



University of Kentucky
UKnowledge

University of Kentucky Doctoral Dissertations

Graduate School

2006

DYNAMIC FREEWAY TRAVEL TIME PREDICTION USING SINGLE LOOP DETECTOR AND INCIDENT DATA

Jingxin Xia

University of Kentucky, jx4k@virginia.edu

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Xia, Jingxin, "DYNAMIC FREEWAY TRAVEL TIME PREDICTION USING SINGLE LOOP DETECTOR AND INCIDENT DATA" (2006). *University of Kentucky Doctoral Dissertations*. 315.
https://uknowledge.uky.edu/gradschool_diss/315

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

ABSTRACT OF DISSERTATION

Jingxin Xia

The Graduate School
University of Kentucky

2006

DYNAMIC FREEWAY TRAVEL TIME PREDICTION USING
SINGLE LOOP DETECTOR AND INCIDENT DATA

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering
at the University of Kentucky

By

Jingxin Xia

Lexington, Kentucky

Co-Director: Dr. Mei Chen, Professor of Civil Engineering
and Dr. Nick Stamatiadis, Professor of Civil Engineering

Lexington, Kentucky

2006

Copyright © Jingxin Xia 2006

ABSTRACT OF DISSERTATION

DYNAMIC FREEWAY TRAVEL TIME PREDICTION USING SINGLE LOOP DETECTOR AND INCIDENT DATA

The accurate estimation of travel time is valuable for a variety of transportation applications such as freeway performance evaluation and real-time traveler information. Given the extensive availability of traffic data collected by intelligent transportation systems, a variety of travel time estimation methods have been developed. Despite limited success under light traffic conditions, traditional corridor travel time prediction methods have suffered various drawbacks. First, most of these methods are developed based on data generated by dual-loop detectors that contain average spot speeds. However, single-loop detectors (and other devices that emulate its operation) are the most commonly used devices in traffic monitoring systems. There has not been a reliable methodology for travel time prediction based on data generated by such devices due to the lack of speed measurements. Moreover, the majority of existing studies focus on travel time estimation. Secondly, the effect of traffic progression along the freeway has not been considered in the travel time prediction process. Moreover, the impact of incidents on travel time estimates has not been effectively accounted for in existing studies.

The objective of this dissertation is to develop a methodology for dynamic travel time prediction based on continuous data generated by single-loop detectors (and similar devices) and incident reports generated by the traffic monitoring system. This method involves multiple-step-ahead prediction for flow rate and occupancy in real time. A seasonal autoregressive integrated moving average (SARIMA) model is developed with an embedded adaptive predictor. This predictor adjusts the prediction error based on traffic data that becomes available every five minutes at each station. The impact of incidents is evaluated based on estimates of incident duration and the queue incurred.

Tests and comparative analyses show that this method is able to capture the real-time characteristics of the traffic and provide more accurate travel time estimates particularly when incidents occur. The sensitivities of the models to the variations of the flow and occupancy data are analyzed and future research has been identified.

The potential of this methodology in dealing with less than perfect data sources has been demonstrated. This provides good opportunity for the wide application of the proposed method since single-loop type detectors are most extensively installed in various intelligent transportation system deployments.

KEYWORDS: Travel Time, SARIMA, Adaptive Kalman Filter, Multi-Step-Ahead Prediction, Incident

Jingxin Xia

12-10-2006

DYNAMIC FREEWAY TRAVEL TIME PREDICTION USING
SINGLE LOOP DETECTOR AND INCIDENT DATA

By

Jingxin Xia

Dr. Mei Chen

Co-Director of Dissertation

Dr. Nick Stamatiadis

Co-Director of Dissertation

Dr. Kamyar C. Mahboub

Director of Graduate Studies

12-10-2006

DISSERTATION

Jingxin Xia

The Graduate School
University of Kentucky
2006

DYNAMIC FREEWAY TRAVEL TIME PREDICTION USING
SINGLE LOOP DETECTOR AND INCIDENT DATA

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering
at the University of Kentucky

By

Jingxin Xia

Lexington, Kentucky

Co-Director: Dr. Mei Chen, Professor of Civil Engineering
and Dr. Nick Stamatiadis, Professor of Civil Engineering

Lexington, Kentucky

2006

Copyright © Jingxin Xia 2006

To my daughter

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude for having the opportunity to complete this dissertation. In the trials and tribulations of this process, it would never be possible without the support, guidance, encouragement, and confidence from my advisor Dr. Mei Chen, who exemplifies the high quality scholarship to which I aspire. Her challenges and comments benefit this dissertation and will benefit my future research career as well

I want to thank Dr. Nick Stamatiadis, Dr. Jerry G. Rose, Dr. Paul L. Cornelius, and Dr. Louis M. Brock, for severing on my advisory committee. I am grateful to Dr. Nick Stamatiadis for his constructive suggestions on this dissertation that improves its quality greatly. My special thanks goes to Dr. Paul L. Cornelius for the suggestions and discussions on the statistical analyses in this dissertation. I admire his keen insight and broad knowledge in statistics.

My final thanks goes to California Performance Measurement System for the generous sharing of the archived traffic operations data to third party researchers.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES	ix
LIST OF FILES	xi
CHAPTER 1 INTRODUTION.....	1
1.1 BACKGROUND	1
1.2 THEORETICAL TRAVEL TIME DEFINITION.....	2
1.3 STATEMENT OF THE PROBLEM.....	3
1.4 RESEARCH OBJECTIVES AND CONTRIBUTION	4
1.5 RESEARCH METHDOLOGIES	5
1.6 ORGANIZATION OF THE DISSERTATION.....	6
CHAPTER 2 LITERATURE REVIEW	7
2.1 INTRODUCTION	7
2.2 TRAVEL TIME ESTIMATION FROM DIRECT TRAVEL TIME MEASUREMENT	7
2.2.1 Travel Time Estimation from Floating Car or Test Vehicle.....	8
2.2.2 Travel Time Estimation from ITS Probe Vehicle.....	9
2.2.3 Travel Time Estimation from License Plate Matching.....	11
2.3 TRAVEL TIME ESTIMATION FROM TRAFFIC MEASUREMENTS OTHER THAN TRAVEL TIMES	12
2.3.1 Travel Time Estimation from Dual Loop Detectors.....	12
2.3.2 Travel Time Estimation from Single Loop Detectors.....	14
2.4 TRAVEL TIME ESTIMATION FROM DATA FUSION TECHNOLOGIES	19
2.5 SHORTCOMING S OF EXISTING RESEARCHES.....	20
CHAPTER 3 DATA COLLECTION AND PRE-PROCESSING	22
3.1 DESCREPTION OF STUDY CORRIDOR	22
3.2 DATA DESCRIPTION	23
3.2.1 Flow Data Description	23
3.2.2 Incident Data Description	26
3.3 DATA PRE-PROCESSING	26
3.3.1 Traffic Flow Data Pre-Processing.....	27
3.3.2 Incident Data Pre-Processing.....	30
CHAPTER 4 DYNAMIC TRAFFIC FLOW PREDICTION.....	32
4.1 OVERVIEW	32
4.2 ADOPTION OF THE SARIMA MODEL	34
4.2.1 SARIMA Model Introduction.....	36

4.2.2	Data Transformation	38
4.2.3	Model Decomposition.....	39
4.3	SEASONAL OPERATOR.....	41
4.4	SHORT-TERM OPERATOR.....	42
4.4.1	Kalman Filter Design.....	42
4.4.2	Adaptive Kalman Filter Implementation	43
4.5	PROPOSED SARIMA MODEL ADEQUACY ANALYSIS	46
4.5.1	Inspection of Seasonally Adjusted Series W_t	46
4.5.2	Inspection of Residual Series ε_t from the Short-Term Operator	49
4.6	TRAFFIC FLOW PREDICTION	54
4.6.1	One-Step-Ahead Prediction	54
4.6.2	Multi-Step-Ahead Prediction.....	61
4.7	SUMMARY	66
CHAPTER 5	CORRIDOR TRAVEL TIME PREDICTION WITHOUT CONSIDERING INCIDENTS	67
5.1	OVERVIEW	67
5.2	MEVL ESTIMATION.....	67
5.2.1	Free-Flow-Speed Estimation	68
5.2.2	MEVL Estimation.....	69
5.3	CORRIDOR TRAVEL TIME ESTIMATION MODEL.....	70
5.3.1	Link Travel Time Derivation.....	70
5.3.2	Corridor Travel Time Prediction	71
5.4	PERFORMANCE ANALYSIS	73
5.4.1	MEVL Estimation.....	73
5.4.2	Corridor Travel Time Prediction	77
5.5	SUMMARY	84
CHAPTER 6	CORRIDOR TRAVEL TIME PREDICTION CONSIDERING INCIDENTS	85
6.1	OVERVIEW	85
6.2	PREDICTION OF INCIDENT DURATION.....	86
6.3	CORRIDOR TRAVEL TIME ADJUSTMENT UNDER AN INCIDENT.....	88
6.3.1	Identification of Incident Impact on Traffic	88
6.3.2	Proposed Method for Corridor Travel Time Adjustment	89
6.4	PERFORMANCE ANALYSIS	97
6.4.1	Structure of Corridor Travel Time Prediction Adjustment System	97
6.4.2	Performance Analysis	99
6.5	SENSITIVITY ANALYSIS	105
6.5.1	Flow Rate.....	106
6.5.2	Occupancy.....	108
6.5.3	Mean Effective Vehicle Length.....	110
6.6	COMPARATIVE EVALUATION.....	112
6.6.1	Choice of Algorithms for Comparison	112
6.6.2	Implementation of Algorithms.....	114
6.6.3	Performance	114
6.7	DESIGN OF THE FULL ON-LINE CORRIDOR TRAVEL TIME PREDICTION SYSTEM.....	115
6.8	SUMMARY	118

CHAPTER 7 CONCLUSIONS AND FUTURE RESEARCH	119
7.1 CONCLUSIONS.....	119
7.2 RECOMMENDATIONS FOR FUTURE RESEARCH.....	120
APPENDIX A OPTIMAL VALUES OF BOX-COX TRANSFORMATION PARAMETER	122
APPENDIX B CORRIDOR TRAVEL TIME PREDICTION AND ADJUSTMENT UNDER INCIDENTS.....	123
REFERENCES	128
VITA.....	139

LIST OF TABLES

Table 3.1 Traffic Flow Data File Format.....	24
Table 3.2 Selected Dual Loop Detector Configurations within the Study Corridor....	25
Table 3.3 Incident Data File Format.....	26
Table 3.4 Data Screening Criteria Used in This Study.....	29
Table 3.5 Definitions of Categories of the Potential Variables.....	31
Table 4.1 Autocorrelation Check of Flow Rate Residuals from ARMA(1,1) Model at VDS 401079 on May 11, 2006.....	53
Table 4.2 Autocorrelation Check of Occupancy Residuals from ARMA(1,1) Model at VDS 401079 on May 11, 2006.....	54
Table 4.3 Performance Evaluation of Flow Rate and Occupancy Prediction.....	61
Table 4.4 Performance of Multi-Step-Ahead Prediction of Flow Rate at VDS 401079 on May 12, and May 13, 2006.....	65
Table 5.1 Results of Free-Flow-Speed Estimation.....	75
Table 5.2 Performance of MEVL Estimation at All Vehicle Detector Stations.....	76
Table 5.3 Predicted and Actual Corridor Travel Time from 17:45 to 18:50 on June 7, 2006.....	81
Table 5.4 Traffic Measurements from 9:20 am to 9:35 am at VDS 401239, 401052 and 400329.....	82
Table 5.5 Performance of the Corridor Travel Time Prediction by Time of Day.....	83
Table 6.1 Multi-Way ANOVA Results of Incident Duration.....	99
Table 6.2 Look up Table of Incident Duration.....	100
Table 6.3 Incidents Identified as Having Great Impacts on the Traffic.....	101
Table 6.4 Sensitivity Analysis to Upstream Flow Rate.....	107
Table 6.5 Sensitivity Analysis to Upstream Occupancy.....	109
Table 6.6 Sensitivity of MAPE to Upstream MEVL.....	111
Table 6.7 Results of Comparative Analysis.....	115
Table B.1 Corridor Travel Time Prediction and Adjustment under Incident 1 on June 1, 2006, Thursday.....	123

Table B.2 Corridor Travel Time Prediction and Adjustment under Incident 2 on June 7, 2006, Wednesday	124
Table B.3 Corridor Travel Time Prediction and Adjustment under Incident 3 on June 9, 2006, Friday	125
Table B.4 Corridor Travel Time Prediction and Adjustment under Incident 5 on June 19, 2006, Monday	126
Table B.5 Corridor Travel Time Prediction and Adjustment under Incident 3 on June 30, 2006, Friday	127

LIST OF FIGURES

Figure 1.1 A Temporal and Spatial Illustration of Link Travel Time	2
Figure 2.1 General Traffic Changes over a Short Link during a Short Time Interval.	17
Figure 3.1 Study Corridor Map.....	23
Figure 3.2 Sample of 5-Minute Traffic Flow Data	27
Figure 4.1 Components of the Proposed Methodology for Corridor Travel Time Prediction	33
Figure 4.2 Weekly and Daily Variation in Flow Rate Series	35
Figure 4.3 Adaptive Kalman Filter Operation	45
Figure 4.4 Autocorrelation Function Plots of the Seasonally Adjusted Series W_t at VDS 401079 on May 11, 2006	48
Figure 4.5 Seasonally Adjusted Series of Flow Rate and Occupancy at VDS 401079 on May 11, 2006	50
Figure 4.6 Autocorrelation Plots of the flow rate and Occupancy Residuals from ARMA(1,1) Model at VDS 401079 on May 11, 2006	52
Figure 4.7 Structure of Online One-Step-Ahead Traffic Flow Prediction System.....	56
Figure 4.8 Predicted and Observed Flow Rate at VDS 401079	58
Figure 4.9 Predicted and Observed Occupancy at VDS 401079	59
Figure 4.10 Structure of Online Multi-Step-Ahead Traffic Flow Prediction System..	64
Figure 5.1 Typical Link of Freeway	70
Figure 5.2 Illustration of Corridor Travel Time Prediction	72
Figure 5.3 Cluster Results in the Plots of Flow Rate versus Occupancy at VDS 400329.....	74
Figure 5.4 Structure of Online One-Step-Ahead Corridor Travel Time Prediction System.....	78
Figure 5.5 Predicted and Real Corridor Travel Time by Time-of-Day on June 11, 2006, Sunday.....	80
Figure 5.6 Predicted and Real Corridor Travel Time by Time-of-Day on June 7, 2006, Wednesday	80
Figure 6.1 Illustration of Incident Duration.....	86

Figure 6.2 Incident Duration Frequency Distributions on the Study Corridor	88
Figure 6.3 Corridor Travel Time Components under an Incident	90
Figure 6.4 Time-Space Diagram of Congestion Caused by an Incident.....	92
Figure 6.5 Structure of Online Corridor Travel Time Prediction Adjustment System....	98
Figure 6.6 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 1	102
Figure 6.7 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 2	103
Figure 6.8 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 3	104
Figure 6.9 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 5	104
Figure 6.10 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 6.....	105
Figure 6.11 Cumulative Traffic Count on 06/04/2006 at VDS 401079 and 401239.	113
Figure 6.12 Structure of the Full Online Corridor Travel Time Prediction System..	117

LIST OF FILES

DissertationJingxinXia.pdf

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Travel time can be defined as the period of time to transverse a route between any two points of interest. It is a fundamental measure in transportation. Travel time is also one of the most readily understood and communicated measure indices by a wide variety of users, including transportation engineers, planners, and consumers, yet it is rigorous enough for technical analyses.

Travel time data is useful for a wide range of transportation analyses including congestion management, transportation planning, and traveler information. Congestion management systems commonly use travel time-based performance measures to evaluate and monitor traffic congestion. Planners use travel time to evaluate transportation facilities and plan improvements. In addition, some metropolitan areas provide real-time travel time prediction as part of their advanced traveler information systems (ATIS). By obtaining short-term predictive travel times for several candidate routes between their origin and destination, travelers are able to make smart decisions on route choice, and hence possibly avoid congestion. Alternatively, as a key input for dynamic route guidance systems, travel time-based measurements also enable generation of the shortest paths between an origin and destination.

Travel time data can be obtained through a number of methods. Some of the methods involve direct measures of travel times along with test vehicles, license plate matching technique, and Intelligent Transportation System (ITS) probe vehicles. Additionally, various sensors (e.g. inductance loop detectors, acoustic sensors) in ITS deployment collect a large amount of traffic data every day, especially in metropolitan areas. Such data can be used for travel time estimation for extensive applications when direct measurements of travel times are not available.

1.2 THEORETICAL TRAVEL TIME DEFINITION

On a macroscopic level, the temporal and spatial characteristics of traffic are usually described by three fundamental variables: flow rate (the number of vehicles passing a point, such as vehicle detector station, per unit time), speed (the distance per unit time), and density (the number of vehicles per unit distance) or occupancy (the percentage of time the detector is occupied by vehicles). Among the three traffic variables, speed is closely related to the concept of travel time. Travel time is the inverse of space-mean speed.

A link is defined as the section between two consecutive stations with positions x_u and x_d in a discrete time-space domain as shown in Figure 1.1, the representative travel time of the link during the time interval from t and $t+1$ can be calculated as the mean travel time within the area (i.e zone ABCD).

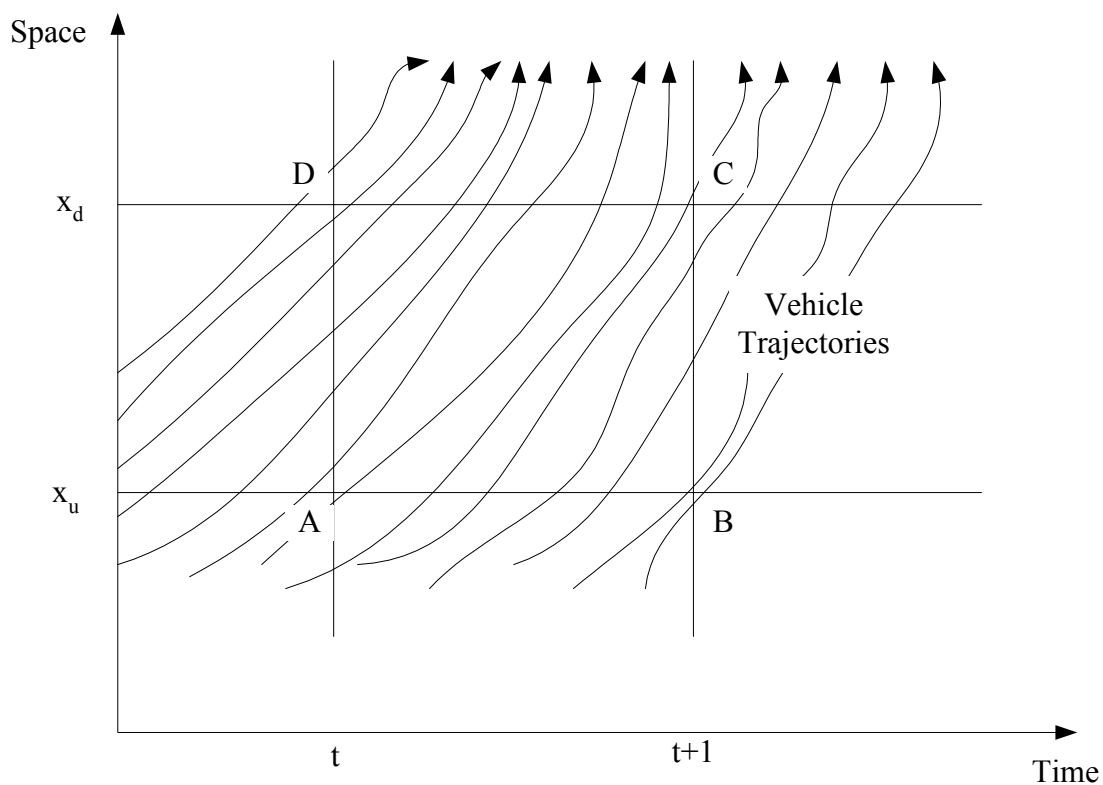


Figure 1.1 A Temporal and Spatial Illustration of Link Travel Time

From time t to $t+1$, assume that there are N vehicles traversing the link between x_u and x_d , the true space-mean speed for vehicles is equal to total traveled distance divided by

the total travel time of all vehicles in zone ABCD (Gerlough and Huber 1975, ITE 1976). An estimate of the space-mean speed over the link is defined as:

$$\bar{V} = \frac{\sum_{i=1}^N \{ \min(x_{t+1}^i, x_d) - \max(x_t^i, x_u) \}}{\sum_{i=1}^N \{ \min(t+1, t_d^i) - \max(t, t_u^i) \}} \quad (1.1)$$

where,

\bar{V} : space-mean speed of the link during the time interval from time t to time $t+1$;

x_t^i : position of the vehicle i at time t ;

t_d^i : time when vehicle i passes through the downstream station;

t_u^i : time when vehicle i passes through the upstream station.

The true travel time calculated from the space-mean speed is defined as:

$$tt = \frac{|x_u - x_d|}{\bar{V}} \quad (1.2)$$

where,

tt : representative travel time of the link during the time interval from time t to time $t+1$;

\bar{V} : space-mean speed of the link during the time interval t to $t+1$.

1.3 STATEMENT OF THE PROBLEM

Since the trajectories of individual vehicles are not available, various methods have been proposed to approximately estimate the travel time from traffic measurements collected by the advance surveillance technologies. Most of these methods were developed based on data generated by dual-loop detectors, which contain average spot speeds. Such methods are limited for extensive applications because single-loop detectors (and similar devices) are the most commonly used devices in traffic monitoring systems.

Currently, a reliable methodology has not been developed for travel time prediction based on data generated by single loop detectors due to the lack of speed measurements.

Much of the work performed to date focuses on travel time estimation instead of prediction. These works merely report the travel time at the time traffic data are collected. While this is important for the application of traffic system monitoring and performance evaluation, there is a need for research that studies the short-term prediction of travel time for various applications such as a route guidance system.

Previous work has not effectively considered the effects of traffic progression along the freeway in the travel time estimation process as well. The corridor travel time is simply estimated as the total of the link travel times estimated during the same time interval. This may cause significant divergence of the estimated travel time from the “ground truth”. Consideration of traffic progression along a freeway is clearly of value for corridor travel time estimation and prediction. In this sense, there is also a need to acknowledge the effects of traffic progression in the short-term prediction of corridor travel time.

Furthermore, most previous methods for travel time estimation were developed under normal traffic conditions or during recurrent congestion. The impact of incidents on travel time estimates has not been effectively considered. Under incident situations, sudden changes are often observed in traffic measurements, and long, unanticipated delays are often caused by accidents when non-recurrent congestion forms. Although incidents have nearly negligible effects on the overall performance of these methods because of the low probability of incident occurrence on freeways, for traveler information purposes, incident impacts need to be accounted for in the development of a travel time prediction model.

1.4 RESEARCH OBJECTIVES AND CONTRIBUTION

This research aims to develop a robust online short-term corridor travel time prediction system based on continuous traffic flow data generated by single loop detectors and incident data generated by the traffic monitoring system. The research objectives are expected to provide valuable information in the areas of:

- Dynamic prediction of flow rate and occupancy over time using traffic measurements from single loop detectors;
- Short-term prediction of corridor travel time integrating the dynamic traffic flow predictor and accounting for the effects of traffic progression along a freeway in model development; and

- Short-term corridor travel time adjustments under an incident.

Compared with previous work for corridor travel time prediction from single loop detectors, there are several contributions from this study. First, this dissertation integrates dynamic traffic flow prediction and the effects of traffic progression along a freeway into the model development for corridor travel time prediction. Moreover, this study integrates the impacts of incidents on the traffic in the corridor travel time prediction system for online implementation. Both contributions are valuable to improve corridor travel time prediction accuracy, particularly under congested traffic conditions, or with the occurrence of an incident. Additionally, the proposed methodology can work with less than perfect data sources. This provides a good opportunity for the wide application of the proposed method.

1.5 RESEARCH METHDOLOGIES

A corridor is selected for data collection in this dissertation. Both traffic flow measurements and incident information are collected from California Freeway Performance Measurement System (PeMS) along I-80 in the Bay Area of California. The collected traffic flow data are aggregated in 5-minute intervals. Each record provides aggregated flow rate, occupancy, and average spot speed. Flow rate and occupancy are used as traffic measurements for model development, while the average spot speed is used for the estimation of actual corridor travel time for model performance evaluation. The incident data includes the start time, location, actual incident duration, and incident type. After these data are obtained, they are stored and pre-processed in an operational database for model testing.

The proposed corridor travel time prediction model starts with the model development for dynamic traffic flow prediction. This model involves the multi-step-ahead prediction of flow rate and occupancy in real time. A seasonal autoregressive integrated moving average (SARIMA) model is developed with an embedded adaptive predictor. This predictor adjusts the prediction error based on traffic flow data that becomes available every five minutes at each vehicle detector station.

The corridor travel time prediction model is developed based on short-term prediction of link travel times, in which a link is defined as the section between two consecutive vehicle detector stations. Beginning from the first link, the corridor travel time prediction model predicts the link travel time in sequence until the last link travel time has been predicted. To

consider the effects of traffic progression along the corridor, the dynamic traffic flow predictor is embedded in the corridor travel time prediction model with a varied number of steps in advance.

The proposed model also considers the incident impacts on traffic. When an incident happens on the corridor, its impact on the traffic is first identified. If the incident affects traffic significantly, the predicted corridor travel times within the influence time of the incident are adjusted. Otherwise, it is assumed that the incident has little impact on the traffic. To adjust the corridor travel times, incident duration is predicted based on a look-up table by incident type and day-of-week from historical incident information. The final corridor travel time prediction adjustment is performed based on shock wave analysis.

1.6 ORGANIZATION OF THE DISSERTATION

This dissertation has been organized into seven chapters. Chapter 1 includes an introduction to the research and discusses the background, problem statement, research objectives, methodologies, contribution of research, and the organization of the dissertation. Chapter 2 provides a literature review of previous work on travel time data collection with emphasis on the travel time estimation from ITS deployment. Chapter 3 presents the study corridor and data collection for testing the proposed corridor travel time prediction methodology. Chapter 4 presents the dynamic traffic flow prediction model. Model adequacy and performance analysis are included. Chapter 5 presents the corridor travel time prediction model without considering the incident impacts on the traffic and the testing results. Chapter 6 presents the method for the corridor travel time prediction adjustment due to the occurrence of an incident. Sensitivity analysis is also included to identify the most important factors for using the proposed method in corridor travel time prediction. Furthermore, comparative analysis is also performed. Finally, conclusions and recommendations for future researches are presented in Chapter 7.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

There have been various studies involving travel time estimation and prediction. The methods used can generally be grouped based on data source. For applications with direct travel time measurements, such as those obtained from probe vehicles or license plate matching techniques, the method of averaging measured travel times has been used. When direct measurement of travel time is not available, various methods have been employed to perform an estimate using traffic measurements from advanced traffic surveillance technologies such as dual loop detectors and single loop detectors. These methods were mainly developed based on relationships between traffic variables.

Relevant literature on travel time estimation and prediction for both groups of methods is described in this chapter. Previous works on travel time estimation from direct measurement of travel times is presented along with travel time estimates from floating cars or test vehicles, ITS probe vehicles, and license plate matching. Methods for travel time estimation from traffic measurements (e.g. flow rate, occupancy, and speed) measured by field sensors and similar devices are also presented. Furthermore, previous work related to obtaining more accurate travel times by fusing multiple data sources from multiple technologies is reviewed.

2.2 TRAVEL TIME ESTIMATION FROM DIRECT TRAVEL TIME MEASUREMENT

Techniques for direct travel time data collection include methods utilizing floating cars or test vehicles, intelligent transportation system (ITS) probe vehicles, and license plate matching. The relevant literature using these technologies for travel time estimation is described as follows.

2.2.1 Travel Time Estimation from Floating Car or Test Vehicle

The floating car or test vehicle technique for travel time data collection and estimation was adopted in the 1920's, but the first comprehensive research was performed in the late 1940's (Berry and Green 1949) and early 1950's (Berry 1952). Travel time estimation from floating car or test vehicle was the most common method for travel time data collection and estimation for early research. It utilizes data collection vehicles containing an observer who records cumulative travel time at predefined checkpoints along a travel route.

In order to accurately estimate travel time using test vehicles, a minimum sample size is necessary. This dictates a required number of test vehicles must transverse a given roadway during the time period of interest, such as time-of-day and day-of-week. The equation to calculate the minimum sample size is given as $(\frac{T - statistics \times c.v}{e})^2$ (Turner et al. 1998), in which T-statistic is the value from the Student's distribution for $n - 1$ degrees of freedom, $c.v$ is the coefficient of variation, and e is the maximum specified relative error.

The coefficient of variation for travel time varies widely, depending upon the physical and traffic control characteristics as well as traffic conditions. A study by Berry and Green (1949) of three arterial corridors in California found that the coefficient of variation for urban arterials ranged from 9 to 16 percent. A subsequent study by Berry (1952), which utilized the same methods, found that the coefficient of variation for urban arterials ranged from 5 to 17 percent. Several other empirical studies indicate that the coefficient of variation ranged from 8 to 17 percent (May 1990). A recent study by the National Cooperative Highway Research Program (NCHRP) (Lomax et al. 1997) not only confirmed these estimates, but also suggested that the coefficient of variation for freeways ranges from 9 to 17 percent, depending upon the average daily traffic volume per lane.

Given the ranges of coefficients of variation, a range representing the number of test vehicles can be obtained, but the maximum number of vehicles is often adapted with test vehicles driving on roadways with evenly distributed headway of 30 minutes (Turner et al. 1998). The Institute of Transportation Engineers (ITE 1994) suggested that the calculation of sample size should be based upon the average range of the coefficient of variation. Subsequent research by Quiroga and Bullock (1998) questioned the validity of ITE's sample size, but both sets of research suggested conducting several travel time runs and then computing range values and corresponding travel times. The Federal Highway

Administration (FHWA) provided completely updated sample size information as well as the procedure for travel time data collection using test vehicles for both arterial and freeway corridors (Turner et al. 1998).

2.2.2 Travel Time Estimation from ITS Probe Vehicle

Since the early 1990's, travel time estimation from probe vehicles has drawn lots of concerns with the increasing development of ITS technologies. Compared with floating cars or test vehicles designed for travel time data collection only, intelligent transportation system (ITS) probe vehicles were not initially designed for real time travel time data collection, but for other specific data collection purposes, such as real-time traffic operations monitoring, incident detection, and route guidance (Turner et al. 1998). However, the information collected by ITS probe vehicles may be used for travel time estimations as well.

Depending on the technologies used in the ITS probe vehicle system, the methods for travel time data collection can be classified into two groups: automatic vehicle location (AVL) systems and automatic vehicle identification (AVI) systems. AVL systems measure travel times by identifying probe vehicle positions through in-vehicle systems. Specific technologies include ground-based radio navigation (Vaidya et al. 1996), global positioning systems (GPS) (Guo and Poling 1995, Roden 1996, Gallagher 1996, Laird 1996, Quiroga and Bullock 1999, James et al. 2000, Choi and Chung 2001, Yim and Cayford 2001, Ngo 2005), and cellular phone tracking (Levine et al. 1993, Larsen 1996, Ygnace et al. 2000, Yim and Cayford 2001, Smith et al. 2003). Ground-based radio navigation is often used for transit or commercial fleet management, where data are collected by communication between probe vehicles and a radio tower infrastructure. Similarly, probe vehicles equipped with GPS receivers send and receive signals from earth-orbiting satellites to collect travel times between two locations along the roadway. Cellular phone tracking methods collect travel time by discretely tracking cellular telephone call transmissions. Comparatively, AVI systems identify vehicles through fixed roadside systems. One example of AVI is probe vehicles that are equipped with electronic tags, which can be used to communicate with roadside transceivers to identify unique vehicles and collect travel times between transceivers.

The key idea behind using ITS probe vehicles for travel time estimation is that a probe vehicle traveling in traffic should be a reasonable representation of the characteristics of the traffic. A sufficiently large number of probe vehicles should be representative of the traffic conditions experienced. There have been several studies discussing the appropriate

probe vehicle percentage, as well as reporting frequency to ensure reliable travel time estimation (Van Aerde et al. 1993, Turner 1995, Srinivasan and Jovanis 1995, Sen et al. 1997, Hellinga and Lu 1999, Chen and Chien 2000, Cheu et al. 2002).

Van Aerde et al. (1993) developed an analytical expression for the reliability of probe vehicle travel times for signalized links and verified these expressions using simulated data. This study indicated that as the number of probe vehicles increases, the sample mean approaches the population mean. A similar result was reached by Srinivasan and Jovanis (1995), but concluded that the number of probe vehicles increases non-linearly as the reliability criterion is made more stringent, and that more probe vehicles are required for shorter measurement periods. A subsequent study by Sen et al. (1997) questioned the conclusion reached by Van Aerde et al. (1993) and Srinivasan and Jovanis (1995), and concluded that the standard error was not substantially improved by making the number of probe vehicles much larger. All conclusions reached by Van Aerde et al (1993) Srinivasan and Jovanis (1995), and Sen et al. (1997) were proved to be correct, but each was appropriate only for specific traffic and sampling conditions (Hellinga and Fu 1999).

Factors that affect the minimum sample size required for probe vehicle travel time were also studied. Based on the examination of probe vehicle travel time data from the Houston AVI traffic monitoring system, Turner (1995) recommended the minimum required sample size for different roadways in the Houston area, and concluded that the average segment speed would be required to estimate the probe vehicle sample sizes. The coefficient of determination, R^2 , between the travel time variation and the average speed was found to be 0.60, which means that 60% of the variability in travel times can be described as the average segment speed. Subsequent studies showed that other variables might also be used for the variation of probe vehicle travel times. For example, a study by Chen and Chien (2000) suggested that link travel time variation depends on traffic demand levels and the geometric conditions.

Due to the limitation of minimum sample size requirements for probe vehicle travel time estimation, numerous attempts were made to find alternatives for probe vehicles. Sanwal and Walrand (1995) developed a framework for the operation of a scheme using moving vehicles as traffic probes. The evaluation results indicated that the fractions of vehicles required to serve as probes is a function of the desired performance. Some recent work performed at the University of Washington developed a method to estimate travel time using transit vehicles as probes in which a mass transit tracking system was developed based

on the automatic vehicle location data and Kalman filter to estimate vehicle position and speed, and thus travel times (Cathey and Dailey 2002, Dailey and Cathey 2002, Cathey and Dailey 2003).

2.2.3 Travel Time Estimation from License Plate Matching

Estimation of travel time from license plate matching is another common method for directly collecting travel time. It consists of collecting vehicle license plate numbers and arrival times at various checkpoints along the roadway. By matching license plate numbers between two consecutive checkpoints, travel times can be calculated from the difference in arrival times. Earlier studies include that performed by Berry and Green (1949) and Berry (1952), in which a thorough analysis and comparison of license plate matching and floating car and test vehicle techniques was performed.

Compared with the technologies of test vehicles and ITS probe vehicles, license plate matching does not suffer from the concern of sample size because data collection includes a large number of vehicles. However, the number and location of checkpoints should be designed along a route varying with the character of the roadway and street network configurations. The FHWA provided a complete, updated list for number and location of checkpoints for travel time data collections using license plate matching techniques for both arterial and freeway corridor (Turner et al 1998).

There are various ways for collecting license plates. Typically, manual, portable computer and video can be used. Manual methods collect license plate numbers via pen and paper. Guidelines for collecting travel time data from the manual collection of license plates were provided by Schaefer (1988), in which practical issues and detailed considerations were discussed. Portable computer methods collect license plate numbers in the field using portable computers that automatically provide an arrival time stamp. Such a method drew a lot of concerns in the 1990's. Studies include those by Rickman et al. (1990) and Washburn and Nihan (1997) in Seattle area, and Bailey and Rawling (1991) in Chicago area, Liu and Haines (1996) in Seattle, Washington, and Lexington, Kentucky, and Lomax (1997) and Turner et al (1994) in Texas around Houston area. Due to the limitation of under-developed technology for transcribing license plates automatically, this method was not commonly used until a decade ago with the increasing developments of image recognition technologies. From then on, various algorithms for license plate recognition were developed (Shuldiner et

al. 1996, Turner and Woodson 1996, Washburn and Nihan 1999, Dailey and Li 2000, Angel and Hickman 2002, Gupta et al. 2002, Chang et al. 2004).

2.3 TRAVEL TIME ESTIMATION FROM TRAFFIC MEASUREMENTS OTHER THAN TRAVEL TIMES

Although direct measures of travel time can be employed to obtain more accurate travel time, they have shown several shortcomings: 1) special equipment is commonly required which may only cover a small area of the highway system if placed on roadsides, or is installed in a low percentage of vehicles; 2) limited data is available; and (3) active participation from vehicle owners is generally required, which may compromise personal privacy. With the increasing applications of advanced surveillance technologies, which provide traffic flow data enhanced in terms of consistency and efficiency, estimation of travel times from aggregated traffic measurements was proposed by a lot.

Although there are a large number of advanced surveillance technologies that can be used to generate traffic measurements, most of them, such as ultrasonic- or infrared-based surveillance, provide only point measurements of flow rate, occupancy, and average speed (Nam et al. 1996). Consequently, the methods for travel time estimation were developed mostly from these point traffic measurements.

Methods for estimating travel time from traffic measurements other than travel time can be classified into two groups. One group uses point traffic measurements of flow rate, occupancy and average spot speed from dual loop detectors (or similar devices). The other estimates travel time merely by using point traffic measurements of flow rate and occupancy from single loop detectors (or similar devices).

2.3.1 Travel Time Estimation from Dual Loop Detectors

Dual loop detectors are placed on a freeway a fixed distance apart, approximately 12 feet, producing a more accurate spot speed estimate between the loop detectors. Provided that the average spots speed over short periods is accurate (if the detector spacing is small) estimates of the true travel time can be obtained based upon the known distance between two adjacent loop detectors (Coifman 2002, Van Lint and Van Der Zijpp 2004, Chen 2004).

A widely used method to estimate travel time from dual loop detector measurements is using piece-wise constant speeds. Assuming an individual vehicle driving on a link during period h at a constant speed V_h , the travel time required for the vehicle to pass through the link can be described by

$$tt = \frac{l}{V_h} \quad (2.1)$$

where l is the link length. This equation allows one to compute the travel times on the link during different periods. However, the constant speed V_h may fluctuate among different drivers (Van Lint and Van Der Zijpp 2004).

Since the time-mean speeds averaged over fixed time intervals are provided by the dual loop detectors, the individual speeds V_h are often substituted by a harmonic time-mean speed given by (Thijs et al. 1999, Van Der Zijpp and Lindveld 1999, Lindveld et al. 2000, Kazimi et al. 2000, Van Lint and Van Der Zijpp 2004)

$$V_h = 2 \left(\frac{1}{V(d,h)} + \frac{1}{V(d+1,h)} \right)^{-1} \quad (2.2)$$

in which the link is considered as the segment enclosed between detector d and $d+1$, and $V(d,h)$ equals the time-mean speed measured at detector d during period h .

Lindvels and Thijs (1999) demonstrated that when the time-mean speeds over fixed time intervals are stationary, the harmonic time-mean speed is equal to the space-mean speed. Otherwise, significant bias might be caused by the difference between the space-mean and time-mean speeds. This may deteriorate the travel time estimation performance.

Travel time estimation using the piece-wise constant speed results in piece-wise speeds over the different roadway links, where vehicles are thought to instantaneously change their driving speed once they enter a new roadway link. In order to utilize the average speed in a smoother fashion, a liner function of speed is provided by Van Lint and Van Der Zijpp (2004) to substitute the constant speed V_h . The function is given by

$$V(t) = V(d,h) + \frac{x(t) - x_d}{l_c} [V(d+1,h) - V(d,h)] \quad (2.3)$$

where,

$V(t)$: the average spot speed at time t during period h ;

$x(t)$: the location of vehicle at time t during period h on the link;

x_d : the location of detector d .

Evaluation of the travel time estimation using one-minute aggregated data from simulation shows that the method using a linear function of speed outperformed that using the harmonic time-mean speed, where residual error is significantly reduced (Van Lint and Van Der Zijpp 2004).

Chen (2004) improved the method by interpolating the average speed as a linear function of both distance and time by iteratively calculating the actual link travel times. First, the time mean speeds at detector d and $d+1$ respectively at time t during period h as a function of time can be estimated as

$$V(d, t) = V(d, h-1) + \frac{t-t_0}{\Delta T} [V(d, h) - V(d, h-1)] \quad (2.4)$$

$$V(d+1, t) = V(d+1, h) + \frac{t-t_0}{\Delta T} [V(d+1, h) - V(d+1, h-1)] \quad (2.5)$$

where,

$V(d, t)$: the average spot speed at time t during period h at detector d ;

$V(d+1, t)$: the average spot speed at time t during period h at detector $d+1$;

t_0 : the starting time of period h ;

ΔT : the duration of report period.

Second, the time-mean speed is estimated as a function of distance as

$$V(t, x) = V(d, t) + \frac{x(t) - x_d}{l} [V(d+1, t) - V(d, t)] \quad (2.6)$$

in which, l is the link length with boundaries at detector d and $d+1$.

2.3.2 Travel Time Estimation from Single Loop Detectors

Single loop detectors are placed on a freeway with random distance apart depending upon roadway geometry, and on-ramp and off-ramp locations, etc. Similar to dual loop

detectors, only point traffic measurements are provided by single loop detectors. These measurements include flow rate and occupancy over fixed time intervals.

Previous work for travel time estimation from single loop detectors can be classified into two groups: 1) estimating travel time by estimating speed from flow rate and occupancy; and 2) directly estimating travel time from flow rate and occupancy without estimating average speed.

2.3.2.1 Travel Time Estimation from Speed Estimation

A commonly adopted method of estimating average speed from flow rate and occupancy is to seek the fundamental relationship among space-mean speed, flow rate and occupancy. Assuming a mean effective vehicle length (MEVL) at detector d during period h , the basic form used to describe such relationship among average speed, flow rate, and occupancy is given as

$$V(d, h) = \frac{N(d, h)}{\Delta T \times o(d, h) \times g(d, h)} \quad (2.7)$$

where

$V(d, h)$: the time-mean speed at detector d during period h ;

$N(d, h)$: the vehicles at detector d during period h ;

ΔT : the duration of report period;

$o(d, h)$: the measured average occupancy at detector d during period h ;

$g(d, h)$: the mean effective vehicle length at detector d during period h .

This formulation was first developed in the 1960's, beginning with an estimator that uses a constant MEVL over time to estimate the time-mean speed, based on the assumption that occupancy is linearly proportional to density (Athol 1965). Because of its simplicity, it has been applied extensively to estimate the mean speed from single loop outputs (Mikhalkin et al. 1972, Gerlough and Huber 1976, Courage et al. 1976, Hall and Persaud 1989, Dailey 1997, Ishimaru and Hallenbeck 1999, Wang and Nihan 2000, Jia et al 2001, Coifman 2001, Eisele (2001), Lin et al. 2003, Van Zwet et al. 2003).

The process treating MEVL as a constant has been challenged extensively. Collecting data from several traffic stations, Hall and Persaud (1989) plotted MEVL versus occupancy

and found that the value of MEVL is not a constant but varies with occupancy. Jia et al. (2001) stated that the constant mean effective vehicle length might result in error in speed estimates of more than 50 percent. The analysis also showed that the mean effective vehicle length for the same detector could vary by as much as 50 percent over a 24-hour period. Because of this, algorithms have been proposed for estimating dynamic MEVL based on the traffic characteristics investigation.

An example of the use of the dynamic MEVL for the mean speed estimation is given by Wang and Nihan (2000). By collecting traffic flow data from dual loop detectors in which the mean effective vehicle length can be directly estimated, the mean effective vehicle length was found to be a function of occupancy and flow rate. The relationship among the mean effective vehicle length, flow rate, and occupancy was then applied to the mean speed estimation from single loop measurements. Results indicated that the standard error of the mean speed can be improved from 4.17 to 3.47 mph.

Coifman (2001) presented another example to provide the dynamic value of mean effective length for speed estimation. Assuming the value of free-flow speed of a specific freeway, the mean effective vehicle length under uncongested traffic conditions can be calculated using equation (2.7). Under congested traffic conditions, the mean effective vehicle length was extended from the uncongested traffic conditions. This method was developed based on the verification of little variation of mean effective vehicle length from uncongested to congested traffic conditions.

2.3.2.2 Directly Estimating Travel Time from Single Loop Measurements

Methods assuming a constant or dynamic mean effective vehicle length may cause flawed estimates for average speed. In order to avoid using the flawed estimate speed for travel time estimation, new attempts were made to estimate travel time directly from flow measurements.

The most commonly used method estimates travel time directly from single loop measurements by estimating density from the difference of cumulative traffic counts by applying the principle of vehicle conservation (Nam and Drew 1996, Chu and Recker 2004). Initially proposed by Lighthill and Whitham (1957), the concept of vehicle conservation states that the difference between the number of vehicles entering the link and those leaving it during the time interval equals to the changes in the number of vehicles traveling on the link

(Gerlough and Huber 1975). In other words, the change of number of vehicle on the roadway link over a short time interval equals to the difference of flow rates at upstream and downstream stations. Given a typical roadway link and traffic condition changes, as shown in Figure 2.1, the principle of vehicle conservation can be represented in the form (Nam and Drew 1996)

$$\frac{f(x_1, t + \Delta t) - f(x_2, t + \Delta t)}{\Delta x} = \frac{k(t + \Delta t) - k(t)}{\Delta t} \quad (2.8)$$

where,

f : the flow rate(vehicles/hour/lane);

k : is the density on the link (vehicles/mile/lane);

x : the location;

t : the time.

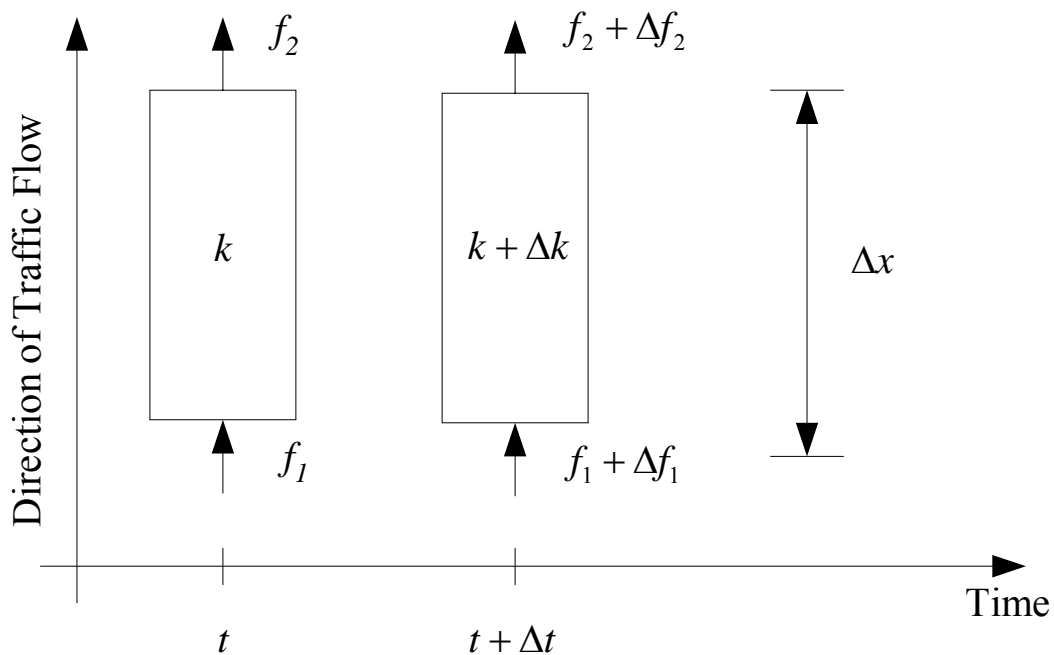


Figure 2.1 General Traffic Changes over a Short Link during a Short Time Interval

When $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$, the general form of vehicle conservation can be obtained (Lighthill and Whitham 1957, Richards 1956) as

$$\frac{\partial f(x,t)}{\partial x} + \frac{\partial k(x,t)}{\partial t} = 0 \quad (2.9)$$

For a link with detector stations i and $i+1$ at its boundaries, the traffic flow passing this link is often estimated as a linear function of the traffic flow passing the upstream and downstream station flow rates as

$$f = \varpi f_i + (1 - \varpi) f_{i+1} \quad (2.10)$$

where,

f : the flow rate passing through the link;

f_i : the measured flow rate at upstream station i ;

f_{i+1} : the measured flow rate at downstream station $i + 1$;

ϖ : a smoothing parameter.

Given the density and flow rate of the link during period h , the estimate of travel time can be computed based on the relationship $f = k \times V$ as

$$tt = \frac{l}{V_h} = \frac{l \times k_h}{f_h} \quad (2.11)$$

where,

tt : the travel time on the link;

l : the short link length;

V_h : the average speed on the link during period h ;

k_h : the density(vehicles/mile/lane) on the link during period h ;

f_h : the flow rate(vehicles/hour/lane) on the link during period h .

In Equation (2.11), the density on the link is obtained from the difference between the cumulative traffic counts at the upstream and downstream stations.

Applications of the principle of vehicle conversation by earlier works usually assume the same values for inflow rate and outflow rate at time t and time $t + \Delta t$ (Nam and Drew 1996). By definition, this is correct only when the link length Δx and the small change of time Δt approach zero. In reality, this is not true, and the link length and small change of

time are often processed as the detector-controlled link length l and the aggregated interval of traffic measurements, both of which might be much larger than zero. In other words, the temporal fluctuations were ignored and spatial fluctuations were only partially considered by making such an assumption (Nam and Drew 1996).

Two other shortcomings have also been shown using the principle of vehicle conservation for travel time estimation or other applications (Nam and Drew 1996, Chu and Recker 2004). First, density estimation from the principle of vehicle conservation requires a known initial number of vehicles on the link, which is very hard to estimate. Second, the system error in field detectors often causes an unrealistic density in the real world. Although adjustment by introducing a feedback control mechanism can be made to make up under-or over-measured traffic counts at the downstream station (Chu and Recker 2004), the full system error cannot be completely solved.

2.4 TRAVEL TIME ESTIMATION FROM DATA FUSION TECHNOLOGIES

Methods presented so far use data sources from only single surveillance technology such as ITS probe vehicle. Numerous attempts have also been made to estimate travel time by fusing data sources from single or multiple technologies to improve the estimation accuracy.

Depending upon the data fusion technologies used, methods for travel time estimation from fusing multiple data sources include weighted average models (Tarko and Roupail 1993, Barka et al. 1995, Choi and Chung 2001, and Xie et al. 2004), Kalman filtering models (Takahashi et al. 1996, Arem et al. 1997, Al_Deek 1998, Pourmollem et al. 2004, and Chu and Recker 2004), and historic and real-time profile models (Hoffman and Janko 1990, Rilett 1992, Boyce et al. 1993, and Tarko et al. 1993).

The basic idea of weighted average approach is that travel time is estimated by assigning different weights to the travel time estimates from different data sources. A typical example of this approach is given by a recent research by Choi and Chung (2001). Consider there are n estimates of travel times with each estimator $TT_i(x)$ given a value at $x = (x_i)_{i=1,2,n}$, the resulting value of travel time estimation TT from the weighted average approach is given by

$$TT = \frac{\sum_{i=1}^n w_i TT_i(x)}{\sum_{i=1}^n w_i} \quad (2.12)$$

where w_i is the weight reflecting the reliability of the travel time estimates from a specific data sources.

Kalman filtering approaches focus on the construction of state and measurement equations for travel time estimation from different data sources. Chu and Recker (2004) estimates travel times from both ITS probe vehicle data and single loop measurements. By constructing the state and measurement equations using the principle of vehicle conservation, and considering the travel time estimate from ITS probe vehicle as the measurement variable, travel time estimation accuracy is enhanced from the single loop detector measurements.

Historical profile approaches assume that historical traffic flow data can be used to reflect traffic characteristics over days. Thus, the classification of days into day types with similar profiles can be developed. Similar to these models, both historical and real time profiles are integrated in the neural network models developed by Nelson et al. (1993), Rouphail et al. (1993), Florio and Mussone (1994), Hua and Faghri (1994), Blue et al. (1994), Pourmollem et al. (1997), Park and Rilett (1998), Park et al. (1998), and Rilett and Park (1999) for travel time estimation and prediction. In all of these studies, historical profiles are used for developing the travel time estimation and prediction model. Based on a well-developed model, travel time can be estimated or predicted based on the given real-time traffic measurements.

2.5 SHORTCOMINGS OF EXISTING RESEARCH

Given the extensive availability of traffic data collected by intelligent transportation systems, a variety of travel time estimation methods have been developed. Despite limited success under light traffic conditions, traditional corridor travel time prediction methods have suffered various drawbacks.

The first shortcoming is that most of the methods for travel time estimation are developed based on data generated by dual-loop detectors which contain average spot speeds. However, single-loop detectors (and other devices that emulate its operation) are the most

commonly used devices in traffic monitoring systems. There has not been a reliable methodology for travel time prediction based on data generated by such devices due to the lack of speed measurements.

Second, the majority of existing studies focus on travel time estimation instead of travel time prediction when using traffic measurements from single loop detectors. These methods merely report the travel time at the time the traffic flow data are collected. Moreover, the effect of traffic progression along the freeway has not been considered in the corridor travel time prediction process.

Finally, all the methods for travel time estimation were developed under normal conditions or for recurrent congestion. The performance therein is not good under traffic congestion, especially when an incident occurs (Xie et al. 2004). Sudden changes are often observed in traffic measurements, and long, unanticipated delays are often caused by accidents where non-recurrent congestion forms. Although incidents have nearly negligible effects on the overall performance of these methods because of the low probability of incident occurrence on freeways, they do affect travel times from a traveler's perspective. Existing research on travel time estimation/prediction has given limited attention to the integration of incident information. Among the most relevant work, Cohen (1999) presented a method to estimate delay due to accidents as a function of the volume-to-capacity ratio and fitted equations in different situations. However, the equations are not applicable when a queue occurs even if there is no incident. Skabardonis et al. (2003) investigated a methodology to measure total recurrent, and non-recurrent (incident related) delay on urban freeways using data from loop detectors and calculated the average and the probability distribution of delays. This method is promising, but it merely provides the estimation of average incident delay. For a corridor travel time prediction at different time steps, this method is not suitable.

CHAPTER 3

DATA COLLECTION AND PRE-PROCESSING

The implementation and testing of corridor travel time prediction model requires extensive traffic flow data at very detailed levels over a reasonably long corridor with corresponding supplementary incident data. The data collection effort in this case consists of two different phases executed simultaneously. One phase consists of collecting traffic flow data that contains flow rate and occupancy. The other phase consists of collecting incident-related data. The data collected in both phases are described in detail in the following subsections. Data pre-processing including data storage, screening, and imputation for further corridor travel time prediction is also described.

3.1 DESCRIPTION OF STUDY CORRIDOR

The corridor selected for this study is located in the Bay area of California on eastbound Interstate-80. It is a 9.006-mile long corridor in the urban area, consisting of four-lane freeway sections with abs milepost between 11.95 and 20.956. This is one of the more heavily traveled and heavily congested corridors in this urban area freeway system. Commuter traffic predominates during the morning and afternoon peaks, and a number of incidents are often observed. Furthermore, this is a freeway corridor with dual loop detectors installed, which not only provide the flow rate and occupancy used for testing the proposed methodology, but also provide average spot speed for the model evaluation.

Along the study corridor, there are a total of 25 vehicle detector stations (VDS) where one loop detector is embedded underground per lane at each station. The distance between two consecutive vehicle detector stations varies with a range from 0.05 to 0.91 miles. A map of the Bay area is provided in Figure 3.1 to give an overall view of the location of the study corridor, which is denoted by a bold black curve with the origin VDS number 401079 and destination VDS number 400865.

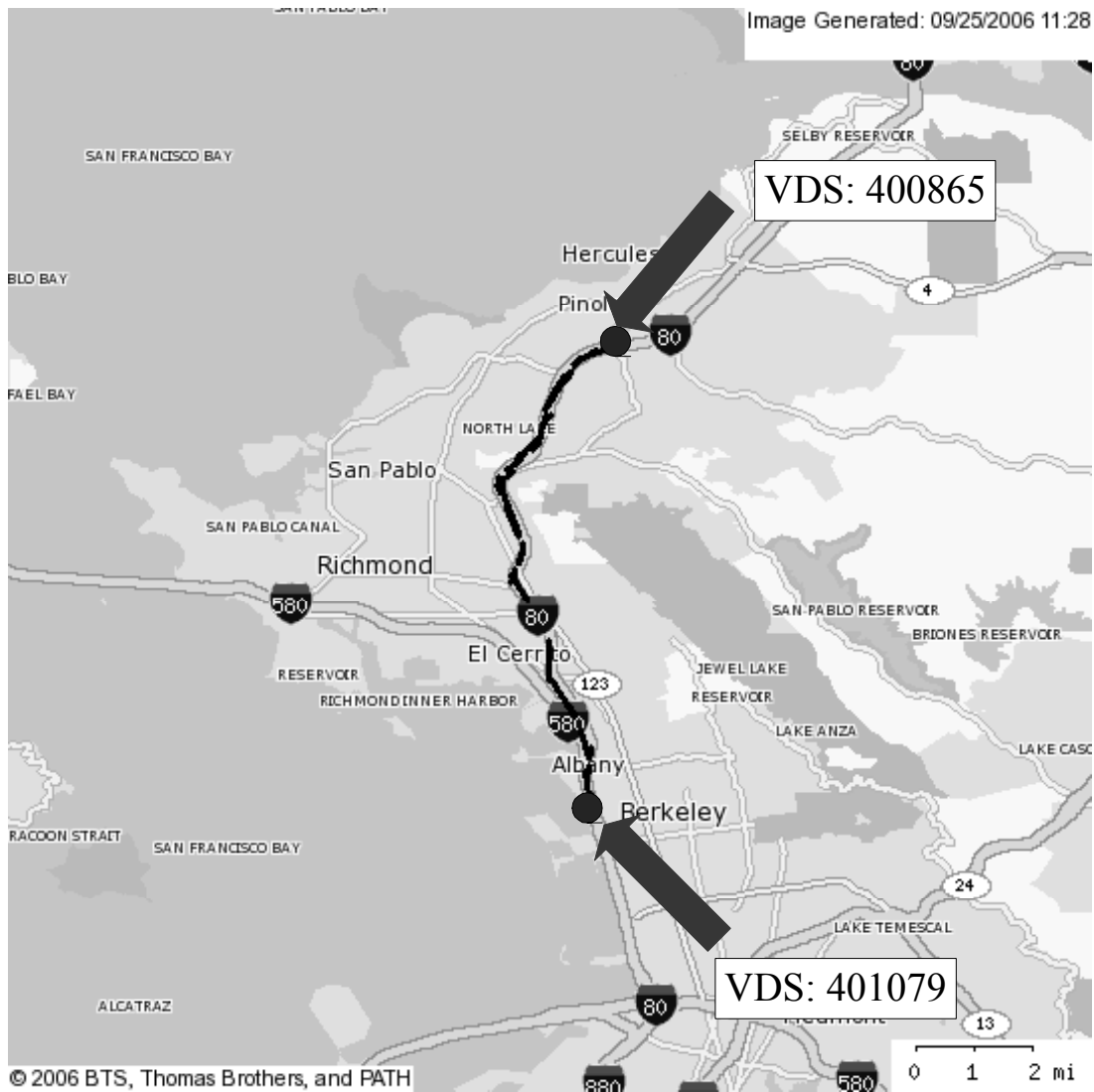


Figure 3.1 Study Corridor Map

3.2 DATA DESCRIPTION

3.2.1 Flow Data Description

In order to enhance freeway systems productivity based on real time traffic surveillance, traffic flow data including flow rate and occupancy on the study corridor are collected and stored by California Performance Measurement System (PeMS) — a system that continuously collects and stores data from California embedded lane specific loop

detectors, and converts these data into useful information. By extracting traffic flow data in the background, lane-specific traffic flow data are reported to the California PeMS every 30 seconds. This data is aggregated at 5-minute intervals. Both 30-second and 5-minute traffic flow data can be downloaded from the California PeMS site (<ftp://128.32.48.245/>) where traffic flow data is stored in the format of comma-delimited ASCII text. In this study, the 5-minute flow data are selected for the testing of the proposed corridor travel time prediction model. The detailed fields in a typical sample of 5-minute traffic flow data are demonstrated in Table 3.1 as follows:

Table 3.1 Traffic Flow Data File Format

Column	Description
Time_Stamp	Start time of the sample, e.g., 01/28/2006 13:00:00. The data is reported by the start time of sampling and not the end time of the sample. So for data between 13:00:00 and 13:05:00 the start time would be 13:00:00.
VDS_ID	Unique identifier of vehicle detector station, e.g. 401079.
FLOW	Flow (vehicles/5-minutes).
OCCUPANCY	Average occupancy as a percentage (0 - 1)
Speed	Flow-weighted average of lane speeds
VMT	Total vehicle miles traveled over this section of freeway
Q	Measure of freeway quality (VMT/VHT)
TRAVEL_TIME	Not in use
DELAY	Vehicle hours of delay
NUM_SAMPLES	Number of samples received in the 5-minute period
PCT_OBSERVED	Percentage of individual lane points from working detectors that were rolled into the station's 5-minute values.

Traffic flow data collected from May 4th, 2006 to July 3rd, 2006 are used for this study. Among the 25 vehicle detector stations, 20 vehicle stations are selected for the testing of the corridor travel time prediction system. The other 5 vehicle detector stations are discarded because they were either closed or did not work properly during the data collection period. Configurations of the selected vehicle detector stations and loop detectors are described in Table 3.2, which contains VDS identification, loop detector identification, and both California and abs mileposts.

Table 3.2 Selected Dual Loop Detector Configurations within the Study Corridor

VDS ID	Lane	Loop ID	CA MP	ABS MP	VDS ID	Lane	Loop ID	CA MP	ABS MP
401079	1	404241	6.64	11.95	401228	1	403477	4.06	17.406
	2	404242	6.64	11.95		2	403478	4.06	17.406
	3	404243	6.64	11.95		3	403479	4.06	17.406
	4	404244	6.64	11.95		4	403480	4.06	17.406
401239	1	404119	7.35	12.66	400081	1	401668	4.2	17.546
	2	404120	7.35	12.66		2	401669	4.2	17.546
	3	404121	7.35	12.66		3	401670	4.2	17.546
	4	404122	7.35	12.66		4	401671	4.2	17.546
401052	1	401797	0.06	13.406	400770	1	404855	4.3	17.646
	2	401798	0.06	13.406		2	404856	4.3	17.646
	3	401799	0.06	13.406		3	404857	4.3	17.646
	4	401800	0.06	13.406		4	404858	4.3	17.646
401329	1	402796	0.44	13.786	401243	1	404264	5.5	18.846
	2	402797	0.44	13.786		2	404265	5.5	18.846
	3	402798	0.44	13.786		3	404266	5.5	18.846
	4	402799	0.44	13.786		4	404267	5.5	18.846
401195	1	401489	1.12	14.466	401209	1	402104	5.94	19.286
	2	401490	1.12	14.466		2	402105	5.94	19.286
	3	401491	1.12	14.466		3	402106	5.94	19.286
	4	401492	1.12	14.466		4	402107	5.94	19.286
401558	1	405108	1.54	14.886	401260	1	405136	6.2	19.546
	2	405109	1.54	14.886		2	405137	6.2	19.546
	3	405110	1.54	14.886		3	405138	6.2	19.546
	4	405111	1.54	14.886		4	405139	6.2	19.546
400378	1	403018	2.05	15.396	400976	1	405734	6.57	19.916
	2	403019	2.05	15.396		2	405735	6.57	19.916
	3	403020	2.05	15.396		3	405736	6.57	19.916
	4	403021	2.05	15.396		4	405737	6.57	19.916
400445	1	403337	2.62	15.966	400838	1	405174	6.89	20.236
	2	403338	2.62	15.966		2	405175	6.89	20.236
	3	403339	2.62	15.966		3	405176	6.89	20.236
	4	403340	2.62	15.966		4	405177	6.89	20.236
400443	1	403330	2.97	16.316	400430	1	403274	7.29	20.636
	2	403331	2.97	16.316		2	403275	7.29	20.636
	3	403332	2.97	16.316		3	403276	7.29	20.636
	4	403333	2.97	16.316		4	403277	7.29	20.636
401221	1	402905	3.41	16.756	400865	1	405275	7.61	20.956
	2	402906	3.41	16.756		2	405276	7.61	20.956
	3	402907	3.41	16.756		3	405277	7.61	20.956
	4	402908	3.41	16.756		4	405278	7.61	20.956

3.2.2 Incident Data Description

The website maintained by California Traffic Incident Information System (<http://cad.chp.ca.gov/>) provides valuable information regarding time, location, incident type and severity of real-time incidents on the freeway network under surveillance where real-time incident information is updated every 1-minute. The California Incident Information System is also integrated in the California PeMS. Table 3.3 presents a description of the fields from the database maintained by PeMS from which data is obtained for this study:

Table 3.3 Incident Data File Format

Column	Description
District	District number, defined by California PeMS.
Area	County name that an incident happens.
Freeway	Freeway name and its traffic direction.
Start	Incident start time.
Duration (mins)	Time that an incident takes to be cleared.
CA Postmile	CA milepost of an incident on a freeway.
Abs Postmile	Abs milepost of an incident on a freeway.
Location	Description of an incident location, e.g. “EB I80 at I580”.
Description	Description of incident type, e.g. “Disabled vehicle”.

Although the incident information is updated per minute on the website, the incident description is found in a text format and is quite ambiguous. This may degrade the utilization of more detailed incident information. In addition, this description is not integrated in the database of California PeMs. Therefore, excessive manual efforts are required for collecting more incident data such as the number of lanes blocked. In this study, only incident information available from the California PeMS database is applied for the corridor travel time prediction.

3.3 DATA PRE-PROCESSING

Data pre-processing consists of traffic flow data pre-processing and incident pre-processing for extracting valuable information to test the proposed methodology for corridor travel time prediction. Traffic flow data pre-processing consists of data storage, data screening, and data imputation for missing and erroneous data records. While incident data pre-processing consists of extracting valuable information that affect the incident duration.

3.3.1 Traffic Flow Data Pre-Processing

3.3.1.1 Data Storage

The first step of traffic flow data pre-processing involves data storage for further analysis. The raw data comes into the data server as compressed comma-delimited ASCII files which contain individual data for all detectors for each timestamp. Each can be identified by a file name, for example: “5minagg-yyyyymmddhhmmss.txt”. Each file contains traffic flow data at a timestamp for all detectors in the district as defined in the California PeMS. Figure 3.2 shows a sample of the ASCII data file, which begins with timestamp “02/28/2005 06:00:00” and then follows the data fields of VDS_ID, FLOW, OCCUPANCY, SPEED, and so on. These data files are updated on the PeMS website every 30-minute. Given the downloaded ASCII files, traffic flow data on the study corridor are extracted and stored in the MySQL database for this study.

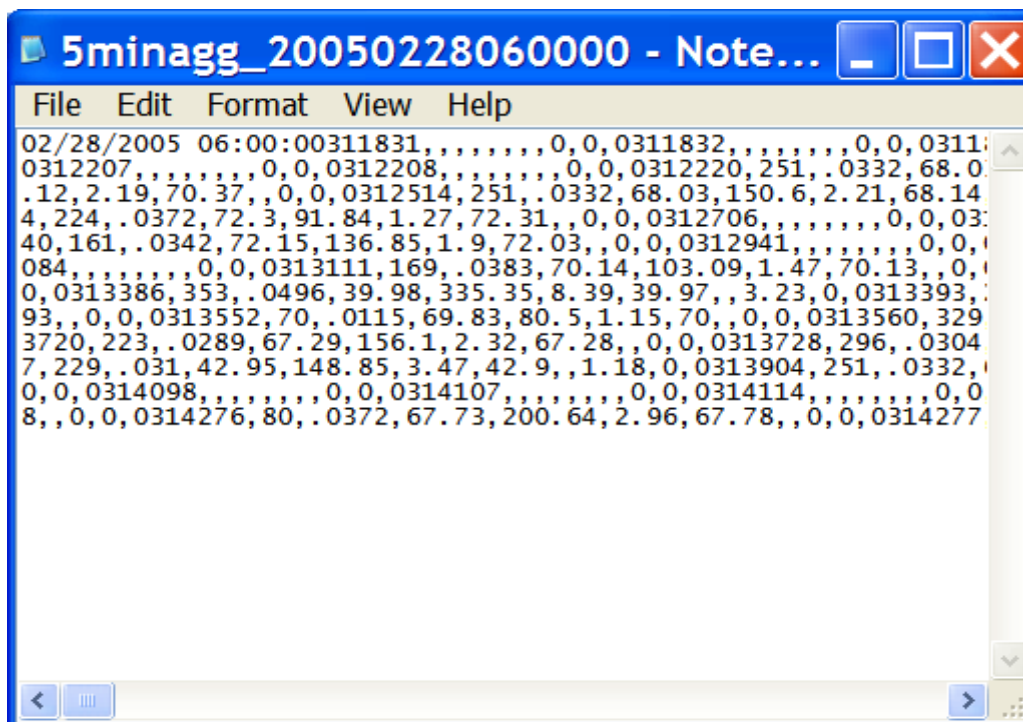


Figure 3.2 Sample of 5-Minute Traffic Flow Data

3.3.1.2 Data Screening

Traffic flow data screening in this study is performed to identify and eliminate suspicious or erroneous data such that the accuracy of corridor travel time prediction that relies on these data can be improved.

In this study, a set of data screening criteria are extracted from those developed by Lomax et al. (2004) for the mobility monitoring program, in which both consistencies between traffic variables, and threshold values for traffic variables are provided. The data screening process consists of 6 tests, as shown in Table 3.4. The first test checks the volume and occupancy values to identify whether data records are missing. The second test checks whether there are duplicate records. Test 3 checks the volumes to identify whether there are continuous identical values. Test 4 and test 5 check the volume and occupancy values against a maximum value threshold. Test 6 checks the consistency between volume and occupancy.

Table 3.4 Data Screening Criteria Used in This Study

Quality Control Test and Description	Sample Code with Threshold Values	Action
Test 1: No vehicles present Indicates that no vehicles passed the detection zone during the detection time period.	If VOLUME=0 and OCC=0	Assign QC flag to VOLUME and OCCUPANCY; set VOLUME and OCCUPANCY to erroneous
Test 2: Duplicate records Caused by errors in data archiving logic or software process.	Detector and date/time stamp are identical.	Remove/delete duplicate records.
Test 3: Consecutive identical volume values Research and statistical probability indicates that consecutive runs of identical data values are suspect. Typically caused by hardware failures.	No more than 8 consecutive identical volume values	Assign QC flag to VOLUME and OCCUPANCY; set VOLUME and OCCUPANCY to erroneous
Test 4: Maximum volume Traffic flow theory suggests a maximum traffic capacity.	If VOLUME > 250 (5 min)	Assign QC flag to VOLUME, set VOLUME to erroneous
Test 5: Maximum occupancy Empirical evidence suggests that all data values at high occupancy levels are suspect. Caused by detectors that may be “stuck on.”	If OCC > 80% (5 min.)	Assign QC flag to VOLUME and OCCUPANCY; set VOLUME and OCCUPANCY to erroneous
Test 6: Multi-variate consistency Zero values when occupancy is non-zero. Unknown cause.	If VOLUME = 0 and OCC > 0	Assign QC flag to VOLUME and OCCUPANCY; set VOLUME and OCCUPANCY to erroneous

Sources: Monitoring urban roadways in 2002: using archived operations data for reliability and mobility measurement (Lomax et al. 2004)

3.3.1.3 Data Imputation

The final step of data pre-processing involves imputation and replacement of missing and erroneous or suspicious traffic flow data, which is also recommended by ITS as a data resource: Preliminary Requirements for a User Service (FHWA 1998).

A review of existing imputation methods indicate that there are a number of imputation methods that can be applied for continuous data. Considering that the missing volumes and occupancy are randomly distributed in the whole data set, the methods of

average of surrounding time periods (SurT), average of surrounding detectors (SurD), and historic average (Hist) are applied. The first two methods are developed based on the method of temporal and spatial nearest neighbors, respectively. Comparatively, the method of historical average is developed based on the classification method where missing volumes and occupancies are imputed from the historic averages for a given time-of-day and day-of-week. The priority sequence for implementing these imputation methods is determined as SurT, SurD, and Hist based on the research by Chen and Xia (2006). More details about these methods are described as follows.

Average of Surrounding Time Periods (SurT)

In this method, the missing value is imputed using the average of its preceding and succeeding values. For example, if the 45th time period flow rate of a Tuesday in February is missing, then that flow rate is imputed by the average of 44th time period and 46th time period flow rates of the same Tuesday of that week in February.

Average of Surrounding Detectors (SurD)

The method of average of surrounding detectors imputes missing flow rates and occupancies by averaging the flow rates or occupancies collected at upstream and downstream stations during the same period. For example, if 45th time period flow rate of a Tuesday in February is missing, then the missing flow rate is imputed and replaced by the average of the flow rates corresponding to the 45th time period of the same Tuesday of February of immediate upstream and downstream detectors.

Historical Average (HIST)

Historical average mainly uses the available historic data related to the missing values. It assumes that the flow rate tend to be stable over time. In this study, the time of a day and the day of a week of the correct data are considered, where the missing flow rate or occupancy was imputed and replaced by averaging flow rates or occupancies with the same time-of-day and day-of-week.

3.3.2 Incident Data Pre-Processing

Incident data pre-processing involves defining the categories for categorical variables for further analysis. In this study, incident type, time of day, and day of week of an incident are considered as the potential factors that affected its duration, which is required for further

corridor travel time adjustment under an incident. The categories of these variables are defined as shown in Table 3.5. It is noted that there are more incident types found in the incident database maintained by California PeMS, but few incident records on the study corridor during the data collection period for this study is found to be related to them. These incident types contained debris, loose animal, pedestrian, and traffic control. These either had little impact on traffic or had known duration. Therefore, incident types of debris, loose animal, pedestrian, and traffic control are not defined in the categories of incident type for this study.

Table 3.5 Definitions of Categories of the Potential Variables

Variable	Category	Description
Time of Day	1	6:00~9:00 and 16:00~19:00
	2	Others
Day of Week	1	Weekday
	0	Weekend
Incident Type	1	Disabled vehicle
	2	Traffic hazard of vehicle
	3	Hit and run
	4	Collision with ambulance responding
	5	Collision with no details
	6	Collision with property damage
	7	Vehicle fire

CHAPTER 4

DYNAMIC TRAFFIC FLOW PREDICTION

4.1 OVERVIEW

The accurate estimation of travel time is valuable for a variety of transportation applications such as freeway performance evaluation and real-time traveler information. Given the extensive availability of traffic data collected by intelligent transportation systems, a variety of travel time estimation methods have been developed. Despite limited success under light traffic conditions, traditional corridor travel time prediction methods have suffered various drawbacks. For example, most of these methods were developed based on data generated by dual-loop detectors, which contain average spot speeds. However, single-loop detectors (and other devices that emulate its operation) are the most commonly used devices in traffic monitoring systems. There has not been a reliable methodology for travel time prediction based on data generated by such devices due to the lack of speed measurements. The majority of existing studies focus on travel time estimation instead of prediction. Additionally, the effect of traffic progression along the freeway has not been considered in the travel time prediction process. Moreover, the impact of incidents on travel time estimates has not been effectively accounted for in existing studies. Much of previous work performed to date took no consideration of the incident impacts for the estimation of travel time.

The objective of this dissertation is to develop a methodology for dynamic corridor travel time prediction based on continuous data generated by single-loop detectors (and similar devices) and incident reports generated by the traffic monitoring system. This method involves multiple-step-ahead prediction for flow rate and occupancy in real time. A seasonal autoregressive integrated moving average (SARIMA) model is developed with an embedded adaptive predictor. This predictor adjusts the prediction error based on traffic data that becomes available every five minutes at each station. Based on this, a corridor travel time prediction model is developed integrating the embedded predictor for flow rate and occupancy. In order to incorporate the incident impacts on traffic, a corridor travel time adjustment model is also developed by conducting queuing analysis based on the prediction of incident duration.

The components of the proposed methodology are illustrated in Figure 4.1 to achieve the final goal of predicting corridor travel time.

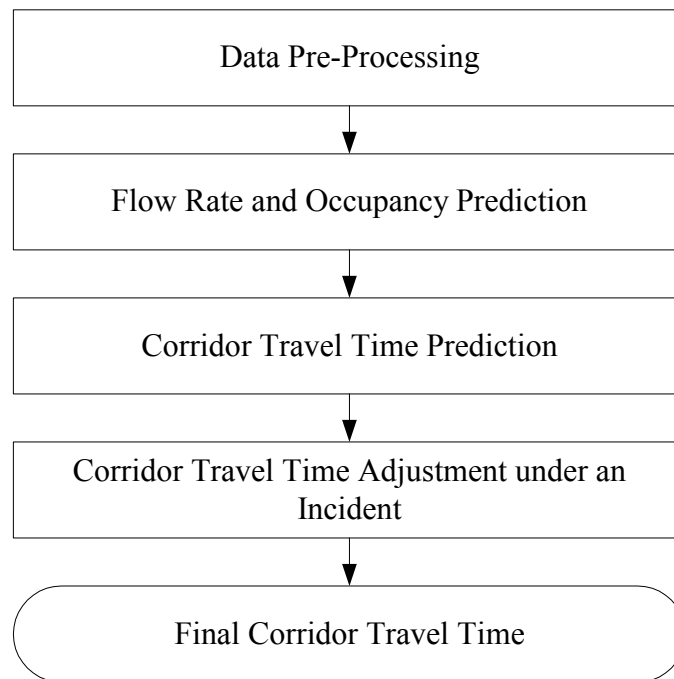


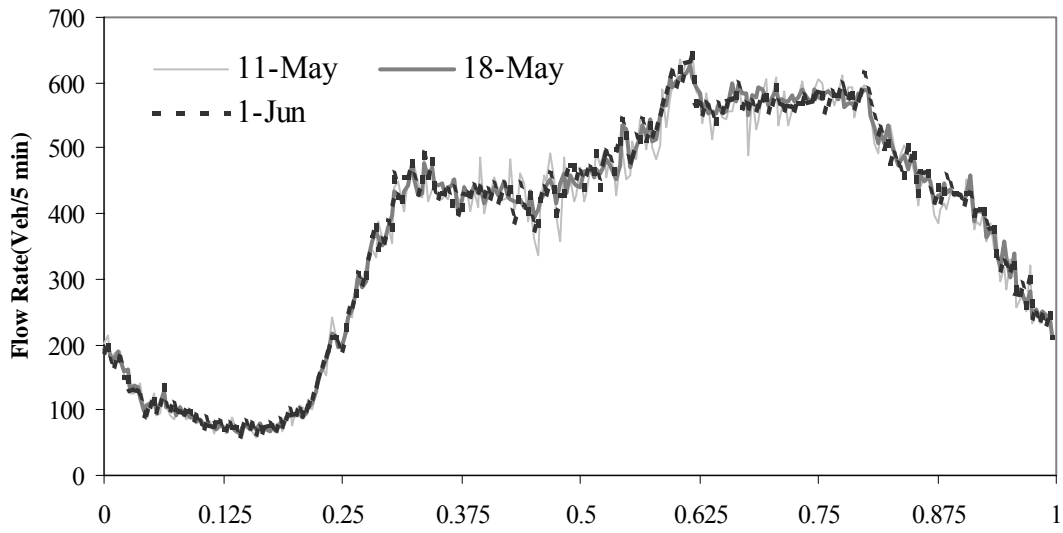
Figure 4.1 Components of the Proposed Methodology for Corridor Travel Time Prediction

The problem addressed in this chapter is that of dynamic prediction of flow rate and occupancy since it is of great importance for further corridor travel time prediction and adjustment under an incident. The rest of this chapter is organized as follows. Section 4.2 presents the adoption of SARIMA model for flow rate and occupancy prediction. Section 4.3 and section 4.4 present the seasonal operator and short-term operator of the SARIMA model, respectively. Section 4.5 presents the model adequacy analysis. Section 4.6 presents the traffic flow prediction, in which both one-step and multi-step-ahead prediction of flow rate and occupancy is performed. Finally, a summary of traffic flow prediction is presented in section 4.7.

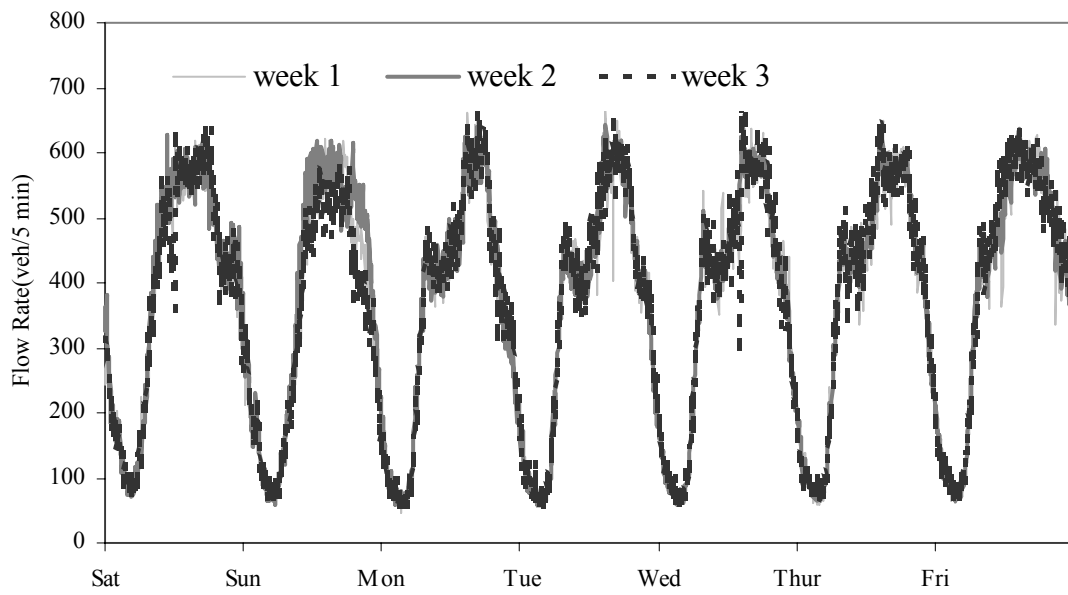
4.2 ADOPTION OF THE SARIMA MODEL

The problem of multi-step-ahead prediction of flow rate and occupancy is defined as: given 5-minute flow rate and occupancy series $X_t = [f_t, o_t]'$ up to time step t , to predict X_{t+m} , where m is the number of steps in advance.

There are various ways to predict future traffic flow values. Using 15-minute flow rate series, Willams (1999) and Guo (2005) provided a thorough analysis of the SARIMA model based on the Box-Jenkins approach, and a weekly model SARIMA(1,0,1)(0,1,1)₆₇₂, was identified to be the most suitable for flow rate prediction. However, both of them predicted flow rate values with only one-step in advance, while multiple-step-ahead prediction was not proposed. In this study, a weekly SARIMA model is adopted using 5-minute flow rate and occupancy series since data for both show weekly seasonality and local variation. As illustrated in Figure 4.2 when using the flow rate data collected at VDS 401079 between May 6, 2006 and June 2, 2006, we can find daily similarity and instant dynamics in both daily and weekly patterns.



Daily Pattern



Weekly Pattern

Figure 4.2 Weekly and Daily Variation in Flow Rate Series

Due to the fact that 2016 records of 5-minute traffic flow data were collected for one week by a field vehicle detector, this study proposes a SARIMA(1,0,1)(0,1,1)₂₀₁₆ model for flow rate and occupancy prediction

4.2.1 SARIMA Model Introduction

To introduce a SARIMA model, the key concept of stationarity in time series is necessary to be introduced firstly since it is the basic requirement in time series models.

There are two kinds of stationarity in time series defined as X_t : strictly stationary and weakly stationary. A process X_t is strictly stationary if its statistical properties do not change over time, i.e. the probability distributions of the process are time-invariant. While, a process X_t is weakly stationary if it satisfies conditions: (1) the expectation of the value of X_t is a constant μ for all t ; (2) the variance of the process given by $\sigma^2(t) = Var(X_t) = E(X_t - \mu(t))^2$ is a constant for all t ; (3) the covariance of the process given by $\gamma(t_1, t_2) = Cov(X_{t_1}, X_{t_2}) = E(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))$ is a function of $(t_1 - t_2)$ only. In practice, a time series X_t is very difficult to meet the conditions of strict stationarity. A stationary process herein always refers to a weakly stationary process.

For stationary time series, the fundamental assumption of time series modeling is that the value of the series at time t depends on the its p previous values and on a random disturbance as defined as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \ddot{Z}_t \quad (4.1)$$

where $\{\phi_1, \phi_2, \dots, \phi_p\}$ are real constants, \ddot{Z}_t is the disturbance at time step t , usually modeled as a linear combination of a zero-mean white noise process as

$$\ddot{Z}_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} \dots + \theta_q Z_{t-q} \quad (4.2)$$

where q is the number of past disturbance values, and $\{Z_t\}$ is a white noise process with zero mean and variance σ^2 . Constants $\{\phi_1, \phi_2, \dots, \phi_p\}$ and $\{\theta_1, \theta_2, \dots, \theta_q\}$ are called autoregressive (AR) and moving average (MA) coefficients, respectively. Combining equation (4.1) and (4.2), time series X_t can be modeled as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q} \quad (4.3)$$

This defines a zero-mean autoregressive moving average (ARMA) process of orders p and q or a $ARMA(p, q)$. By introducing the backshift operator B , which is defined by $B^j X_t = X_{t-j}$, equation (4.3) can be rewritten as

$$X_t - \phi_1 B X_t - \phi_2 B^2 X_t - \dots - \phi_p B^p X_t = Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_t + \dots + \theta_q B^q Z_t \quad (4.4)$$

An autoregressive polynomial $\phi(B)$ can be defined as $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ and a moving average polynomial $\theta(B)$ can be defined as $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$, then equation (4.4) is written in the form of

$$\phi(B) X_t = \theta(B) Z_t \quad (4.5)$$

When $q = 0$, only the AR part of the equation remains. Equation (4.5) would reduce to a pure autoregressive process of order p , denoted by $AR(p)$. Similarly, if $p = 0$, a pure moving average process of order q denoted by $MA(q)$ can be obtained.

If the series X_t is not stationary, an ARMA model cannot be used directly. However, if the differenced series $\{M_t : M_t = (1 - B)^d X_t\}$ is stationary, then we can obtain

$$(1 - B)^d \phi(B) X_t = \theta(B) Z_t \quad (4.6)$$

where d is the order of differencing on X_t . Using the definition of the backshift operator B , one order of differencing time series is defined as $(1 - B)X_t = X_t - X_{t-1}$. Similarly, two orders of differencing of time series is defined as

$$(1 - B)^2 X_t = (1 - B)(X_t - X_{t-1}) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}.$$

Equation (4.6) defines an ARIMA process of orders p , d , q , or simply $ARIMA(p, d, q)$. When $d = 0$, it is an ARMA model. Therefore, an ARMA model is a special case of the ARIMA model.

A SARIMA model is further expanded from an ARIMA model when seasonal or cyclic components exist in a time series X_t . A seasonal autoregressive integrated moving average process of a time series X_t with regular and seasonal AR orders p and P , and regular and seasonal MA orders q and Q , is defined as

$$(1-B)^d(1-B^s)^D\phi(B)\Phi(B^s)X_t = \theta(B)\Theta(B^s)Z_t \quad (4.7)$$

where,

B : the backward shift operator;

d and D : the regular and seasonal differencing orders;

ϕ and Φ : the regular and seasonal AR parameters;

θ and Θ : the regular and seasonal MA parameters;

s : the seasonal process period that a season cycle covers.

This defines a SARIMA(p, d, q)(P, D, Q) $_s$ model. When $s=0$, i.e. there are no seasonal trends existing in the process of series X_t , a SARIMA model becomes an ARIMA model. Therefore, both ARMA models and ARIMA models are special cases of SARIMA models.

4.2.2 Data Transformation

After the process of seasonal and regular differencing, traffic flow data series, including flow rate and occupancy, is required to be stationary to be fitted in an ARMA model. Unfortunately, real-world data does not often satisfy this condition because of the nonlinear relationship or heteroscedasticity in flow rate and occupancy series. To make the traffic flow series X_t “well-behaved”, data transformation is incorporated in flow rate and occupancy prediction such that the transformed traffic flow series Y_t can be modeled by a zero-mean stationary ARMA type of process, after performing the process of seasonal and regular differencing.

In this study, Box-Cox transformation (Box and Cox 1964), a general class of power transformations, is applied. Given the traffic flow series X_t , the Box-Cox transformation is defined as

$$Y_t = \begin{cases} \frac{(X_t)^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \ln(X_t) & \text{for } \lambda = 0 \end{cases} \quad (4.8)$$

where,

X_t : the traffic flow series;

Y_t : the transformed traffic flow series;

λ : the transformation parameter;

The parameter λ controls the shape of transformation. If $\lambda = 0$, it produces a logarithmic transformation, while if $\lambda = 0.5$, it results in a square root transformation.

The value of λ is determined by fitting the SARIMA(1,0,1)(0,1,1)₂₀₁₆ model typically in a range from -3 to 3 (MathWorks 1994). The λ value producing the maximum likelihood is reported as the optimal Box-Cox transformation. Given the observations X_1, X_2, \dots, X_n , the log-likelihood of original data series X_t is constructed from the log-likelihood for the transformed traffic flow series Y_t as:

$$l_x(\cdot) = l_y(\cdot) + (\lambda - 1) \left(\sum_{i=1}^n \ln(X_i) \right) \quad (4.9)$$

where $l_x(\cdot)$ is the log-likelihood of the original traffic flow series, $l_y(\cdot)$ is the log-likelihood of the transformed traffic flow series.

Although the log-likelihood is a continuous function of λ and the local optimal value for λ can be obtained by solving the differential equation of λ , simple method by specifying a range of λ and trying different λ values with a small increment difference is often performed. This estimation method is used, because the practical difference between 0.5 and 0.55, for example, is likely to be very small (MathWorks 1994). This is also verified by this study, in which the range of the λ values is specified from -2 to 2 with an increment change of 0.05, and for each λ value, the log likelihood for the original series are collected.

At different vehicle detector stations, flow rate and occupancy series produce a different value of λ . The value of λ merely depends upon the flow rate or occupancy series itself at the specific vehicle detector station.

4.2.3 Model Decomposition

Following the notations of a SARIMA model, a SARIMA(1,0,1)(0,1,1)₂₀₁₆ model for the process of transformed traffic flow series Y_t is defined as

$$(1 - B^{2016})\phi(B)Y_t = \theta(B)\Theta(B^{2016})Z_t \quad (4.10)$$

where,

B : the backward shift operator;

2016: the seasonal process period;

ϕ : the short-term AR parameters;

θ : the short-tem MA parameters;

Θ : the seasonal MA parameters;

Z_t : the random error or noise.

The proposed SARIMA(1,0,1)(0,1,1)₂₀₁₆ model can be interpreted as a cascade of a seasonal operator and a short-term operator. The seasonal operator separated from the SARIMA(1,0,1)(0,1,1)₂₀₁₆ model is used to extract the seasonal trend in the flow rate and occupancy series. It is characterized by nonzero correlations only at lag 2016 in the form of

$$(1 - B^{2016})Y_t = (1 + \Theta B^{2016})W_t \quad (4.11)$$

where,

W_t : the random error series from the seasonal operator;

Θ : the seasonal MA parameters.

After subtracting the seasonal trend in traffic flow series, the short-term operator is interpreted by an ARMA(1,1) model to the seasonally adjusted series W_t . The short-term operator of the proposed SARIMA model is used to capture the local variation in traffic flow series in the form of

$$(1 - \phi B)W_t = (1 + \theta B)\varepsilon_t \quad (4.12)$$

where,

W_t : the seasonally adjusted series;

ϕ : the short-term AR parameter;

θ : the short-tem MA parameter;

ε_t : the random error or noise series output from the short-term operator.

4.3 SEASONAL OPERATOR

Following the notations of a SARIMA model, equation (4.11) can be rewritten as

$$Y_t = B^{2016}Y_t + \Theta B^{2016}W_t + W_t \quad (4.13)$$

Using the definition of backward shift operator B , it is obtained that $B^{2016}Y_t = Y_{t-2016}$ and $B^{2016}W_t = W_{t-2016}$. Thus, equation (4.13) can be written as

$$Y_t = Y_{t-2016} + \Theta W_{t-2016} + W_t \quad (4.14)$$

Since W_t is the random error of the predicted values at time step t , the term $Y_{t-2016} + \Theta W_{t-2016}$ is actually the predicted value of Y_t due to the process of seasonal operator. Let $S_t = Y_{t-2016} + \Theta W_{t-2016}$, we can obtain $W_t = Y_t - S_t$. Thus, equation (4.14) can be written in an iterative format as

$$S_t = Y_{t-2016} + \Theta(S_{t-1} - Y_{t-2016}) \quad (4.15)$$

Equation (4.15) is a standard form of simple exponential smoothing, indicating that we can fit a simple exponential smoothing model for the process of the seasonal operator of the proposed SARIMA(1,0,1)(0,1,1)₂₀₁₆ model.

Let $\alpha = 1 - \Theta$, equation (4.15) is written as

$$S_t = \alpha Y_{t-2016} + (1 - \alpha)S_{t-2016} \quad (4.16)$$

This is the final equation for the seasonal operator of the flow rate and occupancy prediction model. The parameter α is called the smoothing constant with a value between 0 and 1. If $\alpha = 1$, equation (4.16) describes a random walk model, where the predicted value S_t simply equals the observed value Y_{t-2016} . Comparatively, if $\alpha = 0$, the predicted value S_t remains the predicted value at time step $t - 2016$, i.e., S_{t-2016} . In this study, the best value for α is selected as 0.15 for three reasons. Firstly, due to the findings in Williams (1999) and Williams and Hoel (2003), α is fairly stable at 0.15 across locations. Secondly, the sensitivity analysis conducted by Guo (2005) concludes that different α values have little impact on traffic flow prediction when using SARIMA model. Additionally, using the limited traffic flow data collected on the study corridor, this study also presents a similar conclusion where α is selected as value which results in the smallest mean square error

(MSE). Given the transformed traffic flow series Y_t with observations Y_1, Y_2, \dots, Y_n , the

mean square error is calculated as
$$\frac{\sum_{i=1}^n (Y_i - S_i)^2}{n}$$

4.4 SHORT-TERM OPERATOR

Following the notation of an ARMA model, the short-term operator of the proposed SARIMA model as described in equation (4.12) can be written in the form of

$$W_t = \phi W_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \quad (4.17)$$

In this equation, the parameters ϕ and θ are required to be estimated from time to time to predict the value of W_t . In order to estimate these two time-varying parameters, adaptive algorithms such as the Kalman filter, least mean square (LMS), and recursive least square (RLS) can be adopted. In mean square sense, the adaptive Kalman filter is optimal and it is applied in this study since both LMS and RLS are the special cases of a Kalman filter method (Travainen et al. 2003).

4.4.1 Kalman Filter Design

The Kalman filter was first introduced in the early of 1960's (Kalman 1960). Since that time, the Kalman filter has been the subject of extensive research and application, particularly in the area of autonomous or assisted navigation (Welch and Bishop 1995). The process of Kalman filtering involves two steps. First, an estimate of the present state of the system is refined based on some observations of the system. Second, the refined estimate of the present state is extrapolated to the next observation by the use of the evolution operator (DeSila 2006). These two steps are repeated when new observations are available.

The use of Kalman filtering techniques requires deriving a stochastic state-space representation of the robot model and of the measurement process. For our case, the measurement equation is derived from the time-varying ARMA(1,1) model to series W_t in the form of

$$W_t = \phi_t W_{t-1} + \theta_t \varepsilon_{t-1} + \varepsilon_t \quad (4.18)$$

where ϕ_t and θ_t are the time-varying AR and MA parameters, and ε_t is the random noise.

Let $\delta_t = [\phi_t, \theta_t]'$ and $\varphi_t = [W_{t-1}, \varepsilon_{t-1}]'$, equation (4.18) can be written as

$$W_t = \varphi_t' \delta_t + \varepsilon_t \quad (4.19)$$

This equation forms a linear observation model with φ_t' being the regression vector, ε_t being the observation error, and δ_t being a state vector. When there is no prior information available, the state vector is typically described as a random walk model in the form of

$$\delta_{t+1} = \delta_t + \omega_t \quad (4.20)$$

where ω_t is the state noise.

Equations (4.19) and (4.20) form a structure of the general state-space equations for the adaptive Kalman filter. ε_t and ω_t denote the measurement and state noises with zero means, with covariance C_{ε_t} and C_{ω_t} , respectively.

4.4.2 Adaptive Kalman Filter Implementation

In order to start the process of an adaptive Kalman filter, initialization of δ_t , φ_t , C_{ε_t} , C_{ω_t} , and the state error covariance C_{δ_t} are required. In this study, all the initializations are achieved by fitting the one-day seasonally adjusted series W_t in an ARMA(1,1) model, in which both AR and MA parameters ϕ and θ are treated as constants, using the maximum likelihood method. Given an ARMA(1,1) model as described in equation (4.12), the log likelihood function of the model can be written as

$$-\frac{1}{2\sigma^2} (W - 1u)' \Sigma^{-1} (W - 1u) - \frac{1}{2} \ln(|\Sigma|) - \frac{n}{2} \ln(\sigma^2) \quad (4.21)$$

where, n is the number of observations, $\sigma^2 \Sigma$ is the variance of as a function of the ϕ and θ parameters, $|\bullet|$ denotes the determinant, u is the mean, and the vector W is the time series W_t written as a column vector.

By treating the maximum likelihood estimate (MLE) of σ^2 as $s^2 = \frac{1}{n}(W - 1u)' \Sigma^{-1}(W - 1u)$, the log likelihood concentrated with respect to σ^2 can be written as

$$-\frac{n}{2} \ln(W - 1u)' \Sigma^{-1}(W - 1u) - \frac{1}{2} \ln(|\Sigma|) \quad (4.22)$$

Let H be the lower triangular matrix with positive elements on the diagonal such that $H'H = \Sigma$, and e be the vector $H^{-1}(W - 1u)$. Equation (4.22) can be written as

$$-\frac{n}{2} \ln(e'e) - \frac{1}{2} \ln(|H|) = -\frac{n}{2} \ln(|H|^{1/n} e'e |H|^{1/n}) \quad (4.23)$$

The MLE is produced by using a Marquardt algorithm (Moré 1978) to minimize the sum of squares $|H|^{1/n} e'e |H|^{1/n}$ (SAS 2002).

After the ARMA(1,1) model is fitted, initializations of δ_t ; C_{δ_t} ; φ_t ; C_{ε_t} are performed. δ_t is initialized as the estimates of AR and MA parameters. C_{δ_t} is initialized as the AR and MA parameter error covariance, which is also provided in the fitness of the ARMA(1,1) model. φ_t is initialized as the vector of predicted W_t and the error ε_t . C_{ε_t} is initialized as the covariance matrix of the predicted error ε_t . C_{ω_t} is initialized as 0.95 (Karjalainen 1996).

The state-space equations can then be readily solved using the well-known time-update recursions given in Figure 4.3 (Myers and Tapley (1976)).

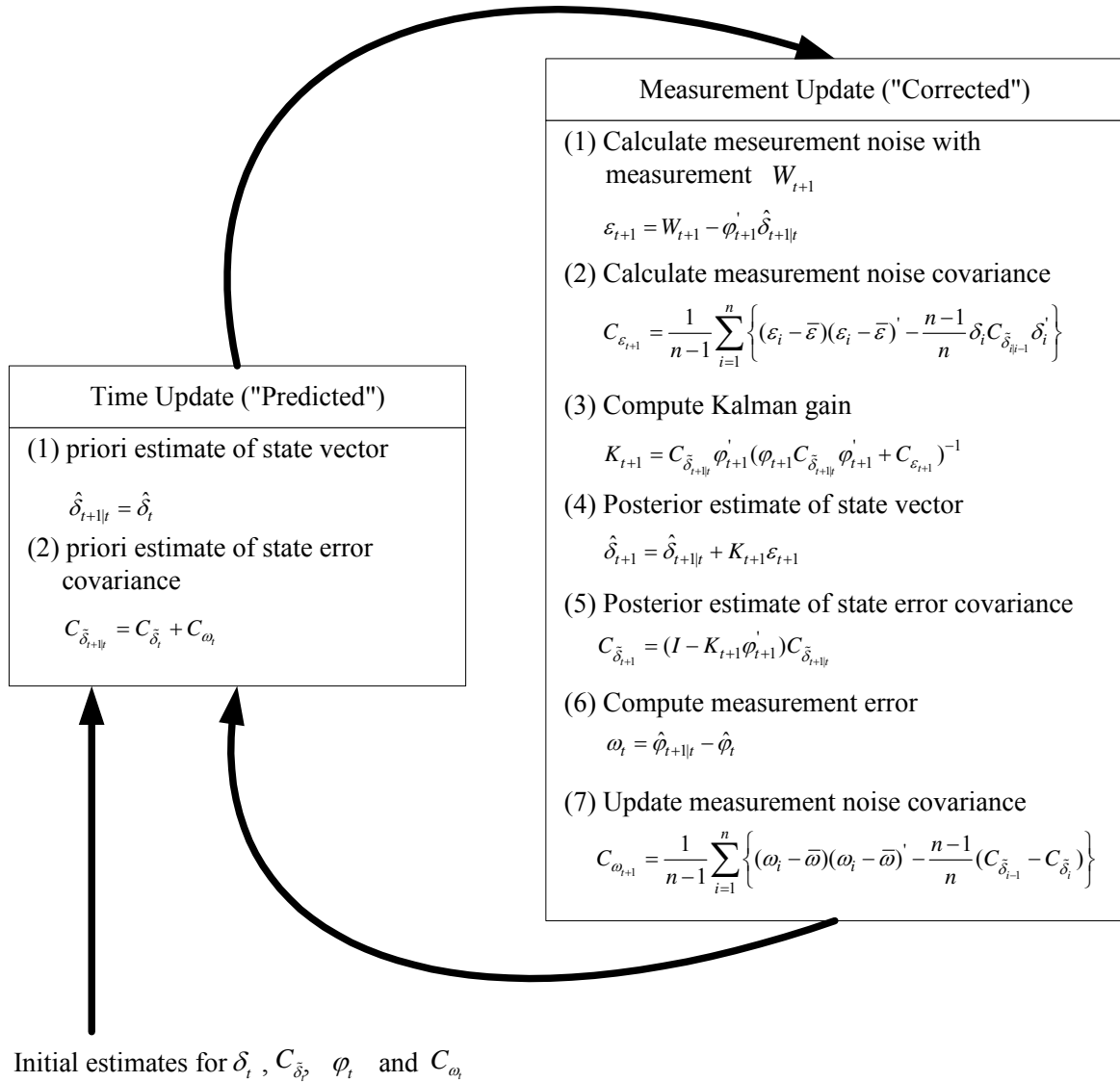


Figure 4.3 Adaptive Kalman Filter Operation

In this diagram, n is a prescribed parameter indicating the memory size. The memory size of the estimator is given by $n = \frac{2}{1-\kappa}$, where κ is the forgetting factor with typical values between 0.90 and 1.0 (Tarvainen et al. 2003). If $\lambda=1.0$, n will be infinitely large, and both measurement error covariance and state noise covariance will be

calculated based on all samples up to time step t . In this work, a 0.95 value of κ is applied, and the number of memory size is calculated as $\frac{2}{1-0.95} = 40$.

4.5 PROPOSED SARIMA MODEL ADEQUACY ANALYSIS

Model adequacy analysis is the most important step in the model building sequence. It is also one of the most overlooked. To check whether the proposed SARIMA(1,0,1)(0,1,1)₂₀₁₆ model is adequate for 5-minute flow rate and occupancy prediction, model adequacy analysis is performed using 19 full days traffic flow series collected on the study corridor.

The model adequacy analysis for the proposed SARIMA model consists of two steps. First, the seasonally adjusted series W_t is inspected through the autocorrelation function (ACF) plots to verify that the series are stationary as inputs for an ARMA(1,1) model. Secondly, the residual series ε_t from the short-term operator is checked to verify that they are white noise, indicating that there is no need to use a more complicated model on the samples.

4.5.1 Inspection of Seasonally Adjusted Series W_t

The purpose of the inspection of residual series W_t is to check whether the series is stationary as required by an ARMA(1,1) model through the check of sample autocorrelation function (ACF) plots. The plots are called autocorrelation functions because they show the degree of correlation with past values of the series as a function of the number of lags in the past at which the correlation is computed.

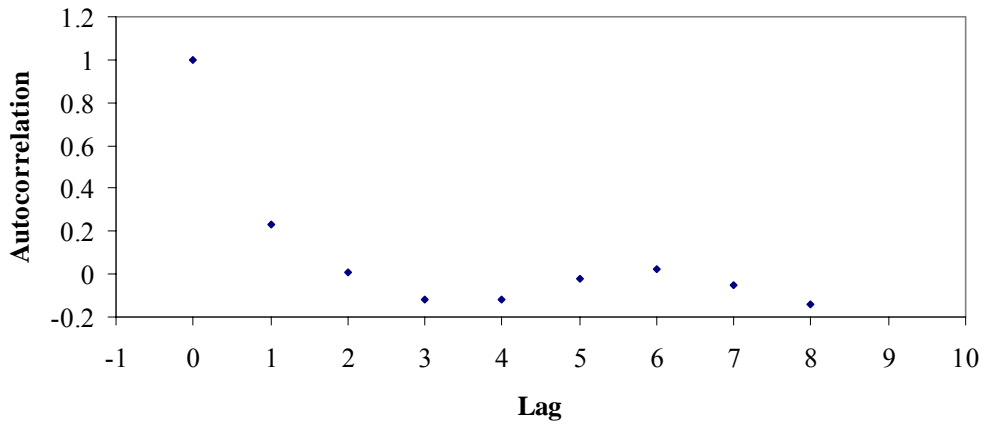
The autocorrelation function plots are formed by a vertical axis: autocorrelation coefficient and a horizontal axis: time lag L ($L=1, 2, 3, \dots$). The autocorrelation coefficient is calculated by $\rho_L = C_L / C_0$, where C_L is the autocovariance function and C_0 is the variance function given by following equations, respectively:

$$C_L = \frac{1}{n} \sum_{i=1}^{n-L} (W_i - \bar{W})(W_{i+L} - \bar{W}) \quad (4.24)$$

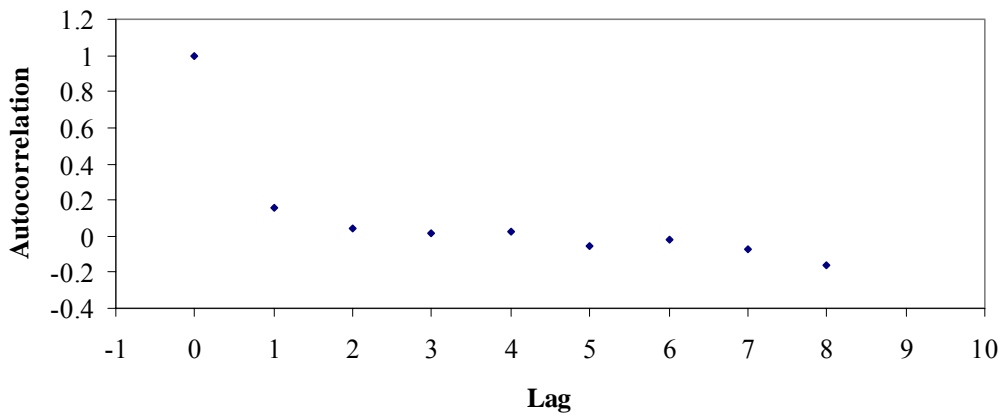
$$C_0 = \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 \quad (4.25)$$

where n is the number samples and \bar{W} is the mean of samples.

Figure 4.4 shows the sample autocorrelation function plots using the seasonally adjusted series W_t at VDS 401079 on May 11, 2006. This figure shows how values of the series are correlated with past values of the series at different lags. For example, the autocorrelation value 0.23213 for lag 1 of the flow rate plots means that the correlation between flow rate and the past value at lag 0 is 0.23213. By examining these plots, one can judge whether the series is stationary or not. If the ACF decays slowly from lag 0 to other lags, the series is nonstationary. Otherwise, it is stationary. In this case, inspection of the autocorrelation function plots indicates that both flow rate and occupancy series after removing the seasonal trends are stationary since the ACFs decrease rapidly from lag 0 (correlation = 1.0) to other lags.



Flow Rate



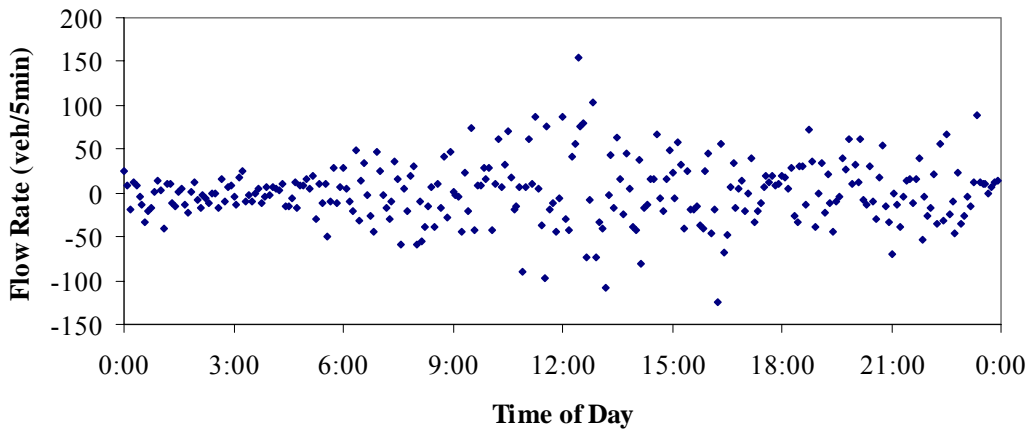
Occupancy

Figure 4.4 Autocorrelation Function Plots of the Seasonally Adjusted Series W_t at VDS 401079 on May 11, 2006

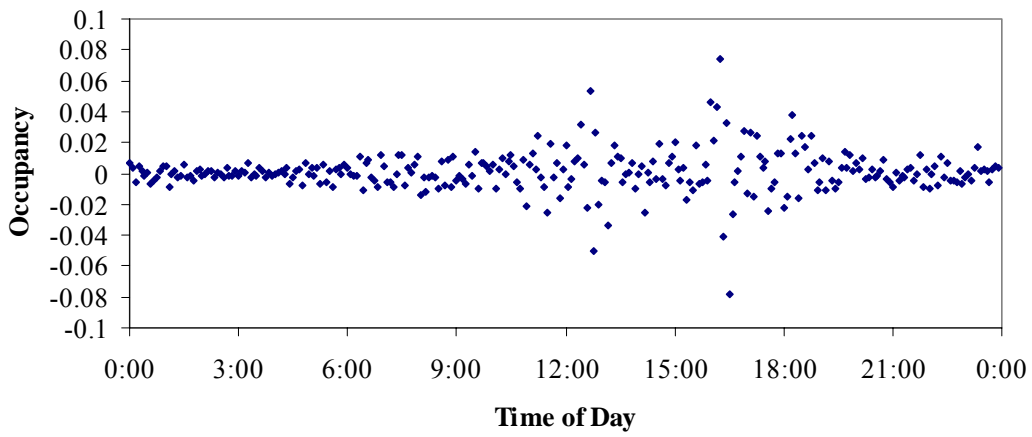
4.5.2 Inspection of Residual Series ε_t from the Short-Term Operator

The purpose of the inspection of residual series ε_t from the short-term operator is to statistically identify whether the series ε_t are uncorrelated (white noise) or contain additional information that might be utilized by a more complex model.

Figure 4.5 shows both flow rate and occupancy residual series output from the short-term operator using traffic flow data collected at VDS 401079 on May 11, 2006 on the study corridor. Both residual series of flow rate and occupancy seem to be random (i.e. white noise) distributed around the value of zero, indicating that an ARMA(1,1) model for capturing the local variation in the flow rate and occupancy series might be appropriate.



Flow Rate



Occupancy

Figure 4.5 Seasonally Adjusted Series of Flow Rate and Occupancy at VDS 401079 on May 11, 2006

To test whether the residuals series from the ARMA(1,1) model are uncorrelated, a value $\pm 2/\sqrt{n}$ is used as rough guide to see if the correlations are significant from zero (Wolfram Research Inc. 2006). Figure 4.6 show the autocorrelation plots of the residuals series from the ARMA(1,1) model using both flow rate and occupancy data collected at VDS 401079 May 11, 2006. Since a full day of traffic flow data consists of a total number of 288 5-minute traffic flow series, the bound values are calculated as $\pm 2/\sqrt{n} = \pm 2/\sqrt{288} = \pm 0.1179$. The figures show that all sample autocorrelations, except those for flow rate at lag 7 and those for occupancy at lag 9 and lag 46, fall inside the bound values. This indicates that both flow rate and occupancy residuals from the ARMA(1,1) model appear to be random, and there is no need to utilize a more complex model to capture the local variation in the traffic flow series.

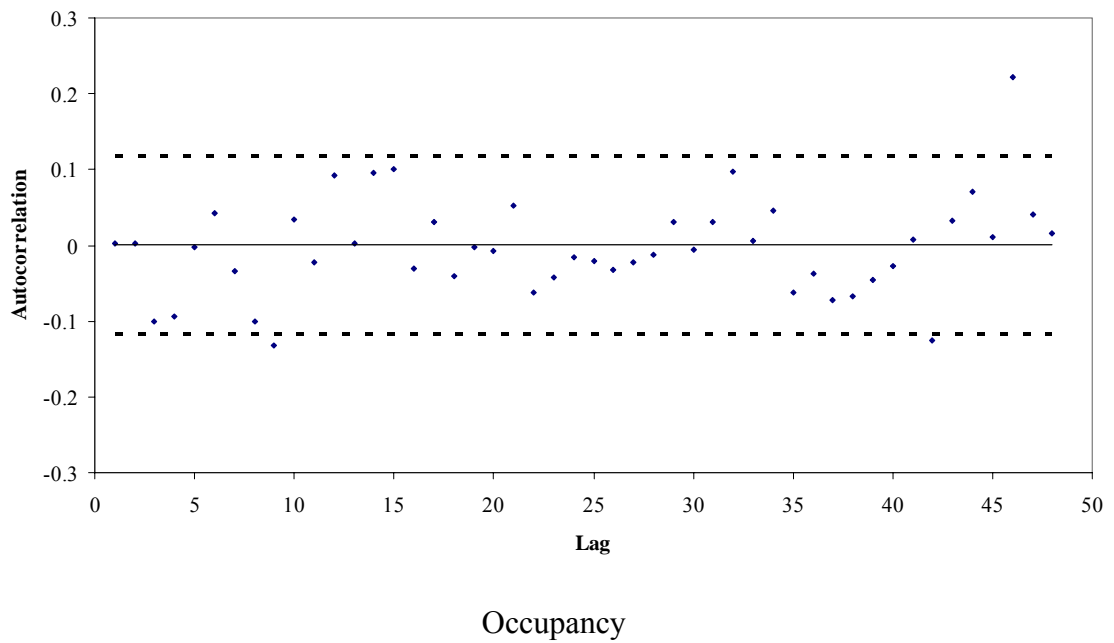
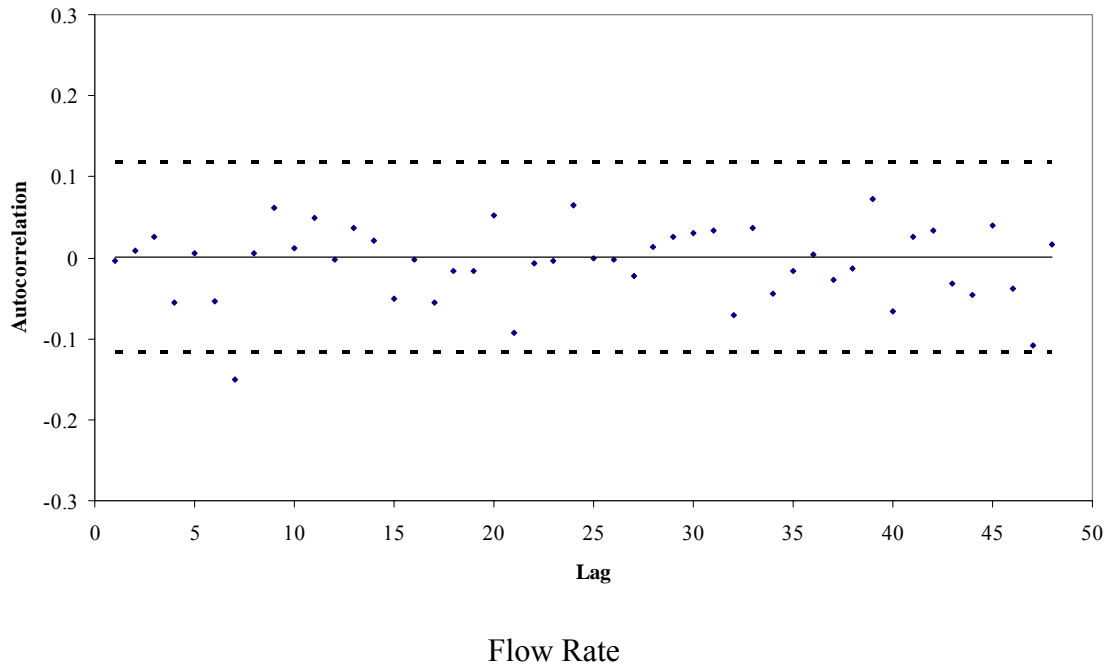


Figure 4.6 Autocorrelation Plots of the flow rate and Occupancy Residuals from ARMA(1,1) Model at VDS 401079 on May 11, 2006

Alternatively, the first ξ correlation values together can be checked and tested to see if the first ξ correlations are significant from zero (null hypothesis) based on the statistics $Y_\xi = j(j+2) \sum_{i=1}^{\xi} \rho_i^2 / (j-i)$, which has an asymptotic χ^2 distribution with $\xi - p - q$ degrees of freedom (Wolfram Research Inc. 2006). At a significance level of $\Gamma = 0.05$, if $Y_\xi > \chi_{1-\Gamma}^2(\xi - p - q)$, the null hypothesis is rejected, indicating that an ARMA(1,1) model is not adequate for a short-term operator of the proposed model.

Table 4.1 and Table 4.2 present the χ^2 test results of the autocorrelation of the residual series from the ARMA(1,1) model in this study using the traffic flow data collected at VDS 401079. The test statistics accept the no-autocorrelation hypothesis of the residual series (all p-values are greater than 0.05). This means that the residual series from the ARMA(1,1) are random (white noise) at a significance level 0.05. Hence, the ARMA(1,1) model is an adequate model for the short-term operator of the proposed SARIMA model for flow rate and occupancy prediction.

Table 4.1 Autocorrelation Check of Flow Rate Residuals from ARMA(1,1) Model at VDS 401079 on May 11, 2006

To Lag	Chi-Square	DF	P-Value	Autocorrelation					
6	1.12	4	0.8914	0.000	-0.004	0.009	0.026	-0.055	0.006
12	10.69	10	0.3820	-0.053	-0.151	0.005	0.062	0.012	-0.055
18	12.93	16	0.6775	-0.003	0.036	0.021	-0.050	-0.002	-0.055
24	16.67	22	0.7815	-0.017	-0.017	0.052	-0.093	-0.007	-0.004
30	18.42	28	0.9151	0.065	-0.001	-0.002	-0.022	0.014	0.025
36	21.92	34	0.9455	0.031	0.034	-0.071	0.037	-0.044	-0.016
42	25.66	40	0.9617	0.004	0.027	-0.013	0.072	-0.066	0.025
48	32.31	46	0.9368	0.034	-0.032	-0.046	0.039	-0.038	-0.109

Table 4.2 Autocorrelation Check of Occupancy Residuals from ARMA(1,1) Model at VDS 401079 on May 11, 2006

To Lag	Chi-Square	DF	P-Value	Autocorrelation					
6	1.12	4	0.1914	0.002	0.003	-0.101	-0.094	-0.003	0.042
12	10.69	10	0.0594	-0.034	-0.100	-0.132	0.034	-0.023	0.092
18	12.93	16	0.0748	0.003	0.096	0.101	-0.031	0.030	-0.041
24	16.67	22	0.1931	-0.003	-0.007	0.052	-0.063	-0.043	-0.016
30	18.42	28	0.4391	-0.021	-0.032	-0.023	-0.013	0.030	-0.006
36	21.92	34	0.4523	0.030	0.098	0.005	0.045	-0.062	-0.038
42	25.66	40	0.3093	-0.073	-0.068	-0.045	-0.028	0.008	-0.125
48	32.31	46	0.4350	0.033	0.071	0.010	0.222	0.040	0.016

4.6 TRAFFIC FLOW PREDICTION

The dynamic traffic flow prediction model is tested in this section. The testing emphasis consists of one-step-ahead and multi-step-ahead prediction.

Traffic flow data used in this testing contains 60 full days aggregated flow rate and occupancy series collected by all 20 vehicle detector stations on the study corridor. Station-specific transformation parameter λ is estimated using two full months traffic flow data. The estimation results are presented in the Appendix A. For simple exponential smoothing, traffic flow data collected between May 4, 2006 and May 10, 2006 are used for initialization, in which both flow rate and occupancy prediction series S_t due to the process of seasonal operator are set to be the transformed series Y_t . Traffic flow data collected on May 11, 2006 are used for the initializations of the adaptive Kalman filtering. The actual dynamic traffic flow prediction starts from May 12, 2006 through July 3, 2006.

4.6.1 One-Step-Ahead Prediction

4.6.1.1 Structure of One-Step-Ahead Prediction System

The structure of the proposed one-step-ahead traffic flow prediction system is presented in Figure 4.7. Note that all symbols in the figure are consistent with those described previous sections. For seasonal operator, both S_t and S_{t+1} should be obtained from the simple exponential smoothing model. S_t is used to calculate the seasonally adjusted

residual W_t for the process of short-term operator. While the S_{t+1} is used for the calculation of final predicted value \hat{X}_{t+1} . For short-term operator, the parameters ϕ_t and θ_t are predicted for system update. Given the parameters ϕ_t and θ_t , \hat{W}_{t+1} is generated for calculating the final predicted flow rate and occupancy.

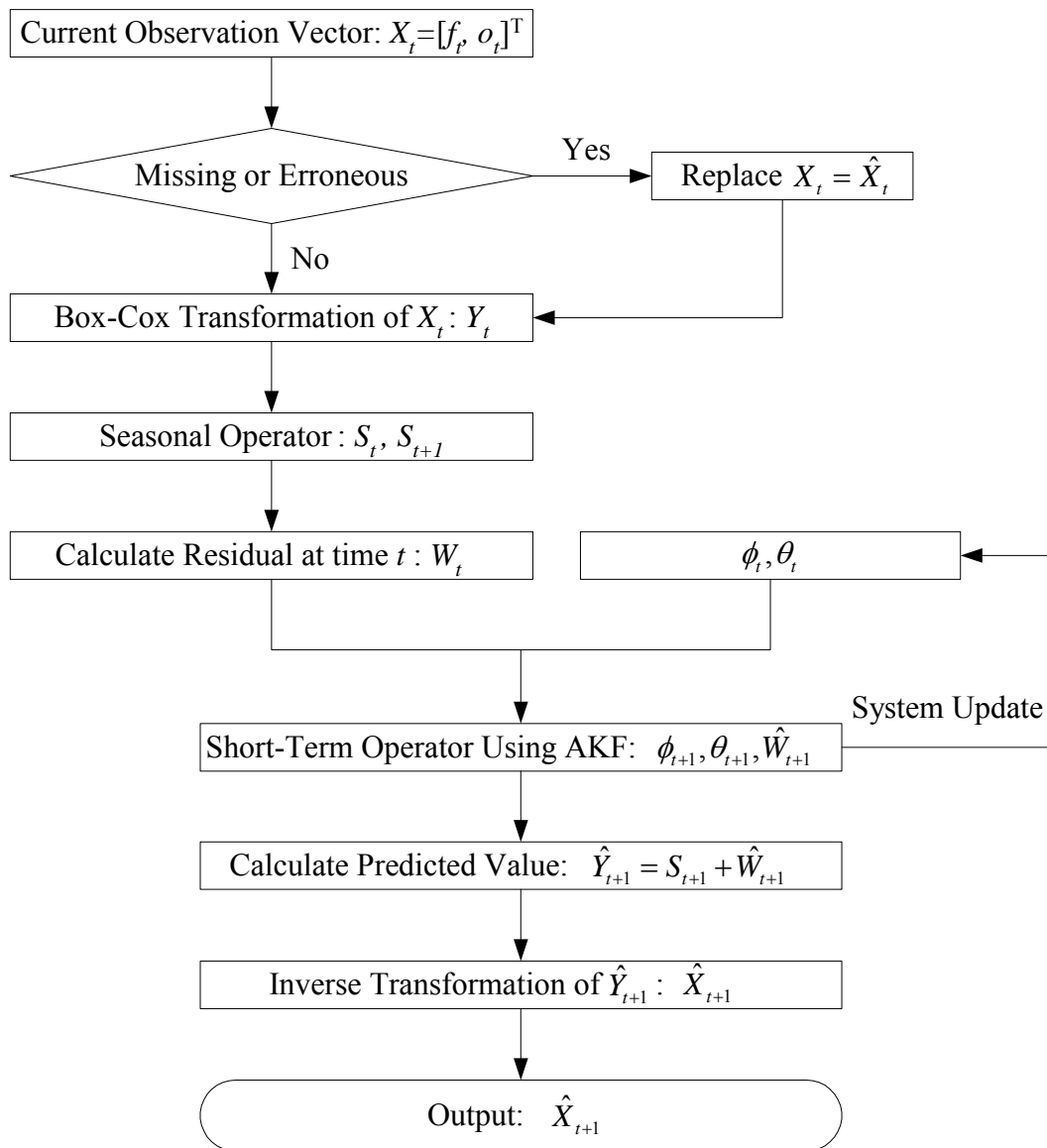
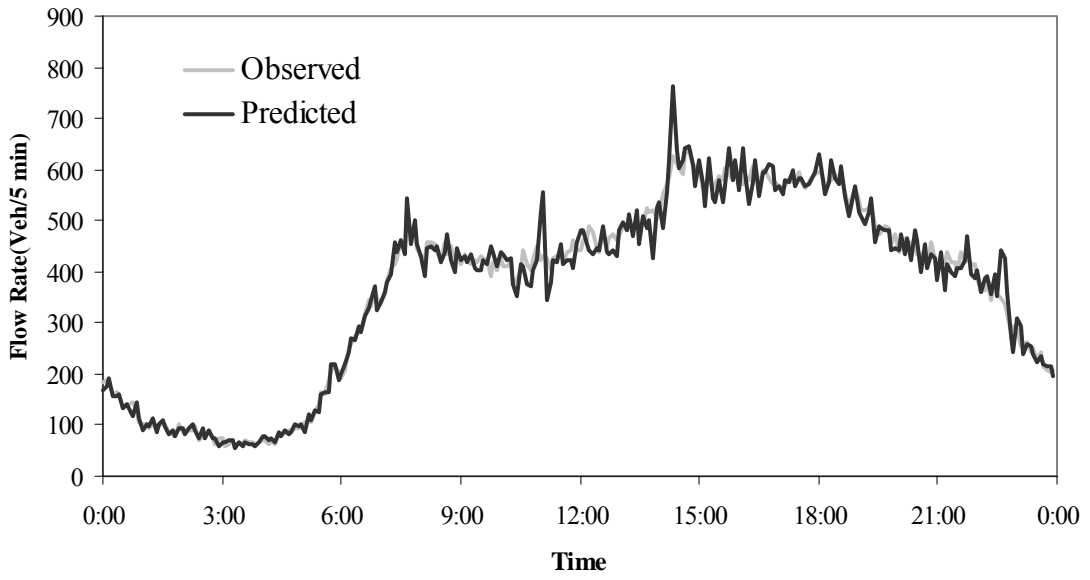


Figure 4.7 Structure of Online One-Step-Ahead Traffic Flow Prediction System

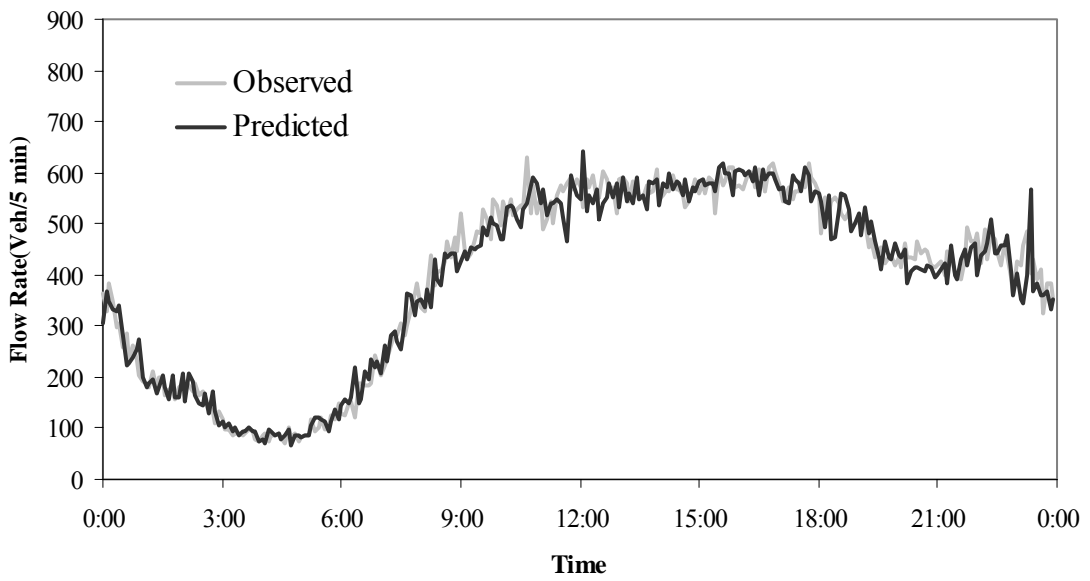
4.6.1.2 Performance Analysis

Results Illustration

A results illustration of the one-step-ahead (5-minute) prediction of flow rate and occupancy uses two full days traffic flow data on both weekday (05/17/2006, Wednesday) and weekend (05/13/2006, Saturday) at vehicle detectors station 401079, as shown in Figure 4.8 and Figure 4.9.

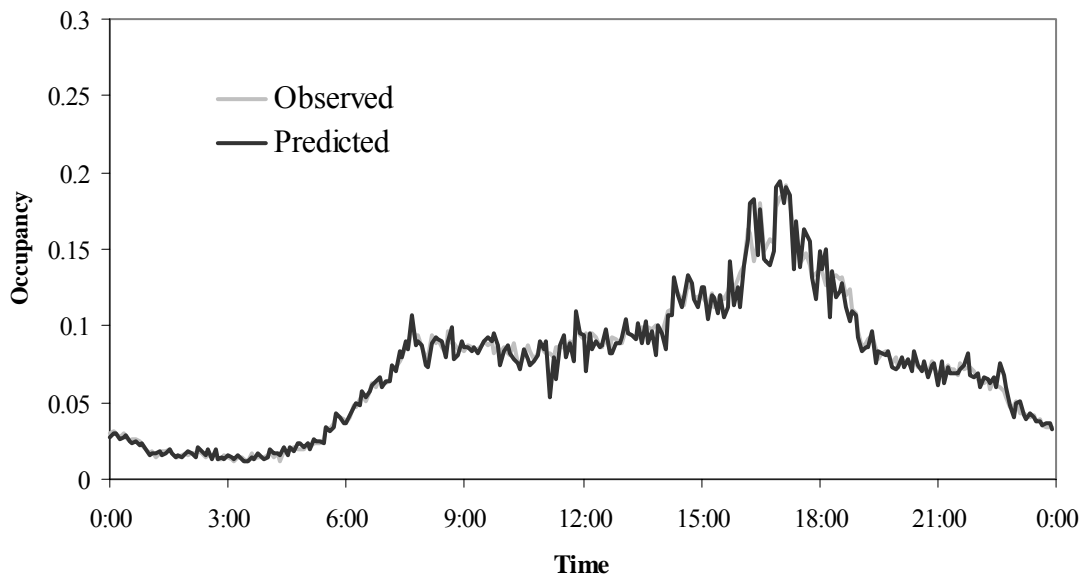


May 17, 2006, Wednesday

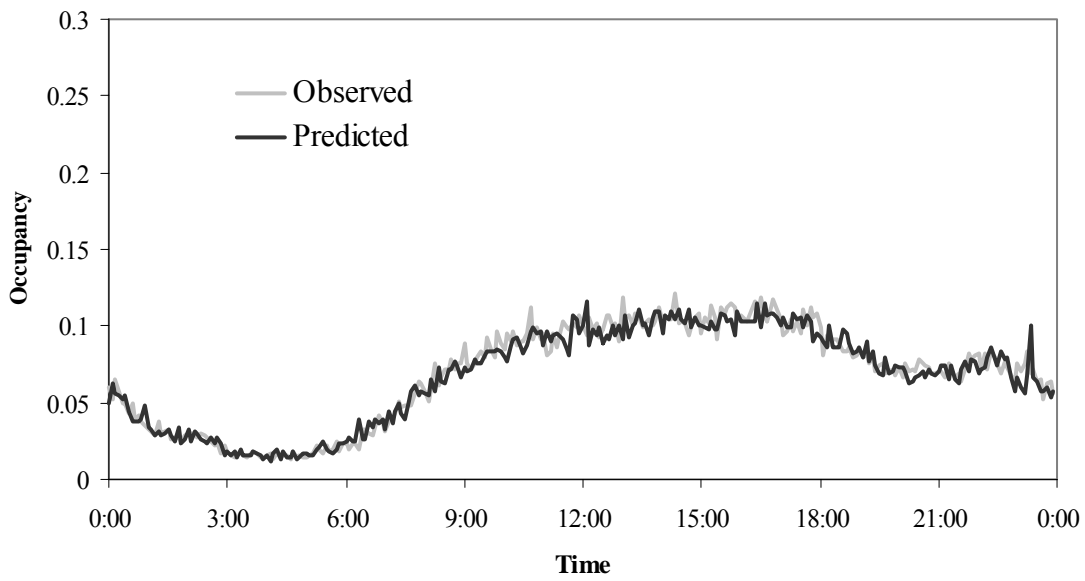


May 13, 2006, Saturday

Figure 4.8 Predicted and Observed Flow Rate at VDS 401079



May 17, 2006, Wednesday



May 13, 2006, Saturday

Figure 4.9 Predicted and Observed Occupancy at VDS 401079

From these two figures, several observations are found. Under the lighter traffic conditions with flow rates not as large as those collected on Wednesday midday and afternoon, sudden changes of flow rate and occupancy between two consecutive time points are seldom found, and the predicted flow rates and occupancies match the observed values very well. Comparatively, when sudden changes between two consecutive flow rates and occupancies are observed, the worse prediction performance reaches. This is caused by the mechanism of the adaptive Kalman filter. For recursive algorithms such as the proposed adaptive Kalman filtering algorithm, the underlying idea is to calculate the gradient of a cost function at time t under the assumption that the gradient at time $t-1$ is zero (that is, the previous parameter vector is optimal) so that a cost function of the prediction errors is minimized (Bohn and Unbehauen 2001). Taking the flow rates collected from 10:45 to 11:10 as an example, the observed flow rates at 10:45, 10:50, 10:55, 11:00, 11:05, and 11:10 are 402 veh/5min, 421 veh/5min, 449 veh/5min, 422 veh/5min, 430 veh/5min, and 427 veh/5min, respectively, without incidents occurring. From 10:45 to 10:55, the flow rates show an increasing trend. By capturing this pattern, the Kalman filtering algorithm predicts the flow rate at 11:00 with a value 516 veh/5min. Since the flow rates from 10:55 to 11:00 decreases from 449 veh/5min to 422 veh/5min, a bad prediction performance reaches at 11:00.

Flow rate and occupancy prediction at all other vehicle detector stations shows similar observations. This indicates that except for a few random sudden changes occurring in the observed measurements, the model can capture the weekly and local variation in flow rate and occupancy series.

Performance Evaluation

The performance of one-step-ahead flow rate and occupancy prediction at all vehicle detector stations is presented in Table 4.3. It can be seen that the performances of both flow rate and occupancy prediction accuracy vary across vehicle detector stations. The MAPE values of flow rate and occupancy prediction range from 7.13% to 10.63% and 9.45% to 12.49%, respectively. The worst performance for flow rate and occupancy prediction happens at vehicle detector stations 400976 and 401209 with MAPE values of 10.63% and 12.49%, respectively.

Table 4.3 Performance Evaluation of Flow Rate and Occupancy Prediction

VDS_ID	Flow Rate			Occupancy		
	MAPE (%)	MAE (veh/5min)	RMSE (veh/5min)	MAPE (%)	MAE	RMSE
401079	7.13	22.07	39.52	9.45	0.0063	0.0120
401239	9.36	23.58	34.24	10.63	0.0037	0.0070
401052	8.52	23.02	33.03	11.11	0.0064	0.0170
400329	9.03	21.96	32.79	11.76	0.0074	0.0245
401195	8.84	23.31	40.53	11.92	0.0085	0.0196
401558	8.77	23.09	34.29	10.33	0.0061	0.0116
400378	8.18	17.80	25.77	9.99	0.0038	0.0067
400445	8.76	23.04	37.11	11.57	0.0084	0.0421
400443	8.18	22.29	35.60	11.23	0.0088	0.0288
401221	8.45	24.81	40.03	11.3	0.0088	0.0204
401228	8.02	21.90	31.25	10.37	0.0065	0.0133
400081	8.96	22.27	35.34	11.51	0.0069	0.0158
400770	8.89	22.17	33.81	10.93	0.0054	0.0116
401243	9.43	24.41	37.81	12.48	0.0084	0.0191
401209	9.19	22.67	33.54	12.49	0.0083	0.0208
401260	8.83	22.56	34.43	11.78	0.0076	0.0251
400976	10.63	20.86	31.41	11.73	0.0065	0.0164
400838	8.85	22.02	32.42	11.40	0.0077	0.0323
400430	9.79	24.15	39.75	12.06	0.0072	0.0164
400865	8.77	22.31	63.62	11.57	0.0078	0.0180

4.6.2 Multi-Step-Ahead Prediction

There are two ways for predicting values of a variable over several steps: direct multi-step prediction and iterated multi-step prediction. Direct multi-step prediction is constructed based on the process of treating the several steps as a longer step in horizon (Chevillon and Hendry 2004). However, the number of steps is required to be determined before developing a direct multi-step prediction model. Furthermore, as the number of steps increases, the performance of a direct multi-step prediction model may fade rapidly by increasing the step size, where the features of the series data existing in finer steps are smoothed. Comparatively, the iterated multi-step prediction model is constructed based on one-step prediction model, iterated forward to the desired number of steps in horizon.

In this study, the iterated multi-step prediction model is applied for two reasons. First, the number of steps m in advance is required to be predetermined by a direct multi-step prediction model. However, the number of m cannot be determined because it varies with the links of the corridor for the corridor travel time prediction. Second, only 5-minute flow

rate and occupancy in the future are used for the travel time prediction, but a direct multi-step prediction model predicts the future values over a longer step.

4.6.2.1 Iterated Multi-Step Prediction Method

Seasonal Operator

Based on the equation (4.16), the multi-step-ahead prediction of the seasonal trends of the flow rate and occupancy is defined in the form:

$$S_{t+m} = \alpha Y_{t+m-2016} + (1-\alpha)S_{t+m-2016} \quad (4.26)$$

where m is the number of steps in advance from current time t . Since both $X_{t+m-2016}$ and $S_{t+m-2016}$ are available at time t , multi-step prediction of the seasonal trends of flow rate and occupancy can be obtained.

Short-Term Operator

For short-term operator of the SARIMA(1,0,1)(0,1,1)₂₀₁₆ model, an adaptive Kalman filter model has been developed. With repeated substitution into the measurement equation $\delta_{t+1} = \delta_t + \omega_t$ for m times, multiple step prediction of the state vector for the short-term operator is obtained as

$$\hat{\delta}_{t+m} = \hat{\delta}_{t+1} + \sum_{i=1}^{m-1} \omega_{t+i} \quad (4.27)$$

where m is the number of steps in advance and $\hat{\delta}_{t+m}$ is the multi-step predictions of δ_{t+m} at time t .

Similarly, with repeated substitution into the state error covariance equation $C_{\tilde{\delta}_{t+1}} = C_{\tilde{\delta}_t} + C_{\omega_t}$ for m times, the expectation of the multi-step-ahead prediction error covariance for the time-variant system is given by

$$C_{\tilde{\delta}_{t+m}} = C_{\tilde{\delta}_{t+1}} + \sum_{i=1}^{m-1} C_{\omega_{t+i}} \quad (4.28)$$

Equations (4.27) and (4.28) show a cumulative error occurring in both state vector prediction and the prediction error covariance. This is reasonable since the uncertainty increases with the increase of the number of steps in advance.

Given these calculations, the procedure of AKF can be used for multi-step-ahead prediction of the series W_{t+m} .

4.6.2.2 Structure of Proposed Multi-Step-Ahead Prediction System

The structure of the proposed multi-step-ahead traffic flow prediction system is presented in Figure 4.10. In this diagram, i represents the index of number of steps in advance, and it is initialized as zero before the algorithm starts. For seasonal operator, both S_{t+i} and S_{t+m} should be obtained from the simple exponential smoothing model. S_{t+1} is used to calculate the seasonally adjusted W_{t+i} from the seasonal operator and thus for the process of short-term operator. S_{t+m} is used for the calculation of final predicted value \hat{X}_{t+m} . For the process of short-term operator, the time varying parameters of the ARMA(1,1) model are updated at all steps 1, 2, ..., m . At the same time, $\hat{W}_{t+1}, \hat{W}_{t+2}, \dots, \hat{W}_{t+m}$ are generated since the iterated multi-step prediction is constructed by repeating the process of one-step-ahead prediction.

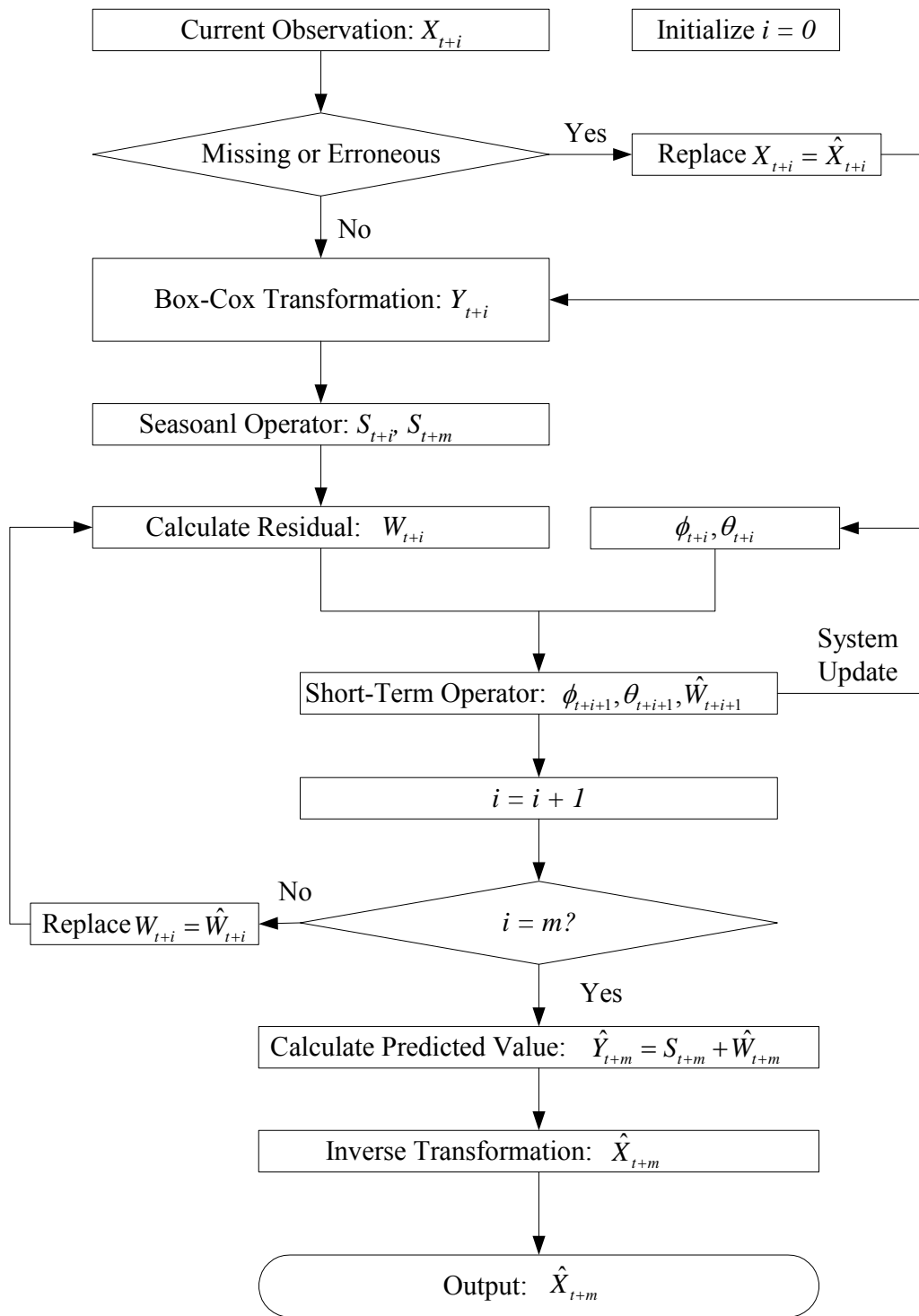


Figure 4.10 Structure of Online Multi-Step-Ahead Traffic Flow Prediction System

4.6.2.3 Performance Analysis

To illustrate the prediction accuracy evolution with the increase in number of steps, steps from 1 to 15 are tested at vehicle detector station 401079 using two full days' flow rate data collected on May 12, 2006 and May 13, 2006. The test results are presented in Table 4.4.

Table 4.4 Performance of Multi-Step-Ahead Prediction of Flow Rate at VDS 401079 on May 12, and May 13, 2006

Step	MAE(veh/5min)	MAPE(%)	RMSE(veh/5min)
1	27.69	8.51	37.95
2	36.26	11.09	49.51
3	35.68	10.93	47.76
4	35.83	10.84	47.77
5	34.86	10.58	46.46
6	35.08	10.64	46.20
7	35.10	10.62	46.58
8	34.78	10.59	45.81
9	34.83	10.60	46.04
10	34.99	10.61	46.38
11	34.69	10.53	45.88
12	34.53	10.51	45.75
13	34.7	10.55	45.77
14	34.46	10.51	45.33
15	34.73	10.55	45.85

With the increase of number of steps, it is expected that the performance of the multi-step prediction model fades with increasing uncertainty. However, the test results do not show this expectation. Although the performance degrades significantly from step 1 to step 2, the performance from step 2 to other steps is fairly stable. Equation $\hat{\delta}_{t+m} = \hat{\delta}_{t+1} + \sum_{i=1}^{m-1} \omega_{t+i}$ indicates that the expectation of the parameters of the time-varying ARMA(1,1) model is a constant. With the increase of number of steps in advance, the term $\sum_{i=1}^{m-1} \omega_{t+i}$ might either increase or decrease. The increasing uncertainty as show in equation (4.28) only increases the confidence interval of the predicted value at a given significance level.

4.7 SUMMARY

Using the traffic flow data collected by single loop detectors, this chapter presents a dynamic traffic flow prediction model, which can be readily implemented in real time. This method involves multi-step-ahead prediction for flow rate and occupancy in real time. A seasonal autoregressive integrated moving average (SARIMA) model is developed with an embedded adaptive predictor. This predictor adjusts the prediction error based on traffic data that becomes available every five minutes at each vehicle detector station. The involvement of multi-step-ahead prediction provides good opportunity for further development of a corridor travel time prediction system considering the traffic progression along the corridor, which is often ignored by other methods when using traffic measurements from single loop detectors.

Model adequacy is also analyzed using the traffic flow data collected on the study corridor. Testing results show that the proposed model is adequate for dynamic flow rate and occupancy prediction. Performance analysis shows that the proposed model with the embedded predictor can provide good estimates of flow rate and occupancy with one or more steps in advance.

CHAPTER 5

CORRIDOR TRAVEL TIME PREDICTION WITHOUT CONSIDERING INCIDENTS

5.1 OVERVIEW

The objective of this chapter is to develop a methodology that integrates the dynamic traffic flow predictor in the model development while considering the traffic progression along the corridor.

This proposed methodology involves the multiple-step-ahead prediction of flow rate and occupancy and corresponding mean effective vehicle length (MEVL) estimation. For a specific link defined as the freeway segment between two consecutive vehicle detector stations, the number of steps varying from link to link is used to account for the traffic progression along the corridor for further link travel time prediction. The final 5-minute-ahead prediction of corridor travel time is achieved by adding all predicted link travel times.

5.2 MEVL ESTIMATION

There are various ways to estimate MEVL using traffic flow data from single loop detectors, but the basic idea is to seek the fundamental relationship among the average speed, flow rate and occupancy. The basic form to present such a relationship is given by

$$g = \frac{V \times o}{f} \quad (5.1)$$

where g is the mean effective vehicle length, V is the average speed, o is the occupancy, and f is the flow rate.

It is acknowledged that the mean effective vehicle length varies with traffic conditions or time-of-day. Using a fixed free-flow-speed (FFS), Coifman (2001) estimated the MEVL under uncongested traffic conditions using equation (5.1), and extended the predicted MEVL to congested time periods. This requires free-flow-speed to be known. A free-flow-speed

that cannot represent the average speed under light traffic may result in significant errors under both uncongested and congested traffic conditions.

This work adopts Coifman's approach but incorporates an accurate estimation of free-flow-speed using traffic flow data from single loop detectors. The basic assumption of Coifman's approach is that the traffic compositions tend to be stable under congested traffic conditions. This is reasonably accepted and has been verified by Coifman (1996).

5.2.1 Free-Flow-Speed Estimation

Free-flow-speed is defined by the Highway Capacity Manual (HCM) as the average speed of vehicles on a given facility measured under low-volume conditions, when drivers are able to drive at their desired speed and are not constrained by control delays (TRB 2000). This indicates that the estimation of FFS requires a low-volume condition on the given facility.

To define the free-flow condition, traffic characteristics have been studied extensively. In this study, a nested clustering technique presented by Xia and Chen (2006) is applied for classifying freeway operation conditions into different flow phases. The method is advantageous in that it does not require a prior knowledge of the number of clusters. Stating with the set of traffic flow data collected at each vehicle detector station, the number of clusters (i.e., subsets of data with distinct traffic characteristics) that these data points should be grouped into is automatically determined based on the statistical characteristics of the data.

Under free-flow traffic conditions, it is observed that almost all heavy vehicles drive in the shoulder and/or middle lane. Seldom do heavy vehicles drive in the median lane. This provides a good opportunity to estimate FFS by using sensor data collected from the median lane when we assume that there is no large difference among the speeds at different lanes. Knowing that that the mean effective length of an average passenger car is about 20 feet (Kwon 2003), we estimate the FFS on the median lane at time step t by

$$FFS_t = \frac{f_t \cdot g}{o_t} \quad (5.2)$$

where,

FFS_t : the FFS of a given roadway station at time step t ;

f_t : the flow rate of a given roadway station at time step t in the median lane;

g : the MEVL of a passenger car;

o_t : the occupancy of a given roadway station at time step t for median lane.

The station-specific free-flow-speed can be calculated by

$$FFS = \frac{1}{n} \sum_{i=1}^n FFS_i \quad (5.3)$$

where n is the number of samples representing the free-flow traffic conditions.

5.2.2 MEVL Estimation

The detailed method for the estimation of MEVL presented by Coifman is described as follows.

In the case of free-flow traffic conditions, station-specific MEVL at time step t is calculated by

$$\hat{g}_t = \frac{FFS \cdot o_t}{f_t} \quad (5.4)$$

where \hat{g}_t is the mean effective vehicle length at time step t , FFS is the estimated station-specific free-flow-speed, o_t is the occupancy at time step t , and f_t is the flow rate.

In the case of congested conditions, the estimate of MEVL from immediately previous uncongested traffic condition is extended. This is because the traffic compositions tend to be stable over short time periods and the MEVL varies little under the congested traffic conditions (Cofman 1996). The stable features of traffic compositions under congested traffic conditions is also verified using real vehicle classification data measured in 2003 on freeways around Louisville, Kentucky. It is observed that the percentage of vehicles other than passenger cars varies significantly during light traffic conditions varying from 4% to 11%. Comparatively, the percentage of vehicles other than passenger cars varies less under congested traffic conditions ranging from 4% to 6%.

The MEVL estimation method can be used for each vehicle detector station corresponding to the predicted flow rate and occupancy at time step t .

5.3 CORRIDOR TRAVEL TIME ESTIMATION MODEL

5.3.1 Link Travel Time Derivation

Figure 5.1 shows a typical corridor link i between two consecutive vehicle detector stations without showing on- and/or off-ramps. $(f_{i,t}, o_{i,t}, g_{i,t})$ and $(f_{i+1,t}, o_{i+1,t}, g_{i+1,t})$ represent flow rate, occupancy, and MEVL entering and exiting the link during time step t at upstream station and downstream stations, respectively.

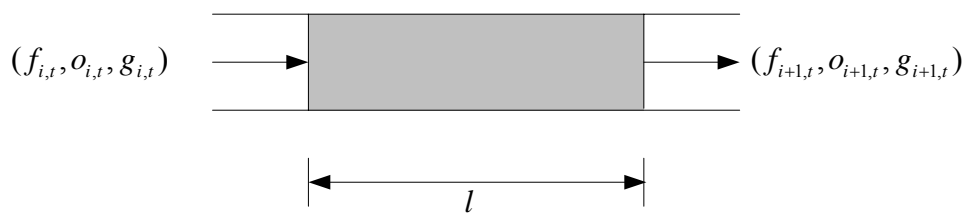


Figure 5.1 Typical Link of Freeway

Given the flow rate, occupancy, and estimated MEVL at time step t , the average spot speed at a specific station can be calculated as

$$V_{i,t} = \frac{g_{i,t} \cdot f_{i,t}}{o_{i,t}} \quad (5.5)$$

The travel time of link i at time step t can be calculated as the link length divided by the space-mean speed $V(i,t)$. When $l_i \rightarrow 0$, the link space-mean speed can be approximated as the harmonic average spot speeds at stations i and $i+1$ as

$$V(i,t) = 2 \left(\frac{1}{V_{i,t}} + \frac{1}{V_{i+1,t}} \right)^{-1} \quad (5.6)$$

The link travel time at time step t can be represented as

$$tt = \frac{l_i}{V(i,t)} = \frac{l_i}{2} \left(\frac{1}{V_{i,t}} + \frac{1}{V_{i+1,t}} \right) \quad (5.7)$$

Note that in the real world, it is impossible for the link length to approach zero. However, it is reasonable to maintain these approximations if the link length l_i and the

reported duration ΔT (5 minutes in this study) of the time step satisfy the condition

$$\frac{l_i}{FFS} < \Delta T \leq 5-10 \text{ min (Nam and Drew 1996)}.$$

5.3.2 Corridor Travel Time Prediction

One-step-ahead prediction of corridor travel time at time step $t+1$ with traffic measurements up to time step t can be readily obtained by adding all link travel times based on the predicted flow rate, occupancy, and corresponding MEVL at time step $t+1$. However, this corridor travel time prediction method takes no consideration for traffic progression along the corridor. Since the driver needs some time to arrive at the upstream station of a specific link i from the start point of the corridor, the link travel time should be calculated using the predicted traffic measurements at the time step $t + m_{i-1} + 1$, where m_{i-1} is the number of steps that the time covers for the driver arriving at the upstream station of link i .

The proposed corridor travel time prediction model is used to predict link travel times while considering the effects of traffic progression along the corridor. As shown in Figure 5.2, suppose that we have a corridor consisting of j links, and that a driver would like to leave the origin of the corridor at time step $t+1$. Let tt_i denote the predicted travel time of link i , and TT_i denote the predicted travel time from the start point of the corridor to the downstream of link i . Let t_0 denote the actual time at the end of time step $t+1$. Given the traffic measurements up to time step t , the procedure for predicting the corridor travel time at time step $t+1$ is described as follows.

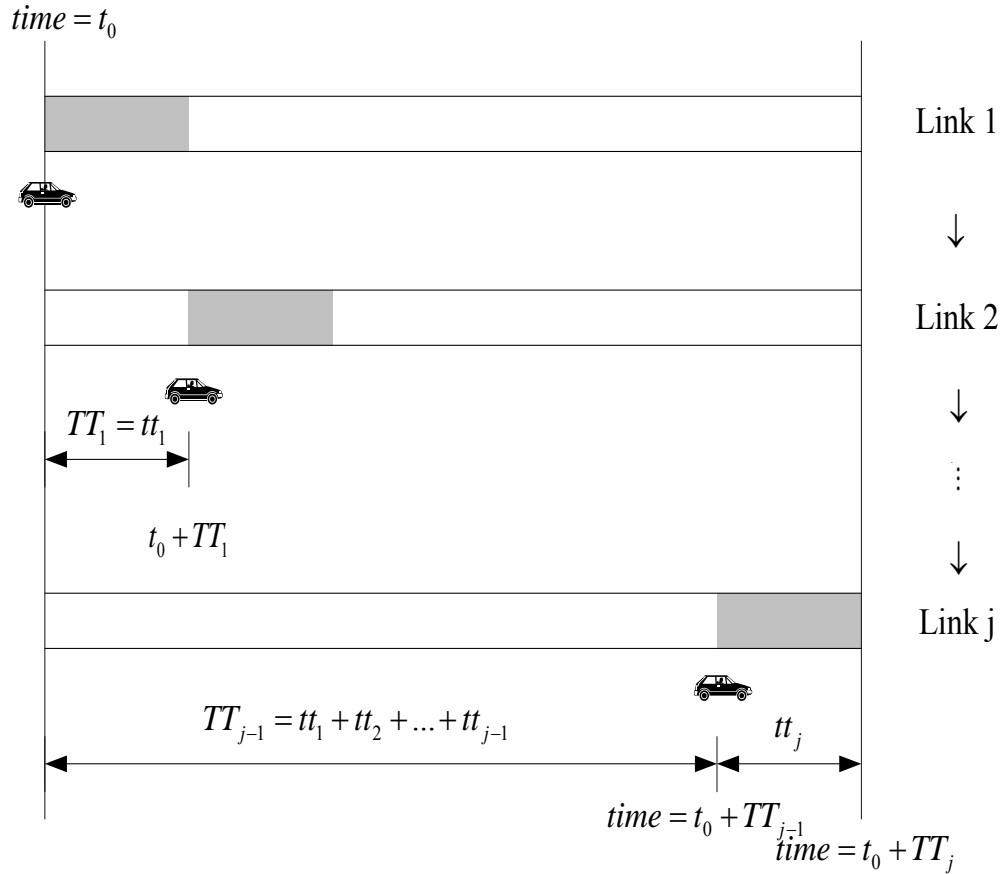


Figure 5.2 Illustration of Corridor Travel Time Prediction

For the first link of the corridor, only one-step ahead prediction of flow rate and occupancy at stations 1 and 2 is required for estimating the corresponding MEVL $g_{1,t+1}$ and $g_{2,t+1}$, and thus computing the travel time tt_1 as

$$tt_1 = \frac{l_1}{V(1,t+1)} = \frac{l_1}{2} \left(\frac{1}{V_{1,t+1}} + \frac{1}{V_{2,t+1}} \right) \quad (5.8)$$

For the second link of the corridor, it is known that the driver needs time $TT_1 = tt_1$ to reach its upstream. Thus, the flow rate and occupancy at stations 2 and 3 at time step $t + m_2$ are required to be predicted for calculating its travel time, where m_2 is the number of steps that TT_1 covers. The number of steps m_2 is obtained by rounding $\frac{TT_1}{\Delta T} + 1$ to an integer.

Repeat the process of link 2. The travel times of link 3,4,...,j can be predicted in sequence consistent with the time steps $t + m_3, t + m_4, \dots, t + m_j$. m_3, m_4, \dots, m_j are

calculated by rounding $\frac{TT_2}{\Delta T} + 1, \frac{TT_3}{\Delta T} + 1, \dots, \frac{TT_{j-1}}{\Delta T} + 1$ to integers, where $TT_i, i = 2, 3, \dots, j$, is calculated as $TT_i = TT_{i-1} + tt_i$ in sequence. The final predicted corridor travel time is TT_j at time step $t+1$ given the flow rate and occupancy measurements up to time step t .

5.4 PERFORMANCE ANALYSIS

5.4.1 MEVL Estimation

5.4.1.1 Free-Flow-Speed Estimation Results

Station-specific free-flow-speeds are estimated using all 3 months of traffic flow data, including 5-minute aggregated flow rate and occupancy data, collected for the median lane of the study corridor.

Figure 5.3 presents the two cluster results in the plots of flow rate versus occupancy at VDS 400329. It is observed that a linear relationship exists in the left portion of the diagram, indicating light traffic conditions. While the low-volume and low occupancy as plotted in Figure 5.3 represents the free-flow conditions. Note, in this case, the boundary value occupancy and flow rate is about 0.05 and 60~70 veh/5min/lane. Variation of the plots representing the flow-flow traffic conditions also represents the traffic variation under free-flow traffic conditions.

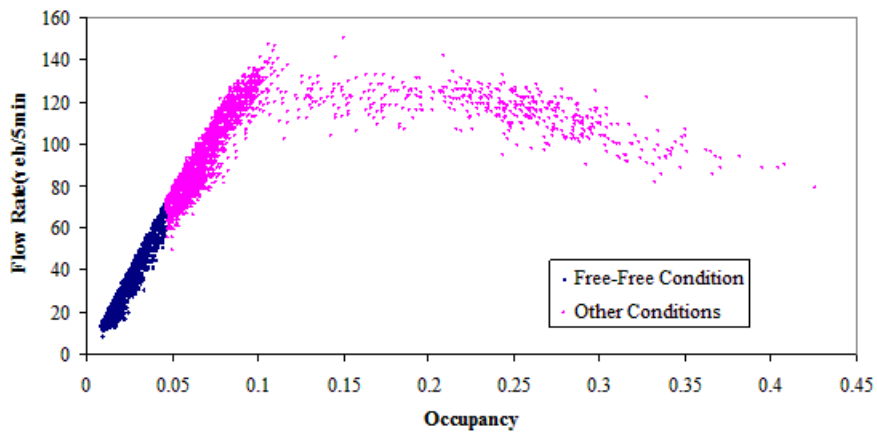


Figure 5.3 Cluster Results in the Plots of Flow Rate versus Occupancy at VDS 400329

Station-specific free-flow-speed estimates and their standard deviations are calculated using equation (5-3) and presented in Table 5.1. It can be seen that the estimates of free-flow-speeds vary at different vehicle detector stations. Even for two consecutive vehicle detector stations with a short distance between them, the difference between the estimates of free-flow-speed may be very large. Take VDS 400081 and VDS 400770 as examples, the distance between them is 0.10 mile, but the estimates of free-flow-speed for daytime are 65.64 mph and 77.97 mph, respectively.

Table 5.1 Results of Free-Flow-Speed Estimation

VDS_ID	ABS Milepost	Estimated		“Actual”	
		FFS (mph)	Standard Deviation (mph)	FFS (mph)	Standard Deviation (mph)
401079	11.95	74.11	6.16	72.59	3.69
401239	12.66	104.19	5.46	68.97	2.26
401052	13.406	90.19	7.60	71.04	2.30
400329	13.786	70.36	5.84	70.15	3.56
401195	14.466	66.09	6.30	69.61	3.62
401558	14.886	68.04	7.88	70.36	3.29
400378	15.396	79.47	4.34	68.06	2.63
400445	15.966	69.88	5.98	69.89	3.55
400443	16.316	64.19	5.32	69.54	3.61
401221	16.756	66.46	6.66	69.69	3.58
401228	17.406	64.88	3.68	69.76	2.38
400081	17.546	65.64	5.29	70.67	3.45
400770	17.646	77.97	5.48	69.64	2.76
401243	18.846	64.88	5.93	70.46	3.55
401209	19.286	63.42	6.55	69.90	3.70
401260	19.546	67.39	7.08	70.14	4.95
400976	19.916	70.64	5.27	69.83	3.33
400838	20.236	66.15	6.08	72.71	3.52
400430	20.636	74.31	6.04	68.93	3.47
400865	20.956	64.68	6.07	73.09	3.21

The “actual” free-flow speeds are also calculated using the measured spot speeds. Comparison between the estimated and actual free-flow speeds shows that the overall performance of the free-flow speed estimation is good. However, at some specific stations, such as VDS 401239 and 401052, the estimates of free-flow speed are very poor, largely deviating from the actual free-flow speeds. The check of the traffic flow data shows that the data quality is suspicious although they both pass the criteria for flow rate and occupancy data screening. The mean effective vehicle lengths calculated from the measured flow rate, occupancy, and spot speed are 13.24 ft and 15.75 ft at VDS 401239 and 401052, respectively. Considering the possible data quality problems in estimating free-flow speed, the mean free-flow speed across all vehicle detector stations is taken for further travel time prediction, as 70 mph since the mean estimates of free-flow speeds across all vehicle detector stations is about 71.65 mph.

5.4.1.2 MEVL Estimation Results

The testing of mean effective vehicle length estimation uses traffic flow data from May 12, 2006 to July 3, 2006, collected on the study corridor for all vehicle stations. Actual mean effective vehicle lengths are computed using equation (5.1), where the average speed is measured from dual loop detectors.

Table 5.2 presents the estimation performance at all vehicle detector stations based on 53 full days of traffic flow data. It is observed that the proposed method has the worst performance at VDS 400430. Its MAPE value under congested traffic conditions is 8.93%. In considering that the difference between the estimate and “actual” free-flow speeds (70 mph and 68.93 mph, respectively) at VDS 400430 is not as large as those stations such as VDS 400430 and VDS 400838, the worst performance might be caused by the extension of MEVL from uncongested traffic conditions to congested traffic conditions.

Table 5.2 Performance of MEVL Estimation at All Vehicle Detector Stations

VDS_ID	MAE (ft)	MAPE (%)	RMSE (ft)
401079	1.26	5.73	1.71
401239	0.58	3.88	0.95
401052	0.77	4.14	1.68
400329	1.24	5.46	1.86
401195	1.60	6.61	2.52
401558	1.26	5.61	1.78
400378	1.18	6.24	1.58
400445	1.30	5.70	1.98
400443	1.56	6.37	2.38
401221	1.54	6.67	2.50
401228	0.96	3.95	1.34
400081	1.34	5.49	1.87
400770	0.91	4.42	1.48
401243	1.57	6.46	2.27
401209	1.63	7.69	2.76
401260	1.55	6.68	2.13
400976	1.14	4.96	1.81
400838	1.72	7.15	2.68
400430	1.79	8.93	2.36
400865	1.44	5.66	2.29

5.4.2 Corridor Travel Time Prediction

Testing of the proposed corridor travel time prediction model uses the traffic flow data collected from May 4, 2006 to July 3, 2006 on the study corridor. The traffic flow data collected from May 4, 2006 to May 11, 2006 is used to obtain the parameters for stating the embedded dynamic traffic flow predictor. The actual corridor time prediction starts from May 12, 2006. The prediction results from June 1, 2006 to June 30, 2006 are presented in Appendix B for further reference in data analysis.

In order to compare the predicted corridor travel times with the actual corridor travel times for performance analysis, actual travel times should be provided. However, the actual travel times are generally not available. On the other hand, the dual loop detectors embedded on the study corridor collect spot speeds from vehicle detector stations at 5-minute increments. Using the measured average spot speed, the actual travel times are computed in this study using the method presented by Chen (2004) as described in Chapter 2, Section 2.3.1. Research has shown that with intelligent interpolation between the measurement locations and times, accurate estimates of the true travel time can be obtained (Coifman 2002, Van Lint and Van Der Zijpp 2004).

5.4.2.1 Structure of Corridor Travel Time Prediction System

The structure of the proposed 5-minute-ahead corridor travel time prediction system integrating the dynamic traffic flow predictor and MEVL estimator is presented in Figure 5.4. Note that all symbols in Figure 5.4 are consistent with those described in section 5.2.

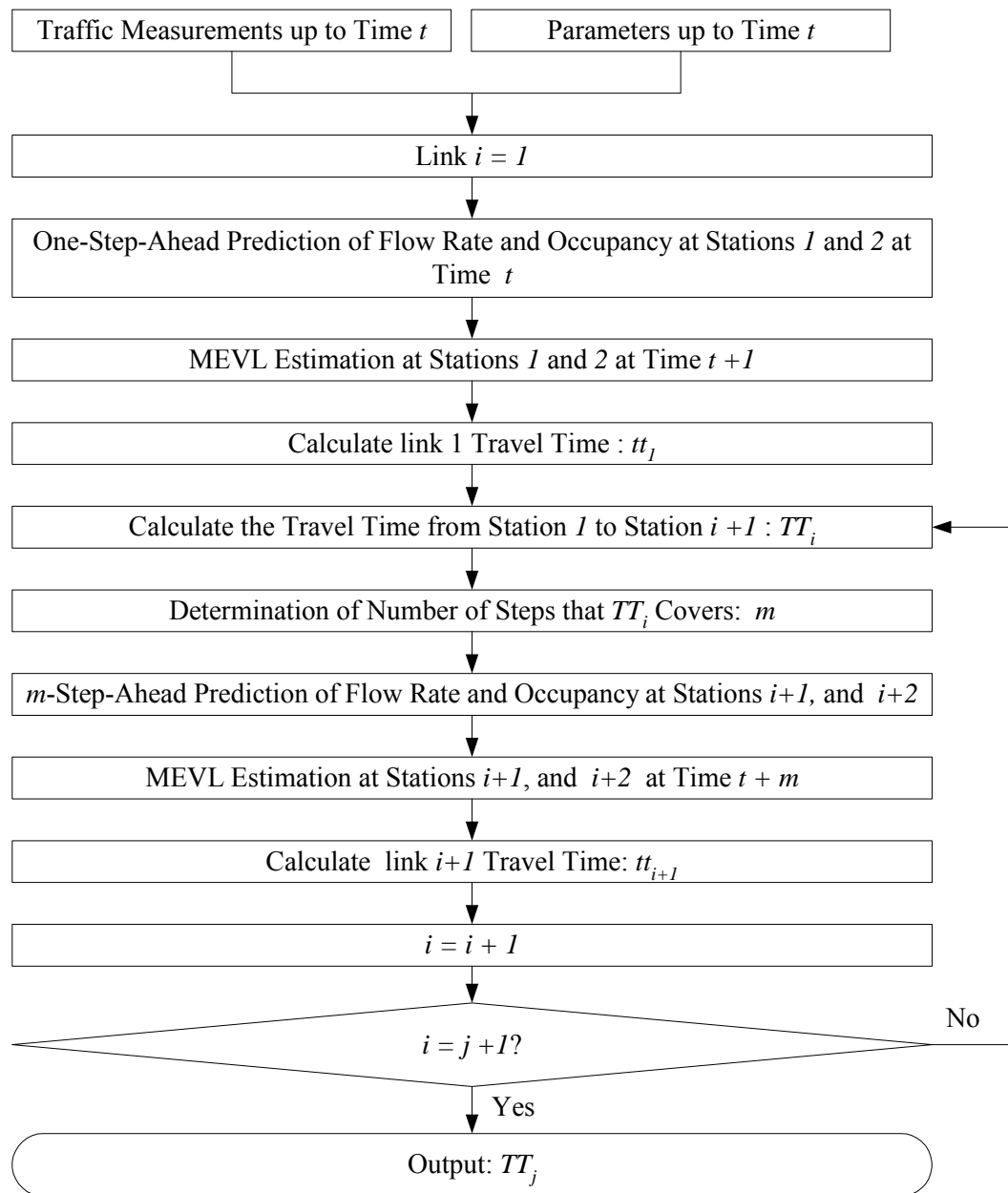
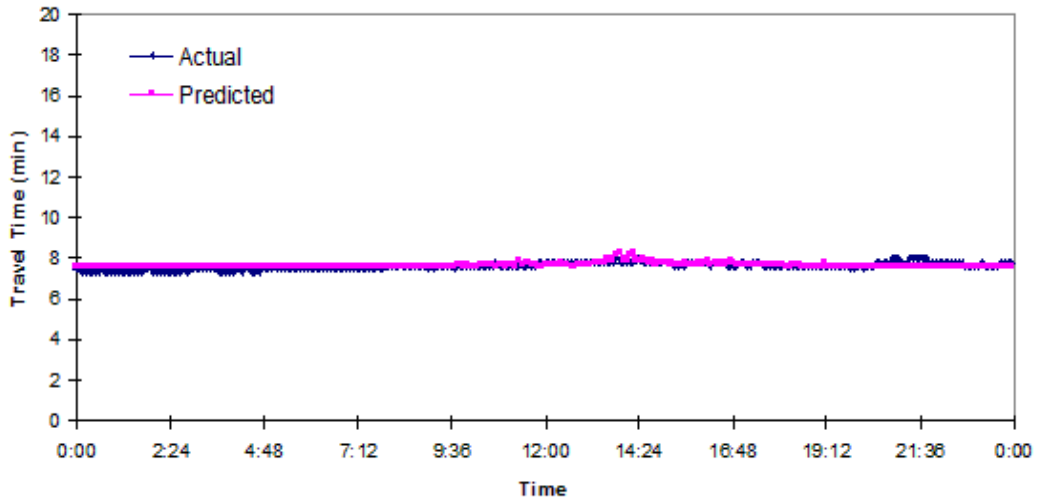


Figure 5.4 Structure of Online One-Step-Ahead Corridor Travel Time Prediction System

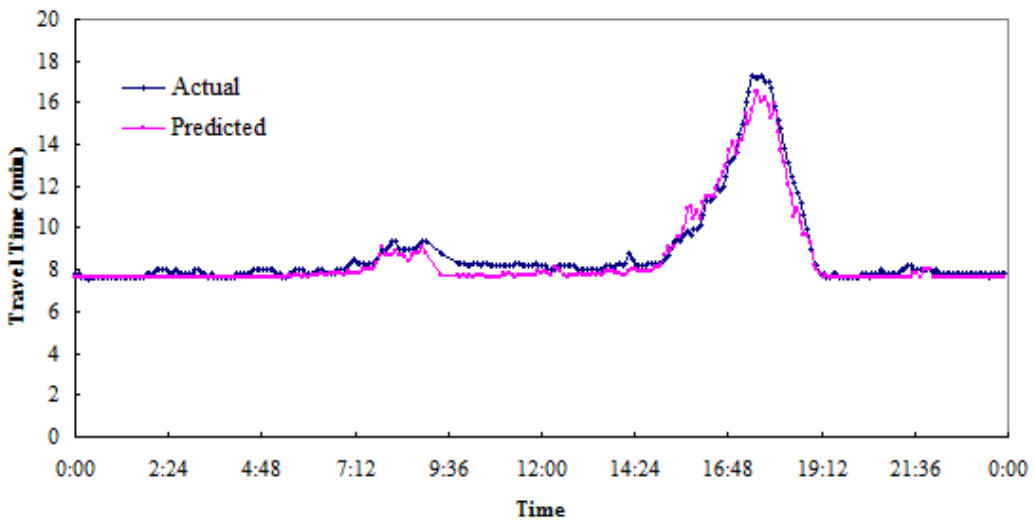
5.4.2.2 Performance Analysis

Results Illustration

Figure 5.5 and Figure 5.6 demonstrate the predicted and actual corridor travel times by time-of-day, where a weekend day and a weekday are randomly selected; on Sunday, June 11, 2006 and Wednesday, June 7, 2006. In each diagram, the actual and predicted corridor travel times are plotted.



**Figure 5.5 Predicted and Real Corridor Travel Time by Time-of-Day on June 11, 2006,
Sunday**



**Figure 5.6 Predicted and Real Corridor Travel Time by Time-of-Day on June 7, 2006,
Wednesday**

From Figure 5.5, it can be seen that the proposed corridor travel time prediction model works well on weekends without a significant difference between the predicted and real corridor travel times. The maximum difference between the predicted and actual travel times on June 11, 2006 is 0.41 minutes with predicted and actual travel times of 8.24 minutes and 7.83 minutes, respectively.

On the other hand, a larger difference between the predicted and actual travel times is observed on the weekday of June 7, 2006, Wednesday in Figure 5.6. The maximum difference occurs at around 18:30 at ABS milepost 16.63. Inspection of the incident log shows that a traffic collision occurred at that time starting at 17:46 and being cleared at 17:55. Correspondingly, the prediction performance becomes worse from 17:50 until 18:45 as presented in Table 5.3. Except for the time at 18:00, the predicted travel times are always smaller than the actual one during this period. This means that the proposed corridor travel time prediction method without considering the incident effects tends to underestimate the travel times.

Table 5.3 Predicted and Actual Corridor Travel Time from 17:45 to 18:50 on June 7, 2006

Time	Predicted Corridor Travel Time (min)	Actual Travel Time (min)	Difference between the Predicted and Actual Travel Time (min)
17:45:00	16.23	17.00	0.77
17:50:00	15.82	17.00	1.18
17:55:00	15.29	16.67	1.38
18:00:00	15.96	15.83	-0.13
18:05:00	14.55	15.17	0.62
18:10:00	13.76	14.83	1.07
18:15:00	13.15	13.83	0.68
18:20:00	12.09	13.17	1.08
18:25:00	11.62	12.5	0.88
18:30:00	10.53	12.17	1.64
18:35:00	10.96	11.67	0.71
18:40:00	10.24	11.17	0.93
18:45:00	9.66	10.67	1.01
18:50:00	9.70	10.00	0.30
18:55:00	9.32	9.00	-0.32
19:00:00	8.18	8.33	0.15

Another big difference between the predicted and actual times is observed at around 9:25. Inspection of the incident log indicates that no incident happened at that time. A check of the traffic flow measurements from single loop detectors shows that the difference

between the estimate of free-flow speed and measured spot speed leads to the large error in corridor travel time prediction. The underestimated corridor travel time is caused by the travel time prediction of two consecutive links between vehicle detector stations 401239 and 400329. The first link is between VDS 401221 and 401052, and the second link is between VDS 401052 and 400329. Table 5.4 presents the traffic measurements including flow rate, occupancy and average spot speed for these stations. From time 9:20 to 9:35, both flow rates and occupancies are very low, and the traffic conditions are classified as free-flow traffic conditions. However, when looking at the measured spot speeds, it is found that such measured spot speeds are also lower than the estimated free-flow speed. When using free-flow speed for link travel time prediction, the underestimated link travel time is obtained. Since the overestimated or underestimated free-flow speed is unavoidable because of the uncertain factors such as the weather conditions, the overestimated or underestimated corridor travel time is also understandable.

Table 5.4 Traffic Measurements from 9:20 am to 9:35 am at VDS 401239, 401052 and 400329

VDS ID	Date Time	Milepost	Flow Rate (veh/5min)	Occupancy	Speed (mph)
401239	2006-06-07 09:20:00	12.66	351	0.0415	68.3
401239	2006-06-07 09:25:00	12.66	353	0.0428	67.1
401239	2006-06-07 09:30:00	12.66	378	0.0465	66.8
401239	2006-06-07 09:35:00	12.66	317	0.0430	62.9
401052	2006-06-07 09:20:00	13.406	449	0.0687	59.8
401052	2006-06-07 09:25:00	13.406	448	0.0673	64.4
401052	2006-06-07 09:30:00	13.406	457	0.0726	64.3
401052	2006-06-07 09:35:00	13.406	441	0.0683	65.3
400329	2006-06-07 09:20:00	13.786	283	0.0593	65.5
400329	2006-06-07 09:25:00	13.786	319	0.0698	63.3
400329	2006-06-07 09:30:00	13.786	311	0.063	65.1
400329	2006-06-07 09:35:00	13.786	328	0.065	66.3

Performance Evaluation

Table 5.5 presents the performance of the corridor travel time prediction model by time-of-day using the 30 days of traffic flow data from June 1, 2006 to June 30, 2006. It is observed that the performance of corridor travel time prediction varies from time to time. The best performance can be seen at night, while the worst performances are obtained during the daytime particularly during afternoon peak periods. Based on the 30 day corridor travel

time prediction results, the worst performance occurs during the period between 16:00 to 17:00 with the MAE, MAPE, and RMSE values of 1.16 minutes, 9.84%, and 1.64 minutes, respectively. The check of measured traffic flow data indicates that it is a peak afternoon period, especially on weekdays, when the mean effective vehicle length is estimated by extending the mean effective vehicle length from previous uncongested traffic conditions. The other possible reason that causes the large error in corridor travel time prediction is the more steps involved in flow rate and occupancy prediction for link travel time prediction in the peak period when compared to the off-peak period.

Table 5.5 Performance of the Corridor Travel Time Prediction by Time of Day

Time of Day	MAE(min)	MAPE(%)	RMSE(min)
0:00~1:00	0.21	2.70	0.26
1:00~2:00	0.27	3.40	0.35
2:00~3:00	0.34	4.12	0.44
3:00~4:00	0.43	5.16	0.61
4:00~5:00	0.50	5.79	0.70
5:00~6:00	0.46	5.37	0.63
6:00~7:00	0.46	5.44	0.60
7:00~8:00	0.57	6.43	0.73
8:00~9:00	0.74	8.25	0.98
9:00~10:00	0.56	6.40	0.78
10:00~11:00	0.55	6.43	0.66
11:00~12:00	0.56	6.64	0.65
12:00~13:00	0.51	6.12	0.58
13:00~14:00	0.56	6.54	0.69
14:00~15:00	0.57	6.60	0.49
15:00~16:00	0.71	6.76	1.12
16:00~17:00	1.16	9.84	1.64
17:00~18:00	0.91	7.53	1.33
18:00~19:00	0.63	6.40	0.85
19:00~20:00	0.21	2.58	0.33
20:00~21:00	0.21	2.56	0.36
21:00~22:00	0.32	3.65	0.67
22:00~23:00	0.18	2.21	0.21
23:00~24:00	0.15	1.86	0.19

5.5 SUMMARY

Using the traffic flow data collected by single loop detectors (and other devices that emulate its operation), this chapter presents a dynamic corridor travel time prediction model, which can be readily implemented in real time integrating the dynamic predictor for traffic flow prediction. Traffic progression along corridor is considered by adjusting the number of steps for dynamic flow rate and occupancy prediction.

Testing of the proposed corridor travel time prediction is performed using the traffic flow data collected on the study corridor. Performance results shows that the proposed corridor travel time prediction model can capture both the seasonal and local variation of traffic flow. The performance also shows that the proposed model has the best performance at light traffic conditions, but degrades for the congested traffic conditions. The larger error in corridor travel time prediction for congested traffic is caused by two reasons. The first reason is less accuracy of MEVL estimation for congested traffic, where the MEVL is estimated by extending the estimate of MEVL from previously uncongested traffic. The other reason is more steps are involved in flow rate and occupancy prediction for link travel time prediction under congested traffic when compared to uncogested traffic.

CHAPTER 6

CORRIDOR TRAVEL TIME PREDICTION

CONSIDERING INCIDENTS

6.1 OVERVIEW

Once an incident occurs on the freeway and has significant impacts on the traffic, traffic begins backing up behind the incident. The amount and the extent of the backup is a function of factors including the approaching traffic volume, number of lanes blocked, incident type, time-of-day, day-of-week, etc. Consequently, the backup of physical queue affects corridor travel times. Chapter 5 demonstrated that the proposed method without considering incident impacts for corridor travel time prediction is not fully capable of capturing the traffic characteristics under an incident, causing some large difference between the actual and predicted corridor travel times within incident influence time. The limitation of the method without considering incidents is that it uses the multiple-step-ahead predicted flow rate and occupancy to calculate travel times for those links that are actually affected by the incident. Unfortunately, there is a large difference between the multiple-step-ahead predicted and measured flow rate and occupancy because the sudden changes in flow rate and occupancy cannot be fully predicted by the dynamic traffic flow predictor.

As an example, a selected incident occurred at 17:01 on June 30, 2006 on the study corridor at milepost 16.43. According to the corridor travel time prediction method, as presented in Chapter 5, the travel time prediction for the link, on which the incident occurred, involves the prediction of flow rate and occupancy two steps in advance. At the upstream station to the incident, flow rate prediction results indicate that the maximum absolute percentage error reaches 31% within the incident influence time. By using such predicted flow rate and occupancy, the predicted link travel time greatly diverges from the “actual” link travel time because both flow rate and occupancy prediction are not so accurate as that those without incident impacts on traffic.

In this chapter, a methodology is developed that adjusts the predicted corridor travel time within the incident influence time. This method involves the estimation of incident duration from historical incidents collected on the study corridor, such that the duration of a new incident occurring in real time can be predicted. The proposed method also involves

identifying the impact of an incident on traffic. If an incident significantly affects the traffic, the predicted corridor travel time is adjusted. Otherwise, the predicted corridor travel time is not adjusted. Adjustment of corridor travel time within incident influence time is performed based on the queuing analysis.

6.2 PREDICTION OF INCIDENT DURATION

Figure 6.1 presents the timeline for the duration of an incident (Hagen 2005). Incident duration is the time from the occurrence of the incident until all of the responders have left the scene. It directly determines how long the incident affects the traffic.

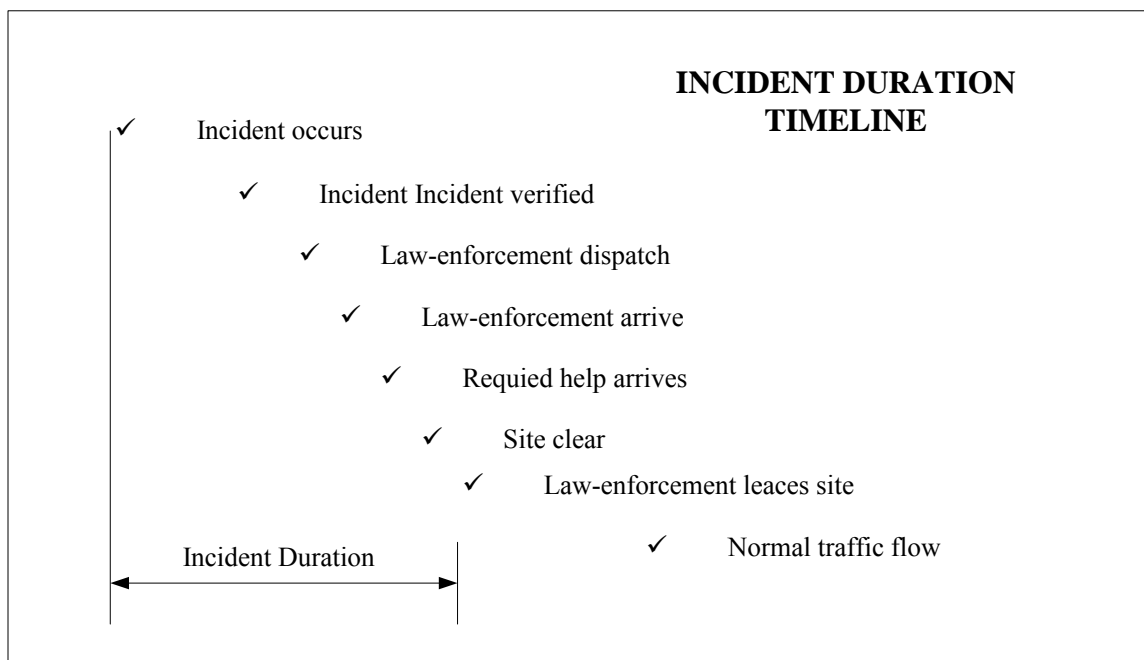


Figure 6.1 Illustration of Incident Duration

The basic idea of this work, to predict the duration of a new incident in real time, is to establish a look-up table, providing estimated incident duration by a set of factors extracted from historical incidents. Duration of an incident in real time is predicted as the duration in the look-up table with the same determinant factors.

Historical incidents collected on the study corridor provide information on incident type, starting time, location, and actual incident duration. To establish the look-up table, potential factors that may affect incident duration are extracted from such information as incident type, time-of-day, and day-of-week. Those factors that have significant impacts on the actual incident duration are then identified.

There are numerous methods for dealing with a problem to identify the great factors from a set of continuous and categorical factors. A multi-way analysis of variance (ANOVA) is often used. This method is also adopted in this study. Based on the Multi-way ANOVA, the median of actual durations of historical incidents grouped by identified factors is obtained for constructing the look-up table. Note that the median, not the mean, of the actual durations of grouped incidents is selected because the frequency of actual durations of historical incidents on the study corridor show a χ^2 distribution, as shown in Figure 6.2, rather than a normal distribution.

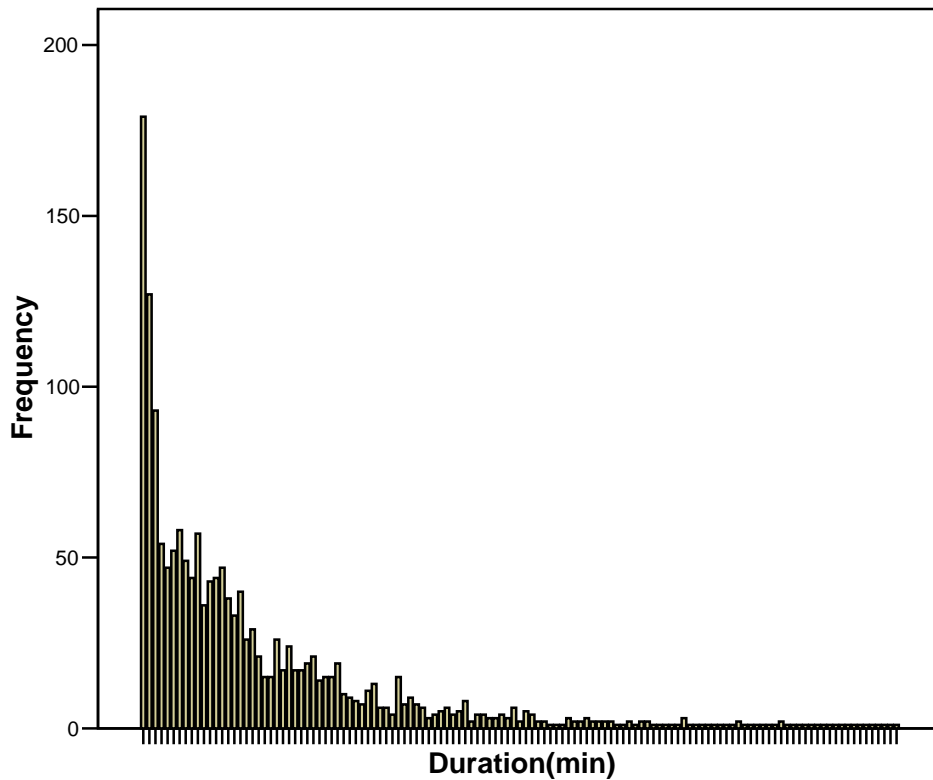


Figure 6.2 Incident Duration Frequency Distributions on the Study Corridor

6.3 CORRIDOR TRAVEL TIME ADJUSTMENT UNDER AN INCIDENT

6.3.1 Identification of Incident Impact on Traffic

Identification of incident impacts on traffic generally utilizes the variation of traffic measurements. Sudden changes of traffic measurements are usually observed when an incident has significant impacts on traffic. Based on this observation, numerous incident detection algorithms were developed to identify an incident from field traffic measurements such as flow rate and occupancy (Parkany and Xie 2005, Cheu et al. 1991). Inversely, an incident detection algorithm can be used to determine whether an incident has significant impacts on traffic when an incident already exists. In this study, the basic California algorithm, originally developed for incident detection, is applied for such purposes as identifying whether an existing incident has significant impacts on traffic in real time. The

basic principle of the algorithm is that an incident will create increased occupancy levels upstream of the incident and a decrease downstream (Payne et al. 1976).

According to the basic California algorithm, an incident occurring on link i within time interval t is determined to have significant impacts on traffic when traffic measurements from field detectors satisfy the conditions as follows.

The difference between upstream station occupancy ($o_{i,t}$) and downstream station occupancy ($o_{i+1,t}$) should be greater than the threshold value 0.08;

The ratio of the difference in the upstream and downstream occupancies to the upstream station occupancy $\frac{o_{i,t} - o_{i+1,t}}{o_{i,t}}$ should be greater than the threshold value 0.5; and

The ratio of the difference in the upstream and downstream occupancies to the upstream station occupancy $\frac{o_{i,t} - o_{i+1,t}}{o_{i+1,t}}$ should be greater than the threshold value 0.16.

The threshold values introduced in these conditions were calibrated from empirical data in the California algorithm. Only those incidents whose associated traffic flow measurements satisfy these conditions are considered as having significant impacts on traffic conditions.

6.3.2 Proposed Method for Corridor Travel Time Adjustment

6.3.2.1 Introduction

Assume that an incident occurs at location x on link i with upstream and downstream stations i and $i+1$ at time T_0 within time interval h . If the incident is identified as having significant impacts on traffic, the corridor is considered to be divided into three segments: 1) segment 1 from the origin of the corridor to the back of the queue; 2) segment 2 occupied by the physical queue; and 3) segment 3 from the bottleneck to the destination of the corridor, as illustrated in Figure 6.3. Correspondingly, the proposed method should predict the travel times of these three segments for each time interval.

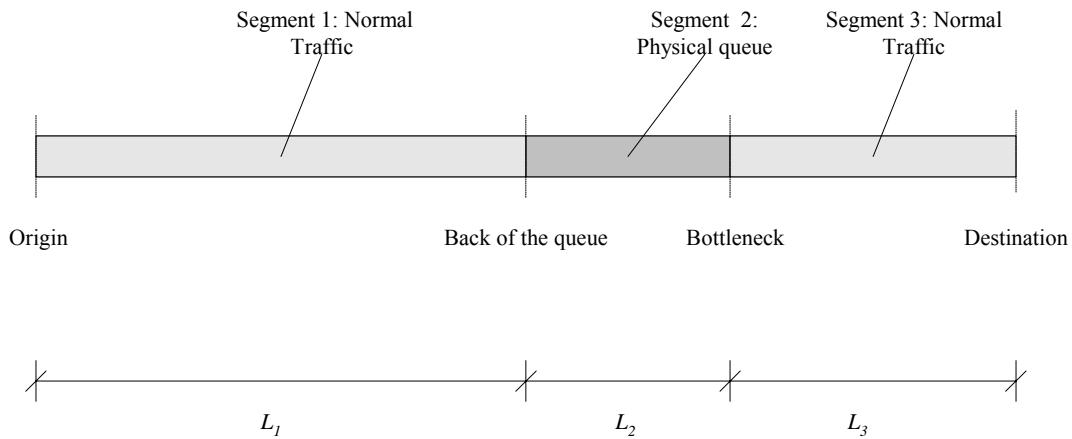


Figure 6.3 Corridor Travel Time Components under an Incident

Denote the lengths of the three segments as L_1 , L_2 , and L_3 , respectively. Both L_1 and L_2 vary with time and satisfy $L_1 \geq 0$ and $L_2 > 0$, but $L_1 + L_2$ is constant.

Although an incident might cause reduced flow on segment 3 such impact is difficult to measure, and therefore, the impacts of incidents on traffic on segment 3 are ignored. Under this situation, the temporal traffic characteristics at all involved vehicle detector stations on segments 1 and 3 can still be captured by the dynamic traffic flow predictor. Thus, the travel time of segments 1 and 3 can be predicted using the same method as presented in Chapter 5.

For segment 2, although the traffic on the links occupied by the physical queue can be predicted by the dynamic traffic flow predictor, the predicted flow rate and occupancy may greatly differ from the actual flow rate and occupancy when multiple-step-ahead prediction is involved.

6.3.2.2 Proposed Method

Travel time prediction considering incidents is conducted for each time interval between the incident start time T_0 and the time when the incident is cleared. The proposed method is then continued after the incident is cleared until traffic reaches normal flow. The basic idea is to estimate the shock wave speed based on the measured flow rate, occupancy up to current time interval, and extend the estimated shock wave speed to the next several time intervals such that the links occupied by the physical queue can be identified. Since the traffic in the physical queue is fairly stable over time, the queuing density estimated from

current traffic measurements is extended to next time intervals as well for travel time prediction for those links occupied the physical queue.

Again, we assume that an incident occurs on link i with upstream and downstream stations i and $i+1$ at time T_0 during time interval h . Let T_t denote the end time of the time period t , Λ'_c denote the estimated incident duration, and Λ_c denote the actual incident duration. For a driver leaving the origin of the corridor at time T_{t+1} , $t = h, h+1, \dots$, with traffic observations up to time step t available, the proposed method for corridor travel time adjustment is presented as follows based on the queuing analysis.

Estimation of the Physical Queue Length

Figure 6.4 presents the time-space diagram of the congestion caused by an incident on a corridor without considering the impacts of ramps. L_2 is the physical queue length. w is the shock wave speed. The physical queue length changes over time. The maximum physical queue occurs when the approaching flow is smaller than the service flow rate at the bottleneck. The shock wave speed changes over different time intervals as well, depending upon the approaching flow and queuing flow, which are unavailable due to the lack of traffic measurements on ramps. On the other hand, the existing shock wave speed can be approximated from the changes of physical queue length up to time T_t based on the traffic measurements at vehicle detector stations along the corridor. Conversely, such a shock wave speed can be used to estimate physical queue length at next time intervals as described follows.

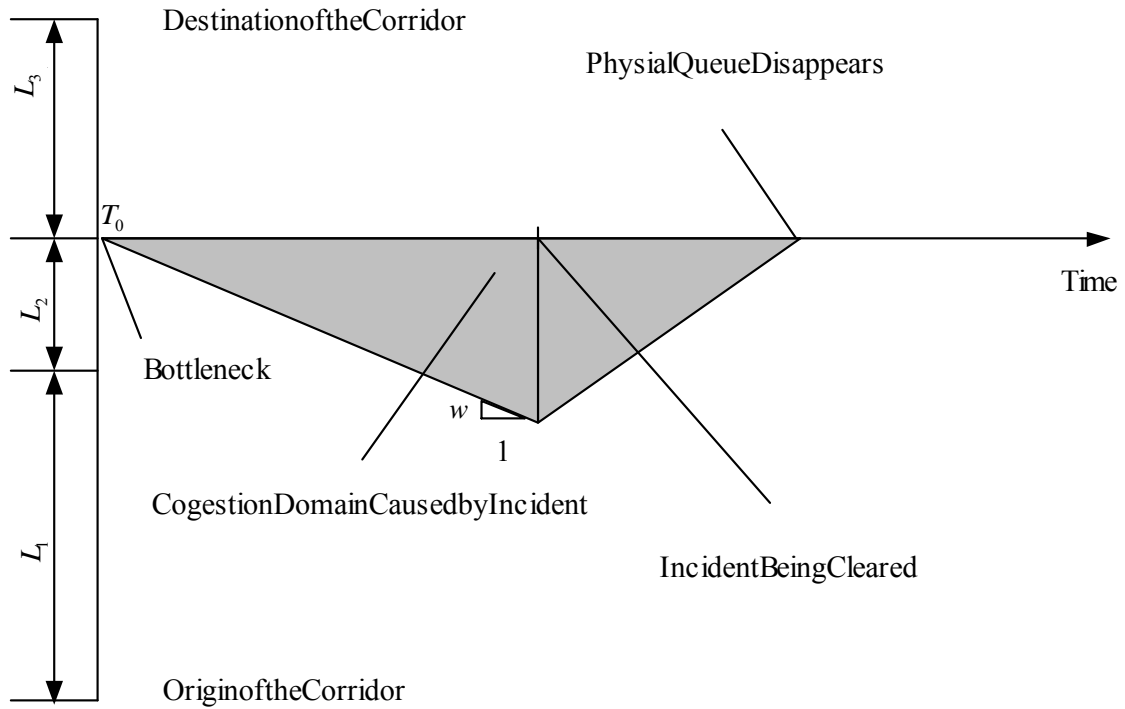


Figure 6.4 Time-Space Diagram of Congestion Caused by an Incident

(1) Before the incident is actually cleared

In this case, the actual incident duration Λ_c is unknown, and the incident duration is predicted as Λ'_c from the established incident duration look-up table. At time T_t , $T_t < T_0 + \Lambda'_c$, for a driver leaving the origin of the corridor, there are two possibilities when the driver arrives at the back of the physical queue. The first possibility is that the driver arrives at the back of the physical queue before the incident is cleared. Or, the driver may arrive at the back of the physical queue after the incident is cleared.

For the first scenario, the shock wave speed at the back of the physical queue can be estimated as the rate of the physical queue length accumulated over the period from T_0 to T_t in the form of

as

$$w_t = \frac{L_{2,t}}{T_t - T_0} \tag{6.1}$$

where,

w_t : the shock wave speed at time T_t ;

$L_{2,t}$: the physical queue length at time T_t ;

$T_t - T_0$: the time taken to form the physical queue.

The physical queue length $L_{2,t}$ at time T_t is estimated as the total length of the upstream links that are occupied by the physical queue. An upstream link is assumed to be occupied by the physical queue if the average measured flow rate and estimated density satisfy the following conditions: 1) the flow rate is smaller than 2400 vehicle/hour/lane and 2) the density is greater than 45 vehicle/minute/lane (TRB 2000). Given the measured flow rate, occupancy, and estimated mean effective vehicle length, the average link density and flow rate are estimated in the form of

$$k_t = \frac{k_{1,t} + k_{2,t}}{2} \quad (6.2)$$

$$f_t = \frac{f_{1,t} + f_{2,t}}{2} \quad (6.3)$$

where,

k_t : the estimated link density at time T_t ;

$k_{1,t}$: the estimated density from traffic measurements at the upstream station of the link at time T_t in the form of $k_{1,t} = \frac{o_{1,t} \cdot 5280}{g_{1,t}}$, where $o_{1,t}$ and $g_{1,t}$ are the measured occupancy and estimated mean effective vehicle length, respectively;

$k_{2,t}$: the estimated density from traffic measurements at the downstream station of the link at time T_t ;

f_t : the flow rate by averaging the measured flow rates at upstream and downstream stations at time T_t ;

$f_{1,t}$: the measured flow rate at the upstream station of the link;

$f_{2,t}$: the measured flow rate at the downstream station of the link.

The physical queue length at any time T_1 satisfying $T_t < T_1 \leq T_0 + \Lambda'_c$ is estimated by extending the estimated shock wave speed w_t to time T_1 in the form of

$$L_2(T_1) = w_t \cdot (T_1 - T_0) \quad (6.4)$$

For the second scenario, the driver arrives at the back of the queue at the time that the incident is cleared. Under this situation, the physical queue length at any time T_1 , $T_t < T_0 + \Lambda'_c < T_1$, cannot be estimated by equation (6.4) because the shock wave speed changes due to the increased capacity at the bottleneck. However, it can be predicted as the physical queue length at time $T_0 + \Lambda'_c$ plus the estimated distance the shock wave will travel from $T_0 + \Lambda'_c$ to T_1 in the form of

$$L_2(T_1) = L_2(T_0 + \Lambda'_c) + w'_t \cdot [T_1 - (T_0 + \Lambda'_c)] \quad (6.5)$$

where $L_2(T_0 + \Lambda'_c)$ is the predicted physical queue length at time $T_0 + \Lambda'_c$ when the incident is predicted as being cleared, w'_t is the predicted shock wave speed at the physical queue between $T_0 + \Lambda'_c$ and T_1 from current time T_t . The value of $L_2(T_1)$ might be zero at some time, indicating that at time T_1 there is no physical queue before him.

Assuming that the approaching flow rate and its corresponding density, and the density in the physical queue do not change over time, the equation to estimate w'_t at time T_t , $T_t < T_0 + \Lambda'_c < T_1$, is given as

$$w'_t = \frac{F - f_A}{k_Q - k_A} \quad (6.6)$$

where,

w'_t : the predicted shock wave speed during $T_0 + \Lambda'_c$ and T_1 at time T_t ;

F : the freeway capacity, set as 2400 vehicle/hour/lane in this study;

f_A : the average approaching flow at time T_t ;

k_Q : the density in the physical queue at time T_t , which can be estimated as the length-weighted average of the link densities in the form of equation (6.2);

k_A : the density of the approaching flow at time T_t .

Knowing that w_t at the back of the physical queue at time T_t can also be represented as $w_t = \frac{f_d - f_A}{k_Q - k_A}$ where f_d is the bottleneck capacity at time T_t , w_t is rewritten as

$$w_t' = \frac{F - f_A}{f_d - f_A} \cdot w_t \quad (6.7)$$

Equation (6.7) indicates that the approaching flow rate f_A and the bottleneck capacity f_d are required to be estimated. In this study, f_d is estimated as the mean of the measured flow rate from T_0 to T_t . While, f_A is estimated from equation

$$f_A \cdot (T_t - T_0) - f_d \cdot (T_t - T_0) = k_Q \cdot L_2(T_t) \cdot c \quad (6.8)$$

where c refers to the number of lanes, $f_A \cdot (T_t - T_0)$ refers to the total number of vehicles entering the physical queue from T_0 to T_t , and $f_d \cdot (T_t - T_0)$ refers to the number of vehicles leaving the bottleneck from time T_0 to T_t . $f_A \cdot (T_t - T_0) - f_d \cdot (T_t - T_0)$ is the number of vehicles in the physical queue, which also can be represented as $k_Q \cdot L_2(T_t) \cdot c$.

(2) After the incident is actually cleared

Once the incident has been cleared at current time T_t , the shock wave speed can be estimated as the physical queue length changes over one time interval in the form of

$$w_t = \frac{L_2(T_{t-1}) - L_2(T_t)}{\Delta T} \quad (6.9)$$

This shock wave speed can be extend to estimate the physical queue length at any time T_1 in next time intervals in the form of

$$L_2(T_1) = L_2(T_t) + w_t \cdot (T_1 - T_t) \quad (6.10)$$

Segment Travel Time Estimation

Given the method to predict the physical queue length at any time in the future, the method to adjust the travel times for segments 1, 2 and 3 are given as follows.

(1) Travel Time Prediction for Segment 1

Let tt_{s1} denote the travel time of segment 1. The basic idea to predict tt_{s1} is to use the relationship between L_1 and L_2 as

$$L_1(T_{t+1} + tt_{s1}) + L_2(T_{t+1} + tt_{s1}) = L_{12} \quad (6.11)$$

where $L_1(T_{t+1} + tt_{s1})$ is the length of segment 1 at time $T_{t+1} + tt_{s1}$, $L_2(T_{t+1} + tt_{s1})$ is the physical queue length at time $T_{t+1} + tt_{s1}$, L_{12} is the constant representing the distance from the corridor origin to the incident location.

Both $L_1(T_{t+1} + tt_{s1})$ and $L_2(T_{t+1} + tt_{s1})$ are functions of tt_{s1} . The approximate solution of tt_{s1} for equation (6.11) is obtained through the steps listed below.

Step1: initialize link index $a = 1$.

Step 2: predict travel times tt_a of link a and calculate the travel time TT_a from the origin of the corridor to station $a + 1$ using the travel time prediction method without considering incident impacts.

Step 3: predict the physical queue length $L_2(T_{t+1} + TT_a)$ at time $T_{t+1} + TT_a$ using equation (6.1) through equation (6.10).

Step 4: if $L_2(T_{t+1} + TT_a) + l_a \geq L_{12}$ indicating the driver arrives at the back of the physical queue, then go to step 5. Otherwise, $a = a + 1$ and go to step 2.

Step 5: set $tt_{s1} = TT_a$.

(2) Travel Time Prediction for Segment 2

Let tt_{s2} denote the travel time of segment 2. tt_{s2} is the total travel times of the links occupied by the physical. For a specific link, the travel time can be estimated as

$$tt = \frac{l}{V} = \frac{l \cdot k_t}{f_t} \quad (6.12)$$

where,

l : the link length;

V : the average speed;

k_t : the link density at time T_t ;

f_t : the flow rate by averaging the measure flow rates at upstream and downstream stations of the link at time T_t .

(3) Travel Time Prediction for Segment 3

Given the predicted tt_{s_1} and tt_{s_2} , it is known that the driver passes through the bottleneck at time $T_{i+1} + tt_{s_1} + tt_{s_2}$. Let tt_{s_3} denote the travel time of segment 3. tt_{s_3} can be obtained using the same methods as presented in Chapter 5, consistent with the arrival time at each upstream station.

(4) Corridor Travel Time

The final corridor travel times affected by the incident are adjusted to $tt_{s_1} + tt_{s_2} + tt_{s_3}$.

6.4 PERFORMANCE ANALYSIS

6.4.1 Structure of Corridor Travel Time Prediction Adjustment System

The structure of the proposed corridor travel time prediction adjustment system is presented in Figure 6.5.

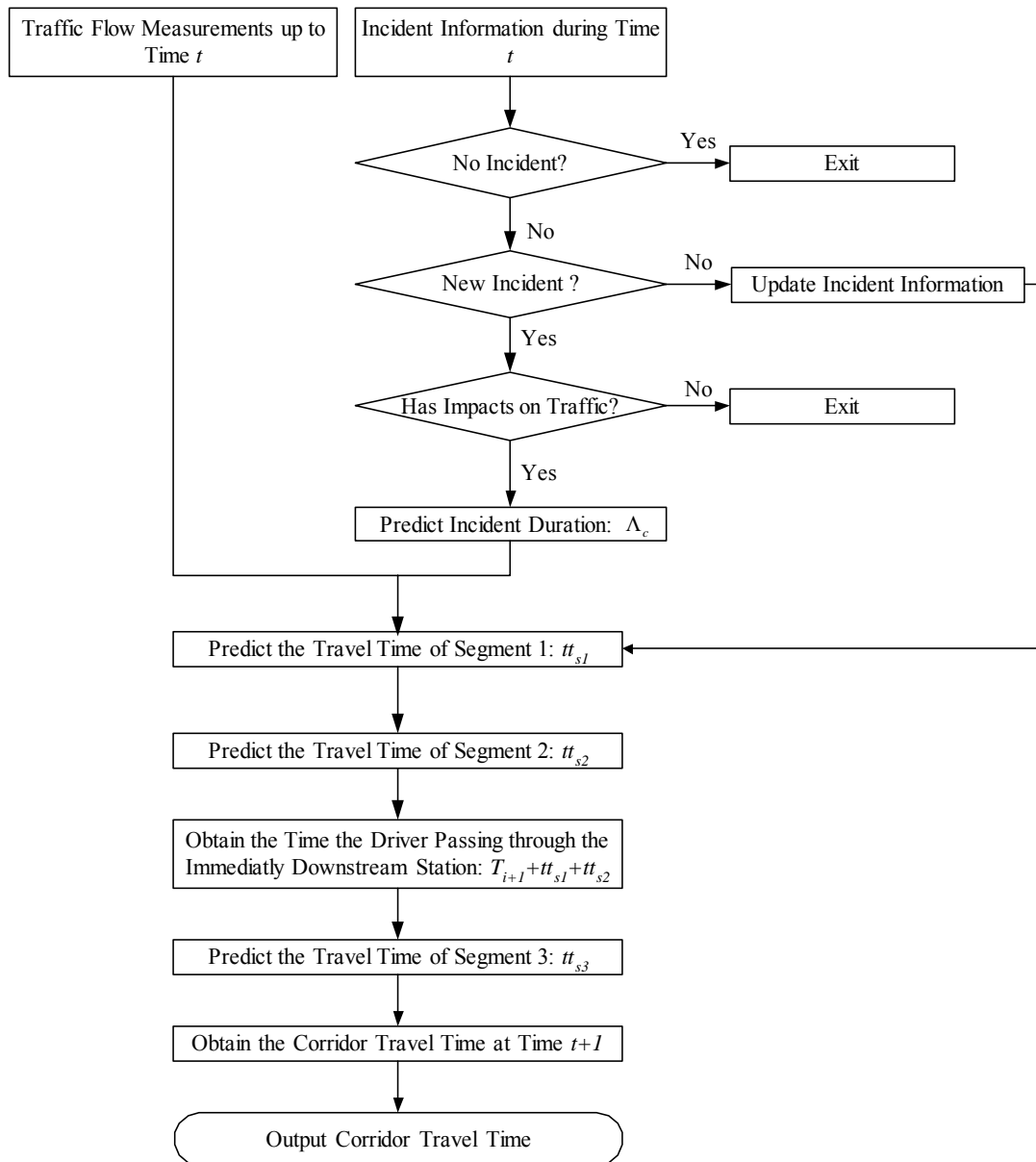


Figure 6.5 Structure of Online Corridor Travel Time Prediction Adjustment System

6.4.2 Performance Analysis

6.4.2.1 Incident Duration Estimation

In order to establish a look-up table to predict the corridor travel time under incident situations, 1623 historical incidents on the study corridor are used. The factors that might affect the actual incident duration drawn from the available incident information include time-of-day, day-of-week, and incident types. In order to evaluate the impacts of these factors on incident duration, these factors and the actual incident durations are fitted in a general univariate linear model, in which a multi-way ANOVA is provided for the identification of factors that have impacts on incident duration.

Table 6.1 presents the multi-way ANOVA results where the potential factors are listed in the column “Factor”. The factors that have impacts on the actual duration are identified as day-of-week and incident type, while time-of-day does not have a significant impact at a significance level 0.05.

Table 6.1 Multi-Way ANOVA Results of Incident Duration

Factor	Sum of Squares	Degree of Freedom	Mean Square	F-Test	P-Value.
Day-of-week	5113.053	1	5113.053	6.83	.009
Time-of-day	1368.918	1	1368.918	1.83	.177
Incident type	20982.688	6	3497.115	4.67	.000

Based on identified factors as day-of-week and incident type, the historic incidents are classified into different groups by day-of-week and incident type. The median and standard deviation of the actual durations of each group are obtained as presented in Table 6.2.

Table 6.2 Look up Table of Incident Duration

Day-of-Week	Incident Types	Incident Duration (min)	Standard Deviation of Incident Duration (min)
Weekday	Disabled Vehicle	10.0	18.2
	Traffic Hazard: Vehicle	12.5	19.2
	Hit and Run	18.0	14.2
	Traffic Collision: Ambulance Responding	32.0	31.8
	Traffic Collision: Property Damage	14.0	25.5
	Traffic Collision: No Detail	10.0	26.7
	Vehicle Fire	11.0	15.7
Weekend	Disabled Vehicle	11.0	15.6
	Traffic Hazard: Vehicle	15.0	18.9
	Hit and Run	44.0	106.0
	Traffic Collision: Ambulance Responding	23.0	39.4
	Traffic Collision: Property Damage	15.5	51.1
	Traffic Collision: No Detail	17.5	27.0
	Vehicle Fire	4.5	3.5

6.4.2.2 Corridor Travel Time Adjustment

From June 1, 2006 to June 30, 2006 on the study corridor, there are a total of 52 incidents. Among these incidents, 6 incidents are identified that have great impacts on the traffic, as described in Table 6.3. Among the 6 incidents identified as having great impacts on traffic, the actual duration of incident 4 is too short to obtain the capacity of the bottleneck from the measured flow rate at the immediately downstream station of the incident. Thus, this incident is not considered in the adjustment of predicted corridor travel times. The predicted corridor travel times within the influence time of the remaining 5 incidents are adjusted. The adjustment results for those time intervals within the 5 incident influence time are presented in Appendix B.

Table 6.3 Incidents Identified as Having Great Impacts on the Traffic

Incident ID	Start Date	Start Time	Day of Week	Duration (min)	Abs Milepost	Incident Type
1	2006-06-01	17:01:00	Thursday	27	16.43	Traffic Collision-Property Damage
2	2006-06-07	17:46:00	Wednesday	9	16.63	Traffic Collision-No Details
3	2006-06-09	15:58:00	Friday	42	18.29	Traffic Collision-Ambulance Response
4	2006-06-18	19:45:00	Sunday	2	12.22	Traffic Collision-Property Damage
5	2006-06-19	17:19:00	Monday	10	16.10	Traffic Collision- No Details
6	2006-06-30	21:32:00	Saturday	90	13.25	Traffic Collision-No Details

Incident 1 occurred at 17:01 on June 1, 2006 at ABS milepost 16.43. The latest time to obtain the estimate of bottleneck capacity is 17:10. Thus, the predicted corridor travel times are considered to be adjusted starting from 17:15. Figure 6.6 presents the results of the corridor travel times, in which the actual, predicted without and with considering incident impact. It is observed that the method without considering the incident impacts on traffic generally underestimates the actual corridor travel times. Comparatively, the adjustment method generally improves the corridor travel times although it still slightly underestimates the actual corridor travel times from 17:45 to 18:50.

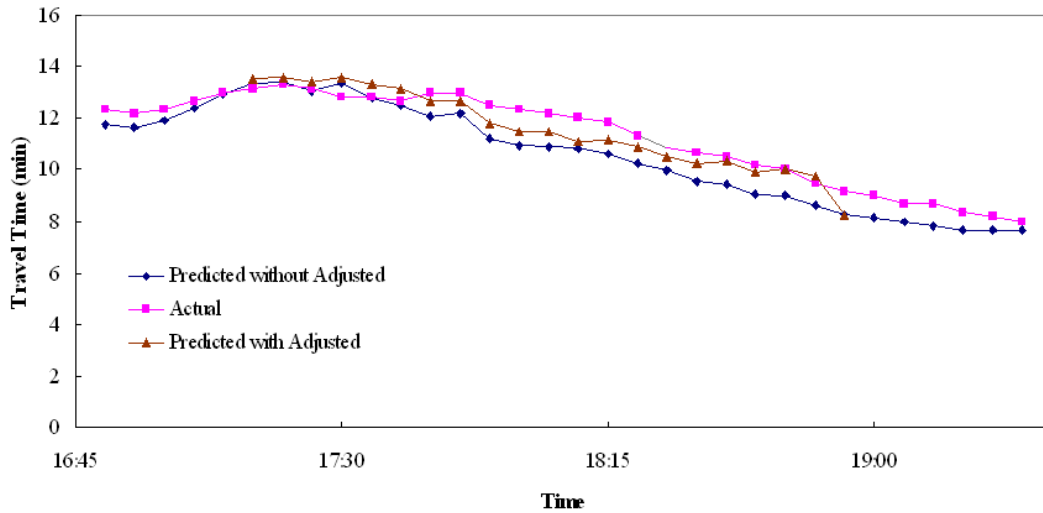


Figure 6.6 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 1

Illustration of the adjustment method for incident 2 is plotted in Figure 6.7. Incident 2 happened at 17:46 on June 7, 2006. This leads to the latest time to obtain the estimate of bottleneck capacity at 17:55. Thus, the predicted corridor travel times are considered to be adjusted starting from 18:00. Similar to incident 1, the predicted corridor travel times are underestimated within the incident influence time before they are adjusted. When compared to incident 1, the occurrence of incident 2 affects the traffic greatly when looking at the plots from 17:45 to 17:55, where the big difference between the actual and predicted corridor travel times exist.

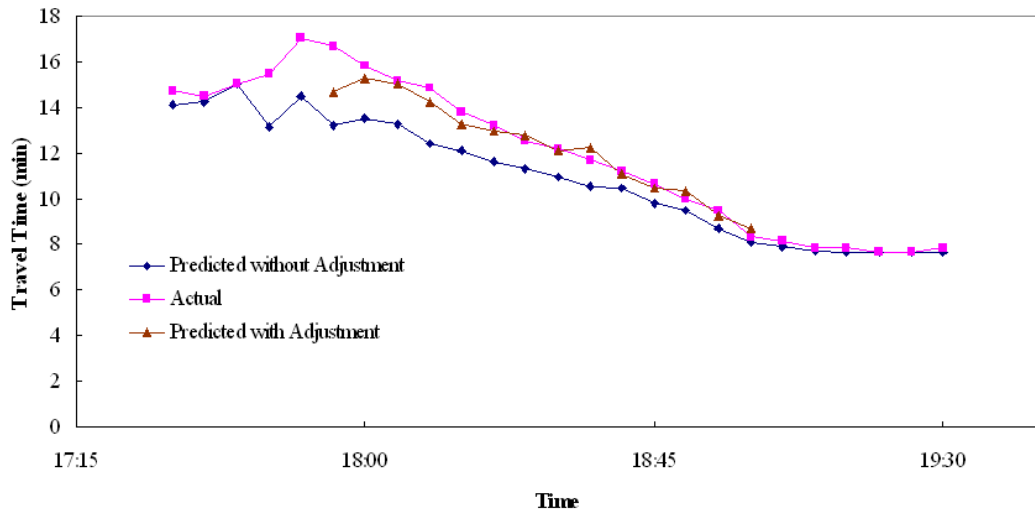


Figure 6.7 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 2

Figure 6.8 through Figure 6.10 illustrates the corridor travel time adjustment of the other three incidents. Similar to incident 1 and incident 2, all corridor travel times are underestimated before the corridor travel times are adjusted. Note, although the actual duration of incident 6 is 90 minutes, its impacts on traffic were not as long as its actual duration. This means that before the incident was cleared at 23:02, the traffic conditions had been recovered.

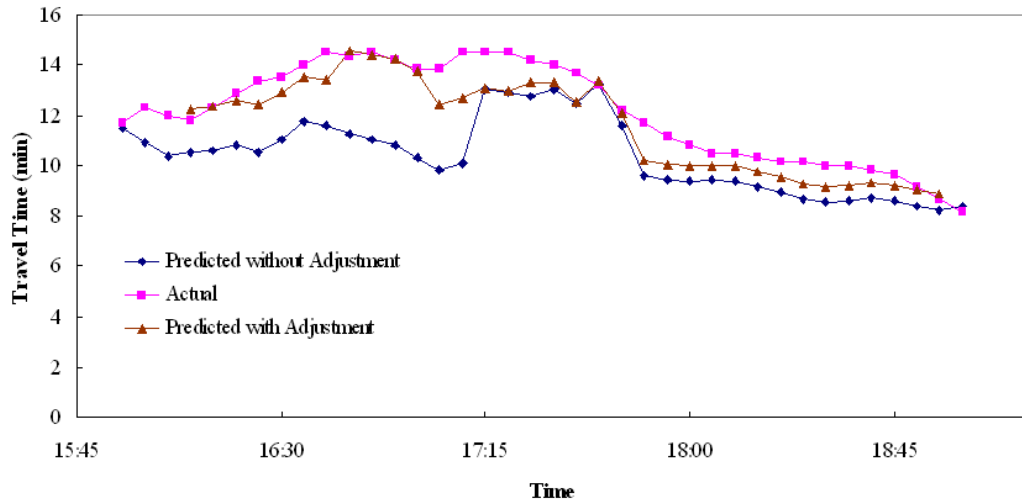


Figure 6.8 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 3

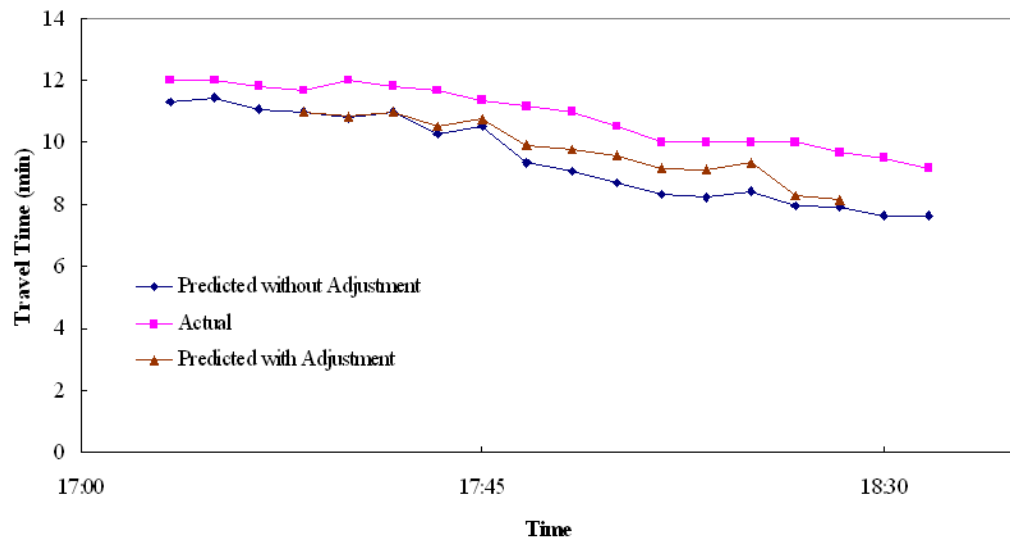


Figure 6.9 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 5

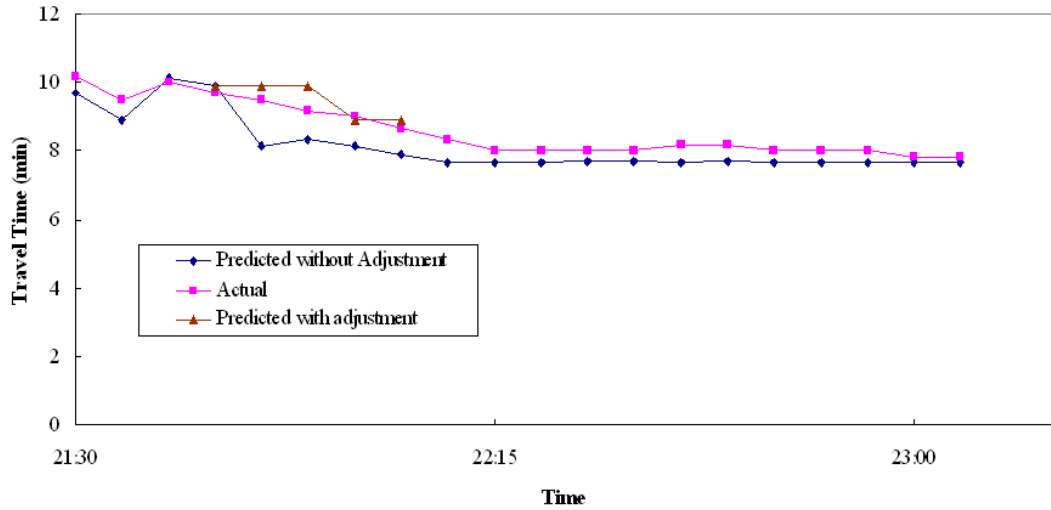


Figure 6.10 Actual, Predicted without and with Adjustment of Corridor Travel Times for Incident 6

6.5 SENSITIVITY ANALYSIS

Sensitivity analysis helps in quantifying the uncertainty associated with estimated model parameters. Since the proposed corridor travel time prediction model predicts corridor travel time by adding all estimates of link travel times, the sensitivity analysis in this study is performed on the link travel time function derived from equation (5.7) in the form of

$$tt = \frac{l}{2} \left(\frac{o_1}{g_1 \cdot f_1} + \frac{o_2}{g_2 \cdot f_2} \right) \quad (6.13)$$

where,

l : the constant of link length;

o_1 : the occupancy at upstream vehicle station;

o_2 : the occupancy at downstream vehicle station;

g_1 : the mean effective vehicle length at upstream vehicle station;

g_2 : the mean effective vehicle length at downstream vehicle station;

f_1 : the flow rate at upstream vehicle station;

f_2 : the flow rate at downstream vehicle station.

As seen in equation (6.13), there are three pairs of parameters (i.e., flow rate, occupancy, and mean effective vehicle length) in the basic system, and they are selected for sensitivity analysis. Testing of the sensitivity analysis is performed on a 1.20-mile long link between VDS 400770 and 401243.

6.5.1 Flow Rate

Based on equation (6.13), the sensitivity analysis is performed for the upstream flow rate using the measured speed (i.e., V_1 and V_2), the measured occupancy (i.e., o_1 and o_2), and the measured downstream flow rate (i.e. f_2), and the predicted upstream flow rate (i.e. \hat{f}_1) to calculate the link travel time. The calculated link travel time is then compared to the actual link travel time.

The sensitivity analysis strategy follows three stages: 1) the different traffic conditions are defined by specifying different ranges of occupancy values; 2) the flow rate prediction accuracy is set into different groups using the performance measure of MAPE under the different traffic conditions; and 3) the MAPE values of link travel time prediction results are collected for the different groups of flow rate prediction accuracy under the different traffic conditions.

The sensitivity analysis results to upstream flow rate under different traffic conditions are presented in Table 6.4. It can be observed that the link travel time performance is fairly stable across different traffic conditions when occupancy is smaller than 0.12. Comparatively, the performance becomes a little worse when the occupancy is greater than 0.12. This indicates that under uncongested traffic conditions, the link travel time is less sensitive to flow rate than under congested conditions. Furthermore, it is also observed that the larger the MAPE of flow rate, the worse the link travel time prediction accuracy.

Table 6.4 Sensitivity Analysis to Upstream Flow Rate

Traffic Condition by Measured Occupancy (%)	Flow Rate Prediction Accuracy by MAPE (%)	MAPE of Link Travel Time (%)
0~5	0~2	0.057
	2~4	0.174
	4~6	0.306
	6~8	0.352
	8-10	0.443
	10-12	0.592
	12-15	0.693
	15-20	0.981
	>20	3.406
5~8	0~2	0.042
	2~4	0.142
	4~6	0.235
	6~8	0.287
	8-10	0.428
	10-12	0.514
	12-15	0.608
	15-20	0.727
	>20	3.68
8~12	0~2	0.053
	2~4	0.153
	4~6	0.197
	6~8	0.318
	8-10	0.421
	10-12	0.367
	12-15	0.448
	15-20	0.509
	>20	2.032
>12	0~2	0.065
	2~4	0.285
	4~6	0.323
	6~8	0.488
	8-10	0.666
	10-12	1.156
	12-15	1.312
	15-20	1.364
	>20	4.231

6.5.2 Occupancy

Using the same sensitivity strategy as flow rate, sensitivity analysis results to upstream occupancy are presented in Table 6.5. It can be observed that the link travel time performance is very sensitive to the prediction accuracy of the occupancy. However, with the same range of occupancy prediction accuracy, the link travel time prediction performance is fairly stable to the occupancy.

Table 6.5 Sensitivity Analysis to Upstream Occupancy

Traffic Condition by Measured Occupancy (%)	Occupancy Prediction Accuracy by MAPE (%)	MAPE of Link Travel Time (%)
0~5	0~2	0.383
	2~4	1.242
	4~6	1.653
	6~8	3.082
	8-10	3.837
	10-12	4.756
	12-15	5.821
	15-20	7.556
	>20	12.573
5~8	0~2	0.441
	2~4	1.302
	4~6	2.112
	6~8	3.059
	8-10	4.037
	10-12	4.821
	12-15	5.657
	15-20	7.689
	>20	13.562
8~12	0~2	0.468
	2~4	1.327
	4~6	1.983
	6~8	3.142
	8-10	3.976
	10-12	4.803
	12-15	5.985
	15-20	7.547
	>20	13.214
>12	0~2	0.438
	2~4	1.232
	4~6	2.218
	6~8	2.978
	8-10	3.856
	10-12	4.645
	12-15	6.043
	15-20	8.431
	>20	15.346

6.5.3 Mean Effective Vehicle Length

The Results of the sensitivity analysis to upstream mean effective vehicle length under different traffic conditions are presented in Table 6-6. The results show that the link travel time performance is very sensitive to the mean effective vehicle length. A worse prediction of mean effective vehicle length leads to worse link travel time prediction accuracy. When compared to the flow rate and occupancy, the sensitivity analysis shows that the mean effective vehicle length and occupancy are much more sensitive to travel time prediction. Although the flow rate may also lead to bad prediction of travel time, its effects are not as significant as the mean effective vehicle length.

Table 6.6 Sensitivity of MAPE to Upstream MEVL

Traffic Condition by Measured Occupancy (%)	MEVL Prediction Accuracy by MAPE (%)	MAPE of Link Travel Time (%)
0~5	0~2	0.683
	2~4	1.134
	4~6	2.732
	6~8	3.894
	8-10	4.899
	10-12	5.642
	12-15	7.912
	15-20	8.913
	>20	—
5~8	0~2	0.698
	2~4	1.438
	4~6	2.642
	6~8	3.574
	8-10	4.568
	10-12	5.463
	12-15	7.124
	15-20	8.453
	>20	—
8~12	0~2	0.842
	2~4	1.654
	4~6	2.492
	6~8	3.536
	8-10	4.442
	10-12	5.532
	12-15	6.605
	15-20	7.786
	>20	14.569
>12	0~2	3.047
	2~4	3.320
	4~6	2.592
	6~8	5.684
	8-10	5.675
	10-12	5.642
	12-15	6.224
	15-20	9.862
	>20	17.466

6.6 COMPARATIVE EVALUATION

6.6.1 Choice of Algorithms for Comparison

As described in Chapter 2, Section 2.3.2, generally two groups of methods can be used for travel time prediction.

The first group of methods predicts travel time by predicting the space-mean speed. Assumptions of free-flow-speed and mean effective vehicle length are often made by this category of methods. Furthermore, these methods don't predict travel time but report travel times for the time the traffic flow data was collected.

The second group of methods estimate travel time from the cumulative traffic counts at both upstream and downstream stations, thereby obtaining the density of the freeway links. However, this method requires the estimation of an initial number of vehicles on the freeway links. Furthermore, system errors exist in the collection of traffic flow data from field detectors, and are very difficult to adjust. Figure 6.11 illustrates a typical phenomenon of cumulative traffic counts caused by system errors between the upstream (VDS 401079) and downstream (VDS 401239) stations using the measured flow rates on 06/04/2006. It is observed that the daily difference in cumulative traffic counts between two vehicle detector stations is about 14,000 vehicles, which is too large to be held in the very short freeway link with a length of 0.71 miles. Although adjustments by introducing some feedback control mechanism can be made to make up under-or over-measured traffic counts at the downstream station, the full system error cannot be completely solved. The worst situation is that, at certain time steps, the cumulative traffic counts at upstream and downstream stations cannot converge and, thus, lose the ability to estimate the link density. Therefore, the second category of travel time prediction methods is not considered in this study.

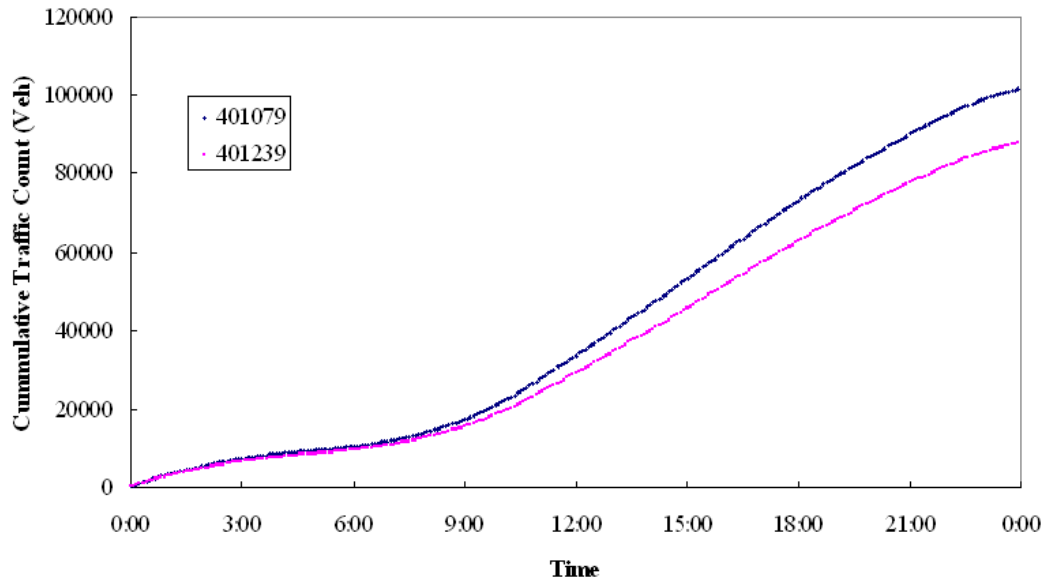


Figure 6.11 Cumulative Traffic Count on 06/04/2006 at VDSs 401079 and 401239

Focusing on the first category of corridor travel time prediction methods, two methods are selected as alternative methods for comparative evaluation.

Method 1 reports the corridor travel time for the time the traffic flow data is collected. Therefore, the effects of traffic progression are not considered. This method reports the link travel times by estimating the space-mean speed using equation (2.7), and obtains the corridor travel time by adding all link travel times. This method was originally used by the Illinois Department of Transportation (IDOT) to estimate current travel time on major freeways in the Chicago area using data aggregated at 5-minute increments from single loop detectors. Later research by Dailey (1997) used 25.63 ft as the mean effective vehicle length based on six-day period estimates to estimate the space-mean speed and thus for travel time estimation.

Method 2 predicts the corridor travel times, but does not consider the traffic progression along the corridor. The predictor for dynamic traffic flow prediction is used for one-step-ahead prediction of flow rate and occupancy.

6.6.2 Implementation of Algorithms

Both alternative methods are tested on the same platform using the same datasets to ensure conformity and congruity among these two methods and the proposed method. As an addendum to the descriptions of the algorithms provided in the previous section, some more implementation details are provided for ease of reference.

Both methods designed for comparative evaluation estimate space-mean speed from flow rate and occupancy in the form of $speed = \frac{MEVL \cdot flow}{occupancy}$. For method 1, mean effective length is directly estimated from current flow rate and occupancy using the presented MEVL estimation method. In the case of the second method, 5-minute-ahead prediction of flow rate and occupancy are performed using the proposed SARIMA model, and thus the corresponding mean effective vehicle length is estimated using the proposed MEVL estimation. Based on this, one-step-ahead prediction of all link travel times can be obtained. For both methods, the corridor travel time is obtained by adding all link travel times.

6.6.3 Performance

The overall performance of the three methods is presented in Table 6.7. Method 1 has the worst performance for prediction accuracy. When compared to method 1, method 2 improves prediction accuracy in terms of MAPE although it merely uses one-step-ahead predicted flow rate and occupancy and corresponding mean effective vehicle length. The proposed method in this study produces the best performance with the minimum values of MAE, MAPE, and RMSE. The performance results prove that both flow rate and occupancy should be predicted for the corridor travel time prediction instead of the reported corridor travel time at the time traffic flow was collected. It also proves that traffic progression should be considered in the corridor travel time prediction. A longer corridor may cause worse performance in both methods 1 and 2.

Table 6.7 Results of Comparative analysis

Method	MAE(min)	MAPE(%)	RMSE(min)
Method 1	1.14	10.24	1.31
Method 2	0.78	7.16	0.82
Proposed method	0.49	5.34	0.75

6.7 DESIGN OF THE FULL ON-LINE CORRIDOR TRAVEL TIME PREDICTION SYSTEM

In this section, a full on-line corridor travel time prediction system is presented to facilitate the reproducibility of the proposed corridor travel time prediction model. In this system, all methods are integrated including operations data screening, flow rate and occupancy prediction, mean effective vehicle length estimation, and corridor travel time prediction and adjustment.

The full on-line corridor travel time prediction system has a modular architecture. The modules are identified as follows: 1) parameter loading; 2) data loading; 3) operation data screening; 4) operation data transformation 5) corridor travel time prediction; 6) predicted corridor travel time adjustment.

The module of parameter loading is used to load the parameters that are predetermined before real-time implementation of the proposed corridor travel time prediction model for different tasks. The parameters required are list below.

- Data screening criteria. These parameters are used for the operations data screening;
- Station-specific data transformation parameters λ for both flow rate and occupancy;
- Station-specific simple exponential smoothing parameters α for both flow rate and occupancy;
- Initialization of parameter values for the adaptive Kalman filtering;
- Station-specific free-flow-speed;
- Duration of incident by day-of-week and incident type.

The data loading module obtains the station-specific traffic flow data when a new data record becomes available. Real-time incident information is also loaded.

The data screening module is used to determine whether the loaded traffic flow data is corrected using the data screening criteria. The module of data transformation is used to transform the loaded traffic data to input the corridor travel time prediction and adjustment system. The module of corridor travel time prediction integrates all the methods of multi-step ahead prediction of flow rate and occupancy, mean effective vehicle length estimation, and the corridor travel time prediction without considering incident. The predicted corridor travel time adjustment module integrates the determination of incident impacts on traffic, incident recovery time prediction, and the prediction of corridor travel time under an incident. This module only works when there is existing incident information at the current time. This module is not triggered for a new incident just loaded in the system.

The full system works in a time-updated style, but always predicts the corridor travel time 5 minutes in advance. The whole structure of the full on-line corridor travel time prediction system is presented in Figure 6.12.

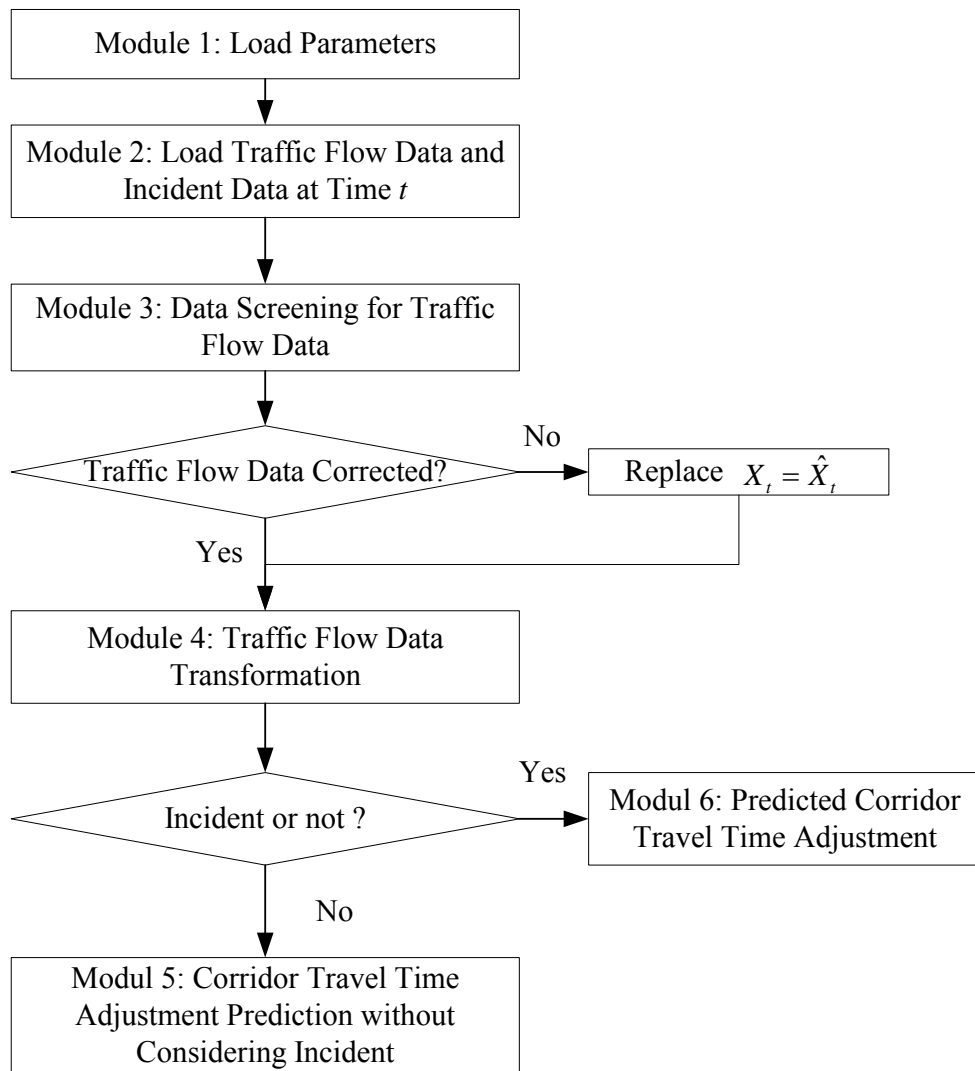


Figure 6.12 Structure of the Full Online Corridor Travel Time Prediction System

6.8 SUMMARY

Due to the fact that the corridor travel times under an incident which significantly affects traffic is not predicted as accurate as those incident-free traffic conditions, this chapter focuses on the development of a methodology to adjust the predicted corridor travel time under an incident based on queuing analysis. To test the proposed method, an incident occurring on the freeway is considered to determine whether it has a significant impact on traffic. If it does affect the traffic, the proposed method is applied. Otherwise, the corridor travel times under such an incident are not adjusted.

Testing of the proposed method is performed on the study corridor starting from 06/01/2006 to 06/30/2006. Performance analysis on the testing results indicates that the proposed corridor travel time adjustment method can improve the corridor travel time prediction accuracy, although it adjusts the corridor travel times from the next step of the occurrence of an incident.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 CONCLUSIONS

Travel time data is useful for a wide range of transportation analyses including congestion management, transportation planning, and traveler information. Direct collection of travel time data through the techniques of test vehicles, ITS probe vehicles, and license plate matching is time-consuming and limited to wide applications. With the implementation of ITS deployments, travel time estimation and prediction has attracted many concerns about the continuously generated traffic measurements from such devices. Despite limited success under light traffic conditions, traditional corridor travel time prediction methods have suffered various drawbacks. There has not been a reliable methodology for travel time prediction based on data generated by single loop detectors.

In this dissertation, a methodology is developed to provide more accurate short-term corridor travel time information based on traffic flow data from single loop detectors and incident information from traffic monitoring systems. The proposed method uses relationships among traffic variables such as flow rate, occupancy, speed, and density, and underlying traffic features over time to calculate link or corridor travel time. As a result, the accuracy of the final results depends upon the relationships used for travel time calculation and the investigation of traffic characteristics over time.

Tests of the corridor travel time prediction methodology are conducted on a study corridor, and some conclusions can be drawn from the test results. It is concluded that traffic flow can be quite accurately predicted with a weekly SARIMA model. The MAPE values of flow rate and occupancy prediction range from 5.89% to 7.83% and 5.90% to 7.93%, respectively, across all vehicle detector stations. It is also concluded that the proposed dynamic traffic flow predictor can be used for multiple-step-ahead prediction of flow rate and occupancy. Test results show that multiple-step-ahead prediction performance for flow rate and occupancy degrades little with

increased number of steps, which is important for the corridor travel time prediction considering the effects of traffic progression along the corridor.

Testing a corridor travel time prediction method without considering incident shows that multiple-step-ahead prediction of flow rate and occupancy under incident situations leads to a large difference between the predicted and actual travel times. This indicates that the sudden changes in traffic measurements under an incident as compared to incident-free conditions cannot be fully captured by the dynamic traffic flow predictor, particularly under congested traffic conditions.

Sensitivity analysis concludes that the occupancy and mean effective vehicle length have much more significant impacts on the final corridor travel time performance than does flow rate. This is a potentially fruitful area for future research, placing more emphasis on occupancy prediction and the development of more robust methods for mean effective vehicle length estimation.

In summary, the proposed method is able to capture the real-time characteristics of traffic and provides more accurate travel time estimates when compared to alternative methods. Particularly, adjustment of the predicted corridor travel time enhances the prediction accuracy under an incident.

7.2 RECOMMENDATIONS FOR FUTURE RESEARCH

Though this dissertation provides several contributions to transportation literature in the area of corridor travel time prediction, there are several areas in which future work is needed.

This dissertation uses aggregated 5-minute traffic flow data as well as incident data obtained along a 11.096 mile long California corridor. Within the testing period, the ramp data are unavailable; there is a need for similar work that performs on a corridor with ramp data available to incorporate the ramp impacts on traffic behavior.

The proposed method merely accounts for the temporal characteristics in traffic flow data with a seasonal autoregressive integrated moving average (SARIMA) model. Future research might find it useful to take both temporal and spatial characteristics in traffic flow measurements into consideration in corridor travel time

model development. This may improve the corridor travel time prediction accuracy when using less accurate traffic flow data collected by single loop detectors.

Furthermore, the relationship between flow rate and occupancy is not considered in the model development for dynamic traffic flow prediction. Attempting to include consideration of the relation between flow rate and occupancy might make prediction of dynamic traffic flow better. For example, taking the linear relationship between the flow rate and occupancy under light traffic conditions might smooth the effects of the sudden changes in consecutive flow rates or occupancies on the flow rate and occupancy prediction. Future work may consider using a new variable (e.g. the ratio of flow rate and occupancy) for time series modeling for both link and corridor travel time prediction.

The proposed method for corridor travel time adjustment under an incident assumes that the approaching traffic flow of the physical queue does not change over time. An incident may affect traffic for a long time; this assumption may degrade the adjustment accuracy. Future work should be performed to analyze both spatial and temporal characteristics of traffic flow at the incident location.

Finally, testing of the proposed methodology is performed on the samples that pass through the data quality screening criteria. However, some traffic flow data still seems abnormal. It is found that the mean effective vehicle length is smaller than 14 feet for some traffic flow data when calculated from the measured flow rate, occupancy, and average speed. Related research is needed to develop more robust data screening criteria for wide applications of traffic flow data.

**APPENDIX A OPTIMAL VALUES OF BOX-COX
TRANSFORMATION PARAMETER**

VDS	Flow Rate	Occupancy
401079	0.45	0.15
401239	0.55	0.25
401052	0.40	0.00
400329	0.45	0.10
401195	0.40	0.05
401558	0.45	0.15
400378	0.45	0.25
400445	0.50	0.10
400443	0.40	0.10
401221	0.40	0.05
401228	0.45	0.25
400081	0.45	0.15
400770	0.40	0.15
401243	0.45	0.10
401209	0.40	0.15
401260	0.45	0.15
400976	0.60	0.15
400838	0.35	0.05
400430	0.45	0.10
400865	0.40	0.10

**APPENDIX B CORRIDOR TRAVEL TIME PREDICTION AND
ADJUSTMENT UNDER INCIDENTS**

**Table B.1 Corridor Travel Time Prediction and Adjustment under Incident 1 on
June 1, 2006 Thursday**

TIME	Actual Travel Time (min)	Predicted Corridor Travel Time without Considering Incident 1 (min)	Predicted Corridor Travel Time Considering Incident 1 (min)
16:45	12.17	11.49	N/A
16:50	12.33	11.40	N/A
16:55	12.17	11.27	N/A
17:00	12.33	11.54	N/A
17:05	12.67	11.73	N/A
17:10	13.00	12.46	N/A
17:15	13.17	13.39	13.50
17:20	13.33	13.36	13.55
17:25	13.17	13.48	13.42
17:30	12.83	13.14	13.60
17:35	12.83	12.73	13.30
17:40	12.67	12.45	13.13
17:45	13.00	12.24	12.65
17:50	13.00	12.28	12.65
17:55	12.50	11.65	11.81
18:00	12.33	11.27	11.45
18:05	12.17	11.06	11.49
18:10	12.00	10.80	11.12
18:15	11.83	11.05	11.15
18:20	11.33	10.40	10.87
18:25	10.83	10.19	10.50
18:30	10.67	9.85	10.26
18:35	10.50	9.54	10.37
18:40	10.17	9.22	9.92
18:45	10.00	9.07	10.04
18:50	9.50	8.87	9.73
18:55	9.17	8.35	8.26
19:00	9.00	8.13	N/A
19:05	8.70	8.04	N/A
19:10	8.67	7.82	N/A
19:15	8.33	7.64	N/A
19:20	8.17	7.64	N/A

Table B.2 Corridor Travel Time Prediction and Adjustment under Incident 2 on June 7, 2006, Wednesday

TIME	Actual Travel Time (min)	Predicted Corridor Travel Time without Considering Incident 1 (min)	Predicted Corridor Travel Time Considering Incident 1 (min)
17:30	14.73	14.11	N/A
17:35	14.49	14.20	N/A
17:40	14.99	14.99	N/A
17:45	15.43	13.14	N/A
17:50	17.00	14.50	N/A
17:55	16.67	13.21	14.65
18:00	15.83	13.53	15.26
18:05	15.17	13.26	15.00
18:10	14.83	12.43	14.24
18:15	13.83	12.12	13.23
18:20	13.17	11.64	12.95
18:25	12.50	11.31	12.74
18:30	12.17	10.92	12.13
18:35	11.67	10.50	12.20
18:40	11.17	10.46	11.06
18:45	10.67	9.78	10.45
18:50	10.00	9.49	10.35
18:55	9.50	8.67	9.24
19:00	8.33	8.1	8.68
19:05	8.17	7.91	N/A
19:10	7.83	7.72	N/A
19:15	7.83	7.69	N/A
19:20	7.67	7.66	N/A
19:25	7.67	7.66	N/A
19:30	7.83	7.65	N/A

Table B.3 Corridor Travel Time Prediction and Adjustment under Incident 3 on June 9, 2006, Friday

TIME	Actual Travel Time (min)	Predicted Corridor Travel Time without Considering Incident 1 (min)	Predicted Corridor Travel Time Considering Incident 1 (min)
16:00	12.33	10.91	N/A
16:05	12.00	10.35	N/A
16:10	11.83	10.56	12.27
16:15	12.33	10.57	12.35
16:20	12.83	10.84	12.57
16:25	13.33	10.55	12.39
16:30	13.50	11.05	12.91
16:35	14.00	11.77	13.51
16:40	14.50	11.58	13.39
16:45	14.33	11.24	14.59
16:50	14.50	11.05	14.42
16:55	14.17	10.83	14.21
17:00	13.83	10.30	13.72
17:05	13.83	9.84	12.39
17:10	14.50	10.11	12.70
17:15	14.50	13.04	13.06
17:20	14.50	12.93	12.97
17:25	14.17	12.75	13.29
17:30	14.00	13.02	13.28
17:35	13.67	12.48	12.54
17:40	13.17	13.25	13.34
17:45	12.17	11.57	12.08
17:50	11.67	9.59	10.22
17:55	11.17	9.43	10.03
18:00	10.83	9.39	9.97
18:05	10.50	9.43	9.98
18:10	10.50	9.40	10.01
18:15	10.33	9.16	9.78
18:20	10.17	8.93	9.54
18:25	10.17	8.65	9.25
18:30	10.00	8.55	9.16
18:35	10.00	8.59	9.20
18:40	9.83	8.70	9.31
18:45	9.67	8.61	9.24
18:50	9.17	8.39	9.05
18:55	8.67	8.23	8.89
19:00	8.17	8.40	N/A

Table B.4 Corridor Travel Time Prediction and Adjustment under Incident 5 on June 19, 2006, Monday

TIME	Actual Travel Time (min)	Predicted Corridor Travel Time without Considering Incident 1 (min)	Predicted Corridor Travel Time Considering Incident 1 (min)
17:10	12.00	11.29	N/A
17:15	12.00	11.43	N/A
17:20	11.83	11.07	N/A
17:25	11.67	10.97	11.00
17:30	12.00	10.81	10.83
17:35	11.83	10.96	10.97
17:40	11.67	10.29	10.52
17:45	11.33	10.53	10.74
17:50	11.17	9.37	9.89
17:55	11.00	9.06	9.75
18:00	10.50	8.72	9.56
18:05	10.00	8.31	9.17
18:10	10.00	8.24	9.11
18:15	10.00	8.41	9.33
18:20	10.00	7.94	8.28
18:25	9.67	7.91	8.14
18:30	9.50	7.65	N/A
18:35	9.17	7.65	N/A

Table B.5 Corridor Travel Time Prediction and Adjustment under Incident 3 on June 30, 2006, Friday

TIME	Actual Travel Time (min)	Predicted Corridor Travel Time without Considering Incident 1 (min)	Predicted Corridor Travel Time Considering Incident 1 (min)
21:30	10.17	9.70	N/A
21:35	9.50	8.91	N/A
21:40	10.00	10.11	N/A
21:45	9.67	9.89	9.89
21:50	9.50	8.15	9.89
21:55	9.17	8.35	9.89
22:00	9.00	8.13	8.91
22:05	8.67	7.91	N/A
22:10	8.33	7.65	N/A
22:15	8.00	7.64	N/A
22:20	8.00	7.66	N/A
22:25	8.00	7.69	N/A
22:30	8.00	7.68	N/A
22:35	8.17	7.67	N/A
22:40	8.17	7.69	N/A
22:45	8.00	7.67	N/A
22:50	8.00	7.66	N/A
22:55	8.00	7.66	N/A
23:00	7.83	7.66	N/A
23:05	7.83	7.66	N/A
23:05	10.17	9.70	N/A

REFERENCES

- Al_Deek, M., M. D'Angelo, and M. Wang. Travel Time Prediction with Non-Linear Time Series. In Proceedings of the 5th International Conference on Applications of Advanced Technologies in Transportation Engineering, America Society of Civil Engineers, Newport Beach, California, 1998, pp. 26-29.
- Angel, A. and M. Hickman. Experimental Investigation of Travel Time Estimation Using Geo-Referenced Arterial Video. Presented at the Transportation Research Board 81st Annual Meeting [CD-ROM], Washington, DC, 2002.
- Arem, B., M. Vlist, M. Muste, and S. Smulders. Travel Time Estimation in GERDIEN Project. In International Journal of Forecasting, Volume 13, Number 1, 1997, pp. 73-85.
- Athol, P. Interdependence of Certain Operational Characteristics within a Moving Traffic Stream. In Highway Research Record Proceedings, Volume 72, HRB, National Research Council, Washington, DC, 1965, pp. 58-87.
- Bailey, M. and F.G. Rawling. A Computerized Travel Time Study for Northeastern Illinois: Methodology and Commentary. Chicago Area Transportation Study Working Paper, Number 91-01, Chicago, Illinois, 1991.
- Berka, S., X. Tian, and A. Tarko. Data Fusion Algorithm for ADVANCE Release 2.0. ADVANCE Working Paper Series, Number 48, University of Illinois at Chicago, Chicago, Illinois, 1995.
- Berry, D.S. Evaluation of Techniques for Determining Over-All Travel Time. In Highway Research Board Proceedings, Volume 31, HRB, National Research Council, Washington, DC, 1952, pp. 429-440.
- Berry, D.S. and F.H. Green. Techniques for Measuring Over-All Speeds in Urban Areas. In Highway Research Board Proceedings, Volume 29, HRB, National Research Council, Washington, DC, 1949, pp. 311-318.
- Bohn, C and H. Unbehauen. Sensitivity Models for Nonlinear Filters with Application to Recursive Parameter Estimation for Nonlinear State-Space Models. In IEE Proceedings on Control Theory Applications, Volume 148, Number 2, 2001, pp. 137-145.

- Box, G. E. P. and Cox, D. R. An Analysis of Transformations. In *Journal of the Royal Statistical Society, Series B: Methodological*, Volume 26, Number 2, 1964, pp.211-252.
- Boyce, D., N. Rouphail, and A. Kirson. Estimation and Measurement of Link Travel Times in the ADVANCE Project. In *Proceedings of IEEE-IEE: Vehicle Navigation and Information Systems Conference*, Ottawa, Ontario, Canada, 1993, pp. 62-66.
- Bule, V., F. List, and J. Embrechts, Neural Network Freeway Travel Time Estimation. In *Proceedings of Intelligent Engineering Systems through Artificial Neural Networks*, American Society of Mechanical Engineers, Saint Louis, Missouri, 1994, pp. 1135-1140.
- Cathey, F.W. and D.J. Dailey. Transit Vehicles as Traffic Probe Sensors. Presented at the Transportation Research Board 81st Annual Meeting [CD-ROM], Washington, DC, 2002.
- Cathey, F.W. and D.J. Dailey. Corridor Travel Time Using Transit Vehicles as Probes. Presented at the Transportation Research Board 82nd Annual Meeting [CD-ROM], Washington, DC, 2003.
- Chang, S.L., L.S. Chen, Y.C. Chung, and S.W. Chen. Automatic License Plate Recognition. In *IEEE Transactions on Intelligent Transportation Systems*, Volume 5, Number 1, 2004, pp. 42-53.
- Chen, C. and S.I.J. Chien. Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation. In *Transportation Research Record 1719*, TRB, National Research Council. Washington, DC, 2000. pp. 61-68.
- Chen C. Travel Times on Changeable Message Signs: Pilot Project. California PATH Working Paper, UCB-ITS-PRR-2004-5, University of Berkeley, Berkeley, California, 2004.
- Chen, C. and J. Xia. Developing a Strategy for Imputing Missing Volume Data. Presented at the Transportation Research Board 85th Annual Meeting [CD-ROM], Washington, DC, 2006.
- Cheu, R.L., S.G. Ritchie, W. Recker, B. Bavarian. Investigation of a Neural Network Model for Freeway Incident Detection. Report No. UCT-ITS-TS-WP-91-1, University of California at Irvine, Irvine, California, 1991.

- Cheu, R.L., C. Xie, and Der-H. Lee. Probe Vehicle Population and Sample Size for Arterial Speed Estimation. In *Computer-Aided and Infrastructure Engineering*, Volume 17, Number 1, 2002, pp. 53-60.
- Chevillon, G. and D.F. Hendry. Non-Parametric Direct Multi-Step Estimation for Forecasting Economic Processes. *Economics Series Working Papers*, Number 196, Department of Economics, Oxford, United Kingdom, 2004.
- Chu, L. and W. Recker. Micro-Simulation Modeling Approach to Applications of On-Line Simulation and Data Fusion. *California PATH Working Paper*, UCB-ITS-PRR-2004-1, University of Berkeley, Berkeley, California, 2004.
- Choi, K. and Y. Chung. Travel Time Estimation Algorithm Using GPS Probe and Loop Detectors Data Fusion. Presented at the Transportation Research Board 80th Annual Meeting [CD-ROM], Washington, DC, 2001.
- Cohen, H. On the Measurement and Valuation of Travel Time Variability Due to Incidents on Freeways. In *Journal of Transportation and Statistics*, Volume 2, Number 2, 1999, pp.123-131.
- Coifman, B. A New Methodology for Smoothing Freeway Loop Detector Data: Introduction to Digital Filtering. In *Transportation Research Record 1554*, TRB, National Research Council, Washington, DC, 1996, pp. 142-152.
- Coifman, B. Improved Velocity Estimation Using Single Loop Detectors. In *Transportation Research, Part A: Policy and Practice*, Volume 35, Number 10, 2001, pp. 863-880.
- Coifman, B. Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors. In *Transportation Research, Part A: Policy and Practice*, Volume 36, Number 4, 2002, pp. 351-364.
- Courage K.G., S. Bauer, and D.W. Ross. Operating Parameters for Main Line Sensors in Freeway Surveillance Systems. In *Transportation Research Record 601*, TRB, National Research Council, Washington, DC, 1976, pp. 19-28.
- Dailey, D.J. Travel Time Estimates Using a Series of Single Loop Volume and Occupancy Measurements. Presented at the Transportation Research Board 76th Annual Meeting [CD-ROM], Washington, DC, 1997.
- Dailey, D.J. and L. Li. An Algorithm to Estimate Vehicle Speed Using Un-Calibrated Cameras. In *IEEE Transactions on Intelligent Transportation Systems*, Volume 1, Number 2, 2000, pp. 98-107.

- Dailey, D.J. and F.W. Cathey. AVL-Equipped Vehicles as Traffic Probe Sensors. Final Research Report No. WA-RD-534.1, Washington State Department of Transportation, Olympia, Washington, DC, 2002.
- deSilva, C.J.S. A Generalization of the Kalman Filter. Australian Research Center for Medical Engineering, Murdoch University, Murdoch, Australian 2006.
- Eisele, L. Estimating Travel Time and Variance Using Intelligent Transportation Systems Data for Real-Time and Off-line Transportation Applications. Ph.D. Dissertation, Texas A&M University, College Station, Texas, 2001.
- Florio, L., and L. Mussone. Neural Network Models for Classification and Forecasting of Freeway Traffic Flow Stability. In Symposium of Transportation Systems: Theory and Application of Advanced Technology, International Federation of Automatic Control, 1994, pp. 773-784.
- Gallagher, J. Travel Time Data Collection Using GPS. Presented at National Traffic Data Acquisition Conference, Albuquerque, New Mexico, 1996, pp. 147-161.
- Gerlough, D.L. and M.J. Huber. Traffic Flow Theory. Special Report 165, TRB, National Research Council, Washington, DC, 1976.
- Guiroga, C.A. and D. Bullock. Travel Time Information Using GPS and Dynamic Segmentation Techniques. Presented at the Transportation Research Board 78th Annual Meeting [CD-ROM], Washington, DC, 1999.
- Guo, B. and A.D. Poling. Geographic Information Systems/ Global Positioning Systems Design for Network Travel Time Study. In Transportation Research Record 1497, TRB, National Research Council, Washington, DC, 1995, pp. 135-144.
- Guo, J. Adaptive Estimation and Prediction of Univariate Vehicular Traffic Condition Series. PhD dissertation, North Carolina State University, 2005.
- Gupta, R., J. D. Fricker, and D.P. Moffett. Reduction of Video License Plate Data. In Transportation Research Record 1804, TRB, National Research Council, Washington, DC, 2002, pp. 31-38.
- Hagen, L.T. Best Practices for Traffic Incident Management in Florida. Final Research Report No. BC353-47, University of South Florida, Tampa, Florida, 2005.
- Hall, F.L. and B.N. Persaud. Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems. In Transportation

- Research Record 1232, TRB, National Research Council, Washington, DC, 1989, pp. 9-16.
- Hellinga, B. R. and L. Fu. Assessing Expected Accuracy of Probe Vehicle Travel Time Reports. In *Journal of Transportation Engineering*, Volume 125, Number 6, 1999, pp. 524-530.
- Hoffman, G. and J. Janko. Travel Time as a Basis of the LISB Guidance Strategy. In *Proceedings of IEEE 3rd International Conference on Road Traffic Control*, London, United Kingdom, 1990, pp. 6-10.
- Hua, J. and Faghri A. Application of Artificial Neural Networks to IVHS. In *Transportation Research Record 1453*, TRB, National Research Council, Washington, DC, 1994, pp. 98-104.
- ITE. *Transportation and Traffic Engineering Handbook*, Second Edition. Institute of Transportation Engineers, Washington, DC, 1976.
- ITE. *Manual of Transportation Engineering Studies*. Institute of Transportation Engineers, Washington, DC, 1994.
- Ishimaru, J.M. and M.E. Hallenbeck. Flow Evaluation Design. Technical Report No. WA-RD 466.2, Washington State Department of Transportation, Olympia, Washington, 1999.
- Jia, Z., C. Chen, B. Coifman, and P. Varaiya. The PeMS Algorithms for Accurate, Real-Time Estimates of g-Factors and Speeds from Single-Loop Detectors. In *Proceedings of IEEE Intelligent Transportation System Conference*, Oakland, California, 2001, pp. 536-541.
- Kalman, R.E. New Results in Linear Filtering and Prediction Theory Problems. In *Journal of Basic Engineering*, Transactions of the ASME, Volume 82, 1960, pp.35-45.
- Karjalainen, P.A. Estimation Theoretical Background of the Root Tracking Algorithms with Applications to EEG. Department of Applied Physics, University of Kuopio, Kuopio, Finland, 1996.
- Kazimi C., A. Brownstone, T.F. Gosh, and van Amelsfort. Willingness-to-Pay to Reduce Commute Time and Its Variance: Evidence from the San Diego I-15 Congestion Pricing Project. Presented at the Transportation Research Board 79th Annual Meeting [CD-ROM], Washington, DC, 2000.

- Kwon, J. Joint Estimation of the Speed and Mean Vehicle Length from Single-Loop Detector Data. Presented at the Transportation Research Board 82nd Annual Meeting [CD-ROM], Washington, DC, 2003.
- Laird, D. Emerging Issues in the USE of GPS for Travel Time Data Collection. Presented at National Traffic Data Acquisition Conference, Albuquerque, New Mexico, 1996, pp. 117-123.
- Larsen, R. Using Cellular Phones as Traffic Probes. In *Traffic Technology International*, August/September 1996.
- Levine, S.Z., W.R. McCasland, and D.O. Smalley. Development of a Freeway traffic Management Project Through a Public-Private Partnership. In *Transportation Research Board 1394*. TRB, National Research Council. Washington, DC, 1993.
- Lighthill, M. J. and G. B. Whitham. On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. In *Proceedings of the Royal Society, Series A* 229, London, United of Kindom 1957, pp. 317-345.
- Lin, W., J. Dahlgram, and H. Huo. An Enhancement to Speed Estimation with Single Loops. California PATH Working Paper, UCB-ITS-PWP-2003-14, University of Berkeley, Berkeley, California, 2003.
- Linveld, C.D.R. and R. Thijs. Online Travel Time Estimation Using Inductive Loop Data: the Effect of Instrumentation Peculiarities on Travel Time Estimation Quality. In *Proceedings of the 6th ITS World Congress*, Toronto, Canada, 1999.
- Linveld, C. D.R., R. Thijs, P.H.L. Bovy, and N.J. van der Zijpp. Evaluation of Online Travel Time Estimators and Predictors. In *Transportation Research Record 1719*, TRB, National Research Council, Washington, DC, 2000, pp. 45-53.
- Liu, T.K. and M. Haines. Travel Time Data Collection Field Tests-Lessons Learned. Report No. FHWA-PL-96-010, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 1996.
- Lomax, T., S. Turner, G. Shunk, H.S. Levinson, R.H. Pratt, P.N. Bay, and G.B. Douglas. Quantifying Congestion: User's Guide. NCHRP Report 398, TRB, National Research Council, Washington, DC, 1997.
- Lomax, T., S. Turner, and R. Margiotta. Monitoring Urban Roadways in 2002: Using Archived Operations Data for Reliability and Mobility Measurement, Report

- No. FHWA-HOP-04-011, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 2004.
- May, A.D. *Traffic Flow Fundamentals*. Prentice Hall Inc., Englewood Cliffs, New Jersey, 1990.
- Mikhalkin, B., H.J. Payne, and L. Isaksen. Estimation of Speed from Presence Detectors. In *Highway Research Record Proceedings, Volume 388, HRB, National Research Council, Washington, DC, 1972, pp. 73-83.*
- Moré, J.J. *The Levenberg-Marquardt Algorithm: Implementation and Theory. Lecture Notes in Mathematics 630, Springer-Verlag, Berlin, 1978, pp. 105-116.*
- Myers, K.A. and B.D. Tapley. Adaptive Sequential Estimation with Unknown Noise Statistics. In *IEEE Transactions on Automatic Control, Volume 21, Number 4, 1976, pp. 520-523.*
- Nam, Do H. and D.R. Drew. Analyzing Freeway Traffic Under Congestion: Traffic Dynamic Approach. In *Journal of Transportation Engineering, Volume 124, Number 3, 1996, pp. 208-212.*
- Nelson, P. and P. Palacharla, A Neural Network Model for Data Fusion in DAVANCE. In *Proceedings of Pacific Rim Conference, America Society Civil Engineers, Volume 1, 1993, pp. 237-243.*
- Ngo, K.G. *GPS based Travel Time Study Data Collection Operation Manual*. Florida State University, Tallahassee, Florida, 2005.
- Park, D. and R. Rilett, Forecasting Freeway Link Travel Times with Modular Neural Networks. Presented at the Transportation Research Board 77th Annual Meeting [CD-ROM], Washington, D.C., 1998.
- Park, D., R. Rilett, and G. Han, Forecasting Multiple-period Freeway Link Travel Times using Neural Networks with Expanded Input Nodes. In *Proceedings of the 5th International Conference on Applications of Advanced Technology in Transportation Engineering, America Society Civil Engineers, 1998, pp. 325-332.*
- Parkany, E. and C, Xie. A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and What Doesn't. Report No. NETCR37, University of Massachusetts, Amherst, Massachusetts, 2005.
- Payne, H. J., E.D. Helfenbein, and H.C. Knobel. *Development and Testing of Incident Detection Algorithms: Vol. 2—Research Methodology and Detailed Results.*

- Technical Report No. FHWA-RD-76-12, Federal Highway Administration, Washington, DC, 1976.
- Pourmollem, N., T. Nakatsuji, and A. Kawamura. A Neural-Kalman Filtering Method for Estimating Traffic States on Freeways. In *Journal of Infrastructure Planning and Management*, Volume 36, Number. 569, 1997, pp.105-114.
- Quiroga, C.A. and D. Bullock. Determination of Sample Sizes for Travel Time Studies. In *ITE Journal*, August, Institute of Transportation Engineers, Washington, DC, 1998.
- Richards, P. I. Shockwaves on the Highway. In *Journal of Operation Research*, Volume 4, 1956, pp. 42-51.
- Rickman, T.D., M.E. Hallenbeck, and M. Schroeder. Improved Method for Collecting Travel Time Information. In *Transportation Research Record 1271*. TRB, National Research Council, Washington, DC, 2000. pp. 79-88.
- Rilett, R. and D. Park. Direct Forecasting of Freeway Corridor Travel Time Using Spectral Neural Networks. Presented at the Transportation Research Board 78th Annual Meeting [CD-ROM], Washington, D.C., 1999.
- Roden, D. Travel Time Data Collection Using GPS Technologies. Presented at National Traffic Data Acquisition Conference, Albuquerque, New Mexico, 1996, pp. 163-182.
- Rouphail M., A. Targo, C. Nelson, and P. Palacharla. Travel Time Data Fusion in ADVANCE-A Preliminary Design Concept. ADVANCE Working Paper Series, Number 21, University of Illinois at Chicago, Chicago, Illinois, 1993.
- Sanwal, K.K. and J. Walrand. Vehicles as Probes. California PATH Working Paper, UCB-ITS-PWP-95-11, University of Berkeley, Berkeley, California, 1995.
- SAS Institute Inc. SAS OnlineDoc®, Version 9, SAS Institute Inc., Cary, NC, 2002
- Schaefer, M.C. License Plate Matching Surveys: Practical Issues and Statistical Considerations. In *ITE Journal*, July, Institute of Transportation Engineers, Washington, DC, 1988, pp. 37-42.
- Sen, A., P. Thakuriah, Xia-Q. Zhu, and A. Karr. Frequency of Probe Reports and Variance of Travel Time Estimates. In *Journal of Transportation Engineering*. Volume 123, Number 4, 1997, pp. 290-297.
- Shuldiner, P.W., S.A. D'Agostino, and J.B. Woodson. Determining Detailed Origin-Destination and Travel Time Patterns Using Video and Machine Vision

- License Plate Matching. In Transportation Research Record 1551, TRB, National Research Council, Washington, DC, 1996, pp. 8-17.
- Skabardonis, A., P. Varaiya and K. Petty. Measuring Recurrent and Nonrecurrent Traffic Congestion. In Transportation Research Record 1856, TRB, National Research Council, Washington, DC, 2003, pp 118-124.
- Smith, B.L., H. Zhang, and M. Frontaine, and M. Green. Cellphone Probes as an ATMS Tool. Report No, STL-2003-01, University of Virginia, Charlottesville, Virginia, 2003.
- Srinivasan, K.K. and P.P. Jovans. Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network. In Transportation Research Record 1537, TRB, National Research Council. Washington, DC, 1995. pp. 15-22.
- Tarko, A. and N. Roupail, Travel Time Data Fusion in ADVANCE. In Proceedings of 3rd International Conference on Applications of Advanced Technologies in Transportation Engineering, America Society Civil Engineers, 1993, pp. 36-42.
- Takahashi, Y., K. Ikenoue, K. Yasui, and Y. Kunikata. Travel Time Information System (TTIS) in Mie Prefecture. In Proceedings of the 3rd Annual World Congress on Intelligent Transportation Systems, Orlando, Florida, 1996.
- Tarvainen, M.P., J.K. Hilunen, P.O. Ranta-aho, and P.A. Karjalainen. Estimation of Nonstationary EEG with Kalman Smoother Approach: an Application to Event-Related Synchronization (ERS). Department of Applied Physics, University of Kuopio, Kuopio, Finland, 2003.
- The MathWorks Inc. Model Brower Documentation. The MathWorks Inc., Natick, Massachusetts, 1994.
- Thijs, R., C.D.R. Linveld, N.J. van der Zijpp, and P.H.L. Bovy. Evaluation of Travel Time Estimation and Prediction algorithms: Evaluation Results from the DACCORD Project. Transportation Planning and Traffic Engineering Section, Delft University of Technology, Delft, Netherlands, 1999.
- TRB. Highway Capacity Manual, 4th Edition. Special Report 209, TRB, National Research Council, Washington, DC, 2000.
- Turner, S.M., C.J. Naples, III, and R.H. Henk. Travel Time Reliability of HOV Facilities. In Compendium of 64th ITE Annual Meeting, Institute of Transportation Engineers, Washington, DC, 1994, pp. 349-353.

- Turner, S.M. and D.J. Holdener. Probe Vehicle Sample Sizes for Real-Time Information: The Houston Experience. In Proceedings of the Vehicle Navigation & Information Systems Conference, Seattle, Washington, IEEE, 1995, pp. 3-10.
- Turner, S.M. and J.B. Woodson. Use of Advanced Technology in HOV Lane Enforcement. Presented at National Traffic Data Acquisition Conference, Albuquerque, New Mexico, 1996, pp. 456-466.
- Turner, S.M., W.L. Eisele, R.J. Benz, and D.J. Holdener. Travel Time Data Collection Handbook. Report No. FHWA-PL-98-035, Federal Highway Administration, U.S. Department of Transportation, 1998.
- Vaidya, N., L.L. Higgins, and K.F. Turnbull. An Evaluation of the Accuracy of a Radio-Trilateration Automatic Vehicle Location System. In Proceedings of the 1996 Annual Meeting of ITS America, Intelligent Transportation Society of America. Washington, DC, 1996.
- Van Aerde, M., B. Hellinga, L. Fu, and H. Rakha. Vehicle Probes as Real-Time ATMS Sources of Dynamic O-D and Travel Time Data. In Proceedings of the ATMS Conference, Petersburg, Florida, 1993, pp. 207-230.
- Van Der Zijpp, N.J. and C. D.R. Linveld. Evaluation of Queue Length Display at the Amsterdam Orbital Motorway. In Proceedings of the 6th ITS World Congress, Toronto, Canada, 1999.
- Van Lint, J.W.C. and N.J. Van Der Zijpp. An Improved Travel-time Estimation Algorithm Using Dual Loop Detectors. Presented at the Transportation Research Board 82nd Annual Meeting [CD-ROM], Washington, DC, 2003.
- Wang, Y. and N.L. Nihan. Freeway Traffic Speed Estimation with Single-Loop Outputs. In Transportation Research Record 1727, TRB, National Research Council, Washington, DC, 2000, pp. 120-126.
- Washburn, S.S. and N.L. Nihan. Using Voice Recognition to Collect License Plate Data for Travel Time Studies. In Transportation Research Record 1593. TRB, National Research Council, Washington, DC, 1997, pp. 41-46.
- Washburn, S.S. and N.L. Nihan. Estimating Travel Time with the Mobilizer Video Image Tracking System. In Journal of Transportation Engineering, Volume 125, Number 1, 1999, pp. 15-20.

- Welch, G. and G. Bishop. An Introduction to the Kalman Filter. Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina, 2004.
- Williams, B.M. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process. PhD dissertation, University of Virginia, 1999.
- Williams, B. M. and L.A. Hoel, L. A. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA: Theoretical Basis and Empirical Results. In Journal of Transportation Engineering, Volume 129, Number6, 2003, pp.664-672.
- Wolfram Research Inc. Time Series Documentation. Wolfram Research Inc., Champaign, Illinois, 2006.
- Xia, J. and C. Chen. A Nested Clustering Technique for Freeway Operating Condition Classification. Accepted by Computer-Aided and Infrastructure Engineering.
- Xie, C., R. Cheu, and D. Lee. Improving Arterial Link Travel Time Estimation by Data Fusion. Presented at the Transportation Research Board 83rd Annual Meeting [CD-ROM], Washington, D.C., 2004.
- Ygnace, Jean-L., C. Drane, Y.B.Yim, and R. de Lacvivier. Travel Time Estimation on the San Francisco Bay Area Network Using Cellular Phones as Probes. California PATH Working Paper, UCB-ITS-PWP-2000-18, University of Berkeley, Berkeley, California, 2000.
- Yim, Y.B. and R. Cayford. Investigation of Vehicles as Probes Using Global Positioning System and Cellular Phone Tracking: Field Operational Test. California PATH Working Paper, UCB-ITS-PWP-2001-9, University of Berkeley, Berkley, California, 2001.
- Zwet, E., C. Chen, and J. Kwon. A Statistical Method for Estimating Speed from Single Loop Detectors. Presented at the Transportation Research Board 82nd Annual Meeting [CD-ROM], Washington, DC, 2003.

VITA

DATE AND PLACE OF BIRTH

November 13, 1976

Lianshui County, Jiangsu Province, P.R.China

EDUCATION

M.S., Transportation Engineering, Southeast University, Nanjing, P. R. China, April 2002.

B.S., Transportation Engineering, Southeast University, Nanjing, P. R. China, July 1999

RESEARCH EXPERIENCE

Graduate Research Assistant, University of Kentucky, 2002 to Present.

Graduate Research Assistant, Southeast University, P. R. China, 1999 to 2002.

REFEREED JOURNAL PUBLICATIONS

Xia, J. and M. Chen. A Nested Clustering Procedure for Traffic Condition Classification. Accepted by Journal of Computed-Aided Civil and Infrastructure Engineering.

Chen, M. and J. Xia. Develop a Strategy for Imputing Missing Volume Data. Accepted by Journal of Transportation Research Forum.

Chen, M., X. Liu, and J. Xia. Dynamic Prediction Method with Schedule Recovery Impact for Bus Arrival Time. In Transportation Research Record 1923, TRB, National Research Council, Washington, DC, 2005, pp. 208-217.

Chen, M., X. Liu, J. Xia, and S. Chien. A Dynamic Bus Arrival Time Prediction Model Based on APC Data. In Journal of Computer-Aided Civil and Infrastructure Engineering, Vol. 19, No. 5, 2004, pp 364-376.

Xia, J. and W. Huang. Research on Expressway Project Construction Management. In Journal of Southeast University, Vol.32, Issue 1, P. R. China, 2001, pp 65-69.

Jingxin Xia

12-10-2006