MD-07-SP508B4D

Robert L. Ehrlich, Jr., *Governor*
Michael S. Steele, *Lt. Governor*

Robert L. Flanagan, *Secretary*
Neil J. Pedersen, *Administrator*

STATE HIGHWAY ADMINISTRATION

RESEARCH REPORT

DEVELOPMENT AND FIELD EVALUATION OF A REAL-TIME TRAVEL TIME PREDICTION SYSTEM

GANG-LEN CHANG, NAN ZOU, AND JIANWEI WANG

**UNIVERSITY OF MARYLAND
COLLEGE PARK**

**Project number MD-07-SP508B4D
FINAL REPORT**

December 2006

Technical Report Documentation Page

1. Report No. MD-07-SP508B4D	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle **Development and Field Evaluation of a Real-Time Travel Time Prediction System**		5. Report Date December 2006	
		6. Performing Organization Code	
7. Author/s **Gang-Len Chang, Nan Zou, Jianwei Wang**		8. Performing Organization Report No.	
9. Performing Organization Name and Address University of Maryland 1172 Glenn Martin Hall Department of Civil and Environmental Engineering College Park, MD 20742		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. SP508B4D	
12. Sponsoring Organization Name and Address Maryland State Highway Administration Office of Policy & Research 707 North Calvert Street Baltimore MD 21202		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency Code (7120) STMD - MDOT/SHA	
15. Supplementary Notes			

16. Abstract

This study has developed a real-time travel time prediction system for a highway segment with sparsely-distributed detectors. The entire system consists of the following principal modules: detector location optimization model, travel time estimation models, missing data imputation model, accident detection models, and travel time prediction models. A field demonstration of the developed system was implemented on the I-70 freeway segment connected to I-695 in the Baltimore metropolitan area. Its performance in providing a reliable travel time prediction was evaluated independently by a consulting firm working for the Maryland Highway Administration. The evaluation results revealed that it is likely to develop a reliable travel time prediction system for commuting corridors if all carefully calibrated detectors have been deployed at proper locations and sufficient historical traffic pattern data have been collected and modeled in advance. The field demonstration has also indicated the need to contend with the critical missing data issues incurred frequently in the real-time operations, especially using the wireless communication systems.

17. Key Words Travel time, prediction, freeway	18. Distribution Statement: No restrictions This document is available from the Research Division upon request.		
19. Security Classification (of this report) None	20. Security Classification (of this page) None	21. No. Of Pages 125	22. Price

Form DOT F 1700.7 (8-72) Reproduction of form and completed page is authorized.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Due to the increasing congestion in most urban networks, providing reliable trip times to commuters has emerged as one of the most critical challenges for all existing advanced traveler information systems (ATIS). However, predicting travel time is a very complex and difficult task, as the resulting accuracy varies with many variables of time-varying nature, including the day-to-day traffic demands, responses of individual drivers to daily commuting congestion, conditions of the road facility, weather, incidents, and reliability of available detectors. To contend with these issues, transportation professionals have proposed and implemented a variety of systems for providing travel times in the past two decades. However, most real world systems provide travel times based on only the current traffic conditions, not the predicted travel time for en-route trips or for pre-trip planning.

Traditionally, travel time prediction models are based on historical travel times concurrently collected by various measurement systems such as electronic toll systems or vehicles with GPS systems. However, due to the high costs associated with collecting a large sample with such systems, most models developed for travel time prediction have not been implemented in practice.

As an alternative, considerable efforts are found in the literature to estimate travel times from traffic detectors, which are relatively cost-effective for implementation in practice, but demand some advanced theoretical models to produce the predicted travel time from limited point measurement information offered by detectors under the potential impacts of various critical factors.

For example, to formulate a reliable travel time model for prediction, one needs to be able to reliably capture the traffic dynamics between detector stations. The complexity of such a task increases with the distance between detectors, and the percentage of missing or faulty data during the detection period. The prediction shall also take into account the future traffic demand generated to the downstream segments of en-route trips, as the surge in volume in the projected time horizon may incur the traffic congestion that is difficult to be estimated with the data from the existing detectors. To reliably predict the future time-varying traffic demand, however, is also quite a complex task, and it necessitates the proposed model not only to best use the available historical data, but also to dynamically account for the day-to-day variation due to the experience of drivers or their responses to the perceived traffic conditions.

In brief, the complex interrelations between detector hardware, historical data, and traffic flow dynamics have made the prediction of travel time one of the most challenging tasks in ATIS. This is also one of the primary reasons that most ATIS for highway systems only provide estimated travel times based on the current traffic condition.

1.2 Research Objectives

Theoretically, a cost-efficient travel time prediction system ready for use in practice on freeways should have the following desirable features:

- The required input variables should be obtainable from traffic detectors, which may be sparsely distributed.
- The system may take advantage of some actual travel times from the field, but not rely on a large number of such data.
- The system should be capable of operating under non-recurrent congestion conditions and effectively dealing with related issues during real-time operations.

Intending to embody all the above desirable features in the proposed travel time prediction system, this study has the following principal objectives:

- Develop a travel time estimation module to provide reliable estimates of completed trips under all types of recurrent traffic patterns with sparsely distributed traffic detectors.
- Construct a travel time prediction module for freeway segments with a large detector spacing, and take full advantage of historical travel times and traffic patterns.
- Integrate a missing data estimation module to deal with various missing data patterns that often occur in a real-world system.
- Calibrate an incident detection module to switch the travel time prediction system to a different mode (i.e., display delay warnings instead of predicted travel times) when an incident has been detected.

1.3 Organization of the Report

Based on the proposed research objectives, this report consists of eight chapters. The interrelations among those tasks are illustrated in Figure 1.1. A brief introduction of each completed chapter is presented next.

Chapter 2 presents a comprehensive review of literature related to the travel time prediction system, including travel time estimation models, travel time prediction models, and simulated/real-world systems. Advantages and limitations of those models with respect to their potentials for use in a real-world system with sparsely distributed detectors are also discussed in this chapter.

The primary task of Chapter 3 is to introduce the framework of the proposed travel time prediction system with sparsely distributed detectors. The system's flowchart and its operational logic will be presented in detail in this chapter. The proposed travel time prediction system consists of four principal modules: travel time estimation module, travel time prediction module, missing data estimation module, and incident detection module.

Chapter 4 focuses on developing a hybrid travel time estimation model on a freeway with sparsely distributed detectors. The proposed hybrid travel time estimation model employs a clustered linear regression model as the main model and an enhanced trajectory-based model as its supplemental model to circumvent the limitations on long links identified in the literature review. To contend with the impacts due to various geometric features and traffic patterns, the hybrid model first categorizes traffic conditions into pre-specified groups, and then applies the

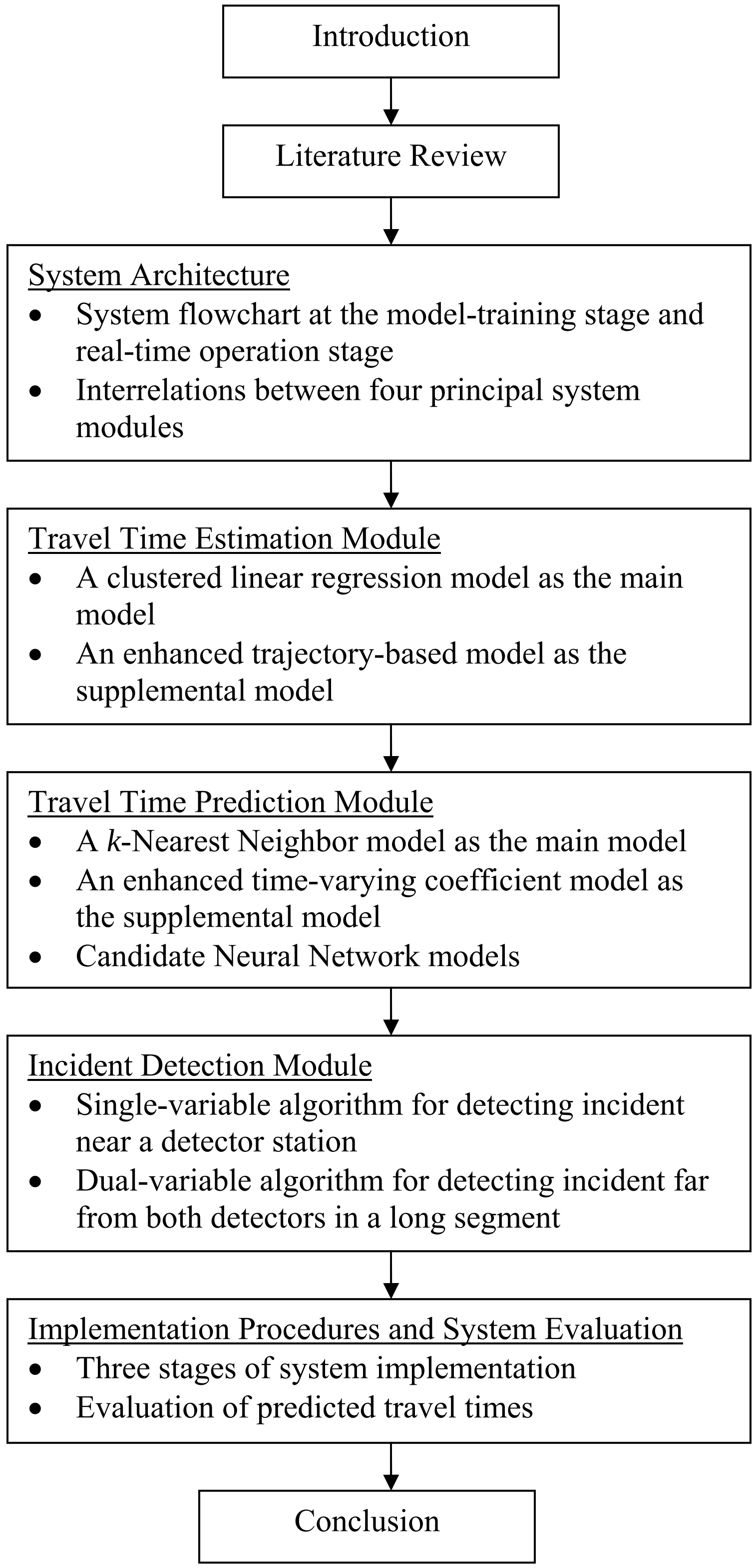


Figure 1.1 Interrelations between primary research tasks

linear regression model to clusters with a sufficient size of sample travel times. The enhanced trajectory-based model takes strengths of both traffic propagation relations and piecewise linear speed-based model to provide the reliable estimation of travel times for clusters without sufficient samples.

Chapter 5 proposes a hybrid travel time prediction model for freeway segments with widely spaced detectors. The hybrid travel time prediction model takes full advantages of a *k*-Nearest Neighbor model and an enhanced time-varying coefficient model to provide reliable travel time prediction under recurrent congestion patterns. The *k*-Nearest Neighbor model, which is the main model of the hybrid model structure, has been modified and modeled to take into account traffic characteristics as well as both daily and weekly traffic patterns. An enhanced time-varying coefficient model serves as the supplemental component to deal with the traffic scenarios that do not have sufficient similar cases in the historical database. This enhanced time-varying coefficient model functions to predict travel times, based on the improved preliminary estimates derived from the daily and weekly traffic patterns.

Chapter 6 details the incident detection algorithms employed in this study to contend with nonrecurrent congestion patterns. Its interrelations with the developed algorithms for travel time estimation and prediction constitute the core of this chapter. The proposed module consists of two components: a single-station incident detection algorithm for detecting incidents near one detector station and a dual-station incident detection algorithm for detecting incidents far from either detector. When a major incident is detected, the incident detection module will inform the travel time prediction system to switch to a different operation mode (i.e., displaying congestion warning messages) for the impacted links.

Chapter 7 reports the system demonstration on the I-70 freeway segment between MD27 and I-695. The presentation includes the pre-deployment system calibration, on-line system demonstration, and off-line system evaluation and model revision. A step-by-step description of the system deployment procedures for field demonstration is the main focus of this chapter. Field evaluation results for travel time prediction will also be discussed in this chapter.

Chapter 8 summarizes the research findings for this study and reports some critical issues to be addressed in the future development of such a system; recommendations for further system enhancement and full-scale deployment will also be addressed in this chapter.

Chapter 2: Literature Review

2.1 Introduction

Most existing studies associated with providing trip travel times on freeways can be classified into two categories: travel time estimation and travel time prediction. The former studies are used to estimate travel times from the traffic data collected during the time in which the trip has been completed. This type of study is essential for a travel time prediction system, which does not directly measure travel times. In contrast, travel time prediction models are for trips that have not departed and will be completed in the future. Thus, future traffic conditions have to be predicted, which makes predicting travel time a challenging task. Embedding a missing data estimation module in a travel time prediction system can significantly improve its reliability and functionality, the accuracy of which is frequently impaired by missing and/or delayed data. This section will first review travel time estimation models and travel time prediction models in the literature. Then it will summarize some systems implemented in simulated environments and in real-world applications.

2.2 Travel Time Estimation Models

As reported in the literature, most studies of travel time estimation fall into one of the following categories: flow-based models, vehicle identification approaches, and trajectory-based models.

2.2.1 Flow-based Models

Flow-based models have been applied to freeway mainline segments without ramps and having uniform travel speeds across all lanes. This type of model estimates travel times by comparing upstream and downstream flow counts, based on the assumption of first depart, first arrive. For example, Dailey (1993) estimated travel times by using a cross-correlation technique to determine the maximum correlation between densities, which are computed from flow measurements.

Nam and Drew (1996) developed a flow-based travel time estimation model by analyzing the number of vehicles that have entered and exited the link in the same time interval, $m(t_n)$. The authors applied a stochastic process model to the upstream and downstream flow counts under generalized conditions of flow conservation and then estimated travel times for the traffic condition in which $m(t_n)$ is positive. A case study showed that the estimated average segment travel speed was consistent with detected upstream and downstream speeds.

By extending Dailey's work, Petty et al. (1998) estimated freeway travel times using flow and occupancy information, based on a simple stochastic model, by analyzing probability distributions of travel times. However, the model results have been verified using only the upstream detector speed, which is not sufficiently reliable to serve as the ground truth value of travel time.

Liu et al. (2006) established a linear relation between travel time and the combination of the number of vehicles in the segment and the average downstream speed. To solve the model, the authors provided an iteration-based method to establish the interrelation between input and output variables. The estimated travel times from two cases generated in a simulation environment were found to be reliable in two distinct types of traffic conditions.

In comparison, existing flow-based models require uniform travel speeds across all lanes and therefore cannot be reliably applied to segments with ramps or complex traffic patterns, i.e., spillback from a downstream off-ramp. Another issue that makes this type of model unsuitable for real-world applications is detector errors. In practice, even the most advanced and properly calibrated detectors still cannot be guaranteed to operate at a desirable level of high detection accuracy. Unpredictable measurement errors for traffic count may dramatically reduce the model accuracy. Nam and Drew (1996) considered an hourly adjustment factor to overcome the drifted flow count. However, detector errors are most likely nonsystematic in nature, and the error patterns remain difficult to model well.

2.2.2 Vehicle Identification Approaches

Vehicle identification approaches estimate travel time by matching the sequence of vehicles in a single lane. The key concept of this type of method is to find vehicles' signatures from the upstream and the downstream detectors in order to calculate their travel times.

In the literature, significant efforts have been made to group vehicles into classes and then match their sequences to estimate travel times. These models (Pfannerstill, 1984; Kühne and Immes, 1993; and Kühne et al., 1997) often require new detection hardware that can provide additional signatures. MacCarley (1998) proposed a method using vehicles' visual signatures from overhead cameras to obtain travel times. The evaluation results indicate that such systems can achieve a high degree of accuracy in daylight, but have a low match rate and a high false-match rate at nighttime.

Coifman et al. (Coifman, 1998; Coifman and Cassidy, 2002; Coifman, 2003; and Coifman and Ergueta, 2003) estimated travel times with a vehicle re-identification (VRI) model, which matches the sequence of individual vehicles or a sub-sampling of vehicles (for example, trucks) with their occupied durations when they pass the upstream and the downstream loop detectors. The VRI model worked well under both free-flow conditions and congested conditions with a very low lane-changing rate. It is reported that the model produces results having the same quality as other travel time estimation methods. However, due to its reduced detection resolution at high vehicle speeds, its match rate is generally quite low under free-flow conditions.

In general, vehicle identification models performed well in one single lane with a low lane-changing rate. They cannot provide reliable travel time estimations for freeway segments near ramps. Using vehicles' visual signatures may potentially improve the model's ability to deal with ramp traffic. However, all VRI models require either improved detection technology or a high bandwidth to transfer the raw data needed to extract vehicle signatures, which will result in high system costs and long system processing times.

2.2.3 Trajectory-based Models

The common features of trajectory-based models are estimating temporal and spatial traffic conditions within a link from upstream and downstream detector data and drawing a target vehicle's trajectory so as to provide the estimated travel time.

One of the typical studies in this category is by Coifman (2002), who estimated the vehicle in-segment speed based on the speed data from a detector placed at one end of a 1/3-mile segment and the traffic propagation relations. With the assumption that the traffic state at one detector location changes discretely and equal to vehicles' headways, the following relations exist for the j^{th} state with an assumed constant traffic propagation speed.

$$\tau_j = \frac{h_j}{1 + v_j / u_c} \tag{2.1}$$

$$x_j^* = v_j \cdot \tau_j \tag{2.2}$$

where τ_j = the travel time;

h_j = the headway;

v_j = the vehicle velocity;

u_c = the traffic propagation speed; and

x_j^* = the distance traveled.

The link travel time of the k^{th} vehicle, T_k, can then be estimated by finding the largest N_k to satisfy (2.3),

$$l \geq \sum_{j=k}^{k+N_k} x_j \tag{2.3}$$

$$T_k = p \cdot \tau_{k+N_k+1} + \sum_{j=k}^{k+N_k} \tau_j \tag{2.4}$$

$$p = \frac{(x_{k+N_k+1} + \sum_{j=k}^{k+N_k} x_j^*) - l}{x_{k+N_k+1}^*} \tag{2.5}$$

where l is the length of the link; and

p is a weighting factor.

This model assumes a constant traffic propagation speed through the entire link and thus is not suitable for use in some conditions, where a dramatic change in traffic state occurs within a link (i.e., presence of a traffic queue or delays caused by traffic weavings near a ramp).

Some researchers have made efforts to use both the upstream and downstream detector information for estimating travel times with piecewise constant-speed-based (PCSB) methods (van Grol et al., 1997; Lindveld et al. 2000; and Cortes, 2002), which assume a constant travel speed within the link. Van Lint and van der Zijpp (2003) estimated travel times with a piecewise linear-speed-based (PLSB) model, which is reported to outperform PCSB models in simulated cases. In the PLSB model, the vehicle's in-segment speed is determined by the convex combination of the speeds obtained at the upstream and downstream detectors at the same time as shown below:

$$v(x,t) = v_d(t) + \frac{x - x_d}{x_{d+1} - x_d} v_{d+1}(t) \tag{2.6}$$

where x is the location of the vehicle, $x_d \leq x \leq x_{d+1}$;

$v(x,t)$ is the estimated speed of the vehicle at location x at time t;

d is the detector ID;

$v_d(t)$ is the speed detected at detector d at time t; and

x_d is the location of detector d.

Note that existing piecewise models do not consider traffic propagation relations, which use the detected speeds at the upstream and downstream detectors at the same time to estimate travel times in short segments (i.e., 0.5 miles).

In summary, many studies use the trajectory-based models to estimate vehicles' in-segment speeds, and thereby compute their travel times. This type of method is relatively applicable to long links and can better tolerate detector errors than the flow-based models. With proper modifications, this type of model has the potential for use on segments with non-uniform travel speeds.

2.3 Travel Time Prediction Models

Predicting travel times usually requires a longer prediction horizon than predicting traffic variables (i.e., flow and speed), because the information of travel times will not be available until vehicles departing at the current time complete their trips. Researchers over the past decades have attempted to implement both parametric models and nonparametric models to forecast travel times and other traffic variables. Among parametric models, time-series models and Kalman filter models have received more attention than other model structures. Some researchers have also devoted considerable attention to Neural Network models, one of the nonparametric prediction models, due to their well-known learning and pattern recognition abilities and their robust performance. The following section will review existing works on travel time prediction and other related forecasting models, including Neural Network models and other nonparametric models. This section will also discuss some attempts made by researchers to combine two or more models.

2.3.1 Parametric Models

Among parametric models, time-series models are widely used in the transportation field for predicting traffic variables, due to the time-series nature of most transportation-related information. Linear regression models and time-varying coefficient models are reported to be efficient as well. Researchers have developed parametric models for travel time prediction, which are mostly for highway systems capable of directly measuring travel times.

Time-Series Models

In the transportation literature associated with travel time studies, the earliest time-series models were developed by Ahmed and Cook (1979) and Levin and Tsao (1980), who predicted traffic volume and occupancy with autoregressive integrated moving-average (ARIMA) models (Box and Jenkins, 1970). Their study showed that ARIMA models outperform simple smoothing methods and historical average values in forecasting single-detector data. They concluded that the optimal form of ARIMA model is site-specific.

Given a time series of data X_t (where t is integer valued and X_t are real numbers), an ARIMA (p, d, q) model has the following standard form (Box and Jenkins, 1970):

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d X_t = (1-\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t \qquad (2.7)$$

where L is the lag operator, $LX_t = X_{t-1}$ for all t>1;

ε_t is the error term, $\varepsilon_t \sim N(0,\sigma^2)$; and

p, d, q are the order of the autoregressive, integrated and moving average parts of the model, respectively.

Due to its strength in capturing the time-series trend, the ARIMA model is widely used to predict traffic volume and occupancy for a single detector location in a highway segment (Oda, 1990; Davis et al., 1991; Hamed et al., 1995; Smith and Demestky, 1996; van der Voort et al., 1996; Ishak and Al-Deek, 2002; Stathopoulos and Karlaftis, 2003).

As reported in the literature, ARIMA models predict mainly the mean values and often fail to deal with large variations due to some congested patterns or incidents. Hence, seasonal ARIMA models have also been developed in various studies (Smith and Demetsky, 1997; Williams et al., 1998; Smith et al., 2000; Chung and Rosalion, 2001; Smith et al., 2002) to take into account the temporal patterns of the traffic data, such as weekly patterns.

Applications of the ARIMA model in predicting travel times (Anderson, 1995; Yang, 2005) are limited to one-link-only cases, based on collected travel times or detector data at both ends. The seasonal ARIMA model has not been reported to implement in practice for travel time prediction.

Due to the complexity in dealing with multiple time-series datasets, time-series models have not been successfully applied to predict travel times for trips that consist of several links. In contrast, nonparametric models are widely seen in this type of application.

Linear Regression Models

There are few attempts in the traffic literature to employ the linear regression model in forecasting travel times. Kwon et al. (2000) developed a linear regression model for travel time prediction in which the independent variables are available occupancy, flow, departure time and day of the week. They reported that their proposed linear regression model performed better than both a regression-tree model and a Neural Network model. However, they did not discuss the most appropriate function forms of the departure time and day of the week.

Due to the high uncertainty of traffic characteristics, it is difficult to fit the entire-day traffic pattern to a global linear regression model. Many studies have tried to divide the data into subsets and then employed different independent variables and/or varying coefficients with a linear regression structure. For example, Danech-Pajouh and Aron (1991) developed a layered statistical approach by first clustering the data and then fitting each group of data to a linear regression model.

Another category of linear models, time-varying coefficient models (TVC), assumes a global linear relation structure between the travel time $T(t)$ and the *status travel time* $T^*(t)$ with time-varying coefficients throughout the day (Zhang and Rice, 2003). The *status travel time* is defined as the time needed for the current departures to complete their trips if traffic conditions remain unchanged and vehicles can maintain their speeds from one detector to its adjacent downstream detector.

$$T(t) = \alpha(t) + \beta(t) \cdot T^*(t)$$

$$T^*(t, \Delta) = \sum_{d=1}^{D-1} \frac{x_{d+1} - x_d}{v_d(t - \Delta)} \qquad (2.10)$$

where $\alpha(t)$ and $\beta(t)$ are time-varying coefficients;

D is the total number of detectors on the segment; and

Δ is the delay caused by data transmission.

It is reported that the time-varying coefficient model can provide reliable travel time predictions under certain traffic conditions with detectors placed 1/3 to 2/3 miles apart (Zhang and Rice, 2003; and Kwon and Petty, 2005).

Despite the reported performance quality, there are two critical issues associated with the time-varying coefficient model that need to be addressed. First, the TVC model ignores day-to-day traffic variations and the spatial distribution of the congestion within each highway segment; therefore, prediction reliability may significantly decrease when the target traffic conditions are significantly different from those in the historical data. Secondly, when detectors are far apart or some in/out flows (i.e., ramps located between two adjacent detectors) interfere with the traffic patterns, a linear relation may not exist between the actual travel time and the *status travel time* originally observed in the data collected from only one site (Zhang and Rice, 2003).

Kalman Filter Models

With its learning ability to update parameters from real-time data, Kalman filter algorithm has been used by some researches in the literature to improve travel time and traffic pattern predictions (Okutani and Stephanedes, 1984; Whittaker et al., 1997; Chien and Chen, 2001; Chien and Kuchipudi, 2003; Chu et al., 2005).

One potential issue associated with the Kalman filter model arises when applying the model to a long segment that has large variations in its travel times. This is due to the fact that actual travel times will be available only after vehicles finish their trips. Thus, the employed Kalman filter model may not have the actual value to update its parameters to contend with a dramatic change in the target time-varying travel time. As a result, the model's prediction performance could be degraded drastically during transition periods.

2.3.2 Neural Network Models

The Neural Network model is one of the most popular nonparametric models reported in the literature on travel time predictions because of its well-known capability of pattern recognition and its robustness. It has been widely applied in many other transportation areas as well (Dougherty, 1995).

A basic, fully connected backpropagation multilayer perceptron (MLP) consists of one input layer, one hidden layer and one output layer. This topology has been implemented to predict travel times or traffic variables in several studies (Clark et al., 1993; Kown and Stephanedes, 1994; Smith and Demetsky, 1994; Park and Rilett, 1999, Zhang, 2000; Huisken and van Berkum, 2003) and has been reported to achieve good performance.

A variety of complex structures for Neural Network models has also been found in the literature, including MLP with a Kalman filter learning rule (Vythoulkas, 1993), time-delay neural networks (TDNN) (Yun et al., 1997; Abdulhai et al., 1999; Lingras and Mountford, 2001), Jordan's sequential networks (Yasdi, 1999), finite impulse response networks (Yun et al., 1997), radial basis function neural networks

(Park et al., 1998), multirecurrent neural networks (Park et al., 1999), modular neural networks (Park and Rilett, 1998), dynamic neural networks (Ishak and Alecsandru, 2004), and partially connected MLP (van Lint, 2002), etc.

Among these complex structures, the TDNN models have received the most discussion in the literature. The basic TDNN model incorporates one tapped delay line in the input layer to better fit the nature of the time-series data (Figure 2.1); therefore, input time-series data items will travel through the tapped delay line to provide the TDNN with a better short-term memory. One can use the backpropagation through time (BPTT) or real-time recurrent learning (RTRL) algorithms to train the TDNN either offline or online. Due to its strong short-term memory unit, TDNN lacks the ability to forget irregular input data. One irregular data point, which may be caused by either highly fluctuating traffic variables or a detection error, will stay and impact the prediction result in the tapped delay line until it reaches the end of the delay line.

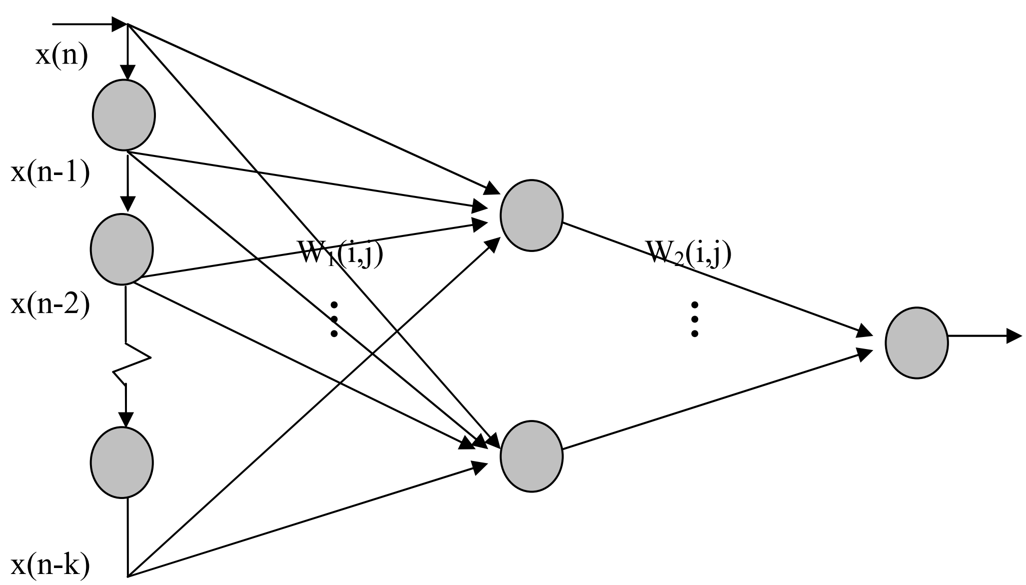


Figure 2.1 Example topology of a Time-Delay Neural Network

Except for the time-delay recurrent Neural Network models, most of such model structures have only been verified with data from one site. The comparison results of Neural Network models with other models are not consistent in the literature. In general, fine-tuning a Neural Network model is always time consuming, but critical to its performance. It may be one of the most significant factors that cause the poor performance of Neural Network models reported in some literature (Smith and Demetsky, 1996; Kirby et al., 1997).

In the literature, many researchers have made considerable attempts to combine Neural Network models with other models to improve prediction reliability; those works will be discussed later in the section of hybrid models.

2.3.3 Other Nonparametric Models

In addition to Neural Network models, various nonparametric models have been applied to forecast travel times, traffic volumes, speeds etc., due to the fact that transportation-related data is often hard to fit in a pre-specified model structure. Commonly used nonparametric models in this area include *k*-Nearest Neighbor models, kernel models, and local regression models.

Most nonparametric models for travel time prediction share a common feature — that is, to search a collection of historical observations for one or more records that are similar to the system's current state and use such data to perform the prediction. Two classes of nonparametric models, kernel models (Nadaraya, 1964; Priestley and Chao, 1972; and Watson, 1964) and *k*-Nearest Neighbor models (Benedetti 1977; Stone, 1977; Tukey, 1977), are widely used (Altman, 1992), especially in the transportation literature (Davis and Nihan, 1991; Smith and Demetsky, 1996; Smith and Demetsky, 1997; Smith et al., 2000; Oswald et al., 2001; Clark, 2003; and Rice and van Zwet, 2004). In the literature, some efforts have been made to use the local regression models (Cleveland, 1979; Cleveland and Devlin, 1988; Hastie and Loader, 1993; and Fan and Gijbels, 1996) on forecasting as well (Sun et al., 2003; and Sun et al., 2004).

A nonparametric model usually consists of three components, including a historical database, a search or classification procedure, and a forecast function (Oswald et al., 2001). With different forms of search/classification procedures and forecast functions, the following three types of nonparametric models are available in the literature: *k*-Nearest Neighbor, kernel and local regression models. A brief description of each model is presented below.

k-Nearest Neighbor Models

In a *k*-Nearest Neighbor model, a set of *K* variables is first determined in the search procedure to describe the system state. The similarity between two records, historical record *p* and the current case *q*, can be defined as their Euclidean distance, $dist_{EUC}(p,q)$:

$$dist_{EUC}(p,q) = \sqrt{\sum_{i=1}^{K}(p_i - q_i)^2} \tag{2.7}$$

where p_i is the value of the i^{th} variable in the historical record; and

q_i is the value of the i^{th} variable in the current state.

Nonuniform weighting factors, w_i, can also be used to define the distance between two records such as in (2.8).

$$dist_{NUW}(p,q)=\sqrt{\sum_{i=1}^{K} w_i (p_i - q_i)^2} \tag{2.8}$$

Other forms of distance, for example Manhattan distance and max distance, have also been used in the literature (Oswald, 2001).

In the forecast function, the *k*-Nearest Neighbor model takes the average of the top *k* nearest neighbors as the prediction result $\hat{V}$:

$$\hat{V}=\frac{1}{k}\sum_{i=1}^{k} V_i \tag{2.9}$$

where V_i is the future value in the i^{th} historical match.

This type of forecast function is available in most transportation-related applications of nonparametric models.

As reported in the literature, *k*-Nearest Neighbor models are capable of providing reliable predictions in many transportation-related literatures (Davis and Nihan, 1991; Smith and Demetsky, 1996; Smith and Demetsky, 1997; Smith et al., 2000; Oswald et al., 2001; Clark, 2003; and Rice and van Zwet, 2004). However, the results of performance comparisons between *k*-Nearest Neighbor models and other prediction models vary with differences in their applications.

Another form of forecast function includes weighting factors that are usually proportional to the distance between two sets of data. Smith et al. (2000) proposed various weighting schemes for traffic condition forecasting.

Nonparametric Kernel Regression Models

With the classification function in a nonparametric model, one can apply a kernel function (i.e., linear, polynomial or radial basis function [RBF]) as the forecast function to a subset of data for predicting future values.

Faouzi (1996) predicted traffic variables by kernel regression. As reported by Sun et al. (2003), one must make additional efforts to avoid frequent outputs of zero when applying the kernel regression model to a small database or in an application with frequent irregular data points. With a support vector machine (SVM) serving as the classification procedure, Wu et al. (2004) applied various kernel functions and produced reliable predictions on travel times for three long segments of between 45 km and 350 km in distance.

Local Regression Models

The local regression model (Cleveland, 1979; Cleveland and Devlin, 1988; Hastie and Loader, 1993; and Fan and Gijbels, 1996) combines the simplicity of linear regression models and the flexibility of nonparametric models to fit a local segment of a dataset without a global function. As reported by Müller (1987), nonparametric local linear regression and nonparametric kernel regression are equivalent for regularly distributed data. However, local regression models can better handle the irregular distributed data often seen in transportation applications; therefore, they are more reliable than kernel regression models in a single-model system.

Similar to the time-varying coefficient models, a local regression model determines its data subsets by the distance of the covariates' spaces, usually with a Nearest Neighbor model, instead of the departure times used in TVC models. Sun et al. (2003) applied the local linear regression model to predict traffic speed at one detector location. It is reported to achieve some improvements by incorporating an empirical bootstrap method (Sun et al., 2004). The prediction results are reported to be reliable when the prediction horizon is short (i.e., 5 to 15 minutes).

Care must be exercised in determining two critical parameters for the nonparametric models: the number of input variables and the bandwidth of the search/classification procedure. Fan and Gijbels (1996) suggested using the basic cross-validation approach to determine these two parameters. However, such a method may not work efficiently for travel time prediction, which usually has a large amount of available data from multiple traffic detectors in a large time horizon. Analyzing other related information — for example, segment geometry and historical traffic patterns — may help to determine the optimal values of these critical parameters.

In the scenario where insufficient good matches are found in the historical database, the nonparametric model may fail to output a reliable prediction. This type of case exists in almost every travel time prediction system. Therefore, at least one alternative method is required to ensure the reliability of a travel time prediction system that utilizes a nonparametric model in order to deal with such situations.

2.3.4 Hybrid Models

Another type of forecasting, usually referred as a hybrid method, involves the use of multiple models. Similar to the nonparametric approaches, hybrid methods generally incorporate a clustering approach and then assign one model structure to each cluster with locally fitted parameters. Related studies for forecasting traffic volume, speed or occupancy are available in the literature by Danech-Pajouh and Aron (1991), van der Voort et al. (1996), Abdulhai et al. (1999), Chen et al. (2001), Lingras and Mountford (2001), Yin et al. (2002), Ishak and Alecsandru (2004), Zheng et al. (2006), etc. Among the aforementioned hybrid models, those combining the Neural Network model with a clustering model or an improved learning model seem to show more potential than the others.

In predicting travel time, some other hybrid models have also been reported in the literature. You and Kim (2000) proposed a combination of nonparametric model and machine learning to improve the accuracy of travel time predictions. Kuchipudi and Chien (2003) developed a travel time prediction system that switches between a path-based prediction model and the link-based prediction model using the Kalman filter algorithm.

The most important technical issue associated with the use of hybrid models is the clustering criteria. Genetic algorithm (GA) and other data-driven methods have been reported in the literature. However, due to the impacts of site-specific factors such as geometry features, regional traffic patterns and driving behaviors, it is often difficult to have a generalized set of calibration procedures for such models to various locations.

2.4 Experimental and Real-world Application Systems

Many experimental systems have been implemented worldwide to provide the travel time information for commuters. Efforts have also been made to develop simulated systems in laboratory environments with data from actual traffic detectors. A review of both types of system is reported in this section, with the focus on detector distribution, method to obtain travel times, and travel time prediction model.

2.4.1 Simulated Systems

Kwon et al. (2000) developed and tested a travel time prediction system for peak hours with data (flow and occupancy) from 19 detectors in each direction of a 10-km segment of freeway. Detector data was first redistributed to ten equidistance virtual detector stations with interpolation. Missing values were estimated by a simple interpolation method to construct the dataset for model training and evaluation. Four traffic scenarios were identified by traffic direction and morning/evening peak hours to cluster the dataset. Two candidate prediction models, a tree method and a linear regression model, were trained with about 200 data points in each subdataset. A cross-validation test showed that both prediction models provided reliable travel time predictions with a prediction horizon of less than 20 minutes in the morning, while the prediction results in two afternoon datasets were not as expected.

The system by Rice and van Zwet (2004) was based on traffic data (flow occupancy) collected from 116 detectors over a freeway segment of 48 miles, where the missing data was estimated with interpolation. Traffic speeds were computed from flow and occupancy information using a method suggested by Jia et al. (2001) to estimate travel times and serve as model inputs. It is reported that the proposed time-varying coefficient model outperformed the historical average method and a *k*-Nearest Neighbors (*k*=2) model.

Chen et al. (2003) developed a travel time prediction system similar to that of Rice and van Zwet (2004) on two 20-mile two-way freeway segments, one having 135 detectors and the other one with 120 detectors. A trajectory-based travel time estimation method was used to estimate historical travel times for model training. By comparing with the data from probe vehicles, they found some large errors in evaluation because of missing data, a severe incident and other unknown reasons.

Shien and Kuchipudi (2003) developed two Kalman filter models based on data collected from electronic toll devices on a 17-mile segment. The time periods with low detection rate were filled with historical average data. The performance of the link-based model and the path-based model was reported to vary under different scenarios.

Note that all of the aforementioned simulated systems were developed based on prefiltered datasets without missing or faulty data.

2.4.2 Real-world Systems

Over the past decades, several real-time travel time display systems have been implemented worldwide. Some systems display travel times to roadside or overhead variable message signs (VMS), and others have web-based output interfaces.

TranStar in Houston, TX, USA, collects travel times from nearly two million EZ-Tags and posts the average travel times from these completed trips onto dynamic message signs (DMS) in real time (http://traffic.houstontranstar.org).

The travel time system in Chicago, IL, USA (Illinois State Toll Highway Authority, 2005), is based on two sources of travel time estimations: travel times computed from electronic toll readers and those estimated from traffic detectors. When more than one source is available, one type of data will be chosen based on operational experience and judgment.

Several states have used the Georgia Navigator software to display the travel times computed by the current average speeds collected from each link, including Atlanta and Macon, GA, USA (http://www.georgia-navigator.com/trips), Portland, OR, USA (Oregon Department of Transportation, 2005), and Nashville, TN, USA (Tennessee Department of Transportation, 2005). Such systems generally will be shut down if no data is reported from one detector station for a period of time.

Washington State Department of Transportation, USA, determines travel times with the current speeds computed by detected flow and occupancy information from detectors at an average spacing of 0.5 miles (http://www.wsdot.wa.gov/Traffic/seattle/questions/traveltimesdetail.htm).

Similar systems have also been implemented in the United Kingdom, the Netherlands, and Japan. However, most of these systems provide travel times with simple estimation or prediction algorithms. No report of incorporating advanced algorithms for filtering and estimating missing data has been found in these actual systems.

2.5 Conclusions

This chapter reviewed the existing approaches for travel time estimation and prediction, including some simulated and real-world travel time prediction systems.

Among the three types of travel time estimation models, the flow-based models, which need high accuracy of detector data and uniform geometric features, are the least applicable for use in a real-world system. Vehicle identification models need new detection hardware or take raw detector signals as input and therefore may incur high system costs and the need for a large data transmission bandwidth.

In contrast, the trajectory-based model for travel time estimation is relatively promising, since it has the potential to fit with long segments and more complex geometric features.

Overall, nonparametric models are able to provide more reliable travel time predictions than parametric models in a single-model system structure. Hybrid models are reported to be able to further improve prediction reliability.

In conclusion, to advance the existing models for real-world applications, one must overcome the following critical issues:

- A travel time estimation model shall be able to deal with all types of geometric features and traffic patterns when the direct measurement of travel times is not available;
- A travel time prediction model shall function reliably under both commonly seen traffic conditions and less frequently observed traffic patterns;
- A real-time missing data estimation model is needed to improve the system's reliability; and

- The system needs to have a monitoring function that can identify situations where reliable predictions cannot be provided due to model limitations and/or missing data.

Chapter 3: The Architecture of a Reliable Travel Time Prediction System with Sparsely Distributed Detectors

3.1 Introduction

As is well recognized, densely distributed traffic detectors can help travel time prediction systems achieve high reliability. The literature review has shown that there lacks the study on developing models for a freeway segment with sparsely distributed detectors, as most existing works are based on the detection spacing of 0.5 miles. The costs of detector purchase, installation, communication and maintenance constitute the majority of the system costs. Therefore, the lower the number of detectors needed to cover the targeted freeway segment for travel time prediction, the more likely for the responsible agency to deploy such a system.

Since travel time information is sensitive to the public, a system using fewer traffic detectors still needs to (1) build a reliable historical travel time database even without direct measurements of travel times; (2) take commonly available data from various types of traffic detectors for better system compatibility; (3) estimate missing or delayed data to extend the system's reliability; and (4) identify the conditions in which a reliable prediction may not be attainable during real-time operations.

The flowchart for system operations, along with the introduction of each principal component and their interrelations, will be described in the rest of this chapter.

3.2 System Flowchart

The proposed system architecture aims to provide reliable travel time prediction using sparsely distributed detectors. The system comprises four principal components: a travel time estimation module, a travel time prediction module, a missing data estimation module, and an incident detection module. The proposed system has two operational stages: the model-training stage and the real-time operation stage. The operational flowcharts for these two stages are briefly presented below.

3.2.1 Model-training Stage

Figure 3.1 shows the system's operational flowchart for the model-training stage. Before the proposed travel time prediction system can start to operate, one must take the following five steps to calibrate all system parameters and construct the historical travel time database.

Step 1: Calibrate all detectors to a reliable state

This step is essential to all intelligent transportation systems that take data from traffic detectors. Without proper calibration, an unreliable detector can significantly degrade the system's reliability.

Step 2: Long-term collection of traffic data

In the model-training stage, the system needs to collect long-term traffic data for training models and constructing the historical travel time database for its on-line operation. For better system performance at the real-time operation stage, the travel time prediction module also uses the information from the weekly traffic patterns. Therefore, it needs to have a fairly long data collection period to make sure that a sufficient number of samples are available for each weekday. For example, a continuous three-month data collection period will yield about 12 to 14 samples for each weekday.

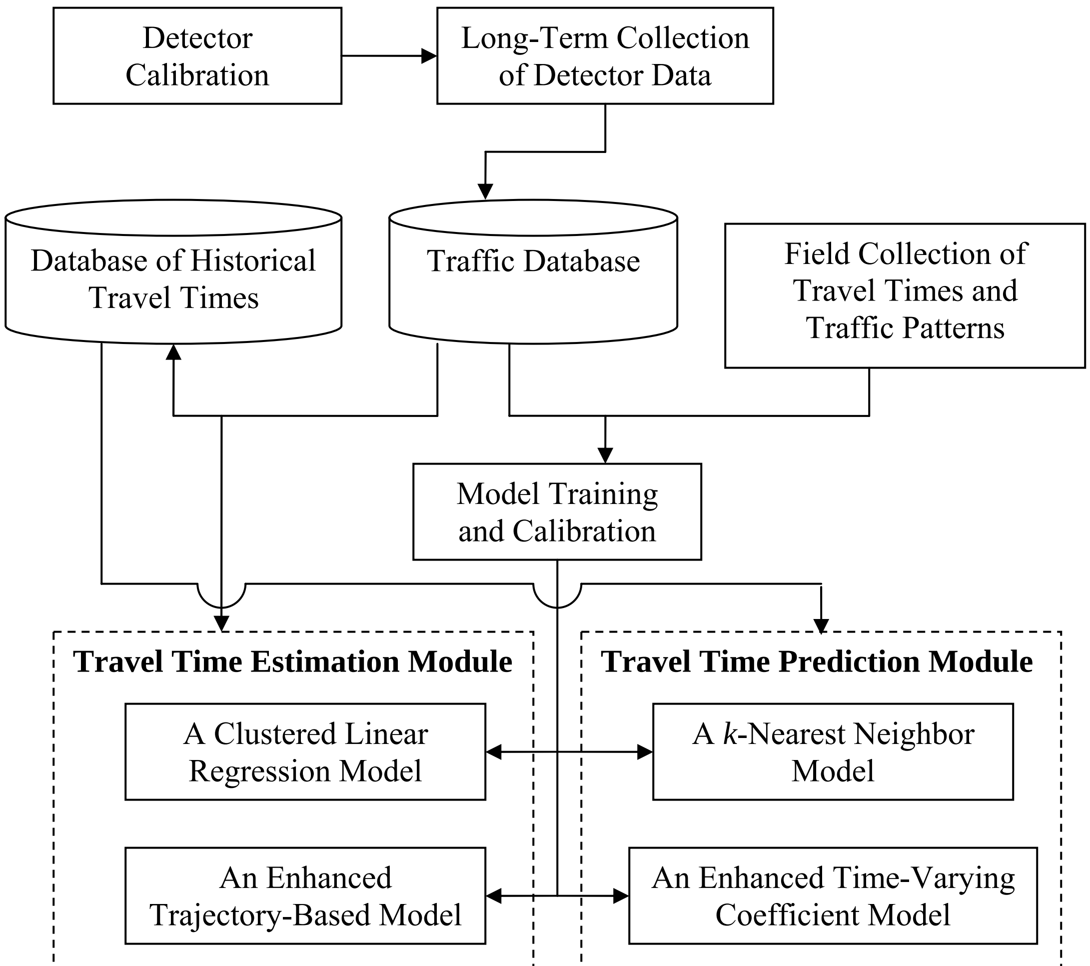


Figure 3.1 System flowchart for the model-training stage

Step 3: Collection of traffic patterns and actual travel times

The proposed system needs information about recurrent traffic patterns to determine the critical lanes before training its model components. The travel time estimation module also requires actual travel time information to train its clustered linear regression model and calibrate its enhanced trajectory-based model. Actual travel times can also help evaluate the actual performance of the travel time prediction module, which is based on estimated travel times.

Step 4: Parameter calibration for the travel time estimation module

The main model of the travel time estimation module, a clustered linear regression model, requires sufficient actual travel times in each cluster to determine its best fit coefficients. The supplemental model, an enhanced trajectory-based model for travel time estimation, does not require actual travel times for calibration, but requires actual speed information to construct the occupancy-speed relations.

Step 5: Construction of the historical travel time database

Once the travel time estimation module has been properly trained and calibrated, one can apply it to the long-term collected set of traffic data to construct the historical travel time database, which is used to support the travel time prediction module.

Step 6: Parameter calibration for the travel time prediction module

In the hybrid model structure of the travel time prediction module, the *k*-Nearest Neighbor model requires the analysis of the historical traffic patterns in critical lanes to determine its parameters. The training process for the supplemental model, an enhanced time-varying coefficient model, is based on the data in both the traffic database and the historical travel time database.

After the entire training process is completed, the proposed travel time prediction system is ready for real-time operations.

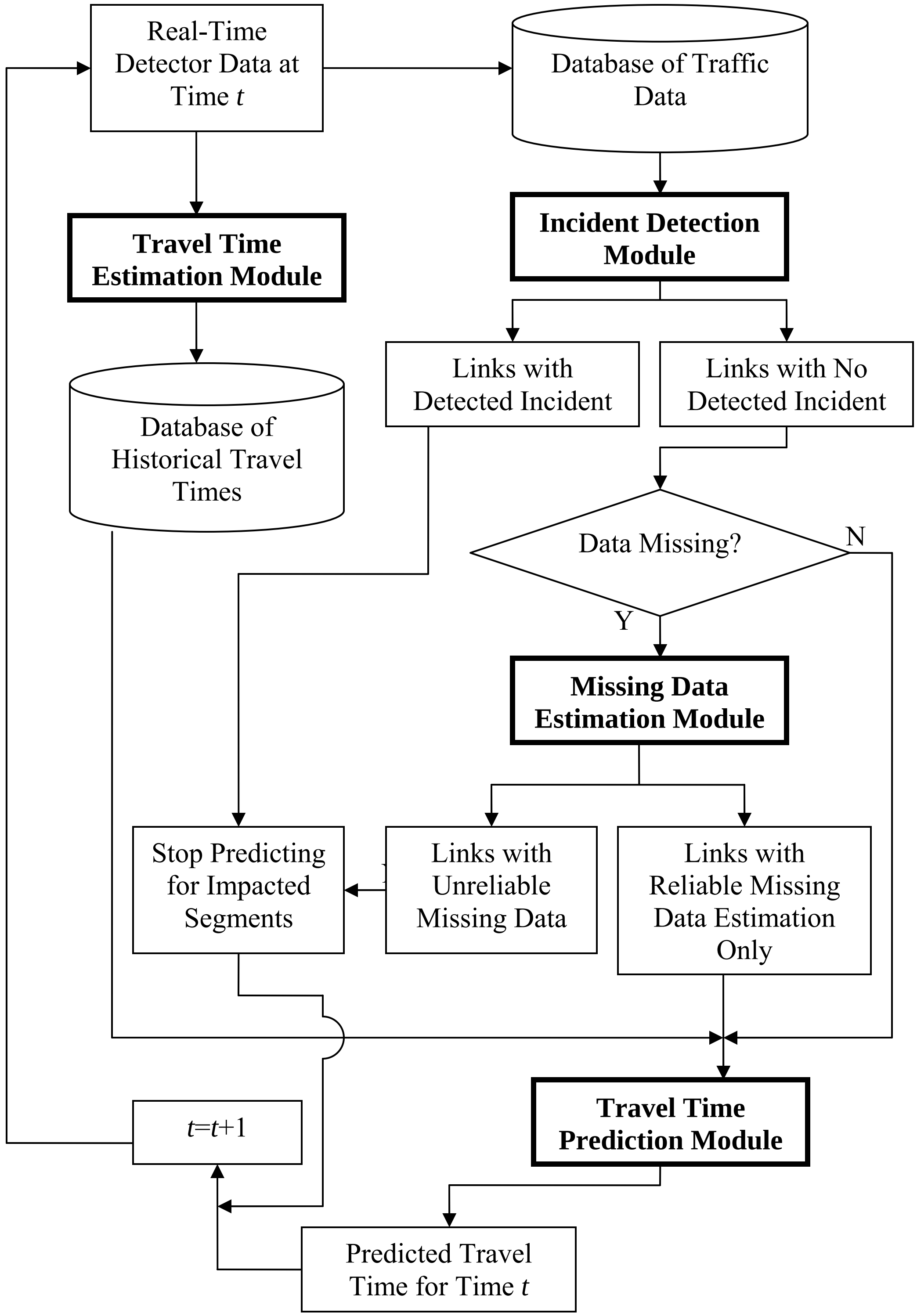


Figure 3.2 System operational flowchart for the real-time operation stage

3.2.2 Real-time Operation Stage

Figure 3.2 shows the operational flowchart of the proposed travel time prediction system at the real-time operation stage.

The entire real-time operation consists of the following steps.

Step 1: Data acquisition

At time *t*, the system will receive the real-time data from all detectors and then store them in the traffic database.

Step 2: Incident detection

The proposed travel time prediction system will first apply its incident detection module to the traffic data. If one or more incidents are detected, the system will inform the control center and stop the predictions for those segments plagued by detected incidents.

Step 3: Missing data estimation

The missing data estimation module will perform a test on those links experiencing no incident and evaluate if any required input data is missing, and then execute the missing data estimation if needed. If the module detects that data missing on one or more links cannot be reliably estimated at the current time, it will then notify the system to stop the prediction of travel times on those segments.

Step 4: Travel time prediction

The travel time prediction module, which has a hybrid model structure, will provide travel time predictions for segments that do not experience incidents or unreliable missing data from traffic detectors.

Step 5: Update of the database of historical travel times

The travel time estimation module will take the most recent available detector data to estimate the travel times of completed trips. The information of the most recently completed trips will be available immediately for use by the travel time prediction module in the next time interval.

The proposed travel time prediction system will then repeat the same process from Step 1 for the next time interval.

3.3 Principal Functions of System Modules

As discussed above, the proposed travel time prediction system consists of four principal modules: a travel time estimation module, a travel time prediction module, a missing data estimation module, and an incident detection module. The following section will briefly describe the key function of each module.

3.3.1 Travel Time Estimation Module

The travel time estimation module will estimate travel times from detector data and update the historical travel time database when there is no direct measurement of the travel time available in the system. To ensure the system's high compatibility, this module shall be capable of receiving data from any commonly used traffic detector. In order to achieve high reliability with fewer detectors, the proposed system will best use the information of geometric features and common traffic patterns to perform travel time estimation.

To contend with inevitable data deficiencies, the proposed travel time estimation module employs a hybrid model structure. The main model, a clustered linear regression model, is used to provide estimated travel times for traffic scenarios that have been frequently observed. In contrast, an enhanced trajectory-based model will serve as the supplemental model, designed to deal with scenarios that lack sufficient field data for model calibration. In real-time operations, the travel time estimation module will concurrently estimate travel times from all completed trips and store them in the database for use by the travel time prediction module.

3.3.2 Travel Time Prediction Module

Similar to the travel time estimation module, the main input variables of the travel time prediction module shall be readily available from most existing traffic detectors. The proposed module employs a hybrid model structure that integrates one *k*-Nearest Neighbor model with an enhanced time-varying coefficient model to improve prediction accuracy.

With the improved *k*-Nearest Neighbors model for travel time prediction, the system can take full advantage of historical travel times and on-line detected traffic conditions. With an improved searching function, the *k*-Nearest Neighbors model can best match the detected traffic conditions with those in the historical data set, based on characteristics of traffic patterns and geometric features of the target segment. An enhanced time-varying coefficient model will serve as the supplemental model in the hybrid model structure for travel time prediction. It functions to capture the relations between the daily and weekly time-dependent travel time patterns.

3.3.3 Missing Data Estimation Module

Missing data is a critical issue that often plagues any on-line system. Most models developed in the literature are based on an assumption of no missing data. Missing just one item in the critical data stream may prevent the system from functioning properly. The proposed travel time prediction system contains a missing data estimation module to deal with the missing and/or delayed data that frequently occurs due to detector malfunctions and/or communication problems.

The missing data estimation approaches in this module are developed specifically to fit the hybrid model structure used in the travel time prediction module and can evaluate the reliability of the estimated missing data. If the estimated missing data may significantly degrade the prediction quality, the proposed travel time prediction system will suspend the prediction on the affected segments until reliable data becomes available.

3.3.4 Incident Detection Module

Note that most travel times during nonrecurrent congestion (such as an incident) vary with a number of response- and operations-related factors, and most prediction systems are designed for recurrent daily congestion only. However, to determine when to switch to the incident report mode, the proposed travel time prediction system needs to have an incident detection module that can employ different algorithms in parallel to ensure its detection accuracy on long links. In real-time operation, this module will evaluate results from all algorithms and then make a final decision. The system will not offer predicted travel times for segments that contain links plagued by detected incidents.

3.4 Summary

This chapter presented the operational flowcharts of the travel time prediction system with sparsely distributed detectors at both the model-training and the real-time operation stages. To contend with many technical and compatibility issues, the proposed system is designed to consist of four main modules: a travel time estimation module, a travel time prediction module, a missing data estimation module, and an incident detection module. The travel time estimation module estimates travel times from detector data to construct the historical travel time database and then continuously update that database in real-time during operations. The travel time prediction module takes real-time traffic data from the detectors and from the historical database to predict travel times for different destinations. The missing data estimation module is designed to estimate missing and/or delayed data in real time so as to avoid system interruption. The incident detection module is used to detect incidents during real-time operations and prevent the system from providing unreliable predictions under nonrecurrent congestions such as incidents.

Chapter 4: A Hybrid Model for Reliable Travel Time Estimation on a Freeway with Sparsely Distributed Detectors

4.1 Introduction

As is well recognized, travel times are essential information for traffic controls, operations, transportation planning, and advanced traveler information systems (ATIS). Several measurement methods have been used in practice to estimate travel times, including probe vehicles, vehicle identification with in-vehicle devices (i.e., electronic toll tags), and vehicle identification without in-vehicle devices (i.e., video-based vehicle identification and license plate recognition). However, due to the limited sample sizes the probe vehicle method can provide and the high costs associated with both types of vehicle identification methods, it is not cost-effective for any responsible agency to sustain ATIS operations with those methods.

With recent advances in vehicle detection technologies, more and more studies emerge to provide better estimates of travel times using new traffic detectors, which can provide reliable measurements of cumulative traffic flows and occupancy for any prespecified time interval. As reported in the literature, most existing models for travel time estimation are developed and tested for short links (i.e., detectors placed less than 0.5 miles apart). These models may not work properly on long links due to the fact that their embedded assumptions may not be valid when detector spacing is longer than 0.5 miles, as in most existing highway systems. In this chapter, all critical issues associated with travel time estimation on long links will be discussed in Section 4.2, followed by the introduction of input variables and other available information for the proposed hybrid travel time estimation module in Section 4.3. Sections 4.4 and 4.5 will present two proposed model structures: a clustered regression model and an enhanced trajectory-based model.

4.2 Challenges in Estimating Travel Times on Long Links

In review of the literature, it is clear that providing a reliable estimate of travel times remains a challenging task, especially for highway segments with long detector spacing (e.g., > 0.5 miles). Some critical issues associated with travel time estimation are discussed below.

Spatial distribution of the congestion patterns

Despite the tremendous efforts made by traffic flow researchers over the past decades in modeling the evolution of congestion patterns, it remains quite difficult for any existing method to reliably estimate or predict the propagation of traffic patterns under both recurrent and nonrecurrent congestion patterns. A failure to capture the temporal and spatial distributions of traffic patterns will actively degrade the quality of any model for travel time estimation or prediction.

Impacts of geometric features

Changes in geometric features often result in different roadway capacity and traffic patterns. Example congestion patterns incurred due to changes in freeway geometric features are summarized below:

- *Lane drop*

Figure 4.1 shows an example of traffic conditions commonly seen near a lane drop point. During congested periods, traffic conditions in four sub segments, A to D, could evolve from a uniform condition to a chaotic state by frequent lane changes and accelerations/decelerations and then move back to a steady state after the merges.

- *Lane addition*

By the same token, traffic conditions as shown in Figure 4.2 may go through a similar evolution process from A to C.

- *On-ramp/off-ramp*

Figure 4.3(a) and (b) show possible traffic conditions near an off-ramp and an on-ramp, respectively. Due to their local knowledge of possible delays and congestions caused by weaving traffic near a ramp, drivers may avoid using the through lane next to the ramp. Figure 4.4 illustrates an example of congestion caused by this phenomenon in two through lanes on I-70 near Exit 87A to US29 southbound (Figure 4.5). One needs to carefully analyze the discrepancy of traffic flow speeds between lanes to estimate the average speed within one segment.

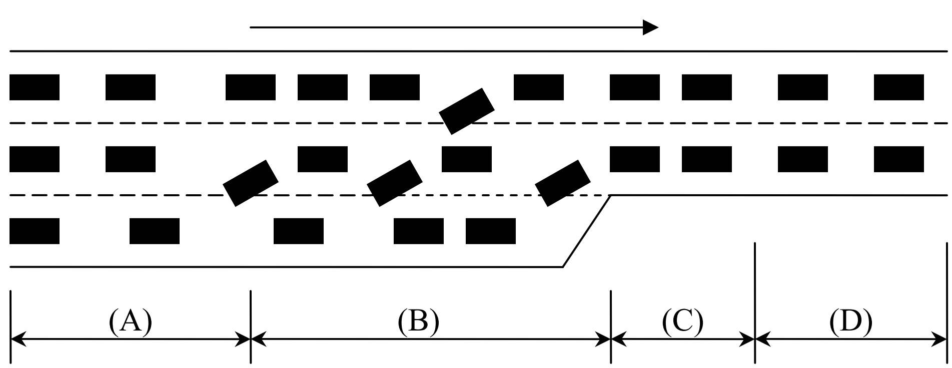


Figure 4.1 Congestion pattern near a lane drop point

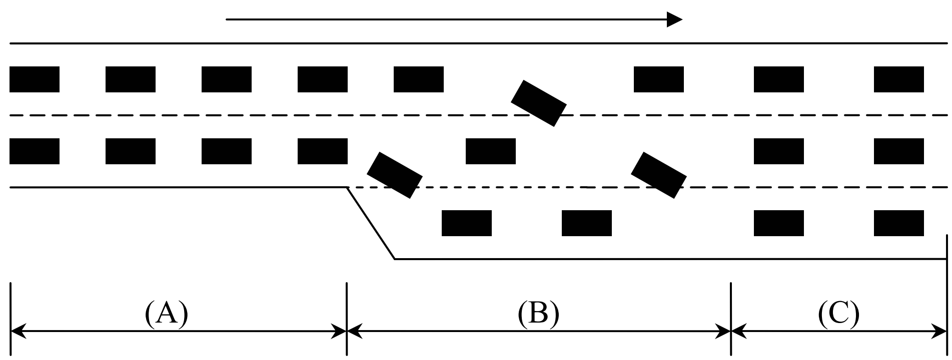
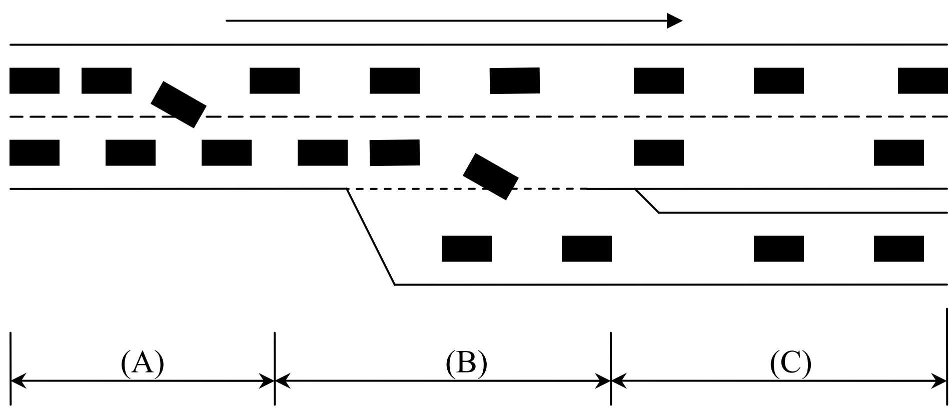
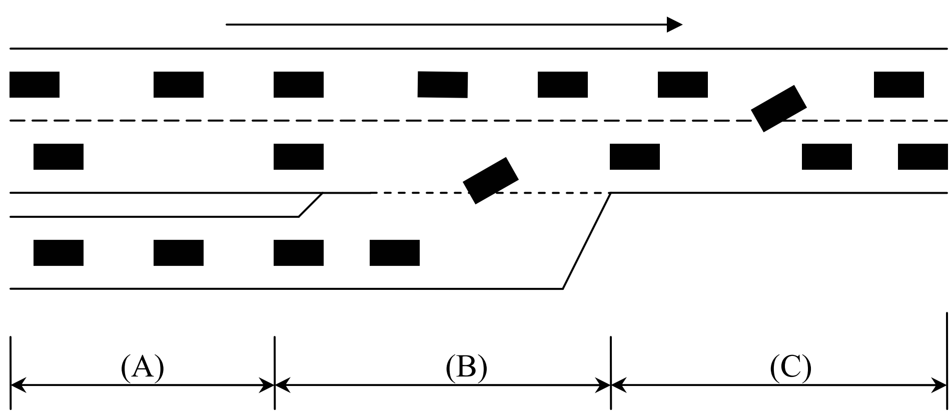


Figure 4.2 Congestion pattern near a lane addition point



(a)



(b)

Figure 4.3 (a) Congestion pattern near an off-ramp;
(b) Congestion pattern near an on-ramp

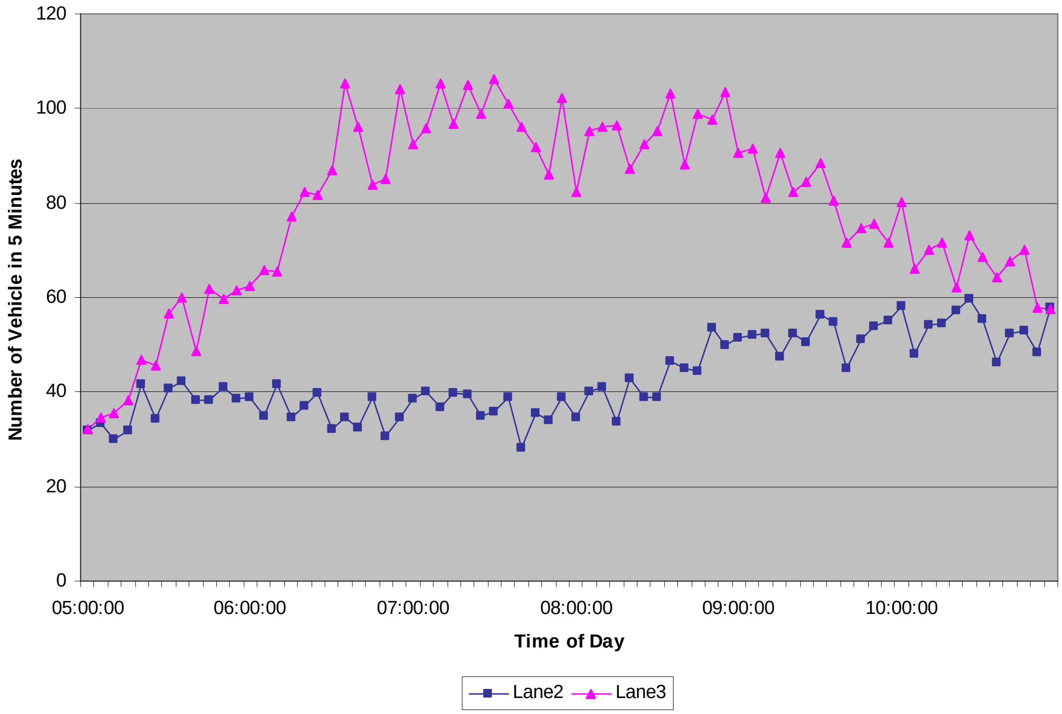


Figure 4.4 Average vehicle counts in 5-minute intervals on four Thursdays in July, 2006 at Exit 87A on I-70

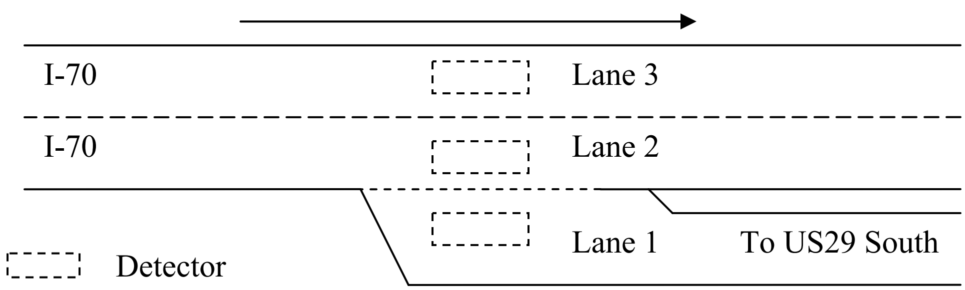


Figure 4.5 Geometry of I-70 at Exit 87A

Other Factors

Aside from the aforementioned factors, the traffic flow patterns and the resulting travel times may also vary with the low visibility caused by weather or sun glare or with poor road surface conditions caused by rain, snow or debris. Quantifying the impacts of those factors, however, has not yet been reported in the literature and is beyond the scope of this study, too.

4.3 A Hybrid Travel Time Estimation Model

This study develops a hybrid model for reliable travel time estimation for long freeway links with widely spaced detectors. This section will present a flowchart of the model and will describe the required input variables.

Flowchart of the Hybrid Model

Figure 4.6 shows the flowchart of the proposed hybrid model, which consists of two main components: a clustered linear regression model and an enhanced trajectory-based model. When applying the hybrid model, the system will first cluster traffic scenarios into predefined categories based on the traffic data. The system will employ the linear regression model if the detected traffic scenario belongs to a category in which a linear regression model has been trained with a sufficiently large sample of historical travel times. Otherwise, it will employ the enhanced trajectory-based model, which does not require pretraining with a large amount of historical data, to produce the travel time estimation.

Model input and available information

As mentioned in Section 3.2.1, both components in the proposed hybrid model employ the cumulative traffic volume and average occupancy in each lane over fixed-length time intervals as the main input variables. Other variables that are collectable with reliable quality are also included in the model development, including roadway geometric features, common daily and weekly traffic patterns, and free-flow travel times. The definitions of variables used to develop the model can be found in Appendix A.

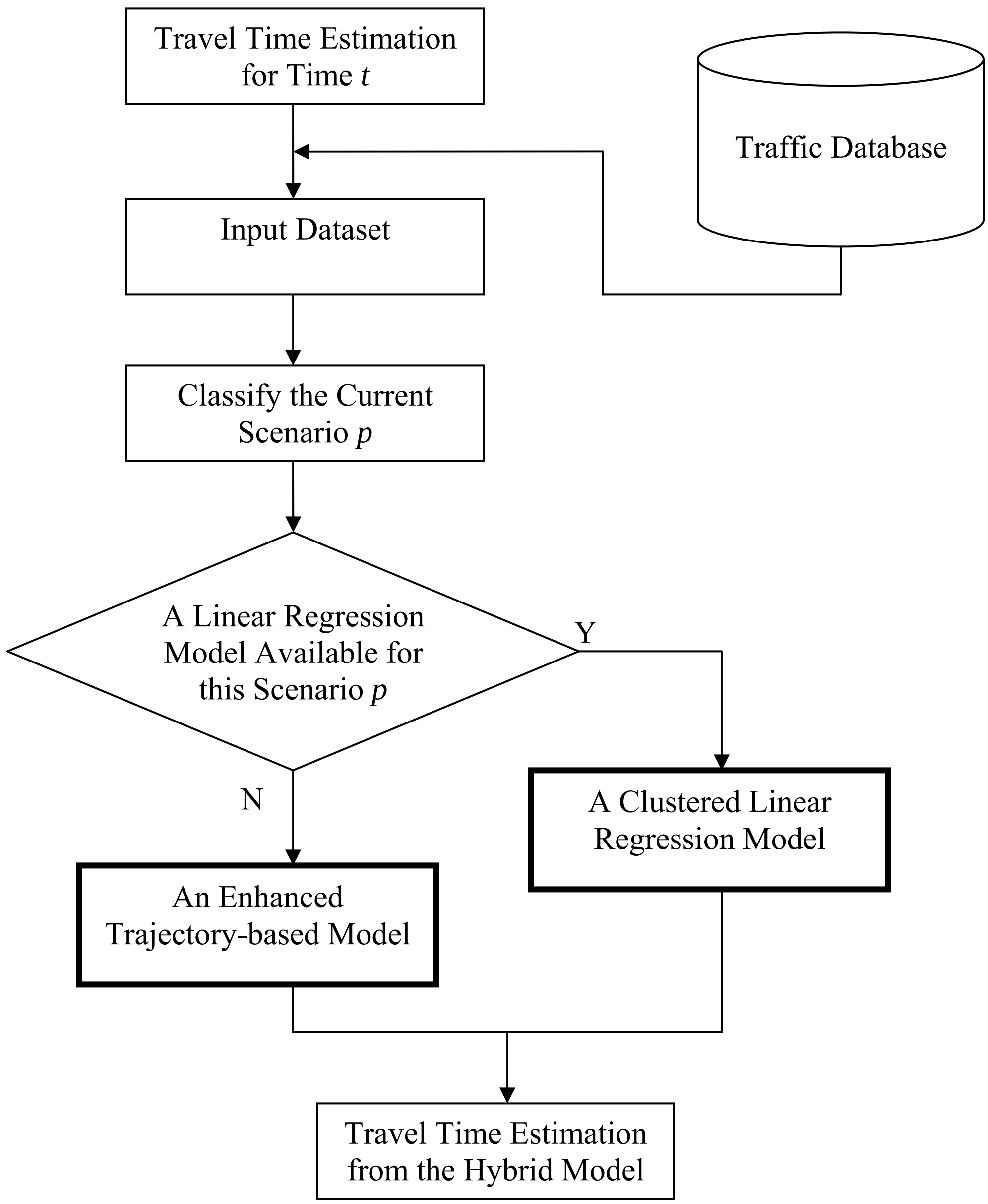


Figure 4.6 Flowchart of the hybrid travel time estimation model

4.4 Clustered Linear Regression Models

When a vehicle is traveling in a link, the range of possible travel times is usually constrained by the traffic pattern. For example, a vehicle can never reach free-flow travel time when there is heavy congestion in the link. Hence, this study first develops a set of clustered linear regression models to categorize traffic conditions into predefined traffic scenarios and then estimates a travel time for each scenario.

4.4.1 Model Formulations

By dividing a link into two equal-length sublinks, one can express a vehicle's travel time as follows:

$$\tau_d(t) = \tau_d^1(t) + \tau_d^2(t) \tag{4.1}$$

where $\tau_d^j(t)$ is the travel time for the vehicle to traverse the first or second half of the link (d, d+1) with departure time t.

Denoting $\bar{u}_d^j(t)$ as the average travel speed in the j^{th} half, one can rewrite Eq. 4.1 as:

$$\tau_d(t) = \frac{L_d}{2\bar{u}_d^1(t)} + \frac{L_d}{2\bar{u}_d^2(t)} \tag{4.2}$$

Coifman (2002) estimated a vehicle's in-segment speeds from the upstream detector data after the departure time, or from the downstream detector data before the vehicle's arrival time, to obtain a travel time estimation. To improve the model's robustness for long segments (e.g., > 0.5 miles), this study assumes a linear relation between a vehicle's average in-segment speed and the average speed of the upstream or downstream through traffic during the same time interval, as follows:

$$\tau_d(t) = \frac{L_d}{2(a_{11}\hat{u}_d^{Thru}(t, \tau_d^1(t)) + a_{12})} + \frac{L_d}{2(a_{21}\hat{u}_{d+1}^{Thru}(t + \tau_d^1(t), \tau_d^2(t)) + a_{22})} \tag{4.3}$$

where a_{ij} are coefficients.

On the right side of Eq. 4.3, the first term is the travel time for a vehicle to traverse the first half of the link (*d*, *d*+1); the second term is for the second half of the link. Similar to the model developed by Liu et al. (2006), Eq. 4.3 has unknown variables on both sides. Liu et al. (2006) provided an iteration-based solution algorithm to solve their problem, which seemed to work well in a simulated traffic environment. However, the performance of their solution algorithm is conditioned on the quality of detector data, which is often undesirably poor in real world systems. Hence, this study uses a preliminary estimate of the travel time to replace the actual travel time information in the independent variables to achieve better robustness. More specifically, assuming that traffic conditions in a link (*d*, *d*+1) can be divided into *P* scenarios with a relatively small range of travel times in each scenario, one can then replace the actual travel time information in independent variables in Eq. 4.3 with a preliminary estimate of travel time for this scenario to obtain Eq. 4.4:

$$\tau_d(t) = \frac{L_d}{2(a_{11}^1\hat{u}_d^{Thru}(t, \gamma_p^d \tau_d^E(p)) + a_{12}^1)} + \frac{L_d}{2(a_{21}^1\hat{u}_{d+1}^{Thru}(t + \gamma_p^d \tau_d^E(p), (1-\gamma_p^d)\tau_d^E(p)) + a_{22}^1)} \tag{4.4}$$

where p is the index of predefined traffic scenarios in link (*d*, *d*+1);

$\tau_d^E(p)$ is the preliminarily estimated travel time in link (*d*, *d*+1) under the p^{th} predefined traffic scenario;

γ_p^d is the estimated proportion of time taken for the vehicle to traverse the first half of the link (d, d+1) under the p^{th} scenario; and

a_{ij}^1 are coefficients.

(4.4) can be reorganized as:

$$\begin{aligned}\tau_d(t) = {} & a_{11}^3 \frac{1}{\hat{u}_d^{Thru}(t, \gamma_p \tau_d^E(p))} \\ & + a_{12}^3 \frac{1}{\hat{u}_d^{Thru}(t + \tau_d^E(p), (1-\gamma_p)\tau_d^E(p))} + a_{13}^3\end{aligned} \tag{4.5}$$

where a_{ij}^k are coefficients.

Note that one can obtain the preliminary estimate of the travel time in various ways. For example, using the average of collected travel times from a sufficient number of samples may be one of the simplest methods. However, for rarely observed traffic scenarios, it is difficult to produce a reliable estimation of the travel time at this preliminary stage. Therefore, the travel time estimation module requires at least one supplemental model to deal with scenarios lacking a reliable preliminary estimate.

Because detector data is usually collected on a lane-by-lane basis, the average speed of through traffic is not directly available from the detector information. Most existing studies either take data from one lane (e.g., the far left lane) as the average condition of the through traffic, or simply compute the average over all through lanes. However, as analyzed in the previous section, traffic conditions in some lanes may not affect the through-flow speed. Therefore, one needs to carefully select critical lanes to obtain the average speed of through traffic flow. This study assumes that the average speed of through traffic flow has a linear relation with those in all critical lanes, which may include both the through lanes (first item on the right side of Eq. 4.6) and the ramp lanes (second item on the right side of Eq. 4.6):

$$\frac{1}{\hat{u}_d^{Thru}(t,\Delta t)} = \sum_{la \in \mathbf{CLT}^{\mathbf{d}}_{d,d+1}(p)} \frac{a_{la}^5}{u_{d,la}(t,\Delta t)} + \sum_{la \in \mathbf{CLR}^{\mathbf{d}}_{d,d+1}(p)} \frac{a_{la}^6}{u_{d,la}(t,\Delta t)} + a_{11}^7 \qquad (4.6)$$

where a_{ij}^k are coefficients.

Note that reliable speed data may not be directly available from one detector and thus needs to be estimated from the available data. A commonly used method to estimate speed is to rely on the relation between traffic flow, occupancy and the average vehicle length.

$$u_{d,la}(t,\Delta t) = g\frac{v_{d,la}(t,\Delta t)}{o_{d,la}(t,\Delta t)} \qquad (4.7)$$

where g is the average vehicle length.

As reported in the literature, Eq. 4.7 may not be valid when the time interval is short, because average vehicle lengths may vary significantly during short intervals. However, the impact of this error decreases with an increase in the length of the selected time interval and/or the traffic volumes. Assuming that, under scenario *p*, a factor g_p can satisfy Eq. 4.7, one can then obtain Eq. 4.8 from Eq. 4.5, Eq. 4.6 and Eq. 4.7 as follows:

$$\begin{aligned}\tau_d(t) = & \sum_{la \in \mathbf{CLT}^{\mathbf{d}}_{d,d+1}(p)} b_{d,la}^{T,p} \frac{o_{d,la}(t,\gamma_p^d \tau_d^E(p))}{v_{d,la}(t,\gamma_p^d \tau_d^E(p))} + \sum_{la \in \mathbf{CLR}^{\mathbf{d}}_{d,d+1}(p)} b_{d,la}^{R,p} \frac{o_{d,la}(t,\gamma_p^d \tau_d^E(p))}{v_{d,la}(t,\gamma_p^d \tau_d^E(p))} \\ & + \sum_{la \in \mathbf{CLT}^{\mathbf{d+1}}_{d,d+1}(p)} b_{d+1,la}^{T,p} \frac{o_{d,la}(t+\gamma_p^d \tau_d^E(p),(1-\gamma_p^d)\tau_d^E(p))}{v_{d,la}(t+\gamma_p^d \tau_d^E(p),(1-\gamma_p^d)\tau_d^E(p))} \\ & + \sum_{la \in \mathbf{CLR}^{\mathbf{d+1}}_{d,d+1}(p)} b_{d+1,la}^{R,p} \frac{o_{d,la}(t+\gamma_p^d \tau_d^E(p),(1-\gamma_p^d)\tau_d^E(p))}{v_{d,la}(t+\gamma_p^d \tau_d^E(p),(1-\gamma_p^d)\tau_d^E(p))} + b_d^{0,p}\end{aligned}$$

(4.8)

where $b^{T,p}_{d,la}$ is the coefficient of the l_ath lane in $\mathbf{CLT}^{d}_{d,d+1}(p)$ at detector *d* under the pth traffic scenario for link (*d*, *d*+1);

$b^{T,p}_{d+1,la}$ is the coefficient of the l_ath lane in $\mathbf{CLT}^{d}_{d,d+1}(p)$ at detector *d*+1 under the pth traffic scenario for link (*d*, *d*+1);

$b^{R,p}_{d,la}$ is the coefficient of the l_ath lane in $\mathbf{CLR}^{d+1}_{d,d+1}(p)$ at detector *d* under the pth traffic scenario for link (*d*, *d*+1);

$b^{R,p}_{d+1,la}$ is the coefficient of the l_ath lane in $\mathbf{CLR}^{d+1}_{d,d+1}$ at detector *d*+1 under the pth traffic scenario for link (*d*, *d*+1); and

$b^{0,p}_{d}$ is the intercept for the pth scenario for link (*d*, *d*+1).

In order to estimate travel times with Eq. 4.8, one needs to estimate γ^{d}_{p}, which is the portion of time it takes one vehicle to traverse the first half of link (*d*, *d*+1).

4.4.2 Defining Traffic Scenarios

Defining the clustering function for a clustered linear regression model for travel time estimation is a challenging task which shall have the following features:

- Travel times in each clustered traffic scenario should always have a relatively small variation;
- The variables used for clustering should be obtainable from detectors;
- The input variables from both the upstream and downstream detectors should be obtained only from critical lanes so as to reflect actual through traffic conditions.

The following guidelines can help define the traffic scenarios under recurrent congestions:

1. Predefine the preliminary types of patterns, based on the congestion level detected by the upstream and the downstream detectors as shown in Table 4.1.

Table 4.1 Four types of basic traffic scenarios in each link

Traffic Condition at Upstream Detector	Traffic Condition at Downstream Detector	Congestion Level in the Link
No congestion	No congestion	Free-flow condition
Congested	No congestion	Moderate congestion or transition period
No congestion	Congested	Moderate congestion or transition period
Congested	Congested	Heavy congestion

2. If the congestion at one end of the link is not always uniformly distributed across lanes, one shall further divide the set of scenarios based on the nature of the congestion — for example, queue spillback caused by an off-ramp.
3. For uniformly distributed traffic conditions, the average of detected data across the same type of lanes shall be used as the input variable for the proposed model.
4. For scenarios with nonuniformly distributed traffic conditions, one shall take data from the lanes that are highly correlated with the observed traffic conditions as the input variables.

4.5 An Enhanced Trajectory-based Model

As it is often difficult to have sufficiently large samples for all possible traffic scenarios from field observations, this research has also developed an enhanced trajectory-based model to serve as a supplemental component for those scenarios with inadequate samples of historical data.

4.5.1 Speed Estimation

Using the trajectory-based model for travel time estimation, one needs to estimate the speed from known traffic data. Because speed data used in most trajectory-based models are for short intervals, Eq. 4.7 cannot provide reliable estimates. Instead, this study proposes the following equations for speed estimation:

$$u(x,t)=\begin{cases} u_{free} & , o(x,t) <= o_{free} \\ u_{cong}+(u_{free}-u_{cong})(1-\dfrac{o(x,t)-o_{free}}{o_{cong}-o_{free}})^{m} & , o_{free} < o(x,t) <= o_{cong} \\ u_{min}+(u_{cong}-u_{min})(1-\dfrac{o(x,t)-o_{cong}}{o_{max}-o_{cong}})^{n} & , o_{cong} < o(x,t) <= o_{\max} \\ u_{min} & , \text{otherwise} \end{cases} \tag{4.10}$$

where $u(x,t)$ is the speed to be computed at location *x* at time *t*;

$o(x,t)$ is the occupancy in the small section near location *x* at time *t*;

o_{free} is the upper bound of occupancy under free-flow traffic conditions;

o_{cong} is the boundary of occupancy between moderately and heavily congested conditions;

o_{max} is the maximum occupancy under recurrent congestion;

u_{free} is the free-flow speed;

u_{cong} is the boundary of the speed between moderately and heavily congested traffic conditions;

$u_{\min}$ is the minimum speed under heavily congested conditions; and

m and n are parameters to be calibrated with field data.

One can calibrate the boundaries of occupancy and speed data with collected travel times and detector data. The method reported by Zou and Wang (2006) is applicable for estimating m and n in Eq. 4.10 with collected field travel time information.

4.5.2 Model Formulations

To provide reliable estimation of travel times for a long link, a trajectory-based travel time estimation model needs to reliably compute the in-segment speed for each target vehicle even if its position is far from either end of the target link.

Unlike the models in the literature for short links (Coifman, 2002; van Lint and van der Zijpp, 2003), this study develops two types of in-segment speed estimation methods, depending on the vehicle's current position in a link. When the vehicle is within a short distance of the upstream detector or the downstream detector, this study considers a possible range of traffic propagation speeds to estimate the in-segment traffic situations from nearby traffic detectors. Otherwise, this study uses a model combining both traffic propagation relations with the piecewise linear speed-based (PLSB) model to achieve better robustness.

As shown in Figure 4.7, the model will first estimate occupancy using the enhanced trajectory-based model at the vehicle's position with Eq. 4.11 and will then apply Eq. 4.10 to compute the vehicle's speed at location *x* at time *t*. The vehicle is assumed to travel at this speed over a short interval, t_{step}, and then its new location at time ($t+t_{step}$) will be updated. The procedure repeats the same steps until the vehicle arrives at the downstream detector.

$$o(x,t) = \begin{cases} o_d\left(t+\frac{x-x_d}{u_c^{\max}}, t+\frac{x-x_d}{u_c^{\min}}\right) & \text{, if } x-x_d < \hat{x} \\ o_{d+1}\left(t-\frac{x_{d+1}-x}{u_c^{\min}}, t-\frac{x_{d+1}-x}{u_c^{\max}}\right) & \text{, if } x_{d+1}-x < \hat{x} \\ o_d\left(t+\frac{\hat{x}-x_d}{u_c^{\max}}, t+\frac{\hat{x}-x_d}{u_c^{\min}}\right) & \\ +\frac{(x-x_d-\hat{x})}{\hat{x}} & \\ \times\left(o_{d+1}\left(t-\frac{x-(x_{d+1}-\hat{x})}{u_c^{\min}}, t-\frac{x-(x_{d+1}-\hat{x})}{u_c^{\max}}\right)\right. & \text{, otherwise} \\ \left.-o_d\left(t+\frac{\hat{x}}{u_c^{\max}}, t+\frac{\hat{x}}{u_c^{\min}}\right)\right) & \end{cases} \tag{4.11}$$

where $\hat{x} = \begin{cases} \min\left(\frac{l_d}{3}, \frac{1}{3}\text{mi}\right) & \text{, when } l_d \geq 1 \text{ mile} \\ \frac{l_d}{3} & \text{, otherwise} \end{cases}$

$x_d \leq x \leq x_{d+1}$; and

$u_c^{\min}$ and $u_c^{\max}$ are the minimum and the maximum traffic propagation speeds.

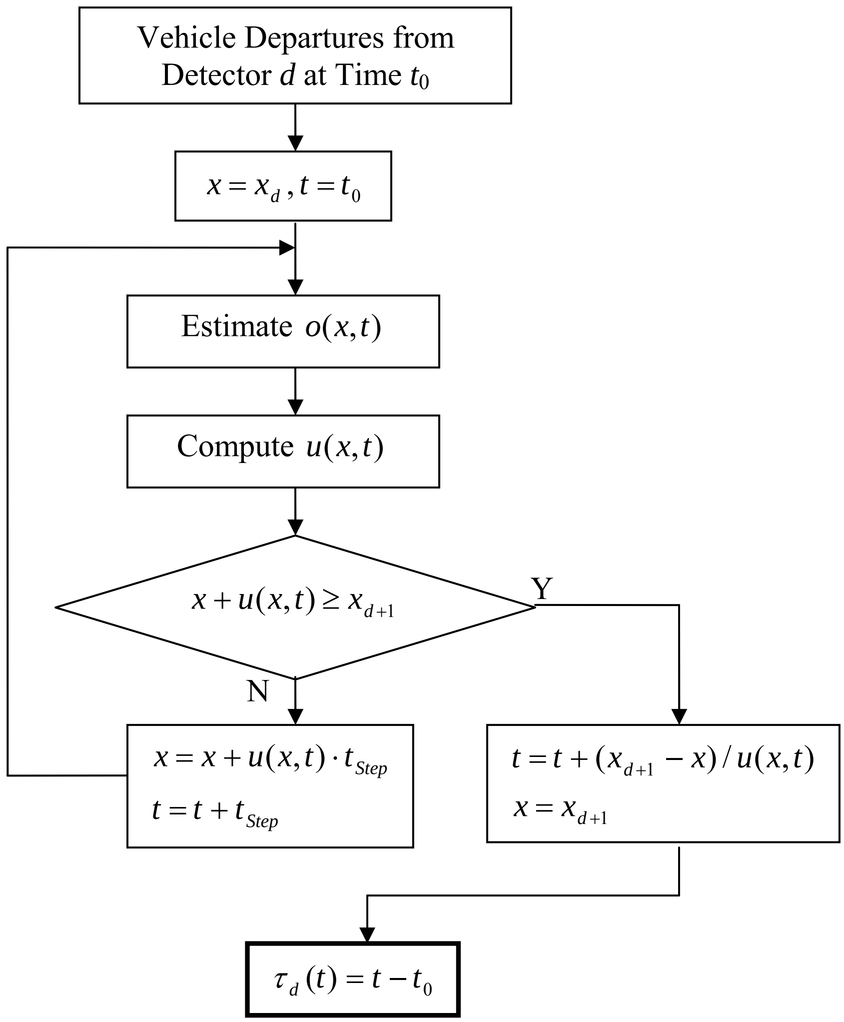


Figure 4.7 Flowchart of the enhanced trajectory-based travel time estimation model

4.6 Performance Evaluation of the Travel Time Estimation Module

The aforementioned travel time estimation module has been successfully calibrated and validated with actual travel time data collected from the field site on a 25-mile stretch of I-70 Eastbound from MD27 to I-695 with 10 detectors. Details of the geometric features of the freeway segment are available in Chapter 7. This section will present the evaluation results of the developed travel time estimation module, which serves as the tool for constructing and concurrently updating the database for historical travel times for the prediction module.

4.6.1 Dataset for Model Development

The research team at the University of Maryland conducted field surveys for each link between two neighboring detectors for calibrating the travel time estimation module. As shown in Table 4.2, the research team designed the survey plan based on the observed daily traffic patterns in the target freeway segment. For example, Segments 3-4, 4-5, 5-6 and 6-7 are often very congested in the morning, but usually not congested in the evening, therefore no evening surveys were conducted for these segments. In contrast, severe congestion is frequently observed in Segments 7-8, 8-9 and 9-10 during both morning and evening peak hours. The research team therefore collected the data for both AM and PM periods for those segments.

Please note that multiple surveys were conducted for certain links due to the encountering of nonrecurrent congestion patters such as accidents. Hence, the research team will generally first filter out the data points under incident/accident impacts, and then calibrate the travel time estimation module with samples in each link that exhibited different recurrent congestion patterns.

Table 4.2 Field surveys conducted by the research team for calibrating the travel time estimation module

Date and Time		Link								
		1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
12/1/2005	AM	Y	Y	Y	Y	Y	Y	Y	Y	Y
1/19/2006	AM					Y				
1/20/2006	AM						Y			
1/20/2006	PM									Y
2/1/2006	AM			Y						
2/2/2006	AM			Y						
2/7/2006	PM								Y	
2/28/2006	AM			Y	Y	Y	Y			
3/1/2006	PM							Y	Y	Y
3/7/2006	AM							Y	Y	Y
3/9/2006	PM							Y	Y	Y

Note: "Y" indicates that a survey has been conducted on the date and time listed on the first column.

4.6.2 Dataset for Performance Evaluation

After calibrating the travel time estimation module, the research team conducted additional field surveys for evaluating its performance. The evaluation covered the subsegments between Detector 3 (at the split of I-70 and US40) and Detector 10 (at the start of ramp to I-695) on the target freeway segment. The subsegment from Detector 3 to Detector 10 often incurs heavy congestion in the morning peak hours on Tuesdays and Thursdays. Therefore, the research team conducted two travel time surveys in the morning peak hours on April 6th, 2006 and April 20th, 2006 for this subsegment from Detector 3 to Detector 10. The actual travel times were obtained by matching vehicles from two videos taken at the beginning and end of the sub-segment. There were a total of 71 data points collected on April 6th, 2006 and 114 data points collected on April 20th, 2006. The surveys covered both transition periods between congestion and free-flow state, as well as heavily congested periods. Figure 4.8 shows the distribution of collected data samples in the subsegment from Detectors 3 to Detector 10 during the survey periods.

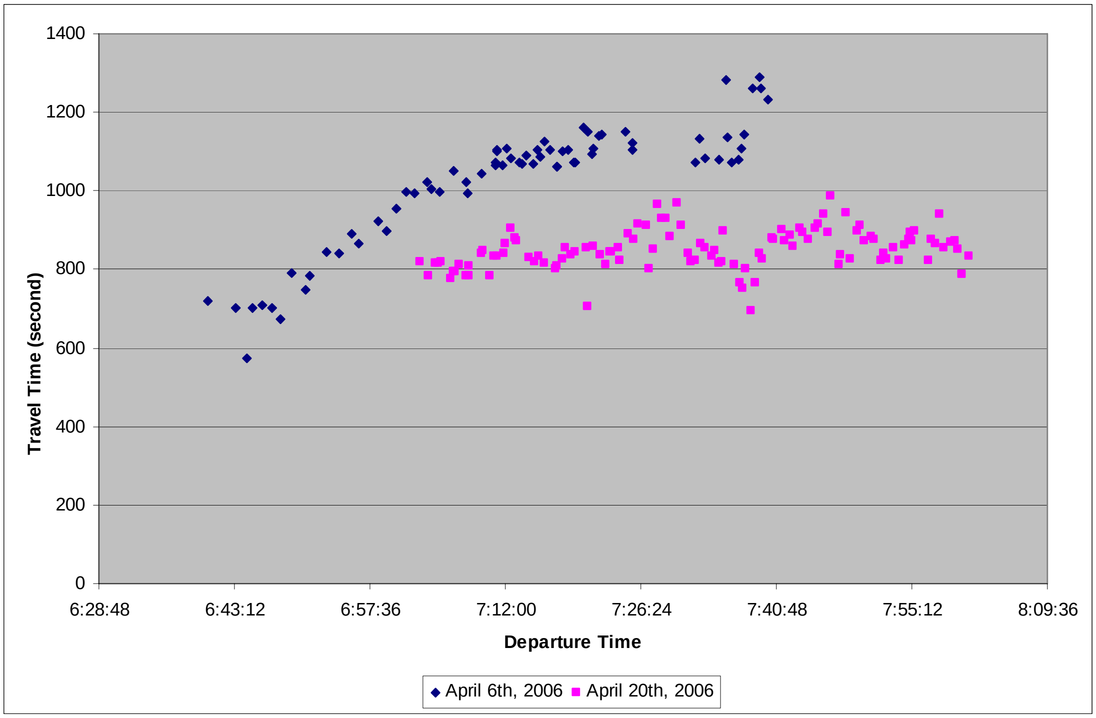


Figure 4.8 The distribution of collected travel times in April 6th, 2006 and April 20th, 2006

Note that the subsegment from Detector 1 to Detector 3 always exhibits no congestion and therefore one can use the free-flow speed to approximate its travel time.

4.6.3 Performance Evaluation

Subsegment from Detector 3 to 7 is about 4-mile in distance that consists of two interchanges and two ramps (Figures 7.1 and 7.2). Complex geometric features and high variation in traffic volumes have made this subsegment difficult for developing travel time estimation model. The research team categorized the congestion into different levels based on travel time ranges so as to have a detailed evaluation of the performance under various traffic conditions. As shown in Figure 4.8 and Table 4.3, congestion was much heavier on April 6th, 2006 having the maximum collected travel time of 1290 seconds (21.5 minutes) on the subsegment from Detector 3 to 10, which had a free-flow travel time of 520 seconds (8.7 minutes). Data collected on April 20th, 2006 were mostly between 800 seconds and 1,000 seconds, which exhibited quite fluctuated pattern between 7:15AM and 8:00AM.

Tables 4.3(a) and (b) summarize the performance of the estimation module on the subsegment from Detector 3 to Detector 10 against the actual data collected on two different days. Figures 4.9(a) and (b) show the distribution of estimated and actual travel times vs. departure time for two days, where the estimated travel times showed a similar trend to the actual travel times. The results from the travel time estimation module showed satisfactory results in all travel time categories during those two days with an average of less than 8.8% relative absolute error. Even in the transition periods, the module was still able to estimate travel times with an error of less than 70 seconds. In heavily congested cases, in which travel times are mostly greater than twice of the free-flow travel time (520 seconds), the developed module can still provide estimates with average absolute error of less than 90 seconds.

Table 4.3(c) shows the overall evaluation results for the transition periods (travel times between 520 seconds and 800 seconds), moderate congestion (travel times between 800 and 1000 seconds) and heavy congestion (travel times greater than 1,000 seconds). For all the 184 collected actual cases, the travel time estimation module successfully yielded the estimated travel times with the acceptable accuracy.

Table 4.3 Performance evaluation of the travel time estimation module

(a) Performance evaluation of travel time estimation module on the subsegment from Detector 3 to Detector 10 on April 6th, 2006

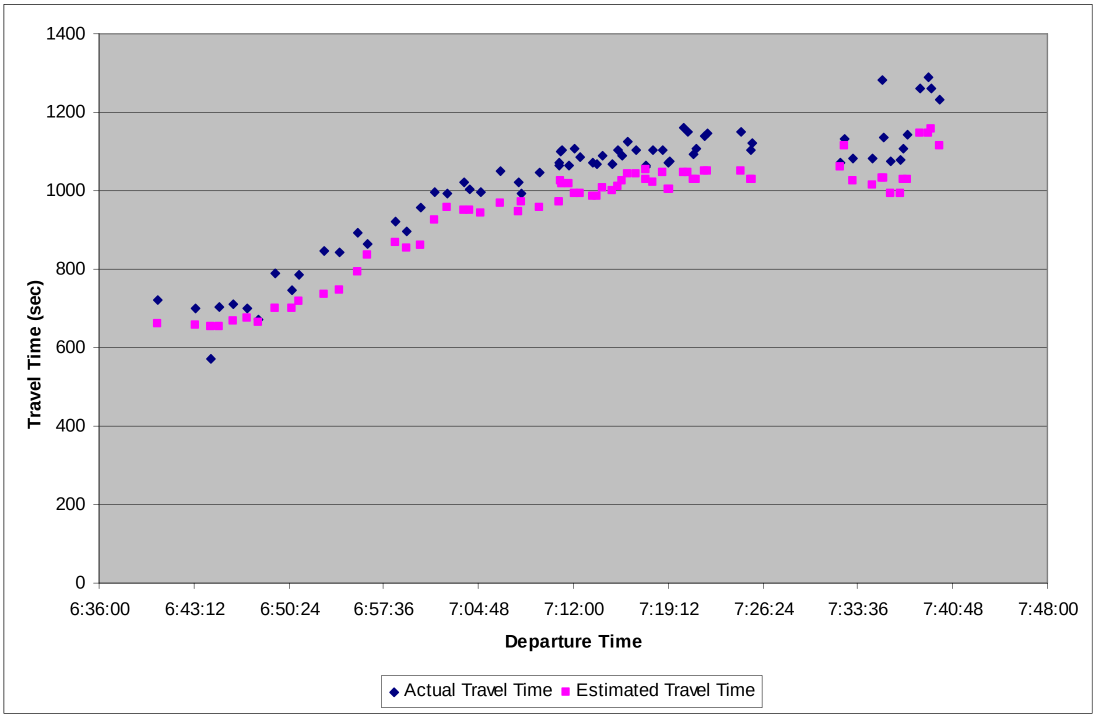
	Travel Time Range (sec)		
	520 to 800	800 to 1000	>1,000
Sample Size	10	12	49
Maximum Travel Time (sec)	791	998	1,290
Average Travel Time (sec)	710	928	1,109
Average Absolute Error (sec)	51.9	60.3	83.6
Average Relative Error (%)	7.3%	6.6%	7.4%

(b) Performance evaluation of travel time estimation module on the subsegment from Detector 3 to Detector 10 on April 20th, 2006

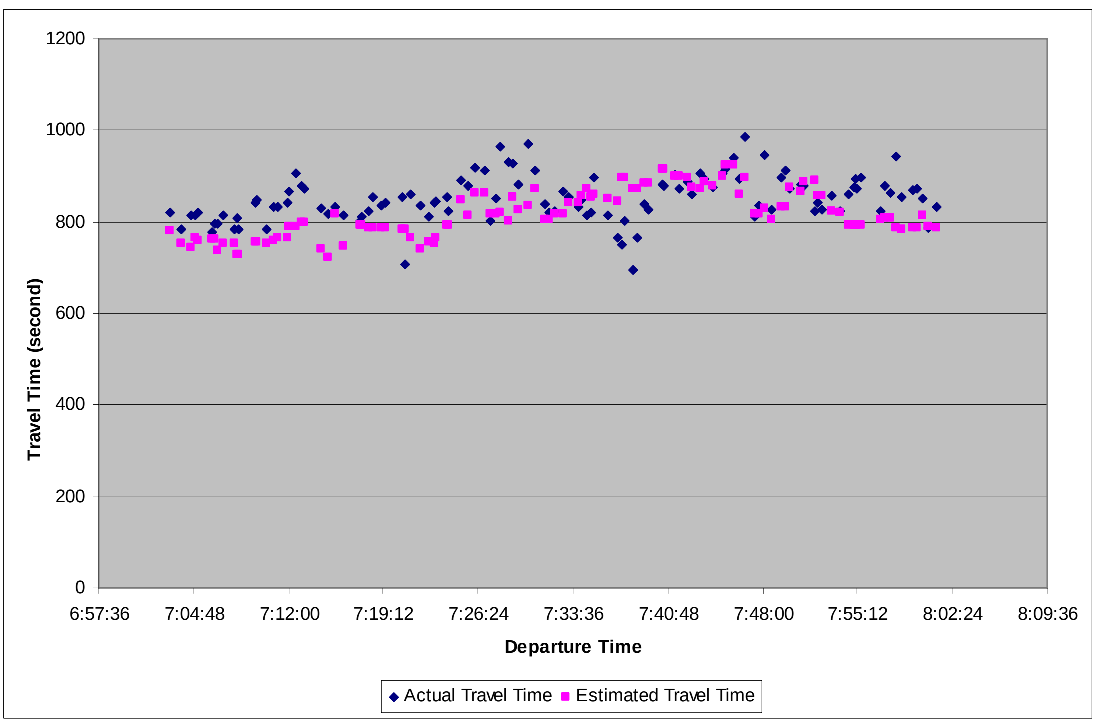
	Travel Time Range (sec)		
	520 to 800	800 to 900	900 to 1000
Sample Size	13	84	17
Maximum Travel Time (sec)	796	898	985
Average Travel Time (sec)	767	847	929
Average Absolute Error (sec)	65.2	49.4	73.0
Average Relative Error (%)	8.7%	5.8%	7.8%

(c) Overall Performance evaluation of travel time estimation module on the subsegment from Detector 3 to Detector 10 on April 6th and April 20th, 2006

	Travel Time Range (sec)		
	520 to 800	800 to 1000	> 1000
Sample Size	23	112	49
Maximum Travel Time (sec)	796	998	1290
Average Travel Time (sec)	742.3	847.2	1109.1
Average Absolute Error (sec)	58.5	54.5	83.6
Average Relative Error (%)	8.1%	6.3%	7.4%



(a) Comparison between actual and estimated travel times in the subsegment from Detector 3 to Detector 10 on April 6th, 2006



(b) Comparison between the actual and estimated travel times in the subsegment from Detector 3 to Detector 10 on April 20th, 2006

Figure 4.9 Comparisons between actual and estimated travel times in the subsegment from Detector 3 to Detector 10 on April 6th, 2006 and April 20th, 2006

4.7 Conclusions

This chapter presents a hybrid travel time estimation model that uses a clustered linear regression model as the main model, and an enhanced trajectory-based model as its supplemental component. The clustered linear regression model functions to categorize traffic conditions in a link into several scenarios, based on the exhibited congestion patterns. One can then construct the input dataset with selected critical lanes. The primary reason for using an enhanced trajectory-based model as a supplemental component is to contend with the lack of sufficient samples for some relatively uncommon traffic scenarios. The proposed supplemental model can take advantage of the traditional trajectory-based methods grounded on traffic propagation relations and piecewise linear-speed-based models to provide reliable travel time estimations on long links.

An extensive comparison between the collected and estimated travel times clearly indicate that the developed module is able to provide reliable estimates under transition periods, moderate congestion, and heavy congestion with an average relative absolute error less than 8.8%. During transition periods in the subsegment from Detector 3 to Detector 10, the developed module may yield a relatively large error, but it remains within the range of one minute. Overall, the module is capable of providing reliable travel times estimates from on-line detector data, and serving as a tool for constructing the historical travel time database.

Chapter 5: A Hybrid Model for Travel Time Prediction with Widely Spaced Detectors

5.1 Introduction

Due to deteriorating traffic conditions in most urban networks, providing reliable trip times to commuters has emerged as one of the most critical challenges for all existing advanced traveler information systems (ATIS). However, designing and implementing such a system to achieve the desired level of performance is quite a difficult task, as its resulting accuracy varies with many variables, including day-to-day traffic demands, responses of individual drivers and their commuting patterns, conditions of the road facility, weather, incidents, and reliability of available traffic detectors etc.

As discussed in Section 2.3, many studies have developed travel time prediction models for highway segments that have simple geometric features and densely distributed traffic detectors (e.g., every half-mile). The large number of detectors required for those models has limited their potential applications because of the diminishing resources for infrastructure development. This study intends to develop a travel time prediction model that can provide reliable travel time predictions in a sparsely distributed detector environment. The proposed model takes into account the geometric features of the target highway segment and the historical time-varying traffic patterns.

5.2 Model Structure

To reliably capture the variability of day-to-day congestions, this study proposes a hybrid model structure that employs a *k*-Nearest Neighbor model as the main model and an enhanced time-varying coefficient model as the supplemental model for traffic scenarios that do not have sufficient historical data. The *k*-Nearest Neighbor model can take full advantage of similar historical travel times. It uses the “distances” between the current traffic condition and historical cases to assess the quality of the model output and to determine the need for selecting a different model structure. During the operation, this system can continuously update the historical travel time database and the parameters of the supplemental model to improve the model’s accuracy. Figure 5.1 shows the flowchart of the proposed model for travel time prediction.

The travel time prediction system will first construct the input dataset of the *k*-Nearest Neighbors model from the current real-time traffic data. If there exist at least *k* historical cases within the similarity threshold (*TH*) for the current condition, then the hybrid model’s output will be the main prediction result, which is the average of those *k* best historical matches. Otherwise, the prediction system will reorganize the input data for the supplemental time-varying coefficient model and then output the prediction result from the supplemental model. The real-time data will be concurrently processed to update the database of historical travel times.

Sections 5.3 and 5.4 will present the core logic of the proposed *k*-Nearest Neighbor model and the supplemental enhanced time-varying coefficient model respectively.

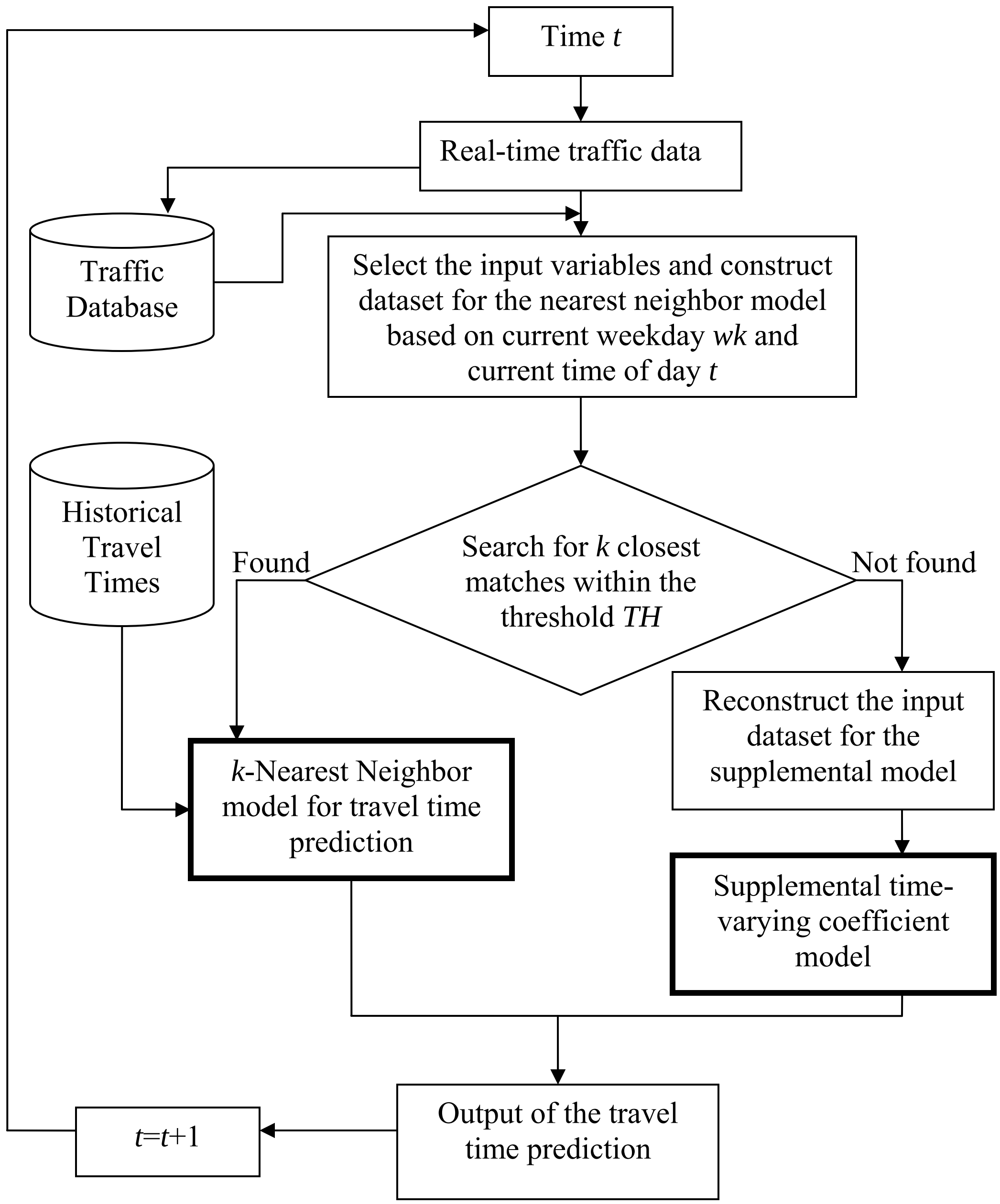


Figure 5.1 Flowchart of the hybrid travel time prediction model

5.3 *k*-Nearest Neighbor Model

To ensure the efficiency of the proposed *k*-Nearest Neighbor model, one needs to carefully analyze the following four key issues: the definition of the similarity, the selection of input variables, the searching window and time range, and the weighting factors. Each of these four key issues is discussed in sequence below:

Definition of the similarity

In a traditional *k*-Nearest Neighbor model, a distance is defined to reflect the similarity between two cases (Eq. 2.8). However, this definition for travel time prediction needs to be revised due to the fact that two cases with substantially different detected traffic data may still have similar travel times. Based on Eq. 2.8, this study has proposes the following procedures to compute the distance between the current and the historical cases.

The proposed model first categorizes the traffic conditions with detected occupancy information. One can then use the following equation to define the traffic conditions into three types, free-flow condition, moderate congestion, and heavy congestion.

$$TC_d^{l_a}(t, t+\Delta t) = \begin{cases} -1 & \text{, when } o_d^{l_a}(t, t+\Delta t) \le OF_d^{l_a} \\ 1 & \text{, when } o_d^{l_a}(t, t+\Delta t) \ge OC_d^{l_a} \\ 0 & \text{, otherwise} \end{cases} \qquad (5.2)$$

Where $TC_d^{l_a}(t, t+\Delta t)$ is the traffic type in lane l_a at detector d from time t to $t+\Delta t$;

$o_d^{l_a}(t, t+\Delta t)$ is the average occupancy in lane l_a at detector d from time t to $t+\Delta t$; and

$OF_d^{l_a}$ and $OC_d^{l_a}$ are the upper bound of free-flow occupancy and lower bound of heavy congestion occupancy for lane l_a at detector d respectively.

The model then defines the modified distance *mdis* between the current case and one historical case as:

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (p_i^* - q_i^*)^2} \tag{5.3}$$

Where $$p_i^* = \begin{cases} p_i & \text{, when } TC_d^{l_a}(t, t+\Delta t) = 0 \\ OC_d^{l_a} & \text{, when } TC_d^{l_a}(t, t+\Delta t) = 1 \\ OF_d^{l_a} & \text{, when } TC_d^{l_a}(t, t+\Delta t) = -1 \end{cases}$$

$$q_i^* = \begin{cases} q_i & \text{, when } TC_d^{l_a}(t_h, t_h+\Delta t) = 0 \\ OC_d^{l_a} & \text{, when } TC_d^{l_a}(t_h, t_h+\Delta t) = 1 \\ OF_d^{l_a} & \text{, when } TC_d^{l_a}(t_h, t_h+\Delta t) = -1 \end{cases}$$

and t_1 and t_2 are the time of day of the current case and historical cases, respectively.

Selection of the input variables

Most existing applications of the *k*-Nearest Neighbor model for travel time prediction simply take all available information to compute the distance between the current case and each candidate historical case. As discussed in Section 4.2.2, only information in critical lanes contributes to a reliable model output, especially when detectors are far apart. This study proposes the following procedures to best identify the most critical variables for computing the similarity distance, *mdis*:

1. Eliminate the data from those lanes that are well recognized by drivers in through traffic for their potential for disturbance incurred by on-ramp or off-ramp flows — for example, a through lane next to an off-ramp lane that most through traffic may avoid due to drivers' knowledge of the possible congestion caused by the queue spillback from the off-ramp lane.
2. Eliminate lanes that have no direct impact on the path travel time. duch as the right lane of a two-lane off-ramp.
3. Compute the average value for all through lanes at one detector location with the same traffic conditions, and then use it as the model input.

Note that those lanes with light historical traffic pattern are still needed in the input dataset for those scenarios having abnormal congestion patterns.

Searching window and time range

Both searching window and data intervals are important parameters for efficient operation of the *k*-Nearest Neighbor model. The searching window is the duration of time from the current time to the past in which a time series of the same variable is selected as the model input.

As is well recognized in most prediction literature, to perform a reliable prediction with a longer horizon usually requires more historical and/or on-line data. To predict the travel time on a segment with multiple likes, one needs to predict traffic conditions in a shorter prediction horizon for the detector location which is closer to the departure point than those detectors that are farther from the departure point. Therefore, the searching window of traffic information at each detector may increase with the distance from the origin point. To ensure the computing efficiency, one needs to set an upper limit for the size of the searching window so as to reduce the total number of input variables for the model, based on the local traffic pattern.

Note that various traffic patterns may exist in a segment during a day, and thus result in different travel times. For example, it is possible for two cases with similar detected traffic conditions to have different travel times. Very often, one morning case and one evening case may have similar detected traffic flows, which however go to different destinations. Therefore, the searching procedure should only look for historical cases that are within a reasonable range from the current time of day for better prediction accuracy. Hence one can add this constraint to Eq. 5.3 to obtain the following equation.

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (\hat{p}_i - q_i^*)^2} \tag{5.4}$$

Where $\hat{p}_i = \begin{cases} M & \text{, if } |t - t_h| > T_{th}(d,t) \\ p_i^* & \text{, otherwise} \end{cases}$

M is a very large number;

$T_{th}(d,t)$ is the time-varying range for searching at detector *d*; and

t and t_h are the time-of-day of the current case and the historical case respectively.

Note that one needs to determine $T_{th}(d,t)$ based on the day-to-day traffic patterns for a given time of day at detector *d*. For example, $T_{th}(d,t)$ may be different in morning peak hours, evening peak hours and off-peak hours.

To improve the model's reliability, besides the use of time of day information, this study further modifies Eq. 5.4 to search for cases that are in a weekday that usually has similar traffic patterns. Weekdays with similar traffic patterns are first grouped together into *S* sets. One can then modify Eq. 5.4 to obtain Eq. 5.5:

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (\hat{p}_i - q_i^*)^2} \tag{5.4}$$

Where $\hat{p}_i = \begin{cases} M & \text{, if } |t - t_h| > T_{th}(d,t) \\ p_i^* \times \hat{w} & \text{, otherwise} \end{cases}$

$$\hat{w} = \begin{cases} 1 & \text{, if } \exists s, wk_h \in \boldsymbol{W_s} \text{ and } wk_c \in \boldsymbol{W_s} \ (1 \le s \le \mathrm{S}) \\ M & \text{, otherwise} \end{cases}$$

$$\bigcup_{s=1}^{S} \boldsymbol{W}_s = \{\text{all weekdays}\}$$

M is a very large number; and

wk_c and wk_h are weekdays of the current case and the historical case, respectively.

Similarly, one needs to determine how to group weekdays based on traffic patterns reflected from the historical data.

Weighting factors

The model uses weighting factors to reflect how traffic conditions in each critical lane contribute to the target prediction. This study implements the following procedures to determine the weighting factors and the searching window for the *k*-Nearest Neighbor model.

Step 1: Divide one day into three traffic periods: morning peak hours, evening peak hours, and off-peak hours.

Step 2: Determine the input variable set for each traffic period in each weekday group, based on the revealed traffic patterns. (i.e., through lanes with uniform traffic conditions at the same detector location can be combined as one variable).

Step 3: Assign weighting factors for each variable during one traffic period in one weekday group, according to the frequency and severity of the congestion.

Step 4: Determine the searching window for each variable and the time-varying searching range for each weekly traffic scenario by analyzing the historical traffic patterns.

5.4 An Enhanced Time-varying Coefficient Model

This research has also developed an enhanced time-varying coefficient model, which serves as the supplemental model structure in the travel time prediction module for cases that do not have sufficient similar historical data. To ensure the model's reliability for long segments with widely spaced detectors, this study modifies the definition of the *status travel time* in Eq. 2.10 with the weekly traffic patterns to the existing time-varying coefficient models (Zhang and Rice, 2003; Kwon and Petty, 2005), which establish a linear relation between the *status travel time* and the actual travel time.

Figure 5.2 shows the flowchart of the enhanced time-varying coefficient model for travel time prediction. The model first constructs an input dataset by combining the currently available real-time data and the historical average of detector data in the same weekday after departure time *t*. The hybrid travel time estimation model developed in Chapter 4 will provide the estimated travel time, $\hat{T}(t)$, which is the modified *status travel time*, based on the input dataset. One can then apply the weekly time-varying coefficients to compute the predicted travel time as follows:

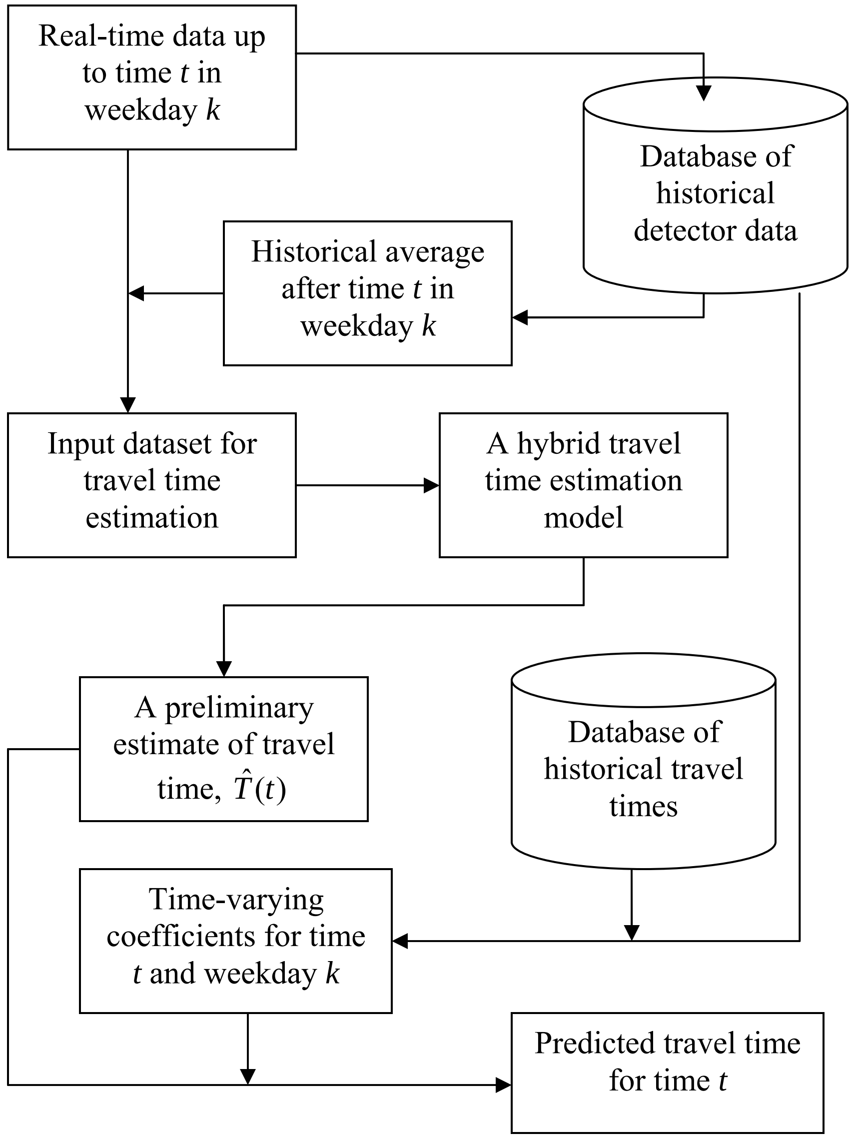


Figure 5.2 Flowchart of the enhanced time-varying coefficient model

$$T(t) = a_{t_i}^k \hat{T}(t) + b_{t_i}^k \tag{5.5}$$

Where $T(t)$ is the estimated travel time; and

$a_{t_i}^k$ and $b_{t_i}^k$ are the weekly time varying coefficients for the t_i^{th} interval of the current weekday, k.

The modified *status travel time* serves as a preliminary estimate of travel time for a long segment more reliably than the original *status travel time* proposed mainly for short links. The use of weekly time-varying coefficients ensures that the model can capture weekly traffic patterns. The coefficients will be updated as soon as the new travel time information is available — after some trips have been finished. Note that the updated coefficient will not be used until the same weekday next week because the time interval Δt is usually smaller than the travel time $T(t)$.

5.5 Conclusions

This study develops a hybrid travel time prediction module for reliable real-time travel time prediction. Based on the historical travel times obtained from the travel time estimation module presented in Chapter 4, the proposed prediction module uses a *k*-Nearest Neighbor Model as the main model, and an enhanced time-varying coefficient model as the supplemental model for traffic conditions which do not have sufficient similar patterns in the historical database. The performance evaluation for the travel time prediction module and the overall system will be presented in Chapter 7.

Chapter 6: Incident Detection Module

6.1 Introduction

The presence of an incident may cause significant delays on travel times, which cannot be reliably predicted with models developed for recurrent traffic conditions due to a variety of factors. However, as the reliability and the robustness are the top priority issues in a travel time prediction system, control center operators need an incident detection module to switch the system to a different mode automatically in response to a detected incident. Such a detection system is especially critical for a prediction system deployed with sparsely distributed detectors. This chapter details the incident detection module developed for the field demonstration of the real-time travel time prediction system, tested between June 4^{th}, 2006 and August 4^{th}, 2006. The developed incident detection module, which focuses on the major incident/accident, has the following key features:

- Has more than one algorithm to work in parallel for best detection performance;
- Embed a false alarm detection function to capture the impacts of geometric features on the traffic condition; and
- Interact with a well-document database for recurrent congestion patterns, including both daily and weekly patterns, to reduce the false alarm rate.

6.2 Model Structure

To reliably detect the incident impacts, which usually are mixed with the day-to-day congestions in links with various geometric features, this study develops a rule-based hybrid incident detection algorithm for use on freeways having long detector spacing.

Figure 6.1 shows the flowchart of the developed incident detection module, where its single-station detection algorithm is used to detect incidents close to the detector station at either end of a link, and the dual-station detection algorithm is designed to identify incidents that are far from either detector.

The operation logic of the developed incident detection module is to first forecast the current traffic condition at each detector location, based on the information collected prior to the current time interval for the two embedded algorithms. The single-station incident detection model and the dual-station incident detection model will then work in parallel to detect incidents for each link over the entire target freeway segment. When one incident is detected by any of the two algorithms, the incident detection module will inform the travel time prediction system to stop the prediction for the impacted segments, and then switch to a different mode such as replacing the predicted travel times with the incident warning messages.

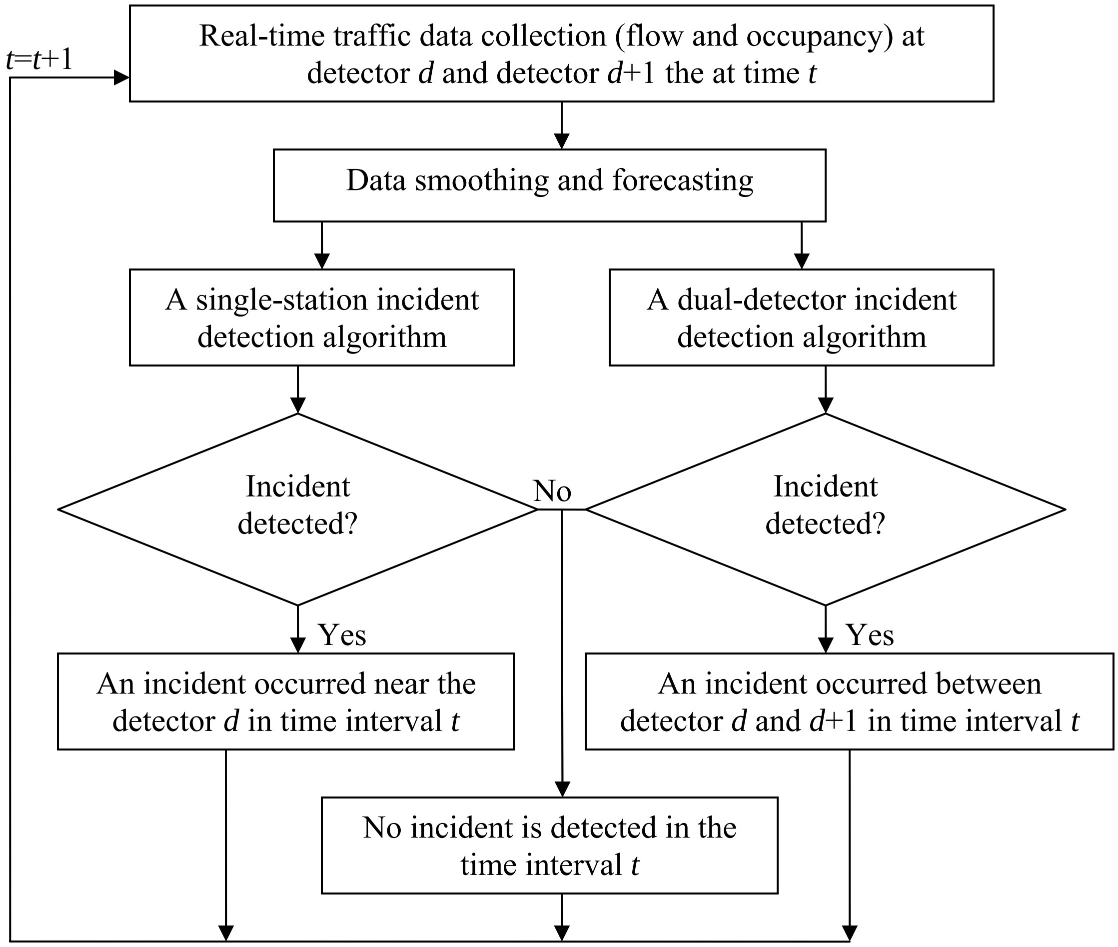


Figure 6.1. System flowchart of the incident detection module

6.3 Single-station Incident Detection Algorithm

The incident detection module has the same type of real-time input information as other modules discussed in the previous chapters. The traffic flow and occupancy over the pre-determined time intervals will be collected from traffic detectors. The algorithm will be used to focus on monitoring the abnormal changes in traffic flow and occupancy data, based on the real-time data and other obtainable information such as the time-varying historical traffic patterns. The core concept of the single-station incident detection algorithm is to detect the abnormal temporal change in the traffic conditions that do not match any historical traffic patterns with the following two types of control variables, Type *A* and *B*, as defined in Eq. 6.1 and 6.2:

$$A_{d,i}^{Flow}(t) = \frac{Y_d^{Flow}(t-i+1) - v_d^{thru}(t-i+1)}{v_d^{thru}(t-i+1)}$$

$$A_{d,i}^{Occ}(t) = \frac{Y_d^{Occ}(t-i+1) - o_d^{thru}(t-i+1)}{o_d^{thru}(t-i+1)} \qquad (6.1)$$

where: $A_{d,i}^{Flow}(t)$ is the i^{th} Type *A* control variable for traffic flow at time *t* at detector *d*

$A_{d,i}^{Occ}(t)$ is the i^{th} Type *A* control variable for occupancy at time *t* at detector *d*

$i_{\max}$ is the maximum number of intervals to compare.

i is the index of the control variable, which is numbered from the current time interval to the past. Therefore, i=1 is for the current time interval, and i=2 is for the first time interval prior to the current time interval, etc. $1 \le i \le i_{\max}$.

$Y_d^{Flow}(t)$ is the predicted flow data at time *t* at detector *d*, based on the actual detector flow data obtained in intervals prior to time *t*

$Y_d^{Occ}(t)$ is the predicted occupancy data at time *t* at detector *d*, based on the actual detector occupancy data obtained in intervals prior to time *t*

$$B_{d,i}^{Flow}(t) = \frac{v_d^{thru}(t-i+1) - \bar{v}_d^{thru}(t-i+1)}{S_d^{FT}(t-i+1)}$$

$$B_{d,i}^{Occ}(t) = \frac{o_d^{thru}(t-i+1) - \bar{o}_d^{thru}(t-i+1)}{S_d^{OT}(t-i+1)} \qquad (6.2)$$

where: $B_{d,i}^{Flow}(t)$ is the i^{th} Type *B* control variable for traffic flow at time *t* at detector *d*

$B_{d,i}^{Occ}(t)$ is the i^{th} Type *B* control variable for occupancy at time *t* at detector *d*

$\overline{v}_{d}^{thru}(t)$ is the mean value of the detected through traffic flow at detector *d* in the past *p* intervals before the time interval *t*. $\overline{v}_{d}^{thru}(t) = \frac{1}{p}\sum_{i=1}^{p} v_{d}^{thru}(t-i)$

$\overline{o}_{d}^{thru}(t)$ is the mean value of the detected occupancy of through traffic at detector *d* in the past *p* intervals before the time interval *t*.

$$\overline{o}_{d}^{thru}(t) = \frac{1}{p}\sum_{i=1}^{p} o_{d}^{thru}(t-i)$$

$S_{d}^{FT}(t)$ is the standard deviation of the flow data for through traffic at station *d* in the past *p* intervals before the time interval *t*.

$$S_{d}^{FT}(t) = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(v_{d}^{thru}(t-i) - \overline{v}_{d}^{thru}(t))^{2}}$$

$S_{d}^{OT}(t)$ is the standard deviation of the occupancy for through traffic at station *d* in the past *p* intervals before the time interval *t*.

$$S_{d}^{OT}(t) = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(o_{d}^{thru}(t-i) - \overline{o}_{d}^{thru}(t))^{2}}$$

Type *A* control variable is defined to detect the absolute difference between the detected traffic information and the predicted traffic information, based on the data collected before the current time interval. Type *B* control variable is for use in detecting a relative change in the traffic characteristic data in the time series. As revealed in Chapter 2, there exist several forecast methods that can reliably forecast $Y_{d}^{Flow}(t)$ and $Y_{d}^{Occ}(t)$ from the traffic data obtained before the time interval *t*, including both the parametric and non-parametric methods. The developed *k*-Nearest

Neighbors model for travel time prediction in Chapter 5 is one of the potential candidates, which can take best advantages of historical daily and weekly traffic patterns for a reliable prediction. The rule-based single-station detection algorithm is designed for the following two incident scenarios with two types of rules:

Incident at the downstream segment of one detector location

When one incident occurs at the downstream of detector *d*, the reduced roadway capacity at the incident location may cause queue spillback to the upstream and result in a lower flow rate and a higher occupancy at the location of detector *d*. In the single-station detection algorithm, the following rules are for detecting the anomaly in the traffic volume and the occupancy at detector *d*:

IF $A_{d,i}^{Occ} > T_{Occ1}^{D}$ for all *i*, THEN

IF $A_{d,i}^{Flow} > T_{Flow1}^{D}$ for all *i*, THEN

One incident is detected at the downstream segment of detector d

ELSE

No incident can be reliably detected.

END

ELSE

No incident can be reliably detected.

END (6.3)

where $1 \le i \le i_{\max}$,

T_{Occ1}^{D} and T_{Flow1}^{D} are pre-determined thresholds.

To reduce the false alarm rate, the algorithm requires $i_{\max}$ consecutive intervals to confirm a detected incident as shown in Eq. 6.3.

Incident at the upstream segment of one detector location

An incident incurred at the upstream location of a detector may cause a substantial decrease in volume that shall be much lower than that in the previous intervals and historical patterns. However, one needs to distinguish the similar traffic pattern caused by the traffic spillback from an off-ramp or a lane drop. Therefore, the impacts of these factors need to be considered when predicting $Y_d^{Flow}(t)$ and $Y_d^{Occ}(t)$, i.e., the traffic conditions in these critical lanes need to be included in the input dataset. The following rules are used to detect incidents at the upstream segment of one detector location:

IF $A_{d,i}^{Occ} < T_{Occ1}^{U}$ for all *i*, THEN

IF $A_{d,i}^{Flow} > T_{Flow1}^{U}$ for all *i*, THEN

One incident is detected at the upstream segment of detector d

ELSE

No incident can be reliably detected.

END

ELSE

No incident can be reliably detected.

END (6.4)

where $1 \le i \le i_{\max}$,

T_{Occ1}^{U} and T_{Flow1}^{U} are pre-determined thresholds.

Notice that the developed rule-based single-station incident detection algorithm may not efficiently detect some incidents, especially those occurred at a location far from both detectors in a long link. Therefore, the following dual-station incident detection algorithm is proposed for incidents in the middle of the long link under certain types of geometric features and traffic patterns.

6.4 Dual-station Incident Detection Algorithm

The proposed dual-station incident detection algorithm, as shown in Figure 6.2, consists of two components: a filtering function and an incident detection model. A k-Nearest Neighbors model serves as the filtering function to identify whether the current traffic scenario has been observed in the historical dataset. Due to the complex interactions between many factors and the long detector spacing, traffic scenarios with different spatial patterns may yield the similar volume and/or occupancy at the detector location over some periods. Hence, a filtering function based on the historical data is necessary to minimize the potential false alarms. To do so, one needs to select all lanes across both upstream and downstream detectors to construct the input dataset of the k-Nearest Neighbors model, which will output N_S, the number of similar historical cases found in the predefined searching window. If $N_S \geq k$, the dual-detector incident detection algorithm will stop and output to the incident detection module to indicate that no incident can be reliably detected at this time. Otherwise, the second component will be executed to detect the existence of an incident.

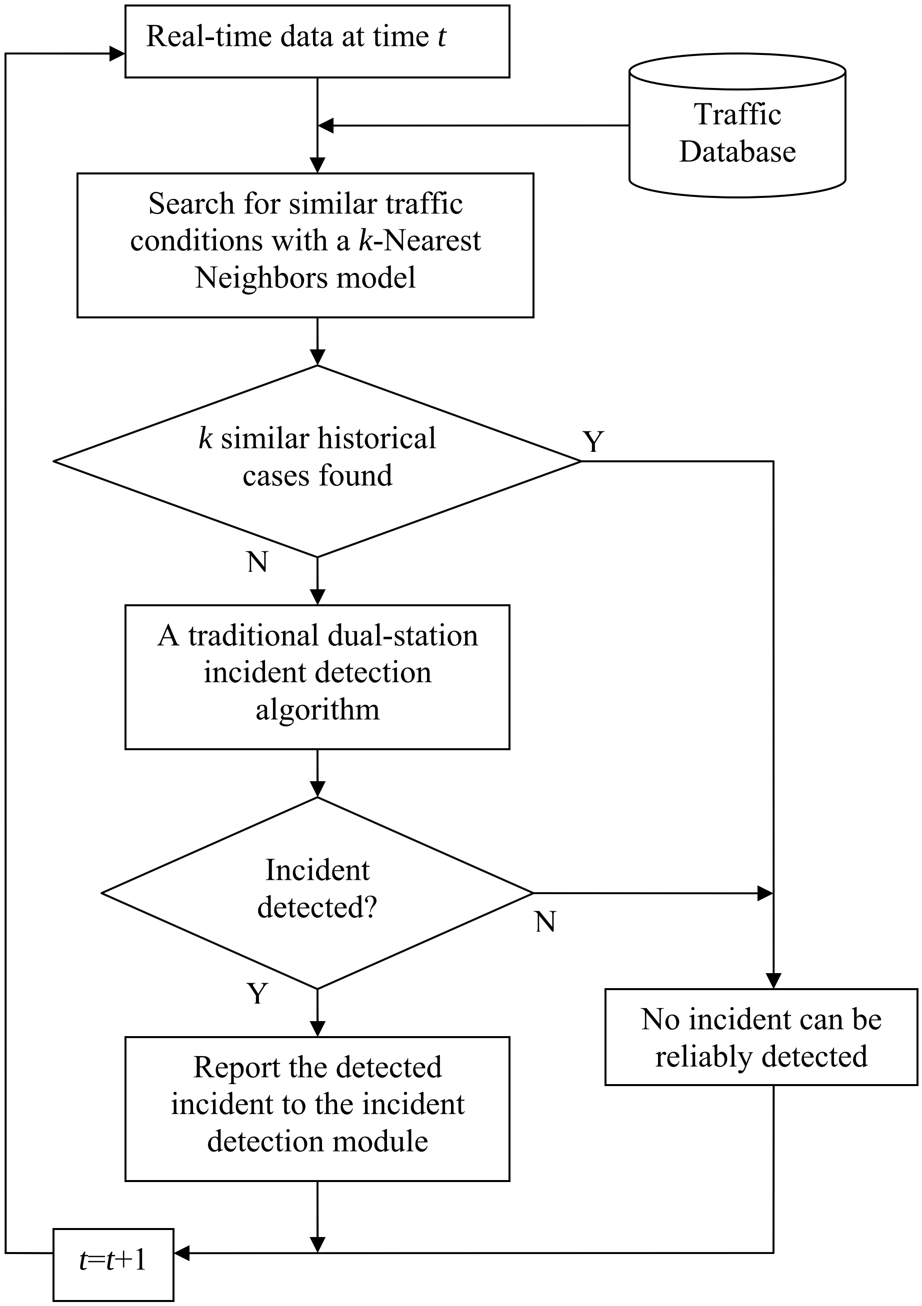


Figure 6.2. Flowchart of the dual-station incident detection algorithm

This study implemented the dual-station detection algorithm by Stephanedes and Hourdakis (1993) for cases that do not follow any traffic patterns observed in the history. Based on the comparison between the upstream data and the downstream data, Stephanedes and Hourdakis's model (1993) was reported to efficiently detect incidents under certain traffic conditions. In applying the dual-station incident detection algorithm, one shall first compute the differences between smoothed upstream and downstream occupancy data before and after the occurrence of the incident. A general form of the difference between smoothed data is defined in Eq. 6.5.

$$\Delta OCC_d(t,n) = OCC_d(t,n) - OCC_{d+1}(t,n) \tag{6.5}$$

where: $OCC_d(t,n)$ is the smoothed occupancy at time t+n-1 from the detected actual occupancy data between time t and t+n-1 at detector d.

The dual-station incident detection algorithm will then define two time periods, m intervals before the incident occurrence for representing the current normal recurrent congestion pattern and n intervals after the time when incident occurs to represent the time delay for the resulted impact from the incident to be detected. One may first obtain the maximum smoothed occupancy from actual detector data under the recurrent congestions from Eq. 6.6:

$$\max OCC_d(t) = \max(OCC_d(t-m,m) - OCC_{d+1}(t-m,m)) \tag{6.6}$$

Then, the algorithm will use the following two rules for the congestion test (Eq. 6.7) and the incident test (Eq. 6.8).

$$\Delta OCC_d(t,n) / \max OCC_d(t) > T_C \tag{6.7}$$

where T_C is a pre-determined threshold.

$$(\Delta OCC_d(t,n) - \Delta OCC_d(t-m,m)) / \max OCC_d(t) > T_I \tag{6.8}$$

where T_I is a pre-determined threshold.

If both tests pass, the dual-detector incident detection algorithm will report to the incident detection module that there exists one incident with a detection lag of *n* intervals.

6.5 Summary

This chapter develops an incident detection module for the travel time prediction system. This module consists of two incident detection algorithms, a single-station incident detection algorithm that is efficient for detecting incidents close to a detector station and a dual-station incident detection algorithm for detecting incident far from both detectors in a long link. The single-station incident detection algorithm predicts the current traffic condition with previous collected data and monitors both absolute and relative differences between the predicted traffic data and the detected traffic data at the current time interval to detect the existence of one incident. The dual-station incident detection algorithm first identify if the current traffic pattern has been observed in the history or not with its filtering function, a *k*-Nearest Neighbors model. If no sufficient historical matches can be found, the module will apply the traditional dual-station incident detection algorithm. If any of the two embedded algorithm returns with a positive detection result, the incident detection module will output a detection of one on-going incident to the travel time prediction model.

Chapter 7: System Implementation Procedures and Performance Evaluation

7.1 System Implementation Procedures

This chapter presents the principal procedures for implementing the automatic real-time travel time prediction system (ARAMPS) for work-zone operations, followed by the evaluation of the travel time prediction performed by ARAMPS, which integrates advanced models, cost-efficient traffic detection system, wireless communications, and its developed control and operation software in a portable, automated, and user-friendly system.

The system implementation process consists of three stages: system preparation, model training, and real-time operation stages. The rest of this section will describe core tasks in each of the three stages in sequence.

Stage 1: System Preparation

The system preparation stage is for identifying potential time-varying impacts due to recurrent traffic congestions, work-zone operations, and geometric features, which are the important criteria for determining optimal detector locations. Another principal task at this stage is to calibrate detectors and test other required hardware, including communication devices, database servers, and web servers.

Review of the work-zone plan

To take into account both the geometry impacts and traffic patterns in design of the travel time prediction system, it is very important to carefully analyze the following issues associated with the work-zone operational plan:

- The lanes scheduled to close, including mainline lanes and ramps, and the length of the closed segment;
- The geometric changes that may take place in the target segment, including lane shifts, changes on auxiliary lanes, and ramp closure etc.; and
- The work schedule for lane closure, including both the time-of-day and the operational duration in each day.

Site surveys

A careful site survey is an essential task to collect the following information:

- Geometric features of the target segment, including both mainlines and ramps;
- Time-varying traffic patterns during the congested periods (for example, morning and evening peak hours); and,
- Main ramps and critical lanes that carry a large volume of traffic.

In the survey, one also needs to identify the approximate locations of frequently observed queues due to the combined impact of traffic patterns, ramps, geometry change, and other factors. This type of information can be collected through field observations, traffic counts at key locations, or from a well-calibrated traffic simulation environment.

Determining the detector locations

With the information collected from the previous two tasks, one can determine candidate locations for deploying detectors based on the work-zone plans, geometric features of the target segment, daily and/or weekly traffic patterns, and the strengths of candidate models developed for ARAMPS.

Determining the model structures

Based on the revealed traffic patterns and geometric features in the target segment, one can determine the number of models needed to predict travel times under different levels of traffic conditions. For example the model structure for frequently recurrent congestion patterns may be different from the model for those relatively less observed traffic conditions.

Installing and calibrating the detection system

Once detectors have been installed at designated locations, one needs to calibrate all detectors under all types of traffic conditions. Both calibration and validation are essential tasks and cannot be skipped so as to prevent dramatic performance deterioration due to unreliable detector data under one or more types of traffic conditions or the impact of certain vehicle types (for example, long trucks). The validation process shall include verifications on volume counts and speed under different congestion levels

Installing communication devices, variable message signs and other required equipments

ARAMPS system requires communication between detectors and servers, from servers to variable signs, and servers to servers. To prevent the system from displaying untimely messages and unreliable predictions due to communication failures, one needs to properly address the following issues:

- Placing a timer on each VMS to shutdown the display if no message has been received over a prespecified interval (i.e., 10 minutes in ARAMPS);
- Installing a program on the data server to detect missing data; and
- Developing system monitors to detect the loss of communication to signs, service interruption of servers and accidents.

Stage 2: Model Training

The main tasks at this stage are to calibrate the system parameters and to build the historical travel time database. The purpose is to ensure that the detectors are ready to transfer traffic data to the server in either the real-time mode or store them in their local storage devices. This stage consists of the following primary tasks.

Field surveys on travel times

To train and/or calibrate all models in the four system modules in ARAMPS with data from detectors, one needs to synchronize the survey time with the internal times embedded in all detectors prior to the survey so as to avoid data inconsistency. At the model-training stage of the ARAMPS system, survey personnel always used the time from GPS satellites to synchronize times between all devices.

To collect a sufficient amount of travel time samples for model training and/or calibration in ARAMPS, this study has employed several survey methods, including vehicle matching with videos from both upstream and downstream detectors, license plate number matching with videos taken at upstream and downstream detectors, and driving in the target segment with GPS devices. To capture the impacts of both daily and weekly traffic patterns, the surveys conducted for ARAMPS covered both morning peak hours, evening peak hours, and non-peak hours of different weekdays.

Long-term collection of traffic data

ARAMPS system needs a long period of data to build its historical traffic database and then apply the travel time estimation module in the next step to construct the historical travel time database prior to the field demonstration. This study has taken a four-month period to gather sufficient information of various traffic patterns for use in model development.

Calibration of system parameters and the construction of the databases

As mentioned in the previous chapters, all principal modules need training and/or calibration. The travel times collected in field surveys will be divided into two groups, one group for model training/calibration and the rest for model verification. Models in the travel time estimation module were calibrated first, and then applied to estimate travel times for the construction of the historical travel-time database. The estimated historical travel times will in turn be used along with actual travel times collected in the field surveys to develop the travel time prediction model.

Implementation and testing of system monitors

To ensure the system's reliability during real-time operations, one needs to carefully evaluate the sensitivity and functionalities of the system monitors. All system monitors are expected to report failure or critical conditions of any of the following items in a timely manner.

- Detectors
- VMS signs
- Databases of traffic data and historical travel times
- System servers, including both data servers and web servers
- Communications between detectors, VMS and servers
- Incident detection
- Missing and/or delayed data

Stage 3: Real-line Operation

At this stage, all system components are in the real-time operation mode. The system will output its prediction results on the VMS signs and the system website (http://i70.umd.edu); and send system alerts to administrators' emails and cell phones.

Review of the automated daily report

This study has designed an automated system daily report to present important information associated with the system's current status, which may also contain early signs of device failures. For example, if missing data occurs at one detector location frequently, it probably indicates a potential failure of either the detector or the attached communication device.

System performance evaluation with field data

After successfully operating ARAMPS in real time over the initialization period, the research team has conducted extensive field studies to assess the system's accuracy and reliability. The system parameters may require a re-calibration if some significant inconsistencies between surveyed travel times and system outputs exist.

7.2 System Performance Evaluation

ARAMPS system has been implemented on I-70 for demonstrating its potential to provide reliable travel time prediction with sparsely distributed detectors for work-zone and other freeway operations. Ten traffic detectors were installed at locations determined by the research team of University of Maryland in May 2005. It took about 8 months for the contractor to finish the calibration of the detectors with the help from the manufacturer, and to setup up the communication between detectors and University of Maryland servers. The research group conducted data validation several times in this 8-month period to ensure the reliability of detector data. The real-time data transmission started to operate in late January 2006. Once 4-month complete data has been collected, ARAMPS system started to operate in the real-time mode from June 4th, 2006. The demonstration lasted for two months until August 4th, 2006. This section will present the system implementation site and the performance evaluation of prediction results.

7.2.1 System Site

ARAMPS system has been implemented on a 25-mile stretch of I-70 eastbound between MD27 and I-695 (Figure 7.1). As shown in Figure 7.2, 10 detectors have been placed unevenly in the target segment to maximize the model efficiency. Table 7.1 summaries the relative location of each detector to its nearby cross road or mileage marker. There are a total of five VMS signs placed in the target segment at locations determined by Maryland State Highway Administration for concurrently displaying predicted travel times from each sign to I-695 in real time. Figure 7.3 shows the locations of five VMS signs, whose geographical coordination and relative locations have been summarized in Table 7.2.

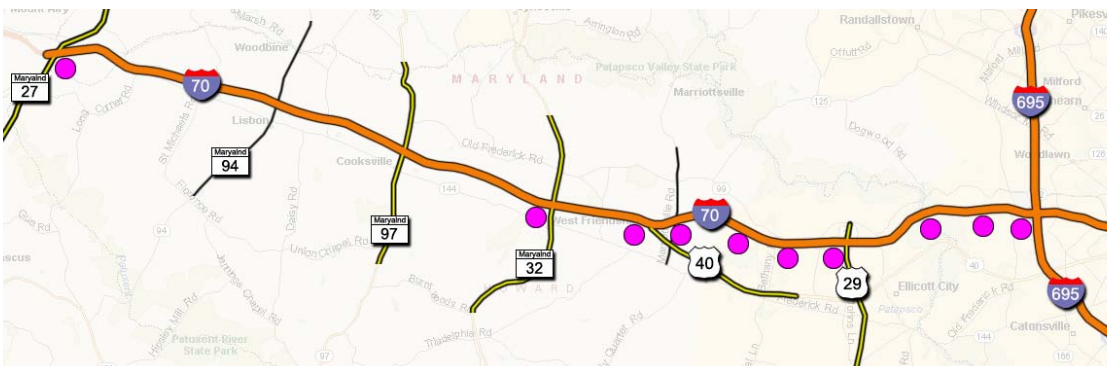


Figure 7.1 Locations of ten detectors in ARAMPS

7.2.2 Performance Evaluation

This section presents the performance evaluation of the developed travel time prediction system based on the actual travel times collected by a consulting company hired by the Maryland State Highway Administration. To extensively assess the accuracy of the predictions during congested traffic scenarios, the consulting firm conducted the data collection in both morning and evening peak hours from May 15th, 2006 to May 19th, 2006, and on July 25th, 2006 and July 26th, 2006. As shown in

Table 7.3, the field travel-time surveys also covered the evening peak hours on July 13th, 2006 and July 18th, 2006. During the field surveys, vehicles were dispatched at a headway of 10 to 15 minutes to travel over the target freeway segment to produce the actually travel times. In order to best estimate the system's performance under recurrent congestion patterns, the evaluation dataset did not include the field travel times that were impacted by incidents or accidents. Drivers conducting the field surveys were asked to record the arrival times to Signs 3, 4 and 5, and to the destination, I-695.

Table 7.1 Description and geographic locations of ten detectors

Detector ID	Location	Longitude	Latitude
1	About 1000 feet pass MD27	-77.163174	39.359605
2	About 500 feet pass the on-ramp from MD32 to I-70EB	-76.941133	39.307418
3	Right before the split of I-70 and US40	-76.918053	39.304853
4	At the acceleration area of the on-ramp from Marriottsville Rd. to I-70EB	-76.894104	39.304877
5	Between the mileage marker 84 and 85	-76.874133	39.302298
6	At the mileage marker 86	-76.848583	39.295600
7	At the deceleration area of the off-ramp to US29 Southbound	-76.830809	39.296183
8	At 2-mi to I-695 sign	-76.790894	39.306034
9	At 1-mi to I-695 sign	-76.771548	39.306553
10	At the split of I-70 to Park and Ride and to I-695	-76.752429	39.306717

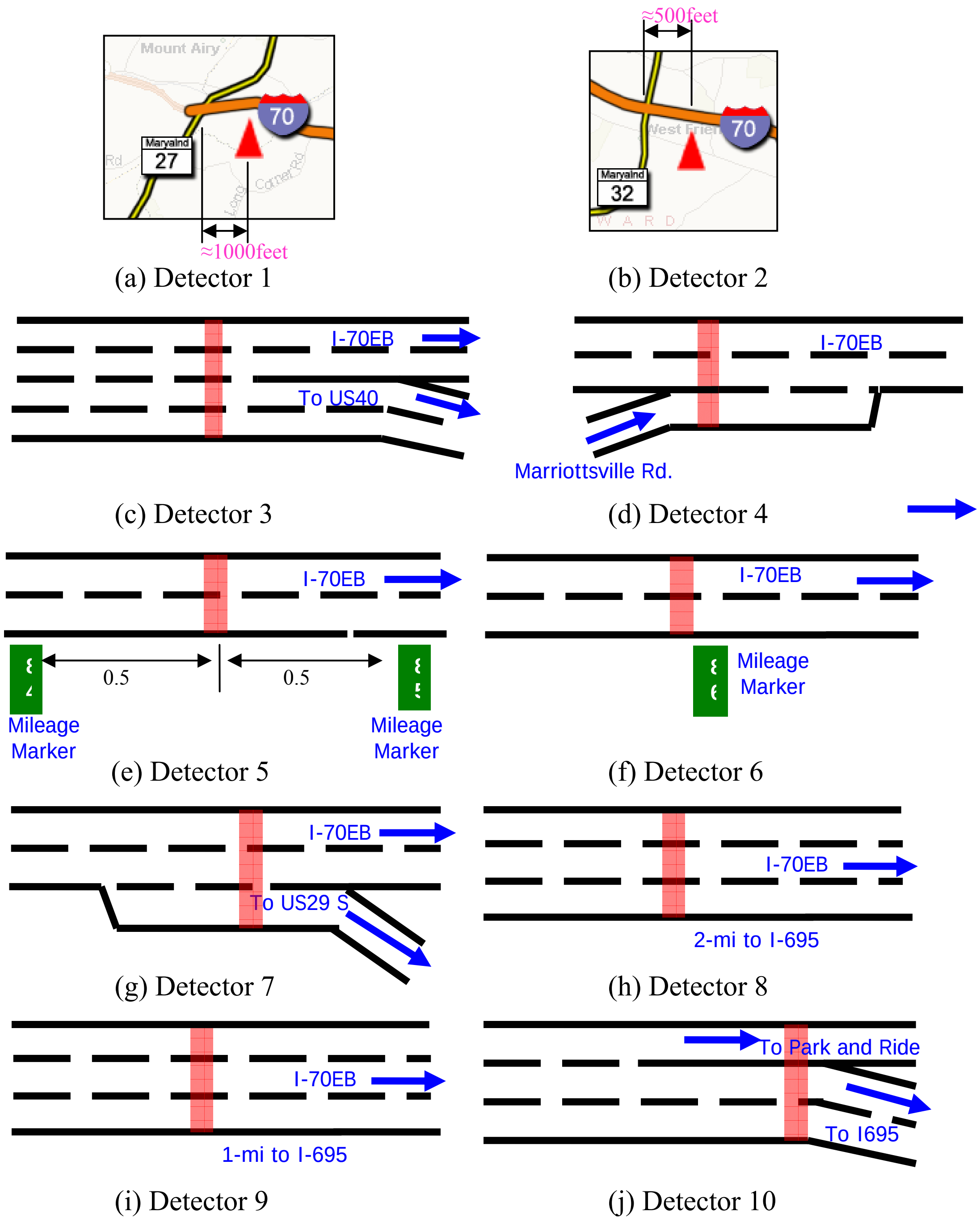


Figure 7.2 (a) to (j) detailed location of each detector

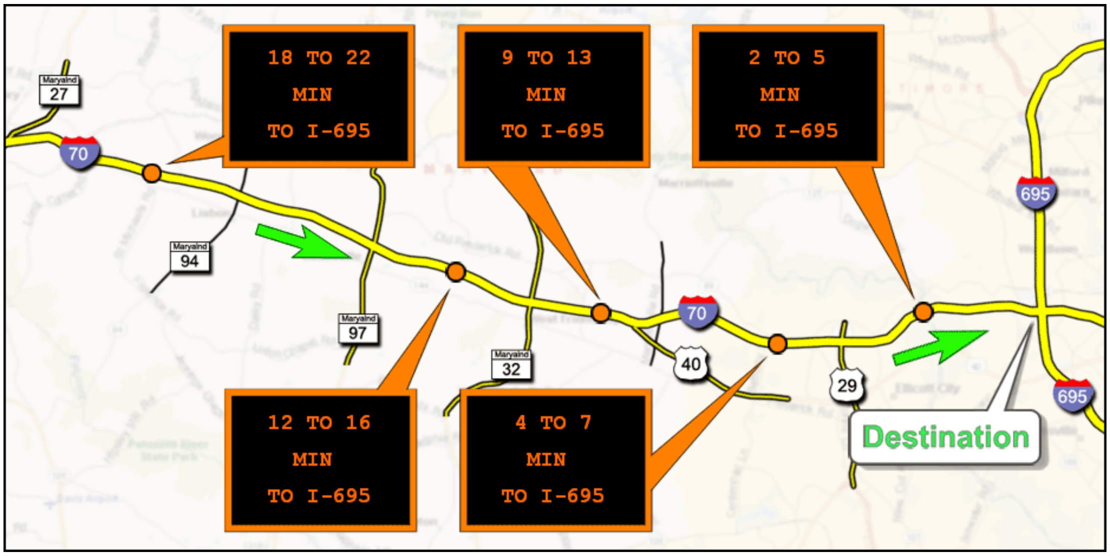


Figure 7.3 Locations of five VMS signs in ARAMPS

Table 7.2 Description and geographic locations of five VMS signs

Sign ID	Location	Distance To I-695 (Miles)	Longitude	Latitude
1	West of Attraction sign for MD 94	20.78	-77.111216	39.350566
2	West of Weigh Station	13.41	-76.981783	39.318216
3	800 feet East of U.S. 40 Sign	10.06	-76.921850	39.305333
4	East of Bethany Lane	6.17	-76.852233	39.295650
5	In Median - East of Patapsco River Bridge	2.77	-76.793550	39.305700

Table 7.3 shows the information of collected travel times on each survey day, which clearly indicate that the subsegment from Sign 5 to I-695 was heavily congested in the evening peak hours. Its maximum travel time in the evening peak hours during some field observation days were 4 times more than that during the morning peak hours. The observed maximum travel time in the subsegment from Sign 3 to I-695 in the morning is between 726 to 958 seconds, compared to 1,288 seconds in the evening peak hours. The high variation of traffic conditions in the

target freeway segment made the travel time prediction a very challenging task.

Table 7.3 Field-measured travel times on each survey day by segment

Date and Time		Sample Size	Travel Time (sec)					
			Sign 3 to I-695 (10.1 mi)		Sign 4 to I-695 (6.1 mi)		Sign 5 to I-695 (2.8 mi)	
			Max.	Avg.	Max.	Avg.	Max.	Avg.
5/15/2006	AM	7	958	740	465	400	230	169
5/15/2006	PM	8	1161	843	996	625	821	462
5/16/2006	AM	14	768	615	370	330	180	135
5/16/2006	PM	14	1088	712	882	514	705	335
5/17/2006	AM	14	907	711	455	362	240	156
5/17/2006	PM	6	934	823	751	609	564	433
5/18/2006	AM	14	785	614	363	326	170	130
5/18/2006	PM	10	1288	1138	1068	927	880	739
5/19/2006	AM	14	765	620	442	355	219	153
5/19/2006	PM	7	906	813	689	566	493	399
7/13/2006	PM	14	764	590	558	385	379	204
7/18/2006	PM	14	883	618	659	399	476	221
7/25/2006	AM	14	726	616	406	346	198	146
7/25/2006	PM	14	782	597	578	379	397	198
7/26/2006	AM	14	827	618	561	361	338	157
7/26/2006	PM	14	1173	826	948	604	756	422
Total		192						

As requested by the Maryland State Highway Administration, the field demonstration system displays the predicted travel time in a range of ±1.5 minutes around the predicted value, if it is less than 7 minutes, in a range of ±2 minutes for all other conditions. (see Figure 7.4).

Hence, the performance evaluation for ARAMPS was based on its displayed travel time ranges and the travel times collected from field trips for the aforementioned three subsegments, respectively. As shown in Table 7.4, ARAMPS seemed to achieve 97.9% prediction accuracy for the travel time for Sign 3 to I-695, 96.9% prediction accuracy for Sign 4 to I-695, and 92.7% prediction accuracy for Sign 5 to I-695. The system only failed to provide reliable travel times for 4, 6 and 14 out of 192 cases of these three subsegments, respectively. As seen in Table 7.4, most unsuccessful predictions in the subsegment from Sign 3 to I-695 incurred on the day having long average travel times.

Figure 7.4 Display of the predicted travel time range

Overall, ARAMPS system was able to reliably predict the travel times for the target freeway segment under light traffic and moderately congested conditions. More detectors may be needed to place on those heavily congested segments to increase the prediction accuracy.

Table 7.4 Performance evaluation of the developed travel time prediction system

	Sign 3 to I-695			Sign 4 to I-695			Sign 5 to I-695		
Time	Avg. T.T.	# of Correct Cases	# of All Cases	Avg. T.T.	# of Correct Cases	# of All Cases	Avg. T.T.	# of Correct Cases	# of All Cases
05/15 AM	740	7	7	400	7	7	169	7	7
05/15 PM	843	7	8	625	7	8	462	7	8
05/16 AM	615	14	14	330	14	14	135	14	14
05/16 PM	712	14	14	514	13	14	335	13	14
05/17 AM	711	14	14	362	14	14	156	14	14
05/17 PM	823	5	6	609	6	6	433	5	6
05/18 AM	614	14	14	326	14	14	130	14	14
05/18 PM	1138	9	10	927	9	10	739	9	10
05/19 AM	620	14	14	355	14	14	153	14	14
05/19 PM	813	7	7	566	7	7	399	7	7
07/13 PM	590	14	14	385	14	14	204	14	14
07/18 PM	618	14	14	399	14	14	221	14	14
07/25 AM	616	14	14	346	14	14	146	11	14
07/25 PM	597	14	14	379	14	14	198	13	14
07/26 AM	618	14	14	361	14	14	157	11	14
07/26 PM	826	13	14	604	11	14	422	11	14
Total		188	192		186	192		178	192

7.3 Conclusions

This chapter presents the step-by-step description of procedures for developing and implementing a real-time travel time prediction system for freeway segments, including the performance evaluation results.

The implementation includes three stages: system preparation stage, model-training stage, and real-time operation stage. The system preparation stage usually takes about one month for system designers to determine detection locations, based on geometric features and daily traffic patterns in each segment. It is recommended that one should collect four-month data to build the historical travel time database for developing the prediction models. The system has the capability to concurrently update its parameters in different models based on the most recent traffic data collected during the real-time operations.

Overall with 10 detectors over the segment of 25 miles, ARAMPS system has demonstrated its capability in reliably predicting the travel times within an acceptable range. Most other travel time prediction systems would need about 50 detectors to achieve the same level of performance. The performance of ARAMPS was found to be satisfactory under light or moderate traffic conditions. The predicted range of travel times was found to be correct in 188, 186 and 178 out of the 192 actual travel time samples collected by Maryland State Highway Administration over the subsegment from Signs 3, 4 and 5 to I-695, respectively.

Chapter 8: Conclusions and Recommendations

8.1 Research Summary

This chapter presents all research tasks that have been completed, followed by a summary of valuable lessons learned from the development, implementation and operation of ARAMPS. This study has accomplished the following major tasks:

Performed an in-depth review of literature associated with travel time prediction:

The task of literature review has covered state of the art and practice on the following topics: travel time estimation, travel time prediction, existing simulated and real-world application systems. The review showed that most existing studies for travel time estimation and prediction are for short links with densely distributed detectors. Nonparametric models, such as *k*-Nearest Neighbor Model, are reported to achieve better performance than the parametric models. Previous studies also indicated that a proper combination of different models may improve the system's reliability. However, due to the complex nature of prediction, all the existing systems implemented in practice use the estimated travel times, based on the current traffic conditions, as the displayed travel time information.

Developed a modeling framework for constructing a travel time prediction system with sparsely distributed detectors

This task has proposed an automatic travel time prediction system for use on freeway segments with various geometric features and traffic patterns. The proposed system does not require concurrent measurements of travel times. With data from sparsely distributed detectors, it uses the travel time estimation module to continuously estimate travel times for those completed trips, and stores them in a database. Its prediction module will take the real-time input from traffic detectors and then perform the prediction with its hybrid model structure. The proposed operating architecture ensures that the travel times of most recently completed trips can be added into the database in the real-time operation, and are immediately available for the system to perform the prediction for the next time interval.

Proposed a hybrid model for estimating travel times for freeways with sparsely distributed detectors

This research has developed a hybrid travel time estimation module by combining a clustered linear regression model and an enhanced trajectory-based model. It first employs a clustering function to categorize traffic patterns observed on a long link, based on its congestion levels in critical lanes, and then applies the travel time estimation module to further calibrate a linear regression model for each cluster that has sufficient samples of field data. For the clusters without adequate field data, this study has developed an enhanced trajectory-based model as a supplemental component that will estimate the in-segment speed with the traffic propagation model or an enhanced piecewise linear speed-based model, based on the distance from the vehicle's location to detectors.

Developed a hybrid model for travel time prediction

This study has developed a hybrid model structure for the travel time prediction on freeways with sparsely distributed detectors. A *k*-Nearest Neighbor Model serves as the main model to take advantage of available historical travel times and then assess their levels of reliability. An enhanced time-varying coefficient model is used to serve as the supplemental component to complete the prediction of travel times. The developed model can take full advantage of real-time traffic data and the historical traffic information estimated with the travel time estimation module for performing the complex travel time prediction task.

Proposed a rule-based hybrid incident detection algorithm for detecting major incidents on long freeways links

This study has also developed a rule-based hybrid algorithm for detecting major incidents on freeways having long detector spacing. The developed hybrid incident detection module combines a single-variable algorithm for detecting major incidents near a detector station and a dual-variable algorithm for major incidents away from either detector. To reduce the false alarm rate due to factors such as long detector spacing, complex geometric features and various traffic patterns, the incident detection module embedded a *k*-Nearest Neighbor model taking into account both daily and weekly traffic patterns to filter potential false alarms.

Implemented the developed system and evaluated its performance

The developed automatic real-time travel time prediction system (ARAMPS) has been implemented on a 25-mile stretch of I-70 eastbound from MD27 to I-695 to demonstrate its applicability. In contrast to most real-world travel time prediction system that requires one detector at every 0.5-mile, the entire system employed only 10 roadside traffic detectors to cover the freeway segment of 25 miles. The field performance evaluation conducted by a third party has indicated that the demonstrated ARAMPS can successfully achieve more than 92 percent of accuracy in predicting travel times during both morning and evening peak periods for the target segment. The system's embedded missing data estimation module and the incident detection module have significantly increased the system's operational reliability.

8.2 Lessons Learned from Developing, Implementing and Operating a Travel Time Prediction System

To take advantages of invaluable experience obtained from the system development and the demonstration project, the research team has summarized the following critical lessons for future deployment of travel-time prediction systems.

Multi-model structure can better capture the impacts of complex geometric features and various traffic patterns on the travel time prediction

The research team found that a single-model structure cannot well capture the impacts on travel times due to the complex interactions between geometric features, recurrent traffic patterns and the sparseness of detector spacing.

Place the detectors at critical locations to maximize their effectiveness

In determining the optimal detector locations, the research team found that the prediction accuracy varies significantly with the distribution of detector locations, especially for a system with long detector spacing. One needs to carefully evaluate the potential detector errors in each link between two neighboring detectors so as to optimize the overall system performance. Usually, segments that may incur high variation of travel times need more detectors than segments with light traffic or less-congested traffic conditions.

Need to have an effective model to contend with the inevitable missing data issue during the real-time operation

Over the entire six months of system reliability test and improvement, the issue of missing data due to either communication failures or detector malfunctions emerges quite often. To avoid frequent system suspensions due to inadequate data for predicting travel times, it is critical to embed an effective module to the deployed travel time prediction system in order to contend with the missing data issues, which will in turn extend the system's operating period.

All deployed detectors require a careful calibration and validation task

Most existing detector manufacturers tend to overstate the effectiveness of their products and understate the required efforts for system calibration. For example, in the ARAMPS project, a 2-week period was scheduled for calibrating all detectors for traffic flow, occupancy, and speeds. However, it took more than 6 months for the experts from the detector manufacturer to complete the calibration for 10 RTMS detectors. It has been found that without proper calibration, RTMS detectors will yield large non-systemic errors on volume counts and speed measurements regardless of the traffic condition. The low quality traffic data provided by poorly calibrated detectors will render the travel time prediction becoming a meaningless task.

Speed data provided by most radar detectors (such as RTMS) are not reliable for use in the travel time estimation and prediction

Another important issue identified in this research is the unreliable speed data measured by the radar detectors (Zou and Wang, 2006). The accuracy of speed data detected by most existing traffic detectors is subject to a rigorous evaluation, especially during congested traffic conditions where vehicles may experience stop-and-go patterns. Unless having some substantial advance in traffic detection technologies, it is strongly recommended that the spot speed data should not be used for developing any travel time prediction system.

Need a careful test of communications between all system components to ensure the high data availability

During the system's testing period, the research team identified that the detector data provided by deployed detectors were frequently not available due to software issues associated with the short data retrieval time interval and the communication mechanism in the system provider's control server. The frequent data loss will certainly have significant impact on the reliability and credibility of a travel time prediction system.

Implement an auto shutdown mechanism for each VMS to ensure the system's credibility

To avoid displaying old information due to a system communication error, it is essential for every VMS to be equipped with a timer to shutdown its display when no new message has been received over a pre-specified interval (e.g., 10 minutes). The VMS will resume its display task when the updated message becomes available. This timer has been found to be very useful in ensuring the VMS's credibility as the cell phone data network used to transfer detector data and to display messages is not as reliable as expected.

Need an automated system monitor in real-time operations

During real-time operations, there exist many issues that may significantly reduce the prediction reliability and/or impact the system's outputs, and thus require the system operators to take the necessary actions in a timely manner. In the ARAMPS demonstration, the research team has designed a system monitor that will send alerts to system administrators via emails and cell phone messages when experiencing undesirable performance data from the following system components:

- Detectors
- VMS signs
- Databases of traffic data and historical travel times
- System servers, including both data servers and web servers

- Communication between detectors, VMS and servers
- Incident detection
- Missing and/or delayed data

Appendix A. Variable Definitions

d : Detector ID. Detectors are numbers from the upstream.
D : Total number of detectors.
l_a : Lane ID. Lane ID is numbered from right to left.
L_d : The length of the link (d, d+1).
$\tau_d(t)$: Travel time for the vehicle to traverse the link (d, d+1) with a departure time t.
$\tau(t)$: Travel time for the vehicle to traverse the entire segment with a departure time t.
$\bar{u}_d^j(t)$: The average speed of the vehicle to traverse the j^{th} half of the link (d, d+1). j=1 or 2.
$\hat{u}_d^{Thru}(t,\Delta t)$: The average speed of the through traffic at the location of detector d in a time interval Δt from time t.
$u_{d,l_a}(t,\Delta t)$: The average speed in lane l_a at detector d in a time interval Δt from time t.
$o_{d,l_a}(t,\Delta t)$: The average occupancy in lane l_a at detector d in a time interval Δt from time t.
$v_{d,l_a}(t,\Delta t)$: The cumulative traffic count in lane l_a at detector d in a time interval Δt from time t.
p: Index of the traffic scenarios.
γ_p^d : The average proportion of time taken for vehicles to traverse the first half of the link (d, d+1).
$\mathbf{CLT}_{d,d+1}^{d}(p) =$ {all critical through lanes at the upstream detector, which significantly contribute to computing the average through traffic condition in link (d, d+1) under traffic scenario p.}
$\mathbf{CLR}_{d,d+1}^{d}(p) =$ {all critical ramp lanes at the upstream detector, which significantly contribute to computing the average through traffic condition in link (d, d+1) under traffic scenario p.}
$\mathbf{CLT}_{d,d+1}^{d+1}(p) =$ {all critical through lanes at the downstream detector, which significantly contribute to computing the average through traffic condition in link (d, d+1) under traffic scenario p.}
$\mathbf{CLR}_{d,d+1}^{d+1}(p) =$ {all critical ramp lanes at the downstream detector, which significantly contribute to computing the average through traffic condition in link (d, d+1) under traffic scenario p.}
$u(x,t)$: The speed of one vehicle when it arrives at location x with a departure time t from the upstream node of a link.
$o(x,t)$: The occupancy of a very small segment at one vehicle's location x with a departure time t from the upstream node of a link.

$TC_d^{l_a}(t, t+\Delta t)$: the traffic type in lane l_a at detector d from time t to $t+\Delta t$

$o_d^{l_a}(t, t+\Delta t)$: the average occupancy in lane l_a at detector d from time t to $t+\Delta t$

$OF_d^{l_a}$ and $OC_d^{l_a}$: the upper bound of free-flow occupancy and lower bound of heavy congestion occupancy for lane l_a at detector d respectively.

mdis: the modified distance for defining the similarity in a *k*-Nearest Neighbors model.

Reference

Abdulhai, B., Porwal, H. and Recker, W. (1999) "Short-term Freeway Traffic Flow Prediction Using Genetically-optimized Time-delay-based Neural Networks," *UCB, UCB-ITS-PWP–99–1*, Institute of Transportation Studies, University of California, Berkeley.

Ahmed, M., and Cook, A., (1979) "Analysis of Freeway Traffic Time-Series Data by Using Box–Jenkins Techniques," *Transportation Research Board* 722, pp. 1–9.

Altman, N., (1992) "An Introduction to Kernel and Nearest Neighbor Nonparametric Regression," *The American Statistician*, Vol. 46(3), pp. 175-185.

Anderson, J., Bell, M., Sayers, T., Busch, F., and Heymann, G., (1994) "The Short-term Prediction of Link Travel Times in Signal Controlled Road Networks," *Transportation Systems: Theory and Application of Advanced Technology, IFAC Symposium*, Tianjin, China, pp 621.

Benedetti, J. (1977) "On the Nonparametric Estimation of Regression Functions," *Journal of the Royal Statistical Society, Series B*, Vol. 39, pp 248-253.

Box, G., and Jenkins. G., (1970) "Time Series Analysis: Forecasting and Control," California Holden-Day.

Chen, M., and Chien, S., (2001) "Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based," *Transportation Research Record* 1768, pp. 157-161.

Chen, H., Grant-Muller, S., Mussone, L. and Montgomery, F., (2001) “A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting,” *Neural Computing and Applications*, Vol. 10, pp. 277–286.

Chen, C, Skabardonis, A and Varaiya, P,, (2003) “A System for Displaying Travel Times on Changeable Message Signs”, *Working Paper*, PATH Database.

Chien, S., and Kuchipudi, C., (2003) “Dynamic Travel Time Prediction with Real-time and Historic Data”, *Journal of Transportation Engineering*, 2002, Vol. 129(6), pp. 608–616.

Chu, L., Oh., J., and Recker W., (2005) “Adaptive Kalman Filter Based Freeway Travel Time Estimation,” *84th Annual Transportation Research Board Meeting*, Washington D.C.

Chung, E., and Rosalion, N., (2001) “Short Term Traffic Flow Prediction,” *24th Australian Transportation Research Forum*, Hobart, Tasmania.

Clark, S., Dougherty, M., and Kirby, H. (1993) “The Use of Neural Networks and Time Series Models for Short-term Traffic Forecasting: a Comparative Study,” *PTRC 21st Summer Annual Meeting*.

Clark, S., (2003) “Traffic Prediction using Multivariate Nonparametric Regression,” Journal of Transportation Engineering, Vol. 129(2), pp. 161-168.

Cleveland, W., (1979) “Robust Locally Weighted Regression and Smoothing Scatterplots,” *Journal of the American Statistical Association*, Vol. 74.

Cleveland, W., and Delvin, S., (1988) “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of American Statistical Association*, Vol. 83(403), pp. 596-610.

Coifman, B., (1998) "Vehicle Re-Identification and Travel Time Measurement in Real-Time on Freeways Using Existing Loop Detector Infrastructure," *Transportation Research Record* 1643, pp 181–191

Coifman, B., (2002) "Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors," *Transportation Research: Part A*, Vol. 36(4), pp. 351-364

Coifman, B., (2003) "Identifying the Onset of Congestion Rapidly with Existing Traffic Detectors", *Transportation Research: Part A*, Vol. 37(3), pp. 277-291.

Coifman, B., and Cassidy, M., (2002) "Vehicle Re-identification and Travel Time Measurement on Congested Freeways," *Transportation Research: Part A*, Vol. 36(10), pp 899–917.

Coifman, B., and Ergueta, E., (2003) "Improved Vehicle Re-identification and Travel Time Measurement on Congested Freeways," *Journal of Transportation Engineering*, Vol. 129(5), pp 475-483.

Cortés, C., Lavanya, R., Oh, J., and Jayakrishnan, R., (2002) "A General Purpose Methodology for Link Travel Time Estimation Using Multiple Point Detection of Traffic," *Transportation Research Record* 1802, pp. 181-189.

Dailey, D., (1993) "Travel-Time Estimation using Cross-correlation Techniques," *Transportation Research: Part B*, Vol. 27(2), pp 97-107

Danech-Pajouh, M., and Aron M., (1991), "ATHENA, A Method for Short-term

Inter-urban Traffic Forecasting,” *Report 177, INRETS*, Arcueil, Paris.

Davis, G., and Nihan., N., (1991) “Nonparametric Regression and Short-Term Freeway Traffic Forecasting,” *Journal of Transportation Engineering*, Vol. 117(2), pp 178-188.

Davis, G., Niham, N., Hamed, M. and Jacobson, L., (1991) “Adaptive Forecasting of Freeway Traffic Congestion,” *Transportation Research Record* 1287, pp. 29-33.

Dougherty, M. (1995) “A Review of Neural Networks Applied to Transport”, *Transportation Research, Part C*, Vol. 3, No. 4, pp 247-260.

Fan, J., and Gibels, I., (1996), “Local Polynomial Modeling and Its Applications,” *Chapman & Hall*, New York.

Faouzi, E. (1996). “Nonparametric Traffic Flow Prediction Using Kernel Estimator,” *13th International Symposium on Transportation and Traffic Theory*, Lyon, France, pp. 24-26.

Georgia Navigator software, http://www.georgia-navigator.com/trips, Accessed October, 2006.

Hamed, M., Al-Masaeid, H., and Bani Said, Z., (1995) “Short-term Prediction of Traffic Volume in Urban Arterials,” *Journal of Transportation Engineering*, Vol. 121(3), pp. 249–254.

Hastie, T., and Loader, C., (1993) “Local Regression: Automatic Kernel Carpentry,” *Statistical Science*, Vol. 8(2), pp. 120-129.

Huisken, G. and Van Berkum, E. (2003) “A Comparative Analysis of Short-Range

Travel Time Prediction Methods," *Transportation Research Board 82nd Annual Meeting*, Washington, D.C.

Ishak, S. Al-Deek, H., "Statistical Evaluation of I-4 Traffic Prediction System", *82rd Annual Meeting of the Transportation Research Board*, Washington, D.C., 2003.

Ishak, S. and Alecsandru, C. (2004) "Optimizing Traffic Prediction Performance of Neural Networks under Various Topological, Input and Traffic Condition Settings," *Jornal of Transportation Engineering*, Vol. 130(4), pp. 452-465.

Kirby, H., Dougherty, M. and Watson, S. (1997) "Should We Use Neural Networks or Statistical Models for Short Term Motorway Forecasting," *International Journal of Forecasting*, Vol. 13, pp. 45–50.

Kuchipudi, C., and Chien, S., (2003) "Development of A Hybrid Model for Dynamic Travel-Time Prediction," *Transportation Research Record* 1855, pp. 22-31.

Kühne, R., and Immes, S., (1993) "Freeway Control Systems for Using Section-Related Traffic Variable Detection," *Proc. Pacific Rim TransTech Conference*, Vol, 1993, pp. 56–62.

Kühne, R., Palen, J., Gardner, C., and Ritchie, S., (1997) "Section-Related Measures of Traffic System Performance," *Presented at the 76th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Kwon, J., Coifman, B., and Bickel, P., (2000) "Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data", *Transportation Research Record* 1717, pp 120-129

Kwon, J. and Petty, K. (2005) “A Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes,” *Transportation Research Record* 1935, pp. 147-153

Kwon, E. and Stephanedes, Y. (1994) “Comparative Evaluation of Adaptive and Neural-Network Exit Demand Prediction For Freeway Control,” *Transportation Research Record* 1446, pp. 66–76.

Jia, Z.,Chen, C., Coifman, B., and Varaiya, P., (2001) “The PeMS Algorithms for Accurate, Real-Time Estimates of g-factors and Speeds from Single Loop Detectors,” *4th International ITSC Conference*, pp. 536–541.

Levin, M., and Tsao, Y., (1980) “On Forecasting Freeway Occupancies and Volumes,” *Transportation Research Record* 773, pp. 47–49.

Lindveld, C., Thijs, R., Bovy, P., and van ZIJPP, N., (2000), “Evaluation of Online Travel Time Estimators and Predictors,” *Transportation Research Record* 1719, pp 45-53.

Lingras, P. and Mountford, P. (2001) “Time Delay Neural Networks Designed Using Genetic Algorithms For Short-Term Inter-city Traffic Forecasting,” *IEA/AIE* 2001, LNAI 2070, pp. 290–299.

Liu, Y., Lai, X., and Chang, G., (2006) “Estimation of Freeway Travel Time Based on Sparsely Distributed Detectors,” *9th International Conference on Applications of Advanced Technology*, Chicago, USA.

MacCarley, C., (1998) "Video-Based Vehicle Signature Analysis and Tracking Phase 1: Verification of Concept Preliminary Testing," *California Partners for*

Advanced Transit and Highways (PATH). Working Papers: Paper UCB-ITS-PWP-98-10.

Müller, H., (1987) "Weak and Universal Consistency of Moving Weighted Averages," *Periodica Mathematica Hungarica*, Vol 18(3), pp. 241-250.

Nadaraya, E. (1964), "On Estimating Regression," *Theory of Probability and Its Applications*, Vol. 9, pp. 141-142.

Nam, D., and Drew, D., (1996) "Traffic Dynamics: Method for Estimating Freeway Travel Times in Real Time from Flow Measurements," *Journal of Transportation Engineering*, Vol. 122(3), pp. 185-191.

Oda, T., (1990) "An Algorithm for Prediction of Travel Time using Vehicle Sensor Data," *Third International Conference on Road Traffic Control*, Institute of Electrical Engineers, pp 40-44.

Okutani, I., and Stephanedes, Y., (1984) "Dynamic Prediction of Traffic Volume through Kalman Filtering Theory," *Transportation Research, Part B*, Vol. 18(1), pp. 1–11.

Oswald, R., Scherer, W. and Smith, B. (2001). "Traffic Flow Forecasting Using Approximate Nearest Neighbor Nonparametric Regression," *Research Project Report for ITS Implementation Research*.

Park, D., Messer, C., and Urbanik, T. (1998) "Short-Term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Networks," *Transportation Research Record* 1651, pp. 39–47.

Park, D. and Rilett, L. (1998) Forecasting Multiple-period Freeway Link Travel

Times Using Modular Neural Networks," *Transportation Research Record* 1617, pp. 163–170.

Park, D., and Rilett, L., (1999) "Forecasting Freeway Link Travel Times with a Multilayer Feedforward Neural Network," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 14(5), pp. 357.

Park, D., Rilett, L. and Han, G. (1999) "Spectral Basis Neural Networks for Real-Time Travel Time Forecasting," *Journal of Transportation Engineering*, Vol. 125(6), pp. 515–523.

Petty, K., Bickel, P., Jiang, J., Ostland, M., Rice, J., Ritov Y., and Schoenberg, F., (1993) "Accurate Estimation of Travel Times from Single-Loop Detectors," *Transportation Research: Part A*, Vol. 32(1), pp. 1–17.

Pfannerstill, E., (1984) "A Pattern Recognition System for the Re-identification of Motor Vehicles," *Proc. 7th International Conference on Pattern Recognition*, Montreal, pp. 553–555.

Priestley, M., and Chao, M., (1972), "Non-parametric Function Fitting," *Journal of the Royal Statistical Society. Series B*, Vol. 34(3), pp. 385-392.

Rice, J., and van Zwet, E., (2004) "A Simple and Effective Method for Predicting Travel Times on Freeways," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5(3), pp. 200-207.

Smith, B. and Demetsky, M. (1994) "Short-Term Traffic Flow Prediction: Neural Network Approach," *Transportation Research Board* 1453, pp. 98–104.

Smith, B., and Demetsky, M., (1996) "Multiple-interval Freeway Traffic Flow

Forecasting," *Transportation Research Record* 1554, pp. 136–141

Smith, B., and Demetsky, M., (1997) "Traffic Flow Forecasting: Comparison of Modeling Approaches," *Journal of Transportation Engineering*, Vol. 123(4), pp. 261–266.

Smith, B., Williams, B., and Oswald, K., (2000) "Parametric and Nonparametric Traffic Volume Forecasting," *80th TRB Annual Meeting*, Washington, DC.

Smith, B., Williams, B., and Oswald, K. (2002) "Comparison of Parametric and Non-parametric Models for Traffic Flow Forecasting," *Transportation Research Part C*, Vol. 10(4), pp. 303–321.

Stathopoulos, A., and Karlaftis, M.(2003) "A Multivariate State-space Approach for Urban Traffic Flow Modeling and Prediction," *Transportation Research Part C*, 11(2), pp. 121–135.

Stone, C., (1977) "Consistent Nonparametric Regression," *The Annals of Statistics*, Vol. 5, pp 595-645.

Sun, H., Liu, H., and Ran, B. (2003) "Short Term Traffic Forecasting Using the Local Linear Regression Model," *82nd Annual Meeting of the Transportation Research Board*, Washington, D.C.

Sun, H., Zhang, C., and Ran, B., (2004) "Interval Prediciton for Traffic Time series Using Local Linear Predictor," *IEEE Intelligent Transportation Systems Conference*, October 3-6, 2004, Washington, D.C.

Travel Time Messaging on Dynamic Message Signs - Chicago, IL, (May 2005), http://ops.fhwa.dot.gov/publications/travel_time_study/chicago/chicago_ttm.h tm. Accessed July 07, 2006.

Travel Time Messaging on Dynamic Message Signs - Houston, TX, (May 2005), http://ops.fhwa.dot.gov/publications/travel_time_study/houston/houston_ttm.htm, Accessed July 07, 2006.

Travel Time Messaging on Dynamic Message Signs - Nashville, TN, (May 2005), http://ops.fhwa.dot.gov/publications/travel_time_study/nashville/nashville_ttm.htm. Accessed July 07, 2006.

Travel Time Messaging on Dynamic Message Signs - Portland, OR, (May 2005), http://ops.fhwa.dot.gov/publications/travel_time_study/portland/portland_ttm.htm. Accessed July 07, 2006.

Tukey, J., (1977) "Exploratory Data Analysis," *Reading*, MA: Addison Wesley.

van der Voort, M., Dougherty, M., and Watson, S., (1996) "Combining Kohonen Maps with ARIMA Timeseries Models to Forecast Traffic Flow," *Transportation Research, Part C*, Vol. 4(5), pp. 307–318

van Grol, N., Manfredi, S., and Danech-Pajouh, M., (1997) "On-Line Network State Estimation and Short Term Prediction Deliverable D05.2.," *Telematics Application Programme Project TR1017*, INRETS, Brussels.

van Lint, J., Hoogendoorn, S. and van Zuylen, H. (2002) "Freeway Travel Time Prediction with State-Space Neural Networks: Modeling State-Space Dynamics with Recurrent Neural Networks," *Transportation Research Record* 1811, pp. 30-39.

van Lint, J., and van der Zijpp, N., (2003) "Improving a Travel-Time Estimation

Algorithem by Using Dual Loop Detectors," *Transportation Research Record*, No. 1855, pp. 41-48

Washington State Department of Transportation, "Seattle Area Traffic," http://www.wsdot.wa.gov/Traffic/seattle/questions/traveltimesdetail.htm, Accessed October, 2006.

Watson, G. (1964) "Smooth Regression Analysis," *Sankhya Series A*, Vol 26, pp. 359-372

Williams, B., Durvasula, P., and Brown, D. (1998) "Urban Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models," *Transportation Research Record* 1644, pp. 132–144.

Whittaker, J., Garside, S., and Lindeveld, K., (1997) "Tracking and Predicting Network Traffic Process," *International Journal of Forecasting*, Vol. 13, pp. 51–61.

Wu, C., Ho., J., and Lee, D., (2004) "Travel-time Prediction with Support Vector Regression," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5(4), pp. 276-281.

Yang., J., (2005) "A Time Series Approach to Arterial Travel Time Modeling and Prediction," *Proceeding of Intelligent Systems and Control*, 497-024

Yasdi, R. (1999) "Prediction of Road Traffic using a Neural Network Approach," *Neural Computation and Application*, Vol. 8, pp. 135–142.

Yin, H., Wong, S. and Xu, J., (2002) "Urban Traffic Flow Prediction Using Fuzzy-

Neural Approach," *Transportation Research Part C*, Vol. 10, pp. 85–98.

You, J., and Kim, T., "Development and Evaluation of A Hybrid Travel Time Forecasting Model," *Transportation Research: Part C*, Vol 8(1), pp. 231-256.

Yun, S., Namkoong, S., Rho, J, Shin, S. and Choi, J. (1998) "A Performance Evaluation of Neural Network Models in Traffic Volume Forecasting," *Mathematical Computer Modeling*, Vol. 27(9–11), pp. 293–310.

Zhang, H. (2000) "Recursive Prediction of Traffic Conditions with Neural Networks," *Journal of Transportation Engineering*, Vol. 126(6), pp. 472–481.

Zhang, X., and Rice, J., (2003) "Short-Term Travel Time Prediction Using a Time-Varying Coefficient Linear Model", *Transportation Research: Part C*, Vol. 11, pp. 187-210.

Zheng, W., Lee, D. and Shi, Q., (2006) "Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach", *Jornal of Transportation Engineering*, Vol. 132(2), pp. 114-121.

Zou, N. and Wang, J. (2006) "A Video-based Method for Evaluating Traffic Data from Detectors" *9th International Conference on Applications of Advanced Technology*, Chicago, USA.