not appropriate. Thus, a statistical comparison is performed because it can qualify the differences between actual and simulated data.

It should be noted that the data set used for model validation should be different from that used for model calibration. In this study, the traffic volumes collected at the data station and speeds from the acoustic sensor locations collected during a different time period (from 3:00 pm to 5:55 pm) were used to validate the model.

The difference between the simulated and actual traffic volumes on the data stations are listed in Table 4.9. The results show that at all data stations, the comparison difference is less than 10%, which validates the effectiveness of the simulation model. The statistical analysis was conducted by calculating, the MAPE for the field and simulated speeds. The MAPEs were 10.92%, 14.72%, 8.33%, 15.42%, and 9.51% at the locations of Sensors 1 to 5, respectively. The validation results indicate that the

developed simulation model could reasonably emulate traffic operations in different periods.

Table 4.9 Comparisons between Actual and Simulated Traffic Volumes.

DataStation	3:00 pm to 3:55 pm			4:00 pm to 4:55 pm			5:00 pm to 5:55 pm			Average
Locations	Actual	Simulated	Error %	Actual	Simulated	Error %	Actual	Simulated	Error %	Error %
8899-899	2429	2429	0.018	2915	2915	0.011	2429	2430	0.024	0.017
893-881	2613	2613	0.016	3135	3133	0.067	2613	2614	0.055	0.046
784-769	2808	2875	2.396	3369	3463	2.782	2808	2924	4.142	3.107
740-711	2979	3080	3.385	3575	3702	3.553	2979	3129	5.030	3.990
684-677	1836	1970	7.282	2204	2374	7.735	1836	1965	7.009	7.342
598-591	1526	1650	8.125	1831	1966	7.361	1526	1606	5.242	6.909
585-580	1612	1713	6.284	1934	2033	5.115	1612	1680	4.237	5.212

4.6 Design Scenarios for Testing Prediction Models

It is known from the literature review that a prediction model has different performance under different traffic conditions. The reason for a superior performance under a particular traffic condition is because the prediction model can better represent the underlying processes of traffic operations during that specific time period. In addition, the literature review revealed that current traffic prediction models had not achieved satisfactory performance due to their inability to cope with predictions under recurrent and non-recurrent congested conditions (Alecsandru, and Ishak, 2004). To develop a sound prediction model suitable for various degrees of congestion, different sets of traffic conditions should be designed as a test bed for evaluating the developed base prediction models.

Three scenarios are designed to represent different traffic conditions in the simulation model. Scenario 1 is designed for simulating free-flow condition, while Scenarios 2 and 3 are designed for simulating recurrent and non-recurrent congested

conditions, respectively. Scenario 1 is developed based on the field traffic data collected in the study network, while Scenarios 2 and 3 are hypothetical cases. A discussion on the development of the three scenarios follows.

After calibration and validation with data collected by data stations and acoustic sensors, the simulation model can replicate traffic operations in the study network as scenario 1. In Scenario 1, traffic is in free-flow condition and no congestion occurs. Ten simulations were made with different random number seeds. The resulting travel times can be used to represent the historical free-flow traffic condition.

Another hypothetical case is simulated to represent recurrent congestion. The traffic distribution of Scenario 2 increases the proportion of traffic in the peak hour dramatically and is shown in Figure 4.6. The higher traffic volume will cause serious congestion with abruptly increased travel times during the peak hour. The recurrent congestion in Scenario 2 is caused because traffic demand exceeds the capacity of the highway periodically. Another ten simulations were preformed with the Scenario 2 model using with different random number seeds. The resulting travel time can be used to represent recurrent congested conditions.

Scenario 3 was designed for simulating a non-recurrent congested condition and was constructed by adding an accident in the traffic environment of Scenario 1. The accident was set on Link 740-711 with 1-hour duration from 7:00 am to 8:00 am. There are total 4 lanes on the link where the accident occurs. Two lanes are blocked and one lane has its capacity reduced to half during the accident. Under such a scenario, the traffic congestion is caused by insufficient roadway capacity, instead of traffic demand. Since

the accident was set randomly based on Scenario 1, the historical traffic data are the same as in Scenario 1.

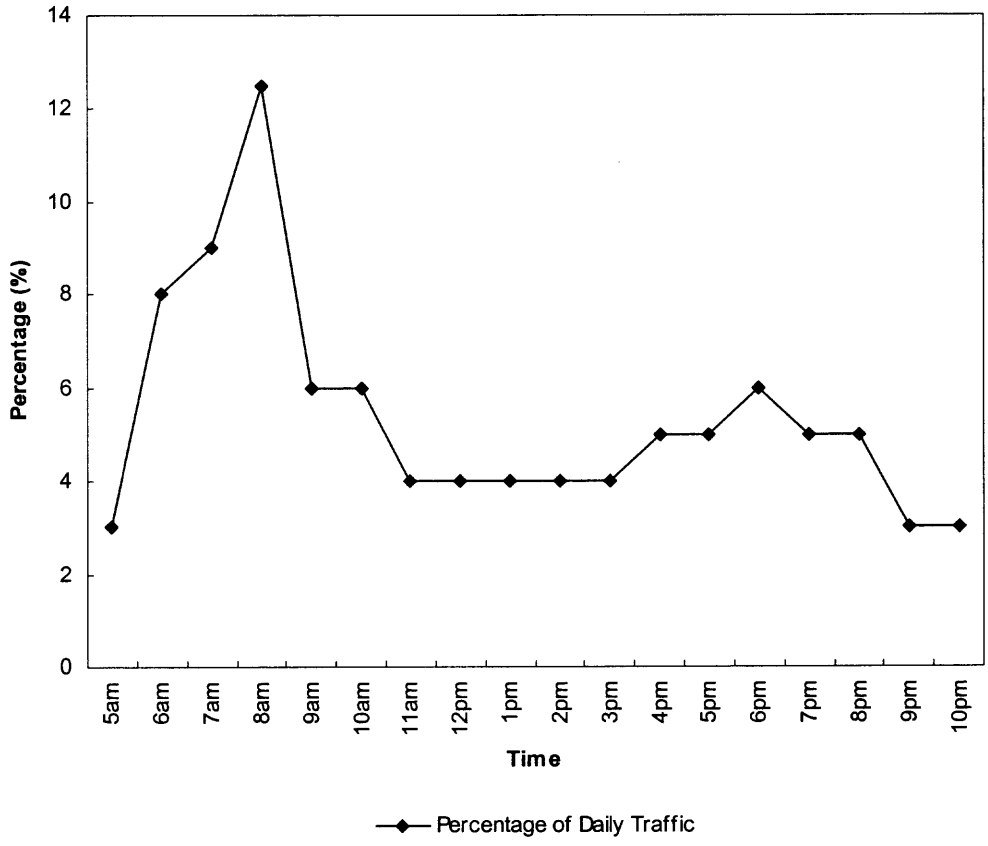

Figure 4.6 Traffic Distributions for Scenario 2.

The travel time for a particular OD under various traffic scenarios can be derived from the corresponding CORSIM simulation output. The path travel time is calculated based on the summation of the travel times on each link connecting the studied OD path. The travel time in the statistics report of CORSIM's output file is cumulative travel time averaging from the beginning of the simulation to the current time. To calculate the actual travel time for each time interval, a Java program is coded to manipulate the

CORSIM output. The data flow diagram and original code for this is contained in Appendix A.

4.7 Evaluation of Base Prediction Models

The performance of the base prediction models (e.g., ESM, MAM, and KFM) under various traffic scenarios is assessed in this Section. The model performances are compared and the properties, logic and application scope of each model are analyzed based on the comparison results.

4.7.1 Statistical Indices

The prediction performances are evaluated using the statistical indices including Mean Absolute Relative Error (MARE), Variance of Absolute Percentage Error (VAPE) and Maximum Relative Error (MRE). The equations of the indices are stated in Equations 4.2, 4.3 and 4.4, respectively.

$$MARE = \frac{1}{N}\sum_{t}\frac{|x(t)-\hat{x}(t)|}{x(t)} \tag{4.2}$$

$$VAPE = Var\left(\frac{|x(t)-\hat{x}(t)|}{x(t)}\right) = \sqrt{\frac{N\sum_{t=1}^{N}\left(\frac{|x(t)-\hat{x}(t)|}{x(t)}\right)^2 - \left(\sum_{t=1}^{N}\frac{|x(t)-\hat{x}(t)|}{x(t)}\right)^2}{N(N-1)}} \tag{4.3}$$

$$MRE = \max_{t}\frac{|x(t)-\hat{x}(t)|}{x(t)} \tag{4.4}$$

Note that N is sample size, while x(t) and $\hat{x}(t)$ represent the actual and simulated travel time, respectively.

The MARE calculates the average relative error between the predicted and actual travel time in the whole prediction period, and can indicate the accuracy of the prediction model. The VAPE calculates the deviations from the average travel time during the prediction horizon in all time intervals, and can reflect the stability of the prediction model. The MRE calculates the maximum value for the absolute percentage error, or the maximum prediction deviation on a particular time interval, and can help to assess how fast the prediction model can catch up with dynamic traffic changes.

4.7.2 Prediction Results

To obtain the best prediction performance of each base model under various traffic scenarios, different parameters were tested to determine their optimal values. The smoothing constant α adopted values from 0.1 to 0.9 in increments of 0.1 in ESM, while the N moving periods considered in MAM ranged from 2 to 4. The noise covariance (R_t, Q_t) in KFM is set with different values in KFM. The average travel time simulated in the same time period by 10 simulations with different random number seeds are used as the historical data for all scenarios.

The evaluation of prediction performance is made with three statistical indices (i.e. MARE, VAPE and MRE). The combinations of different parameter setting is implemented in the three base models to test their performance. The MARE, VAPE, and MRE are shown in Tables 4.10, 4.11 and 4.12, respectively. The prediction performance of the base models with optimal parameters is shown in Table 4.13.

For Scenario 1, the MARE in Table 4.11 increased from 3.20% to 4.08% when the weight α of historical data increased from 0.1 to 0.9 in ESM. It indicates that the

weight of historical data can affect the performance of ESM. Similarly, different combinations of (R, Q) setting in KFM shown in Table 4.11 lead to different prediction performance, with MARE varying between 3.30% and 4.45%. However, the numbers of moving periods N for the MAM have a slight impact on the prediction performance. Finally based on the best parameter settings highlighted in Table 4.11 for each model, the prediction results for Scenario 1 are shown in Figure 4.7.

For Scenario 2, the MARE had more variation when different α values were tested in the ESM. When α changed from 0.1 to 0.9, the MARE varied from 3.94% to 4.89%. The noise covariance (R, Q) in KFM affected the prediction performance potentially. The MARE varied from 3.30% to 4.45% with different settings of combination of (R, Q). Similarly, the numbers of moving periods N for the MAM have a slight impact on prediction performance. The best parameter settings highlighted in Table 4.11 were used to generate the prediction results as shown in Figure 4.8.

For Scenario 3, the prediction results were less sensitive to the values of α in the ESM. When the values of α varied from 0.1 to 0.9, the MARE in ESM changed from 5.29% to 5.84%. However, different combinations of (R, Q) settings played a greater role in affecting the prediction accuracy under this traffic condition. For example, the MARE in KFM varied from 6.09% to 10.46% when different combinations of (R, Q) settings were used as shown in Table 4.11. The numbers of moving periods N for the MAM have a slight impact on prediction performance. Using the best parameter settings identified by the comparisons, the prediction results are shown in Figure 4.9.

Table 4.10 (a) MARE Analysis for ESM, MAM (%).

SCENARIOS	MARE (ESM)									MARE (MAM)		
	α= 0.9	α=0.8	α=0.7	α=0.6	α=0.5	α=0.4	α=0.3	α=0.2	α=0.1	N=2	N=3	N=4
1	**1.31**	1.34	1.38	1.43	1.47	1.53	1.58	1.63	1.68	1.60	1.46	**1.44**
2	**3.19**	3.34	3.54	3.78	4.05	4.35	4.67	5.00	5.33	**5.07**	5.31	5.74
3	3.32	3.16	3.05	2.95	2.87	2.79	2.72	2.68	**2.66**	2.49	**2.36**	2.43

Table 4.10 (b) MARE Analysis for KFM (%).

KFM	MARE (SCENARIO 1)				MARE (SCENARIO 2)				MARE (SCENARIO 3)			
(R,Q)	0.1	1	10	100	0.1	1	10	100	0.1	1	10	100
5	1.37	1.49	1.74	1.93	4.15	4.36	4.96	5.61	3.35	2.80	2.91	3.13
50	1.32	1.39	1.47	1.75	**4.06**	4.19	4.35	4.98	3.94	3.37	2.79	2.94
500	1.30	1.33	1.38	1.49	4.07	4.07	4.15	4.36	3.87	3.93	3.35	**2.78**
5000	**1.29**	1.31	1.30	1.37	4.07	4.08	4.07	4.19	4.95	3.88	3.97	3.33

Table 4.11 (a) VAPE Analysis for ESM, MAM (%).

SCENARIOS	VAPE (ESM)									VAPE (MAM)		
	α= 0.9	α=0.8	α=0.7	α=0.6	α=0.5	α=0.4	α=0.3	α=0.2	α=0.1	N=2	N=3	N=4
1	**2.06**	2.17	2.34	2.54	2.80	3.10	3.46	3.88	4.35	3.38	2.65	**2.55**
2	**8.04**	8.61	9.48	10.55	11.93	13.52	15.43	17.62	20.18	**26.29**	40.77	60.08
3	20.96	17.28	14.06	11.46	9.57	8.39	**7.86**	7.99	8.74	**8.38**	8.55	9.38

Table 4.11 (b) VAPE Analysis for KFM (%).

KFM	VAPE (SCENARIO 1)				VAPE (SCENARIO 2)				VAPE (SCENARIO 3)			
(R,Q)	0.1	1	10	100	0.1	1	10	100	0.1	1	10	100
5	2.10	2.57	3.58	4.43	14.26	14.37	19.43	24.78	13.06	8.85	**8.34**	9.42
50	1.95	2.12	2.54	3.60	14.78	14.29	14.40	19.40	19.46	13.05	8.84	8.35
500	1.90	1.95	2.10	2.57	14.31	14.79	14.26	14.37	29.57	19.45	13.06	8.87
5000	**1.87**	1.91	1.96	2.13	**14.19**	14.32	14.78	14.28	29.56	29.57	19.49	13.03

Table 4.12 (a) MRE Analysis for ESM, MAM (%).

SCENARIOS	MRE (ESM)									MRE (MAM)		
	α= 0.9	α=0.8	α=0.7	α=0.6	α=0.5	α=0.4	α=0.3	α=0.2	α=0.1	N=2	N=3	N=4
1	**7.58**	7.58	7.58	7.85	8.16	8.48	8.79	9.11	9.42	9.66	8.62	**8.10**
2	**12.90**	13.44	14.03	14.63	15.23	15.82	16.60	18.09	19.99	**30.46**	40.35	53.18
3	23.08	20.91	18.74	16.90	15.58	**14.25**	16.07	19.10	22.12	22.91	21.18	**19.77**

Table 4.12 (b) MRE Analysis for KFM (%).

KFM	MRE (SCENARIO 1)				MRE (SCENARIO 2)				MRE (SCENARIO 3)			
(R,Q)	0.1	1	10	100	0.1	1	10	100	0.1	1	10	100
5	7.23	8.70	10.17	10.63	17.75	**15.77**	19.65	22.66	20.96	**20.20**	22.79	24.45
50	6.76	7.26	8.67	10.19	19.36	17.78	15.97	19.63	23.49	20.97	20.24	22.76
500	6.66	6.79	7.23	8.70	20.40	19.40	17.71	15.99	26.13	23.48	20.96	20.18
5000	**6.62**	6.66	6.76	7.28	20.81	20.41	19.35	17.77	27.74	26.18	23.50	20.90

Table 4.13 Overall Performance of Base Models under Various Scenarios (%).

SCENARIOS	1	2	3	1	2	3	1	2	3
	ESM			MAM			KFM		
MAPE	1.31	3.19	2.66	1.44	5.07	2.36	1.29	4.06	2.78
VAPE	2.06	8.04	8.74	2.55	26.29	8.55	1.87	14.78	8.87
MRE	7.58	12.90	22.12	8.10	30.46	21.18	6.62	19.36	20.18

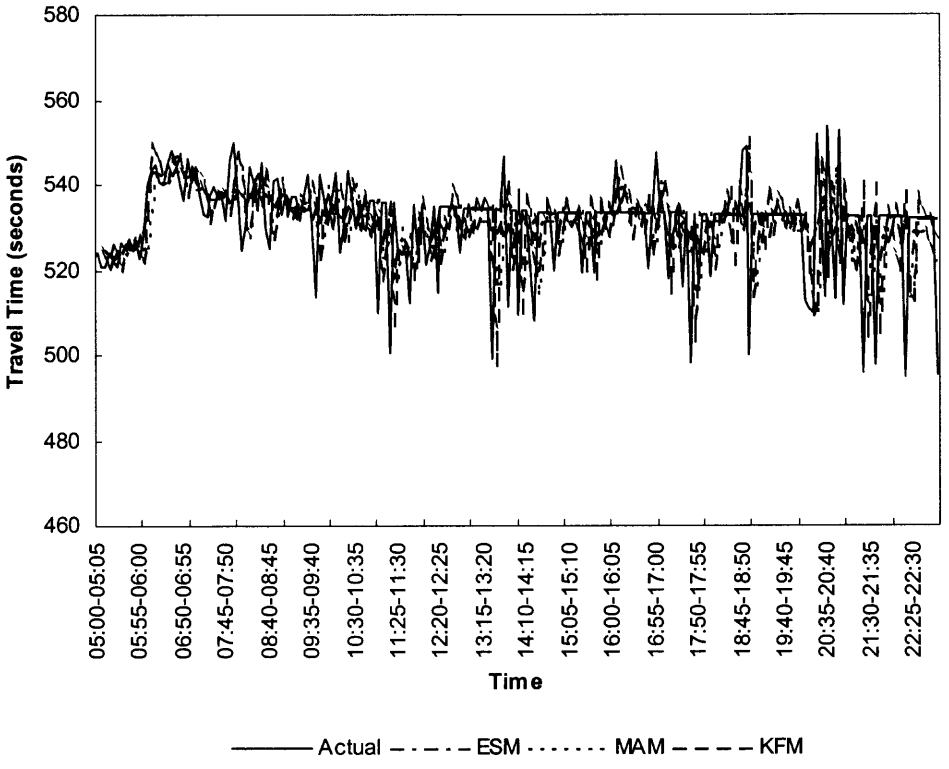

Figure 4.7 Predictions under Free Flow Conditions.

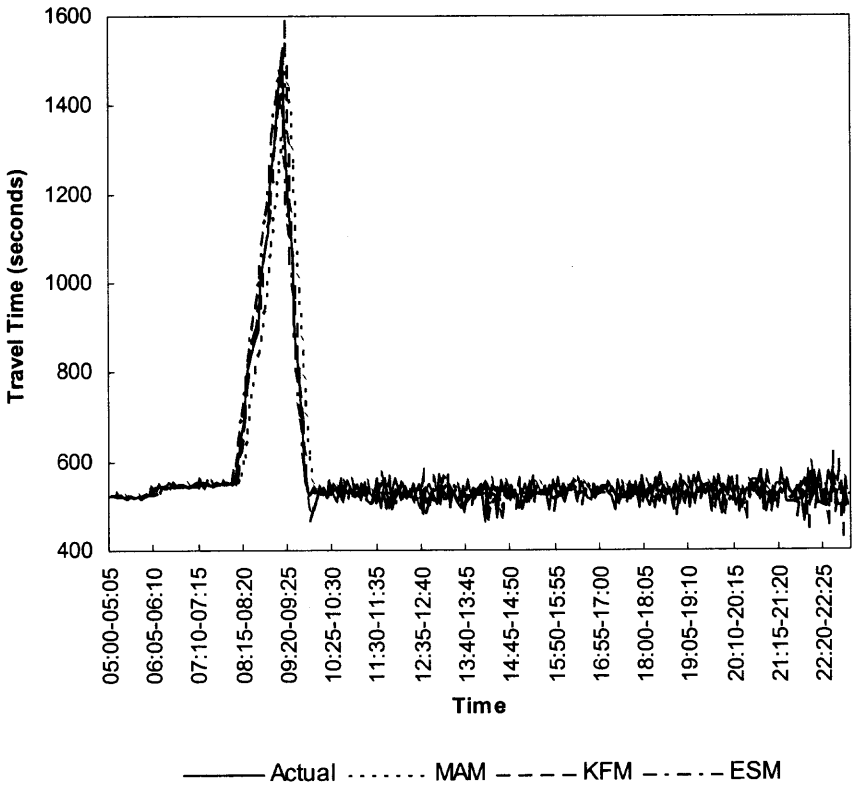

Figure 4.8 Predictions under Recurrent Congested Conditions.

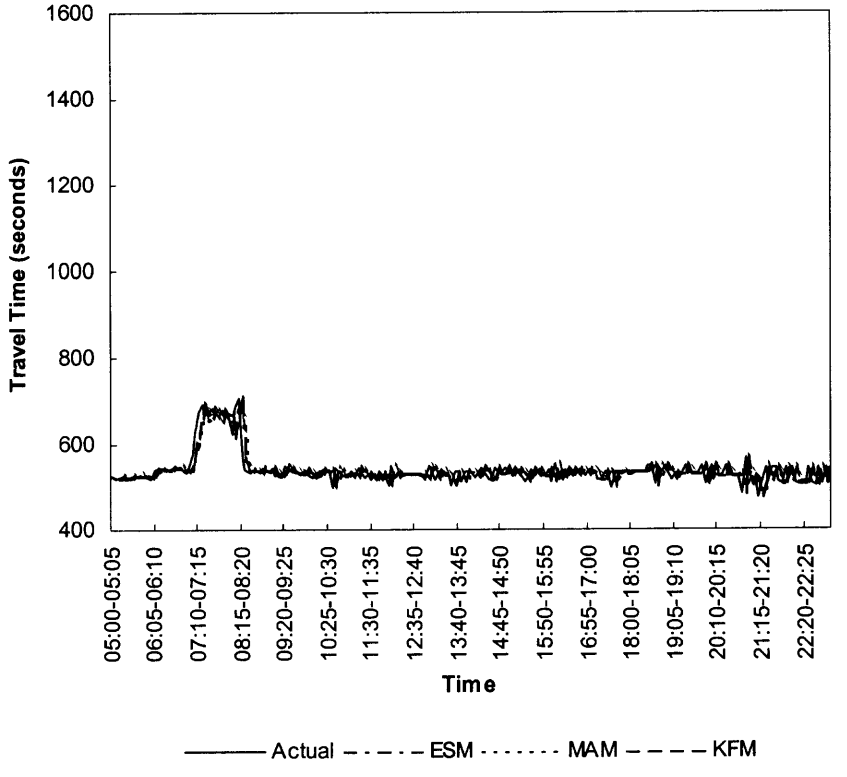

Figure 4.9 Predictions under Non-recurrent Congested Conditions.

4.7.3 Comparison and Analysis

The performance analysis of the base prediction models was conducted by comparing the MARE, VAPE, and MRE for all base prediction models, and the results were included in Tables 4.11, 4.12, 4.13 and 4.14:

- In general, the accuracy of the ESM, MAM, and KFM is degraded when traffic conditions changes from Scenario 1 (free-flow) to Scenario 2 (recurrent congested condition) and Scenario 3 (non-recurrent congested conditions). For example, the MARE of ESM increases from 1.31% to 3.19% and 2.66% when traffic changes from Scenario 1 to Scenarios 2 and 3, respectively as shown in Table 4.13. Similar relations can be observed when using MAM and KFM. The trend of prediction accuracy indicates that traffic congestion significantly impacts the accuracy of the developed base prediction models.

- In Scenario 1, all base models (i.e. ESM, MAM and KFM) can generate results with satisfactory accuracy. It proves that all base models are able to perform well in predicting free-flow traffic. In Scenario 2, ESM outperforms other models significantly with smaller MARE, while KFM can generate better results than MAM. It demonstrates that ESM can effectively predict future traffic under recurrent congestion conditions. In Scenario 3, MAM could achieve the best performance with the smallest MARE. This indicates that it may be appropriate to use a different prediction model for travel time prediction under various traffic conditions.

- The investigation showed that the selection of an α value substantially affects the prediction accuracy of ESM. For example, it can be observed from Table 4.10 that under Scenario 2, the MARE of ESM would be increased from 3.19% to 5.33% for different α values. If the selected α is not appropriate for the specific network, other base models (e.g., KFM or MAM) would outperform ESM on overall prediction accuracy. Therefore, it can be concluded that it is critical to get an optimal α value to obtain satisfactory performance for ESM. To successful implement ESM, it is desirable to select parameters dynamically for short-time travel time predictions.

- The reason for the best performance of ESM under various traffic conditions is that ESM could be more "responsive" to changes occurring in the recent past by employing historical data. This increases the prediction accuracy and overcomes the time lag when the trend of future traffic conditions is similar with historical data. Even under non-recurring congested conditions, when the historical traffic data cannot reflect the future traffic trend, the ESM could generate prediction results with compatible accuracy as other models by reducing the weight of the historical data. It

proved that the ESM is an appealing approach for travel time prediction under various traffic conditions.

- The optimal α in ESM is determined based on empirical traffic data on a specific study site. The α cannot be used based on traffic data from other locations since the traffic pattern may differ from location to location. It indicates that the ESM has no capability of transferability for travel time prediction. In addition, even for the same study site, the constant weight parameter α appropriate for one time period may result in big deviations in another time period when traffic patterns change. It demonstrates that the adaptability of ESM is low since constant weight parameters cannot reflect a dynamic traffic evolution.

- VAPE is an index that reflects the variation on the level of prediction errors, which can indicate the stability of the prediction model. For example, in Scenario 2 the VAPE of MAM is much higher than that of ESM and KFM. This is because the estimation of the local mean in MAM lags behind the true value of the local mean by about (N+1)/2 periods. Thus, this time lag causes big variations when traffic experiences dramatic changes as in Scenario 2.

- MRE is an index to represent the maximum relative error of predicted results on one single interval. The more dynamic feature a prediction model has, the smaller MRE is expected of the prediction results. MRE can reveal the model's capability of timeliness to follow the traffic changes. For example, the performances of MAM and KFM are compared for Scenario 3 in Table 4.13. Even though the MAM has better prediction performance with smaller MARE, the MRE of MAM (21.18%) is larger than the KFM (20.18%) of MAM. This is still caused by the time lag inherent in the MAM.

CHAPTER 5

DEVELOPMENT OF DYNAMIC RECURSIVE MODELS

In this Chapter, the dynamic recursive models Dynamic Exponential Smoothing Model (DESM), Improved Dynamic Smoothing Model (IDESM) and Dynamic Moving Average Model (DMAM), are developed by integrating the three base prediction models (i.e., ESM, MAM and KFM) discussed in Chapter 3.

The properties and features of each base model are summarized before proceeding with the formulation of DESM, IDESM and DMAM. The ESM prediction for travel time in time t+1 is based on the historical travel time in that interval as well as the real-time observation at t. Makridakis et al. (1982) proved that the "ESM could give quite accurate forecasts" when the future traffic condition has the same trend as that in the historical profile, such as in recurrent congestion. It was discussed in Chapter 4 that the value of the smoothing constant (or weight parameter) α substantially affects the prediction accuracy. It can be expected to achieve better prediction accuracy by applying a dynamic rather than a constant weight parameter α_t for each time interval to capture the stochastic evolution of traffic conditions. Masliah (2004) concluded, "optimizing the method of ESM becomes a matter of determining the best smoothing constant α_t".

The MAM employs real-time information (e.g., travel time) of N previous intervals and averages their mean value for prediction. Alecsandru and Ishak (2004) stated that "non-recurrent traffic condition is believed to be related to the most recent past information". MAM provides an appealing approach to predict travel time when the future traffic trend cannot be reflected in the historical profile such as in cases of non-

recurrent traffic conditions. Similarly, the weight parameter θ_t for the averaged travel time is expected to improve the accuracy of MAM by reflecting the traffic variations.

The KFM has demonstrated its ability to capture dynamic traffic changes based on the most recent real-time observation, and has been wildly used for traffic trend prediction (Okutani, I., and Stephanedes, 1984, Kuchipudi, C.M. and Chien, 2003 and Jiang, 2003). This model has the ability to accommodate traffic fluctuations adequately with time varying parameters (Kalman Gain) to continuously improve the next prediction. The incentive to develop dynamic recursive models is to optimize the weight parameters with real time information and the error in the previous prediction.

5.1 Dynamic Exponential Smoothing Model (DESM)

In ESM, a conventional weight scheme such as a grid search, can only obtain a static weight value over time by using a trial and error approach as discussed in Chapter 4. Though ESM demonstrated a satisfactory performance as discussed in Chapter 4, it is possible to further improve its performance in a dynamic traffic environment, by optimizing the weight parameter α_t.

The proposed DESM integrates the features of KFM and ESM, which can timely optimize the weight parameter α_t at every time interval by applying KFM while considering the most recent prediction error. In addition, historical data employed in the DESM play an important role to eliminate the impact of the time lag in KFM when traffic is experiencing recurrent change. Thus, α_t is acting not only as the weight parameter in ESM, but also the state variable predicted in the KFM. Therefore, the ESM and KFM are integrated while α_t can be optimized based on real time data.

The weight parameter α_t is treated as the state variable in the KFM, one component of the integrated DESM. To develop a KFM for dynamic α_t prediction, the following assumptions are made:

(1) The state variable (e.g., weight parameter) to be predicted is linear on successive time intervals based on Equation 5.2;

(2) The state variable is normally distributed in the prediction horizon.

Let α_t denote the state variable (e.g., weight parameter) to be predicted at time interval t, ϕ_t denote the transition parameter at time interval t which is externally determined to represent the linear relationship between successive state variables. ω_t denote a noise term for the weight parameter that has a normal distribution with zero mean and a variance of Q_t, the covariance of the weight parameter.

Let z_t denote an observation of the measurement variable (e.g., travel time) in time t, which is the observation of travel time x_t in the DESM. v_t denotes the measurement error at time t and has a normal distribution with zero mean and a variance of R_t, the covariance of the measurement variable. H is the measurement sensitivity, which represents the linear relationship between measurement variable z_t (e.g., travel time) and state variable x_t (weight parameter).

Let P_t denote the covariance of the prediction error at time t. The notation (+) or (-) represents the priori or posteriori value of the studied variables, respectively. For example, P_t^- is the priori error covariance of the predicted weight parameter at time t, $\hat{\alpha}_t^+$ denotes the posteriori (or updated) value of predicted weight parameter when the observation of measurement variable z_t is available at time t.

The ESM, one component of DESM is formulated in Equation 5.1:

$$x_{t,DESM} = \alpha_t x_{t,h} + (1-\alpha_t) x_{t-1} \qquad 0 \le \alpha_t \le 1 \tag{5.1}$$

where $x_{t,DESM}$: Predicted travel time at t,

α_t : Weight parameter of historical data at t,

$x_{t,h}$: Historical travel time at t,

x_{t-1} : Travel time observed at t-1.

The weight of the historical travel time α_t is the key parameter to be optimized to improve the performance of the DESM so that different weights (e.g., α_t , $1-\alpha_t$) could be associated with historical and real-time data to predict future travel times.

The KFM, another component of DESM is formulated as

$$\alpha_t = \phi_{t-1}\alpha_{t-1} + \omega_{t-1} \tag{5.2}$$

$$z_t = H_t \alpha_t + v_t \tag{5.3}$$

Assume that in a linear system $\forall$ i, j, $E[\omega_i * v_j]=0$, which means ω_i and v_j are uncorrelated. The derivation of equations in DESM is discussed in Appendix C. The basic steps of the computational procedure for DESM are shown below:

Step 0: Initialization

$$E[\alpha_0] = \hat{\alpha}_0 = 0.5 \qquad E[(\alpha_0 - \hat{\alpha}_0)^2] = P_0 = 0 \quad (\text{for } t = 0)$$

The initial value of the weight parameter is set to 0.5, which allocates the same weights on historical and real-time observations when there is no priori knowledge about a traffic pattern.

Step 1: Extrapolation

State estimate extrapolation: $\hat{\alpha}_t^- = \Phi_{t-1}\, \hat{\alpha}_{t-1}^+$

Error covariance extrapolation: $P_t^- = \Phi_{t-1} P_{t-1}^+ \Phi_{t-1}^T + Q_{t-1}$

Step 2: prediction of travel time with optimized $\hat{\alpha}_t$

$$x_{t,DESM} = \hat{\alpha}_t x_{t,h} + (1-\hat{\alpha}_t) x_{t-1}$$

The optimized $\hat{\alpha}_t$ and 1- $\hat{\alpha}_t$ are associated with historical and real-time data to predict future travel times.

Step 3: calculate the prediction error

$$e_t = z_t - x_{t,DESM}$$

When a new observation of travel time z_t is available at t, the prediction error is calculated as the difference between observed travel time z_t and predicted travel time $x_{t,DESM}$ at t. When e(t) >0, it means that the predicted travel time is smaller than the actual one., and vice versa. The prediction error will be used to update the state variable in Step 7.

Step 4. Measurement of Sensitivity Parameter:

$$H_t = \frac{\partial x_{t,DESM}}{\partial \hat{\alpha}_t} = x_{t,h} - x_{t-1}$$

Step 5: Kalman Gain Calculation

$$\overline{K}_t = P_t^- H_t^T [H_t P_t^- H_t^T + R_t]^{-1}$$

Step 6: Update state variable and error covariance

State estimate update: $\hat{\alpha}_t^+ = \hat{\alpha}_t^- + \overline{K}_t [z_t - H_t \hat{\alpha}_t^-]$.

To remain within the value range for the weight parameter in DESM, the $\hat{\alpha}_t^+ = 1$ when $\hat{\alpha}_t + K_t e_t > 1$; and $\hat{\alpha}_t^+ = 1$ when $\hat{\alpha}_t + K_t e_t < 0$.

Error covariance update: $P_t^+ = [I - \overline{K}_t H_t] P_t^-$

Step 7: Let t = t + 1 and go to Step 1 until the pre-specified prediction horizon ends.

The measurement variable z_t could be obtained by averaging the travel times reported from probe vehicles or the simulation result generated for time interval t. The state transition parameter ϕ_t represents the linear relationship between the state variable (e.g., weight parameter) of successive intervals in DESM. In this study, the relationship between successive weight parameters is unknown. Thus, we assumed that $\hat{\alpha}_t^- = \hat{\alpha}_{t-1}^+$, that means ϕ_t is set to be 1.

The proposed DESM optimizes weight parameters recursively based on the most recent observation to capture dynamic traffic changes. Since both historical and real-time traffic data are employed in DESM, it is anticipated to perform well for recurrent traffic congestion because the future traffic pattern can be "seen" from the historic profile.

5.2 Improved Dynamic Exponential Smoothing Model (IDESM)

In DESM, the range of the weight parameter α_t is limited between 0 and 1. It means that the predicted value will fall in a range between $x_{t,h}$ and x_{t-1}, but this is not always true in real-world applications. Instead, such a constraint may reduce the performance of the prediction model under certain traffic conditions such as abrupt traffic congestion not captured in the historical traffic profile.

The travel time prediction model in DESM according to Equation 5.1 is:

$$x_{t,DESM} = \hat{\alpha}_t x_{t,h} + (1-\hat{\alpha}_t) x_{t-1}$$

Assume that the actual travel time x_t on interval t is known. Then the "true" weight parameter that can make $x_{t,DESM} = x_t$ is derived from Equation 5.1 as:

$$\alpha_t = \frac{x_t - x_{t-1}}{x_{t,h} - x_{t-1}} \tag{5.4}$$

It can be observed from Equation 5.4 that the optimal value of α_t could be bigger than 1 or less than 0. Therefore, the Improved Dynamic Smoothing Model (IDESM) is proposed to optimize weights with a broader range [e.g., $(-\infty, +\infty)$ instead of (0,1)]. The approaches used for extending the range of the weight parameter can be classified into interpolation and extrapolation methods and are discussed below.

(1) Interpolation

When the value of weight α_t is set within the range (0,1), the predicted travel time $x_{t,DESM}$ in DESM can be interpolated by historical travel time $x_{t,h}$ at t and x_{t-1} observed at t-1. It means that $x_{t,DSM}$ will fall in the range between $x_{t,h}$ and x_{t-1} as shown in Figure 5.1.

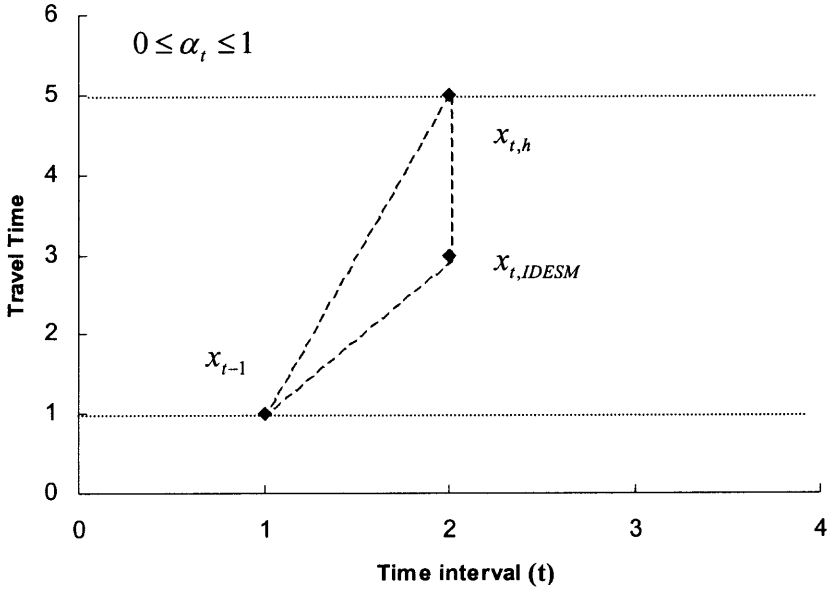

Figure 5.1 Interpolation of α_t between (0,1).

(1) Extrapolation

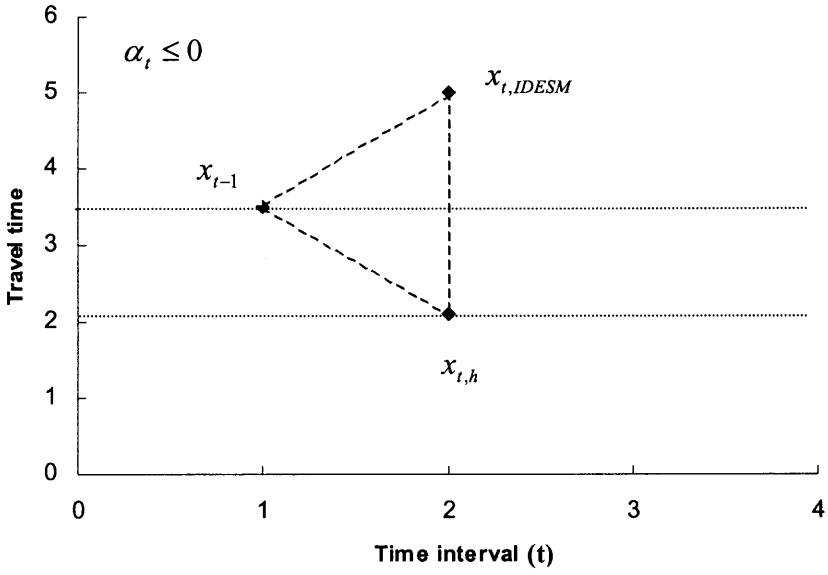

Figure 5.2 Extrapolation of α_t between (-∞, 0)

Under other circumstances (shown in Figures 5.2 and 5.3), the predicted travel time $x_{t,IDESM}$ could fall out of $x_{t,h}$ and x_{t-1} range. The interpolation of $x_{t,DESM}$ between $x_{t,h}$ and $x_{t-1,r}$ cannot satisfy such a condition. Therefore, the enhanced α_t setting proposed in the IDESM can extrapolate the $x_{t,IDESM}$ between $x_{t,h}$ and $x_{t-1,r}$ Thus, the value of weight α_t could be less than 0 and greater than 1.

The formulation of IDESM is the same with that of DESM except for the value range of α_t. Thus one can refer to section 5.1 and Appendix C for the derivation and application procedures.

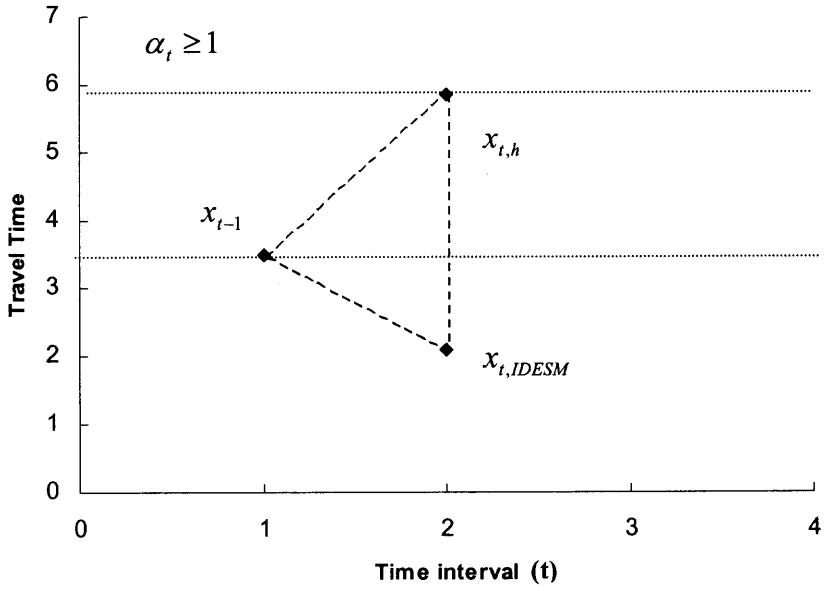

Figure 5.3 Extrapolation of α_t between (0, +∞).

5.3 Dynamic Moving Average Model (DMAM)

The MAM is an approach to improve the prediction when the historical data do not reflect the future traffic trend. In MAM, the static weight θ is not effective to capture the dynamic traffic change. The optimality of the weighted parameter θ_t associated with the traffic data observed in previous N intervals is critical for improving the performance of the MAM. Instead of applying conventional weight schemes such as a trial and error approach, it is desirable to develop a dynamic prediction model that can dynamically optimize the weights of previous observations in MAM based on real-time observations.

An integrated dynamic moving average model (DMAM) that combines the features of KFM and MAM is proposed, which can optimize the time-varying weight parameter θt by KFM and use a dynamic value of weight parameter θt in MAM. Thus, θt is acting not only as the weight parameter in MAM, but also the state variable predicted

in the KFM. Therefore, the ESM and KFM are integrated while θ_t can be optimized based on real time observations.

The weight parameter θ_t is treated as the state variable in the KFM, one component of the integrated DMAM. To develop a KFM for dynamic θ_t prediction, the following assumptions are made:

(1) The state variable (e.g., weight parameter) to be predicted is linear on successive time intervals based on Equation 5.6;

(2) The state variable is normally distributed in the prediction horizon.

Let θ_t denote the state variable (e.g., weight parameter) to be predicted at time interval t, ϕ_t denote the transition parameter at time interval t and is externally determined to represent the linear relationship between successive state variables. ω_t denotes a noise term of the weight parameter that has a normal distribution with zero mean and a variance of Q_t, the covariance of the weight parameter.

Let z_t denote an observation of the measurement variable (e.g., travel time) in time interval t, which is the observation of travel time x_t in the DMAM. v_t denotes the measurement error at time t that has a normal distribution with zero mean and a variance of R_t, the covariance of the measurement variable. H is the measurement sensitivity, which represents the linear relationship between the measurement variable z_t (e.g., travel time) and state variable x_t (weight parameter).

Let P_t denote the covariance of the prediction error at time t. The notation (+) or (-) represents the priori or posteriori value for the studied variables, respectively. For example, P_t^- is the priori error covariance of the predicted weight parameter at time t,

$\hat{\theta}_t^+$ denotes the posteriori (or updated) value of the predicted weight parameter when an observation of the measurement variable z_t is available at time t.

The MAM, one component of DMAM is formulated in Equation 5.5:

$$x_{t,DMAM} = \theta_t \left(\sum_{n=1}^{N} x_{t-n} \right) / N \tag{5.5}$$

Where

$x_{t,DMAM}$: Predicted travel time at time t,

x_{t-n} : Observed travel time at time t-n,

N: number of time periods included in the moving average,

θ_t : Weight parameter at time t.

The weight parameter θ_t is the key parameter to be optimized to improve the performance of the DMAM so that different weights θ_t could be associated with real-time observations to predict future travel times.

The KFM, another component of DMAM is formulated as

$$\theta_t = \phi_{t-1}\theta_{t-1} + \omega_{t-1} \tag{5.6}$$

$$z_t = H_t\theta_t + v_t \tag{5.7}$$

Assume that in a linear system $\forall$ i, j, $E[\omega_i * v_j]=0$, which means that ω_i and v_j are uncorrelated. The derivation of equations in DMAM is discussed in Appendix D. The basic steps of the computational procedure for DMAM are shown below:

Step 0: Initialization

$$E[\theta_0] = \hat{\theta}_0 = 1 \qquad E[(\alpha_0 - \hat{\alpha}_0)^2] = P_0 = 0 \quad (\text{for } t = 0)$$

The initial value of the weight parameter is set to 1, when there is no priori knowledge about a traffic pattern.

Step 1: Extrapolation

State estimate extrapolation: $\hat{\theta}_t^- = \Phi_{t-1}\ \hat{\theta}_{t-1}^+$

Error covariance extrapolation: $P_t^- = \Phi_{t-1}\ P_{t-1}^+\ \Phi_{t-1}^T + Q_{t-1}$

Step 2: Predict travel time with optimized weight parameter $\hat{\theta}_t$

$$x_{t,DMAM} = \hat{\theta}_t \left(\sum_{n=1}^{N} x_{t-n} \right) / N$$

The optimized state variable $\hat{\theta}_t$ is applied on real time travel time observations in N previous time intervals in DMAM for travel time prediction.

Step 3: Calculate the prediction error

$$e_t = z_t - x_{t,DMAM}$$

When a new observation of travel time z_t is available at t, the prediction error is calculated as the difference between observed travel time z_t and predicted travel time $x_{t,DMAM}$ at t. When e(t) >0, it means that the predicted travel time is smaller than actual one and vice versa. The prediction error will be used to update the state variable on Step 7.

Step 4. Measurement of Sensitivity Parameter:

$$H_t = \frac{\partial x_{t,DMAM}}{\partial \hat{\theta}_t} = \left(\sum_{n=1}^{N} x_{t-n} \right) / N$$

Step 5: Kalman Gain Calculation

$$\overline{K}_t = P_t^-\ H_t^T [\, H_t\ P_t^-\ H_t^T + R_t \,]^{-1}$$

Step 6: Update of state variable and error covariance

State estimate update: $\hat{\theta}_t^+ = \hat{\theta}_t^- + \overline{K}_t [\, z_t - H_t \, \hat{\theta}_t^- \,]$

Error covariance update: $P_t^+ = [\mathrm{I} - \overline{K}_t \, H_t \,] \, P_t^-$

Step 7: Let t = t + 1 and go to Step 1 until the pre-specified prediction horizon ends.

The measurement variable z_t could be obtained by averaging the travel times reported from probe vehicles or simulation results generated for time t. The state transition parameter ϕ_t represents the linear relationship between the state variable (e.g., weight parameter) of successive intervals in DMAM. In this study, the relationship between successive weight parameters is unknown. Thus, we assume that $\hat{\theta}_-^t = \hat{\theta}_{t-1}^+$, it means ϕ_t is set to 1.

The proposed DMAM applies optimized weights recursively based on the most recent observation. Thus, the weight parameter can be optimized for averaging real-time observations in N previous time intervals to reflect dynamic traffic changes. Since no historical data is employed in DMAM, it is expected to perform well for non-recurrent traffic congestion because the future traffic pattern cannot be "seen" from the historical profile in such a situation.

CHAPTER 6

CASE STUDY II

6.1 Background Introduction

As discussed in the beginning of this document, the objective of this study is to develop dynamic recursive models for short-term travel time prediction. Three base models including ESM, MAM and KFM, have been developed in Chapter 3, while an evaluation of the base models was presented in Chapter 4. Each base model had demonstrated its performance through its MARE, VAPE, and MRE under various traffic conditions. Results showed that the static weights embedded in ESM and MAM and the inherent time lag in KFM reduce their prediction accuracy and stability, especially when traffic conditions change dramatically.

In Chapter 5, two dynamic recursive models, called DESM and DMAM were developed by integrating ESM and MAM with KFM, respectively. In addition, IDESM is also developed by extending the feasible range of the weight parameter embedded in the DESM. The developed three dynamic recursive models are able to adapt to real-time traffic changes based on the optimized time-varying weight parameters.

In this Chapter, the same experiments designed in Chapter 4 are used for testing the three dynamic recursive models. The evaluation processes and their prediction accuracy are analyzed and compared with those of the base models. First, each dynamic recursive model is tested under the three traffic scenarios (i.e., free-flow condition, recurrent and non-recurrent congested conditions) with the developed CORSIM simulation model discussed in Chapter 4. Three statistical indices (e.g., MARE, VAPE, and MRE) are adopted to evaluate the prediction results, with MARE being the primary

MOE to assess the prediction accuracy, while VAPE and MRE are secondary MOEs to represent the stability of the prediction model. Then, a comparative analysis is made of the dynamic recursive models and their corresponding base models (e.g., DESM vs. ESM, IDESM vs. DESM, DMAM vs. MAM).

Finally, the prediction interval is evaluated, while the impact of the interval duration to the prediction accuracy is analyzed. Traffic data collected in different durations of prediction intervals are used with the dynamic recursive models under various traffic scenarios.

6.2 Testing of the Developed Dynamic Recursive Models

6.2.1 DESM

To fairly compare the prediction accuracy between the dynamic and base models, the same level of traffic data detail is used. The data were collected and processed every 5 minutes for the entire evaluation period of 18 hours from 5:00 am to 11:00 pm. To evaluate the models in a more microscopic way, the overall prediction period is divided into three periods for Scenario 2 (i.e., recurrent congested condition) and Scenario 3 (i.e., non-recurrent congested condition). The purpose of the separate periods is to analyze the impact of time-varying optimal weights on the model's prediction capabilities at times when traffic conditions change at different states. Since there is no significant traffic change in Scenario 1, the prediction accuracy for the entire evaluation period (18 hours) is analyzed.

The separation of the overall evaluation period for Scenarios 2 and 3 is shown in Figures 6.1 and 6.2, respectively. In Scenario 2 (Figure 6.1), the entire evaluation period (5:00 am-11:00 pm) is divided into three time periods (TPs), including TP 1 (5:00 am-7:30 am), TP 2 (7:30 am-10:30 am), and TP 3 (10:30 am-11:00 pm). Traffic experiences recurrent congestion during TP 2.

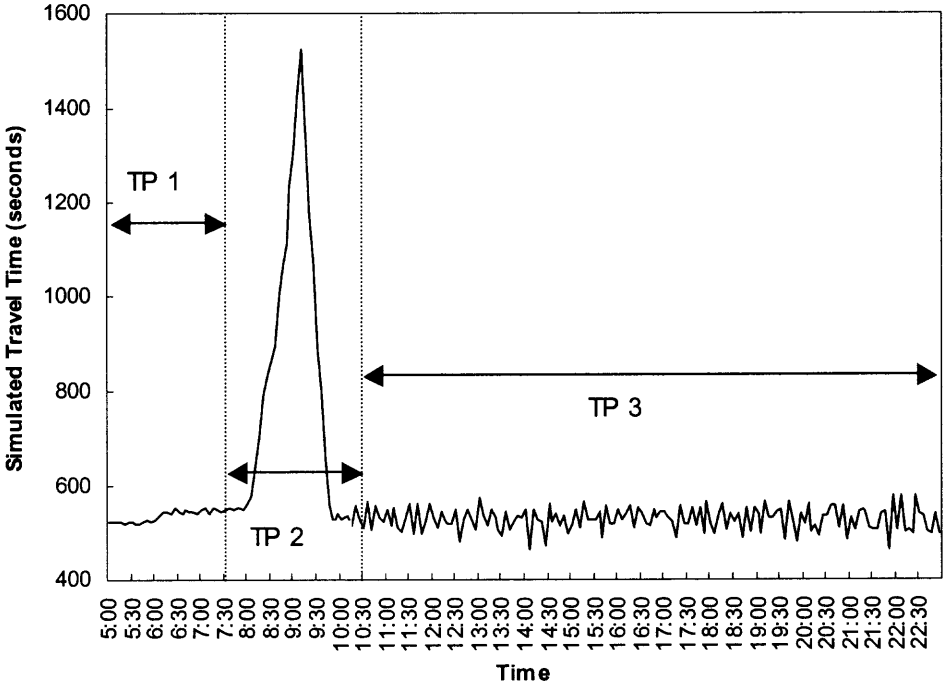

Figure 6.1 Configuration of the Evaluation Period in Scenario 2.

In Scenario 3, the entire evaluation period is also divided into three TPs, including TP 1 (5:00 am-6:30 am), TP 2 (6:30 am-8:30 am) when traffic experiences non-recurrent congestion, and TP 3 (8:30 am-11:00 pm),.

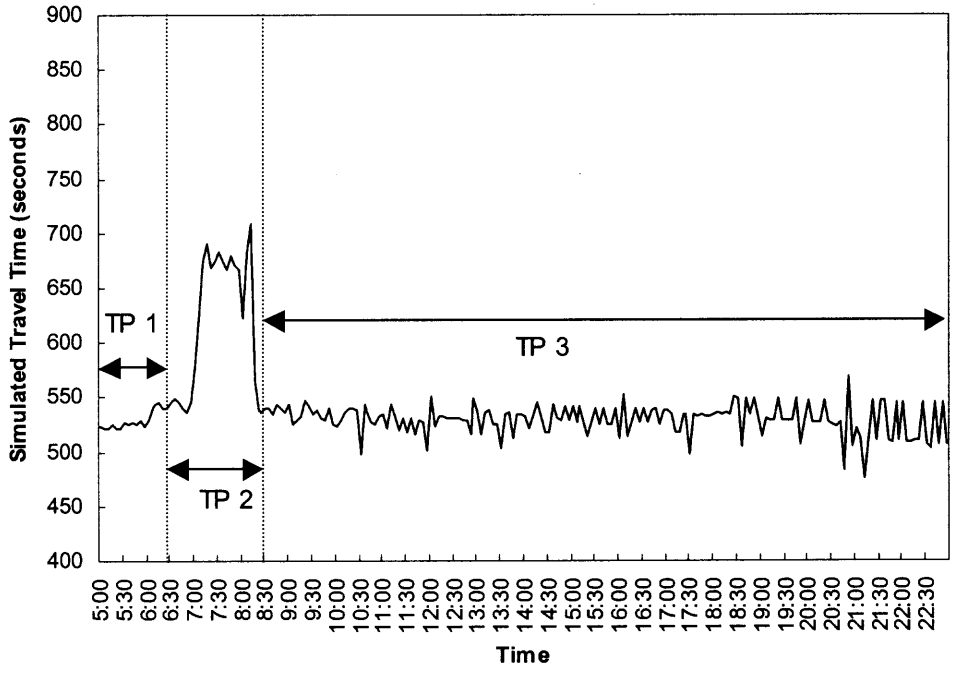

Figure 6.2 Configuration of the Evaluation Period in Scenario 3.

The DESM has been implemented for all three Scenarios. To demonstrate the computation procedure, a DESM application example based on Scenario 2 is shown in Table 6.3, where α_t acts as a "slope" in interpolating or extrapolating the historical and real-time traffic data. α_t is sensitively affecting the accuracy of predicted travel times. A minor change of α_t could lead to a major change of predicted travel time. Considering that, the covariance of process noise Q_t should be minor to keep the variation of α_t small. A numerical search was made to optimize the combination of (R_t, Q_t). The results are shown in Table 6.2 and Figure 6.3, where the suggested value ranges for (Rt, Qt) are presented.

Table 6.1 Prediction Travel Time with DESM under Scenario 1.

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Time	x_t	$x_{t,h}$	$\hat{\alpha}_t^+$	Error (%)	ϕ_t	R_t	Q_t	K_t	$\hat{x}_t$	$\hat{\alpha}_t$	P_t^-	P_t^+	Measured	H_t
05:00-05:05	524	524	0.50		1	50000	1					0	524	
05:05-05:10	521	525	0.50	0.72	1	50000	1	0	524	0.500000	1.00	1.00	521	1.4
05:10-05:15	521	524	0.50	0.32	1	50000	1	0	523	0.499990	2.00	2.00	521	3.5
05:15-05:20	524	523	0.50	0.40	1	50000	1	0	522	0.499966	3.00	3.00	524	2.6
22:45-22:50	532	525	0.95	1.36	1	50000	1	0	525	0.971275	20.93	20.47	532	-7.3
22:50-22:55	532	521	0.90	2.02	1	50000	1	0	522	0.949915	21.47	20.41	532	-11.0
22:55-23:00	495	521	1.03	5.28	1	50000	1	0	522	0.902886	21.41	20.31	495	-11.2

Note:

(0) Time interval (5 minutes).

(1) Real time travel time x_t at t.

(2) Historical travel time $x_{t,h}$ at t.

(3) Updated state variable, $(3)_t=(10)_t+(8)_t*[(13)_t-(9)_t]$.

(4) Prediction error percentage, $(4)_t=|(9)_t-(1)_t|/(1)_t*100\ \%$.

(5) State transition matrix ϕ_t is assumed 1.

(6) Covariance matrix of observational (measurement) uncertainty, $(6)_t=50000$.

(7) Covariance matrix of process noise in the system state dynamics, $(7)_t=1$.

(8) Kalman gain matrix, $(8)_t=(11)_t*(14)_t*[(14)_t*(11)_t*(14)_t+(6)_t]^{-1}$.

(9) Predicted travel time (seconds), $(9)_t=[1-(10)_t]*(1)_{t-1}+(10)_t*(2)_t$.

(10) State estimates, $(10)_t=(5)_{t-1}*(3)_{t-1}, (10)_t \in [0,1]$.

(11) Estimation error covariance, $(11)_t=(5)_{t-1}*(12)_{t-1}*(5)_{t-1}+(7)_{t-1}$.

(12) Updated estimation error covariance, (12)t=[1-(8)t*(14)t]*(11)t.

(13) Measured travel time (seconds), (13)t=(1)t.

(14) Measurement sensitivity, (14)t=(2)t-(1)t-1.

Table 6.2 MARE of Combination (R_t, Q_t) in DESM (Scenario 1).

DESM	MARE			
(R,Q)	1	10	100	1000
50000	**1.32**	1.33	1.36	1.58
5000	1.34	1.36	1.58	1.97
500	1.36	1.58	1.97	2.81
50	1.58	1.97	2.81	5.79

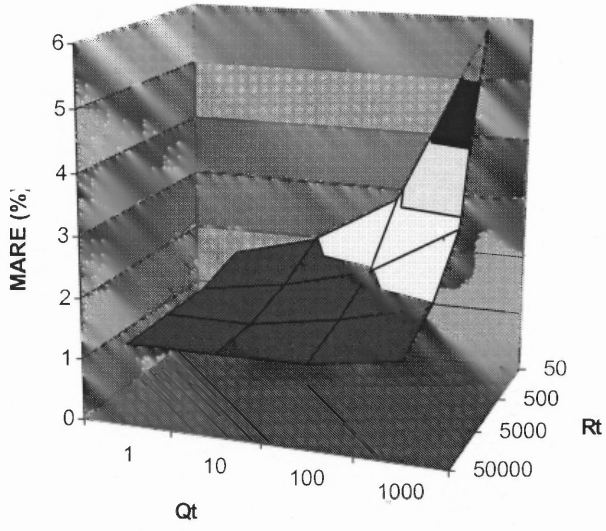

Figure 6.3 Sensitivity Analysis of Combination (R_t, Q_t) in DESM.

The state variable α_t whose feasible range is between 0 and 1 is dynamically optimized in the DESM. The optimized weight parameters are shown in Figures 6.4, 6.5, and 6.6 for Scenarios 1, 2, and 3, respectively. The prediction results in different TPs are calculated by using Equations 4.2, 4.3, and 4.4, while the statistical indices (i.e., MARE, VAPE, and MRE) are shown in Table 6.3. The prediction accuracy comparison under the three traffic scenarios is summarized in Figure 6.7.

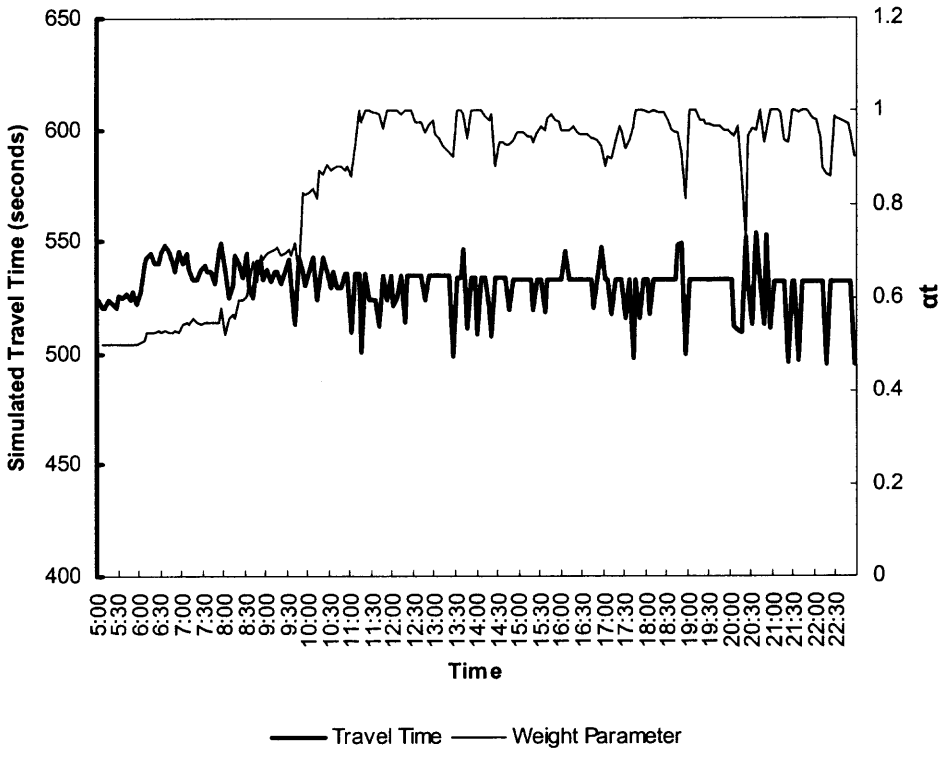

Figure 6.4 Optimized DESM Weight Parameter vs. Time (Scenario 1).

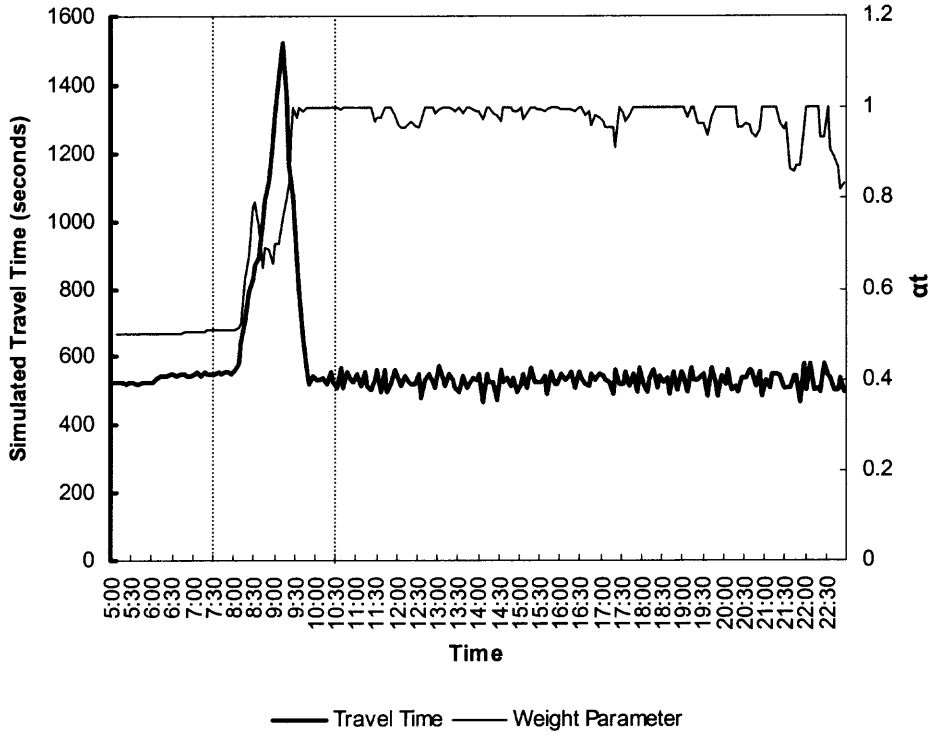

Figure 6.5 Optimized DESM Weight Parameter vs. Time (Scenario 2).

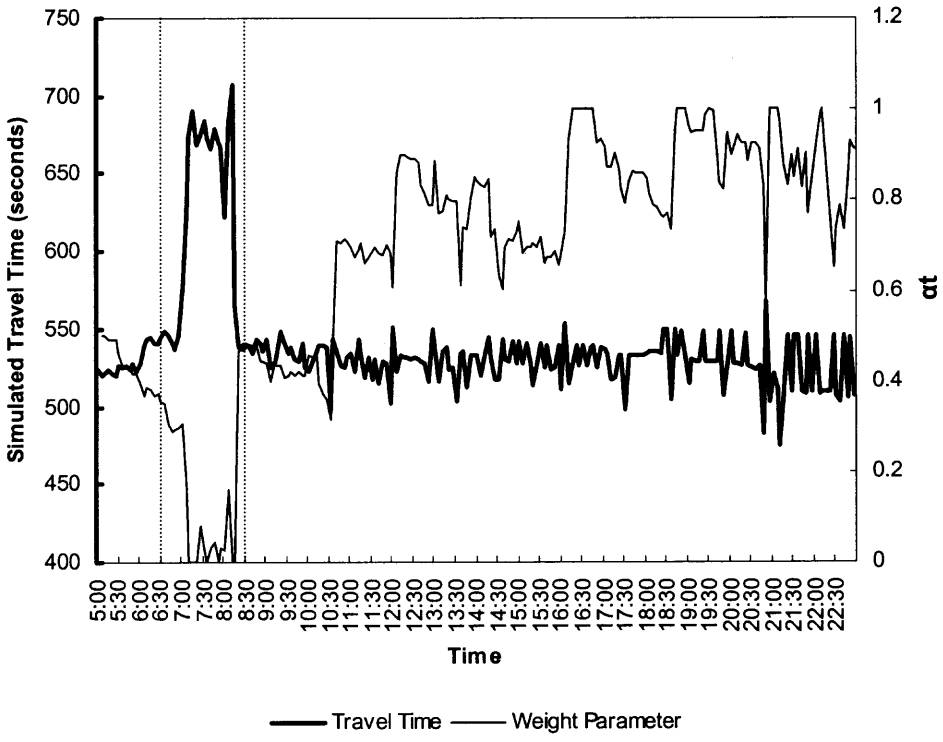

Figure 6.6 Optimized DESM Weight Parameter vs. Time (Scenario 3).

In Figure 6.4, the variation of travel times for Scenario 1 is minor during the entire prediction period. The optimized weight parameter varies between 0.5 and 1 from 5:00 am to 11:00 am, and then fluctuates between 0.8 and 1 during 11:00 am to 11:00 pm. At some spots between 20:10 pm and 21:10 pm, active traffic variation causes significant variation of the weight parameter.

In Figure 6.5, the optimized weight parameter increases dramatically from TP 1 to TP 2. Because the historical traffic data are similar to the future traffic trend in recurrent congested conditions, the increased weight parameter can guarantee the historical data has more influence on the predicted information. It has been demonstrated that the DESM could achieve better accuracy by employing an optimized weight parameter when recurrent traffic congestion occurs.

In Figure 6.6, a decreasing optimized weight parameter is observed in TP 2, where a non-recurrent congestion event occurs. Under this situation, the historical data would be different from the future traffic trend. Thus, the decreased weight parameter implies that the prediction results would be influenced more by the real-time data rather than the historical data. It demonstrated that the DESM could employ the optimized weight under a non-recurrent congested condition and perform substantially well.

Table 6.3 Results of Prediction Accuracy (DESM).

Scenario	Time Period	MARE (%)	VAPE (%)	MRE (%)
1	Overall	1.32	2.02	7.58
2	1	0.57	0.25	1.85
	2	**2.19**	**2.49**	**6.00**
	3	**3.84**	**8.46**	**12.75**
	Overall	3.11	7.72	12.75
3	1	1.11	0.41	2.64
	2	**3.89**	**28.96**	**25.14**
	3	**2.28**	**4.25**	**11.44**
	Overall	2.37	6.97	25.14

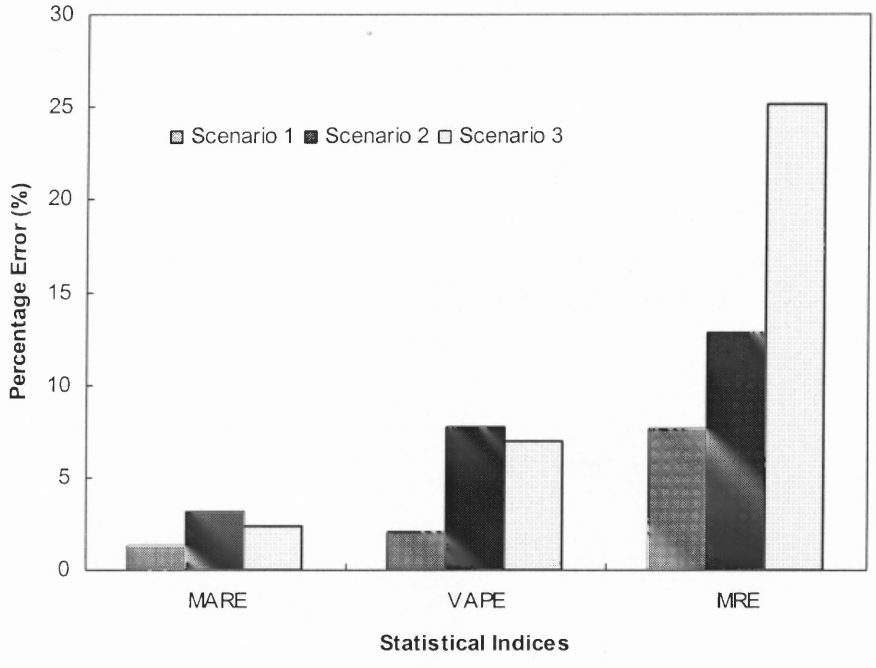

Figure 6.7 Prediction Accuracy of DESM (5:00am-11:00pm).

It is observed in Figure 6.7 that the MAREs of Scenarios 2 and 3 are higher than that of Scenario 1, and so are the VAPE and MRE. However, the prediction accuracy in Scenario 1 is superior to that in Scenarios 2 and 3. This implies that traffic congestion could reduce the accuracy of DESM.

The MARE and VAPE in Scenario 2 are slightly higher than in Scenario 3. This indicates that the DESM could achieve satisfactory performance in terms of accuracy and stability in both recurrent and non-recurrent congested conditions. However, the MRE of DESM in Scenario 3 is significantly increased in comparation with the other two scenarios in Figure 6.7 due to a non-recurrent congestion event (i.e., incident).

It should be noted that in Scenario 2 (see Table 6.3), the prediction accuracy and stability of the DESM in TP 2 is better than in TP 3 (i.e., free-flow condition). Given the travel times in individual TPs, the reason causing the difference of performance is the higher variation of travel times in TP 3.

6.2.2 IDESM

The IDESM is an extended DESM model. The difference between IDESM and DESM is that the IDESM enables α_t to adopt a value in the range of $(-\infty, +\infty)$ instead of [0,1]. The extended α_t is able to predict travel time by extrapolating the historical and real-time data to efficiently adapt to substantial traffic changes.

The IDESM is tested under three traffic scenarios. The statistical indices (i.e., MARE, VAPE, and MRE) are summarized in Table 6.4. Figure 6.8 shows the overall performance indices of the IDESM under the three traffic scenarios.

Table 6.4 Results of Prediction Accuracy (IDESM).

Scenario	Time Period	MARE (%)	VAPE (%)	MRE (%)
1	Overall	1.29	1.95	7.59
2	1	0.57	0.25	1.85
	2	**2.21**	**2.79**	**6.00**
	3	**3.76**	**8.30**	**12.36**
	Overall	3.06	7.56	12.36
3	1	0.46	0.05	1.03
	2	**4.23**	**38.29**	**28.70**
	3	**1.94**	**3.44**	**9.93**
	Overall	2.08	7.64	28.70

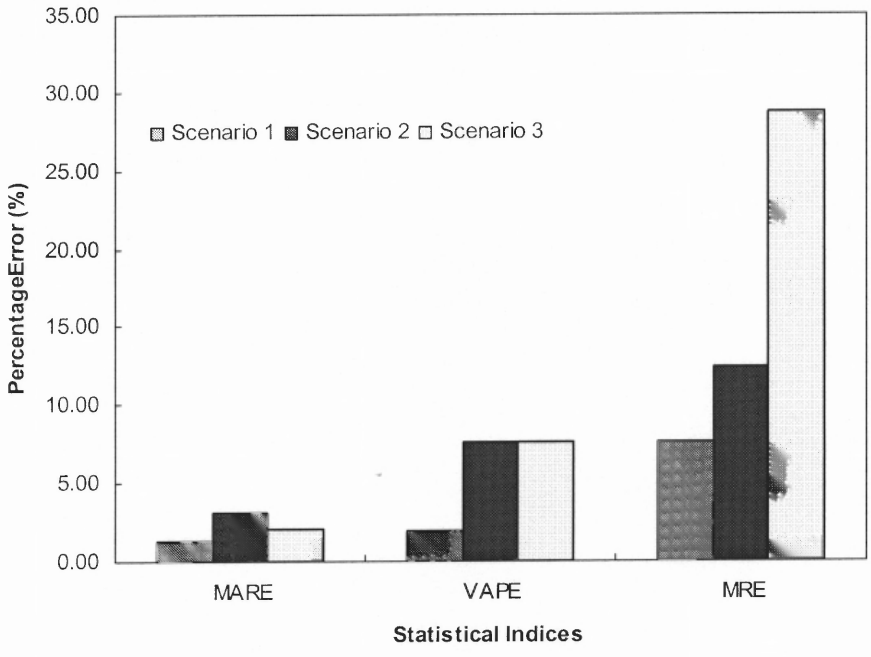

Figure 6.8 Prediction Accuracy of IDESM (5:00 am-11:00 pm).

As in Table 6.4 and Figure 6.4, it can be observed that the trend of overall prediction performance of the IDESM is similar to that of the DESM in various traffic conditions.

6.2.3 DMAM

To demonstrate the computation procedure, an example of DMAM application in Scenario 1 is shown in Table 6.5, in which the system parameters (Rt, Qt) are set at (50000000, 0.1) based on the same approach as discussed in Section 6.2.1. The DMAM is implemented under three traffic scenarios for travel time prediction. The statistical indices (i.e., MARE, VAPE, and MRE) are summarized in Table 6.6. The overall performance indices of the DMAM are shown in Figure 6.9.

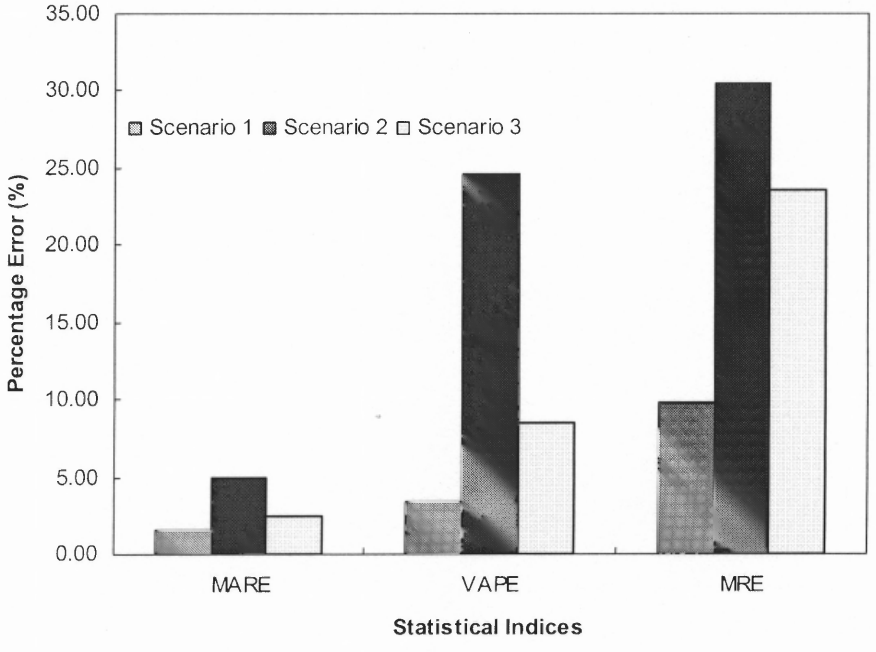

Figure 6.9 Prediction Accuracy of DMAM (5:00 am-11:00 pm).

Table 6.5 Prediction Travel Time with DMAM under Scenario 1.

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Time	x_t	$\hat{\theta}_t^+$	Error (%)	ϕ_t	R_t	Q_t	K_t	$\hat{x}_t$	$\hat{\theta}_t$	P_t^-	P_t^+	Measured	H_t
05:00-05:05	524			1	50000000	0.1						524	524
05:05-05:10	521	0.000		1	50000000	0.1	0.000			0.000	0.000	521	522
05:10-05:15	521	1.000	0.27	1	50000000	0.1	0.000	522	1.0000000	0.100	0.100	521	522
05:15-05:20	524	1.000	0.65	1	50000000	0.1	0.000	521	0.9999985	0.200	0.200	524	521
22:45-22:50	532	1.000	0.01	1	50000000	0.1	0.000	532	0.9999200	4.28	4.18	532	532.2
22:50-22:55	532	1.000	0.01	1	50000000	0.1	0.000	532	0.9999219	4.28	4.18	532	532.2
22:55-23:00	495	0.998	7.46	1	50000000	0.1	0.000	532	0.9999237	4.28	4.18	495	532.2

Note:

(0) Time interval (5 minutes).

(1) Real time travel time x_t at t.

(2) Updated state variable, $(2)_t=(9)_t+(7)_t*[(12)_t-(8)_t]$.

(3) Prediction error percentage, $(3)_t=|(8)_t-(1)_t|/(1)_t*100$ %.

(4) Prediction error percentage, (3)t=abs [(8)t- (1)t]/(1) t *100 %.

(5) State transition matrix ϕ_t is assumed to be 1.

(6) Covariance matrix of observational (measurement) uncertainty, (5)t=50000000.

(7) Covariance matrix of process noise in the system state dynamics, (6)t=0.1.

(8) Kalman gain matrix, (7)t=(10)t*(13)t*[(13)t*(10)t*(13)t +(5)t]-1.

(9) Predicted travel time (seconds), (8)t=[(1)t-1+(1)t-2]*(9)t/2.

(10) State estimates, (10)t=(4)t-1*(2)t-1.

(11) Estimation error covariance, (10)t=(4)t-1*(11)t-1*(4)t-1+(6)t-1.

(12) Updated estimation error covariance, (11)t=[1-(7)t*(13)t]*(10)t.

(13) Measured travel time (seconds), (12)t=(1)t.

(14) Measurement sensitivity, (13)t=[(1)t-1- (1)t-2]/2.

Table 6.6 Results of Prediction Accuracy (DMAM).

Scenario	Time Period	MARE (%)	VAPE (%)	MRE (%)
1	Overall	1.62	3.43	9.78
2	1	0.80	0.38	2.26
	2	**7.99**	**64.36**	**30.49**
	3	**5.03**	**14.43**	**17.02**
	Overall	4.95	24.57	30.49
3	1	0.69	0.44	2.76
	2	**4.83**	**34.82**	**23.57**
	3	**2.38**	**4.72**	**11.36**
	Overall	2.53	8.49	23.57

As shown in Table 6.5, the lower MARE, VAPE, and MRE in Scenario 1 demonstrate that the DMAM could achieve the best prediction in free-flow condition. In Scenario 2, the MARE, VAPE, and MRE of the DMAM are significantly higher than in Scenarios 1 and 3, which imply that the accuracy and stability of the DMAM is reduced dramatically under recurrent congested condition. This result indicated the fact that DMAM does not have historical data similar to those currently occurring to predict the future trend. In Scenario 3, the MARE and VAPE of the DMAM are improved compared with those of Scenario 2. This demonstrates that in a non-recurrent congested condition, the DMAM could perform well to predict travel time without employing historical data. The high MRE of the DMAM in Scenario 3 is caused by the time lag inherent in MAM.

It can be observed in in Table 6.5 that in both Scearnios 2 and 3 the MOEs, including MARE, VAPE, and MRE of the DMAM in TP 2 are much higher than those in other TPs. Thus, the DMAM is less capable of predicting travel time in both recurrent and non-recurrent congested conditions.

6.3 Evaluation of the Dynamic Recursive and Base Models

6.3.1 DESM and ESM

The DESM is developed based on its base model ESM and both models have identical forms. The ESM uses the smoothing constant α, while the DESM employs the optimized weight parameter α_t for travel time prediction. The optimal α of ESM is determined by a grid search to find the best prediction results in terms of minimum prediction error. The optimal α_t is derived from the KFM to produce the minimum mean squared error, and assumes that α_t is normal distributed and is a linear on the successive intervals.

Based on the prediction results, the comparison between DESM and ESM of MOEs for the three Scenarios is shown in Figure 6.10. In addition, the statistical indices for the DESM and ESM in different TPs of the three scenarios are shown in Table 6.7.

It can be seen from Figure 6.10 that the MARE, VAPE, and MRE of the DESM are lower than those of the ESM under the three traffic scenarios, except for a higher MRE in Scenario 3. This demonstrates that the DESM outperforms the ESM under various traffic conditions.

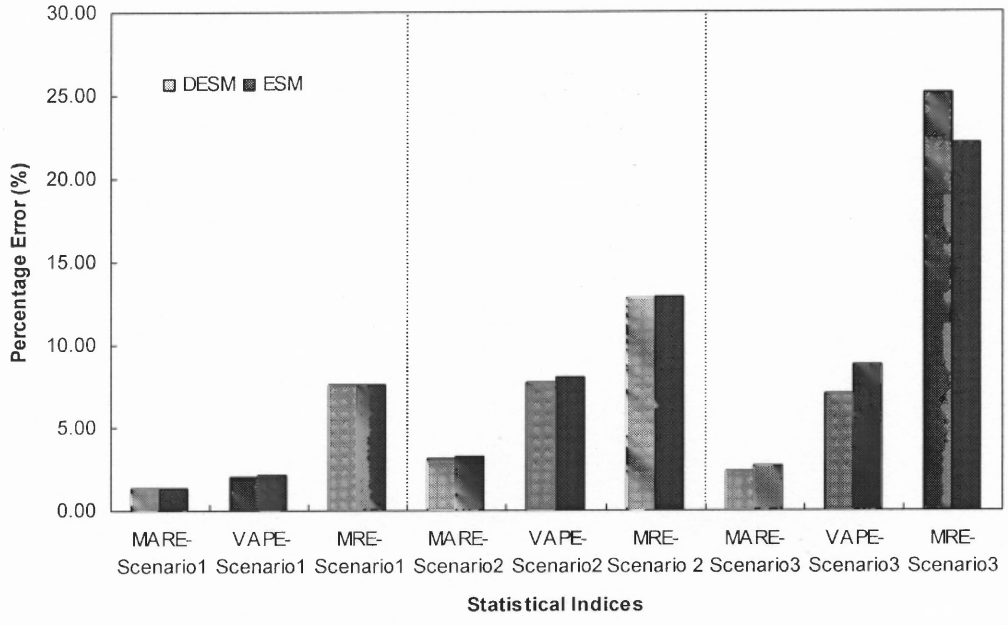

Figure 6.10 Prediction Accuracy of DESM and ESM (5:00am-11:00pm).

Table 6.7 Prediction Accuracy of DESM and ESM.

Scenario	Time Period	MARE		VAPE		MRE	
		DESM	ESM	DESM	ESM	DESM	ESM
1	Overall	1.32	1.31	2.02	2.06	7.58	7.58
2	1	0.57	0.58	0.25	0.17	1.85	1.57
	2	**2.19**	**2.12**	2.49	3.02	6.00	6.12
	3	3.84	3.97	8.46	8.61	12.75	12.90
	Overall	3.11	3.19	7.72	8.04	12.75	12.90
3	1	**1.11**	**0.65**	0.41	0.18	2.64	2.03
	2	3.89	3.99	28.96	23.31	25.14	22.12
	3	2.28	2.68	4.25	7.07	11.44	14.22
	Overall	2.37	2.66	6.97	8.74	25.14	22.12

In Table 6.7, the MARE, VAPE, and MRE generated by the DESM over the ESM can be observed in overall and most of individual TPs. Two exceptions can be observed as highlighted in Table 6.7, including TP 2 in Scenario 2 and TP 1 in Scenario 3. The reason is that in those situations, the selected smoothing constant in ESM coincidently matches "true" weight parameter and generates better prediction results than the dynamic

weight parameters in the DESM in the particular TP. Considering the consistent performance in most of periods, it can be concluded that the DESM outperforms the ESM in terms of prediction accuracy and stability.

6.3.2 IDESM and DESM

The IDESM was developed with the same methodology and computation procedure as the DESM. The only difference is that the IDESM allows for a wider range of the weight parameter [i.e., $(-\infty, +\infty)$ instead of [0,1]]. The statistical indices of the predicted travel time with the developed IDESM and DESM are summarized by different TPs and traffic scenarios and shown in Figure 6.11 and Table 6.8.

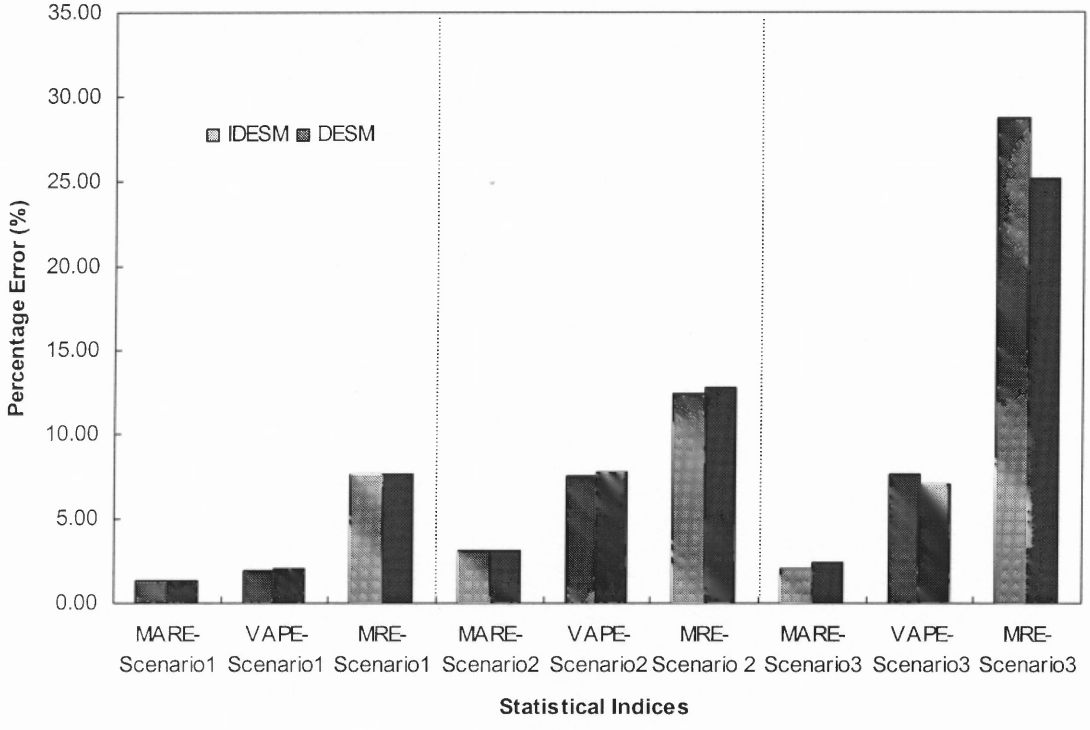

Figure 6.11 Prediction Accuracy of IDESM and DESM (5:00am-11:00pm).

Table 6.8 Prediction Accuracy of IDESM and DESM.

Scenario	Period	MARE		VAPE		MRE	
		IDESM	DESM	IDESM	DESM	IDESM	DESM
1	Overall	1.29	1.32	1.95	2.04	7.59	7.58
2	1	0.57	0.57	0.25	0.25	1.85	1.85
	2	2.21	2.19	2.79	2.49	6.00	6.00
	3	3.76	3.84	8.30	8.46	12.36	12.75
	Overall	3.06	3.11	7.56	7.72	12.36	12.75
3	1	0.46	1.11	0.05	0.41	1.03	2.64
	2	4.23	3.89	38.29	28.96	28.70	25.14
	3	1.94	2.28	3.44	4.25	9.93	11.44
	Overall	2.08	2.37	7.64	6.97	28.70	25.14

As shown in Figure 6.11, the MARE, VAPE, and MRE of the IDESM are lower than those of the DESM for Scenarios 1 and 2. In Scenario 3, the MARE of the IDESM is lower than that of the DESM, but the VAPE and MRE of the IDESM are higher than those of the DESM. This implies that in Scenario 3, the IDESM could improve the prediction accuracy but its stability was reduced. The reason is that the extended value of the weight parameter enables IDESM to quickly adapt to the traffic change. However, the tradeoff is during the initial time interval when the traffic pattern starts to change, the extended weight parameter could predict poorly and caused a higher prediction error.

Table 6.8 indicates that the IDESM outperforms the DESM in most of the TPs. However, as highlighted in Table 6.7, the accuracy of the IDESM is worse than that of the DESM in TP 2 under Scenarios 2 and 3. The reason may be the abrupt traffic changes in TP 2 that make the traffic follow a distribution other than the normal, and against the basic assumption for the optimization of the weight parameter.

6.3.3 DMAM and MAM

The DMAM is developed based on its prototype model MAM, whose weight parameter θ_t is optimized to dynamically reflect the real-time traffic trend. Based on prediction results in different TPs and traffic scenarios, the statistical indices of the DMAM and MAM are generated and shown in Figure 6.12 and Table 6.9.

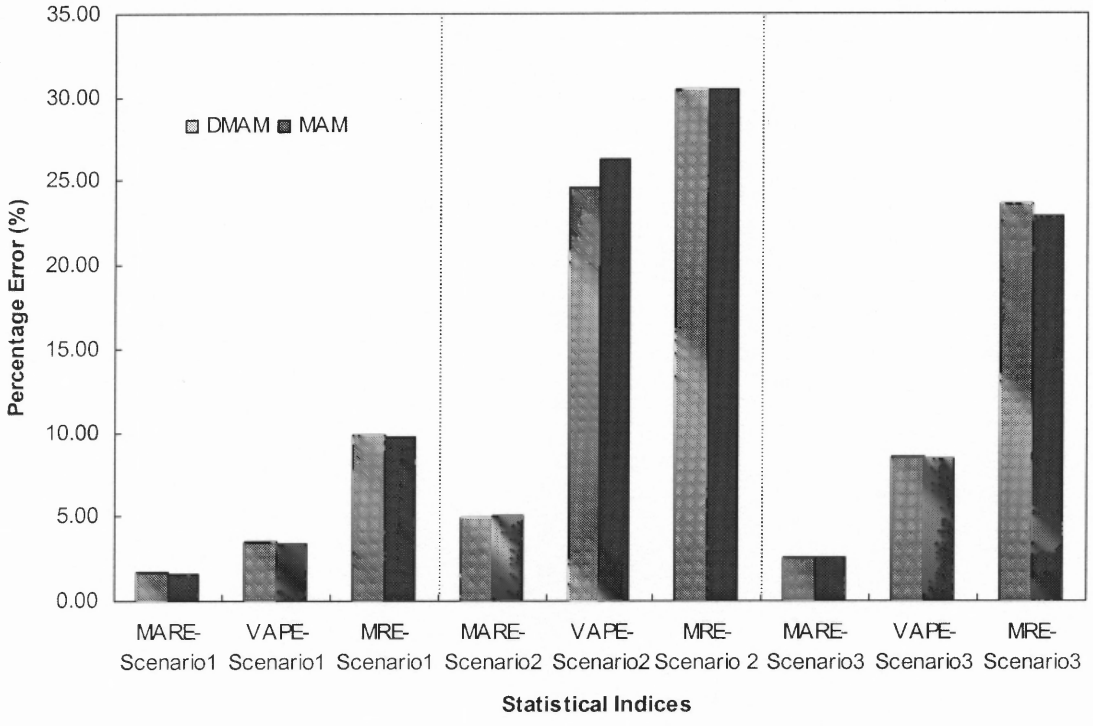

Figure 6.12 Prediction Accuracy of DMAM and MAM (5:00 am-11:00 pm).

In Figure 6.12, the MAREs of the DMAM are very close to those of the MAM in all three scenarios. This means that in general, the DMAM cannot improve the prediction accuracy of the MAM. Similarly, the VAPE and MRE of DMAM are fairly close to that of MAM.

Table 6.9 Prediction Accuracy of DMAM and MAM.

Scenario	Period	MARE		VAPE		MRE	
		DMAM	MAM	DMAM	MAM	DMAM	MAM
1	Overall	**1.62**	**1.60**	3.43	3.38	9.78	9.66
2	1	**0.80**	**0.78**	0.38	0.39	2.26	2.17
	2	7.99	9.36	64.36	73.52	30.49	30.46
	3	**5.03**	**4.87**	14.43	12.26	17.02	16.74
	Overall	4.95	5.07	24.57	26.29	30.49	30.46
3	1	0.69	0.69	0.44	0.45	2.76	2.77
	2	**4.83**	**4.76**	34.82	34.48	23.57	22.91
	3	**2.38**	**2.35**	4.72	4.66	11.36	11.14
	Overall	**2.53**	**2.49**	8.49	8.38	23.57	22.91

This may arise from two reasons. First, the time-varying weight parameter θ_t may not follow a normal distribution in the developed DMAM, which limits the performance of KFM to optimize θ_t. Second, in DMAM, θ_t is a weight parameter for information collected in the N previous intervals. Based on this setting, the travel times in previous intervals have the same impact to the future travel time, which may not reflect the real relationship between successive travel times.

6.3.4 Evaluation of the Dynamic Recursive Models

The statistical indices (i.e., MARE, VAPE, and MRE) of the three dynamic recursive models and the entire evaluation period shown in Figure 6.13, the indices by different TPs and traffic scenarios are summarized in Table 6.10.

It is shown in Figure 6.13 that the IDESM predicts future travel time well as indicated by its low MARE, VAPE, and MRE under both Scenarios 1 and 2. In Scenario 3, the IDESM has the largest VAPE and MRE although its MARE is still the best one.

As shown in Table 6.10, the DMAM performed well in Scenario 1 with low MARE, VAPE, and MRE of 1.62%, 3.43%, and 9.78%, respectively. Under recurrent congested condition (Scenario 2), the DMAM generates a significantly high deviation that can be attributed to its high MARE.

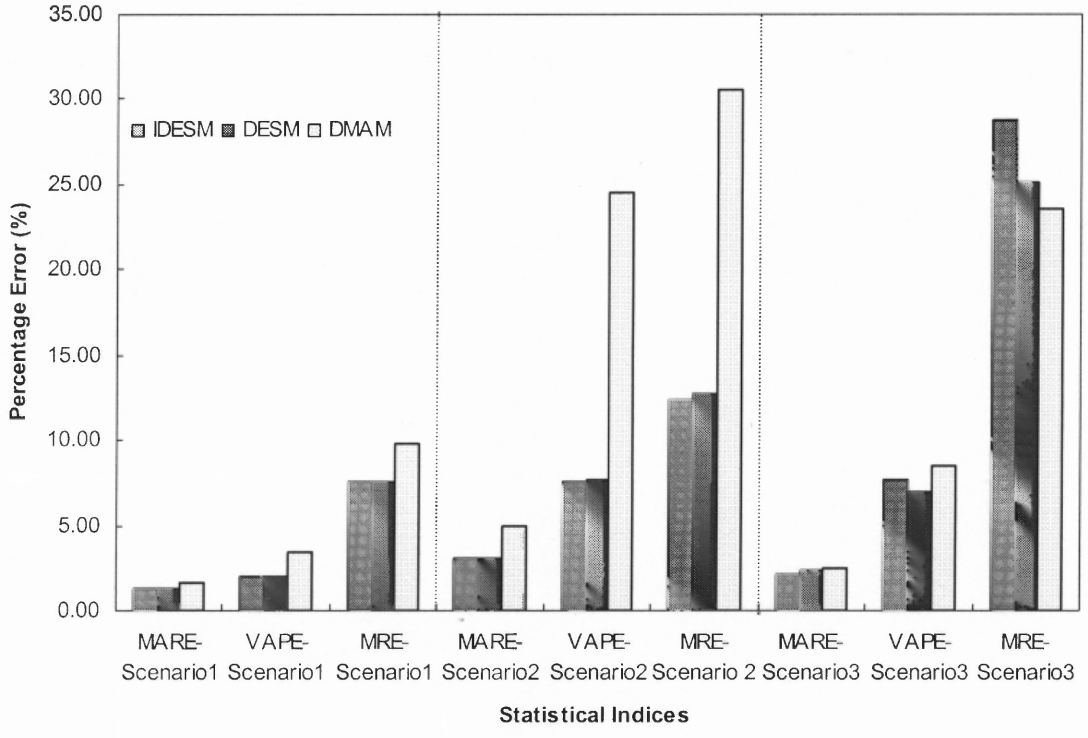

Figure 6.13 Prediction Accuracy of IDESM, DESM, and DMAM (5:00am-11:00pm).

Table 6.10 Prediction Accuracy of IDESM, DESM, and DMAM.

Scenario	Period	MARE			VAPE			MRE		
		IDESM	DESM	DMAM	IDESM	DESM	DMAM	IDESM	DESM	DMAM
1	Overall	1.29	1.35	1.62	1.95	2.06	3.43	7.59	7.59	9.78
2	1	0.57	0.57	0.80	0.25	0.25	0.38	1.85	1.85	2.26
	2	**2.21**	**2.19**	**7.99**	2.79	2.49	64.36	6.00	6.00	30.49
	3	3.76	3.84	5.03	8.30	8.46	14.43	12.36	12.75	17.02
	Overall	3.06	3.11	4.95	7.56	7.72	24.57	12.36	12.75	30.49
3	1	0.46	1.11	0.69	0.05	0.41	0.44	1.03	2.64	2.76
	2	**4.23**	**3.89**	**4.83**	38.29	28.96	34.82	28.70	25.14	23.57
	3	1.94	2.28	2.38	3.44	4.25	4.72	9.93	11.44	11.36
	Overall	2.08	2.37	2.53	7.64	6.97	8.49	28.70	25.14	23.57

Under the non-recurrent congested condition (Scenario 3), the DMAM demonstrates promising performance in terms of accuracy and stability compared with the IDESM and DESM. It is suggested that the DMAM may not be the first choice to be used under recurrent congested condition, but the DMAM could be an alternative approach when there is no historical data for prediction development, especially under a non-recurrent congested condition.

During the overall evaluation period, the developed IDESM is the best among the three dynamic recursive models due to its least MARE. The effectiveness of IDESM can be further demonstrated by its low VAPE in most of the TPs. Even in TP 2 of Scenarios 2 and 3, the DESM can generate slightly better prediction results than the IDESM. The consistent superior performance of the IDESM under various traffic conditions proves that the IDESM becomes the prediction model to be recommended.

6.4 Evaluation of the Prediction Interval

6.4.1 Introduction

It was indicated in the literature review that the stochastic traffic characteristics may significantly affect the prediction accuracy. Various traffic scenarios were designed in the case study to test and evaluate the performance of the developed prediction models. A sensitivity analyses using different prediction intervals is performed to evaluate the impacts of travel time variation to the prediction accuracy. Traffic data can be collected during different prediction time intervals.

In this study, travel time information is collected under varying intervals lengths. The optimal time interval which produces results with minimum error can be identified. The analysis of interval duration and its impact to prediction accuracy is discussed next.

6.4.2 DESM

Travel time information used in the DESM collected during prediction intervals of different lengths (i.e., 5 minutes, 10 minutes and 15 minutes). The indices are categorized by the individual and overall evaluation periods in Table 6.11, while Figure 6.14 shows the overall prediction accuracy for various intervals. Figures 6.15, 6.16, and 6.17 show the prediction error of the DESM with different prediction intervals in Scenarios 1, 2, and 3, respectively.

Table 6.11 Performance Indices with Various Intervals of DESM.

Scenario	Interval	5 min			10 min			15 min		
	Indices	MARE	VAPE	MRE	MARE	VAPE	MRE	MARE	VAPE	MRE
1	Overall	1.32	2.02	7.58	0.83	0.41	3.36	0.67	0.29	2.55
2	1	0.57	0.25	1.85	0.46	0.08	0.96	0.25	0.04	0.58
	2	2.19	2.49	6.00	1.80	2.46	4.48	2.15	5.41	7.30
	3	3.84	8.46	12.75	1.86	2.03	6.12	1.56	1.16	3.92
	Overall	3.11	7.72	12.75	1.66	2.04	6.12	1.49	1.94	7.30
3	1	1.11	0.41	2.64	0.35	0.04	0.68	0.31	0.02	0.47
	2	**3.89**	28.96	25.14	**7.68**	42.82	20.20	**10.31**	45.31	19.40
	3	2.28	4.25	11.44	1.49	1.51	5.10	0.88	0.53	3.72
	Overall	2.37	6.97	25.14	2.10	9.73	20.20	1.90	14.09	19.40

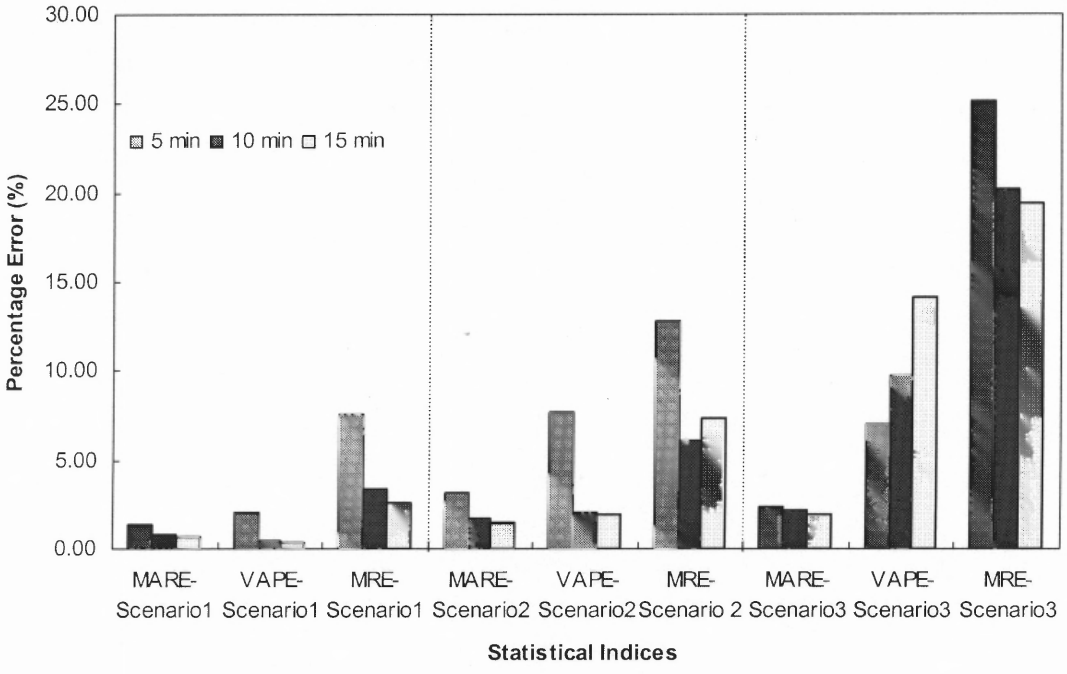

Figure 6.14 Prediction Accuracy of DESM with Various Intervals.

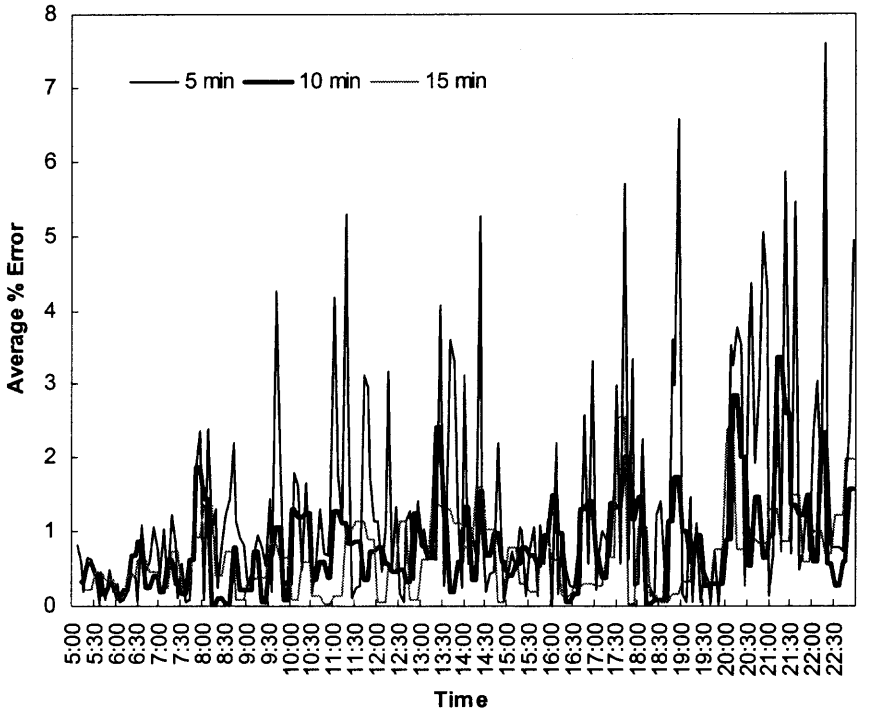

Figure 6.15 Prediction Error of DESM (Scenario 1).

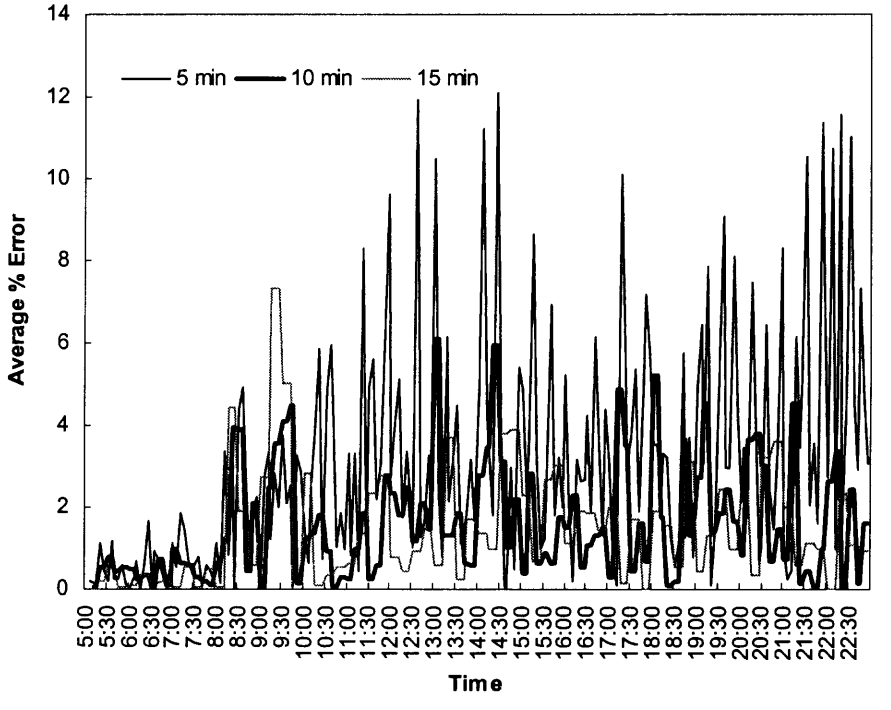

Figure 6.16 Prediction Error of DESM (Scenario 2).

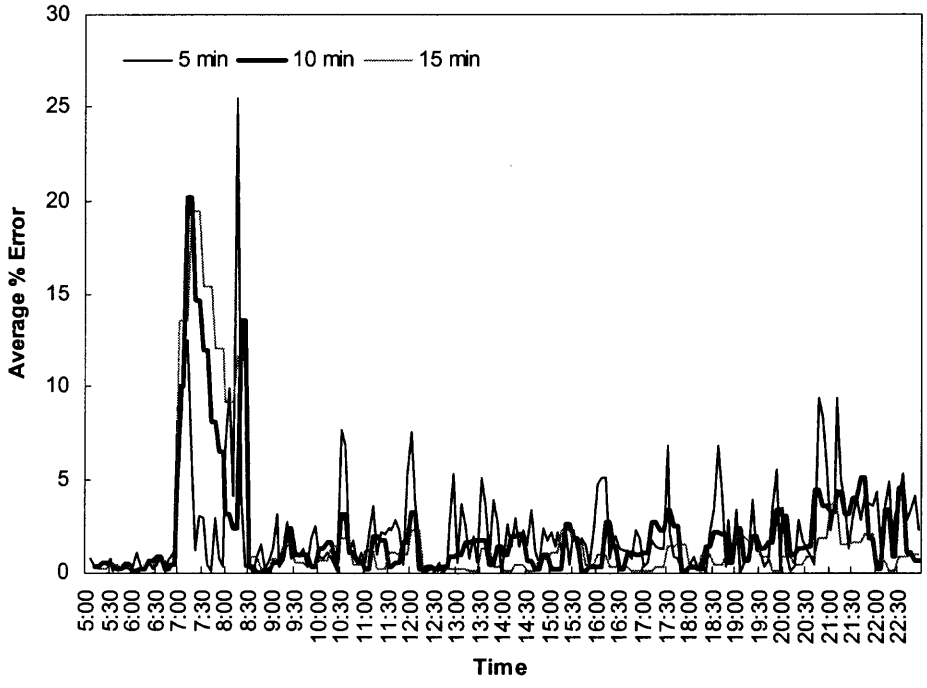

Figure 6.17 Prediction Error of DESM (Scenario 3).

It can be observed in Figure 6.14 that the increased prediction interval reduces the travel time variation and improves the prediction accuracy in the overall evaluation period under the three scenarios. The stability of the prediction results is also improved, in general, as the duration of prediction interval increases. However, in TP 2 of Scenario 3 when a non-recurrent event occurs, the reverse result has been observed.

Figure 6.15 indicates that the prediction error decreases as the duration of the prediction interval increases in Scenario 1. By observing the DESM prediction MOEs of Scenario 2 in Table 6.11, it is found that with shorter time interval data (e.g., 5-minute), the MARE of the DESM in TP 2 is 2.15% but changes to be 1.80% for a 10-minute interval and 2.15% for a 15-minute interval. This indicates that the optimal prediction interval is 10-minute.

The MARE of the DESM in TP 2 in Table 6.11 significantly decreases when the data collected during a shorter prediction interval are used. This can be easily seen in Figure 6.17, where the boundary line of the prediction error with 5-minute interval data falls under the range of that with 10-minute interval. It can be concluded that the traffic data with longer interval could reduce the DESM's accuracy in Scenario 3.

As highlighted in Table 6.11, the MARE of 5-minute data is 3.89%, while the MARE of 15-minute data is 10.31% in TP 2. To investigate the reason that causes this difference, Figure 6.18 is generated to analyze the actual and predicted travel times with 5-minute and 15-minute intervals.

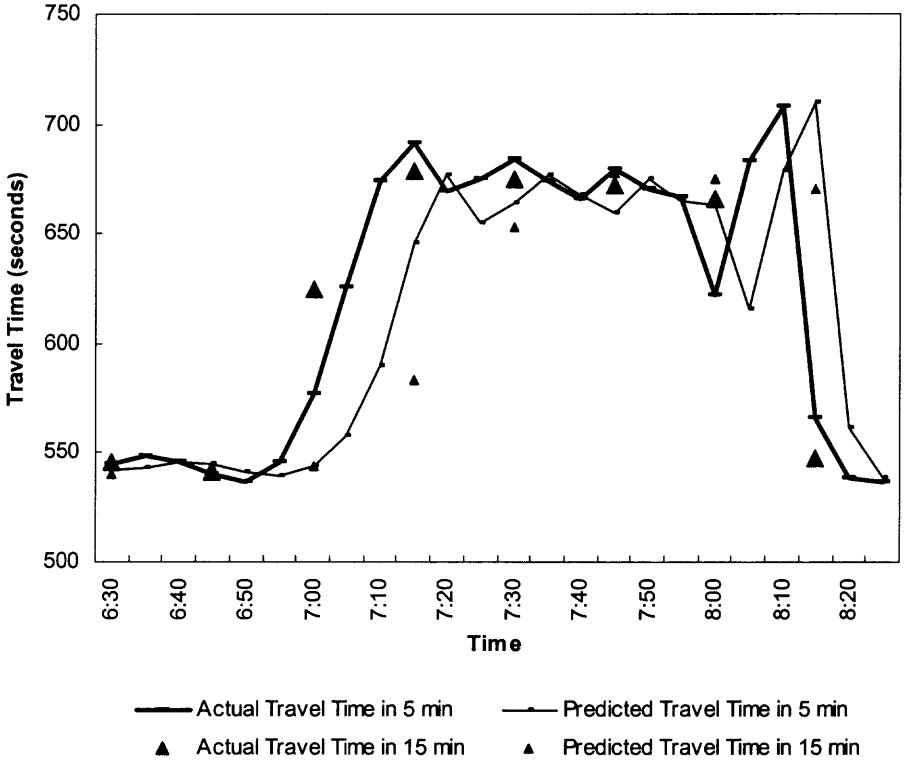

Figure 6.18 Predicted Travel Time of DESM (Scenario 3).

Figure 6.18 shows that from 7:00 am to 7:20 am, the traffic trend changed gradually that can be monitored clearly with 5-minute interval data, while with 15-minute interval, the collected data changes abruptly without any obvious trend. Thus, under non-recurrent congested condition it is preferred to use the shorter time interval data when dramatic traffic change occurs.

6.4.3 IDESM

After the travel times with various prediction intervals are used in the IDESM, the statistical indices were categorized and shown in Figure 6.19 and Table 6.12. Similar performance indices for the IDESM can be obtained. The prediction errors of the IDESM with various prediction intervals used in Scenarios 1, 2, and 3 are slightly different from those of the DESM. Therefore, the impact of the prediction interval to the prediction accuracy of the IDESM as well as the prediction errors in each scenario are similar to those in the DESM, which were already presented in the Section 6.4.2.

Table 6.12 Performance Indices with Various Intervals of IDESM.

Scenario	Interval	5 min			10 min			15 min		
	Indices	MARE	VAPE	MRE	MARE	VAPE	MRE	MARE	VAPE	MRE
1	Overall	1.29	1.95	7.59	0.83	0.41	3.36	0.67	0.29	2.55
2	1	0.57	0.25	1.85	0.46	0.08	0.96	0.25	0.04	0.58
	2	2.21	2.79	6.00	1.78	2.51	4.48	2.14	5.73	7.30
	3	3.76	8.30	12.36	1.78	2.09	6.09	1.56	1.16	3.91
	Overall	3.06	7.56	12.36	1.61	2.07	6.09	1.49	1.99	7.30
3	1	0.46	0.05	1.03	0.35	0.04	0.68	0.31	0.02	0.47
	2	4.23	38.29	28.70	7.68	42.82	20.20	10.31	45.31	19.40
	3	1.94	3.44	9.93	1.49	1.51	5.10	0.88	0.53	3.72
	Overall	2.08	7.64	28.70	2.10	9.73	20.20	1.90	14.09	19.40

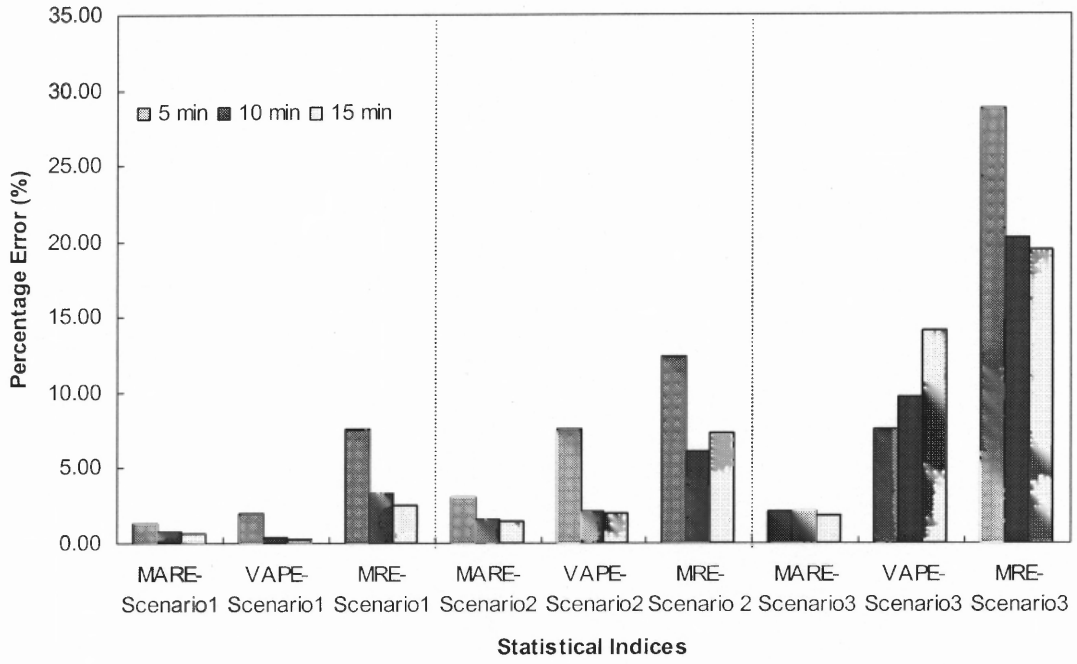

Figure 6.19 Prediction Accuracy of IDESM with Various Intervals.

6.4.4 DMAM

Travel times with various prediction intervals were used in the DMAM, the statistical indices were calculated as shown in Figure 6.20 and Table 6.13.

Figure 6.20 shows the overall performance of the DMAM with various time intervals for the three traffic scenarios. Figures 6.21, 6.22, and 6.23 show the prediction errors of the DMAM with different prediction intervals for all Scenarios.

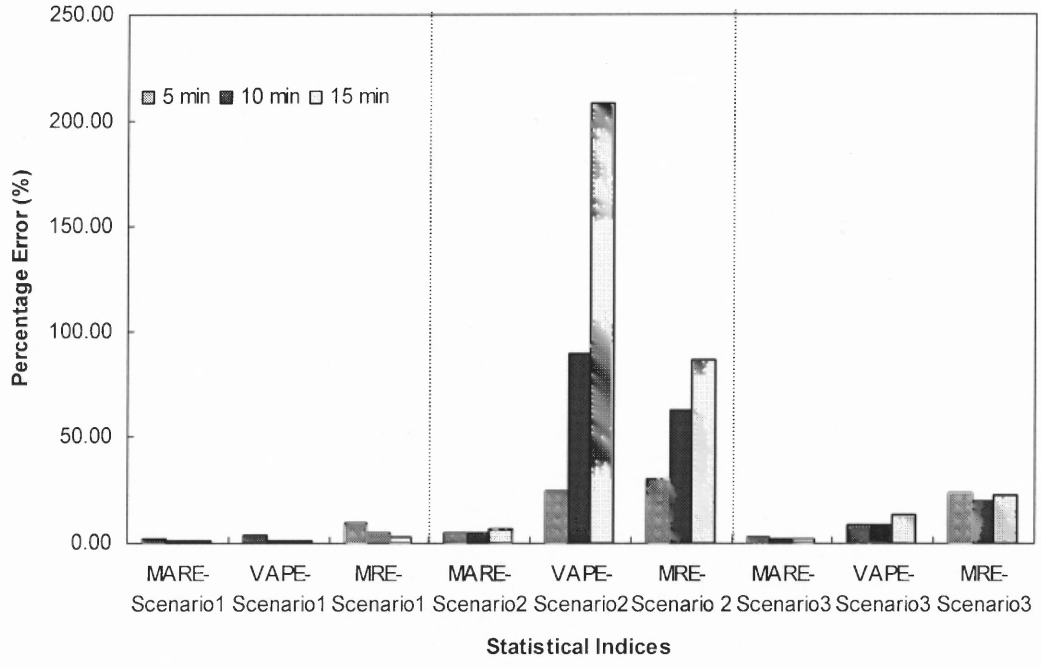

Figure 6.20 Prediction Accuracy of DMAM with Various Intervals.

Table 6.13 Prediction Accuracy with Various Intervals of DMAM.

Scenario	Interval	5 min			10 min			15 min		
	Indices	MARE	VAPE	MRE	MARE	VAPE	MRE	MARE	VAPE	MRE
1	Overall	1.62	3.43	9.78	1.07	1.02	4.87	0.91	0.53	2.84
2	1	0.80	0.38	2.26	0.81	0.58	2.62	1.05	1.07	2.85
	2	7.99	64.36	30.49	17.45	332.55	63.21	27.54	704.83	86.29
	3	5.03	14.43	17.02	2.59	3.93	7.66	2.12	2.23	6.56
	Overall	4.95	24.57	30.49	4.90	89.54	63.21	6.36	208.37	86.29
3	1	0.69	0.44	2.76	0.84	0.89	2.49	1.30	1.04	2.43
	2	4.83	34.82	23.57	5.10	44.74	19.88	7.06	68.79	22.36
	3	2.38	4.72	11.36	1.63	2.55	8.67	1.23	3.11	12.84
	Overall	2.53	8.49	23.57	1.97	8.13	19.88	1.90	13.08	22.36

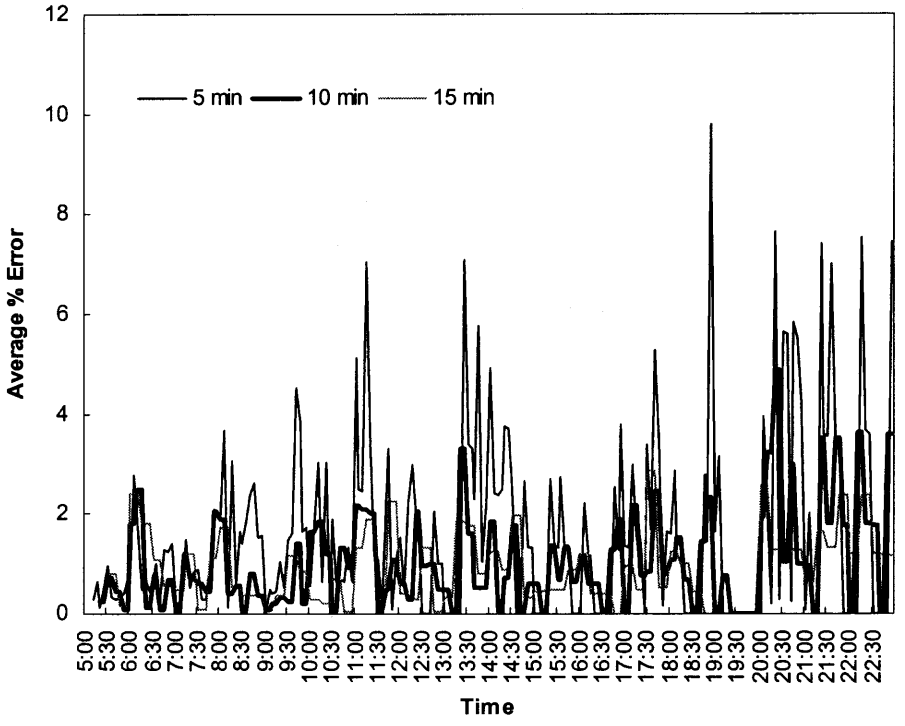

Figure 6.21 Prediction Error of DMAM (Scenario 1).

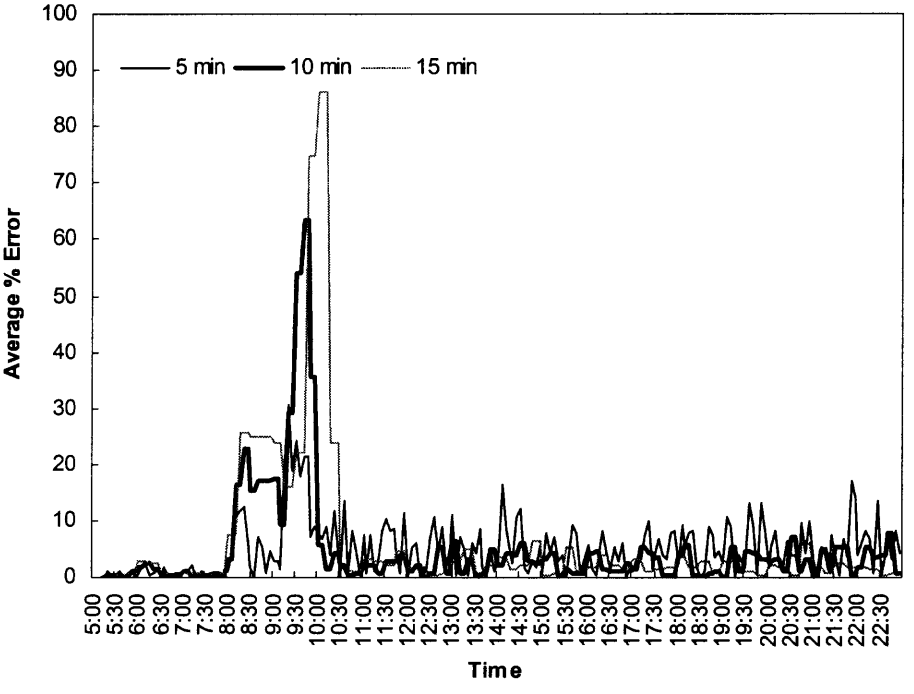

Figure 6.22 Prediction Error of DMAM (Scenario 2).

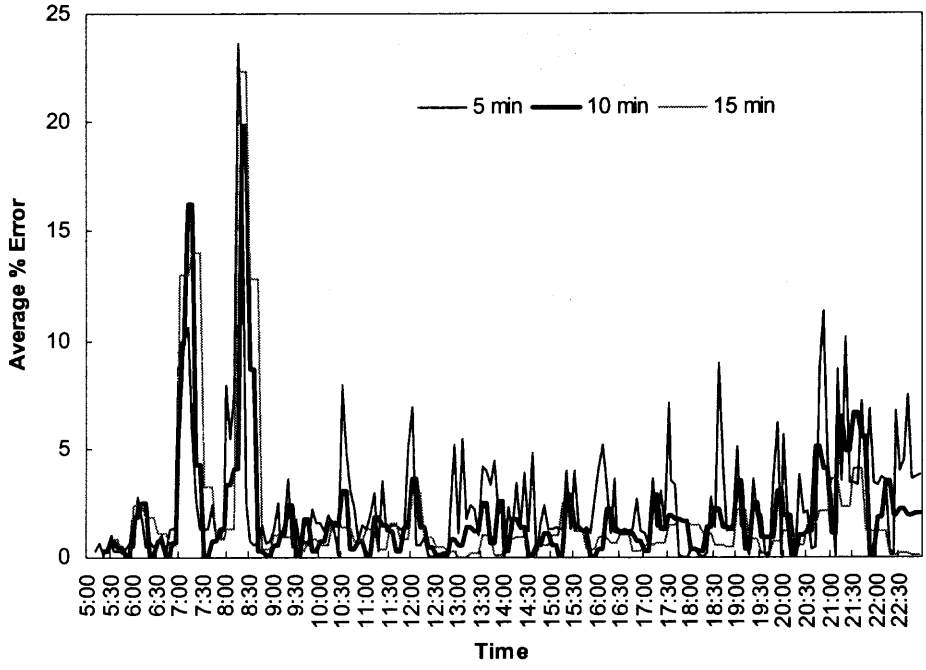

Figure 6.23 Prediction Error of DMAM (Scenario 3).

It can be seen in Figure 6.20 that the MARE decreases as the length of the prediction interval increases. However, in Scenario 2 the increased prediction interval also increases the VAPE and MRE. This indicates that increasing the prediction interval may cause a significant reduction on the stability of the prediction performance for Scenario 2.

It can be seen from Table 6.12 that the MOEs keep decreasing when the prediction interval becomes longer for Scenario 1. It means that in a free-flow condition, a longer prediction interval can improve the DMAM accuracy and reduce the variation of the prediction results.

In TP 2 under Scenarios 2 and 3, the MAREs increase as the prediction interval increases. This indicates that the traffic data collected from a longer prediction interval significantly reduce the prediction accuracy when recurrent and non-recurrent traffic

congestion occurs. In Figures 6.22 and 6.23, the boundary lines of the prediction error of the 10-minute or 15-minute interval data are out of the range of the 5-minute interval data.

In general, increasing the prediction interval could eliminate unstable traffic disturbances and reduce traffic variation, and in turn it can improve the prediction accuracy. However, it is not appropriate for the transition period of Scenarios 2 and 3. To investigate the reason for the higher prediction errors due to longer prediction intervals, Figure 6.20 shows the actual and predicted travel times with 5-minute and 15-minute intervals for the traffic transition period of Scenario 2.

Figure 6.24 plots the actual and predicted traffic time with different prediction intervals in TP 2 during which traffic experiences recurrent congested condition. Under such a condition, the traffic flow is in an unstable state (i.e., non-stationary traffic) during the congestion period (i.e., peak hour in this case). If the prediction time interval is too long, the abrupt travel time change may fall in the middle of an interval. Thus, the measurement may be meaningless. As shown in Figure 6.24, the travel time increased continuously from 8:00 am to 9:00 am and reached the highest value at time 9:10am and then started to decrease. However, with a 15-minute interval, the collected travel time cannot capture the travel time "turning point" at 9:10 am because that moment falls in one of its intervals (9:00am to 9:15am). Therefore, the predicted travel time cannot monitor the traffic changes instantly, and most likely this causes the prediction deviation.

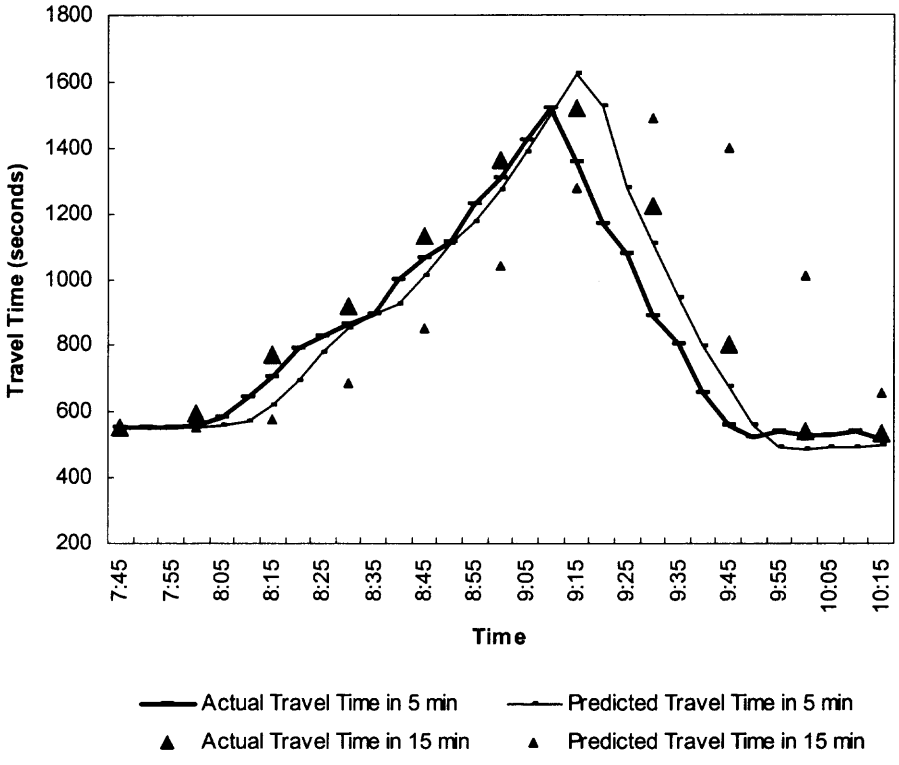

Figure 6.24 Predicted Travel Time of DMAM in Scenario 2.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

Predicting short-term travel time information on freeways has been desired for ATIS and many other ITS applications (e.g., dynamic traffic assignment, ramp metering control). It can essentially affect pre-trip planning and en-route optimization to mitigate traffic congestion and improve transportation network efficiency. In this study, three dynamic recursive models (i.e. DESM, IDESM, and DMAM) were developed based on the three wildly used base models (i.e. ESM, MAM, and KFM) and the performance of each model was analyzed.

Subject to the limitation of traffic information collected from the studied site, a calibrated and validated simulation model developed with CORSIM was used to generate travel times for the study network. Three traffic scenarios (i.e. free-flow, recurrent congestion and non-recurrent congestion) were designed and used to test and evaluate the performance of the developed models.

7.1 Conclusions

The major findings of the study are as follows:

- The ESM was proven to be a potential approach for travel time prediction under various traffic conditions. The numerical test of its performance with different values of α indicated that the selection of an α value substantially affected its prediction accuracy. The optimal smoothing constant α was optimized based on a grid search by employing empirical data collected at the study. The optimal α could achieve good performance when the historical traffic condition is similar to the future condition. However, if the traffic experienced unexpected changes, the predicted travel times from the ESM showed significant deviation. This indicated that the static weight parameter embedded in the ESM is the primary reason for reducing its accuracy when the condition changes.

- The IDESM or DESM was developed for travel time prediction with, the combined features of the ESM and KFM. The time-varying smoothing parameter α_t is optimized based on the most recent observation (i.e. prediction error). The benefit from employing an optimized α_t has been verified. It was found that the MAREs under all three scenarios have been decreased when compared with those generated by the ESM. This indicated that the reason why the IDESM and DESM could outperform the ESM. A reduced VAPE also demonstrated an improved stability for the prediction results when using the IDESM and DESM.

- Unlike the static weight parameter in the ESM, the weight parameter α_t in IDESM and DESM could dynamically reflect traffic changes in various traffic conditions. In addition, the IDESM and DESM used an optimized weight parameter α_t based on the most recent traffic observation. Therefore, the developed modeling technique can be applied at different locations, illustrating the transferability of the IDESM and DESM for travel time prediction application.

- In the IDESM, the range limit of the weight parameter is (i.e. $\alpha_t \in [0,1]$). While in DESM it is $\alpha_t \in (-\infty, +\infty)$. This enables the IDESM to predict traffic change in a more effective way. However, when the underlying assumptions (i.e. normal distribution and linear state variable) were violated at certain circumstance (i.e. traffic congestion period), the α_t generated by IDESM was not optimal for that traffic situation. Therefore, the application scope for the IDESM and DESM should be verified before used to assure the successful implementation of the IDESM.

- The DMAM was developed for travel time prediction with the combined features of the MAM and KFM, while θ_t was optimized at each interval based on the most recent observation. The DMAM doesn't show a superior performance over MAM because assuming equal impacts from N previous travel times in DMAM may not reflect the actual traffic flow relationship. Though the MAM and DMAM don't perform well in a recurrent congested condition, they can predict travel time with comparable accurate levels as the DESM and with better accuracy than the base KFM in a non-recurrent congested condition. More importantly, when there are no historical data available, the DAM and DMAM could be recommended as alternative approaches for travel time prediction.

- A sensitivity analysis of the prediction interval was conducted to assess the impacts of different interval durations on the prediction accuracy. The evaluation of the indicated results illustrated that increasing the prediction interval could eliminate unstable disturbances of traffic flow and improve the prediction accuracy in general. However, this is not true when traffic experiences some dramatic changes such as during a traffic congestion condition. Therefore, the distribution type and variation level of travel time should be considered before choosing the most appropriate prediction interval.

7.2 Future Study

Future studies about the developed dynamic recursive prediction models may be focusing on the following aspects: (1) Data, (2) Parameter Selection, (3) Variability Identification. The future extension of each of the above aspects is discussed below.

Data

- In this study, the link-based travel time was calculated by adding up the travel times of vehicles over the links that constitute the path in each time interval simultaneously. It is desirable to apply path-based data (Chen and Chien, 2002) as a substitute to the link-based data for travel time prediction applications. In addition, the travel time correlation between successive links should be considered, because it can be more accurately representing actual travel time in a real world network.

- The actual travel times should be collected from the field to replace the simulated data for further validation of the developed models. While using real world data, the sample size should be chosen in a way that assure that the collected data represent can reflect the actual travel time considering the impact on many factors of prediction accuracy, such as prediction interval and confident interval of the predicted travel time.

Parameter Selection

- The normal distributions of weight parameters α_t and θ_t should be tested while optimizing α_t and θ_t in the IDESM, DESM, and DMAM. Unlike travel time, these weight parameters cannot be directly measured from the collected data. It is desirable to develop a method that can identify the underlying weight parameter distribution under various traffic conditions.

- The linear relationship between successive state variables (i.e. weight parameter) should also be verified to validate the developed dynamic models (i.e. IDESM, DESM and DMAM). When traffic experiences conditions with a non-linear relationship between the weight parameters, it is necessary to transform the non-linear parameter into a linear relationship to ensure that the optimized weight achieves the minimum square error before making each prediction.

- The system internal parameters such as the covariance of state noise Q_t and covariance of observation noise R_t contribute significantly to the prediction accuracy shown in Figure 6.3 In this study, a sensitivity analysis was conducted within a recommended range for Q_t and R_t were recommended. It is desirable to

develop an effective method that can optimize Q_t and R_t to reflect the actual error covariance for the study network.

- A suggestion for an optimal prediction interval was made based on a numerical analysis. This was a qualitative index that cannot provide a feasible procedure for implementation. Considering the distribution type and variation level of travel time, a quantitative index is desirable to determine the optimal prediction interval and improved the prediction accuracy.

Variability Identification

- The prediction model should generate not only the mean value of predicted travel time, but also provide the variance of the predicted travel time. With a confidence interval of the predicted results, it is more reliable and effective to use the predicted travel times in ATIS applications, such as a route guidance system.

- The developed prediction models demonstrated their strength in different application scenarios (e.g., free-flow congested condition, congested traffic condition). For example, in the congestion period, DESM may have superior capability for travel time prediction than IDESM. It would be desirable to develop a hybrid prediction system (Chien and Mouly, 2003) to adapt to various traffic conditions.

APPENDIX A

PATH TRAVEL TIME DERIVATION

The derivation for path travel time from the CORSIM simulation model is discussed in this Appendix. In the CORSIM output file, the travel time on each link is generated. But the generated travel time in each statistics report is a cumulative travel time from the onset of the simulation to the current time interval. It is necessary to retrieve the individual travel time in each prediction interval, and then the path travel time could be calculated by adding up the link travel times.

The following chart shows the data flow of the algorithm used to derive path travel times. The original code programmed in Java is also enclosed.

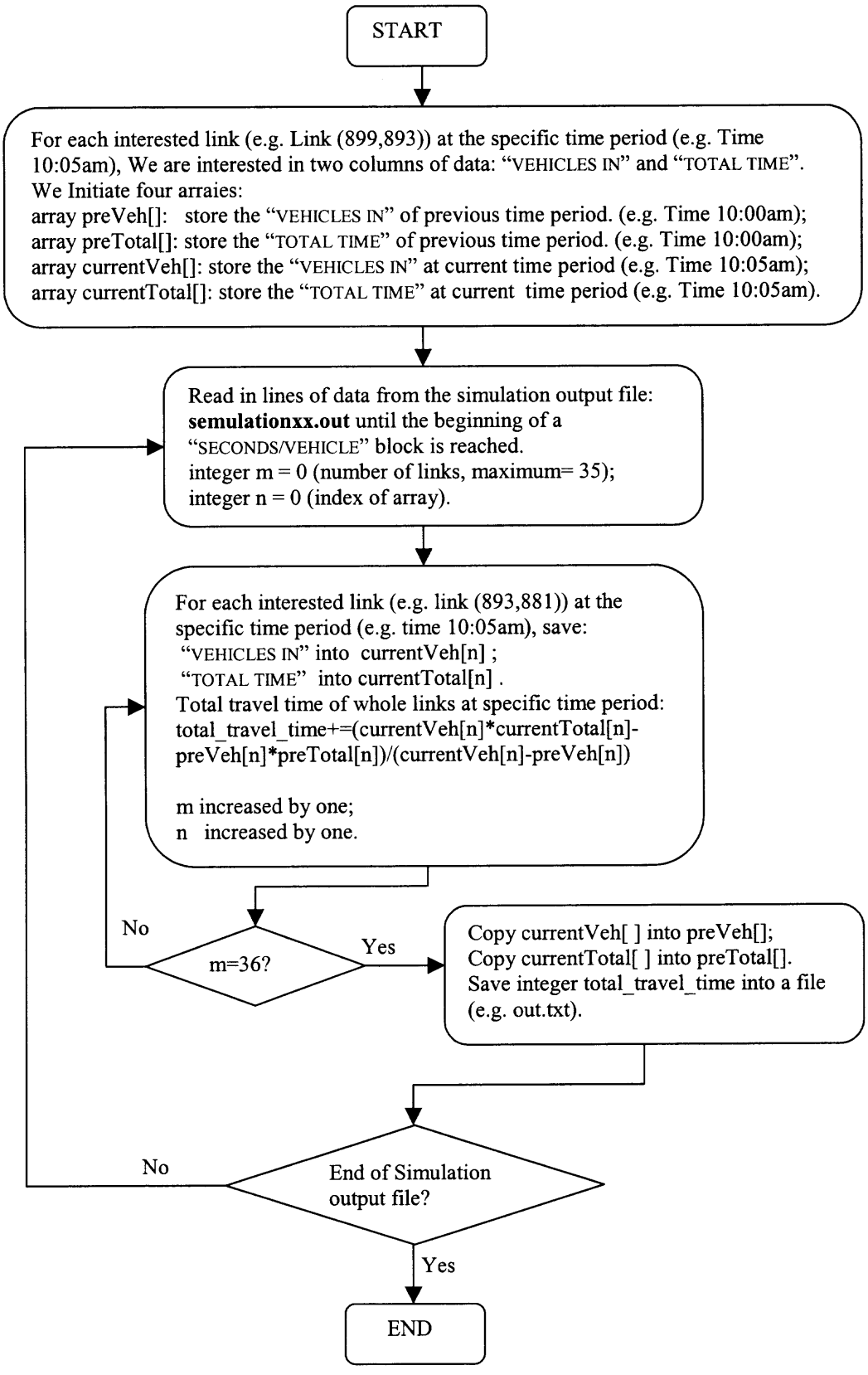

Figure A-1 Data Flow Chart for the Derivation of Path Travel Time.

JAVA PROGRAM: TRAVELTIME.JAVA

Used for retrieving path travel time from simulation output file.

```
import java.io.*;
import java.util.*;

public class TravelTime {

public static void main(String[] args) throws IOException {

System.out.print("please enter your filename:");
BufferedReader br=new BufferedReader(new InputStreamReader(System.in));
String filename=null;
filename=br.readLine();
System.out.println("Processing file: "+filename);

File infile=new File(filename);
File outfile=new File("out.txt");

FileReader a=new FileReader(infile);
FileWriter b=new FileWriter(outfile);

BufferedReader aa=new BufferedReader(a);
BufferedWriter bb=new BufferedWriter(b);

String beginline="  LINK     IN  OUT CHNG CONT CONT MILES   MIN   TIME  TIME TIME
M/T  TOTAL DELAY VEH/LN/HR VEH/LN-MILE  MILE/HR  TYPE";

String result=aa.readLine();

int lines=36;
double preVeh[]=new double[lines];
double preTotal[]=new double[lines];
double currentVeh[]=new double[lines];
double  currentTotal[]=new double[lines];

for(int i=0;i<lines;i++) {
preVeh[i]=0.0;
preTotal[i]=0.0;
currentVeh[i]=0.0;
currentTotal[i]=0.0;
}

while(result!=null) {

if(result.equals(beginline)) {

        aa.readLine();aa.readLine();

        double total_time=0;
        for(int i=0;i<lines;i++) {
        result=aa.readLine();
```

```
        String A1=result.substring(11,17);
        String A2=result.substring(55,63);

        currentVeh[i]=Double.parseDouble(A1);
        currentTotal[i]=Double.parseDouble(A2);

        total_time +=(currentVeh[i]*currentTotal[i]-
        preVeh[i]*preTotal[i])/(currentVeh[i]-preVeh[i]);

        aa.readLine();
        }

        bb.write(""+total_time);
        bb.newLine();

        for(int i=0;i<lines;i++) {
        preVeh[i]=currentVeh[i];
        preTotal[i]=currentTotal[i];
        }
    }

result=aa.readLine();
}

aa.close();
bb.close();
System.out.println("Successful! go to c:\\java\\out.txt for your output");
}
}
```

APPENDIX B

DERIVATION OF KALMAN FILTERING MODEL

The derivation for Kalman Filtering Model (KFM) is introduced in this Appendix based on the study by Grewal and Andrews (1993). In the proposed KFM, x_t denotes the travel time to be predicted at time interval t, ϕ_t denotes the transition parameter at time interval t which is externally determined, and w_t denotes a noise term that has a normal distribution with zero mean and a variance of Q_t, the covariance of the state variable. Let z_t denote the observation of travel time in time interval t, and v_t denotes the measurement error at time interval t that has a normal distribution with zero mean and a variance of R_t, the covariance of the measurement variable. The system model can be formulated as

Measurement model: $z_t = Hx_t + v_t$ (B.1)

System dynamic model: $x_t = \phi_{t-1}x_{t-1} + \omega_{t-1}$ (B.2)

The objective of the developed KFM is to find an estimate of the n state vector x, represented by $\hat{x}_t$, a linear function of the measurements z_i, ..., z_t, that minimizes the weighted mean-squared error.

The observation update problem for a system state estimator

Suppose that a measurement of travel time has been made at time t, and that the information that it provides is to be used in updating the estimate of the state x of a stochastic system at time t. It is assumed that the measurement variable (e.g., travel time) is linearly related to the state variable (i.e., travel time) by an Equation B.1

$z_t = Hx_t + v_t$ (B.1)

where H is the measurement sensitivity matrix and v_t is the measurement noise. The measurement sensitivity reflects the linear relationship between state and measurement variables. Its value is the derivative of measurement variable z_t with respect to the state variable x_t. Since both z_t and x_t are travel time in KFM, H =1.

Estimator in linear form

The optimal linear estimate is equivalent to the general optimal estimator if the variates x and z are jointly Gaussian. Therefore it suffices to seek an updated estimate $\hat{x}_t^+$ based on the observation z_t that is a linear function of the a priori estimate and the measurement z:

$$\hat{x}_t^+ = K_t^1 \hat{x}_t^- + \overline{K}_t z_t \tag{B.3}$$

where $\hat{x}_t^-$ is the a priori estimate of x_t and $\hat{x}_t^+$ is the posteriori value of the estimate.

Optimization problem

The matrices K_t^1 and $\overline{K}_t$ are as yet unknown. We seek those values of K_t^1 and $\overline{K}_t$ such that the new estimate $\hat{x}_t^+$ will satisfy the orthogonality principle. This orthogonality condition can be written in the form

$$E\langle [x_t - \hat{x}_t^+] z_i^T \rangle = 0, \ i = 1,2, \ldots, t\text{-}1 \tag{B.4}$$

$$E\langle [x_t - \hat{x}_t^+] z_t^T \rangle = 0 \tag{B.5}$$

If one substitutes the formula for x_t from Equation B.2 and for $\hat{x}_t^+$ from Equation B.3 into Equation B.4, then one will observe from Equation B.1 and B.2 that the data $z_1, \ldots, z_t$ do not involve the noise term w_t. Therefore, because the random sequences w_t and v_t are uncorrelated, it follows that

$E\, w_t\, z_i^T = 0$ for $1 \le i \le t$. (Independent assumption)

Using this result, one can obtain the following relation:

$$E\langle \Phi_{t-1} x_{t-1} + w_{t-1} - K_t^1 \hat{x}_t^- - \overline{K}_t z_t \rangle z_i^T = 0, \ i = 1, \ldots, t\text{-}1 \tag{B.6}$$

But because $z_t = H_t z_t + v_t$, Equation B.6 can be rewritten as

$$E[\Phi_{t-1} x_{t-1} - K_t^1 \hat{x}_t^- - \overline{K}_t H_t x_t - \overline{K}_t v_t] z_i^T = 0, \ i = 1, \ldots, t\text{-}1 \tag{B.7}$$

We also know that Equation B.5 holds at the previous step, i.e.,

$E\langle [x_{t-1} - \hat{x}_{t-1}^+] z_i^T \rangle = 0, \ i=1,\ldots,t\text{-}1$ and

$E\langle v_t z_i^T \rangle = 0, \ i=1,\ldots,t\text{-}1.$

Then Equation B.7 can be reduced to the form

$$\Phi_{t-1}\,\mathrm{E}\,x_{t-1}\,z_i^T - K_t^1\,\mathrm{E}\,\hat{x}_t^-\,z_i^T - \overline{K}_t\,H_t\,\Phi_{t-1}\,\mathrm{E}\,x_{t-1}\,z_i^T - \overline{K}_t\,\mathrm{E}\,v_t\,z_i^T = 0$$

$$\Phi_{t-1}\,\mathrm{E}\,x_{t-1}\,z_i^T - K_t^1\,\mathrm{E}\,\hat{x}_t^-\,z_i^T - \overline{K}_t\,H_t\,\Phi_{t-1}\,\mathrm{E}\,x_{t-1}\,z_i^T = 0$$

$$\mathrm{E}\langle[\,x_t - \overline{K}_t\,H_t\,x_t - K_t^1\,x_t\,] - K_t^1\,\hat{x}_t^- - x_t)\rangle\,z_i^T = 0$$

$$[\mathrm{I} - K_t^1 - \overline{K}_t\,H_t]\,\mathrm{E}\,x_t\,z_i^T = 0. \tag{B.8}$$

Equation B.8 can be satisfied for any given x_k if

$$K_t^1 = \mathrm{I} - \overline{K}_t\,H_t. \tag{B.9}$$

Clearly, this choice of K_t^1 causes Equation B.3 to satisfy a portion of the condition given by Equation B.4. $\overline{K}_t$ is chosen such that Equation B.5 is satisfied. Let

$$\tilde{x}_t^+ \triangleq \hat{x}_t^+ - x_t \tag{B.10}$$

$$\tilde{x}_t^- \triangleq \hat{x}_t^- - x_t \tag{B.11}$$

$$\tilde{z}_t \triangleq \hat{z}_t^- - z_t$$

$$= H_t\,\hat{x}_t^- - z_t. \tag{B.12}$$

Vectors $\tilde{x}_t^+$ and $\tilde{x}_t^-$ are the estimation errors after and before updates, respectively. From Equation B.5

$$\mathrm{E}[\,x_t - \hat{x}_t^+\,]\,\hat{z}_t^T = 0 \tag{B.13}$$

and also (by subtracting Equation B.1 from Equation B.13)

$$\mathrm{E}[\,x_t - \hat{x}_t^+\,]\,\tilde{z}_t^T = 0 \tag{B.14}$$

Substitute for $x_t, \hat{x}_t^+$ and $\tilde{z}_t$ from Equation B.2, B.3 and Equation B.12, respectively. Then

$$\mathrm{E}[\,\Phi_{t-1}\,x_{t-1} + w_{t-1} - K_t^1\,\hat{x}_t^- - \overline{K}_t\,z_t\,][\,H_t\,\hat{x}_t^- - z_t\,]^T = 0.$$

However, by the system structure

$$\mathrm{E}\,w_t\,z_t^T = \mathrm{E}\,w_t\,\hat{x}_t^{T+} = 0$$

$$\mathrm{E}[\,\Phi_{t-1}\,x_{t-1} - K_t^1\,\hat{x}_t^- - \overline{K}_t\,z_t\,][\,H_t\,\hat{x}_t^- - z_t\,]^T = 0.$$

Substituting for K_t^1, z_t, and $\tilde{x}_t^-$, and using the fact that $\mathrm{E}\,\tilde{x}_t^-\,v_t^T = 0$, this last result can be modified as follows:

$$0 = E\langle[\Phi_{t-1}\, x_{t-1} - \hat{x}_t^- + \overline{K}_t\, H_t\, \hat{x}_t^- - \overline{K}_t\, H_t\, x_t - \overline{K}_t\, v_t][H_t\, \hat{x}_t^- - H_t\, x_t - v_t]^T\rangle$$

$$= E\langle[(x_t - \hat{x}_t^-) - \overline{K}_t\, H_t\, (x_t - \hat{x}_t^-) - \overline{K}_t\, v_t][H_t\, \tilde{x}_t^- - v_t]^T\rangle$$

$$= E\langle[\tilde{x}_t^- - \overline{K}_t\, H_t\, \tilde{x}_t^- - \overline{K}_t\, v_t][H_t\, \tilde{x}_t^- - v_t]^T\rangle.$$

By definition, the a priori covariance (the error covariance matrix before the update) is

$$P_t^- = E\langle \tilde{x}_t^-\, \tilde{x}_t^{T-}\rangle.$$

It satisfies the equation

$$[I - \overline{K}_t\, H_t]\, P_t^-\, H_t^T - \overline{K}_t\, R_t = 0,$$

and therefore the gain can be expressed as

$$\overline{K}_t = P_t^-\, H_t^T\, [H_t\, P_t^-\, H_t^T + R_t]^{-1}, \tag{B.15}$$

which is the solution we seek for the gain as a function of the a priori covariance.

One can derive a similar formula for the posteriori covariance (the error covariance matrix after the update), which is defined as

$$P_t^+ = E\langle[\tilde{x}_t^+\, \tilde{x}_t^{T+}]\rangle. \tag{B.16}$$

By substituting Equation B.9 into EquationB.3, one obtains the equations

$$\hat{x}_t^+ = (I - \overline{K}_t\, H_t)\, \hat{x}_t^- + \overline{K}_t\, z_t$$

$$\hat{x}_t^+ = \hat{x}_t^- + \overline{K}_t\, [z_t - H_t\, \hat{x}_t^-]. \tag{B.17}$$

Substract x_t from both sides of the latter equation to obtain the equations

$$\hat{x}_t^+ - x_t = \hat{x}_t^- + \overline{K}_t\, H_t\, x_t + \overline{K}_t\, v_t - \overline{K}_t\, H_t\, \hat{x}_t^- - x_t$$

$$\tilde{x}_t^+ = \tilde{x}_t^- - \overline{K}_t\, H_t\, \tilde{x}_t^- + \overline{K}_t\, v_t$$

$$\tilde{x}_t^+ = (I - \overline{K}_t\, H_t)\, \tilde{x}_t^- + \overline{K}_t\, v_t. \tag{B.18}$$

By substituting Equation B.18 into Equation B.16 and noting that $E\,\tilde{x}_t^-\, v_t^T = 0$, one obtains

$$P_t^+ = E\{[I - \overline{K}_t\, H_t]\, \tilde{x}_t^-\, \tilde{x}_t^{T-}\, [I - \overline{K}_t\, H_t]^T + \overline{K}_t\, v_t\, v_t^T\, \overline{K}_t^T\}$$

$$= (I - \overline{K}_t\, H_t)\, P_t^-\, (I - \overline{K}_t\, H_t)^T + \overline{K}_t\, R_t\, \overline{K}_t^T. \tag{B.19}$$

This last equation is the so-called "Joseph form" of the covariance update equation derived by P.D. Joseph. By substituting for $\overline{K}_t$ from Equation B.15, it can be put into the following forms

$$P_t^+ = P_t^- - \overline{K}_t\, H_t\, P_t^- - P_t^-\, H_t^T\, \overline{K}_t^T + \overline{K}_t\, H_t\, P_t^-\, H_t^T\, \overline{K}_t^T + \overline{K}_t\, R_t\, \overline{K}_t^T$$

$$= (\mathrm{I} - \overline{K}_t H_t) P_t^- - P_t^- H_t^T \overline{K}_t^T + \underbrace{\overline{K}_t (H_t P_t^- H_t^T + R_t)}_{P_t^- H_t^T} \overline{K}_t^T$$

$$= (\mathrm{I} - \overline{K}_t H_t) P_t^-, \tag{B.20}$$

the last of which is the one most often used in computation. This implements the effect that conditioning on the measurement has on the covariance matrix of estimation uncertainty.

Error covariance extrapolation

Error covariance extrapolation models the effects of time on the covariance matrix of estimation uncertainty, which is reflected in the a priori values of the covariance and state estimate.

$$P_t^- = \mathrm{E}[\tilde{x}_t^- \tilde{x}_t^{T-}]$$

$$\hat{x}_t^- = \Phi_{t-1} \hat{x}_{t-1}^+, \tag{B.21}$$

respectively. Subtract x_t from both sides of the last equation to obtain the equations

$$\hat{x}_t^- - x_t = \Phi_{t-1} \hat{x}_{t-1}^+ - x_t$$

$$\tilde{x}_t^- = \Phi_{t-1} [\hat{x}_{t-1}^+ - x_{t-1}] w_{t-1}$$

$$= \Phi_{t-1} \tilde{x}_{t-1}^+ - w_{t-1}$$

for the propagation of the estimation error, $\tilde{x}_t$. Postmultiply it by $\tilde{x}_t^{T-}$ (on both sides of the equation) and take the expected values. Use the fact that $\mathrm{E}\,\tilde{x}_{t-1} w_{t-1}^T = 0$ to obtain

$$P_t^- \overset{def}{=} \mathrm{E}[\tilde{x}_t^- \tilde{x}_t^{T-}]$$

$$= \Phi_{t-1} P_{t-1}^+ \Phi^T{}_{t-1} + Q_{t-1}$$

$$= \Phi_{t-1} P_{t-1}^+ \Phi^T{}_{t-1} + Q_{t-1}, \tag{B.22}$$

which gives the a priori value of the covariance matrix of estimation uncertainty as a function of the previous posteriori value.

Summary of Equations for the Discrete-Time Kalman Estimator

The equations derived in the preceding section are summarized as follows.

System dynamic model:

$$x_t = \Phi_{t-1}\, x_{t-1} + w_{t-1}$$

$$w_t \sim \mathrm{N}\ (0,\ Q_t)$$

Measurement model:

$$z_t = H_t\, x_t + v_t$$

$$v_t \sim \mathrm{N}\ (0,\ R_t)$$

Initial conditions:

$$\mathrm{E}\langle x_0 \rangle = \hat{x}_0$$

$$\mathrm{E}\langle \tilde{x}_0\, \tilde{x}_0^T \rangle = P_0$$

Independence assumption:

$$\mathrm{E}\langle w_t\, v_j^T \rangle = 0 \text{ for all } t \neq j$$

State estimate extrapolation:

$$\hat{x}_t^- = \Phi_{t-1}\, \hat{x}_{t-1}^+$$

Error covariance extrapolation:

$$P_t^- = \Phi_{t-1}\, P_{t-1}^+\, \Phi_{t-1}^T + Q_{t-1}$$

State estimate observational update:

$$\hat{x}_t^+ = \hat{x}_t^- + \overline{K}_t [\, z_t - H_t\, \hat{x}_t^- \,]$$

Error covariance update:

$$P_t^+ = [\mathrm{I} - \overline{K}_t\, H_t]\, P_t^-$$

Kalman gain matrix:

$$\overline{K}_t = P_t^-\, H_t^T [\, H_t\, P_t^-\, H_t^T + R_t \,]^{-1}$$

APPENDIX C

DERIVATION OF DYNAMIC EXPONENTIAL SMOOTHING MODEL

The Dynamic Exponential Smoothing Model (DESM) is developed in this Appendix by integrating two components (i.e., ESM and KFM). In the proposed DESM, α_t is the key parameter that combines ESM and KFM into one integrated model. α_t acts not only as a weight parameter in ESM, but also the state variable in KFM. The fundamental prediction model in DESM is formulated in Equation C.0:

$$x_{t,DESM} = \alpha_t x_{t,h} + (1-\alpha_t) x_{t-1} \tag{C.0}$$

where $x_{t,h}$: Historical travel time at time t.

x_{t-1}: Observed travel time at time t-1.

$x_{t,DESM}$: predicted travel time by DESM.

The fundamental models in KFM are formulated as

Measurement model: $z_t = H\alpha_t + v_t$ (C.1)

$$v_t \sim N^{**}(0,\ R_t)$$

where z_t denotes the observation of travel time in time interval t, and v_t denotes the measurement error in time interval t that has a normal distribution with zero mean and a variance of R_t, the covariance of the measurement variable.

H is the measurement sensitivity matrix and v_t is the measurement noise. The measurement sensitivity reflects the linear relationship between state and measurement variables. In the DESM prediction model, the state variable is the weight parameter α_t, and the measurement variable is the travel time z_t. The H value is the derivative of the measurement variable z_t with respect to the state variable α_t:

$$H_t = \frac{\partial z_t}{\partial \hat{\alpha}_t} = x_{t,h} - x_{t-1} \tag{C.2}$$

System model: $\alpha_t = \phi_{t-1}\alpha_{t-1} + \omega_{t-1}$ (C.3)

$w_t \sim$ N **$(0, Q_t)$

In the above ϕ_t denotes the transition parameter between successive α_t at time interval t, which is externally determined. Since there is no linear covariance that can be pre-specified for successive weight parameters, ϕ_t adopts a constant value 1 to assume that the α_t is equal to the updated α_{t-1}^{+} in DESM. W_t denotes a noise term that has a normal distribution with zero mean and a variance of Q_t, the covariance of the state variable.

The objective of the developed DESM is to find an estimate of the n state vector α, represented by $\hat{\alpha}_t$, a linear function of the measurements z_i, ..., z_t, that minimizes the weighted mean-squared error.

The observation update problem for a system state estimator

Suppose that a measurement of travel time has been made at time t, and that the information that it provides is to be used in updating the estimate of the weight parameter α of a stochastic system at time t. It is assumed that the measurement variable (i.e., travel time) is linearly related to the state variable (i.e., weight parameter) by Equation C.1

$$z_t = H\alpha_t + v_t \quad \text{(C.1)}$$

$$H_t = \frac{\partial z_t}{\partial \hat{\alpha}_t} = x_{t,h} - x_{t-1} \quad \text{(C.2)}$$

Estimator in linear form

The optimal linear estimate is equivalent to the general optimal estimator if the variates α and z are jointly Gaussian. Therefore it suffices to seek an updated estimate $\hat{\alpha}_t^{+}$ based on the observation z_t that is a linear function of the a priori estimate and the measurement z:

$$\hat{\alpha}_t^{+} = K_t^1 \hat{\alpha}_t^{-} + \overline{K}_t z_t \quad \text{(C.4)}$$

where $\hat{\alpha}_t^{-}$ is the a priori estimate of α_t and $\hat{\alpha}_t^{+}$ is the posteriori value of the estimate.

Optimization problem

The matrices K_t^1 and $\overline{K}_t$ are as yet unknown. We seek those values of K_t^1 and $\overline{K}_t$ such that the new estimate $\hat{\alpha}_t^+$ will satisfy the orthogonality principle. This orthogonality condition can be written in the form

$$E\langle[\alpha_t - \hat{\alpha}_t^+]\, z_i^T\rangle = 0\,,\ i = 1,2,\ldots,t\text{-}1 \qquad (C.5)$$

$$E\langle[\alpha_t - \hat{\alpha}_t^+]\, z_t^T\rangle = 0 \qquad (C.6)$$

If one substitutes the formula for α_t from Equation C.2 and for $\hat{\alpha}_t(+)$ from Equation C.4 into Equation C.5, then one will observe from Equations C.1 and C.2 that the data $z_1,\ldots,z_t$ do not involve the noise term w_t. Therefore, because the random sequences w_t and v_t are uncorrelated, it follows that

$E\, w_t\, z_i^T = 0$ for $1 \le i \le t$ (Independent assumption)

Using this result, one can obtain the following relation:

$$E\langle\Phi_{t-1}\,\alpha_{t-1} + w_{t-1} - K_t^1\,\hat{\alpha}_t^- - \overline{K}_t\, z_t\rangle z_i^T = 0,\ i = 1,\ldots,t\text{-}1 \qquad (C.7)$$

But because $z_t = H_t\, z_t + v_t$, Equation C.7 can be rewritten as

$$E[\Phi_{t-1}\, x_{t-1} - K_t^1\,\hat{\alpha}_t^- - \overline{K}_t\, H_t\, x_t - \overline{K}_t\, v_t]\, z_i^T = 0,\ i = 1,\ldots,t\text{-}1 \qquad (C.8)$$

We also know that Equation C.6 holds at the previous step, i.e.,

$E\langle[x_{t-1} - \hat{x}_{t-1}^+]\, z_i^T\rangle = 0,\ i=1,\ldots,t\text{-}1$ and

$E\langle v_t\, z_i^T\rangle = 0,\ i=1,\ldots,t\text{-}1$

Then Equation C.8 can be reduced to the form

$$\Phi_{t-1}\,E\alpha_{t-1}\, z_i^T - K_t^1 E\hat{\alpha}_t^-\, z_i^T - \overline{K}_t\, H_t\,\Phi_{t-1}\,E\alpha_{t-1}\, z_i^T - \overline{K}_t E v_t\, z_i^T = 0$$

$$\Phi_{t-1}\,E\alpha_{t-1}\, z_i^T - K_t^1 E\hat{\alpha}_t^-\, z_i^T - \overline{K}_t\, H_t\,\Phi_{t-1}\,E\alpha_{t-1}\, z_i^T = 0$$

$$E\langle[\alpha_t - \overline{K}_t\, H_t\,\alpha_t - K_t^1\,\alpha_t] - K_t^1\,(\hat{\alpha}_t^- - \alpha_t)\rangle z_i^T = 0$$

$$[I - K_t^1 - \overline{K}_t\, H_t]E\alpha_t\, z_i^T = 0 \qquad (C.9)$$

Equation C.9 can be satisfied for any given α_t if

$$K_t^1 = I - \overline{K}_t\, H_t \qquad (C.10)$$

Clearly, this choice of K_t^1 causes Equation C.4 to satisfy a portion of the condition given by Equation C.5. $\overline{K}_t$ is chosen such that Equation C.6 is satisfied. Let

$$\tilde{\alpha}_t^+ \triangleq \hat{\alpha}_t^+ - \alpha_t \tag{C.11}$$

$$\tilde{\alpha}_t^- \triangleq \hat{\alpha}_t^- - \alpha_t \tag{C.12}$$

$$\tilde{z}_t \triangleq \hat{z}_t^- - z_t$$

$$= H_t \hat{\alpha}_t^- - z_t \tag{C.13}$$

Vectors $\tilde{\alpha}_t^+$ and $\tilde{\alpha}_t^-$ are the estimation errors after and before updates, respectively. From Equation C.6

$$\mathrm{E}[\alpha_t - \hat{\alpha}_t^+]\hat{z}_t^T = 0 \tag{C.14}$$

and also (by subtracting Equation C.1 from Equation C.14)

$$\mathrm{E}[\alpha_t - \hat{\alpha}_t^+]\tilde{z}_t^T = 0 \tag{C.15}$$

Substitute for $\alpha_t, \hat{\alpha}_t^+$ and $\tilde{z}_t$ from Equation C.2, C.4 and Equation C.13, respectively. Then

$$\mathrm{E}[\Phi_{t-1}\alpha_{t-1} + w_{t-1} - K_t^1 \hat{\alpha}_t^- - \overline{K}_t z_t][H_t \hat{\alpha}_t^- - z_t]^T = 0$$

However, by the system structure

$$\mathrm{E}\, w_t z_t^T = \mathrm{E}\, w_t \hat{\alpha}_t^{T+} = 0$$

$$\mathrm{E}[\Phi_{t-1}\alpha_{t-1} - K_t^1 \hat{\alpha}_t^- - \overline{K}_t z_t][H_t \hat{\alpha}_t^- - z_t]^T = 0$$

Substituting for K_t^1, z_t, and $\tilde{\alpha}_t^-$, and using the fact that $\mathrm{E}\,\tilde{\alpha}_t^- v_t^T = 0$, this last result can be modified as follows:

$$0 = \mathrm{E}\langle[\Phi_{t-1}\alpha_{t-1} - \hat{\alpha}_t^- + \overline{K}_t H_t \hat{\alpha}_t^- - \overline{K}_t H_t \alpha_t - \overline{K}_t v_t][H_t \hat{\alpha}_t^- - H_t \alpha_t - v_t]^T\rangle$$

$$= \mathrm{E}\langle[(\alpha_t - \hat{\alpha}_t^-) - \overline{K}_t H_t (\alpha_t - \hat{\alpha}_t^-) - \overline{K}_t v_t][H_t \tilde{\alpha}_t^- - v_t]^T\rangle$$

$$= \mathrm{E}\langle[\tilde{\alpha}_t^- - \overline{K}_t H_t \tilde{\alpha}_t^- - \overline{K}_t v_t][H_t \tilde{\alpha}_t^- - v_t]^T\rangle$$

By definition, the a priori covariance (the error covariance matrix before the update) is

$$P_t^- = \mathrm{E}\langle\tilde{\alpha}_t^- \tilde{\alpha}_t^{T-}\rangle$$

It satisfies the equation

$$[\mathrm{I} - \overline{K}_t H_t]P_t^- H_t^T - \overline{K}_t R_t = 0$$

and therefore the gain can be expressed as

$$\overline{K}_t = P_t^- H_t^T [H_t P_t^- H_t^T + R_t]^{-1} \quad \text{(C.16)}$$

which is the solution we seek for the gain as a function of the a priori covariance. One can derive a similar formula for the posteriori covariance (the error covariance matrix after update), which is defined as

$$P_t^+ = \mathrm{E}\langle [\tilde{\alpha}_t^+ \tilde{x}_t^{T+}] \rangle \quad \text{(C.17)}$$

By substituting Equation C.10 into Equation C.4, one obtains the equations

$$\hat{\alpha}_t^+ = (\mathrm{I} - \overline{K}_t H_t) \hat{\alpha}_t^- + \overline{K}_t z_t$$

$$\hat{\alpha}_t^+ = \hat{\alpha}_t^- + \overline{K}_t [z_t - H_t \hat{\alpha}_t^-] \quad \text{(C.18)}$$

Substract x_t from both sides of the latter equation to obtain the equations

$$\hat{\alpha}_t^+ - \alpha_t = \hat{\alpha}_t^- + \overline{K}_t H_t \alpha_t + \overline{K}_t v_t - \overline{K}_t H_t \hat{\alpha}_t^- - \alpha_t$$

$$\tilde{\alpha}_t^+ = \tilde{\alpha}_t^- - \overline{K}_t H_t \tilde{\alpha}_t^- + \overline{K}_t v_t$$

$$\tilde{\alpha}_t^+ = (\mathrm{I} - \overline{K}_t H_t) \tilde{\alpha}_t^- + \overline{K}_t v_t \quad \text{(C.19)}$$

By substituting Equation C.19 into Equation C.17 and noting that $\mathrm{E}\,\tilde{\alpha}_t^- v_t^T = 0$, one obtains

$$P_t^+ = \mathrm{E}\{[\mathrm{I} - \overline{K}_t H_t] \tilde{\alpha}_t^- \tilde{\alpha}_t^{T-} [\mathrm{I} - \overline{K}_t H_t]^T + \overline{K}_t v_t v_t^T \overline{K}_t^T\}$$

$$= (\mathrm{I} - \overline{K}_t H_t) P_t^- (\mathrm{I} - \overline{K}_t H_t)^T + \overline{K}_t R_t \overline{K}_t^T \quad \text{(C.20)}$$

This last equation is the so-called "Joseph form" of the covariance update equation derived by P.D. Joseph. By substituting for $\overline{K}_t$ from Equation C.16, it can be put into the following forms

$$P_t^+ = P_t^- - \overline{K}_t H_t P_t^-$$

$$- P_t^- H_t^T \overline{K}_t^T + \overline{K}_t H_t P_t^- H_t^T \overline{K}_t^T + \overline{K}_t R_t \overline{K}_t^T$$

$$= (\mathrm{I} - \overline{K}_t H_t) P_t^- - P_t^- H_t^T \overline{K}_t^T + \underbrace{\overline{K}_t (H_t P_t^- H_t^T + R_t)}_{P_t^- H_t^T} \overline{K}_t^T$$

$$= (\mathrm{I} - \overline{K}_t H_t) P_t^- \quad \text{(C.21)}$$

the last of which is the one most often used in computation. This implements the effect that conditioning on the measurement has on the covariance matrix of estimation uncertainty.

Error covariance extrapolation

Error covariance extrapolation models the effects of time on the covariance matrix of estimation uncertainty, which is reflected in the a priori values of the covariance and state estimate.

$$P_t^- = \mathrm{E}[\tilde{\alpha}_t^-\ \tilde{\alpha}_t^{T-}]$$

$$\hat{\alpha}_t^- = \Phi_{t-1}\ \hat{\alpha}_{t-1}^+ \qquad \text{(C.22)}$$

respectively. Subtract α_t from both sides of the last equation to obtain the equations

$$\hat{\alpha}_t^- - \alpha_t = \Phi_{t-1}\ \hat{\alpha}_{t-1}^+ - \alpha_t$$

$$\tilde{\alpha}_t^- = \Phi_{t-1}\ [\hat{\alpha}_{t-1}^+ - \alpha_{t-1}]\ w_{t-1}$$

$$= \Phi_{t-1}\ \tilde{\alpha}_{t-1}^+ - w_{t-1}$$

for the propagation of the estimation error, $\tilde{\alpha}_t$. Postmultiply it by $\tilde{\alpha}_t^{T-}$ (on both sides of the equation) and take the expected values. Use the fact that $\mathrm{E}\,\tilde{\alpha}_{t-1}\ w_{t-1}^T = 0$ to obtain

$$P_t^- \underset{=}{def} \mathrm{E}[\tilde{\alpha}_t^-\ \tilde{\alpha}_t^{T-}]$$

$$= \Phi_{t-1}\ P_{t-1}^+\ \Phi_{t-1}^T + Q_{t-1}$$

$$= \Phi_{t-1}\ P_{t-1}^+\ \Phi^T{}_{t-1} + Q_{t-1} \qquad \text{(C.23)}$$

This gives the priori value of the covariance matrix of estimation uncertainty as a function of the previous posteriori value.

Summary of Equations for the Discrete-Time Kalman Estimator

The equations derived in the preceding section are summarized as follows.

System dynamic model:

$$\alpha_t = \Phi_{t-1}\ \alpha_{t-1} + w_{t-1}$$

$$w_t \sim \mathrm{N}\ **(0,\ Q_t)$$

Measurement model:

$$z_t = H_t\ x_t + v_t$$

$$v_t \sim \mathrm{N}\ **(0,\ R_t)$$

Initial conditions:

$E\langle \alpha_0 \rangle = \hat{\alpha}_0$

$E\langle \tilde{\alpha}_0 \ \tilde{\alpha}_0^T \rangle = P_0$

Independence assumption:

$E\langle w_t \ v_j^T \rangle = 0$ for all $t \neq j$

State estimate extrapolation:

$\hat{\alpha}_t^- = \Phi_{t-1} \ \hat{\alpha}_{t-1}^+$

Error covariance extrapolation:

$P_t^- = \Phi_{t-1} \ P_{t-1}^+ \ \Phi_{t-1}^T + Q_{t-1}$

State estimate observational update:

$\hat{\alpha}_t^+ = \hat{\alpha}_t^- + \overline{K}_t [z_t - H_t \ \hat{\alpha}_t^-]$

Error covariance update:

$P_t^+ = [I - \overline{K}_t \ H_t] P_t^-$

Kalman gain matrix:

$\overline{K}_t = P_t^- \ H_t^T [H_t \ P_t^- \ H_t^T + R_t]^{-1}$

Measurement sensitivity

$H_t = x_{t,h} - x_{t-1}$

APPENDIX D

DERIVATION OF DYNAMIC MOVING AVERAGE MODEL

The Dynamic Moving Average Model (DMAM) is developed in this Appendix by integrating two components (i.e., MAM and KFM). In the proposed DMAM, θ_t is the key parameter that combines MAM and KFM into one integrated model. θ_t acts not only as a weight parameter in MAM, but also as the state variable in KFM. The fundamental prediction model in DMAM is formulated in Equation D.0:

$$x_{t,DMAM} = \theta_t \left(\sum_{n=1}^{N} x_{t-n} \right) / N \qquad \text{(D.0)}$$

where x_{t-n}: Observed travel time at time t-n.

$x_{t,DMAM}$: Predicted travel time at time t.

N: Number of time periods considered for moving average.

The fundamental models in KFM are formulated as

Measurement model: $z_t = H\theta_t + v_t$ (D.1)

$$v_t \sim \text{N}^{**}(0,\ R_t)$$

where z_t denotes the observation of travel time in time t, and v_t denotes the measurement error at time interval t that has a normal distribution with zero mean and a variance of R_t, the covariance of the measurement variable.

H is the measurement sensitivity matrix and v_t is the measurement noise. The measurement sensitivity reflects the linear relationship between the state and measurement variables. In the DMAM prediction model, the state variable is the weight parameter θ_t, and the measurement variable is the travel time z_t. The H value is the derivative of the measurement variable z_t with respect to the state variable θ_t:

$$H_t = \frac{\partial z_t}{\partial \hat{\theta}_t} = \left(\sum_{n=1}^{N} x_{t-n} \right) / N \qquad \text{(D.2)}$$

System model: $\theta_t = \phi_{t-1}\theta_{t-1} + \omega_{t-1}$ (D.3)

$w_t \sim$ N **$(0, Q_t)$

Here ϕ_t denotes the transition parameter between successive θ_t at time interval t, which is externally determined. Since there is no linear covariance that can be pre-specified for successive weight parameters, ϕ_t adopts a constant value of 1 to assume that the θ_t is equal to the updated θ_{t-1}^{+} in DMAM. W_t denotes a noise term that has a normal distribution with zero mean and a variance of Q_t, the covariance of the state variable.

The objective of the developed DMAM is to find an estimate of the n state vector θ, represented by $\hat{\theta}_t$, a linear function of the measurements z_i, ..., z_t, that minimizes the weighted mean-squared error.

The observation update problem for a system state estimator

Suppose that a measurement of travel time has been made at time t, and that the information that it provides is to be used in updating the estimate of the weight parameter α of a stochastic system at time t. It is assumed that the measurement variable z_t (i.e., travel time) is linearly related to the state variable (i.e., weight parameter θ_t) by Equation D.1

$$z_t = H\theta_t + v_t \quad \text{(D.1)}$$

$$H_t = \frac{\partial z_t}{\partial \hat{\theta}_t} = \left(\sum_{n=1}^{N} x_{t-n}\right) / N \quad \text{(D.2)}$$

Estimator in linear form

The optimal linear estimate is equivalent to the general optimal estimator if the variates θ and z are jointly Gaussian. Therefore it suffices to seek an updated estimate $\hat{\theta}_t^{+}$ based on the observation z_t that is a linear function of the priori estimate and the measurement z:

$$\hat{\theta}_t^{+} = K_t^1 \hat{\theta}_t^{-} + \overline{K}_t z_t \quad \text{(D.4)}$$

where $\hat{\theta}_t^{-}$ is the a priori estimate of θ_t and $\hat{\theta}_t^{+}$ is the a posteriori value of the estimate.

Optimization problem

The matrices K_t^1 and $\overline{K}_t$ are as yet unknown. We seek those values of K_t^1 and $\overline{K}_t$ such that the new estimate $\hat{\theta}_t^+$ will satisfy the orthogonality principle. This orthogonality condition can be written in the form

$$E\langle[\theta_t - \hat{\theta}_t^+]\, z_i^T\rangle = 0\,,\ i = 1,2,\ldots,t\text{-}1 \tag{D.5}$$

$$E\langle[\theta_t - \hat{\theta}_t^+]\, z_t^T\rangle = 0 \tag{D.6}$$

If one substitutes the formula for θ_t from Equation D.3 and for $\hat{\theta}_t^+$ from Equation D.4 into Equation D.5, then one will observe from Equation D.1 and D.3 that the data $z_1,\ldots,z_t$ do not involve the noise term w_t. Therefore, because the random sequences w_t and v_t are uncorrelated, it follows that

$E\, w_t\, z_i^T = 0$ for $1 \le i \le t$ (Independent assumption)

Using this result, one can obtain the following relation:

$$E\langle\Phi_{t-1}\,\theta_{t-1} + w_{t-1} - K_t^1\,\hat{\theta}_t^- - \overline{K}_t\, z_t\rangle z_i^T = 0,\ i = 1,\ldots,t\text{-}1 \tag{D.7}$$

But because $z_t = H_t\, z_t + v_t$, Equation D.7 can be rewritten as

$$E[\Phi_{t-1}\,\theta_{t-1} - K_t^1\,\hat{\theta}_t^- - \overline{K}_t\, H_t\, x_t - \overline{K}_t\, v_t]\, z_i^T = 0,\ i = 1,\ldots,t\text{-}1 \tag{D.8}$$

We also know that Equation D.6 holds at the previous step, i.e.,

$E\langle[\theta_{t-1} - \hat{x}_{t-1}^+]\, z_i^T\rangle = 0,\ i=1,\ldots,t\text{-}1$ and

$E\langle v_t\, z_i^T\rangle = 0,\ i=1,\ldots,t\text{-}1$

Then Equation D.8 can be reduced to the form

$$\Phi_{t-1}\,E\theta_{t-1}\, z_i^T - K_t^1 E\hat{\theta}_t^-\, z_i^T - \overline{K}_t\, H_t\,\Phi_{t-1}\,E\theta_{t-1}\, z_i^T - \overline{K}_t E v_t\, z_i^T = 0$$

$$\Phi_{t-1}\,E\theta_{t-1}\, z_i^T - K_t^1 E\hat{\theta}_t^-\, z_i^T - \overline{K}_t\, H_t\,\Phi_{t-1}\,E\theta_{t-1}\, z_i^T = 0$$

$$E\langle[\theta_t - \overline{K}_t\, H_t\,\theta_t - K_t^1\,\theta_t] - K_t^1\,(\hat{\theta}_t^- - \theta_t)\rangle z_i^T = 0$$

$$[I - K_t^1 - \overline{K}_t\, H_t]E\theta_t\, z_i^T = 0 \tag{D.9}$$

Equation D.9 can be satisfied for any given θ_t if

$$K_t^1 = I - \overline{K}_t\, H_t \tag{D.10}$$

Clearly, this choice of K_t^1 causes Equation D.4 to satisfy a portion of the condition given by Equation D.5. $\overline{K}_t$ is chosen such that Equation D.6 is satisfied. Let

$$\tilde{\theta}_t^+ \triangleq \tilde{\theta}_t^+ - \theta_t \tag{D.11}$$

$$\tilde{\theta}_t^- \triangleq \tilde{\theta}_t^- - \theta_t \tag{D.12}$$

$$\tilde{z}_t \triangleq \hat{z}_t^- - z_t$$

$$= H_t \hat{\theta}_t^- - z_t \tag{D.13}$$

Vectors $\tilde{\theta}_t^+$ and $\tilde{\theta}_t^-$ are the estimation errors after and before updates, respectively. From Equation D.6

$$\mathrm{E}[\theta_t - \tilde{\theta}_t^+]\,\hat{z}_t^T = 0 \tag{D.14}$$

and also (by subtracting Equation D.1 from Equation D.14)

$$\mathrm{E}[\theta_t - \tilde{\theta}_t^+]\,\tilde{z}_t^T = 0 \tag{D.15}$$

Substitute for $\theta_t, \hat{\theta}_t^+$ and $\tilde{z}_t$ from Equation D.3, D.4 and Equation D.13, respectively. Then

$$\mathrm{E}[\Phi_{t-1}\,\theta_{t-1} + w_{t-1} - K_t^1\,\hat{\theta}_t^- - \overline{K}_t\,z_t][H_t\,\hat{\theta}_t^- - z_t]^T = 0$$

However, by the system structure

$$\mathrm{E}\,w_t\,z_t^T = \mathrm{E}\,w_t\,\hat{\theta}_t^{T+} = 0$$

$$\mathrm{E}[\Phi_{t-1}\,\theta_{t-1} - K_t^1\,\hat{\theta}_t^- - \overline{K}_t\,z_t][H_t\,\hat{\theta}_t^- - z_t]^T = 0$$

Substituting for K_t^1, z_t, and $\tilde{\theta}_t^-$, and using the fact that $\mathrm{E}\,\tilde{\theta}_t^-\,v_t^T = 0$, this last result can be modified as follows:

$$0 = \mathrm{E}\langle[\Phi_{t-1}\,\theta_{t-1} - \hat{\theta}_t^- + \overline{K}_t\,H_t\,\hat{\theta}_t^- - \overline{K}_t\,H_t\,\theta_t - \overline{K}_t\,v_t][H_t\,\hat{\theta}_t^- - H_t\,\theta_t - v_t]^T\rangle$$

$$= \mathrm{E}\langle[(\theta_t - \hat{\theta}_t^-) - \overline{K}_t\,H_t\,(\theta_t - \hat{\theta}_t^-) - \overline{K}_t\,v_t][H_t\,\tilde{\theta}_t^- - v_t]^T\rangle$$

$$= \mathrm{E}\langle[\tilde{\theta}_t^- - \overline{K}_t\,H_t\,\tilde{\theta}_t^- - \overline{K}_t\,v_t][H_t\,\tilde{\theta}_t^- - v_t]^T\rangle$$

By definition, the a priori covariance (the error covariance matrix before the update) is

$$P_t(-) = \mathrm{E}\langle\tilde{\theta}_t^-\,\tilde{\theta}_t^{T-}\rangle$$

It satisfies the equation

$$[\mathrm{I} - \overline{K}_t\,H_t]\,P_t^-\,H_t^T - \overline{K}_t\,R_t = 0$$

and therefore the gain can be expressed as

$$\overline{K}_t = P_t^- \; H_t^T \; [H_t \; P_t^- \; H_t^T + R_t]^{-1} \tag{D.16}$$

which is the solution we seek for the gain as a function of the a priori covariance. One can derive a similar formula for the posteriori covariance (the error covariance matrix after update), which is defined as

$$P_t^+ = \mathrm{E}\langle [\tilde{\theta}_t^+ \; \tilde{x}_t^{T+}] \rangle \tag{D.17}$$

By substituting Equation D.10 into Equation D.4, one obtains the equations

$$\hat{\theta}_t^+ = (\mathrm{I} - \overline{K}_t \; H_t) \hat{\theta}_t^- + \overline{K}_t \; z_t$$

$$\hat{\theta}_t^+ = \hat{\theta}_t^- + \overline{K}_t \; [z_t - H_t \; \hat{\theta}_t^-] \tag{D.18}$$

Substract x_t from both sides of the latter equation to obtain the equations

$$\hat{\theta}_t^+ - \theta_t = \hat{\theta}_t^- + \overline{K}_t \; H_t \; \theta_t + \overline{K}_t \; v_t - \overline{K}_t \; H_t \; \hat{\theta}_t^- - \theta_t$$

$$\tilde{\theta}_t^+ = \tilde{\theta}_t^- - \overline{K}_t \; H_t \; \tilde{\theta}_t^- + \overline{K}_t \; v_t$$

$$\tilde{\theta}_t^+ = (\mathrm{I} - \overline{K}_t \; H_t) \tilde{\theta}_t^- + \overline{K}_t \; v_t \tag{D.19}$$

By substituting Equation D.19 into Equation D.17 and noting that $\mathrm{E}\, \tilde{\theta}_t^- \; v_t^T = 0$, one obtains

$$P_t^+ = \mathrm{E}\{[\mathrm{I} - \overline{K}_t \; H_t] \tilde{\theta}_t^- \; \tilde{\theta}_t^{T-} [\mathrm{I} - \overline{K}_t \; H_t]^T + \overline{K}_t \; v_t \; v_t^T \; \overline{K}_t^T\}$$

$$= (\mathrm{I} - \overline{K}_t \; H_t) P_t^- (\mathrm{I} - \overline{K}_t \; H_t)^T + \overline{K}_t \; R_t \; \overline{K}_t^T \tag{D.20}$$

This last equation is the so-called "Joseph form" of the covariance update equation derived by P.D. Joseph. By substituting for $\overline{K}_t$ from Equation D.16, it can be put into the following forms

$$P_t^+ = P_t^- - \overline{K}_t \; H_t \; P_t^- - P_t^- \; H_t^T \; \overline{K}_t^T + \overline{K}_t \; H_t \; P_t^- \; H_t^T \; \overline{K}_t^T + \overline{K}_t \; R_t \; \overline{K}_t^T$$

$$= (\mathrm{I} - \overline{K}_t \; H_t) P_t^- - P_t^- \; H_t^T \; \overline{K}_t^T + \underbrace{\overline{K}_t (H_t P_t^- H_t^T + R_t)}_{P_t^- H_t^T} \; \overline{K}_t^T$$

$$= (\mathrm{I} - \overline{K}_t \; H_t) P_t^- \tag{D.21}$$

the last of which is the one most often used in computation. This implements the effect that conditioning on the measurement has on the covariance matrix of estimation uncertainty.

Error covariance extrapolation

The error covariance extrapolation models the effects of time on the covariance matrix of estimation uncertainty, which is reflected in the a priori values of the covariance and state estimate.

$$P_t^- = \mathrm{E}[\tilde{\theta}_t^- \tilde{\theta}_t^{T-}]$$

$$\hat{\theta}_t^- = \Phi_{t-1} \hat{\theta}_{t-1}^+ \tag{D.22}$$

respectively. Subtract α_t from both sides of the last equation to obtain the equations

$$\hat{\theta}_t^- - \theta_t = \Phi_{t-1} \hat{\theta}_{t-1}^+ - \theta_t$$

$$\tilde{\theta}_t^- = \Phi_{t-1} [\hat{\theta}_{t-1}^+ - \theta_{t-1}] w_{t-1}$$

$$= \Phi_{t-1} \tilde{\theta}_{t-1}^+ - w_{t-1}$$

for the propagation of the estimation error, $\tilde{\theta}_t$. Postmultiply it by $\tilde{\theta}_t^{T-}$ (on both sides of the equation) and take the expected values. Use the fact that $\mathrm{E}\,\tilde{\theta}_{t-1} w_{t-1}^T = 0$ to obtain

$$P_t^- \underset{=}{def} \mathrm{E}[\tilde{\theta}_t^- \tilde{\theta}_t^{T-}]$$

$$= \Phi_{t-1} P_{t-1}^+ \Phi^T{}_{t-1} + Q_{t-1}$$

$$= \Phi_{t-1} P_{t-1}^+ \Phi^T{}_{t-1} + Q_{t-1} \tag{D.23}$$

This gives the a priori value of the covariance matrix of estimation uncertainty as a function of the previous a posteriori value.

Summary of Equations for the Discrete-Time Kalman Estimator

The equations derived in the preceding section are summarized as follows.

System dynamic model:

$$\theta_t = \Phi_{t-1} \theta_{t-1} + w_{t-1}$$

$$w_t \sim \mathrm{N} \; **(0, Q_t)$$

Measurement model:

$$z_t = H_t x_t + v_t$$

$$v_t \sim \mathrm{N} **(0, R_t)$$

Initial conditions:

$$E\langle \theta_0 \rangle = \hat{\theta}_0$$

$$E\langle \tilde{\theta}_0 \, \tilde{\theta}_0^T \rangle = P_0$$

Independence assumption:

$E\langle w_t \, v_j^T \rangle = 0$ for all $t \neq j$

State estimate extrapolation:

$$\hat{\theta}_t^- = \Phi_{t-1} \, \hat{\theta}_{t-1}^+$$

Error covariance extrapolation:

$$P_t^- = \Phi_{t-1} \, P_{t-1}^+ \, \Phi_{t-1}^T + Q_{t-1}$$

State estimate observational update:

$$\hat{\theta}_t^+ = \hat{\theta}_t^- + \overline{K}_t [\, z_t - H_t \, \hat{\theta}_t^- \,]$$

Error covariance update:

$$P_t^+ = [\mathrm{I} - \overline{K}_t \, H_t \,] \, P_t^-$$

Kalman gain matrix:

$$\overline{K}_t = P_t^- \, H_t^T [\, H_t \, P_t^- \, H_t^T + R_t \,]^{-1}$$

Measurement sensitivity

$$H_t = \left(\sum_{n=1}^{N} x_{t-n} \right) / N$$

REFERENCES

1. Abbasi, M. F., Sisiopiku, V. P., and Taylor, W.C., *Development and Testing of a Freeway Recurring Congestion Measure,* Transportation Research Board, 78th Annual Meeting, Washington D. C., 1999.

2. Abdelfattah,A.M., Khan,A.M., *Models for Predicting Bus Delays,* Transportation Research Board, 77th Annual Meeting, Washington D. C., 1998.

3. Ahmed, M. S., and Cook, A. R., *Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques,* Transportation Research Record 722, TRB, National Research Council, Washington, D. C., 1979, pp. 1-9.

4. Akahame, H., and Koshi, M., *Short-Term Prediction of Inflow Volumes of Urban Freeway,* the second international conference on Road Traffic Control, 1986, pp. 35-38.

5. Al-Deek, H. M., Angelo, P. D., Wang, M.C., *Travel Time Prediction with Non-linear Time series,* ASCE Journal of Transportation Engineering, Vol.122, No.6, 1998, pp. 440-446.

6. Alecsandru, C., Ishak, S., *A Hybrid Model-based and Memory-based Traffic Prediction System,* Transportation Research Board, 83nd Annual Meeting, Washington D. C., 2004.

7. Angelo, P. D., Al-Deek, H. M., and Wang, M. C, *Travel-Time Prediction for Freeway Corridors,* Transportation Research Record 1676, TRB, National Research Council, Washington, D. C., 1999, pp. 184-191.

8. Arem, B., Vlist, M., Muste, M., and Smulders, S., *Travel Time Estimation in the GERDIEN Project,* International Journal of Forecasting, Vol. 13, 1997, pp 73-85.

9. Athol, P. *Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream,* Highway Research Record 72, HRB, National Research Council, Washington D.C.,1965, pp 58 – 87.

10. Bae, S., *Dynamic Estimation of Travel Time on Arterial Roads by Using [an] Automatic Vehicle Location (AVL) Bus as a Vehicle Probe,* Transportation Research-Part A, Vol. 31, 1997, pp 60.

11. Bhattacharjee, D., Sinha, K. C., and Krogmeier, J. V., *Modeling the Effects of Traveler Information on Freeway Origin-Destination Demand Prediction,* Transportation Research-Part C, Vol. 9, 2001, pp 381-398.

12. Bowerman, B., O'Connell, R., *Time Series and Forecasting,* North Scituate, Massachusetts: Duxbury Press, 1979.

13. Brown, R.G., *Smoothing and Prediction of Discrete Time Series,* Prentice Hall, Englewood Cliffs, NJ, 1963.

14. Çakanyildirim, M., *Forecasting write up,* retrieved January 2004 from World Wide Web:http://www.utdallas.edu/~metin/Ba3352/Slides/forecast.ppt, 2002.

15. Chabini, I., Ganugapati, S., *Design and Implementation of Parallel Dynamic Shortest Path Algorithms for Intelligent Transportation Systems Applications,* Transportation Research Record 1771, TRB National Research Council, Washington, D. C., 2001, pp. 219-228.

16. Chatfield, C., *Time-series Forecasting,* ISBN 1-58488-063-5, published by Chapman and Hall/CRC, 2002.

17. Chen, M., and Chien, S., *Determining the Number of Probe Vehicles for Freeway Travel Time Estimation Using Microscopic Simulation,* Transportation Research Record 1719, TRB National Research Council, Washington, D. C., 2001, pp. 61~68.

18. Chen, M., and Chien, S., *Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based,* Transportation Research Record 1768, TRB National Research Council, Washington, D. C., 2002, pp.157-161.

19. Chen, M., Liu, X., Chien, S., *A Dynamic Bus Arrival Time Prediction Model Based on APC Data,* Journal of Computer- Aided Civil and Infrastructure Engineering, Accepted for publication in 2003.

20. Chen, H., Grant-Muller, S., *Use of Sequential Learning for Short-term Traffic Flow Forecasting,* Transportation Research-Part C, Vol. 9, 2001, pp. 319-336.

21. Cheu, R. L., X. Jin, K. C. Ng, Y. L. Ng, and D. Srinivasan., *Calibration of FRESIM for Singapore Expressway Using Genetic Algorithm,* Journal of Transportation Engineering, Vol. 124, No. 6, November/December 1998, pp. 526-535.

22. Chestlow, M., Htcher, G., and Patel, M., *An Initial Evaluation of Alternative Intelligent Vehicle Highway Systems Architectures,* MIT Research Report, 92w063, 1992.

23. Chestlow, M., Htcher, G., *Comparison of Advanced Traffic Management and Traveler Information System Architectures for Intelligent Vehicle Highway Systems,* Transportation Research Record 1408, TRB National Research Council, Washington, D. C., 1993, pp. 8-17.

24. Chien, S., Ding, Y., and Wei, C., *Dynamic Bus Arrival Time Prediction Using Artificial Neural Networks,* Journal of Transportation Engineering, ASCE, Vol. 128, No. 5, 2002, pp. 429~438.

25. Cho, H., Rilett, L., R., *Forecasting Train Travel Time,* Transportation Research Board, the 82nd Annual Meeting, Washington, D. C., 2003.

26. Chu, L., Liu, H., Recker, W., Zhang, H., *Development of a Simulation Laboratory for Evaluating Ramp Metering Algorithms,* 2002 TRB Annual Meeting.

27. Cohen, S. L., *Application of Car-Following Systems to the Queue Discharge Problem at Signalized Intersections,* The 81st Annual Meeting, TRB, 2002.

28. Coifman, B., *New Methodology for Smoothing Freeway Loop Detector Data: Introduction to Digital Filtering,* Transportation Research Record 1554, TRB National Research Council, Washington, D. C., 1996, pp. 142-152.

29. Croarkin, C., Tobias, P., Filliben, J., Hembree, B., et.al. *Engineering Statistics Handbook, 2001.* Retrieve October 2002 from the World Wide Web: http://www.itl.nist.gov/div898/handbook/index.htm.

30. Daganzo, C.F., and Lin H-W, *The Spatial Evolution of Queues during the Morning Commute in a Single Corridor,* Transportation Research Board, 73rd Annual Meeting, 1994.

31. Dailey, D. J., *Travel-time Estimation Using Cross-correlation Techniques,* Transportation Research-Part B, Vol. 27, 1993, pp 97-107.

32. Dailey, D.J., and Schoepflin, T.N., *Correlation Technique for Estimating Traffic Speed from Cameras,* Transportation Research Board, 82nd Annual Meeting, Washington, D. C., 2003.

33. Dailey, D.J., Cathey, F., *Corridor Travel Time Using Transit Vehicles as Probes,* Transportation Research Board, 82nd Annual Meeting, Washington, D. C., 2003.

34. Davis, A., and Nihan, L., *Nonparametric Regression and Short-term Freeway Traffic Forecasting,* Journal of Transportation Engineering, ASCE, 1991, pp178-188.

35. *Developing Traveler Information Systems Using the National ITS Architecture, U.S.* Department of Transportation, Intelligent Transportation Systems Joint Program Office, August, 1998.

36. Dia, H., and Rose, G., *Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data,* Transportation Research-Part C: Vol. 5, 1997, pp. 313-331.

37. Eisele, W. L., Rilett, L. R., *Estimating Corridor Travel Time Mean, Variance, and Covariance with Intelligent Transportation Systems Link Travel Time Data,* Transportation Research Board, 81st Annual Meeting, Washington, D. C., 2002.

38. Enders, W., *RATS Handbook for Econometric Time Series,* John Wiley and Sons, Inc, 1996.

39. Fu, L., and L. R. Rilett, *Shortest Path Problems In Traffic Networks With Dynamic and Stochastic Link Travel Times,* Transportation Research-Part B, Vol. 32, 1998, No. 7.

40. Fu, L., *An Adaptive Routing Algorithm for In-vehicle Route Guidance Systems with Real-time Information.* Transportation Research-Part B, Vol. 35, 2001, pp 749-765.

41. Frechette, L. A., Khan, A. M., *Bayesian Regression-Based Urban Traffic Models,* Transportation Research Board 77th Annual Meeting, Washington, D. C., 1998.

42. Gall, A., Hall, F., *Distinguishing Between Incident Congestion and Recurrent Congestion: A Proposed Logic,* Transportation Research Record 1232, 1989, pp. 1-8.

43. Gajewski, B.J., Rilett, L.R., *Estimating Link Travel Time Correlation: An Application of Bayesian Smoothing Splines,* Transportation Research Board 82nd Annual Meeting, Washington, D. C., 2003.

44. Gardner, S., McKenzie, E., *Forecasting Trends in Time Series,* Management Science, 31, 1985, pp.1237-1246.

45. Gerlough, D. L., Barnes, F. C., and Schuhl, A., *Poisson and Other Distributions in Traffic,* Eno Foundation for Transportation, 1971.

46. Golob, T.F., Regan, A.C., *Impacts of Information Technology on Personal Travel and Commercial Vehicle Operations: Research Challenges and Opportunities,* Transportation Research-Part C 9, 2001, pp. 87-121.

47. Gross, D., Craig, R., *A Comparison of Maximum Likelihood, Exponential Smoothing and Bayes Forecasting Procedures in Inventory Modeling,* International Journal of Production Research, Vol. 12, No. 5. 1974.

48. Hall, F., and Persaud, B. N., *Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems,* Transportation Research Record 1232, 1989, pp. 9-16.

49. Hall, R.W., *Route Choice and Advanced Traveler Information Systems on A Capacitated and Dynamic Network,* Transportation Research-Part C, Vol. 4, 1996, pp. 289-306.

50. Hellinga, B., and Gudapati, R., *Estimating Link Travel Times for Advanced Traveler Information Systems,* Proceedings of the Canadian Society of Civil Engineers 3rd Transportation Specialty Conference, London, 2000.

51. Hellinga, B., and Knapp, G., *Automatic Freeway Incident Detection Using Travel Time Data from AVI Equipped Vehicles,* the 6th ITS World Congress, Toronto, Canada November 8 - 12, 1999.

52. Hoffman, G. and Janko, J., *Travel Times as a Basic Part of the LISB Guidance Strategy,* IEEE Road traffic Control Conference, London, May, 1990.

53. Hsiao, C., *The Application of Fuzzy Logic and Neural Networks to Freeway Incident Detection,* Transportation Research-Part A: Vol. 31, pp. 58-59,1997.

54. http://www.azfms.com/, retrieved October 2002.

55. http://roso.epfl.ch/DTMS/, retrieved October 2002.

56. http://www.mto.gov.on.ca/english/traveller/compass/, retrieved October 2002.

57. http://www.palmbeachcotraffic.org/, retrieved October 2002.

58. http://www.fortrantraffic.com/systems/fastracs.htm, retrieved October 2002.

59. http://www.transguide.dot.state.tx.us/index.html, retrieved October 2002.

60. http://www.pdpassociates.com/tips.htm, retrieved October 2002.

61. http://www.fhwa.dot.gov/ohim/timedata.htm, retrieved October 2002.

62. http://www.itl.nist.gov/div898/handbook/, retrieved October 2002.

63. http://www.fhwa.dot.gov/tfhrc/safety/pubs/its/generalits/choosette.pdf, retrieved October 2002.

64. http://www.itsdocs.fhwa.dot.gov/jpodocs/repts_te/37n01!.pdf, retrieved October 2002.

65. http://www.ctre.iastate.edu/pubs/en_route/aug02cer/Kamyab.pdf, retrieved October 2002.

66. http://www.itsa.org/mn.nsf/0/2bc50fbbd01ced6d85256bea00767411?OpenDocument, retrieved October 2002.

67. Intelligent Transportation Society of America, *Choosing the Route to Traveler Information Systems Deployment: Decision Factors for Creating Public-Private Business Plans,* 1998.

68. Ishak, S., Alecsandru, C., *Optimizing Traffic Prediction Performance of Neural Networks under Various Topological, Input, and Traffic Condition Settings,* Transportation Research Record 82nd Annual Meeting, Washington, D. C., 2003.

69. Ishak, S., Al-Deek, H., *Statistical Evaluation of I-4 Traffic Prediction System,* Transportation Research Record 82nd Annual Meeting, Washington, D. C., 2003.

70. Iwasaki, M., and Saito, K., *Short-Term Prediction of Speed Fluctuations on a Motorway Using Historical Patterns,* Transportation Research Board, Annual Meeting.

71. Iwasaki, M. and Shirao, K., *A Short Term Prediction of Traffic Fluctuations Using Pseudo-Traffic Patterns,* The 3rd World Congress on Intelligent Transport Systems, Orland, 1996.

72. Jiang, Y., *Prediction of Freeway Traffic Flows Using Kalman Predictor in Combination with Time Series,* Journal of the Transportation Research Forum, Vol. 56, Issue 2, 2003, pp. 99-118.

73. Jin, X., Cheu, R., and Srinivasan, D., *Development and Adaptation of Constructive Probabilistic Neural Network in Freeway Incident Detection,* Transportation Research-Part C: Vol. 10, 2002, pp. 121-147.

74. Kamarianakis, Y., and Prastacos, P., *Forecasting Traffic Flow Conditions in an Urban Network, Comparison of Multivariate and Univariate Approaches,* Transportation Research Board 82nd Annual Meeting, Washington, D. C., 2003.

75. Kaysi, I., Ben-Akiva, M., and Koutsopoulos, H., *Integrated Approach to Vehicle Routing and Congestion Prediction for Real-Time Driver Guidance,* Transportation Research Record 1408, 1993, pp. 66-74.

76. Khattak, A., F. Koppelman, and J. Schofer, *Stated Preference for Investigating Commuters' Diversion Propensity,* Transportation Research Board, 71st Annual Meeting, Washington, D.C., 1992.

77. Kisgyorgy, L., and Rilett, L.R., *Travel Time Prediction By Advanced Neural Network,* Periodica Polytechnica Ser. Civil Engineering, Vol. 46, No.1, 2002, pp. 15-32.

78. Koutsopoulos, H.H. and Polydoropoulou A. and Ben-Akiva, M., *The Use of Driving Simulators to Investigate the Response to Traffic Information,* Center for Transportation Studies, Massachusetts Institute of Technology, 1993.

79. Kreer, J. B., *A Comparison of Predictor Algorithms for Computerized Control,* Traffic Engineering, Vol. 45; 1975, pp.51-S6.

80. Kuchipudi, C.M. and Chien, S., *Development of a Hybrid Model for Dynamic Travel Time Prediction,* Transportation Research Board 82nd Annual Meeting, Washington, D. C., 2003.

81. Kwon, E., and Stephanedes, Y. J., *Comparative Evaluation of Adaptive and Neural-Network Exit Demand Prediction for Freeway Control,* Transportation Research Record 1446, 1993, pp. 66-76.

82. Lan, C., and Miaou, S., *Real-Time Prediction of Traffic Flows Using Dynamic Generalized Linear Models,* Transportation Research Record 1678, 1998, pp. 168-177.

83. Lee, D. H., Yang, X., and Chandrasekar, P. *Parameter Calibration for PARAMICS Using Genetic Algorithm,* The 80th Annual Meeting, TRB, 2001.

84. Lin, W., Lu, Q., Dahlgren, J., *A Dynamic Procedure for Short-Term Prediction of Traffic Conditions,* Transportation Research Record, 81st Annual Meeting, Washington, D. C., 2002.

85. Lin, W., Kulkarni, A., Mirchandani, P., *Arterial Travel Time Estimation for Advanced Traveler Information Systems,* Transportation Research Record, 82nd Annual Meeting, Washington, D. C., 2003.

86. Lint, J., Hoogendoom, S., Zuylen, H., *Freeway Travel Time Prediction with State-Space Neural Networks,* Transportation Research Record 1811, TRB, National Research Council, Washington, D. C., 2002, pp. 30-39.

87. Lu, J., *Prediction of Traffic Flow by an Adaptive Prediction System,* Transportation Research Record 1287, TRB, National Research Council, Washington, D. C., 1990, pp. 54-61.

88. Makridakis, S., *The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition,* Journal of Forecasting, Vol. 1, 1982, pages 111-153.

89. Masliah, M., *Using the Exponential Smoothing Approach to Time Series Forecasting on 6 DOF Tracking Data,* retrieved October 2002 from: http://etclab.rose.utoronto.ca/people/moman/Smoothing/smoothing.html.

90. McFadden, J., and Durrans, S.R., *Application Artificial Neural Networks to Predict Speeds on Two-lane Rural Highways,* Transportation Research Record, 80th Annual Meeting, Washington, D. C., 2001.

91. Mcshane, W., Roess, R., and Prassas, E., Traffic Engineering, 2nd edition, ISBN-0-13-461336-8, 1998, published by Prentice Hall.

92. Meldrum, D., and Taylor, C., *Freeway Traffic Data Prediction Using Artificial Neural Networks and Development of a Fuzzy Logic Ramp Metering Algorithm,* Final Technical Report T9903, Washington State Department of Transportation, April 1995.

93. Nair, A., Liu, J., Rilett, L., and Gupta, S., *Non-Linear Analysis of Traffic Flow, TransLink Research Reports, 2000.* Retrieved October 2002 from:http://translink.tamu.edu/docs/Research/LinearAnalysisTrafficFlow/chaos1.pdf

94. Nam, K. H. and Drew, D. R., *Automatic Measurement of Traffic Variables for Intelligent Transportation Systems Applications,* Transportation Research-Part B, Vol. 33, 1999, pp 437-457.

95. Nam, K. H. and Drew, D. R., *Traffic Dynamics: Method for Estimating Freeway Travel Times in Real Time from Flow Measurements,* Journal of Transportation Engineering, ASCE, Vol. 122, 1996, pp 185-191.

96. Nau, R., http://www.duke.edu/~rnau/411avg.htm, retrieved December 2003.

97. Nanthawichit, C., Nakatsuji, T., Suzuki, H., *Application of Probe Vehicle Data for Real-Time Traffic State Estimation and Short-Term Travel Time Prediction on a Freeway,* Transportation Research Board, 82nd Annual Meeting, Washington, DC, 2003.

98. Newbold, P., *Statistics for Business and Economics,* Fourth Edition, Prentice-Hall. 1995, pp. 711-712

99. Nihan, N.L., and Holmesland, K.O., *Use of the Box and Jenkins Time Series Technique in Traffic Forecasting,* Transportation, Vol. 9, 1980, pp 125-143.

100. Okutani, I., and Stephanedes, Y. J., *Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory,* Transportation Research-Part B, Vol. 18, 1984, pp 1-11.

101. Pattanamekar, P., Park, D., Rilett, L.R., Lee, J., Lee, C., *Dynamic and Stochastic Shortest Path in Transportation Networks with Two Components of Travel*

Time Uncertainty, Transportation Research-Part C, Vol. 11, 2003, pp. 331-354.

102. Persaud, B., Hall, F., *Catastrophe Theory and Patterns in 30-second Freeway Traffic data Implications for Incident Detection,* Transportation Research-Part A: Vol. 23, 1989, pp. 103-113.

103. Petty, K., Bickel, P., Ostland, M., Rice, J., Schoenberg, F., Jiang, J., and Ritov, Y., *Accurate Estimation of Travel Times from Single-loop Detectors,* Transportation Research-Part A, Vol. 32, 1998, pp 1-17.

104. Rilett, L., Park, D., and Gajewski, B., *Estimating Confidence Interval for Freeway Corridor Travel Time Forecasts,* 6th World Congress on Intelligent Transportation Systems, Toronto, Canada, 1999.

105. Ross, P., *Exponential Filtering of Traffic Data,* Transportation Research Board, 869, TRB, National Research Counil, Washington C.C., 1982, pp. 43-49.

106. Suzuki, H., Nakatsuji, T., Tanaboriboon, Y., and Takahashi, K., *A Neural-Kalman Filter for Dynamic Estimation of Origin-Destination Travel Time and Flow on a Long Freeway Corridor,* Transportation Research Record 1739, TRB, National Research Council, Washington D.C., 2000, pp. 67-75.

107. Sen, A., Thakuriah, P., Zhu, X., Karr, A., *Variances of Link Travel Time Estimates: Implications for Optimal Routes,* International Transactions in Operational Research, 1999, pp 75-87.

108. Schreckenberg, M., Neubert, L., and Wahle, J., *Simulation of Traffic in Large Road Networks,* Future Generation Computer Systems, Vol. 17, 2001, pp 649-657.

109. Sheu, J., *A Stochastic Modeling Approach to Real-Time Prediction of Queue Overflows,* Transportation Science, Vol. 37 Issue 1, 2003, pp. 97-120.

110. Sheu, J., *A Fuzzy Clustering-based Approach to Automatic Freeway Incident Detection and Characterization,* Fuzzy Sets and Systems, Vol. 128, 2002, pp 377-388.

111. Shalaby, A. and Farhan, A., *Bus Travel Time Prediction Model for Dynamic Operations Control and Passenger Information Systems,* Transportation Research Board, 82nd Annual Meeting, Washington, D. C., 2003.

112. Shbaklo, S., Bhat, C., Koppelman, F., Li, J., Thakuriah, P., Sen, A., Rouphail, N., *Short-Term Travel Time Prediction,* Review of Literature and Methods, ADVANCE Project Report TRF-TT-01, 1992.

113. Skabardonis, A., Varaiya, P., Petty, K., *Measuring Recurrent and Non-recurrent Traffic Congestion,* Transportation Research Board, 82nd annual meeting, Washington, D. C., 2003.

114. Skabardonis, A., *Simulation of Freeway Weaving Areas,* The 81st Annual Meeting, TRB, 2002.

115. Smith, B.L., and Demetsky, M.J., *Short-Term Traffic Flow Prediction: Neural Network Approach,* Transportation Research Record 1453, TRB, National Research Council, Washington D.C., 1994, pp. 98-104.

116. Smith, B. L., and Demetsky, M. J., *Multiple-Interval Freeway Traffic Flow Forecasting,* Transportation Research Record 1554, TRB, National Research Council, Washington D.C., 1996, pp. 136-141.

117. Smith B. L., Demetsky, M. J., *Traffic Flow Forecasting: Comparison of Modeling Approaches,* Journal of Transportation Engineering Vol.123, 1997, pp. 261-266.

118. Smith, B. L., *Forecasting Freeway Traffic Flow for Intelligent Transportation Systems Application,* Transportation Research Part A, Vol. 31, 1997, pp. 61.

119. Smith, B.L., Williams, B., M., and Oswald, R., K, *Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting,* Transportation Research-Part C, Vol. 10, 2002, pp. 303-321.

120. Spasovic, L., Hausman, K., Curley, J., *Mobility and the Costs of Congestion in New Jersey,* final report to New Jersey Department of Transportation, 2001.

121. Sisiopiku, V.P. and Rouphail. N. M., *Toward the Use of Detector Output for Arterial Link Travel Time Estimation: A Literature Review,* Transportation Research Record 1457, TRB, National Research Council, Washington D.C., 1994, pp. 158-165.

122. Sisiopiku, V., *Travel Time Estimation From Loop Detector Data for Advanced Traveler Information Systems Applications,* Transportation Research-Part A, Vol. 30, 1996, pp. 61-62.

123. Srinivasan, D., Cheu, R., Poh, Y., and Ng, A., *Development of an Intelligent Technique for Traffic Network Incident Detection,* Engineering Applications of Artificial Intelligence, Vol. 13, 2000, pp 311-322.

124. Stathopoulos, A., Karlaftis, M.G., *A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction,* Transportation Research-Part C, Vol. 11, 2003, pp. 121-135.

125. Stephanedes, Y.J., Michalopoulos, P.G., and Plum, R.A., *Improved Estimation of Traffic Flow for Real Time Control,* Transportation Research Record, 795, TRB, National Research Council, Washington D.C., 1981, pp. 28-39.

126. Stephanedes, Y. J., Kwon, E., and Michalopoulos, P., *On-line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors,* Transportation Research Record 1287, TRB, National Research Council, Washington D.C., 1990, pp.11-19.

127. Stephanedes, Y., and Chassiakos, A., *Freeway Incident Detection Through Filtering,* Transportation Research-Part C: Vol. 1, 1993, pp. 219-233.

128. Sun, H., Liu, H., Xiao, H., He, R., and Ran, B., *Short-Term Traffic Forecasting Using the Local Linear Regression Model,* Transportation Research Record, 82nd Annual Meeting, Washington, D. C., 2003.

129. Sung, K., Bell, M.G., Seong, M., and Park, S., *Shortest Paths in a Network with Time-dependent Flow Speeds,* European Journal of Operational Research Vol. 121, 2000, pp 32-39.

130. Taft, K., *Operations Management,* Chapter 5 Forecasting, retrieved from World Wide Web:http://www.sba.pdx.edu/faculty/karlt/frcst/sld001.htm.

131. Terry, W.R., Lee, J.B., and Kumar, A. *Time Series Analysis in Acid Rain Modeling Evaluation of Filling Missing Values by Linear Interpolation,* Atmospheric Environment, 1986, pp. 1941-1945.

132. Tsai, Y., Wu, M., and Adams, E., *GPS/GIS-Enhanced Roadway Inventory System,* Transportation Research Board, 82nd Annual Meeting, Washington, D. C., 2003.

133. Vemuri, A., Pant, M., *Short-Term Forecasting of Traffic Delays in Highway Construction Zones Using On-Line Approximators,* Mathematical and Computer Modeling, Vol. 27, 1998, pp. 311-322.

134. Wahle, J., Annen, O., Schuster, C., Neubert, L., and Schrechenberg, M., *A Dynamic Route Guidance System Based on Real Traffic Data,* European Journal of Operational Research, Vol. 131, 2001, pp. 302-308.

135. Walker,J., retrieved October 2003 from World Wide Web,

http://www.fourmilab.ch/hackdiet/www/subsubsection1_2_4_0_4_3.html.

136. Wang, Y., and Nihan, N. L., *Freeway Traffic Speed Estimation with Single-loop Outputs,* Transportation Research Record 1727, TRB, National Research Council, Washington D.C., 2000, pp. 120-126.

137. Wang, Y., and Nihan, N. L., *A Robust Method of Filtering Single-Loop Data for Improved Speed Estimation,* Transportation Research Board, 81st Annual Meeting, Washington, D. C., 2002.

138. Williams, B.M., *Multivariate Vehicular Traffic Flow Prediction: An Evaluation of ARIMAX Modeling,* Transportation Research Board, 80th Annual Meeting, Washington, D. C., 2001.

139. Williams, B.M., Durvasula, P.K., and Brown, D.E., *Urban Freeway Traffic Flow Prediction: Application of Seasonal ARIMA and Exponential Smoothing Models,* Transportation Research Board, 78th Annual Meeting, Washington, D. C., 1999.

140. Woodhull, J., *Issues in On-time Performance of Bus Systems, Unpublished manuscript.* Los Angela, CA: Southern California Rapid Transit District, 1987.

141. Xie, C., and Parkany, E., *Signalized Intersection Simulation in CORSIM and SIMTRAFFIC,* The 81st Annual Meeting, TRB, 2002.

142. Yin, H., Wong, S., Xu, J., Wong, C., *Urban Traffic Flow Prediction Using a Fuzzy-neural Approach,* Transportation Research-Part C, Vol. 10, 2002, pp 85-98.

143. You, J., and Kim, T. J., *Development and Evaluation of a Hybrid Travel Time Forecasting Model,* Transportation Research-Part C, Vol. 8, 2000, pp 231-256.

144. Zhang, X., Rice, J.A., *Short-Term Travel Time Prediction,* Transportation Research-Part C, Vol. 11, 2003, pp. 187-210.

145. Zhu, F., *Locations of AVI System and Travel Time Forecasting,* Master thesis, Virginia Polytechnic Institute and State University, 2002.

146. Zietsman, J., Rilett, L., *Aggregate- and Disaggregate-Based Travel Time Estimation: Comparison of Applications to Sustainability Analysis and Advanced Traveler Information Systems,* Transportation Research Record, 1725, TRB, National Research Council, Washington D.C., 2000, pp 84-94.